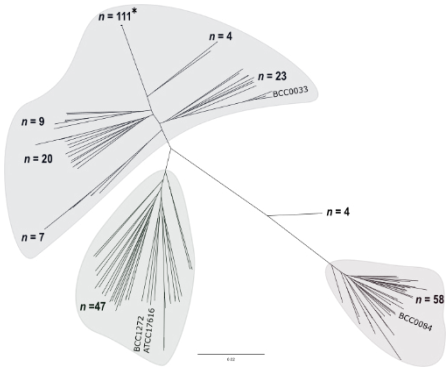
Using next generation sequencing approaches to define the population biology of the neglected cystic fibrosis lung pathogen *Burkholderia multivorans*



Thesis presented by

Kasia Maria Parfitt
BSc (Hons) Microbiology (SW),
MSc Microbiology and Infection

In candidature for the degree of Doctor of Philosophy

Microbiomes, Microbes, and Informatics Group,
School of Biosciences,
Cardiff University

October 2022

“Believe you can and you’re halfway there”

- Theodore Roosevelt

Acknowledgements

I would like to deeply thank the following people:

- My primary PhD supervisor at Cardiff University, Professor Eshwar Mahenthiralingam for all his exceptional knowledge, encouragement, and support throughout my PhD. It has been a pleasure to work under his supervision, also being part of an incredible laboratory team. Words cannot express the experience I have gained throughout this journey, shaping my future scientific career. I guarantee would not be where I am today if I did not have Esh as my primary supervisor.

- My secondary Cardiff University PhD supervisor, Professor Tom Connor, for his advice on bioinformatics and the Cloud Infrastructure for Medical Microbial Bioinformatics (CLIMB)

- My PhD internal assessor at Cardiff University, Professor Pete Kille, for his enthusiasm and advice.

- Dr. Dan Neill and Dr. Angharad Green at Liverpool University for their invaluable knowledge on inhalation models, general assistance, and brilliant hosting during a week-long collaborative placement in the Infection and Global Health laboratories. Thank you for making me feel so welcome and part of the team.

- Dr. Angela Marchbank in the Cardiff University Genome Hub for her help in processing my short-read Illumina samples and performing Tape Station/QC analysis used in this PhD thesis.

- Dr. Paul Williams from University of Nottingham for supplying the *E. coli* bioluminescence reporter strains used in Sections 5.2.7 and 5.2.9.5 of this thesis.

- Dr. Sarah Christofides at Cardiff University for her incredible knowledge on statistics and help solving some of my queries, as well as being an incredible friend. Thanks for all the 'pick me up' coffees, picnics, phone calls and hugs. You were literally my shoulder to cry on when things were tough, and I cannot be more grateful.

- The GW4 BioMed Medical Research Council (MRC) Doctoral Training Partnership as my collaborative partnership PhD funding body.

- The MMI group at Cardiff University, particularly the Esh Lab, who have made this the best experience. I couldn't have imagined spending the last 3.5 years with anyone else – you guys made it for me.

- A special shout out to my PhD microbiology buddies: Yoana Petrova, Aimee Bettridge and Teresa Paradell-Gil. You guys have been my rock throughout the PhD and the best lab friends I could have asked for. Thank you for giving me endless laughs as well as unconditional support. I have enjoyed our adventures in and out of the lab and cannot wait for what the future holds for us all.

- To Rema and Alice for making the first year of my PhD a bundle of laughs whilst you completed your Master's degrees. I am so glad we kept in touch, and I value every moment of our friendship.

- My family for putting up with all the memories, laughter, crying and science talk. I wouldn't have been able to do this without you. You know I'm not a quitter, and you made sure I remembered that every single day. All the phone calls and all the texts, they pushed me. You knew I could do this and believed in me, even when I was struggling to believe in myself. Thanks for helping me accomplish one of my dreams.

- Vašek Špáňa (aka Vance), for being the most level-headed, helpful, and incredible person I know. There's no one else I would rather wake up at 1am to check my gel photos, to get excited about my work with or just to need a hug from. You didn't let me give up, no matter how hard things were. You kept me sane and provided so much more support than you realize.

- My best friends for all your friendship and love through the roller coaster ride. You've put up with so much of my science chat and worries as well as providing 'down time' and breaks from the research lifestyle. You guys know who you are, and I love you all.

- A dedication to my friend Amy Cawthrow, who suffers with cystic fibrosis. Your courage and determination are forever admirable.

Scientific Conferences and Awards

Poster Presentation: Using next generation approaches to define epidemiology and develop therapies for the neglected cystic fibrosis lung bacterium *Burkholderia multivorans.* OnE Away Day, Student led conference, May 2019, Cardiff, Wales.

Oral Presentation: Using next generation approaches to define epidemiology and develop therapies for the neglected cystic fibrosis lung pathogen *Burkholderia multivorans.* 3 Minute Thesis, GW4 BioMed MRC Doctoral Training Event, October 2019, Cardiff, Wales.

Poster Presentation: Using next generation approaches to define epidemiology and develop therapies for the neglected cystic fibrosis lung pathogen *Burkholderia multivorans*. CF Microbiology Consortium Meeting, November 2019, Liverpool, UK.

Poster Presentation: Does the two-lineage evolutionary split of *Burkholderia multivorans* have an impact on its ability to cause cystic fibrosis lung infection? Sfam ECS Symposium, March 2021, Online.

Oral Presentation: Next generation approaches identified an evolutionary split in the population of cystic fibrosis pathogen *Burkholderia multivorans*: are there any impacts on its ability to cause lung infections? Microbiology Society Annual Conference, April 2021, Online.

Oral Presentation: Next generation approaches identify an evolutionary split in the population of cystic fibrosis pathogen *Burkholderia multivorans*: are there any impacts on its ability to cause lung infections? OnE Away Day, Student led conference, May 2021, Cardiff, Wales (online).

Poster Presentation: Identification of an evolutionary split in *Burkholderia multivorans* using phylogenomics: does it have any impacts on the ability to cause CF lung infection? European Cystic Fibrosis Society (ECFS) Annual Conference, June 2021, Online.

Oral Presentation: Next generation approaches identify an evolutionary split in the population of cystic fibrosis pathogen *Burkholderia multivorans*: are there any impacts on its ability to cause lung infections? Bitesize Science Flash Talks organized by University of Leeds, April 2022, Online.

Summary

Burkholderia multivorans is the most frequently isolated *Burkholderia cepacia* complex species recovered from cystic fibrosis lung infection. However, its pathogenesis and species population biology remain elusive. Understanding adaptational factors of *B. multivorans* to the CF lung microenvironment is important for predicting its pathogenesis and disease outcome.

B. multivorans population biology was explored using pan genome analysis, average nucleotide identity and phylogenomic analysis (n = 283). The population split into two major genomic lineages, designated 1 and 2, and four *B. multivorans* model strains were selected to represent them: the soil strain ATCC 17616 (lineage 2a), BCC1272 (lineage 2a), BCC0033 (lineage 2b), and BCC0084 (lineage 1). The latter 3 CF strains were completely genome sequenced to add to the readily available reference genome ATCC 17616. Using gene presence-absence analysis, unique *B. multivorans* lineage-specific genes were identified. This enabled diagnostic PCR design with genes *ghrB_1* and *glnM_2* selected as the lineage 1 and lineage 2 targets, respectively. The PCRs showed 100% lineage-specificity against 48 *B. multivorans* strains.

Phenotypic analysis was performed on a subset of 49 *B. multivorans* strains evaluating their morphology, growth kinetics, motility, biofilm formation, and exopolysaccharide production. The *B. multivorans* phenotype was variable amongst the strains, with no link to genomic lineage. Phenotypic comparison was also performed when *B. multivorans* were mixed with a secondary CF pathogen. The suppression of *P. aeruginosa* LESB58 protease production, when cultured with *B. multivorans,* was identified as an interesting interaction based on an unknown mechanism. Three of the *B. multivorans* model strains (BCC0033, BCC0084, and ATCC 17616) were also evaluated in a murine respiratory infection model and all showed good persistence over 5-days.

Overall, this work has built a foundation of knowledge on the *B. multivorans* phenotype and genotype, enabling associations between lineage, therapeutics testing, and clinical outcome to be studied.

Contents

List of Figures

List of Tables

%	Percent
µg	Microgram
µl	Microlite
AMR	Antimicrobial resistance
ANI	Average Nucleotide Identity
ANIb	ANI (BLAST)
ANIm	ANI (Mummer)
AntiSMASH	Antibiotics and Secondary Metabolite Analysis Shell
ARG-ANNOT	Antibiotic resistance gene-ANNOTation
ATCC	American Type Culture Collection
BCC	*Burkholderia cepacia* complex
BCSA	*Burkholderia cepacian* Selective Agar
BLAST	Basic Local Alignment Search Tool
BLASTn	Nucleotide BLAST
bp	Base pairs
BSM-G	Basal salts media with glycerol supplement (4% w/v)
CARD	Comprehensive Antibiotic Resistance Database
CARD	The comprehensive antibiotic resistance database
CC	Clonal Complex
cci	*cenocepacia* island
c-di-GMP	Cyclic dimeric guanosine monophosphate
CDS	Coding sequences
CF	cystic fibrosis
CFU	Colony forming units
CFU /g	Colony forming units per gram
CFU /ml	Colony forming units per millilitre
CGD	Chronic granulomas disease
CGView	Circular Genome Viewer
CLIMB	Cloud Infrastructure for Microbial Bioinformatics
CLIN	Clinical
COG	Cluster of Orthologous Group
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DDH	DNA-DNA hybridization
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
EggNOG	Evolutionary Genealogy of Genes: Non-supervised Orthologous Groups
ENA	Europeans Nucleotide Archive
ENV	Environmental
ENVH	Environmental (hospital)
EPS	Exopolysaccharide
EU	European Union
GC	Guanine-cytosine
GFP	Green Fluorescent Protein
GI	Genomic Island
GTR	General Time Reversible
Gubbins	Genealogies Unbiased By recomBinations In Nucleotide Sequences
HCP	High copy number plasmid
HGT	Horizontal Gene Transfer

HGT	Horizontal Gene Transfer
Kan	Kanamycin
KASS	KEGG Automatic Annotation Server
Kb	Kilo base pairs
KEGG	Kyoto Encyclopedia of Genes and Genomes
LB	Luria-Bertain
LCP	Low copy number plasmid
LMG	Belgiuan Co-ordinated Collections of Micro-organisms, Gent
Mb	Mega base pairs
MH Agar	Mueller-Hinton Agar
MH Broth	Mueller-Hinton Broth
MIC	Minimum inhibitory concentration
MLST	Multi Locus Sequence Type
MNP	Multiple Nucleotide Polymorphism
NCBI	National Centre of Biotechnology and Information
NCTC	National Collection of Type Cultures
NLV	N Locus Variant
nm	Nanometer
OD	Optical Density
PacBio	Pacific Biosciences
PAI	Pathogenicity Island
PB1	Polymyxin B
PBS	Phosphate Buffer Solution
PCR	Polymerase Chain Reaction
PHASTER	PHAge Search Tool Enhaned Release
pubMLST	Public databases for MLST
QC	Quality Control
QS	Quorum Sensing
QUAST	Quality Assessment Tool for Genome Assemblies
R	R Statistical Software
RAPD	Random Amplification of Polymorphic DNA
RAST	Rapid Annotation Subsystems Technology
RAxML	Randomized Accelerated Maximum Likelihood
RISA	Ribsosomal Intergenic Spacer Analysis
rMLST	Ribosomal MLST
RNA	Ribonucleic acid
RNA-Seq	RNA Sequencing
RND	Resistance Nodulating Division
rpm	Revolutions per minute
rRNA	Ribosomal RNA
SD	Standard deviation
SEM	Standard error mean
SMRT	Single Molecule Real-Time
SNP	Single Nucleotide Polymorphism
ST	Sequence Type
T3SS	Type 3 Secretion System
T4SS	Type 4 Secretion System
T6SS	Type 6 Secretion System

Taq	*Thermus aquaticus*
Tet	Tetracycline
TSA	Tryptone Soya Agar
TSB	Tryptone Soya Broth
TVC	Total Viable Count
UK	United Kingdom
USA	United States of America
UV	Ultraviolet
V	Voltz
VFDB	Virulence factor database
wgMLST	Whole Genome MLST
WGS	Whole Genome Sequencing
WHO	World Health Organization
WT	Wild type
x *g*	Times (x) gravity
YEM	Yeast Extract Mannitol

1 Introduction

1.1 An overview of cystic fibrosis

Cystic fibrosis (CF) was first recognized as a disease after its separation from coeliac disease in 1938 (Andersen 1938). CF is a recessive genetic condition, which mostly affects the Caucasian population, causing early mortality (Davis 2006). In the 1960s, individuals with CF were unlikely to live past the age of 5-years-old (FitzSimmons 1993). As disease management and therapeutics have advanced, the expected life expectancy of people with CF has also increased, particularly in recent decades. The national CF registry reports show median life expectancy of individuals with CF is now around 50-years of age or more (UK-Cystic-Fibrosis-Registry 2021; Cystic-Fibrosis-Canada 2022).

There are newborn screening programs in Europe which can identify CF at birth (Southern *et al.* 2007). However, not all patient registries are consistent, meaning that exact prevalence of CF in Europe may not be completely known (Farrell 2008). Centralized databases, such as the UK CF registry (supported and maintained as part of the CF Trust), reports the number of registered CF patients and relative statistics (UK-Cystic-Fibrosis-Registry 2021) in an attempt to capture more accurate numbers of affected individuals. Most CF cases are in Europe, North America, and Australia. However, its incidence is variable amongst these continents. For example, Ireland has the highest European prevalence with 1/1353 births affected (Farrell 2008). In the United Kingdom (UK), there are more than 10,500 CF-diagnosed individuals (UK-Cystic-Fibrosis-Registry 2021). There have also been very high CF incident rates in the Ohio Amish population (affecting 1/569 births) (Klinger 1983) and in Sanguenay-Lac-Saint-Jean, Qubec (affecting 1/902 births) (Daigneault *et al.* 1991).

CF is caused by a mutation in the chloride-conducting transmembrane conductance regulator (*CTFR*) alleles, present on chromosome seven (Newmark 1985). The CFTR protein regulates water, chloride, and bicarbonate translocation (Kim and Skach 2012; Sabharwal 2016). Lack of this regulation causes a buildup of thick, sticky mucus in the lungs and other organs (Li and Somerset 2014), opportunistic for CF-related pathogens to colonize and thrive (Elborn 2016). Whilst the *CFTR* mutation affects multiple organs in the body, it is the profound effect it has on the lungs which drives morbidity and mortality (Mall and Hartl 2014). This is attributed to lung tissue deterioration, lung function decline (Fauroux *et al.* 2004) and overall respiratory failure in 80-95% of CF patients (Lyczak *et al.* 2002). Infections with a range of bacterial and fungal pathogens is a key driver of lung disease in CF.

It is known that pathogen infection in CF patients can either be dominated by a singular CF pathogen or by polymicrobial communities known as the CF microbiota. These communities affect the CF airways in a chronic nature and cause a cycle of respiratory exacerbation which impacts the overall health of the individual (Jean-Pierre *et al.* 2021). There are multiple pathogens which can infect the CF airway. *Staphylococcus aureus* and *Haemophilus influenzae* are predominantly found in children with

CF. Chronic infection by formidable pathogens such as *Pseudomonas aeruginosa* and members of the *Burkholderia cepacia* complex (BCC) (Razvi *et al.* 2009; Price *et al.* 2013) may then occur later in life. Polymicrobial communities are not just limited to aerobic bacteria, they also include a host of additional microbes including anaerobes (Rogers *et al.* 2010). Understanding of polymicrobial interplay between the CF microbiota remains unclear. Interactions between the pathogens and their links to clinical outcome is complicated (Jean-Pierre *et al.* 2021). Antimicrobial resistance (Harriott and Noverr 2009; Radlinski *et al.* 2017; Vandeplassche *et al.* 2019) and increased virulence (DeLeon *et al.* 2014) have been attributed to polymicrobial interactions in animal infection models. Polymicrobial communities are highly patient-specific, supporting the requirement for personalized therapy.

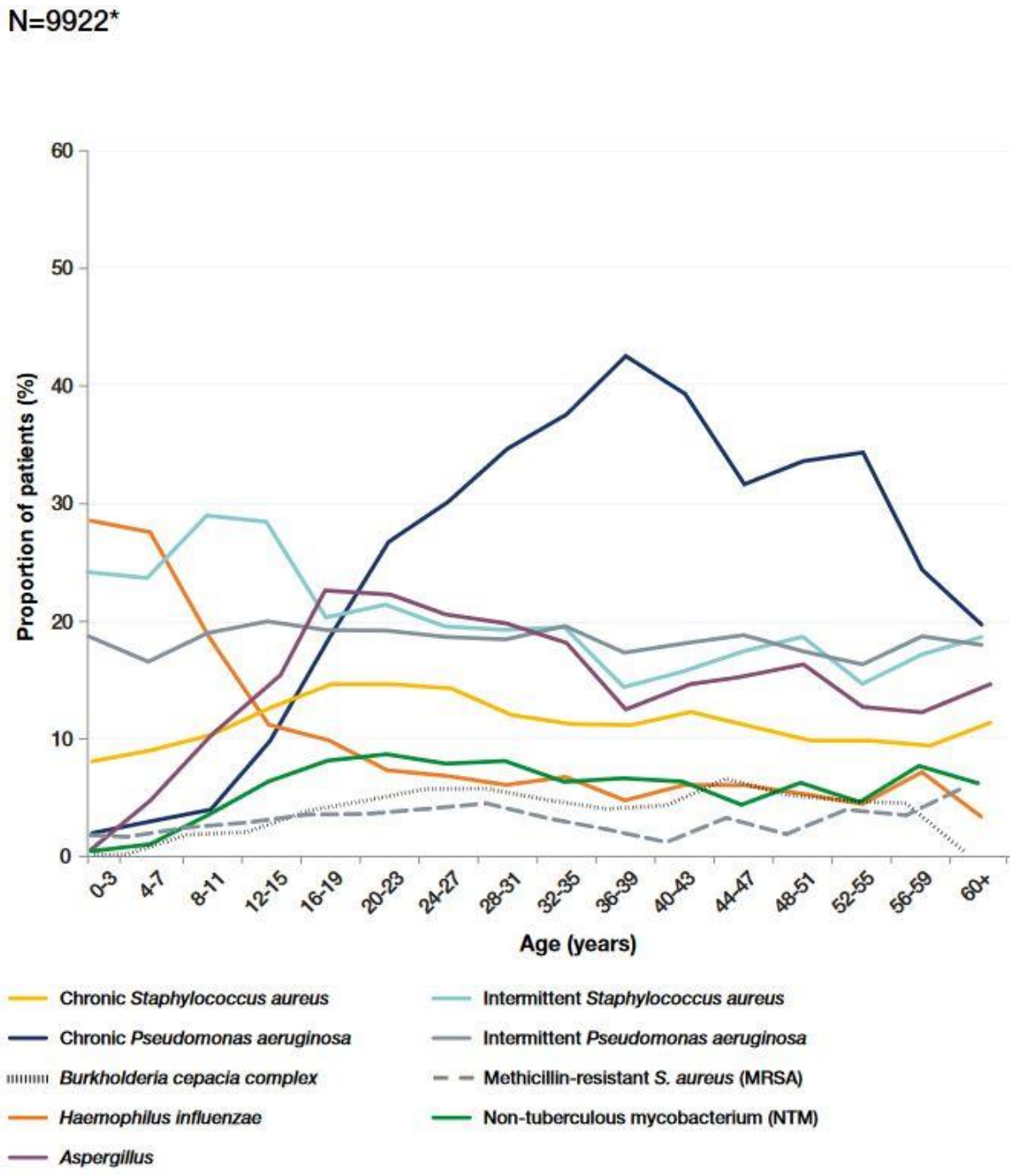


Figure 1 | Microorganisms found in CF lung infections during 2020. The percentage of patients with each microorganism is grouped by age range (years). The sample is from 9,992 isolated bacteria. Figure is from the UK CF patient registry (UK-Cystic-Fibrosis-Registry 2021).

1.2 An introduction to the Burkholderia genus

In 1949, Walter Burkholder discovered *Pseudomonas cepacia*, the causative bacterial agent of onion rot, which was described as a human pathogen a year later (Burkholder 1950). *P. cepacia* was originally rejected under the Approved List of Bacterial Names (Skerman *et al.* 1989) and thus disappeared from bacterial nomenclature. Palleroni and Holmes (1981) rekindled the bacterial name and evaluating its

validity. This was done by comparing *P. cepacia* to *Pseudomonas* ribonucleic acid homology group II (Palleroni and Holmes 1981). Reclassification of *P. cepacia* into the new *Burkholderia* genus was then proposed by Yabuuchi *et al.* (1992). Justification of this genomic reclassification was proven using 16S rRNA gene sequencing (Böttger 1989), DNA-DNA homology (Brenner *et al.* 1969), fatty acid and lipid concentrations, and phenotype (Yabuuchi *et al.* 1992).

Burkholderia species are Gram-negative, non-fermenting rods which group under the class Betaproteobacteria (Coenye *et al.* 2001b). In 1992, the *Burkholderia* genus was comprised of 7 species: *B. cepacia*, *B. gladioli, B. solnacearum, B. mallei, B. pseudomallei*, *B. pickettii,* and *B. caryophylli* (Yabuuchi *et al.* 1992). All 7 of these species corresponded to *Pseudomonas* rRNA group II. However, *B. pickettii* and *B. solnacearum* were later reclassified into the novel genus, *Ralstonia* (Yabuuchi *et al.* 1995). More recently, there have been a split of the *Burkholderia* genus into two taxonomic groupings: one remaining as *Burkholderia* species and one being reclassified as *Paraburkholderia* species (Sawana *et al.* 2014). This split was based on phylogenetic relationships and comparative genomic analyses but was also aligned with overall clinical relevance of each taxonomic grouping. The *Burkholderia* genus holds more animal and human pathogen species, whereas the *Paraburkholderia* genus harbours more environmental species associated with beneficial function (Sawana *et al.* 2014). A further 11 species of the *Burkholderia* genus have also been proposed for transfer to novel genus *Caballeronia* (Dobritsa and Samadpour 2016). To date, there are over 100 named *Burkholderia* species in the approved nomenclature (Skerman *et al.* 1989), with 24 of these species comprising the *Burkholderia cepacia* complex (BCC) group. Jin *et al.* (2020) showed further classification of the BCC into clusters through drawing of a maximum likelihood tree (Figure 2).

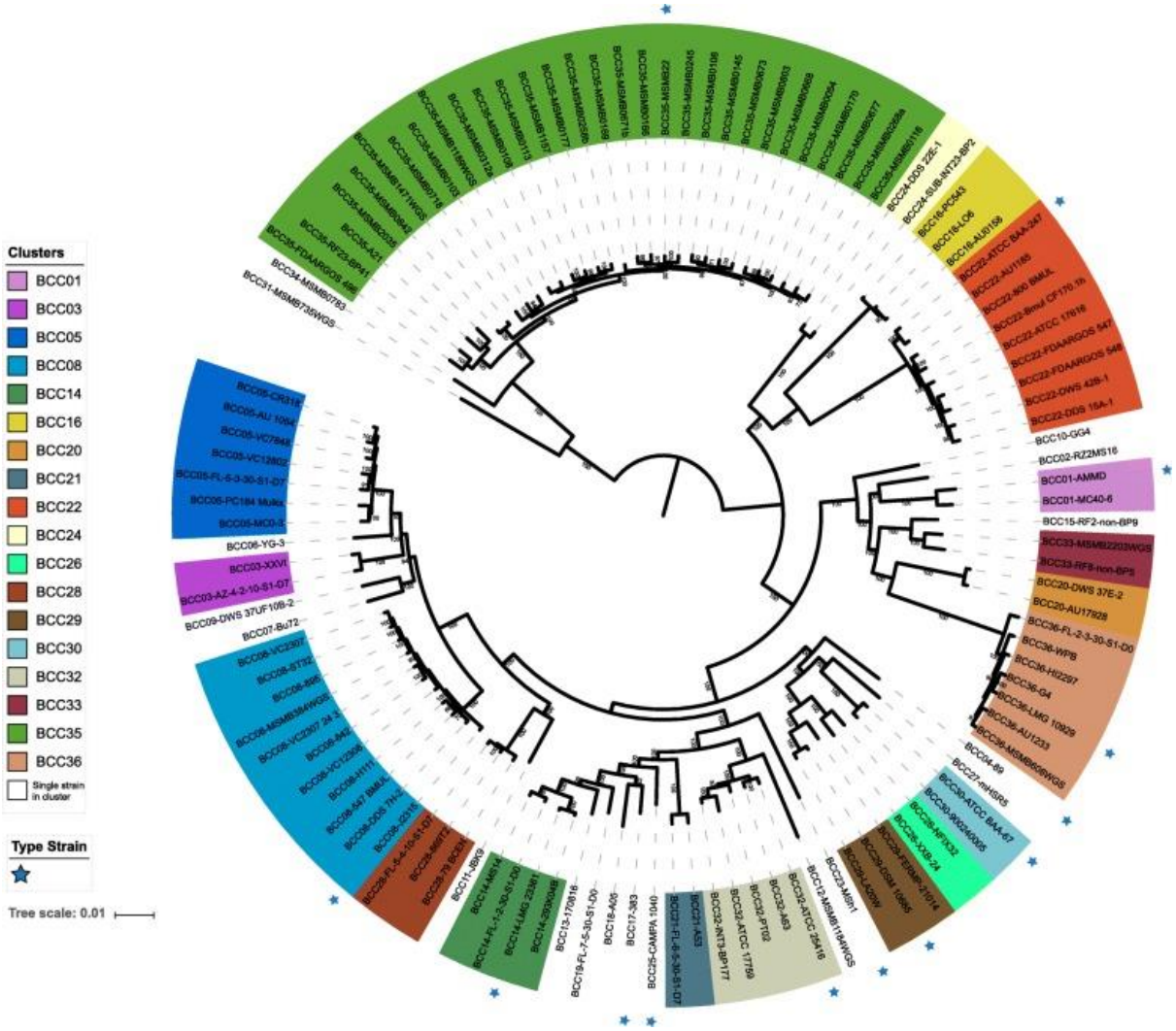


Figure 2 | The re-classification of BCC genomes based on a maximum-likelihood phylogeny (*n* = 116). Phylogenetic tree drawn by Jin *et al.* (2020). Type strains are marked with a blue star on the outer edge of the circle. The figure represents 116 BCC genomes aligned using 1,005 single-copy gene orthologs. Groups associated with clustering are shown by coloured backgrounds.

Accurate identification helps with underpinning both epidemiology and CF lung infections. Due to the high similarity of the BCC species, phenotypic identification is not enough (Devanga Ragupathi and Veeraraghavan 2019; Furlan *et al.* 2019), and overall identification has proven difficult (Jin *et al.* 2020). *B. multivorans,* the species studied in this PhD, can be differentiated from other members of the BCC through a combination of molecular and phenotypic methods such as sucrose and adonitol acidification, bacterial growth at 42°C, and API 20 NE biochemical testing (Henry *et al.* 2001).

Both the 16S rRNA gene has been used for bacterial classification (Eisen 1995) and recA gene (Karlin *et al.* 1995) was found to be a useful marker for BCC identification (Mahenthiralingam *et al.* 2000). With expanding diversity of BCC taxa, the reliability of these single gene markers is not sufficient for

accurate identification (Jin *et al.* 2020). In 2005, expansion to a 7-gene multi-locus sequence typing (MLST) system proved to be a more accurate differentiation tool for the BCC (Baldwin *et al.* 2005), but recent genomic analysis showed there are still gaps in characterization of certain BCC sequence types (STs) (Furlan *et al.* 2019). Identification methods are still developing, but with more advanced next-generation sequencing (NGS) technologies, genomic taxonomy has become more routine. There has been a large increase in the availability of published BCC complete genomes (Sharma *et al.* 2014; Salloum *et al.* 2018; García-Romero and Valvano 2020). By combining this with whole genome sequencing (WGS), the capability to define BCC evolutionary relationships is enhanced (Jin *et al.* 2020).

Interestingly, a key similarity of the BCC genomes is their multipartite replicon structure, comprising 3 primary replicons: C1, C2, and C3 (Mahenthiralingam *et al.* 2005). The *B. multivorans* ATCC 17616 strain was investigated by Cheng and Lessie (1994), identifying its multi-replicon structure in the 1990s. The replicons varied in size at 3.4 Mb, 2.5 Mb, and 0.9 Mb for C1, C2, and C3, respectively (Cheng and Lessie 1994). All 3 replicons were found to hold both insertion sequence (IS) elements and rRNA (Cheng and Lessie 1994). As well as the 3 main BCC replicons, various plasmids are also found within the genus (Mahenthiralingam *et al.* 2005). The C3 replicon of the BCC has previously been described as a conserved, nonessential virulence mega-plasmid (Agnoli *et al.* 2012). It has been thought that the C3 replicon is important in adaptation and survival to varying niches (Yoder-Himes *et al.* 2009; Nishiyama *et al.* 2010). The plasmid may also assist the bacterium in the tolerance to various environmental stresses, providing an overall fitness benefit (Agnoli *et al.* 2014).

1.3 Burkholderia multivorans and its association with cystic fibrosis

The species name *Burkholderia multivorans* (previously known as *B. cepacia* genomovar II) was first proposed by Vandamme *et al.* (1997). *B. multivorans* is now the most common CF-associated BCC infection in the USA (Lipuma 2010), UK (Kenna *et al.* 2017), and arguably worldwide, overtaking *B. cenocepacia* in epidemiological prevalence from the late 1990s onward (Zlosnik *et al.* 2015). In the US during 2010, disease metrics showed *B. multivorans* infected 3 times the number of CF individuals than *B. cenocepacia* does (Figure 3) (Lipuma 2010). In the UK, similar epidemiological numbers are observed, with 162 of 316 (51.3%) of the BCC bacteria isolated from CF infections were *B. multivorans* (Kenna *et al.* 2017). *B. cenocepacia* infections have been reduced through strict and effective control measures, reducing spread between individuals (Mahenthiralingam *et al.* 2001; Speert *et al.* 2002; France *et al.* 2008). However, environmental sources of *B. multivorans* may still contribute to ongoing infection of CF individuals (Turton *et al.* 2003; Baldwin *et al.* 2008; Ramsay *et al.* 2013). The exact niche of *B. multivorans* remains uncertain, with studies reporting sporadic isolation from the natural environment (Lipuma 2010; Depoorter *et al.* 2016).

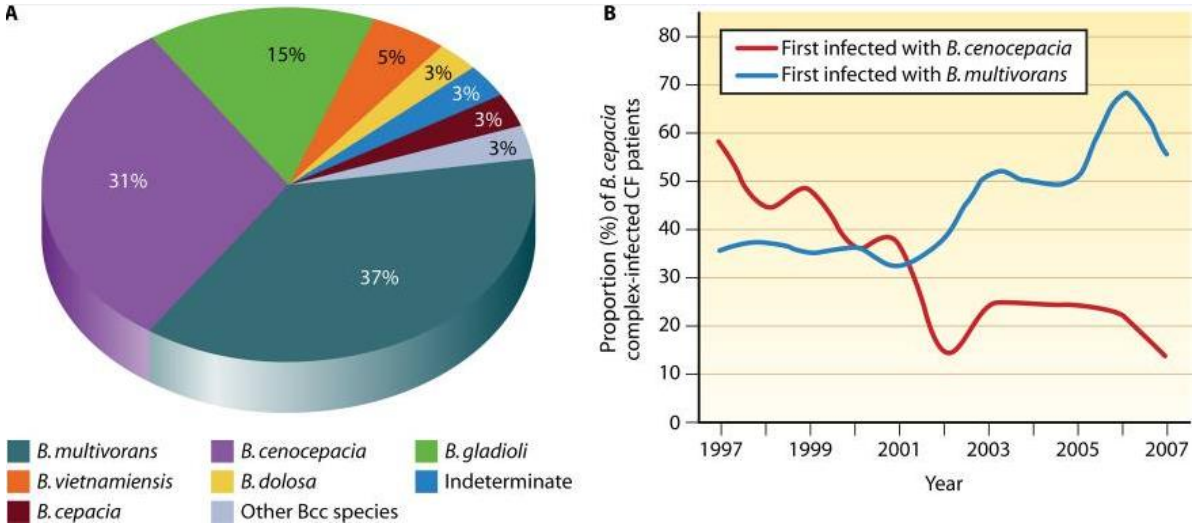


Figure 3 | BCC species infections in CF patients over a 10-year period (1997-2007). Figure obtained from (Lipuma 2010). (A) Distribution of BCC species in CF individuals from the USA (n = 2,024) during the period. (B) Incidence rates of *B. multivorans* (blue line) and *B. cenocepacia* (red line) over the period.

From MLST data, several *B. multivorans* STs (ST-16, ST-17, ST-18, ST-21, ST-24, ST-181, ST-190, ST-195, ST-198, ST-270, ST-274, and ST-375) are globally disseminated, having been found in multiple countries over several decades (Baldwin *et al.* 2008). These prevalent STs have been found in New Zealand, US, Canada, UK, Brazil, Czech Republic, Australia, Belgium, and Portugal (Baldwin *et al.* 2008). Several *B. multivorans* STs associated with outbreaks of CF lung infection have also been reported (Table 1).

Table 1 | *B. multivorans* STs associated with outbreaks of CF lung infection

Sequence Type	Other names	Comments	Reference
27	Glasgow outbreak strain	Hospital outbreak starting in December 1991	(Whiteford *et al.* 1995)
25	OHBM	Outbreak in a US CF centre (35.3% of patients shared the same strain)	(Biddick *et al.* 2003)
180	/	Multiple infections in CF and non-CF patients (Prague, Czech Republic)	(Baldwin *et al.* 2008)
179	TUL2	Outbreak in a US CF centre (62.5% of patients shared the same strain)	(Biddick *et al.* 2003)
199	/	Transient strain shared between two siblings	(Mahenthiralingam *et al.* 2001)
15	South Wales outbreak strain	Hospital outbreak in Llandough, Cardiff where 56% of patients had cross-infection of strains	(Millar-Jones *et al.* 1998)
419	/	Outbreak amongst 32 CF individuals in France	(Baldwin *et al.* 2008)
16	French epidemic clone	Patient-to-patient spread amongst 22 CF individuals in France	(Segonds *et al.* 1999)

Infection of the CF lung with BCC bacteria can be either transient or chronic, with the latter causing lung function decline (Mahenthiralingam *et al.* 2001). Certain individuals with chronic BCC infection may also develop a fatal necrotizing pneumonia known as "*cepacia* syndrome", which may also simultaneously coincide with invasive bacteraemia (Isles *et al.* 1984). This occurs in around 20% of BCC-infected CF individuals (Ledson *et al.* 1998). Bacteraemia is correlated with BCC infection, as other CF pathogens do not have the ability to cross the epithelial barrier (Leitão *et al.* 2010).

The lives of CF individuals are very affected by BCC infection. There is clear evidence of patient-patient spread of BCC bacteria (Millar-Jones *et al.* 1998; Segonds *et al.* 1999; Mahenthiralingam *et al.* 2001; Biddick *et al.* 2003). Therefore, contact with other CF individuals has been limited, with prohibition of two CF individuals meeting in certain countries. This includes hospital settings and social gatherings (Speert *et al.* 2002). More virulent BCC species such as *B. cenocepacia* may also replace an existing BCC strain in CF infection (Mahenthiralingam *et al.* 2001). Mechanisms underlying this replacement remains inconclusive. Therefore, there must be careful control of CF patient interactions to limit BCC spread, but also considering patient health to ensure they are not physically and mentally isolated (Speert *et al.* 2002). As a result of effective infection control, there has been a reduction in overall BCC prevalence and transmission (Saiman *et al.* 2014; Salsgiver *et al.* 2016). However, lack of knowledge of infection, improper infection control or lapses in management can in turn cause outbreaks of BCC infection (Table 1).

1.4 Virulence and pathogenesis of the BCC associated with CF infection

To date, molecular-level understanding of BCC pathogenesis remains limited. We currently know that successful CF infections are established by BCC bacteria through attachment to epithelial cells (Leitão

et al. 2010). The thick mucus layer in the CF lung enhances microbial ability for colonization due to a reduction in antimicrobial peptide efficacy and an enhanced inflammatory response (Boucher 2007). Invasion of the lung epithelia allows the BCC bacteria to evade the host immune response, providing an environment where the bacteria can replicate. Both *in vitro* and *in vivo* modelling of *B. multivorans* strains have shown successful epithelial invasion (Cieri *et al.* 2002).

There are several virulence factors which the BCC use during host infection, but their association with pathogenesis remains unclear. Multiple virulence factors have been identified in the BCC species and those also associated with *B. multivorans* (Table 2). These factors include the secretion of extracellular lipases and proteases (McClean and Callaghan 2009). Multiple lipases are secreted by BCC bacteria (Allan *et al.* 2003), and are thought to help with invasion (Mullen *et al.* 2007). In fact, *Burkholderia* are regarded as a highly important lipase-producing genus, although there is sparse information regarding lipases and virulence, particularly amongst the BCC (Mullen *et al.* 2007). *Burkholderia* secreted lipases can actually alter the cell morphology of macrophages, reducing phagocytic ability (Straus *et al.* 1992).

Initial CF lung infection by the BCC is driven through colonization of the upper respiratory tract (URT), aided by phenotypic traits such as motility, adhesion, biofilm formation, and tissue damage (Mahenthiralingam and Vandamme 2005). Motility and adherence to host cells is mostly driven through flagella, pili and adhesins (Leitão *et al.* 2010). Motility is key for establishment of infection and for movement through liquid environments. Studies have shown a reduction in virulence of BCC pathogens when flagella biosynthesis is impaired (Tomich *et al.* 2002). Pili and their use for mucin and cell adhesion in CF have been studied in *B. cenocepacia* (Sajjan *et al.* 1995; Sajjan *et al.* 2002). For instance, the cable pilus, found in the *B. cenocepacia* ET12 lineage, has been widely studied (Sajjan *et al.* 2002). The CblA pilin subunit (cblA gene) of the *B. cenocepacia* cable pilus was used as a genetic marker to identify virulent isolates (Sun *et al.* 1995a). Whilst pili have been found in *B. cenocepacia*, their presence and requirement for virulence in *B. multivorans* lung infection is uninvestigated. Post-establishment of infection, bacteria then build biofilm structures in the CF airways and sputum over the course of chronic infection (Mahenthiralingam and Vandamme 2005). *In vitro* biofilm-forming abilities are widespread amongst BCC bacteria, with species such as *B. multivorans* and *B. cenocepacia* generally being good biofilm formers (Conway *et al.* 2002).

The Gram-negative lipopolysaccharide (LPS), also known as endotoxin, comprises around 75% of the outer membrane of bacterial cells. This cellular component is necessary for bacterial growth, survival, and pathogenesis (Silipo *et al.* 2007). The BCC LPS is very interesting, particularly as it is unlike other Gram-negative bacteria in several ways. This virulence factor not only helps confer antibiotic resistance, but also induces cell damage through cytokines (De Soyza *et al.* 2004; Sengyee *et al.* 2019). LPS is comprised of O-antigen repeats and conserved lipid A structure. Not all Gram-negative bacteria have the O-antigen component, whereby BCC have a ‘rough LPS’ (Silipo *et al.* 2007). The LPS has

been shown to modify during chronic Gram-negative infection (Maldonado *et al.* 2016), such as cases like *P. aeruginosa* (Pier 2007; Faure *et al.* 2018). This modification protects the bacterium against the host innate immune response (Maldonado *et al.* 2016). The lipid A part of the LPS is altered by adding chemical groups, adjusting overall outer membrane charge (Ortega *et al.* 2009). These lipid A modifications suppress recognition by host immune receptors (Di Lorenzo *et al.* 2015) and can also increase resistance to antimicrobials through reduced membrane permeability (Needham and Trent 2013).

The presence of the O-antigen (OAg) is the most important phenotypic modification of the BCC LPS. Loss of OAg has been correlated with chronic infection of *B. multivorans* and *B. cenocepacia*, where loss was most seen in longer infection periods (Hassan *et al.* 2019). Interestingly, OAg loss also is frequently associated with more virulent lineages. For instance, *B. cenocepacia* recA IIIA lineage strains have a greater loss of OAg when compared to other lineages (Hassan *et al.* 2019). OAg loss in *B. multivorans* has been attributed to three mutations in the wbi gene cluster (Silva *et al.* 2016). Whilst the presence of O-antigen has been shown to help Gram-negative bacteria evade the immune response (Murray *et al.* 2006), there may be increased survival of BCC in host macrophages when OAg is absent (Hassan *et al.* 2019).

Multiple BCC virulence factors are regulated by quorum sensing (QS) systems, and all BCC bacteria harbour at least one QS system, such as well-documented CepIR (Table ***2***) (Leitão *et al.* 2010) which is conserved amongst the BCC (Suppiger *et al.* 2013). The CepIR system (Table ***2***) is dependent on acyl homoserine lactone (AHL) autoinducer molecules to activate or repress gene expression (Suppiger *et al.* 2013). QS systems regulate BCC virulence factor expression (Leitão *et al.* 2010) such as motility and biofilm formation (Venturi *et al.* 2004). Biofilm formation is not always under QS control, and is also dependent on growth conditions, with further research required to better understand important BCC phenotypes (Conway *et al.* 2002). *B. multivorans* has been shown to benefit from a functioning QS system during CF infection (McKeon *et al.* 2011). BCC bacteria can also interact and communicate with QS molecules produced by *P. aeruginosa*, such as within mixed biofilm interactions (Riedel *et al.* 2001). This highlights the importance of QS in BCC pathogenicity both alone and in mixed microbial communities**.**

Table 2 | Virulence factors previously found in the BCC

Virulence factor	Description	References
Flagella	BCC bacteria have either a single flagellum or cluster of polar flagella for swimming motility.	(Mahenthiralingam and Vandamme 2005)
Lipases	*B. multivorans* has been shown to produce high levels of lipase. These lipases may also contribute to lung epithelial invasion.	(Mullen *et al.* 2007)
Fimbrial and afimbrial adhesins	Fimbrial adhesins have been found in the genome sequences of clinical and environmental *B. multivorans*. Afimbrial adhesins, however, are more common in clinical *B. multivorans* strains.	(Denman and Brown 2013)
Outer Membrane Proteins (OMPs)	OmpW has been found in *B. multivorans* and *B. cenocepacia*. This OMP contributes to host-cell attachment[a]. However, immunization of murine models with OmpW have shown to effectively protect against *B. multivorans* infection[a]. OmpR has been shown in *B. multivorans* to help with switching from mucoid-to-nonmucoid in the CF lung[B.] This response regulator makes changes to the *B. multivorans* cell wall under stress conditions[B.]	[a](McClean *et al.* 2016) [b](Silva *et al.* 2018)
Lipoprotein	SlyB is an outer membrane integrity protein found in *B. multivorans* and *B. cenocepacia*.	(Plesa *et al.* 2006)
Lipopolysaccharide (LPS)	The LPS of BCC bacteria inflicts a host immune response, including proinflammatory cytokine induction, damaging host cells in the process. The LPS of BCC species, including *B. multivorans* has lipid A which may have links to successful human infection[B.]	[a](De Soyza *et al.* 2004) [b](Sengyee *et al.* 2019)
O-antigen	Presence of O-antigen may prevent phagocytosis, but loss of O-antigen repeats were linked to chronic infection with *B. multivorans*. As O-antigen may interfere with adhesion and stimulation of an immune response, loss of the O-antigen may be beneficial for the bacterium.	(Saldías *et al.* 2009)
Siderophores	*B. multivorans* was shown to have several secondary siderophores (ornibactin and *cepacia*chelin). These help with iron acquisition by the bacterium.	(Butt and Thomas 2017)

Exopolysaccharide	*B. multivorans* was shown to produce two types of EPS: cepacin and PS1. The former is more abundant in *B. multivorans* and the BCC.	(Herasimenka *et al.* 2007)
Biofilms	Multiple members of the BCC was shown to form biofilms. These structures enhance antibiotic resistance of the bacterium, making eradication more difficult. Biofilms directly correlate with clinical virulence of the pathogen, including chronic colonization of the CF lung microenvironment.	(Caraher *et al.* 2007)

There is currently little understanding about the virulence factors which aid chronic colonization of *B. multivorans* in CF lung infection (Silva *et al.* 2016; Diaz Caballero *et al.* 2018). It has been proposed that adaptation of *B. multivorans* to the CF lung microenvironment is like that of *P. aeruginosa* (Smith *et al.* 2006) and *Burkholderia dolosa* (Lieberman *et al.* 2011), due to the similarity in genomic mutation rate over long-term infection (Silva *et al.* 2016). Genotypic alterations in chronic *B. multivorans* infection have been observed in lipid metabolism and gene expression (Silva *et al.* 2016). In terms of phenotypic adaptations, there were changes to antimicrobial resistance, biofilm formation, and presentation of the lipopolysaccharide (LPS) O-antigen (Silva *et al.* 2016).

As the BCC cause a range of infection in humans, animals and plants, several infection models have been evaluated (Uehlinger *et al.* 2009): *Galleria mellonella* (Seed and Dennis 2008), *Caenorhabditis elegans* (Cardona *et al.* 2005), murine infection, onions, and the alfalfa plant (Uehlinger *et al.* 2009) (Table 3). One issue faced by the research community is that there is not a 'one model fits all' approach for using these strategies to characterize infection. All the infection models have their advantages as well as limitations, and a fully indicative 'true' CF model remains elusive (Leitão *et al.* 2010). Certain investigations have been performed to examine the overlap of BCC virulence factors within different infection models, but there is evidence that these factors are frequently host-specific (Uehlinger *et al.* 2009).

Table 3 | Infection models used to assess the virulence and pathogenicity of different BCC bacteria. Studies include both vertebrates and invertebrates. Table adapted from Leitão *et al.* (2010).

Infection model	Description of study	Reference
Vertebrates		
Murine agar bead model (Cash *et al*. 1979)	Swiss mice were used to model chronic pulmonary BCC infection using pneumonia. This used agar beads to initiate infection via intratracheal inoculation. The result compared the morphology and histopathology of the simultaneous BCC infections. Whilst there is a limitation that this model accurate, natural progression of BCC infection, the immune response and tissue damage in this model was like that observed in CF patients.	(Starke *et al.* 1987)
	Male Sprague-Dawley rats were inoculated with BCC strains using the agar bead model. Lungs from each rat were homogenized on infection days 7 and 21. BCC members, including *B. multivorans,* were able to establish chronic infection. This was used to compare virulence in the rats compared to alfalfa models, where *B. multivorans* was not suitable in this plant infection model.	(Bernier *et al.* 2003)
	C57/Black 6 mice were used for the agar bead infection using *B. cepacia* strains. Lungs and spleens were homogenized after 72-hours of infection. Genomovars containing *B. multivorans* and *B. cenocepacia* showed strains were invasive and noninvasive in each species. This indicated that the assay was strain specific.	(Cieri *et al*. 2002)
	C57BI/6 Mice were used to assess *B. cenocepacia* airway infection using the agar bead intratracheally infected model. Infection in the murine lung, liver and spleens were assessed at 1, 7, 14 and 28-days post-infection. Histological analysis was also performed on the infected murine lungs. The results showed that pathogenicity was not influenced by *B. cenocepacia* isolate origin.	(Pirone *et al.* 2008)
	Murine agar bead models were used to assess the role of type 3 secretion system (T3SS) gene *bscN* in the BCC. B57BL/6 mice were infected with *B. cenocepacia* J2315 WT and a mutant strain of CM56. Infection was measured at 3- and 7-days post-infection, analyzing the spleen and lungs of each mouse using histopathology. The results showed that BCC strains lacking *bscN* had a lower recovery of viable bacteria compared to the WT, supporting the requirement of T3SS in the BCC for full virulence.	(Tomich *et al*. 2003)
	Female C57/BL6 mice were inoculated with agar beads containing *B. cenocepacia* with and without the major flagellin subunit (*fliCII*). The mice were assessed for mortality over a 72-hour period. The aim was to assess whether the presence of flagella contributed to bacterial virulence and pathogenicity. No mortality was described in *B. cenocepacia* strains with a *fliCII* mutant. However, 40% of strains which had the full major flagellin subunit was observed. Therefore, flagella are likely to contribute to *in vivo* virulence.	(Urban *et al.* 2004)
	This project assessed *B. cenocepacia* virulence genes. Male Sprague-Dawley rats were used for the agar bead chronic lung infection models for a 10-day period. Genes were screened to identify those necessary for *in vivo* survival.	(Hunt *et al*. 2004)
	Male Sprague-Dawley rats infected with a clinical isolate of *B. cepacia* were used for this experiment. Rats were assessed at given time points over a 24-hour period for the effect of liposomal tobramycin on the lung infection. The results showed that inhalation of tobramycin improved drug effectiveness and longevity.	(Marier *et al*. 2002)
	B. cenocepacia K56-2 was used to infect male Sprague-Dawley rats using the agar bead model. The rat lungs were assessed for infection burden at 3 and 10-days post-infection. To ensure the *cepI* and *cepR* phenotype was present during infection, viable colonies were tested for protease production. Extracellular protease activity is regulated by the CepIR system, and therefore lack	(Sokol *et al.* 2003a)

	of protease production would indicate that the phenotype had been lost. The results found that the *zmpA* metalloprotease gene, regulated by CepR, may account of bacterial virulence.	
Mice with CF	Mice (between 2-17 months old) with CFTR knock-out mutations were used for this experiment. Mice lungs were removed and histologically examined for infection. A reduction in the clearance of BCC bacteria (*B. cenocepacia* J2315) and *S. aureus* was observed in this CF mouse model. As the models could not clear the infection, this model was deemed appropriate for bacterial pathogen study in the CF lung.	(Davidson *et al.* 1995)
	CFTR knock-out mice and respective WT CFTR mice comparisons (aged 6-8 weeks) were used in this experiment. The infectivity of *B. cenocepacia* BC7 (CF sputum isolate) from the ET12 lineage and environmental isolate ATCC 25416 were assessed. The results showed that the *B. cenocepacia* CF sputum isolate (BC7) was better at causing excessive inflammation and lung infection, over the course of 9 days, compared to environmental isolate ATCC 25416 in the CFTR knock-out mice. By comparing the CFTR mutated mice to the respective mouse WTs, it can be said this model may reflect what happens in human CF lung infection.	(Sajjan *et al.* 2001)
	CFTR knock-out mice and their respective WTs (age 10-12 weeks) were used for this experiment. *B. cenocepacia* K56-2 was assessed for infectivity in the CF and WT mice over a 5-day-period. The lungs and spleens of each mouse were assessed, using histopathology, after the infection period. *B. cenocepacia* K56-2 mutant strains (lacking either *cepI* or *cepR*) had a lower overall virulence compared to the K56-2 WT strain, irrespective of whether the mouse had the CFTR mutation or not. This model was useful in showing that the CepIR QS system of *B. cenocepacia* is required for full virulence in mice.	(Sokol *et al.* 2003a)
Mice with CGD	BCC bacteria were used in X-CGD mice for this experimental injection model. Murine infection was monitored over the course of 72-hours. Mice injected with epidemic *B. cenocepacia* J2315 all died of sepsis within the 3-day timeframe. The other BCC strains proved to be less virulent than *B. cenocepacia* J2315. A limitation of this experiment is how mice with CGD would translate to CF infection, particularly in terms of mucus clearance. However, it is a good model for assessing virulence of different BCC members in a vertebrate model.	(Sousa *et al.* 2007)
Leukopenic mice (leukocyte-deficiency)	Female BAL/c mice (age 6-8 weeks) were used for this experiment. They were treated with cyclophosphamide to create a leukocyte deficiency. The study used several BCC species for experimentation (*B. multivorans, B. cepacia, B. stabalis, B. vietnamensis,* and *B. cenocepacia*). Of the strains tested, *B. multivorans* appeared to be most persistent in the 16-day murine infection, with 5 of 6 strains infecting the mice over the whole period. The varying pathogenicity of each of the BCC strains could be assessed in this model, with differences in clinical outcome. This model is good for assessing *B. multivorans* persistence during infection.	(Chu *et al.* 2002)
	B. cenocepacia strain C1394 (CF isolate) was used to infect BAL/c mice (age 6-8 weeks). Immunosuppression of the mice to cause leukocyte deficiency was performed using cyclophosphamide. The results showed that this model was good for understanding not only bacterial persistence, but also implications in phenotype of the BCC.	(Chung *et al.* 2003)
Zebra fish model (*Danio reiro*)	Zebrafish (6-months old) were infected with *B. cenocepacia* by injection. *In vivo* survival of the bacterial strains was assessed at 24-hour time points. The model was used to also understand the production of a QS-regulated molecule (BDSF) *in vivo*. This is a molecule shown to be produced by *B. cenocepacia* which causes fungal antagonism. BDSF was shown to be important in *B. cenocepacia* virulence in the zebrafish model.	(Deng *et al.* 2009a)
Invertebrates		
C. elegens (nematode)	*B. cenocepacia* H111 (CF isolate) was used for experimentation in the *C. elegans* model. The results showed that bacterial growth medium heavily influences *C. elegans* death, due to differences in mode of killing. On media which supports nematode growth, killing takes a longer time (2-3 days) as bacterial load increases in the model lumen. A more rapid killing	(Köthe *et al.* 2003)

Limitation: not a viable model at 37°C	occurs on high-osmolarity medium. Irrespective of bacterial growth medium, the Cep QS system was required for invasion, but non-essential for overall killing. This contrasts that observed in vertebrate models.	
	Several BCC species (*B. cepacia, B. multivorans B. cenocepacia, B. stabilis, B. dolosa, B. ambifaria, B. anthina, B. pyroccinia,* and *B. vietnamensis*) were evaluated. This spanned both clinical and environmental BCC species and strains. The results showed that infection outcome, using the standard slow-killing assay, was strain-dependent rather than species-dependent. Overall, the nematode model was not effective for predicting BCC pathogenicity types based on their isolation source.	(Cardona *et al.* 2005)
Panagrellus redivivus (roundworm)	Both CF-associated and non-CF pathogens were used to assess the *P. redivivus* infection model. This included use of several *B. multivorans* strains. Overall, the model appeared to be more of a temperature-sensitive one. The *P. redivivus* model has been proposed as an alternative to the *C. elegans* model, due to its viability at 37°C. The results showed that *B. multivorans* could kill *P. redivivus* at 37°C, but this did not occur at 25°C. This gives the indication that a temperature-mediated virulence mechanism is utilized by *B. multivorans* in this model.	(Laws *et al.* 2005)
G. mellonella (wax moth larvae)	This experiment utilized strains from the BCC panels for infection modelling. This included *B. cepacia, B. cenocepacia, B. multivorans,* and *B. stabilis.* The model showed that *B. multivorans* had a higher LD_{50}s than *B. cenocepacia* and *B. cepacia.* This supports other studies which suggest *B. multivorans* is less virulent than these strains. Virulence was relatively similar amongst common CF-associated BCC strains (*B. cenocepacia* and *B. multivorans*), but variation was observed in *B. vietnamensis* and *B. dolosa.* Overall, this optimized infection model is a good alternative for rapidly assessing BCC virulence.	(Seed and Dennis 2008)
	The *G. mellonella* model has also been used to assess therapy treatments against *B. cenocepacia* CF isolates. In this study, experimentation of bacteriophage therapy was assessed as the first study *in vivo* over a 48-hour period. There are various factors which affect phage effectiveness against the infection, such as antibacterial activity, phage persistence, and lysogenic potential. Overall, the results show that this model is good for assessing phage therapy against BCC infection.	(Seed and Dennis 2009)
	The *G. mellonella* model has been used to assess genes required for *B. cenocepacia* virulence. The infection experiment was performed for 72-hours. Genes associated with cell adhesion were identified and tested for their role in *B. cenocepacia* virulence. Knock-out gene mutants were used to compare against the WT in the *G. mellonella* model.	(Holden *et al.* 2009)
Protozoa Limitation: Clinical BCC strains are mostly non-infective	Amoebas have been used to assess intracellular survival of BCC species. *B. vietnamensis* and *B. cenocepacia* were shown to survive inside *A. polyphaga* but lacked the ability to replicate. This study illustrated the importance of amoebae as an environmental survival mechanism., but its broader clinical relevance needs to be determined.	(Lamothe *et al.* 2004)

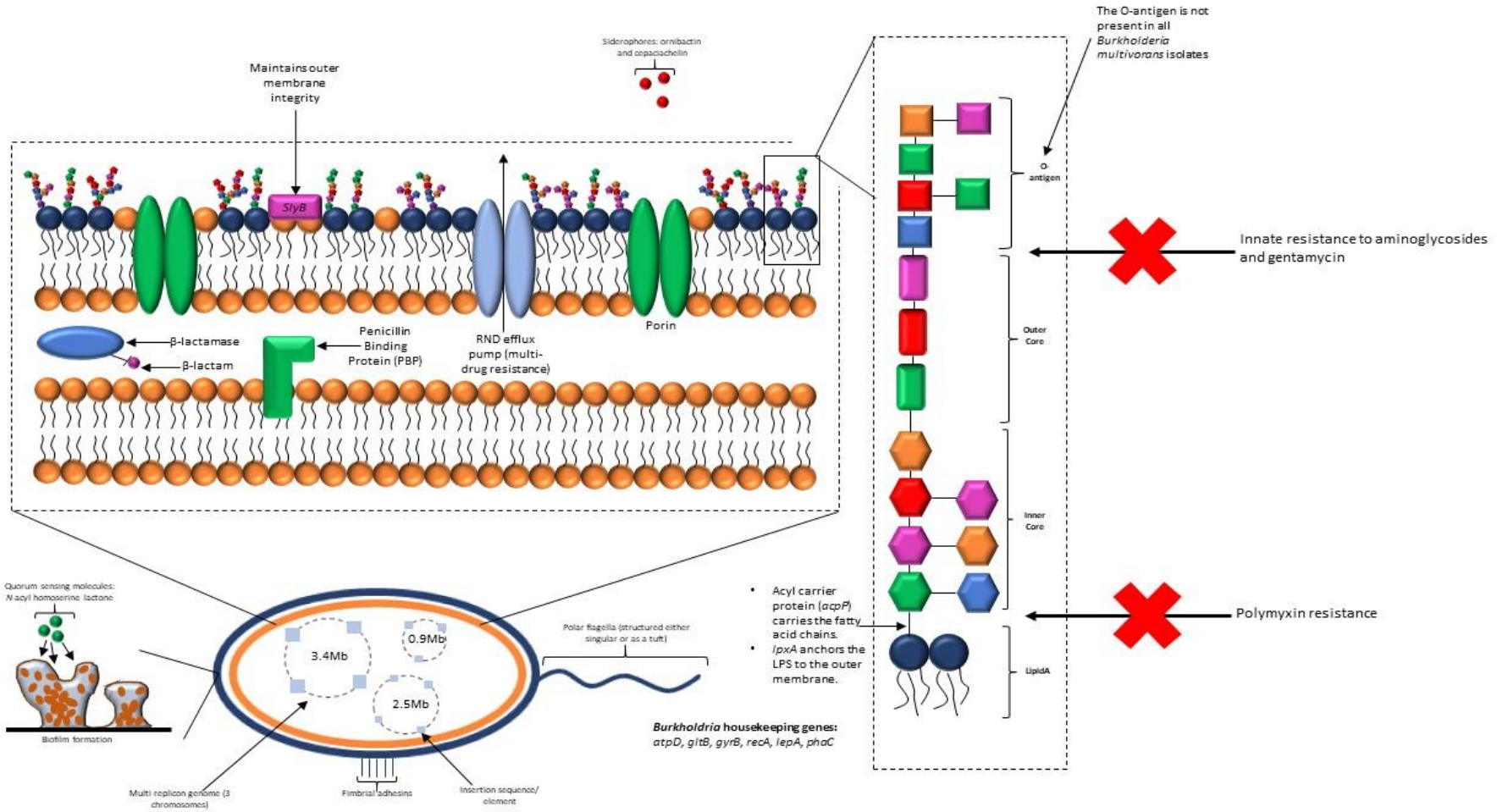


Figure 4 | An overview of *B. multivorans* virulence factors. The overview illustrates the organism's polar flagella, multi-replicon genome harbouring many insertion elements and fimbrial adhesions. The figure provides a more in-depth perspective of the factors associated with the inner and outer cell membranes, also showing that the organism can create biofilms through quorum sensing molecules. The image on the right shows the lipopolysaccharide and O-antigen and how this makes *B. multivorans* intrinsically resistant to polymyxins, through its modified lipid A structure, and innate resistance to aminoglycosides and gentamicin.

1.5 Treatment of BCC infection in CF

Antibiotic treatment is used to combat CF lung infection, suppressing coughing, excessive mucus, and breathlessness symptoms. Understanding of optimal therapies against BCC CF infection is difficult, with antibiotic combinations dependent on individual circumstances and antibiotic susceptibility patterns their infecting strain (Lord *et al.* 2020). As the BCC are multi-drug resistant, including resistance to many of the front-line antibiotics, it makes infections with these bacteria difficult to eradicate (EUCAST 2017). Aaron *et al.* (2000) showed that over 50% of CF individuals with BCC infection had strains that were resistant to the 10 most used antibiotic therapies. *B. multivorans* strains, for example ATCC 17616 (Prince *et al.* 1988), may also express antibiotic-degrading genes such as *penA* ***(Becka et al. 2018)***. This inducible class A ß-lactamase is ubiquitous amongst *Burkholderia* species. The *B. multivorans* PenA carbapenemase can hydrolyze ß-lactamase inhibitors ***(Becka et al. 2018)***. This causes broad-spectrum antibiotic resistance, rendering all ß-lactam and carbapenems ineffective (Nordmann and Poirel 2002; Livermore and Woodford 2006). The BCC also have intrinsic resistance to polymyxins (Loutet and Valvano 2011) and aminoglycosides (Coenye and Vandamme 2007). A modified lipid A confers strong polymyxin resistance in the bacterial genus (Loutet and Valvano 2011; Olaitan *et al.* 2014). Aminoglycoside resistance is through resistance nodulation cell division (RND) efflux pumps in *Burkholderia* species and the BCC (Podnecky *et al.* 2015). Whilst efflux systems in *B. multivorans* remain uninvestigated, analysis has been performed in *B. cenocepacia* strains (Guglierame *et al.* 2006; Buroni *et al.* 2009).

Another major problem is the lack of studies evaluating novel therapies for BCC infection in CF individuals. Studies which do exist have been shown to be inconclusive (Regan and Bhatt 2016), with no current randomized trials for optimum treatment against BCC infection (Horsley *et al.* 2016). The antibiotic susceptibility patterns of the BCC pathogens differ between CF individuals and those infected with the bacteria but do not suffer from CF. This is because of the continual courses of various antibiotics given to treat CF lung infections (Zhou *et al.* 2007; Castellani *et al.* 2018). These alternating antibiotic courses are used to prevent lung deterioration before onset of pulmonary exacerbations (Van de Kerkhove *et al.* 2016). Treatment lengths for CF infection are also patient specific. Inadequate eradication of infection may be a problem during short-course treatments. However, long-term treatment may cause problems with increased probability of allergic reactions (Parmar and Nasser 2005) or increased toxicity (Prayle and Smyth 2010). Long-term treatment with chloramphenicol may also cause blood composition imbalances (Lord *et al.* 2020). As the composition of the CF lung microbiota and antibiotic susceptibility patterns are very person-specific, better understanding of microbiota interactions and host-pathogen interactions are required. Tailoring of antibiotic treatment to individuals and their infecting strain will help improve our management of CF infection.

1.5.1 *Project Aims and objectives*

The overall aim of this research study was to define the population biology of *B. multivorans* as a CF lung pathogen, characterizing the strain using both genomic and phenotypic strategies as follows:

1) **Genome sequencing of a *B. multivorans* strain collection (Chapter 3)**

B. multivorans genomes were sequenced using the Illumina and Pacific Biosciences platforms and combined into a wider collection using sequences available in the public databases. The *B. multivorans* genome collection was then subjected to population structure analysis and construction of a representative sub-strain panel. Two major genomic lineages within *B. multivorans* were identified by this analysis. Model CF *B. multivorans* strains, to focus future research on, were also identified as part of this Chapter.

2) **Key genomic features of the *B. multivorans* (Chapter 4)**

The key genomic features of the *B. multivorans* was evaluated in this Chapter. This work identified virulence and pathogenicity genes, prophages, and plasmids within the sequenced genomes. Unique lineage genes were also identified and led to the development of PCR primers able to target each new *B. multivorans* genomic lineage. Murine infection analysis of the model strains was also performed to define their mammalian pathogenesis over short term infection.

3) **Phenotypic characterisation of the *B. multivorans* strains (Chapter 5)**

In parallel to the genomic analysis, a representative *B. multivorans* sub-strains collection was subjected to phenotypic characterisation using a range of conventional assays including colony morphology, growth rate analysis, motility, biofilm formation, exopolysaccharide production, antibiotic susceptibility testing and protease production. Phenotypic alterations were also assessed when an additional CF pathogen was combined in the presence of *B. multivorans* strains, to begin to model relevant polymicrobial interactions that may occur within lung infection.

2 Materials and Methods

2.1 Bacterial strains, growth media, chemicals, and antibiotics

2.1.1 Bacterial strains and incubation conditions

All bacterial strains used in this study were from the collections held by Mahenthiralingam research group at Cardiff University. All culture suspensions were stored in cryovials containing TSB supplemented with 8% dimethyl sulfoxide (DMSO) and held at -80 °C for long-term storage. A total of 73 *B. multivorans* isolates from the collection were sent for whole genome sequencing (Section 2.6.1; Table 5). A selection of n = 89 (Table 5; footnote a) *B. multivorans* isolates which have been used for colony morphology and n = 50 (Table 5; footnote b) *B. multivorans* isolates phenotypically characterized beyond colony morphology. The *B. multivorans* isolates were recovered from multiple sources denoted as follows: isolates from individuals with cystic fibrosis (CF), clinical infection isolates from a person without CF (NON), unknown, environmental (ENV), environmental hospital isolate (ENVH), clinical isolate but unknown if CF or NON (CLIN), and chronic granulomas disease (CGD) (Table 5). The splits for the strain panels can be found in Table 4. A selected panel of secondary CF pathogens were also chosen for community interaction analysis (Table 6).

Table 4 | Number of *B. multivorans* strains used in each panel for this work.

Panel	Isolate type						
	CF	ENV	CGD	ENVH	NON	CLIN	Unknown
Genomic panel (n = 283)	248	23	6	1	8	3	2
Genomic sub-panel (n = 77)	61	7	3	1	5	0	0
Colony morphology panel (n = 89)	77	3	3	1	5	0	0
Phenotypic panel (n = 49)	43	2	1	0	3	0	0

Table 5 | *B. multivorans* strains and genomes used in this study.

Study Strain Name	Alternative Name	Accession Number	Isolate Source	Origin	Reference
Genome sequences from this study					
BCC0006 [b]	4F/MATLOW	ERZ1645180	CF	Toronto, Canada	This study
BCC0009 [b]	LMG 18824	ERZ1645179	CF	USA	This study and (Speert et al. 1994)
BCC0032 [b]	/	ERZ1645178	CF	British Colombia, Canada	This study
BCC0033 [b]	C5568	ERZ1645177	CF	California, USA	This study and (Schmerk a Valvano 2013)
BCC0043 [b]	C1528	ERZ1645176	CF	Belfast, UK	This study
BCC0047 [b]	C1712	ERZ1645175	CF	Aberdeen, UK	This study
BCC0065 [b]	CEP0600	ERZ1645174	CF	Oregon, USA	This study
BCC0066 [b]	CEP0602	ERZ1645173	CF	British Colombia, Canada	This study
BCC0068 [b]	CEP0604	ERZ1645172	CF	Montreal, Canada	This study
BCC0074 [b]	HI-2308	ERZ1645170	CF	USA	This study
BCC0075 [b]	C1579	ERZ1645169	CF	Glasgow,UK	This study
BCC0079 [b]	C3430	ERZ1645168	CF	British Colombia, Canada	This study
BCC0080 [b]	C4861	ERZ1645167	CF	British Colombia, Canada	This study
BCC0082 [b]	C6100	ERZ1645166	CF	British Colombia, Canada	This study
BCC0084 [b]	C6398	ERZ1645165	CF	British Colombia, Canada	This study
BCC0087 [b]	C6935	ERZ1645164	CF	British Colombia, Canada	This study
BCC0093 [a]	C7363	ERZ1756586	CF	British Colombia, Canada	This study
BCC0096 [b]	C7062	ERZ1645163	CF	British Colombia, Canada	This study
BCC0101 [b]	C8298	ERZ1645162	CF	British Colombia, Canada	This study
BCC0134 [a]	CEP0699	ERZ1756605	CF	Denmark	This study
BCC0141 [b]	CEP074; AU0071	ERZ1645161	CF	USA	This study
BCC0181 [a]	LMG 14273	ERZ1756585	CF	Belgium	This study
BCC0188 [b]	LMG 14276	ERZ1645171	CF	Belgium	This study
BCC0225 [b]	CEP0837	ERZ1645159	CF	Manitoba, Canada	This study
BCC0241 [b]	CEP0935	ERZ1645160	CF	Minnesota, USA	This study
BCC0246 [b]	NU9366	ERZ1645158	CF	New Zealand	This study
BCC0247 [b]	CEP0949	ERZ1645157	CF	New Zealand	This study
BCC0255 [a]	CEP0966	ERZ1756587	CF	Australia	This study

BCC0264 [b]	CEP0992	ERZ1645156	CF	Australia	This study
BCC0266 [b]	CEP0995	ERZ1645155	CF	Australia	This study
BCC0269 [b]	CEP1000	ERZ1645154	CF	Australia	This study
BCC0292 [a]	C1409; CEP0133	ERZ1756588	CF	Edinburgh, UK	This study
BCC0293 [a]	C1406; BCC0293	ERZ1756589	CF	Oklahoma, USA	This study
BCC0300 [a]	CEP0503	ERZ1756590	CF	France	This study
BCC0303 [b]	AU0066	ERZ1645153	CF	USA	This study
BCC0317 [b]	BP102	ERZ1645152	CF	British Colombia, Canada	This study
BCC0321 [a]	C1664; CEP0181	ERZ1756591	CF	Manchester, UK	This study
BCC0375 [b]	AU0607	ERZ1645151	CF	USA	This study
BCC0381 [b]	LMG 16665	ERZ1645150	CF	UK	This study
BCC0384 [b]	FC0769	ERZ1645149	CF	British Colombia, Canada	This study
BCC0470 [a]	MN (#303)	ERZ1756592	CF	Italy	This study
BCC0493 [b]	BELF 1 (#353)	ERZ1645148	CF	Belfast, UK	This study
BCC0497 [b]	BELF 5 (#357)	ERZ1645147	CF	Belfast, UK	This study
BCC0533 [a]	C3168	ERZ1756593	CF	British Colombia, Canada	This study
BCC0553 [a]	C6558	ERZ1756594	CF	British Colombia, Canada	This study
BCC0583 [a]	C9140	ERZ1756595	CF	British Colombia, Canada	This study
BCC0585 [a]	C8814	ERZ1756596	CF	British Colombia, Canada	This study
BCC0702 [b]	IST453	ERZ1645146	CF	Portugal	This study
BCC0704 [a]	IST455	ERZ1756597	CF	Portugal	This study
BCC0710 [b]	A1-4	ERZ1645145	CF	Cardiff, UK	This study
BCC0729 [a]	C1628	ERZ1756598	CF	Cardiff, UK	This study
BCC0737 [b]	C1628	ERZ1645144	CF	UK	This study
BCC0814 [b]	3-454	ERZ1645143	CF	Prague, Czech Republic	This study
BCC0865 [b]	54-1138	ERZ1645142	CF	Prague, Czech Republic	This study
BCC0901 [a]	90-N320	ERZ1756599	CF	Prague, Czech Republic	This study
BCC0904 [b]	93-N331	ERZ1645141	CF	Prague, Czech Republic	This study
BCC0907 [a]	96-N337	ERZ1756600	CF	Prague, Czech Republic	This study
BCC0915 [a]	104-455	ERZ1756601	CF	Prague, Czech Republic	This study
BCC0921 [b]	110-1226	ERZ1645140	CF	Prague, Czech Republic	This study
BCC0962 [a]	151-454	ERZ1756602	CF	Prague, Czech Republic	This study

BCC0968 [a]	157-1226A	ERZ1756603	CF	Prague, Czech Republic	This study
BCC1147 [b]	C9861	ERZ1645139	CF	British Colombia, Canada	This study
BCC1148 [b]	C9862	ERZ1645138	CF	British Colombia, Canada	This study
BCC1177 [b]	D0913	ERZ1645137	CF	British Colombia, Canada	This study
BCC1185 [b]	D1348	ERZ1645136	CF	British Colombia, Canada	This study
BCC1190 [b]	D1442	ERZ1645135	CF	British Colombia, Canada	This study
BCC1271 [a]	AU 4543	ERZ1756604	CF	USA	This study
BCC1272 [b]	HI 2229	ERZ1645134	CF	USA	This study
BCC1367 [a]	AU0623	ERZ1756606	CF	USA	This study
BCC1368 [b]	AU1187	ERZ1645133	CF	USA	This study
BCC1384 [a]	AU0508	ERZ1756607	CF	USA	This study
BCC1385 [b]	AU4608	ERZ1645132	CF	USA	This study
BCC1421 [a]	C1576	ERZ1756608	CF	Glasgow, UK	This study
Genome sequences from public databases					
701_BMUL	/	GCF_001058025.1	CF	Washington, USA	(Roach et al. 2015)
800_BMUL	/	GCF_001058485.1	CF	Washington, USA	(Roach et al. 2015)
ATCC_17616 [b]	/	GCF_000018505.1	ENV	USA	This study and (Stanier et al. 1966)
ATCC_BAA-247	/	GCF_000959525.1	CF	Brussels, Belgium	(Johnson et al. 2015)
AU10047	/	GCF_002980655.1	CF	Unknown	(Becka et al. 2018a)
AU10086	/	GCF_002980675.1	CF	Unknown	(Becka et al. 2018a)
AU10398	/	GCF_002980695.1	CF	Unknown	(Becka et al. 2018a)
AU10897	/	GCF_002980715.1	CF	Unknown	(Becka et al. 2018a)
AU11204	/	GCA_002980785.1	CF	Unknown	(Becka et al. 2018a)
AU11233	/	GCF_002980775.1	CF	Unknown	(Becka et al. 2018a)
AU11358	/	GCF_002981015.1	CF	Unknown	(Becka et al. 2018a)
AU11772	/	GCF_002981035.1	CF	Unknown	(Becka et al. 2018a)
AU1185	/	GCF_001718755.1	CF	USA	(Winsor et al. 2008)
AU12481	/	GCF_002981075.1	CF	Unknown	(Becka et al. 2018a)
AU13919	/	GCF_002981095.1	CF	Unknown	(Becka et al. 2018a)
AU14328	/	GCF_002980825.1	CF	Unknown	(Becka et al. 2018a)
AU14364	/	GCF_002980815.1	CF	Unknown	(Becka et al. 2018a)
AU14371	/	GCF_002980855.1	CF	Unknown	(Becka et al. 2018a)

AU14786	/	GCF_002980875.1	CF	Unknown	(Becka et al. 2018a)
AU15814	/	GCF_002980895.1	CF	Unknown	(Becka et al. 2018a)
AU15954	/	GCF_002980905.1	CF	Unknown	(Becka et al. 2018a)
AU16734	/	GCF_002981615.1	CF	Unknown	(Becka et al. 2018a)
AU17135	/	GCF_002981135.1	CF	Unknown	(Becka et al. 2018a)
AU17534	/	GCF_002980935.1	CF	Unknown	(Becka et al. 2018a)
AU17545	/	GCF_002980995.1	CF	Unknown	(Becka et al. 2018a)
AU18096	/	GCF_002981145.1	CF	Unknown	(Becka et al. 2018a)
AU19518	/	GCF_002981195.1	CF	Unknown	(Becka et al. 2018a)
AU19564	/	GCF_002981155.1	CF	Unknown	(Becka et al. 2018a)
AU19659	/	GCF_002981255.1	CF	Unknown	(Becka et al. 2018a)
AU19729	/	GCF_002981215.1	CF	Unknown	(Becka et al. 2018a)
AU20929	/	GCF_002981635.1	CF	Unknown	(Becka et al. 2018a)
AU21015	/	GCF_003048355.1	CF	Unknown	(Becka et al. 2018a)
AU21596	/	GCF_002981675.1	CF	Unknown	(Becka et al. 2018a)
AU21747	/	GCF_002981315.1	CF	Unknown	(Becka et al. 2018a)
AU22436	/	GCF_002981695.1	CF	Unknown	(Becka et al. 2018a)
AU22892	/	GCF_002981295.1	CF	Unknown	(Becka et al. 2018a)
AU23365	/	GCF_002981715.1	CF	Unknown	(Becka et al. 2018a)
AU23668	/	GCF_002981725.1	CF	Unknown	(Becka et al. 2018a)
AU23690	/	GCF_002981325.1	CF	Unknown	(Becka et al. 2018a)
AU23919	/	GCF_002981335.1	CF	Unknown	(Becka et al. 2018a)
AU23995	/	GCF_002981755.1	CF	Unknown	(Becka et al. 2018a)
AU24277	/	GCF_002981375.1	CF	Unknown	(Becka et al. 2018a)
AU25057	/	GCF_002981775.1	CF	Unknown	(Becka et al. 2018a)
AU25543	/	GCF_002981815.1	CF	Unknown	(Becka et al. 2018a)
AU26250	/	GCF_002981835.1	CF	Unknown	(Becka et al. 2018a)
AU27706	/	GCF_002981395.1	CF	Unknown	(Becka et al. 2018a)
AU28069	/	GCF_002981845.1	CF	Unknown	(Becka et al. 2018a)
AU28442	/	GCF_002981415.1	CF	Unknown	(Becka et al. 2018a)
AU29198	/	GCF_002981455.1	CF	Unknown	(Becka et al. 2018a)
AU30050	/	GCF_002981485.1	CF	Unknown	(Becka et al. 2018a)

AU30438	/	GCF_002981515.1	CF	Unknown	(Becka et al. 2018a)
AU30441	/	GCF_002981535.1	CF	Unknown	(Becka et al. 2018a)
AU30760	/	GCF_002981555.1	CF	Unknown	(Becka et al. 2018a)
AU4507	/	GCF_002981595.1	CF	Unknown	(Becka et al. 2018a)
BCC0005 [a]	LMG 18822; C5393	ERS785011	CF	British Colombia, Canada	This study and (Mahenthiralingam et al. 1996)
BCC0008 [a]	LMG 16660; C1576	ERS784904	CF	Glasgow, UK	This study and (Whiteford et al. 1995; Vandamme et al. 1997)
BCC0010 [a]	LMG 16665; C1962	ERS784919	CF	UK	This study and (Hobson et al. 1995)
BCC0031 [a]	KE1; C0514	ERS785002	CF	British Colombia, Canada	This study and (Schmerk a Valvano 2013)
BCC0037 [a]	CEP0178	ERS785053	CF	California, USA	This study
BCC0050 [a]	J0365	ERS785036	CF	UK	This study
BCC0059 [a]	CEP0494	ERS785022	CF	Montreal, Canada	This study
BCC0067 [a]	CEP0603	ERS784935	CF	Montreal, Canada	This study
BCC0089 [a]	C7274	ERS785033	CF	British Colombia, Canada	This study
BCC0099 [a]	C7510	ERS785069	CF	British Colombia, Canada	This study
BCC0102 [a]	C8467	ERS784954	CF	British Colombia, Canada	This study
BCC0115 [a]	CEP0612	ERS784971	CF	Washington, USA	This study
BCC0149 [a]	FC0328	ERS785050	CF	USA	This study
BCC0175 [a]	FC0442	ERS785024	CF	Colorado, USA	This study
BCC0244 [a]	CEP0938	ERS785004	CF	British Colombia, Canada	This study
CF170.0a	/	GCF_003257435.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.10a	/	GCA_003256675.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.10b	/	GCF_003257315.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.10c	/	GCF_003257325.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.10d	/	GCF_003258065.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.10e	/	GCF_003258115.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.10f	/	GCF_003256685.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.10g	/	GCF_003256695.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.10h	/	GCF_003257345.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.10i	/	GCF_003256725.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.10j	/	GCF_003258145.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.11a	/	GCF_003256735.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.11b	/	GCF_003258135.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)

CF170.11c	/	GCF_003256765.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.11d	/	GCF_003257375.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.11e	/	GCF_003256805.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.11f	/	GCF_003258155.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.11g	/	GCF_003257365.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.11h	/	GCF_003256775.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.11i	/	GCF_003257405.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.11j	/	GCF_003256855.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.1a	/	GCF_003255945.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.1b	/	GCF_003257445.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.1c	/	GCF_003255985.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.1d	/	GCF_003257475.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.1e	/	GCF_003256835.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.1f	/	GCF_003257495.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.1g	/	GCF_003256865.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.1h	/	GCF_003255935.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.1i	/	GCF_003255975.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.1j	/	GCF_003257515.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.2a	/	GCF_003256005.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.2b	/	GCF_003257525.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.2c	/	GCF_003256015.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.2d	/	GCF_003256895.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.2e	/	GCF_003256905.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.2f	/	GCF_003257545.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.2g	/	GCF_003256055.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.2h	/	GCF_003256915.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.2i	/	GCF_003257575.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.2j	/	GCF_003256075.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.3a	/	GCF_003256095.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.3b	/	GCF_003257595.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.3c	/	GCF_003257625.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.3d	/	GCF_003257635.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)

CF170.3e	/	GCF_003258535.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.3f	/	GCF_003257615.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.3g	/	GCF_003256935.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.3h	/	GCF_003256105.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.3i	/	GCF_003257665.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.3j	/	GCF_003257705.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.4a	/	GCF_003257685.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.4b	/	GCF_003256115.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.4c	/	GCF_003257725.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.4d	/	GCF_003256985.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.4e	/	GCF_003256145.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.4f	/	GCF_003256185.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.4g	/	GCF_003257005.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.4h	/	GCF_003256175.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.4i	/	GCF_003257765.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.4j	/	GCF_003256225.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.5a	/	GCF_003256975.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.5b	/	GCF_003256195.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.5c	/	GCF_003256995.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.5d	/	GCF_003257075.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.5e	/	GCF_003257055.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.5f	/	GCF_003256255.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.5g	/	GCF_003256275.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.5h	/	GCF_003256285.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.5i	/	GCF_003257085.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.5j	/	GCF_003256305.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.6a	/	GCF_003257795.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.6b	/	GCF_003257755.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.6c	/	GCF_003257805.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.6d	/	GCF_003256355.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.6e	/	GCF_003257135.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.6f	/	GCF_003257825.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)

CF170.6g	/	GCF_003257845.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.6h	/	GCF_003257095.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.6i	/	GCF_003256335.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.6j	/	GCF_003256475.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.7a	/	GCF_003256525.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.7b	/	GCF_003257165.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.7c	/	GCF_003256515.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.7d	/	GCF_003257155.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.7e	/	GCF_003256365.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.7f	/	GCF_003256395.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.7g	/	GCF_003257865.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.7h	/	GCF_003256485.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.7i	/	GCF_003257175.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.7j	/	GCF_003257885.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.8a	/	GCF_003256545.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.8b	/	GCA_003256555.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.8c	/	GCF_003257915.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.8d	/	GCF_003257935.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.8e	/	GCF_003257955.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.8f	/	GCF_003257205.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.8g	/	GCF_003257235.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.8h	/	GCF_003256565.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.8i	/	GCF_003257255.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.8j	/	GCF_003257275.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.9a	/	GCF_003256605.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.9b	/	GCF_003257975.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.9c	/	GCF_003257985.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.9d	/	GCF_003257995.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.9e	/	GCF_003258005.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.9f	/	GCF_003257295.1	CF	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.9g	/	GCF_003256615.1	CGD	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.9h	/	GCF_003258055.1	CGD	Toronto, Canada	(Diaz Caballero et al. 2018)

CF170.9i	/	GCF_003258075.1	CGD	Toronto, Canada	(Diaz Caballero et al. 2018)
CF170.9j	/	GCF_003256655.1	CGD	Toronto, Canada	(Diaz Caballero et al. 2018)
CF2	/	GCF_000286575.1	CGD	Unknown	(Varga et al. 2012)
CGD1	/	GCF_000182255.1	CGD	Maryland, USA	(Varga et al. 2012)
CGD2	/	GCF_000182275.1	CLIN	Maryland, USA	(Varga et al. 2012)
CGD2M	/	GCF_000182295.1	CLIN	Maryland, USA	/
D2095	/	GCF_000807825.1	CLIN	Vancouver, Canada	(Silva et al. 2015)
D2214	/	GCF_000807815.1	ENV	Vancouver, Canada	(Silva et al. 2015)
DDS 15A-1	/	GCF_000756005.1	ENV	Australia	(Vandamme et al. 1997)
DSOPR54	/	GCA_002222845.1	ENV	Singapore	/
DSOPR57	/	GCA_002222875.1	ENV	Singapore	/
DWS 42B-1	/	GCF_000756965.1	ENV	Unknown	(Daligault et al. 2014)
FDAARGOS_246	/	GCF_003019965.1	ENV	Unknown	/
FDAARGOS_546	/	GCF_003938705.1	ENV	Unknown	/
FDAARGOS_547	/	GCF_003812365.1	ENV	USA	/
FDAARGOS_548	/	GCF_003812585.1	ENV	Unknown	/
HI3534	/	GCF_001528605.1	ENV	USA	/
MSMB1128WGS	/	GCF_001529795.1	ENV	Australia	/
MSMB1272WGS	/	GCF_001529925.1	ENV	Australia	/
MSMB1535WGS	/	GCF_001530485.1	ENV	Australia	/
MSMB1640WGS	/	GCF_001718995.1	ENV	Australia	/
MSMB1641WGS	/	GCF_001531425.1	ENV	Australia	/
MSMB1794WGS	/	GCF_001526715.1	ENV	Australia	/
MSMB1916WGS	/	GCF_001530985.1	ENV	Australia	/
MSMB2008WGS	/	GCF_001528045.1	ENV	Australia	/
MSMB2021WGS	/	GCF_001528425.1	ENV	Australia	/
MSMB575WGS	/	GCF_001534105.1	ENV	Australia	/
MSMB576WGS	/	GCF_001531955.1	ENV	Australia	/
MSMB612WGS	/	GCF_001532145.1	ENV	Australia	/
NCTC 13007	/	GCF_900446205.1	ENVH	/	(Segonds et al. 1999)
NKI379	/	GCF_001302465.1	NON	Taiwan	(Lin et al. 2011)
R-20526	/	GCF_001267755.1	NON	Belgium	/

BCC0033::pIN233-mCherry	/	/	NON	/	This study
BCC0033::pIN301-eGFP	/	/	NON	/	This study
BCC0084::pIN233-mCherry	/	/	NON	/	This study
BCC0084::pIN301-eGFP	/	/	NON	/	This study
BCC1272::pIN301-eGFP	/	/	NON	/	This study
BCC1385::pIN233-mCherry	/	/	NON	/	This study
BCC1385::pIN301-eGFP	/	/	Unknown	/	This study
ATCC 17616::pIN301-eGFP	/	/	Unknown	/	This study

[a]*B. multivorans* which have been subjected to colony morphology analysis; [b] *B. multivorans* which have been subjected to phenotypic analysis beyond colony morphology

Table 6 | Secondary pathogen strains used in the co-culture experiments

Secondary Pathogen	Isolate Source	Experiments used in (Sections)	Reference
P. aeruginosa LESB58	CF	5.2.9.1, 5.2.9.2, 5.2.9.3, 5.2.9.4, 5.2.9.5	(Cheng *et al.* 1996)
P. aeruginosa PA14	Clinical Isolate	5.2.9.1, 5.2.9.2, 5.2.9.3, 5.2.9.4	(Lee *et al.* 2006)
P. aeruginosa PAO1	Clinical Isolate	5.2.9.4	(Stover *et al.* 2000)
P. aeruginosa PAK	Clinical Isolate	5.2.9.4	(Cain *et al.* 2019)
P. aeruginosa CHA	CF	5.2.9.4	(Toussaint *et al.* 1993)
P. aeruginosa Mi162	Clinical Isolate	5.2.9.4	(Pirnay *et al.* 2009)
P. aeruginosa LES 400	CF	5.2.9.4	(Salunkhe *et al.* 2005)
P. aeruginosa LES 431	Non-CF	5.2.9.4	(Salunkhe *et al.* 2005)
P. aeruginosa AA43	CF	5.2.9.4	(Bragonzi *et al.* 2006)
P. aeruginosa C3719	CF	5.2.9.4	(Mathee *et al.* 2008)
P. aeruginosa LMG 14084	Environmental	5.2.9.4	(Pirnay *et al.* 2009)
S. aureus NCTC 12981	Lab strain	5.2.9.1, 5.2.9.2, 5.2.9.3, 5.2.9.4	Antibiotic susceptibility testing reference strain
C. albicans SC5314	Clinical Isolate	5.2.9.2, 5.2.9.3	(Gillum *et al.* 1984)
A. xylosoxidans LMG 1863	Clinical Isolate	5.2.9.1, 5.2.9.2, 5.2.9.3	(Ridderberg *et al.* 2015)
R. mannitolilytica LMG 18103	CF (blood)	5.2.9.2, 5.2.9.3, 5.2.9.4	(Hogardt *et al.* 2009)
S. maltophilia LMG 958	Type Strain	5.2.9.1, 5.2.9.2, 5.2.9.3	(Hogardt *et al.* 2009)
B. cenocepacia J2315-Pavia	CF	5.2.9.1, 5.2.9.2, 5.2.9.3, 5.2.9.4, 4.2.8.3	(Holden *et al.* 2009)
B. cenocepacia K56-2	CF	5.2.9.1, 5.2.9.2, 5.2.9.3, 5.2.9.4, 4.2.8.3	(García-Romero and Valvano 2020)
B. cenocepacia BCC0019 (LMG 18829)	CF	5.2.9.1, 5.2.9.2, 5.2.9.3, 5.2.9.4, 4.2.8.3	(LiPuma *et al.* 1988)
B. ambifaria BCC0207 (AMMD)	Environmental	4.2.8.3	(Coenye *et al.* 2001a)
P. aeruginosa LESB58::pIN233-mCherry	/	4.2.9.2, 4.2.9.3, 5.2.9.4	This study
P. aeruginosa LESB58::pIN301-eGFP	/	4.2.9.2, 4.2.9.3	This study

Culture purity was determined by streaking frozen stocks onto Tryptic Soy Agar (TSA) (Oxoid) and incubating plates for 24 hours at 37 ºC. This was performed at each passage to routinely work with fresh isolate cultures.

Overnight cultures were gown for 18-24 hours in 3ml of Tryptic Soy Broth (TSB) contained in a 15 ml tube and incubated at 37 ºC. Cultures were continually shaken on a rotating platform set to 150 rpm.

2.1.2 Growth media

Unless otherwise stated, all Oxoid (Basingstoke, UK) agar and broth growth media were purchased from Sigma-Aldrich Company Ltd. (Dorset, UK). All media was dissolved in double-deionized water and autoclaved at 121 ºC for 15 minutes to sterilize. Media was prepared as per instructions from the manufacturer, unless noted in other methods. TSA and TSB were used for the routine growth of all *Burkholderia* isolates. In cases of impure culture, *Burkholderia cepacia* selective agar (BCSA) (Oxoid) was used to specifically select for BCC species (Henry *et al.* 1999). Artificial Sputum Medium (ASM) was used in phenotypic growth experiments and was made using a modified method by Kirchner *et al.* (2012). The modification used all in one protein solution rather than individual protein solutions as per the original recipe. Other growth media used in this work for specific methods are noted in the appropriate Sections below.

2.1.3 Total viable counts

Enumeration of all bacterial and fungal cultures used in this work was performed using the Miles and Misra (Miles *et al.* 1938) method. All serial dilutions were performed using TSB media before dropping onto the appropriate culture agar, as stated under each results Section. Dilutions were plated in triplicate onto the agar plates and incubated for 24-hours at 37ºC, unless otherwise stated. Individual colonies were counted using the naked eye before calculating the colony forming units per ml (CFU/ml).

2.1.4 Chemicals and stains

Like media, chemicals and stains were also purchased from Sigma-Aldrich or Fisher Scientific (Loughborough, UK) unless stated in the specific Section. Chemicals were made into aqueous solutions using double-deionized water. If sterile solution was required, sterilization was performed using either autoclaving (121 ºC for 15 minutes) or by filtration through 0.2 µm sized pore Minisart® syringe filters (Sartorius, UK).

2.1.5 Antimicrobial agents

2.1.5.1 Antibiotic stocks

Stock solutions of all antibiotics used in this work were prepared in the appropriate sterile diluent to the given mg/ml (Table 7). Once prepared, all antibiotic solutions were stored at -20 ºC in 1.5 ml

microtubes. Stock solutions were further diluted for work in this thesis, where the final concentration is noted in each results Section.

Table 7 | Antibiotic stock solutions and their appropriate diluents used in this thesis.

Antibiotic	Abbreviated Name	Diluent	Stock concentration (mg/ml)
Trimethoprim	Tm	DMSO	50
Polymyxin B	PB1	Sterile water	50
Tetracycline	Tet	Ethanol	5
Ampicillin	Amp	Sterile water	50
Chloramphenicol	Cm	Sterile water	25

2.2 Statistical analysis

2.2.1 RStudio

All experiments were performed as a minimum of 3 biological replicates unless otherwise stated. This was to enable statistical testing. Multiple experiments also included technical replicates, noted where appropriate.

All statistical analysis in this thesis used R (R-Core-Team 2013), performed in Rstudio (RStudio-Team 2020). Data was checked for normality by plotting the mean data as histograms. This was then confirmed or rejected by checking with the Shapiro test and analysis of standard residuals. An ANOVA was performed in the first instance, to provide initial insight into the dataset. The standard residuals of the ANOVA were found using Pearson correlations. The standard residuals were plotted using qqnorm with qqline. If data was non-normal, it was transformed to ensure that normality could not be achieved before analysis. All statistical analysis in this thesis was non-parametric using a Kruskal-Wallis with post-hoc Dunn Test and Benjamini-Hochberg correction. The standard 95% confidence limit was applied throughout. Significance values were either noted as $p = < 0.05$ or fully stated for the analysis made. Plots were drawn in Rstudio using ggplot2, unless otherwise stated.

2.3 DNA extraction from bacterial cultures

2.3.1 Chelex® 100 resin for rapid DNA extraction

Rapid DNA extraction was performed using 5% (w/v) Chelex 100 resin (Sigma, UK) and basic protocol described by Mahenthiralingam *et al.* (2009). This was performed either on the solid-surface bacterial colony growth from agar plates or from overnight broth culture. For agar growth, multiple colonies were picked from the surface using a 200 µl sterile pipette tip and resuspended in 50 µl of sterile 5% (w/v) Chelex 100 resin in a 200 µl PCR tube. For growth in broth, 10 µl of sample was added to 40 µl sterile 5% Chelex 100 resin. The PCR tubes containing each sample were cycled through two rounds of heating to 95 ºC using a Bio-rad DNA Engine Dyad Peltier Thermal Cycler then cooling on an ice

block at 5 minutes per turn. After the freeze-thaw process was completed, the tubes were briefly left at room temperature to settle the resin before moving 2 µl of clear supernatant into fresh PCR tubes containing the reaction volumes as described previously. Chelex DNA extractions were always used fresh and not held for further storage, reducing the chance of DNA degradation.

2.3.2 The Promega Maxwell® 16 system for automated DNA extraction

Preparation of total bacterial genomic DNA was prepared using an automated DNA extraction system as follows. Overnight cultures were centrifuged for 10 minutes at 2,879 x *g*. The supernatant was carefully discarded, leaving the remaining pellet. The pellet was re-suspended in 500 µl sterile TSB. To each sample, 500 µl of Guanidine isothiocyanate (Promega, UK) was added and vigorously vortexed for 10 seconds. Using the Maxwell® 16 tissue DNA purification kit (Promega, UK), 300 µl of each sample was added into well 7 of the cartridges. One elution tube containing 300 µl of elution buffer was prepared for each cartridge. Samples were loaded into the Maxwell instrument, and the tissue DNA protocol was selected. This program takes approximately 45 minutes to complete extraction and purification of the DNA samples. Upon completion, each DNA sample was transferred into non-stick microtubes and stored at -20 ºC.

2.3.3 Promega Wizard® Genomic DNA purification kit

For long-read sequencing, the Wizard Genomic DNA purification kit (Promega, UK) was used, following the Gram-negative bacterial DNA extraction method. 1 ml of overnight culture was placed into a sterile 1.5 ml microtube. The culture was centrifuged at 16,000 x *g* for 2 minutes to create a pellet. The supernatant was carefully removed and discarded. 600 µl of Nuclei Lysis Solution was added to the microtube containing the pelleted cells. A pipette was used to re-suspend the pellet. The re-suspended cells were immediately placed in a heat block set at 80 ºC. This was incubated for 5 minutes to cause cell lysis before cooling for 12-15 minutes at room temperature. Once the suspension was at room temperature, 3 µl of RNase Solution was added to the tube and mixed by inverting 2-5 times. The tube containing the cell lysate was incubated at 37 ºC for 1 hour before cooling to room temperature. 200 µl of Protein Precipitation Solution was added to the room temperature cell lysate before vortex mixing for 20 seconds. The sample was immediately placed on ice for 5 minutes before spinning in a centrifuge for 3 minutes at 16,000 x *g*. The tube hinges were placed outward so that the protein precipitate would form at the bottom of the tube. 600 µl of room temperature molecular grade isopropanol was placed into a second sterile 1.5 ml microtube. The supernatant containing the bacterial DNA was carefully transferred into the new microtube. Residual liquid which was close to the protein precipitate was left to reduce the chance of contaminating the DNA. The microtube containing the supernatant and isopropanol was gently inverted until a visible thread-like mass was observed. This indicated that the DNA was present. The tube was then spun in a centrifuge again at 16,000 x *g* for 2 minutes. The supernatant was aspirated, and the tube was blotted onto clean, adsorbent Kimberly

Clark™ KIMCARE® medical wipes before leaving to air-dry for 10 minutes. 100 µl of DNA Rehydration Solution was added to the tube, flicking the bottom to ensure the pellet was dislodged from the tube side and suspended in the solution. The DNA was rehydrated by incubating at 65 ºC for 1 hour, flicking the tube after the first 10 minutes to ensure mixing. The tube was then left overnight in the fridge at 4 ºC before QC checks and sequencing.

2.4 Quality control of DNA extractions

2.4.1 Quantification using Qubit

Genomic DNA was quantified using a Qubit™ fluorometer (Invitrogen, Massachusetts, USA). Standards were used to set the standard curve. Neat DNA extraction products were set up and quantified as per manufacturer's instructions. If the DNA concentration was too high, nuclease free water (Severn Biotech Ltd) was used to dilute the sample 1:10 (v/v) and a second read was taken.

2.4.2 Quality assessment using NanoDrop

A NanoDrop Spectrophotometer (Thermo Fisher, USA) was used to measure the absorbance at 260/280 nm and 260/230 nm. High quality DNA is indicated with a ratio ~1.8 for the former absorbance and between 2.0 and 2.2 is considered pure for the latter.

2.4.3 16S ribosomal RNA gene sequencing to confirm bacterial purity

In any cases where bacterial purity was questioned, samples were sent for 16S rRNA gene sequencing to obtain a preliminary identification of the bacterial genera. This was also performed before complete PacBio DNA sequencing of *B. multivorans* strains. DNA was extracted as previously described (see Sections 2.3) before performing a 16S rRNA gene PCR. Universal primers 27F and 907R (Lane *et al.* 1985) (Table 8) were used for the amplification. Products and the forward primer were then sent in together in two vials via the Eurofins Genomics LightRun service for product clean-up and sequencing. The returned results included a text file which was opened in Notepad++. This short sequence(s) was copied and pasted into the NCBI BLASTn web server, with the 16S ribosomal RNA sequence database selected as the search database. All other parameters were left on default. The output file detailed sequence similarity, gaps, and e-value of the top BLAST hits in the database, giving an initial bacterial identification.

Table 8 | Primers and PCR conditions used to amplify the 16S rRNA gene for bacterial identification

Target gene	Primer name	Primer sequence (5' to 3')	Product size (bp)	Reference
16s	27F	AGAGTTTGATCCTGGCTCAG	~1400	(Lane *et al.* 1985)
PCR Stage		**Temperature** (°C)	**Number of cycles**	**Time (mins)**
Initial heating step		95	1	2
Amplification		94	35	0.5
Annealing		52		0.5
Extension		72		1.5
Final extension		72	1	5

2.5 DNA amplification and visualization

2.5.1 Polymerase Chain Reaction (PCR)

A DNA thermal cycler (Dyad®) was used for PCR analysis of the *B. multivorans* DNA. Amplification of the *recA* 1,040bp gene was used to ensure samples were *B. multivorans.* Optimized *B. multivorans*-specific primer sequences from Mahenthiralingam *et al.* (2000) were used, including the applicable annealing temperatures and times (Table 4). A 1kb ladder was used for molecular size comparison. Final reaction volumes were 50 µl per PCR tube. Each reaction consisted of 25 µl DreamTaq Green PCR Master Mix (2x) (Thermo Fisher Scientific, UK), 2µl of primer mixture (1 µl forward primer and 1µl reverse primer), 2 µl template DNA and 21 µl nuclease free water.

Table 9 | Primers and PCR conditions used to amplify the *B. multivorans*-specific *recA* gene

Target gene	Primer name	Primer sequence (5' to 3')	Product size (bp)	Reference
recA	BCRBM1	CGGCTCAACGTGCCGGAT	714	(Mahenthiralingam *et al.* 2000)
	BCRBM2	TCCATCGCCTCGGCTTCGT		
PCR Stage		**Temperature** (°C)	**Number of cycles**	**Time (mins)**
Initial heating step		96	1	3
Amplification		96	35	1
Annealing		56		1
Extension		72		1.5
Final extension		72	1	10

2.5.2 DNA visualisation using gel electrophoresis

Gel electrophoresis was used to visualize PCR products. A 1.5% (w/v) agarose gel was prepared with Tris-Acetate-EDTA (TAE) buffer. The solution was heated, swirling at regular intervals, in a conical flask until clear. Add 10µl SYBRSafe DNA stain per 100ml gel. The gel was poured into the mould, with an appropriately sized comb, whilst in a liquid state. The gel was then left to set for at least 30 minutes. The comb was carefully removed, ensuring the wells had properly formed. The gel was then placed into the electrophoresis tank before submerging with TAE buffer. Each well had 5 µl of sample

which were run at 70V to 80V. Gels were run until the samples were around three-quarters of the way down the gel. Once complete, the gel was removed from the instrument and placed into the visualization chamber and the gel image captured.

2.6 Genome sequencing, assembly, and annotation

2.6.1 Short-read DNA sequencing and assembly

The 73 *B. multivorans* from the collection at Cardiff University (Table 5) were subject to short-read Whole Genome Sequencing (WGS) using the Illumina MiSeq platform (Ravi *et al.* 2018). Genomic reads were assembled and annotated using the shared Cloud Infrastructure for Microbial Genomics (CLIMB) computing facility (Connor *et al.* 2016). Illumina reads were subjected to the Trim Galore v0.4.4 (Krueger 2017) wrapper script. This utilises Cutadapt v1.9.1 (Martin 2009) for automated quality and adapter trimming and FastQC v0.11.4 (Andrews 2009) for quality control. MultiQC v1.7 (Ewels *et al.* 2016) Python package was used to compile a single file report and interactive report for the samples, helping to streamline quality control screening. To assemble the bacterial genomes, we used the Unicycler v0.4.7 (Wick *et al.* 2017) assembly pipeline. This tool acts as a SPAdes (Bankevich *et al.* 2012) optimizer for streamlining *de novo* assembly of the genome contigs.

In addition to the 73 genomes sequenced as part of this thesis, a further 210 sequences were downloaded from the publicly accessible databases (NCBI) on 2nd November 2019 (Table 5). A final strain panel (*n* = 77) was constructed based on ST and phylogenomic analysis (Table 5). The specific strains used for each experiment are noted in the relevant Sections.

2.6.2 Long-read DNA sequencing and assembly

Three model *B. multivorans* strains (CF strains BCC0033, BCC0084 and BCC1272) were selected for long-read genome sequencing using the PacBio Sequel System (Pacific Biosciences). This complete genome sequencing was performed through the sequencing company Novogene (Cambridge, UK). Genomic DNA (see Section 2.3.3) was provided for each strain. After analysis, three output files were provided from the sequencing run: .bam, .pbi and .xml files. Genomic reads were processed using CLIMB (Section 2.6.1) (Connor *et al.* 2016). The Binary Alignment Map (BAM) file contains the raw genome sequencing data, therefore was the one to be processed. The BAM file was converted to a FASTQ file using bedtools bamtofastq (v2.30.0) (Quinlan and Hall 2010). PacBio FASTQ reads were then subjected to the Trycycler pipeline (v0.4.1) (Wick *et al.* 2021). A light read QC was performed on all genomic sequences using filtlong v0.2.0 (Wick 2017). Reads less than 1kb and the worst 5% of reads were discarded. The quality checked reads were subsampled to create multiple maximally independent read sets, based on the genome length. The --genome_size flag was used for a more rapid subsampling, as miniasm (v0.3-r179) (Li 2016) is not required. Read depths were noted for all three genomes. The subreads were then subjected to three different assemblers, creating 4 assemblies from each, to produce

a total of 12 FASTA files. The assemblers used were Flye v2.8.3(Kolmogorov *et al.* 2019), miniasm_and_minpolish.sh (minimap2 v2.21-r1071 (Li 2018); miniasm v v0.3-r179 (Li 2016) and minipolish v0.1.3 (Wick and Holt 2019)) and raven v1.5.1 (Wang *et al.* 2018).

2.6.3 Annotation of DNA sequences

Prokka v1.14.0 (Seemann 2014) was used to rapidly annotate the sequences and specifically identify putative coding sequences (CDS). This was performed on all the *B. multivorans* strain genomes used in this thesis (Table 5) including re-annotation of sequences obtained from the databases. The standard bacterial annotation database was used unless otherwise stated. Prokka highlighted CDS through the use of BLAST+ (Camacho *et al.* 2009) against well-known bacteria, and will also identify other basic genomic features such as the number of rRNAs, tRNAs and tmRNAs (Seemann 2014). It also produced several output files, such as .gff, which were then used for further analysis. RASTk (Aziz *et al.* 2008) was used to supplement the *B. multivorans* annotations, providing the number of RNAs and subsystems (playing functional roles) represented in each genome. The subsystems are identified from a manually curated database (Overbeek *et al.* 2005) and protein-encoding families (Overbeek *et al.* 2014).

2.6.4 Mapping of reads to a reference genome

Read mapping was performed on the *B. multivorans* short-read genomes against appropriate complete genome references (*B. multivorans* strains ATCC 17616, BCC0033, BCC0084, and BCC1272). This was used to check the percentage of reads which both matched and misaligned. Mapping was performed using Minimap2 (Li 2018). The misaligned reads were subjected to BLAST to identify any contamination in the sequences.

2.6.5 Scaffolding of genomic contigs to a reference genome

To add genomic context and approximate locations to the *B. multivorans* draft genomes, CONTIGuator2 (Galardini *et al.* 2011) online software was used to scaffold the genomic contigs to 4 *B. multivorans* complete genome references. The reference genomes used were *B. multivorans* strains ATCC 17616, BCC0033, BCC0084, and BCC1272. The reference genome which provided the smallest number of contigs and simultaneously the greatest scaffolded alignment was used to generate final consensus sequence for each draft genome sequence. Scaffolded replicons were subject to Prokka v1.14.0 (Seemann 2014) annotation as per Section 2.6.3.

2.6.6 Checking for contamination using Kraken2

To confirm the identity of the sequenced DNA, and thus any contaminating reads, the Minikraken database from Kraken2 v2.08-beta (Wood *et al.* 2019) was used. This taxonomically classifies the sequence reads by assigning *k*-mers which are then compared to the database (Wood *et al.* 2019). Any sequence reads which fell outside the *Burkholderiaceae* genus were discarded.

2.6.7 CheckM to analyse the long-read sequence data

All PacBio long-read genome sequences were subjected to CheckM (Parks *et al.* 2015) quality control. This consisted of the consensus FASTA files produced by Trycycler and *B. multivorans* ATCC 17616 complete sequence downloaded from NCBI. CheckM v1.0.18 (Parks *et al.* 2015) was used via the KBase online web server (Allen *et al.* 2017; Arkin *et al.* 2018) to screen the assemblies for taxonomic rank, genome completeness, contamination percentage and strain heterogeneity. The output also details the number of reference genomes used to infer the lineage, the markers and sets of genes that inferred the taxonomy, as well as the number of times the specific marker provides a hit.

2.6.8 QUAST quality control

QUAST v5.0.2 (Gurevich *et al.* 2013) was used to assess the quality statistics for the genomic assemblies. This was run using default settings without a reference genome. QUAST provided the genome size (bp), number of contigs, GC content, longest contig size, L50 and N50 values.

2.7 BLAST analysis for identification of DNA and genome sequences

2.7.1 NCBI BLAST

The Basic Local Alignment Search Tool (BLAST; (Altschul *et al.* 1990; Camacho *et al.* 2009)) was used to identify similarities between DNA sequences. The online NCBI BLAST server was used to compare DNA sequences obtained in this thesis to the RefSeq (Pruitt *et al.* 2006) genome databases. Query sequences were uploaded in FASTA format. The search was set to standard databases and optimised to megablast. Where applicable, the 'organism' search was imputed as *B. multivorans*. In cases where a general search was required, this was left blank.

2.7.2 Local BLAST

To compare DNA sequences to a personalised, custom database, local BLAST searches were used. Local BLAST compared unpublished sequences which were not publicly available. This work was performed in CLIMB (Connor *et al.* 2016) (Section 2.6.1). The local BLAST database was created using a single multiFASTA file, with the nucleotide BLAST type option chosen. The query sequences were then run against the local BLASTn database with -max_hsps set to 1 and -outfmt to 6. This provided an output in a comma separated tabulated format, with only one hit per query.

2.8 Average Nucleotide Identity and genome alignments

2.8.1 Species-level identification and relationships using Average Nucleotide Identity (ANI)

Two methods were used to confirm species validity of the *B. multivorans* strains: PyANI (Pritchard 2014) and FastANI (Jain *et al.* 2018). PyANI v0.2.9 (Pritchard 2014) was used as the initial species-level identification of the *B. multivorans* strains. This used pairwise alignments to illustrate the genomic

relationships within the bacterial species. The application was used with the MUMer algorithm It outputted several graphical heatmaps covering the percentage identity, alignment coverage, alignment length and similarity errors.

FastANI v1.32 (Jain *et al.* 2018) was used as a secondary, alignment-free method for whole-genome ANI analysis of the *B. multivorans* strains. This was performed using the many to many arguments, where a query list was compared to a reference list. The reference genomes used in the analysis were the *B. multivorans* complete genome model strains ATCC 17616, BCC0033, BCC0084, and BCC1272. The reference genome which most closely reflected the genomic lineage and had the closest similarity selected to compare against draft genomes and confirm their ANI grouping.

2.8.2 Roary pan genome analysis and core-gene alignment tool

The Roary (v3.13.0) pipeline (Page *et al.* 2015) was used for a pan-genome analysis and core-gene alignment of 283 *B. multivorans* genomes (Table 5). Prokka-annotated GFF files (Section 2.6.3) were used as the input. The -e parameter were used to create a MAFFT multi-FASTA core-gene alignment (Katoh *et al.* 2002; Page *et al.* 2015) and -r to create R plots.

2.9 Phylogenomic analysis of the whole genome and MLST sequences

2.9.1 FastTree and RAxML

The GFF annotated-file outputs from Prokka (Seemann 2014) (Section 2.6.3) were used in the Roary v3.12.0 (Page *et al.* 2015) pan genome pipeline to assess the core and accessory genome of the strain panel. The command was performed using the default settings (95% minimum percentage ANI for blastp) (Page *et al.* 2015). The core gene alignment output file produced by Roary (Page *et al.* 2015) was created using the MAFFT alignment algorithm. This was then used to create a double-precision maximum-likelihood phylogeny using FastTree v2.0.2.9 (Price *et al.* 2009), utilizing the Generalized Time-Reversible (GTR) evolutionary model (Lanave *et al.* 1984). After determining the root branch point for a phylogeny, a final phylogenetic tree was constructed using maximum likelihood (GTRGAMMA model) Randomized Accelerated Maximum Likelihood (RAxML v8) (Stamatakis 2014), supported by 100 bootstraps.

2.9.2 Determining sequence type and creating phylogenies

Sequence Types (STs) were also determined for the *B. multivorans* strains using the MLST v2.18.1 script by Torsten Seemann (Seemann), performed by matching against the pre-defined Bcc database. A secondary MLST was performed using PubMLST blast schemes, using MLSTcheck (Page *et al.* 2016b), on the final 77 *B. multivorans* panel strains. This uses a nucleotide BLAST of the query genomes against public MLST databases, creating multiFASTA output files for each of the strains. For

comparison to the core gene alignment phylogeny, the multiFASTA file outputs were concatenated into a single alignment file and an MLST tree was drawn using FastTree (Price *et al.* 2009) (Section 2.9.1).

2.9.3 Phyloviz online

The MLST data provided by Section 2.9.2 (Seemann ; Page *et al.* 2016b) was used for analysis using Phyloviz Online (Francisco *et al.* 2012; Ribeiro-Gonçalves *et al.* 2016). A manually curated table of all *B. multivorans* STs, and related strain information (isolate type and country of origin), found in the pubMLST database, and the genomes used in this thesis (Table 5) were compiled and submitted privately to the online interface. Various MLST-based phylogenetic trees were drawn, based on the strain information, to understand the distribution and epidemiology of the *B. multivorans* strains. N Locus Variant (NLV) thresholds were then used to investigate the genomic relatedness of the *B. multivorans* strains. Tree cut-off and NLV graph thresholds were adjusted to identify the closely related nodes (Ribeiro-Gonçalves *et al.* 2016).

2.9.4 Visualisation and editing of phylogenomic trees

All phylogenetic trees were visualized using FigTree v1.4.4 (Rambaut 2009) graphical software. These were then edited using Inkscape v0.92.1 (Project 2020).

2.10 Variant analysis

2.10.1 Use of Snippy to identify variants and create core-SNP phylogenies

Haploid variant calling and core genome alignment was performed using Snippy v3.2-dev (Seemann 2018) Nucleotide polymorphisms were predicted via Snippy by aligning NGS reads to a reference genome using BWA MEM. The complete genomes of *B. multivorans* ATCC 17616 (downloaded from NCBI), BCC0084 and BCC1272 (PacBio sequences) were used as the reference genomes. BCC0084 and BCC1272 were used as the lineage representatives when creating SNP core gene alignments and analysis of the two lineages individually. Resultant variant validity is assessed using Freebayes prior to employing SnpEff to determine SNP effects on the genome. The variants were noted as follows: SNP = Single Nucleotide Polymorphism (only one base change), MNP = Multiple Nucleotide Polymorphism (two or more sequential bases change), INS = Insertion of a base within the string, DEL = Deletion of a base within the string and COMPLEX (a combination of SNPs and MNPs).

2.10.2 Gubbins to address recombination in the variant analysis

To address recombination in the snippy aligned (Section 2.10.1) *B. multivorans* genomes, the Genealogies Unbiased By recomBinations In Nucleotide Sequences (Gubbins) (Croucher *et al.* 2015) algorithm was used. The Snippy clean full alignment was subjected to Gubbins (Croucher *et al.* 2015) using the -p flag. The SNP-sites tool (Page *et al.* 2016a) was next performed on the Gubbins (Croucher *et al.* 2015) filtered polymorphic sites FASTA file to create a clean core SNP alignment. The outputted

alignment was passed through phylogenetic tree building software as above (Section 2.9.1) to give the final Gubbins-filtered core-SNP phylogenies.

2.11 Identification of B. multivorans genes of interest using a pan-GWAS approach

2.11.1 Pan-GWAS scoring of key genes

Scoary was used as a pan-GWAS approach to assign genes identified in the *B. multivorans* genomes to a given 'trait'. The algorithm used the Roary (Page *et al.* 2015) gene presence absence output file and an associated binary trait file assigned manually. The trait files used in this work were to identify a) lineage-specific genes (Section 4.2.7), and b) genes associated with phenotype (Section 5.2.10). For the phenotype analysis, a positive trait was defined as 'motile', 'biofilm forming', 'exopolysaccharide producing', and 'normal growing'. If the strain could exhibit any motility, biofilm, or exopolysaccharide, irrespective of the amount, it was noted as positive. Negative traits were given to 'non-motile', 'non-biofilm forming', 'no exopolysaccharide production', and 'slow growing'. Positive traits were scored '1' and negative traits as '0' in the matrix. Scoary was run using default parameters (p-value <0.05).

2.11.2 Use of a semi-automated shell script to sort the data

A semi-automated approach was used to analyse the lineage-specific gene data. The output of Scoary (Brynildsrud *et al.* 2016) (Section 2.11.1) provided a spreadsheet of genes and their relative number of hits to each lineage. The spreadsheet was manually filtered for genes which were only present in one lineage, acquiring a list of genes unique to both lineages individually. A local BLASTn (Section 2.7.2) was used to check the uniqueness of the list when the *B. multivorans* genes tested increased to all 283 genomes (Table 5). Any hits which were no longer lineage-specific were manually checked for and discarded. The final BLAST spreadsheet was then subjected to a customised shell script (available via Github) which ensured the data was in the correct format for processing, removed any duplicates, sorted the data using parallel (Tange 2018), counted the number of gene hits for each lineage, then outputted the results in a .csv file.

2.11.3 Identification of the lineage-specific gene locations and its stability

The Artemis viewer (Berriman and Rutherford 2003) was used to identify the gene location of the unique lineage genes of interest on the *B. multivorans* complete genomes. The region was checked for stability. This ensured there was no skewed GC content, indicating genomic changes such as horizontal gene transfer. The gene was only considered if it was found on replicons 1 or 2 as these are present in all BCC bacteria.

Once the genomic location was identified and parameters accepted, Bedtools was used to extract the FASTA sequence of the gene from the respective *B. multivorans* complete genome. The gene was then

subject to BLAST against the scaffolded *B. multivorans* strain panel, to ensure that the gene location was consistent across all the genomes. Bedtools (Quinlan and Hall 2010) was used to extract the FASTA sequences for the gene in the *B. multivorans* scaffolds using the BLAST output.

2.12 PCR primer design

2.12.1 *In silico* design to identify mismatches in the sequences

Four primers were designed in total, targeting 4 different *B. multivorans* genes (*ghrB_1, naiP_3,* and *yiaJ_1* against lineage 1 genomes and *glnM_2* against the lineage 2 genomes). The Bedtools (Quinlan and Hall 2010) extracted gene sequences were aligned using MAFFT (Katoh *et al.* 2002) to reveal the conserved and unique nucleotide sequence regions. Manual analysis of the aligned sequences was used to pinpoint the unique regions as possible primers. At least one primer of the pair was designed to have at least 2 mismatches on the 3' end. Primer lengths were kept within a range of 18-22 bases. Primer pairs were checked using both the NCBI primer designing tool and the Eurofins Genomics Oligo Analysis Tool for suitability. BLAST was then used to check primer cross-reactivity with other closely related *Burkholderia* species. An additional check was performed using the online OligoAnalyzer Tool by Integrated DNA Technologies to ensure the primers did not cause hairpin structures and were appropriate for PCR. Successfully designed primer sequences were sent to Eurofins Genomics for synthesis.

2.12.2 Use of SnapGene for *in silico* visualisation of the PCR primers and products

To visualize the forward and reverse primers in the *B. multivorans* complete genomes, SnapGene software (Insightful-Science) was used. The primer pair targeting *ghrB_1* was analyzed in the *B. multivorans* BCC0084 genome. Primers targeting *glnM_2* were analyzed using the *B. multivorans* ATCC 17616 genome. DNA files for the two *B. multivorans* strains were loaded into the software, indicating that the sequence was circular. The designed forward and reverse primers were manually mapped in the software against the respective DNA file. The output provided the genomic location of the forward and reverse primers, as well as annotation of surrounding features. SnapGene was also used to simulate a 1.2% agarose gel *in silico*. The primer pairs were added into the gel lanes, a 1 kb DNA ladder selected, and Taq DNA polymerase chosen.

2.12.3 PCR protocol optimisation

PCR optimization was based around the melting temperature of the primers, and a gradient PCR protocol was used to identify an optimal annealing temperature. The annealing temperature ranged from 48 ºC to 65 ºC. PCR reactions were set up in 200 µl microtubes and the number of cycles used was 35. This cycle number was recommended for DreamTaq DNA polymerases.

2.12.4 Testing of the lineage-specific primers

The PCR primers were initially tested using the *B. multivorans* phenotypic strain panel of 48 strains (Table 5). Negative controls consisted of the opposing *B. multivorans* lineage and DNAse-free H_2O. The PCR reactions were performed using final reaction volumes of 25 µl. Each reaction consisted of 12.5 µl DreamTaq Green PCR Master Mix (2X) (Thermo Fisher Scientific, UK), 1µl of each primer (1 µl forward primer and 1µl reverse primer), 1 µl template DNA and 9.5 µl nuclease free water. Final PCR primers and the associated protocols can be found in results Section 4.2.8. To confirm there was no cross-reactivity in the primers, a second PCR was performed using 2 closely related *B. cenocepacia* strains (K56-2 and BCC0019) and a *B. ambifaria* AMMD strain. PCR products were visualized using a 1.2% agarose gel electrophoresis gel, submerged in TAE buffer, and run at 80V. Gels were then visualized in a gel imager.

2.13 Genetic manipulation of the model B. multivorans strains

2.13.1 Triparental mating of pMLBAD into the *B. multivorans* strains

pMLBAD (Lefebre and Valvano 2002), a well-characterized *B. multivorans* cloning vector, was used to see if the four model *B. multivorans* strains (ATCC 17616, BCC0033, BCC0084, and BCC1272) could take up DNA via conjugal transfer. Conjugative plasmid transfer into *B. multivorans* was achieved using a triparental mating method (Craig *et al.* 1989). This used helper *E. coli* strain HB101 containing the pRK2013 plasmid (Figurski and Helinski 1979; Ditta *et al.* 1980) and *E. coli* DH5α donor (Grant *et al.* 1990). For the initial growth, *B. multivorans* strains ATCC 17616, BCC0033, BCC0084 and BCC1272 and *E. coli* strains were grown in 5ml of TSB and Luria-Bertani (LB) broth, respectively.

Overnight cultures were centrifuged for 10 minutes at 4,000 rpm. The supernatant was carefully discarded, leaving the remaining pellet. The pelleted cells were resuspended in 5ml modified LB broth (supplemented with 10 mM $MgCl_2$) and the centrifugation washing step was repeated. After discarding the supernatant, the *E. coli* cells and *B. multivorans* cells were resuspended in 4 ml and 3ml of modified LB respectively. Each bacterium alone, as a culture, acted as the negative control. Aliquots of 50 µl of each bacterium were plated directly onto LB agar supplemented with 10 mM $MgCl_2$. For the conjugation, the *E. coli* HB101 helper, *E. coli* DH5α donor and *B. multivorans* strain was mixed equally in 50 µl volumes in a 1.5ml microtube. This was mixed using a vortex for 20 seconds before aliquoting 50 µl of each strain mixture onto modified LB agar. The plated drops were allowed to dry completely and incubated at 37 ^{o}C for 24 hours to enable conjugation.

Sterile plastic loops were used to take scrapings and resuspend the growth from the modified mating mixtures into 1ml of sterile TSB. Selective media of TSA containing 600 units of polymyxin B (PB1) and 50 µg/ml of trimethoprim (Tm) were used for plating the single bacterium controls and conjugants.

Suspensions were plated in 100 µl volumes and spread evenly using a sterile spreader. Plates were incubated for 72 hours at 37 ºC. Multiple colonies were individually picked from the mixed conjugation plates and were replated onto selective TSA plates. Plates were incubated overnight at 37 ºC to obtain fresh cultures of transconjugants.

2.13.2 Electroporation of fluorescent reporter gene plasmids into the *B. multivorans* model strains

The fluorescent reporter gene plasmids pIN301 (pIN301-eGFP) (Mesureur *et al.* 2017) and pIN233 (pIN233-mCherry) (Gomes *et al.* 2018) were used as additional tests to see if *B. multivorans* would take up DNA by electroporation and express recombinant proteins. *B. multivorans* cultures were grown in TSB for 18-24 hours before being diluted to OD_{600nm} 0.1 in 3ml TSB. These were then incubated at 37 ºC, shaking at 150 rpm for around 4 hours. This allowed the *B. multivorans* model strains to reach an OD_{600nm} of 1. Cultures were checked for the correct OD_{600nm} by sampling and screening using a spectrophotometer. The remaining 2 ml aliquot was spun down in a centrifuge for 5 minutes at 4000 rpm and washed with an equal volume of room temperature sterile ddH_2O. This was repeated two times in total. Cultures were subject to a final resuspension in 30 µl of sterile room temperature ddH_2O. 10 ng of room temperature vector DNA was added to the suspension. The mix was transferred to electroporation cuvettes and electroporated using 2500V and 2 mm setting. The field capacity was 12.5 kV/cm. Electroporated cell were recovered for 1 hour in 1 ml sterile TSB at 37 ºC, shaking at 150 rpm. The neat cultures were plated on TSA containing 50 µg/ml chloramphenicol (Cm). Plates were incubated statically at 37 ºC overnight before confirming pIN301-eGFP or pIN233-mCherry uptake using blue light. Electroporated strains were captured using a Biospace Lab Photon Imager Optima. This used excitation of 488 nm and emission of 522 nm for pIN301-eGFP. For pIN233-mCherry electroporated strains, an excitation wavelength of 580 nm and emission of 610 nm was used for the imaging. Both pIN301-eGFP and pIN233-mCherry colonies were transferred into cryovials containing TSB supplemented with 8% DMSO at -80 ºC for long-term storage. The same procedure was used to introduce both plasmids into *P. aeruginosa* strain LESB58 using a selection of 50 µg/ml Cm.

2.13.3 PCR to confirm genetically manipulated strains were *B. multivorans*

To confirm that the pMLBAD (Section 2.13.1) pIN301-eGFP and pIN233-mCherry (Section 2.13.2) *B. multivorans* genetically manipulated strains were from the expected parent, a rapid DNA extraction was performed using the Chelex method (Section 2.3.1). This picked a colony from a 50 µg/ml Cm supplemented selective TSA agar plate. Random Amplification of Polymorphic DNA (RAPD) was used to compare the pmlBAD, pIN301-eGFP, and pIN233-mCherry genetically manipulated *B. multivorans* to their respective wild type strains using primer sequences and PCR settings noted in Table 10. The negative experimental control was the nuclease free water alone. A 1.5% agarose gel was used to visualize the amplified DNA products. Banding patterns of the wild type and respective pmlBAD,

pIN301-eGFP or pIN233-mCherry products should be identical as the plasmid should not affect the banding pattern.

Table 10 | Primers and PCR cycles used for RAPD-PCR amplification

Target gene	**Primer name**	**Primer sequence (5' to 3')**	**Reference**	
/	RAPD 272	AGCGGGCCAA	(Mahenthiralingam *et al.* 1996)	
PCR Stage		**Temperature** (°C)	**Number of cycles**	**Time (mins)**
Initial heating step		95	1	2
Amplification		95	45	1
Annealing		36		1
Extension		72		2
Final extension		72	1	4

2.13.4 M13 PCR to confirm plasmid uptake

Plasmid uptake into the *B. multivorans* strains was also checked using an M13 PCR on the pIN301-eGFP and pIN233-mCherry electroporated bacterial strains (Table 11). Colonies were picked from the 50 µg/ml Cm supplemented selective TSA agar plates and subjected to rapid Chelex DNA extraction as above (Section 2.3.1). The M13 primers were used to amplify the plasmid inserts into the *B. multivorans* genomes. A 1.5% agarose gel was used to visualize the amplified DNA products. Banding patterns of the genetically manipulated *B. multivorans* strains were compared to the negative control parent strains.

Table 11 | Primers and PCR cycles used for M13 amplification

Target gene	Primer name	Primer sequence (5' to 3')	Product size (bp)	Reference
/	pUC/M13	GTA AAA CGA CGG CCA GT GGA AAC AGC TAT GAC CAT G	1152	/
PCR Stage		**Temperature** (°C)	**Number of cycles**	**Time (mins)**
Initial heating step		93	1	5
Amplification		94	35	0.5
Annealing		56		0.5
Extension		72		1.5
Final extension		72	1	5

2.13.5 Stability of the pIN301-eGFP and pIN233-mCherry plasmids in the *B. multivorans* model strains

Genetic stability of the pIN301-eGFP and pIN233-mCherry plasmids in the *B. multivorans* model strains and *P. aeruginosa* LESB58 strain were performed using week-long growth passage experiment as follows. Freezer stocks (Section 2.1.1) of each strain were revived on TSA supplemented with 50 µg/ml Cm, grown up to 48-hours at 37 ºC. A few colonies were picked from the revived plates and transferred into BSM-G broth with 50 µg/ml Cm and incubated at 37 ºC (150 rpm) for 18 hours. This was known as the 'revival culture'. Dilution of each culture was prepared in 3 ml BSM-G broth with 50 µg/ml Cm to a final count of 10^5 CFU/ml per strain. This was used as the 'day 0' culture and thus the start point for the assay. The day 0 culture was incubated at 37 ºC shaking at 150 rpm for 24 hours before diluting in a 1 ml microtube to around 10^7 CFU/ml in BSM-G broth. 200 µl of each diluted strain was placed into a 96-well plate, with 4 technical repeats per strain. The 96-well plate was placed into a Tecan plate reader and held for 5-minutes in darkness before reading. The cultures were measured for optical density (OD_{600nm}) and fluorescence of either pIN301-eGFP (excitation: 488_{nm}, emission: 522_{nm}) or pIN233-mCherry (excitation: 550_{nm}, emission: 610_{nm}). Each of the strains were enumerated for viability at each time point as per Section 2.1.3. The day 0 culture was further diluted to a final count of 10^5 CFU/ml for each strain, without antibiotic, in 3 ml BSM-G broth. This gave the day 1 culture which was then incubated at 37 ºC shaking at 150 rpm for 24 hours. The dilution passages without antibiotic selection and OD/fluorescence readings were performed each day for a total of 5 days. Three biological replicates were performed for each strain in total.

The data was normalized by first subtracting the media blank from each reading, then dividing the fluorescent reading by the corresponding OD_{600nm} measurement. To confirm that the cells harboured the plasmid over the course of the experiment, the following equation was used:

$$\frac{\frac{CFU}{ml} on\ TSA\ with\ antibiotic}{\frac{CFU}{ml} on\ TSA\ plate}\ x\ 100$$

2.14 Comparative genomics of the B. multivorans strains

2.14.1 Identification key virulence and pathogenicity genes

Two approaches were used to identify the putative virulence and pathogenicity genes in the *B. multivorans* strain panel. Firstly, the gene presence absence matrix from Roary (Page *et al.* 2015) was manually investigated to group the genes into the following categories: ABC transporters, biocide and chemical resistance, biofilm and adhesion, cation efflux, drug resistance, haemolysins, hypermutators, invasion/flagella, LPS & O-antigen, metabolism, metal and mineral resistance, metalloprotease, OMPs, prophage and IS elements, QS genes, stress proteins, T2SS genes, T3SS genes, T4SS genes, T6SS genes, and Toxin-antitoxin systems.

A second approach utilised the Abricate (Seemann 2017) v1.0.0and multiple databases implemented within this as follows: CARD (Jia *et al.* 2017), Resfinder (Zankari *et al.* 2012), PlasmidFinder (Carattoli *et al.* 2014), VFDB (Chen *et al.* 2016), NCBI AMRFinder (Feldgarden *et al.* 2019), ARG-ANNOT (Gupta *et al.* 2014) and MEGARES (Doster *et al.* 2020). The prokka-annotated scaffolded *B. multivorans* contigs were used for these gene prediction analyses. The virulence and pathogenicity genes in each of the *B. multivorans* contigs were predicted using each of the Abricate (Seemann 2017) databases using default parameters. Outputs were combined, with duplicate hits from multiple databases noted. Abricate (Seemann 2017) was then re-run on the *B. multivorans* complete genomes to identify the exact location of the virulence and pathogenicity genes. Circa software was used to develop Circos (Krzywinski *et al.* 2009) circular graphics plots of the complete genomes which illustrated the virulence and pathogenicity genes of interest, their respective categories, and genomic locations.

2.14.2 Identification of Low Oxygen Locus associated genes in the *B. multivorans* genomes

To identify the genes homologous to the *B. cenocepacia* Low Oxygen Locus Activated (*lxa*) locus in the *B. multivorans* strain panel, a custom Abricate database was created using the *lxa* genes from the *B. cenocepacia* J2315 genome (www.*Burkholderia*.com) described by (Sass *et al.* 2013). This was run against the Prokka-annotated unscaffolded *B. multivorans* genomes. The custom *lxa* database was executed in the command line using a cut-off of 80% identity and 80% coverage. The BLAST output was manually adapted to a gene presence-absence matrix for reporting the results.

2.14.3 Identification of prophages

All 77 *B. multivorans* strain panel genomes were subjected to prophage sequence prediction using the PHAge Search Tool Enhanced Release (PHASTER) (Zhou *et al.* 2011; Arndt *et al.* 2016). The Prokka annotated FASTA files were used to compare against the PHASTER database. Phage region were scored as incomplete (< 70), questionable (70-90) and complete (> 90). Results were illustrated as box plots for overall phage metrics and as a graph comparing phage region size and *B. multivorans* genome size.

2.14.4 Insertion Sequence identification

ISEScan v1.7.2.3 (Xie and Tang 2017) was used to identify insertion sequence (IS) elements in the *B. multivorans* panel strains. This used the Prokka-annotated, unscaffolded *B. multivorans* contigs and was run using default settings. The .csv file outputs were used for analysis.

2.14.5 Identification of plasmids

RFPlasmid v0.0.18 (van der Graaf-van Bloois *et al.* 2021) was used to predict plasmid contigs within the Prokka-annotated *B. multivorans* strain panel. This was performed on both the model strain complete genomes and the un-scaffolded contigs. RFPlasmid was used with default parameters, flagging the species as *Burkholderia.* BLAST was then used to look at the similarity between the C4 replicons identified in the scaffolded *B. multivorans* contigs compared to the *B. multivorans* model strains using a local BLAST search (Section 2.7.2). ACT was finally used to compare the predicted replicon 4 (C4) regions of the *B. multivorans* complete genomes. ATCC 17616 C4 region was used as the main comparator against the C4 region of BCC1272, BCC0084, and BCC0033.

2.14.6 EggNOG to identify COGs within the *B. multivorans* genomes

COG functional categories were assigned using the online EggNOG-mapper online tool v2 (Jensen *et al.* 2008) using database v5 (Huerta-Cepas *et al.* 2019). This was performed on the complete *B. multivorans* genomes (Section 4.2.5). All parameters were left at default. Inputs were amino acid translated coding sequences (CDS) .faa files produced by Prokka. The DIAMOND based homology search was used to assign functional orthologs. This used the bacterial database. The output table was used to extract COG categories and their respective accessions. The well-known categories were split into (i) information storage and processing, (ii) cellular processes and signalling, (iii) metabolism. If a COG category was unknown, the CDSs was noted ‘poorly categorized’.

2.14.7 SNP Variant analysis

Snippy (Seemann 2018) v4.6.0 was used to perform variant calling of the *B. multivorans* strains compared to a respective reference. For the lineage analysis, ATCC 17616 and BCC0084 complete genomes were used as the references. When performing variant analysis of the murine modelling infection isolates, the respective BCC0084::pIN301-eGFP, BCC0033::pIN301-eGFP, BCC0033 WT and ATCC 17616 WT insertion sequence short-read references were used. The core SNP outputs were used to create phylogenetic trees (Section 2.9.1) which was visualised and edited as per Section 2.9.4.

2.15 In vivo modelling of the B. multivorans model strains

2.15.1 Ethics statement

All murine modelling experiments were performed at University of Liverpool under project licence PP2072053 (home office approved). All experiments were performed using 6- to 8-week-old BALB/c female mice. The mice were purchased from Charles River (Margate, UK). An independent member of staff randomly assigned the mice into cages of 4. The mice were left to acclimatise for 7-days in the cages before infection. Most procedures and specifically those involving animals were carried out by Angharad Green and Daniel Neill as the licence holders. A group size of 8 mice per time point (1, 3, and 5 days), with a total of 24 mice for each *B. multivorans* strain analysed, was used as standard parameters sufficient for initial statistical power analysis as shown in previous studies (Green *et al.* 2021)

2.15.2 Experimental evolution in murine infection models

B. multivorans strains ATCC 17616 WT, BCC0084::pIN301-eGFP, BCC0033 WT, and BCC0033::pIN301-eGFP were grown as overnight cultures (16-18 hours growth) in TSB from single colonies. The overnight cultures were then sub-cultures in TSA with 20% foetal bovine serum (FBS). The sub-cultured bacteria were grown for ~6 hours to mid-exponential phase, before creating -80 ^{o}C freezer stocks suspensions (Section 2.1.1). The murine infections were performed based on the protocol by Green *et al.* (2021). The freezer stocks acted as the initial infection stocks for the experiments. Infection stocks were gently thawed before use at room temperature. Bacteria were harvested by centrifugation. The bacteria were suspended in Phosphate Buffered Saline (PBS). Twenty-four mice, per *B. multivorans* strain, were infected intranasally with ~10^7 CFU/50 µl bacterial cells. Mice were under light anaesthesia of mixed oxygen and isoflurane during the infection. Mice were monitored for pain scores and their weight every 24-hours of infection. This was performed until the time point where the mouse was due to be culled to ensure there were no breaches of ethics in terms of animal suffering. Groups of 8 mice, per *B. multivorans* infection strain, were culled at 1-, 3- and 5-days post-infection. The nasopharynx and lungs were removed from each of the culled mice and were homogenised in 2 ml sterile PBS. Serial dilutions of the homogenised murine tissues were serially diluted and plated on *B. cepacia* selective agar (BCSA) (Oxoid, UK). Viable cell counts were performed after 48-hours growth at 37 ^{o}C. The bacterial isolates were pooled for each group of 8 mice. This gave one stock for the nasopharynx and one stock for the lung, per time point, for each *B. multivorans* infection strain. Original murine infection and time point stocks were frozen and stored at -80 ^{o}C at Liverpool University.

Each stock was sent to Cardiff University on a charcoal swab. These were revived within 48-hours of receipt on BCSA media. DNA extraction and quality control of each revived stock were performed as per Sections 2.3.1, 2.3.3, and 2.4.1. Extracted DNA was sent to Novogene (Cambridge, UK) for whole

genome sequencing using the Illumina platform. Genome sequences were assembled and checked for quality as per Sections 2.6.1 and 2.6.6. Snippy (Seemann 2018) v3.2-dev was used to perform the variant analysis (Section 2.10.1).

2.16 Phenotypic characterization of the B. multivorans strains

2.16.1 Colony morphology

Bacterial strains were revived, using quadrant streaks, from the -80 °C freezer stocks on TSA and grown for 48-hours at 37 °C for colony morphology analysis. Results were interpretated by size (small, medium, or large), form (circular or irregular), and appearance (dull or shiny). All colony morphologies were performed on 89 *B. multivorans* strains. This was performed on the 73 *B. multivorans* strains sequenced as part of this thesis (Table 5), as well as a further 16 *B. multivorans* strains which are epidemiologically relevant (Table 5).

TSA supplemented with 0.01% (w/v) congo red dye (Chung *et al.* 2003) was used to identify cell surface differences of 89 *B. multivorans* strains. The *B. multivorans* strains were quadrant streaked onto the plates. The strains were scored for colour between pink, orange, and red. This indicated the amount of congo red dye uptake by the strains after 48-hours growth at 37 °C. The scale of congo red dye uptake by the *B. multivorans* strains ranged from pink to orange to red.

Exopolysaccharide production of the 89 *B. multivorans* strains, to evaluate their mucoid status, was performed using yeast extract-mannitol (YEM) agar. The composition of the agar was 0.05% (w/v) yeast extract, 0.4% (w/v) mannitol, and 1.5% (w/v) agar as described by (Sage 1990). *B. multivorans* were quadrant streaked onto the YEM agar and incubated for 48-hours at 37 °C. After incubation, the exopolysaccharide (EPS) production of the *B. multivorans* strains was measured using categorisation by the naked eye. Scoring of the EPS production was as follows: '-' = non-EPS producing, '+' = partially mucoid, '++' = frankly mucoid, '+++' = mucoid, '++++' = highly mucoid.

2.16.2 Growth rates using a Bioscreen C instrument

Growth analysis of the *B. multivorans* and other CF strains was carried out essentially as previously described (Rushton *et al.* 2020). An automated Bioscreen C instrument was used to determine the bacterial growth dynamics of *B. multivorans* isolates. Microtitre plates comprised of 100 wells in a honeycomb shape. Wells contained 200 µl of bacterial overnight culture, diluted to approximately 10^6 *B. multivorans* CFU/ml in either TSB or Artificial Sputum Medium (ASM). The Bioscreen C was set to the following conditions: wideband filter (450-580nm), 37°C incubation and medium shaking 10 seconds before read. Well readings were taken every 15 minutes for a 48-hour duration. Each experiment was performed in triplicate, with 4 technical replicates. Data collected was manipulated using Microsoft Excel. The technical replicates for each experiment were averaged, subtracting the broth only control from each value. The data was then logarithmically transformed ($\log_{10}$).

The growth rates of the pIN301-eGFP and pIN233-mCherry electroporated bacterial strains were determined using the above method. The strains were grown both with and without Cm (50 µg/µl) antibiotic selection to see whether any differences occur. Comparisons were made to the growth of the wild type (WT) strains.

To examine the growth rates of co-culture interactions, 10^6 CFU/ml of *B. multivorans* was mixed 1:1 with 10^6 CFU/ml of secondary pathogen (Table 6) in TSB. A total of 200 µl of mixed bacterial culture was placed into each Bioscreen C well, with 4 replicates per interaction. Growth interactions were measured in the same way as the *B. multivorans* monocultures above.

2.16.3 Data analysis of *B. multivorans* growth parameters

Growth parameters of all mono- and co-culture growth rates were estimated using the Grofit statistical package designed by Kahm *et al.* (2010). This package is compatible with use in R software. The GcFit function was used to estimate the following: I) maximum growth rate, II) length of the lag phase and III) maximum growth density. Measurements were created through parametric models. Any strains that were modelled using non-parametric tools were excluded.

2.16.4 Swimming and swarming motility assays

Motility of *B. multivorans* was measured using a modified method from (Rashid and Kornberg 2000). Agar plates were prepared 24 hours before use to ensure consistent moisture content. Agar was measured using a falcon tube, with each Petri dish containing 20 ml, and dried on an even surface in a laminar flow cabinet for 30 minutes. Varying concentrations were achieved by adding purified agar powder (Oxoid) to either LB broth (Sigma-Aldrich) or Basal Salts Medium (prepared in-house) supplemented with 0.4% glucose (w/v). Swimming plates were prepared to 0.3% agar with LB, with the inoculum through the agar to the base of the Petri dish. Swarming plates were prepared using 0.5% agar in both media types, with the inoculum on the surface only. Plates were wrapped in petri-dish bags to prevent drying. Each experiment was performed in triplicate, with 2 technical replicates per assay. Plates were incubated for 37 °C and zones were measured at 24-hours, averaging two perpendicular measurements. To ensure the data is robust, all biological replicates must be within a 15mm range. Each isolate was assigned a category: non-motile ≤5 mm, low motility 5.1-25 mm, intermediate motility 25.1-50 mm, high motility ≥50.1 mm.

For the mixed motility assays, the primary *B. multivorans* pathogen was mixed with a secondary pathogen from the CF secondary pathogen panel (Table 6). An aliquot of 500 µl overnight culture of each pathogen was placed into a sterile 2 ml microtube and mixed using a vortex for 10 seconds. The motility of the combined pathogens was assessed using the same method as above.

2.16.5 *Agar plugs from the edge of swimming motility zones*

A 6 mm sterile corkborer was used to recover the CF pathogens from the outside of the swimming motility interaction. The cork borer was placed on the outer edge of the motility zone, after incubation. The agar plugs were vortexed for 20 seconds in 1 ml sterile PBS for DNA extraction (Section 2.3.1). This ensured re-suspension of bacteria from the agar plug.

2.16.6 RISA PCR profiling

To identify the presence of a secondary pathogen in the mixed swimming motility, RISA PCR fingerprinting (Flight *et al.* 2015) was used to visualise banding patterns of the bacteria isolated from the agar plugs (Section 2.16.5). The RISA primers used were those defined by Borneman and Triplett (1997). PCR cycle steps and DNA amplification can be found in Table 12. Amplified PCR products were visualized using a 1.2% agarose gel electrophoresis (Section 2.5.2).

Table 12 | RISA PCR fingerprinting primers and PCR stages

Target gene	**Primer name**	**Primer sequence (5' to 3')**	**Reference**	
/	1406F	TGYACACACCGCCCGT	(Borneman and Triplett 1997; Flight *et al.* 2015)	
	23SR	GGGTTBCCCCATTCRG		
PCR Stage		**Temperature** (°C)	**Number of cycles**	**Time (mins)**
Initial heating step		94	1	2
Amplification		94	30	0.25
Annealing		55		0.25
Extension		72		0.75
Final extension		72	1	2

2.16.7 Disk diffusion for antibiotic susceptibility testing

Disk diffusion assays (Bauer *et al.* 1966) were performed using MASTRING-S® M26/NCE antibiotic rings (Mast Group). Assays were performed as per the European Committee of Antimicrobial Testing (EUCAST) guidelines. Mueller-Hinton agar (Oxoid) was used as the base for all bacterial strains tested. Bacterial strains were diluted to approximately 10^8 CFU/ml (OD 0.5 McFarland standard) in Mueller-Hinton broth (Oxoid). A single cotton swab of diluted bacteria was plated onto the agar in two directions, creating a continuous lawn. Plates were left to partially dry for around 30 seconds before the MASTRING-S® M26/NCE antibiotic rings were added to the centre of each plate using sterile forceps. Once placed, the antibiotic rings were not moved. Each antibiotic disc was gently pressed onto the agar with the sterile forceps to ensure contact. Plates were incubated for 24-hours at 37°C before reading. After incubation, two perpendicular measurements were manually taken, using a ruler, for each antibiotic disc, noting the average. The disc diffusion assay was performed using 2 technical and 3 biological replicates per bacterial strain.

2.16.8 Biofilm formation in a 96-well microtitre plate

A crystal violet assay binding plastic plates (O'Toole 2011) was used to determine the mass biofilm formation of *B. multivorans* isolates. Overnight cultures were diluted to an OD0.001 at 600nm (approximately 10^5 cfu/ml) as a starting inoculum. Aliquots of 100 µl suspension were placed into the 96-well plates into columns 2 to 10, rows B to G. This was performed in columns, giving 6 technical replicates. The broth control was put into column 11. The wells on the outer edge of the plate did not contain any suspension or control to avoid evaporation when incubating. Plates were incubated statically for 24-hours at 37°C. After incubation, the medium was shaken out of the plate and washed three times with autoclaved, ultra-pure H20. The H20 was shaken out after each wash. A final washing step was performed where the plate was submerged in H20 and shaken out. Using a multichannel pipette, 200 µl of 0.1% crystal violet was added into each well and left at room temperature for 20 minutes so that the dye is absorbed. The crystal violet was then washed from the wells using the same method as before. Plates were inverted and dried overnight. To quantify the biofilm mass, 200 µl of 70% absolute ethanol was added to each well and left at room temperature for 15 minutes to solubilise the crystal violet. The plates were shaken at room temperature for 1 minute at 500rpm using a Grant-Bio PHMT Thermoshaker. A Tecan plate reader was used to read the absorbance of the wells at 570 nm. A final absorbance reading was determined by subtracting the average of the broth blanks from all other wells on that plate. The absorbance value was taken as the approximation of the biomass present in each well as described (O'Toole 2011).

For the mixed biofilm assays, the above crystal violet biofilm protocol was also used. In the case of a mixed interaction, a *B. multivorans* strain was mixed in a 1:1 (CFU/ml) ratio with a secondary pathogen from the cystic fibrosis community panel. This would be approximately 10^5 CFU/ml of each pathogen as the starting inoculum. 50 µl suspensions of each pathogen was placed into the 96-well plates to give a 100 µl total volume. Staining and quantification of the biofilms were performed as above.

2.16.9 Protease production on agar

Protease production in the *B. multivorans* isolates was assessed using a modified protocol from Morris *et al.* (2012). Lactose-free skimmed milk powder was obtained from Valio (https://www.valio.com). TSA and the lactose-free skimmed milk were made separately for sterilization. TSA was prepared as per manufacturer's instructions. For the lactose-free skimmed milk, a 10% (w/v) solution was prepared in ultrapure deionized water. Both components were autoclaved for 15 minutes at 121 °C and then placed in a water bath at 55 °C until combining. Protease agar was made by mixing the two components to give a final volume of 1% (w/v) lactose-free skimmed milk.

Overnight cultures were diluted to 10^7 CFU/ml in TSB. Aliquots of 10 µl diluted culture were placed onto the lactose-free skimmed milk agar in triplicate. Plates were left on a flat surface to dry before incubating for 18-hours at 37 °C. Protease activity was determined by taking the average of two

perpendicular measurements for each technical replicate. Assays were performed in at least three biological replicates for all bacterial strains.

An additional protease production agar assay was performed with the *B. multivorans* model strains (ATCC 17616, BCC0084, BCC0033, and BCC1272) and *P. aeruginosa* LESB58. This used brain-heart infusion (BHI) agar (Oxoid) in place of TSA. The lactose-free skimmed-milk supplement was added as above. The assay was performed and assessed in the same way as the TSA-based assay.

To assess the protease production of the secondary CF pathogens (Table 6) when interacted with *B. multivorans* strains, the TSA and BHI agar media was prepared as above. All cultures were grown individually overnight (16-18 hours) and then diluted to 10^6 CFU/ml. 500 µl aliquots of *P. aeruginosa* strain was placed into a sterile 1.5ml microtube. 500 µl of either *B. multivorans* or *B. cenocepacia* was then added to the microtube, mixing the two pathogens in a co-culture, creating a total volume of 1 ml. The mono- and co-cultures were inoculated onto the protease agar, incubated, and assessed as above.

Enumeration of viable cells was performed (Section 2.1.3) on the TSA with 1% lactose-free skimmed-milk agar when *B. multivorans* was interacted with *P. aeruginosa* LESB58. After incubation, one of the inoculated spots per plate was scraped, using a sterile spreader, and re-suspended in 1 ml sterile TSB. Serial dilutions were performed down to 10^{-8} and plated onto (i) TSA for total viable counts, (ii) using TSA with 50 µg/ml of PB1 to select for *B. multivorans,* and (iii) using TSA with 50 µg/ml of Tm to select for *P. aeruginosa*. Plates were incubated for 48-hours at 37 ºC before calculating the CFU/ml from each inoculation.

2.16.10 Use of fluorescent reporter strains for viability assessment in co-culture

Bacterial cultures (*B. multivorans* BCC0084::pIN301-eGFP and *P. aeruginosa* LESB58::pIN233-mCherry) were grown overnight (16-18 hours) at 37 ºC (150 rpm) in BSM-G broth supplemented with 50 µg/ml Cm to ensure plasmid stability. The bacteria were diluted to a starting culture of ~10^6 CFU/ml in BSM-G broth. A 1:1 suspension of diluted *B. multivorans* BCC0084::pIN301-eGFP and *P. aeruginosa* LESB58::pIN233-mCherry was performed in an microtube. The bacterial cultures, both alone and in co-culture, were placed into a microtitre plate, giving 200 µl per well. The microtitre plate was covered with a sterile transparent lid and placed in a Tecan plate reader. Plates were held for 5 minutes in the dark before the first reading was taken. Growth was measured every 30 minutes for a 24-hour period. The plate was subject to shaking before each reading was taken. OD_{600nm} readings were taken at each time point, immediately followed by a fluorescence reading for both pIN301-eGFP (excitation: 488_{nm}, emission: 522_{nm}) and pIN233-mCherry (excitation: 550_{nm}, emission: 610_{nm}). This experiment was performed with 4 technical replicates per plate and 3 biological replicates. The mean was calculated for each time point from the 4 technical replicates. Data was normalized to give the RFU as per Section 2.13.2. The normalization was performed against both monocultures individually.

Enumeration of viable counts was performed using TSA with and without antibiotic selection (50 µg /ml Cm) at time point 0 and 24-hours (Section 2.1.3).

2.16.11 Quorum Sensing

E. coli bioreporter strains (Winson *et al.* 1998) harbouring *lux*-based bioluminescence were used to detect AHLs in the bacterial strains. The reporter plasmid pSB406 was used to measure C_4-HSL and C_6-HSLs and pSB1142 was used to measure 3-oxo-C_{12}-HSLs through the emission of light. Overnight cultures (16-18 hours growth at 37 ºC with 150 rpm shaking) of *E. coli* pSB406 (supplemented with 50 µg/ml ampicillin), *E. coli* pSB1142 (supplemented with 5 µg/ml tetracycline), and the experimental bacterial strains (without supplement) were grown. The cultures were then diluted to ~10^6 CFU/ml in BSM-G. For single bacterial cultures, 100 µl of each strain was mixed 1:1 with 100 µl of appropriate *E. coli* bioreporter strain. For the *B. multivorans* and *P. aeruginosa* mixed culture analysis, 100 µl of *E. coli* bioreporter strain was mixed with 50 µl of *B. multivorans* and 50 µl of *P. aeruginosa.* Total volumes of 200 µl mixed, incubated bacterial culture were loaded into a 96-well black walled, clear-bottom plate (Greiner Bio-one). Plates were incubated at 37 ºC (150 rpm) for 30 minutes before reading. A Tecan plate reader was used to measure the OD_{600nm} and RLU of each well. Technical replicates were averaged to find the mean for each of the 3 biological repeats. Luminescence of the strains was recorded as RLU by normalising against the OD_{600nm} reading. The *E. coli* bioluminescence reporter strains were used as a negative control, but the overall RLU readings were not subtracted from the experimental *B. multivorans* and *P. aeruginosa* strains.

3 Genome sequencing and phylogenomic characterization of a *B. multivorans* strain collection

3.1 Introduction

For certain pathogens there is great variation and genetic diversity in their genomes due to their nature of existing in complex populations. Genetic differences within the population are largely due to recombination in non-clonal groups (Spratt and Maiden 1999). For instance, *P. aeruginosa* has been observed to be highly recombinant in genetic structure causing population diversity (Darch *et al.* 2015). However, clonal microbial populations may still differ in phenotype and genotype (Coutinho *et al.* 2011b; Andam 2019). This diversification can be through means of allelic and genome content, including horizontal gene transfer (Croucher *et al.* 2014).

Genome evolution and population biology within the BCC has been characterized in several studies. Clonality has been observed in *B. multivorans* (Baldwin *et al.* 2008; Hassan *et al.* 2020), with observed clonal complex changes largely due to intra- and interspecies recombination (Baldwin *et al.* 2008). There is also causation of lineage diversification with evidence of clonal *B. multivorans* mutation due to enhanced selective pressures in the CF lung (Hassan *et al.* 2020). Positive selection pressures on the BCC genomes can arise due to the stressful CF lung environment, causing clonal variants with varying phenotypic traits (Madeira *et al.* 2013).

Several clonal complexes have been observed in *B. multivorans* and other *Burkholderia* species. Grouping of STs into clonal complexes is defined as having 5 or more shared MLST loci (Baldwin *et al.* 2008). *B. multivorans* is widespread globally, with examples of 4 clonal complexes (CC1, CC4, CC5 and CC6) found amongst various countries and continents. These include the United Kingdom, USA, France, Portugal, Belgium, New Zealand, and Australia. CC3 is a clinically associated clonal complex, with outbreaks found in the Czech Republic and France (Baldwin *et al.* 2008).

Furthermore, phylogenomic and clonal correlations of other significant CF pathogens, *B. cenocepacia* and *P. aeruginosa* have been documented. *B. cenocepacia* has two defined lineages (IIIA and IIIB) originally found via sequence analysis of the single recA gene (Mahenthiralingam *et al.* 2000; Vandamme *et al.* 2003). The biggest difference between these two *B. cenocepacia* genomic lineages is that IIIA holds multiple MLST types associated with clinical sources and epidemic spread of CF strains, with environmental strains being rare (Drevinek and Mahenthiralingam 2010). This IIIA lineage has also shown to cause more virulent CF infections compared to other members of the BCC (Mahenthiralingam *et al.* 2001; Zlosnik *et al.* 2015; Scoffone *et al.* 2017), including multiple epidemic outbreaks (Jones *et al.* 2004; Drevinek *et al.* 2005). In analogous fashion, *P. aeruginosa* has also adapted and evolved to form two distinct genomic groups (Ozer *et al.* 2019; Weiser *et al.* 2019), with at least

two sub-groups and various clonal lineages (Freschi *et al.* 2019). These two bacteria are comparable to *B. multivorans* since they all cause chronic CF lung infections.

Four lineages (genomovars IIIA, IIIB, IIIC and IIID) comprise the *B. cenocepacia* population (Wallner *et al.* 2019; Hassan *et al.* 2020; Jin *et al.* 2020), which are thought to correspond to genetic and phenotypic adaptations in the CF lung (Coutinho *et al.* 2011a). IIIA was one of the most common and virulent of the *B. cenocepacia* lineages, accounting for several epidemic outbreaks (Drevinek and Mahenthiralingam 2010) and isolation bias from clinical samples (Baldwin *et al.* 2007). Outbreaks of cluster IIIA in Canada (Speert *et al.* 2002) and Europe are often associated with the so-called ET-12 lineage (Sun *et al.* 1995b; Chen *et al.* 2001; Mahenthiralingam *et al.* 2005), an d other epidemic strains such as RAPD types 01, 04 and 06 and the CZ1 clone (Drevinek *et al.* 2005; Drevinek *et al.* 2010). The ET-12 lineage is a widely studied and characterised strain (Chen *et al.* 2001). Studies have identified its epidemic strain marker (Mahenthiralingam *et al.* 1997), type II cable pili (*cblA*) expression for use in epithelial cell attachment (Sajjan *et al.* 1995), and association of the ET-12 lineage and prophages (Bodilis *et al.* 2018). Like IIIA, IIIB also harbours mostly clinical isolates (Drevinek and Mahenthiralingam 2010) which has epidemic clones within the USA (Midwest clone) (Coenye and LiPuma 2002) and Europe (PHDC strain) (Chen *et al.* 2001). In contrast to lineage IIIA, *B. multivorans* was associated with a reduced rate of transmission between CF individuals and decreased mortality rate (Mahenthiralingam *et al.* 2001).

Initially, PCR assays were developed to identify "*B. cepacia*" using reference culture collection DNA sequences, unknowingly ignoring the multitude of species which comprise the BCC (O'Callaghan *et al.* 1994; Campbell *et al.* 1995; Karpati and Jonasson 1996). Methods have since developed, and the species-specific *recA* gene is now used for PCR-based identification of the BCC (Coenye *et al.* 2001b). Research found that there were higher similarities when comparing different members of the BCC using the 16S rRNA gene in comparison to comparisons against other *Burkholderia* species. (Coenye *et al.* 2001b). Using both the 16S rRNA and recA genes, Mahenthiralingam *et al.* (2000) developed novel diagnostic approaches for several of the BCC, creating species-specific PCR primers for *B. multivorans*.

To further evolve *Burkholderia* species identification methods, MLST can be used as a more effective method than *recA* gene analysis. MLST typing is an established way of determining bacterial population structures and epidemiology in the BCC (Baldwin *et al.* 2005; Spilker *et al.* 2009). A study by Cesarini *et al.* (2009) showed that *recA* gene identification of BCC isolates was around 64.5%, increasing to 88.5% identification using MLST. The latter method uses amplification of 7 BCC housekeeping genes (*atpD, gltB, gyrB, recA, lepA* and *phaC*) (Baldwin *et al.* 2005; Spilker *et al.* 2009) which determines the allelic profile and MLST type of the strain (Jolley and Maiden 2010). These genes are reliable in the BCC for sequence typing due to their low mutation rates (Waine *et al.* 2007). MLST analysis was also further developed into ribosomal MLST (rMLST) and whole genome MLST (wgMLST).

Ribosomal MLST uses the 53 genes which encode the bacterial ribosomal subunit, useful for identifying gene variants (Jolley *et al.* 2012) and drawing phylogenetic relationships to better analyze clades across multiple bacterial species compared to the standard MLST scheme (Depoorter *et al.* 2016). One step further uses wgMLST (Maiden *et al.* 2013), taking whole genome sequence data and translating the bases into genes with character data, which can be performed *in silico* (Jolley and Maiden 2010). This method not only confirms the bacterial species ST, but can also identify clonality and SNPs in the population (Hassan *et al.* 2020).

As science is progressing, next generation sequencing is now more cost-effective for many research studies, and genome sequences are becoming more publicly available. This allows large scale genomic analyses as well as high throughput bioinformatic analyses to characterize multiple strains and collections of bacteria. The use of DNA-DNA hybridization (DDH) performed on extracted bacterial DNA was the determinator for the novel species *B. multivorans*, differentiating from the original *B. cepacia* genomovar II name in 1997 (Vandamme *et al.* 1997). For the determination of a single bacterial species, ANI has been proposed as an *in silico* method to replace DDH (Rosselló-Mora 2005). To determine a bacterial species, DDH has a cut-off at 70% relatedness, whereas ANI has a threshold of 95% (Goris *et al.* 2007). BLAST was originally used a the method of determining ANI, however multiple bioinformatics tools have advanced analyses through packages such as PyANI (Pritchard *et al.* 2016) and FASTANI (Jain *et al.* 2018).

As of April 2022, there are 678 *B. multivorans* genome sequences publicly available on NCBI. Soil isolate *B. multivorans* ATCC 17616 (Stanier *et al.* 1966) was the first publicly available *B. multivorans* complete genome released in 2007 and is now used as the well-characterized reference strain. Of the public *B. multivorans* sequences, there are 42 complete genomes, 8 chromosome assemblies, 244 scaffolded assemblies and 384 contig assemblies. Draft genomes are 'open' with sequence gaps, and incomplete in terms of organism genomic content. Contigs and scaffolded assemblies are known as draft genomes (Land *et al.* 2015). Complete genomes, on the other hand, do not have sequence gaps in the genome, representing all genomic content in the organism. This is also referred to as a 'closed' genome (Mardis *et al.* 2002). The median total length (Mb) of the available *B. multivorans* genomes is 6.44681 Mb, median protein count is 5,750 and median GC content is 67.1%.

The BCC also has a unique multi-replicon structure, between two and four circular replicons (Sousa *et al.* 2017). There are three primary replicons (C1, C2 and C3) as well as plasmids (Lessie *et al.* 1996). WGS can be used to identify genes associated with the pan, core, and accessory genomes (Costa *et al.* 2020) in the BCC bacteria, also link how these genetic components are associated with specific replicons. The core genome represents the genes in which are present in all strains examined, with the accessory genome only present in few strains or groups (Tettelin *et al.* 2005). The large multi-replicon genomic content and genetic plasticity of the BCC (Madeira *et al.* 2013) essentially underpins their

ability to adapt to stressful environments such as the CF lung (Harrison 2007; Coutinho *et al.* 2011b; Döring *et al.* 2011).

3.1.1 *Aims and objectives*

This overall aim of this Chapter was to gain genomic insights into the population biology of *Burkholderia multivorans* through the following objectives:

1) To select a genetically diverse collection of *B. multivorans* strains for short-read Illumina whole genome sequencing, with the predominant focus of the collection being strains from CF lung infections
2) To assemble, perform quality control and verify the *B. multivorans* genomes
3) To define the population biology of *B. multivorans* using phylogenomic analyses
4) To identify model CF *B. multivorans* strains for use in further analyses

Hypothesis 1: The population of *B. multivorans* will split into multiple groups

Hypothesis 2: Model CF *B. multivorans* strains can be identified through their phylogenomic placement

3.2 Results

3.2.1 *Short-read DNA extraction, sequencing, and genome assembly*

74 *B. multivorans* strains were revived from freezer stocks held at Cardiff University. DNA extraction was performed using either a Maxwell® automated instrument or the Promega Wizard Kit described in Section 2.3.3. DNA extractions were subjected to short-read sequencing either internally at the Cardiff University genome hub or externally via commercial sequencing by Novogene (Cambridge, UK). Strain BCC0302 did not sequence correctly and was removed from the analyses prior to genome quality control and assembly. The remaining 73 *B. multivorans* strains were processed through adapter trimming, genome assembly using Unicycler and annotation via Prokka (Section 2.6.3), before being used in the genomic and phylogenomic analyses.

3.2.2 Basic features of the 73 *B. multivorans* genomes

3.2.2.1 Genome quality metrics

The full QC statistics of the 73 *B. multivorans* genomes sequenced as part of this thesis are shown in Appendix Table 43. The QUAST assembled contigs (Section 2.6.8) provided genomes which were between 6.02-7.1 Mb in size, with an average of 6.5 M*B.* Mean GC content was 67.2%. RASTk predicted between 5975 and 7374 CDS (mean = 5848.7), 359 and 375 subsystems (mean = 466), 56 and 74 tRNAs (mean = 68.7), and 43 and 67 RNAs per genome (mean = 3.0) (Appendix Table 43). Figure 5 details the genome size (bp), GC content (%) and number of CDS for each genomic lineage of *B. multivorans* as defined in Sections 3.2.4.1, 3.2.4.2 and 3.2.5. Lineage 1 *B. multivorans* isolates had

a significantly (p = 0.0002) smaller mean genome size (6.4 Mb) and significantly lower (p = 0.0163) number of CDS (5743.6) compared to lineage 2a (6.7 Mb; 6050.4 CDS) and 2b (6.6 Mb; 5854.5 CDS) (Figure 5). In terms of GC content, all three lineages had approximately the same percentage as the entire dataset (all non-significant comparisons). Lineage 1, sub-lineage 2a and sub-lineage 2b have a GC content (%) of 67.1%, 67.07% and 67.2% respectively (Figure 5, Appendix Table 43). The mean number of contigs in the 73 genomes was 107, ranging from 52 to 305 contigs. The N50 values for the genomes varied from 92055 (BCC0082) to 640560 (BCC0225), with a mean of 269407 (Appendix Table 43). These figures were comparable to the master *B. multivorans* set of 283 genomes, which included those downloaded from public databases (Section 2.1.1; Table 5). In the 283-genome panel, the mean GC content was 67.04%, mean sequence length 6.5 Mb, and mean N50 value was 338304. For the 283 *B. multivorans* annotations, the mean number of CDS was 5814.7, mean rRNAs was 4.3, and mean tRNAs was 64.6. This provided confidence that the 73 genomes sequenced as part of this work were representative of standard *B. multivorans* genomes, including those already publicly available.

The genome statistics were also analysed based on the isolate source (Table 13). There was the largest variation in sequence size (bp) when analysing the CF isolates, with a range from 6.2 Mb to 7.1 M*B.* Non-CF infection isolates had a similar mean sequence size (6.4 Mb) compared to CF isolates (6.5 Mb), with a difference of 0.1 M*B.* The isolate source with the smallest sequence size were environmental strains isolated from a clinical/hospital environment (ENVH) with a mean size of 6.2 M*B.* The mean GC content (%) of all groups was relatively similar, around 67%. This ranged from 67.1% in the CF isolate group to 67.5% in the environmental group. Greater variation between the groups was seen in the number of CDS. CF isolates had the highest number of CDS with a mean of 5872.4. The group with the smallest mean average of CDS was the clinical environmental isolates at 5579.0 CDS. The dataset contains a greater number of CF isolates compared to other types and therefore must be taken at face value for these results.

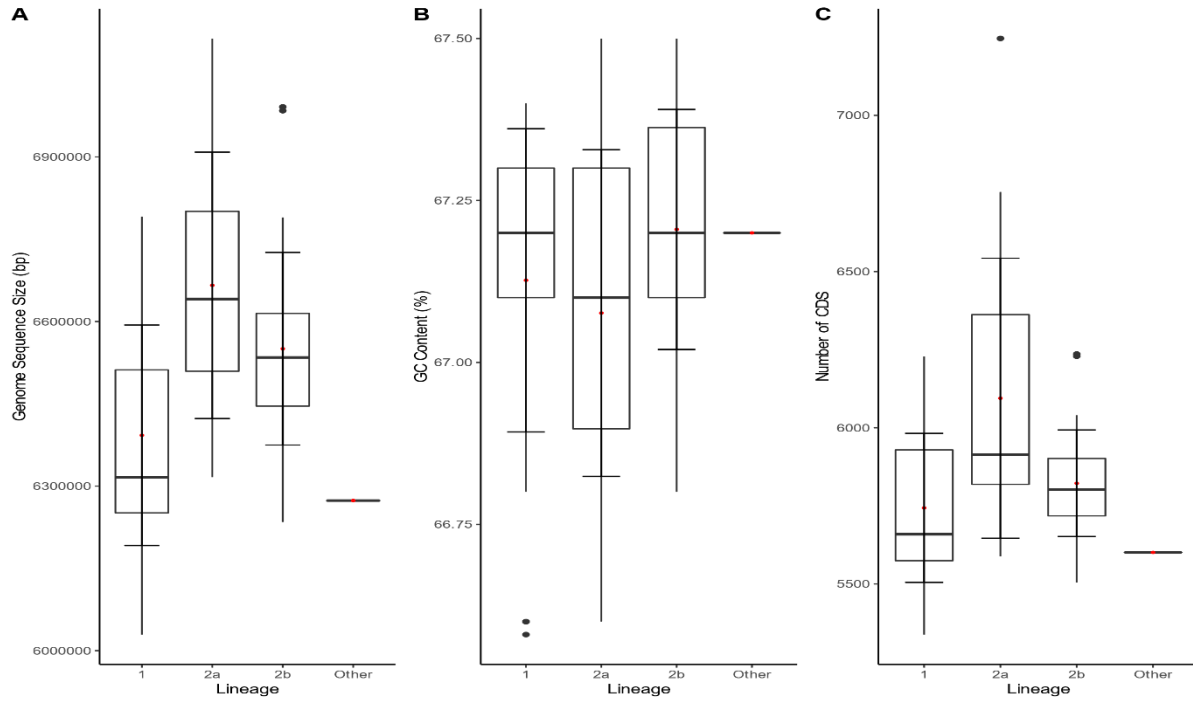


Figure 5 | The genomic properties of the 73 *B. multivorans* genomes sequenced as part of this thesis. **(A)** *B. multivorans* genome sequence size (bp), **(B)** GC Content (%) and **(C)** Number of CDS. All box plots are split by genomic lineage. Error bars show the 95% confidence intervals. The middle line for each plot represents the median average. Outliers in the dataset are represented by single, round points. The red point on each box plot represents the mean average of each lineage. The 'other' lineage contains single strain *B. multivorans* BCC1368. Lineage 1 comprised of 25 genomes and lineage 2 comprised of 47 (2a: n = 18, 2b: n = 29) genomes. Statistical significance was observed for the genome size (bp) between lineage 1 and lineage 2a (p = 0.005) and lineage 1 and lineage 2b (p = 0.0051). There was also significance in the number of CDS between lineage 1 and lineage 2a (p = 0.0049).

Table 13 | Genome quality statistics of the *B. multivorans* strains (*n* = 73) separated by isolate type.

Isolate Type	Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
Genome Sequence Size (bp)						
CF	6.2	6.4	6.5	6.5	6.6	7.1
CGD	6.3	6.3	6.3	6.3	6.3	6.3
NON	6.0	6.2	6.5	6.4	6.7	6.7
ENV	6.4	6.4	6.4	6.4	6.4	6.4
ENVH	6.2	6.2	6.2	6.2	6.2	6.2
GC Content (%)						
CF	66.58	67.10	67.17	67.13	67.30	67.50
CGD	67.20	67.20	67.20	67.20	67.20	67.20
NON	66.93	67.10	67.30	67.23	67.40	67.40
ENV	67.50	67.50	67.50	67.50	67.50	67.50
ENVH	67.32	67.32	67.32	67.32	67.32	67.32
Number of CDS						
CF	5472	5677	5830	5872.4	5975	7246
CGD	5618	5618	5618	5618.0	5618	5618
NON	5338	5531	5793	5689.8	5889	5898
ENV	5606	5606	5606	5606.0	5606	5606
ENVH	5579	5579	5579	5579.0	5579	5579

CF = cystic fibrosis, CGD = chronic granulomas disease, NON = non-CF infection, ENV = environmental, ENVH = environmental from a clinical setting. Statistics performed in R Studio using summary function.

3.2.3 Improving genomic resolution by mapping of sequence reads to a reference genome

3.2.3.1 Mapping of reads to a complete reference genome

Minimap2 (Li 2018) was used to map the 73 *B. multivorans* reads to the well characterised ATCC 17616 strain (FASTA sequence obtained from NCBI). This was to verify the contigs sequenced in all individual genomes assigned correctly to *B. multivorans.* An adapted script was used to provide outputs of both mapped and unmapped short reads. Mapped reads were then converted to .bam files for statistical analysis (Appendix Table 44). Outputs detailed the total reads, mapped reads, the number of reads on the forward and reverse strands, duplicates, failed QC reads, proper paired end reads, cases where both pairs mapped to the reference, and singletons (Appendix Table 44). No reads failed QC or were duplicated in the analysis.

On average 86.23% mapped to the *B. multivorans* ATCC 17616 reference, ranging from 80.16% (BCC0921) to 98.53% (BCC1272; an isolate of the same MLST sequence type, ST-21, as the reference genome). Six strains had >90% contigs mapped to ATCC 17616. The number of 'proper paired' reads and both paired reads being mapped was 82.12% and 84.44%, respectively. There was variation in the number of singletons, from 0.40% to 2.90%, with a mean of 1.79%. The strain with the greatest similarity was BCC1271 at 96.55% mapped reads, 95.12% proper paired reads and 0.40% singletons. Strain BCC1271 also possessed the same MLST (ST-21) as the reference genome (Table 16). Only 70-

90% of reads were observed to align to the reference genome meaning that unmapped reads could be either contaminants, sequence differences between strains, or low-quality reads (Sangiovanni *et al.* 2019).

3.2.3.2 Highlighting unmapped reads in the *B. multivorans* genomes

Unmapped reads (Section 3.2.3) from Minimap2 (Li 2018) were subjected to further analysis to identify whether they were of *B. multivorans* origin. This was performed using a similar script for identification of the mapped reads. Samtools view was used for this with the -h flag to indicate a header in the output as well as -f 4 to give the unmapped reads only (Li *et al.* 2009). The same process was applied to the unmapped reads, giving a final output in FASTA format. PyANI (Section 2.8.1) was performed to quickly check the unmapped reads against the complete *B. multivorans* ATCC 17616 genome.

The Minikraken database from Kraken2 (Wood *et al.* 2019) (Section 2.6.6) was used to identify any contamination of the unmapped reads. Unmapped reads which were identified by Kraken to be of *Burkholderia multivorans* origin were noted to be either genomic sequence differences or low-quality reads, which caused a 'misalignment' to the reference genome (Schbath *et al.* 2012). Studies have shown that genomes with <5% contamination are suitable for processing (Bowers *et al.* 2017). All genomes in this study met the threshold requirements.

The average number of unmapped reads allocated to the *Burkholderia* genus was 508680.6 (53.4%). When looking at *B. multivorans* species level, the average number of unmapped reads decreased to 342101.1 (35.8%). They are both lower than the average number of unclassified reads (380425.8; 42.4%) and hits to other *Burkholderia* species (57339.0; 5.8%). The unmapped reads which were not of *B. multivorans* origin were largely hits against closely related *B. cenocepacia,* as well as *B. vietnamensis, B. ubonensis, B. pseudomultivorans* and *B. cepacia.* As the results of these hits were relatively minimal, these were regarded as incorrect species classification. When there is a high percentage of genomic identity between bacterial species, Kraken 2 may only output correct classification to genus level (Wood *et al.* 2019).

3.2.4 Confirmation that *B. multivorans* is a single genomic species

3.2.4.1 PyANI

The Minikraken database (Section 2.6.6) was also used on the assembled draft genomes (n = 73). This confirmed that all genomes sequenced in this project were *B. multivorans* and suitable for analysis. Further to the 73 *B. multivorans* genomes sequenced in this thesis, a further 210 *B. multivorans* genomes were downloaded from NCBI to create a 'master' panel of 283 genomes. Each of the additional 210 genomes were checked to ensure they were *B. multivorans,* and overall, this additional collection had

genomic metrics of size (mean = 6.47 Mb), GC content (mean = 67.04%), tRNAs (mean = 64.6), rRNAs (mean = 4.3), and predicted CDS (mean = 51814.7) that match the 73 strain sequences obtained in this study (see Section 3.2.2). Overall, this analysis gave confidence in both the collection of isolates sequenced in this study and those that they were compared to from NCBI.

The total collection of 283 *B. multivorans* genomes were subjected to ANI analysis using alignment-based PyANI (Section 2.8.1) (Pritchard *et al.* 2016). The comparative percentage identity output of the 283 genomes gave a strong indication of at least two genomic ANI groups (Figure 65), which were defined as group 1 and group 2. It was clear that lineage 2 could be further broken into at least two sub-groups. The data was skewed by the CF170 strain, which comprised 111 CF isolate genomes from a single infection in Canada (Diaz Caballero *et al.* 2018). PyANI was subsequently applied to the 77-genome strain panel (Section 3.2.5.2), which once again gave a clear picture of the two genomic groups. With the absence of the 111 CF170 strains, and only one representation in place, the two groups with sub-groups 2a and 2b were also more clearly defined (Figure 6). The numerical output of PyANI was investigated for isogenic isolates, highlighting a percentage identity of 0.999% between BCC1147 and BCC1148, and between ATCC 17616 and BCC1272. These two sets of genomes also sat on the same phylogenetic branch in as described below (Section 3.2.5) and hence from now onwards will be described as genomic lineages.

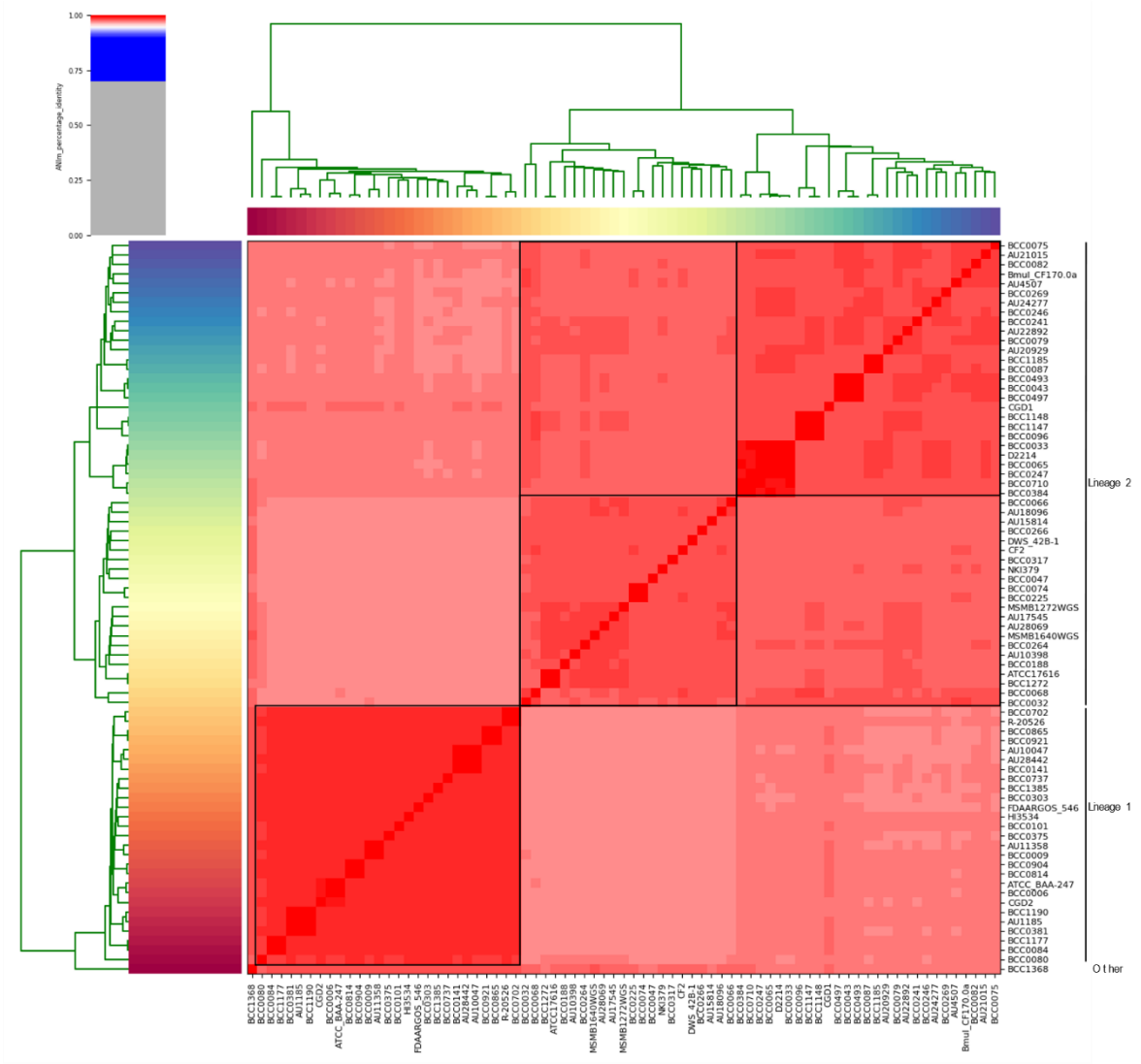


Figure 6 | Average Nucleotide Identity (ANI) of the *B. multivorans* strain panel (*n* = 77). PyANI (Pritchard *et al.* 2016) was used to create the heatmap. The scale of red represents the genomic relatedness of the *B. multivorans* strains. The higher the ANI, the darker the red colour on the heatmap figure. PyANI calculates ANI based on a pairwise analysis between the genomes. No reference genome has been used to create this figure. The two genomic groups have been highlighted on the heatmap using black outlines. Group 1 can be seen in the bottom left of the figure and group 2, with respective sub-groups 2a and 2b highlighted in the top right of the figure. Since the phylogenomic analysis also supported these ANI genomic groups, the two *B. multivorans* lineages have also been noted on the right-hand side of the figure, encompassing the genomes within that specific group. Inkscape was used to edit the figure to add the details.

3.2.4.2 FastANI

FastANI (Jain *et al.* 2018) was used as an alignment-free computational approach to both confirm *B. multivorans* classification of the sequenced isolates and performed species-level comparison between the *B. multivorans* genomes (Section 2.8.1). FastANI was first performed on all 283 *B. multivorans* compared to the ATCC 17616 strain. All genomes had an average ANI 97.8% using this approach, with

a range from 96.8% to 100% identity. This suggested that all the genomes were from the same bacterial species, well above the 95% ANI cut-off defined by Jain *et al.* (2018), and fitting in with recent analysis of 4,000 *Burkholderia* sensu lato genomes (Mullins and Mahenthiralingam 2021).

A secondary analysis was performed on the smaller *B. multivorans* strain panel (n = 77) (Section 3.2.5.2) using the complete PacBio genomes of the identified model strains BCC0033, BCC0084 and BCC1272 (Section 3.2.8) as comparators. Table 14 details the closest reference genome to each query strain and its respective lineage. All strains had a mean ANI of 98.59%, ranging from 97.24% to 100.00%. Query genomes represented in lineage 1 all had the closest ANI to BCC0084, the lineage 1 representative model strain. The average ANI for lineage 1 was 98.90%, ranging from BCC0080 at 98.6% to BCC1177 at 100%. BCC1368, which represents a potential 3rd lineage, denoted 'other', also has the closest match to BCC0084 with 98.1% ANI. All lineage 2 genomes matched closest with one of the lineage 2 model strains: BCC0033, ATCC 17616 and BCC1272. The only difference was that two genomes (BCC0047 and CGD1), which reside in sub-lineage 2a had a slightly higher ANI to the BCC0033 reference for sub-lineage 2b (2 of 48 cases (4.2%)) (Table 14). The average ANI for the whole of lineage 2 is 98.43%, 0.47% lower than the average for lineage 1. When separating lineage 2 into sub-lineages, the mean ANI of lineage 2a decreased to 98.24% whilst lineage 2b slightly increased to 98.54% (Table 14). This indicates a closer genomic ANI relatedness of lineage 1 in comparison to lineage 2, with sub-lineage 2a being the most distantly related in terms of their evolution.

Table 14 | FastANI analysis of the 77 *B. multivorans* genomes against the reference genome with the closest ANI value.

Query Genome	Query Lineage	Reference Genome	Reference Lineage	ANI value
ATCC BAA-247	1	BCC0084	1	98.8461
AU10047	1	BCC0084	1	98.8781
AU11358	1	BCC0084	1	98.8431
AU1185	1	BCC0084	1	98.8149
AU28442	1	BCC0084	1	99.0271
BCC0006	1	BCC0084	1	98.871
BCC0009	1	BCC0084	1	98.8858
BCC0080	1	BCC0084	1	98.5715
BCC0084	1	BCC0084	1	99.9994
BCC0101	1	BCC0084	1	98.8802
BCC0141	1	BCC0084	1	98.8901
BCC0303	1	BCC0084	1	98.8422
BCC0375	1	BCC0084	1	98.8465
BCC0381	1	BCC0084	1	98.9797
BCC0702	1	BCC0084	1	98.947
BCC0737	1	BCC0084	1	98.9285
BCC0814	1	BCC0084	1	98.9406
BCC0865	1	BCC0084	1	98.796
BCC0904	1	BCC0084	1	98.9555
BCC0921	1	BCC0084	1	98.7793
BCC1177	1	BCC0084	1	99.9951
BCC1190	1	BCC0084	1	98.7956
BCC1385	1	BCC0084	1	98.7605
CGD2	1	BCC0084	1	98.7675
FDAARGOS_546	1	BCC0084	1	98.7088
HI3534	1	BCC0084	1	98.8621
R-20526	1	BCC0084	1	98.9732
AU10398	2a	ATCC 17616	2a	98.3835
AU15814	2a	ATCC 17616	2a	98.1344
AU17545	2a	ATCC 17616	2a	98.3403
AU18096	2a	ATCC 17616	2a	98.3027
AU28069	2a	ATCC 17616	2a	98.4635
BCC0074	2a	ATCC 17616	2a	98.1891
BCC0188	2a	ATCC 17616	2a	98.3647
BCC0225	2a	ATCC 17616	2a	98.112
BCC0264	2a	ATCC 17616	2a	98.4597
MSMB1272WGS	2a	ATCC 17616	2a	98.3724
BCC0066	2a	BCC1272	2a	98.3289
BCC0317	2a	BCC1272	2a	98.1395
BCC1272	2a	BCC1272	2a	99.9987
CF2	2a	BCC1272	2a	98.1115
DWS_42B-1	2a	BCC1272	2a	98.1534
MSMB1640WGS	2a	BCC1272	2a	98.4075
NKI379	2a	BCC1272	2a	98.1411
BCC0266	2a	BCC1272	2a	98.1612
BCC0047	2a	BCC0033	2b	97.8901

CGD1	2a	BCC0033	2b	98.1077
AU20929	2b	BCC0033	2b	98.3281
AU21015	2b	BCC0033	2b	98.426
AU22892	2b	BCC0033	2b	98.2999
AU24277	2b	BCC0033	2b	98.4523
AU4507	2b	BCC0033	2b	98.271
BCC0032	2b	BCC0033	2b	98.1766
BCC0033	2b	BCC0033	2b	99.9955
BCC0043	2b	BCC0033	2b	98.3437
BCC0065	2b	BCC0033	2b	99.9759
BCC0068	2b	BCC0033	2b	97.9929
BCC0075	2b	BCC0033	2b	98.3374
BCC0079	2b	BCC0033	2b	98.2695
BCC0082	2b	BCC0033	2b	98.3136
BCC0087	2b	BCC0033	2b	98.447
BCC0096	2b	BCC0033	2b	98.3197
BCC0241	2b	BCC0033	2b	98.3153
BCC0246	2b	BCC0033	2b	98.4302
BCC0247	2b	BCC0033	2b	99.9567
BCC0269	2b	BCC0033	2b	98.5281
BCC0384	2b	BCC0033	2b	99.5843
BCC0493	2b	BCC0033	2b	98.3415
BCC0497	2b	BCC0033	2b	98.3633
BCC0710	2b	BCC0033	2b	99.5928
BCC1147	2b	BCC0033	2b	98.2909
BCC1148	2b	BCC0033	2b	98.2589
BCC1185	2b	BCC0033	2b	98.4796
CF170.0a	2b	BCC0033	2b	98.3007
D2214	2b	BCC0033	2b	99.9934
BCC1368	Other	BCC0084	1	98.1373

3.2.4.3 ANI statistical comparisons between the *B. multivorans* lineages

To confirm the two-lineage split ANI findings, the mean and median ANIs were calculated for the lineages separately using R Studio. The data from the FastANI output was used, taking ANI values of the *B. multivorans* panel strains when compared to the respective reference genomes. BCC1368 and the reference genomes were excluded in the ANI comparisons. This is because BCC1368 does not reside within one of the two defined lineages, and thus has been noted as 'other'. The reference genomes were also excluded from the comparison as they would skew the dataset, having an ANI of 100% with themselves. Therefore, a final panel of 73 genomes have been used as part of the analysis. Statistical analysis comparing the two lineage ANI groupings provided a chi-squared test result of 30.5762. Lineage 1 had a significant difference of $p = <0.0001$ when comparing ANI to both sub-lineages 2a and 2*B*. There was a significance level of $p = 0.0202$ when the ANI of sub-lineage 2a was compared to the ANI of 2*B*. Overall, this highlights the significance of the *B. multivorans* ANI lineage groupings, affirming the idea of at least two secondary groups within lineage 2.

Table 15 | Summary statistics of the Average Nucleotide Identity (%) of the *B. multivorans* lineages (all strains: n=73). Lineage 1: n=25, Lineage 2: n= 47 (Lineage 2a: n=18, Lineage 2b: n=29).

Lineage		Minimum	Maximum	Median	Mean	1st Quartile	3rd Quartile
	All strains	**97.24**	**100.00**	**98.43**	**98.59**	**98.30**	**98.85**
Main Lineages	Lineage 1	98.57	100.00	98.86	98.90	98.80	98.93
	Lineage 2	97.24	99.99	98.32	98.43	98.19	98.43
Sub-lineages	Lineage 2a	97.87	98.47	98.25	98.24	98.15	98.34
	Lineage 2b	97.24	99.99	98.34	98.54	98.29	98.45

*BCC1368 from 'other' lineage not included in main or sub-lineage statistics

**ANI was assessed based on comparison to the most appropriate reference genome

3.2.4.4 Orthologous gene hits

The output of FastANI (Jain *et al.* 2018) gave the number of orthologous gene hits compared to the total number of genes scanned (Appendix Table 45). These ortholog genes are paramount in understanding genomic evolution, related by speciation events, where a gene has descended from a common ancestor (Koonin, 2005). The average percentage of orthologous genes, when comparing the draft genomes to all the reference genomes, was 90.1% (Appendix Table 45). The results were then filtered to only show the highest scoring ortholog match between the draft genome and one respective reference genome. For lineage 1, the percentage of ortholog matches ranged from 88.0% to 99.5%. 25 of 27 (92.6%). *B. multivorans* genomes in lineage 1 had the highest number of orthologous matches to the respective BCC0084 reference. BCC0080 and BCC0384, instead, had the highest number of matches to BCC1272 and ATCC 17616 from lineage 2a, respectively. Regarding lineage 2, five (R-20526, BCC0082, CF170.0a, CGD1 and AU4507) of 51 strains (9.8%) had the highest ortholog similarity to lineage 1. Lineage 2 had a similar ortholog match distribution to lineage 1, ranging from 85.1% to 100% similarity. This showed there was a high percentage of orthologous matches to the total number of sequence fragments, indicating that *B. multivorans* has multiple evolutionary conserved genes. Further analysis on clusters of orthologous groups (COGs) are discussed in Section 4.2.5 of this thesis.

3.2.5 Core-gene phylogenies

3.2.5.1 Master phylogenomic trees

An initial RAxML core-gene phylogeny was created on the 283-genome master panel, aligning a total of 2,998 conserved genes identified using the Roary pipeline (Figure 7). This analysis further illustrated the two-lineage *B. multivorans* population structure, consistent with the previous ANI findings (Section 3.2.4.1 and 3.2.4.2). Lineage 1 harboured 58 of 283 (20.5%) *B. multivorans* strains whilst lineage 2 harboured 221 of 283 (78.1%) of strains. A 3rd potential lineage, currently noted as 'other', was observed with 4 of 283 strains (1.4%). This held genome BCC1368 sequenced in this thesis and 3 other strain genomes obtained from the databases. When dividing lineage 2 into sub-lineages, 2a had 47 of 221

strains (21.3%) and 2b held the other 174 (78.7%). Within sub-lineage 2b, 111 of 174 strains (63.8%) were part of the isogenic CF170 group.

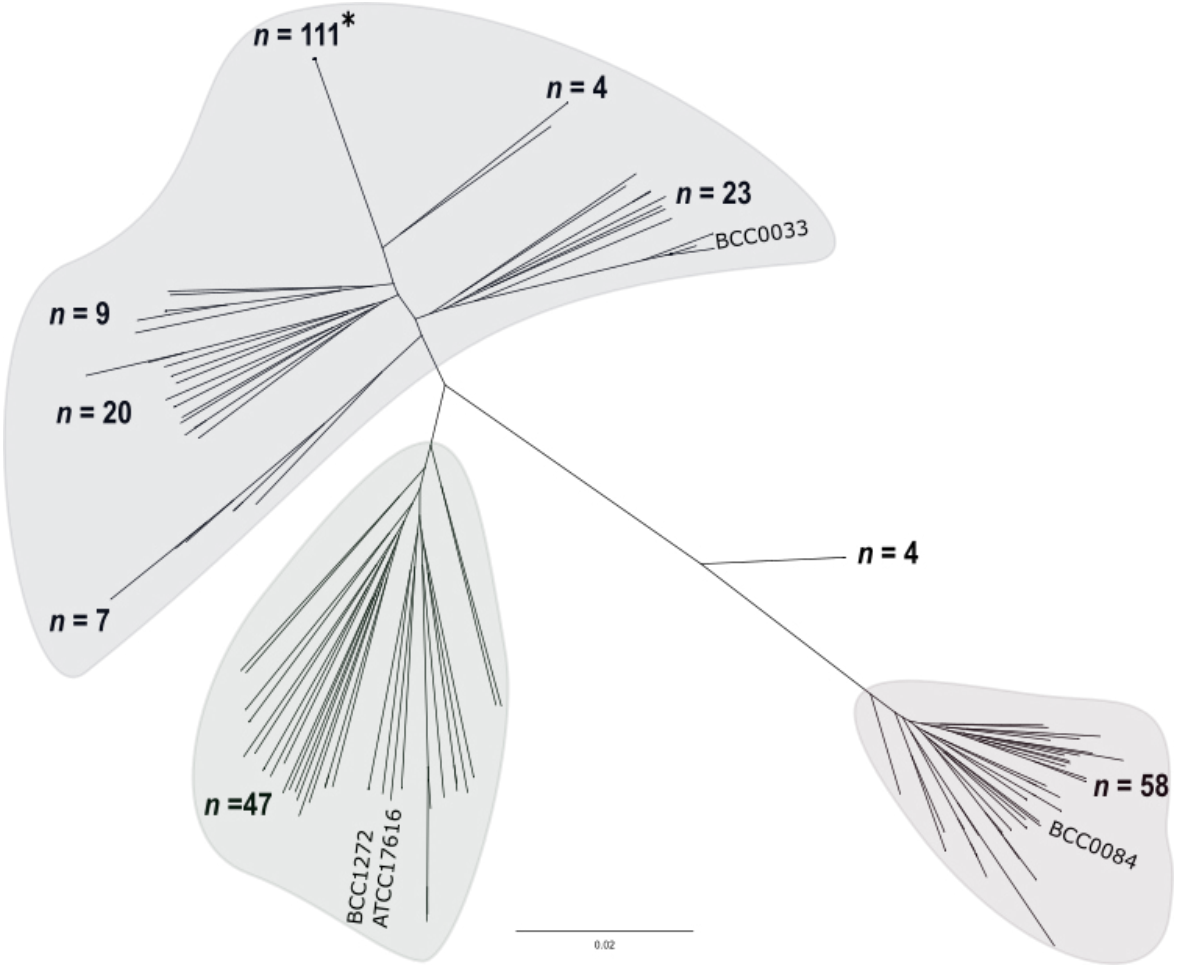


Figure 7 | Master core gene phylogenomic tree of the 283 *B. multivorans* genomes. RAxML phylogenomic tree created by aligning 2,998 core-genes with 100 bootstraps. Genome relatedness is represented by the scale bar on the bottom of the figure, illustrating the evolutionary distance between the *B. multivorans* strains. Lineages have been highlighted by different colours. Lineage 1 = red, lineage 2a = green and lineage eb = blue. Model CF strains characterised in this thesis are noted (BCC0033, BCC0084, BCC1272 and ATCC 17616). The *n* numbers for each lineage/group are also noted at the end of the tip clusters. The asterix next to *n* =111 represents CF170 (Canadian strain) which is a single *B. multivorans* chronic infection of sequential isolates. Scale bar represents the phylogenetic distance of 0.02 nucleotide substitutions.

Looking at strain isolation source, lineage 1 had 46 (79.3%) CF, 4 (6.9%) CGD, 1 (1.7%) ENV, 2 (3.4%) ENVH and 4 (6.9%) NON isolates. Lineage 2 had 193 (87.3%) CF, 2 (0.9%) CGD, 2 (0.9%) CLIN, 20 (9.0%) ENV, 2 (0.9%) unknown and 3 (1.4%) NON isolates. Of the environmental isolates in lineage 2, 16 were from lineage 2a whilst the other 4 were lineage 2*B.* Overall, it appeared that the majority environmental isolates were found within lineage 2 within the collection of genomes examined. All 4 isolates in the 'other' lineage were of CF origin.

3.2.5.2 Construction of a *B. multivorans* strain panel

Once the master phylogenomic tree was drawn, a representative strain panel of 77 *B. multivorans* genomes was selected for all further analysis in this thesis. They were selected based on ST (Section 3.2.6.1), isolation source (Section 2.1.1, Table 4) and phylogenomic placement (Figure 7). In addition, 50 isolates, representative of the total genomic diversity were investigated for differential phenotypic traits as described in Chapter 5. A secondary phylogenomic tree was drawn of the strain panel (Figure 8). *B. dolosa* AU0158 was included in the preliminary construction of the RAxML phylogenomic tree (Appendix Figure 66) as a close-neighbour outgroup. This analysis identified BCC1368 as the nearest *B. multivorans* genome outgroup for rooting of the secondary strain panel phylogenomic tree which used 3,251 core genes. The well-characterised environmental isolate genome, ATCC 17616, placed adjacent to CF isolate BCC1272, providing further evidence that they were isogenic.

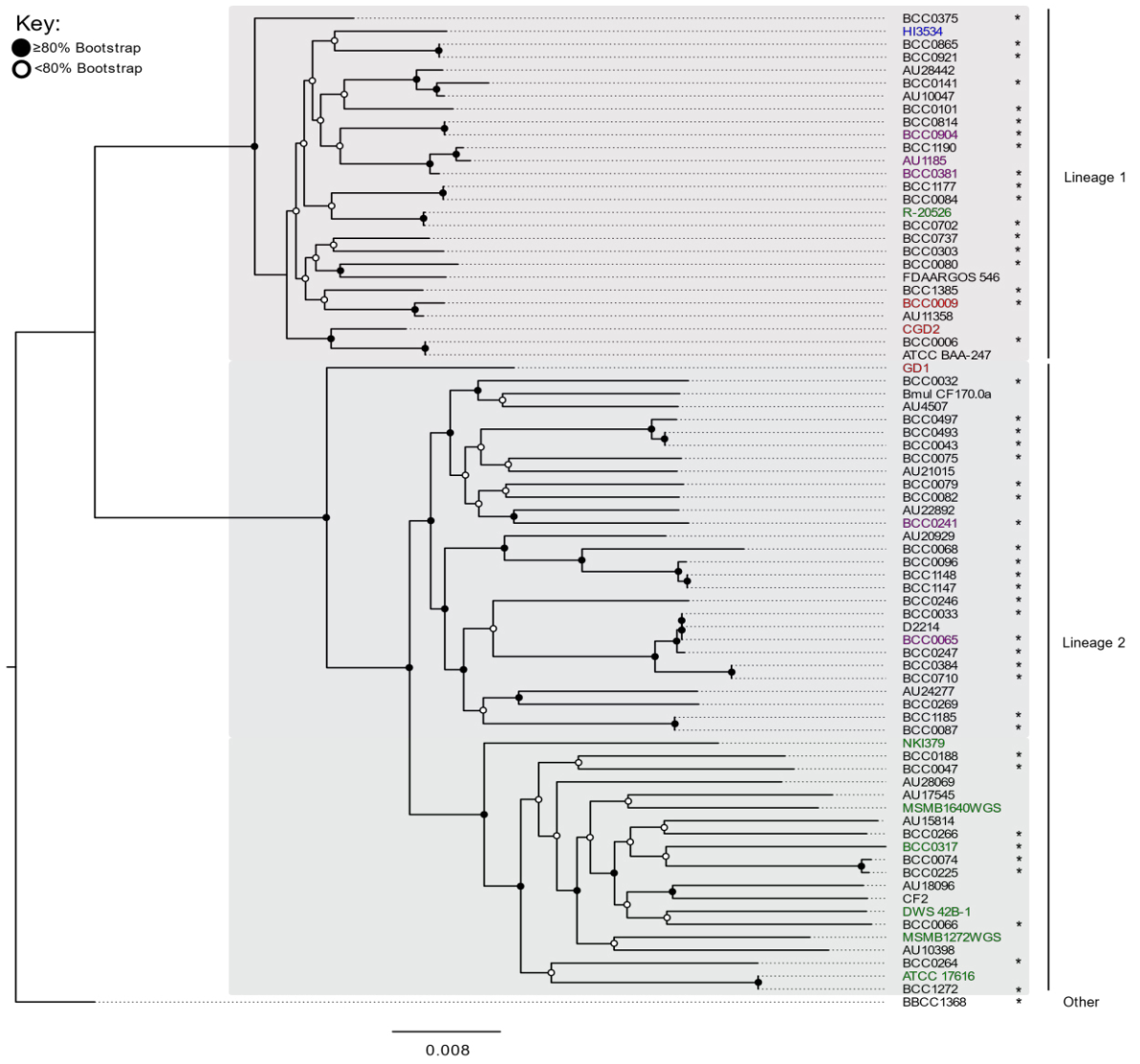


Figure 8 | Phylogenomic analysis of the *B. multivorans* strain panel (*n* = 77). RAxML tree built using the alignment of 3,251 core-genes and 100 bootstraps. Strain genomic relatedness is indicated by the scale bar (0.008). Rooting of the tree was done using BCC1368 as an outgroup. Nodes have been allocated either a filled or hollow circle to represent the bootstrap confidence levels. Filled circles represent a bootstrap of ≥80% and a hollow (white) circle means confidence of ≤80%. Lineages have been marked on the right-hand side. This shows Lineages 1 and 2 as well as 'other'. Colours denote the isolate type for each strain. Black = CF, green = ENV, blue = ENVH, purple = NON, red = CGD. The asterix next to the genome names represents the genomes sequenced as part of this thesis. Scale bar represents the phylogenetic distance of 0.008 nucleotide substitutions.

Core gene phylogenies were then created for the two lineages independently (Figure 8). ATCC 17616 was used as the root for lineage 2 and BCC0084 was used as the root for lineage 1 (Figure 10). The lineage 1 phylogeny showed BCC0084 was most closely related to BCC1177. Lineage 1 diverged into at least three clades, with BCC0702 and R-20526 closely related (clade 3), but more divergent from the rest of lineage 1. BCC0375 (clade 1) was also found on a single divergent branch. The remainder of lineage 1 then separated into at least four further clades (2a-f) (Figure 10).

Lineage 2 had a very clear divergence into the two sub-lineages, further dividing into 3 clades per sub-lineage (clades 1a-c and 2a-c) (Figure 10). The most complex divergence was in clade 2b of lineage 2*B.* Two single *B. multivorans* strains, BCC0079 and BCC0264 in lineage 2b were separated into their own two unique clades 2a and 2c, respectively.

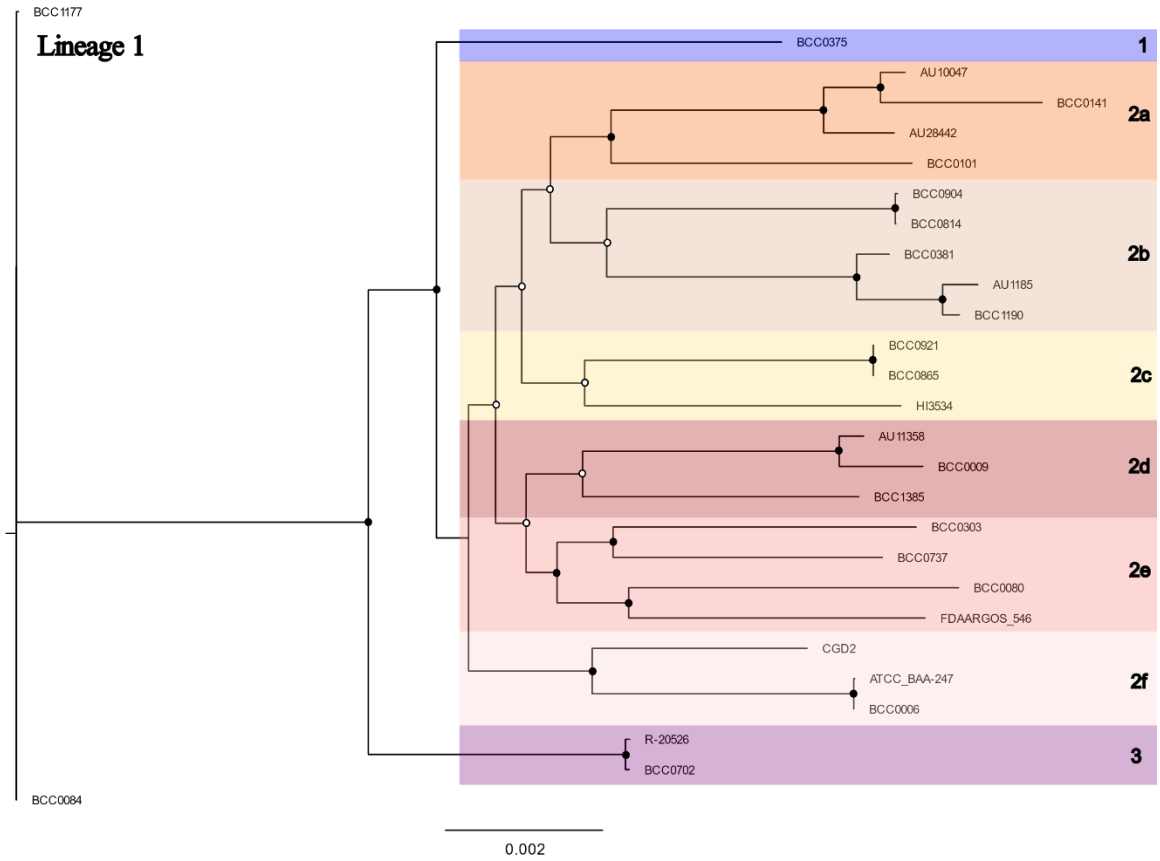


Figure 9 | Phylogenetic tree and lineage 1 using RAxML with 100 bootstraps. The phylogenetic tree used an alignment of 4,124 core-genes from 27 *B. multivorans* strains. The scale bar represents the phylogenetic distance of 0.002 nucleotide substitutions. BCC1368 was left out of the lineage analysis as it was noted to be in an undefined 'other' lineage. Colours on the phylogenetic trees represent the clades in which the *B. multivorans* strains belong. A full black circle on the node division indicates a confidence of ≥80% and a hollow (white) circle means confidence of ≤80%.

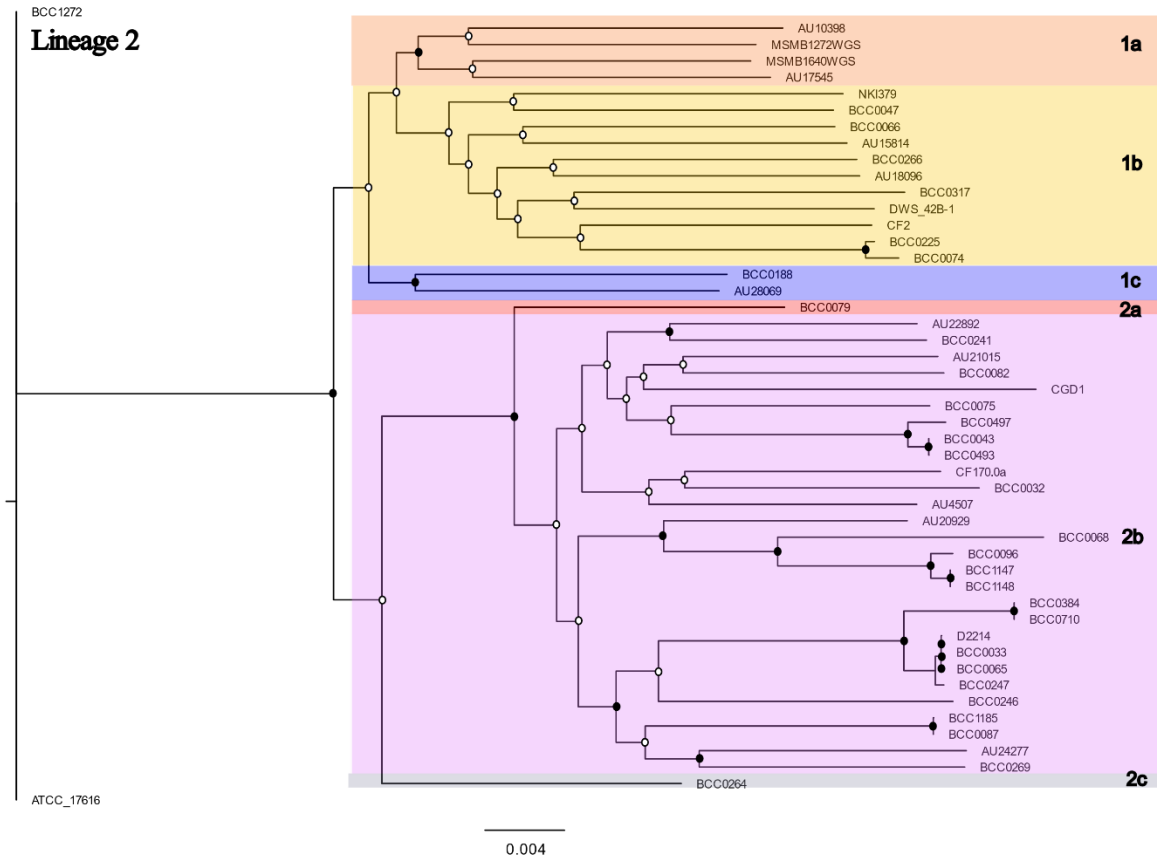


Figure 10 | Phylogenetic tree and lineage 2 using RAxML with 100 bootstraps. The phylogenetic tree used an alignment of 3,793 core-genes from 49 strains. The scale bar represents the phylogenetic distance of 0.004 nucleotide substitutions. BCC1368 was left out of the lineage analysis as it was noted to be in an undefined 'other' lineage. Colours on the phylogenetic trees represent the clades in which the *B. multivorans* strains belong. A full black circle on the node division indicates a confidence of ≥80% and a hollow (white) circle means confidence of ≤80%.

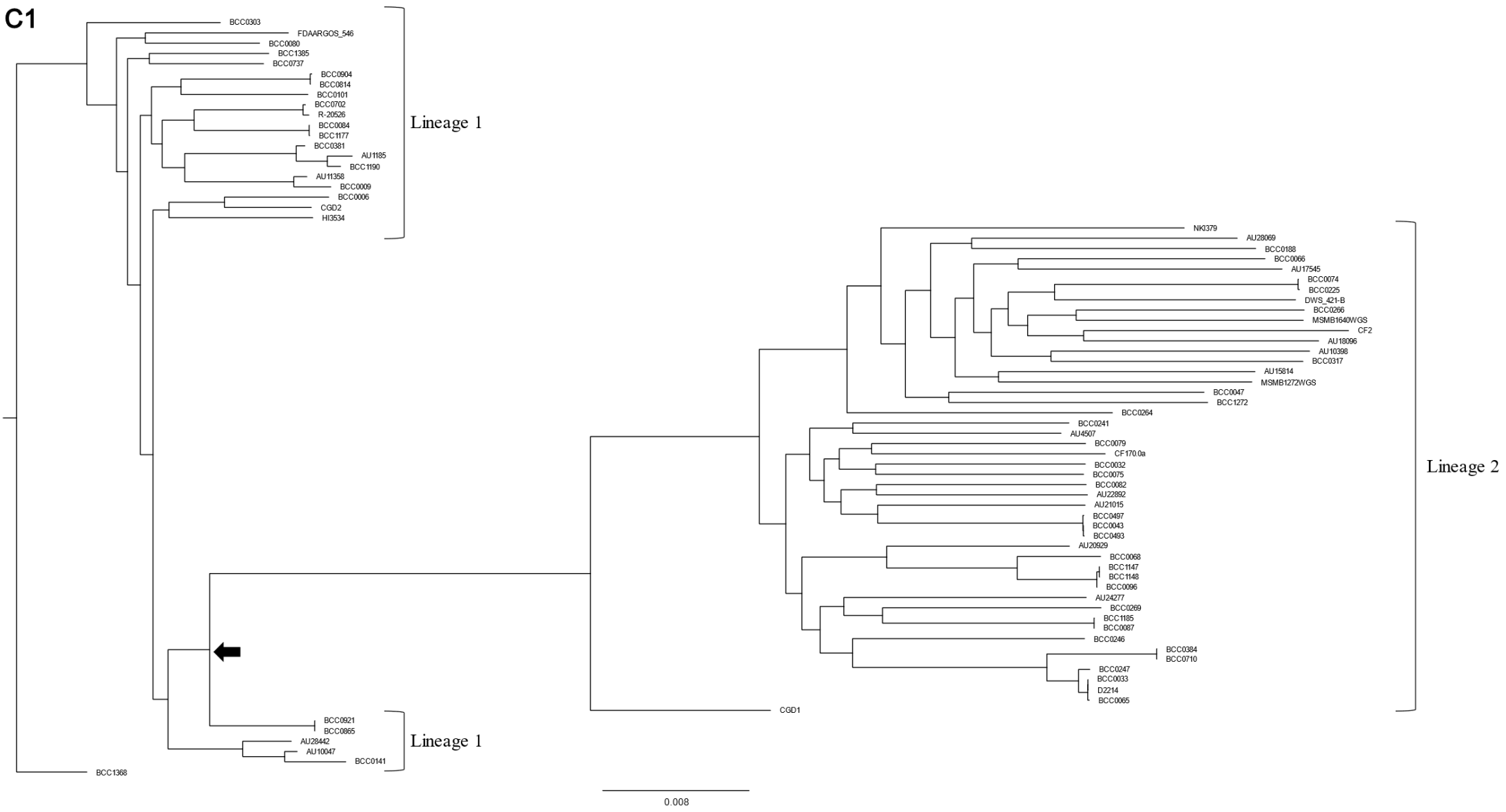
3.2.5.3 Comparison of the replicon phylogenies

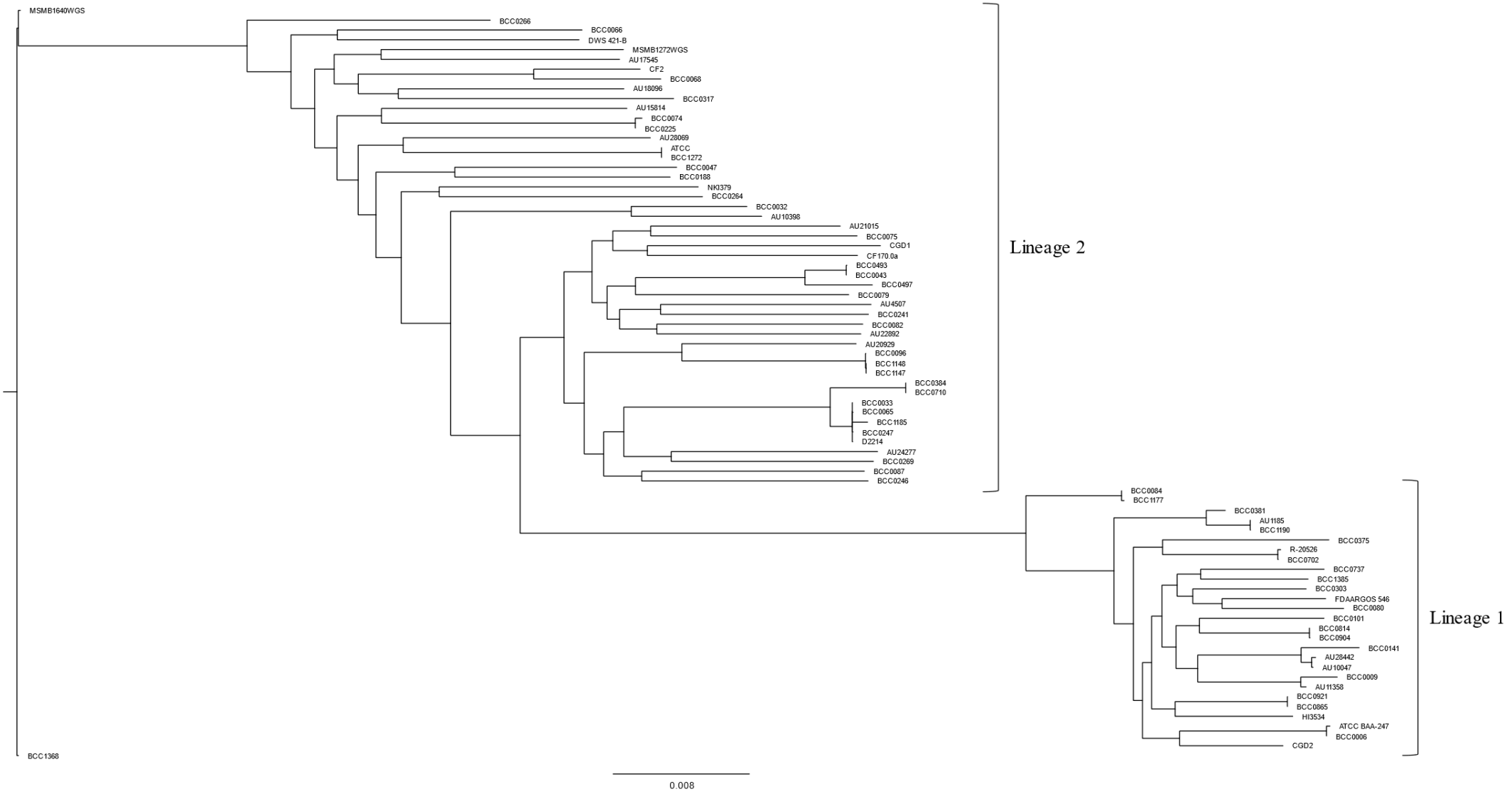
Questions arose as to whether there would be clear genomic differences between the lineages when looking at each replicon. The analysis aimed to observe whether the two-lineage divergence seen at whole genome level also occurred at replicon level within the multi replicon BCC genome (Mahenthiralingam *et al.* 2005). Species level conservation within replicon specific partition (*parAB*) and replication genes *repA* had been observed for the BCC (Drevinek *et al.* 2008a). Core gene phylogenies were drawn using RAxML for each genomic replicon separately, rooting all trees with outgroup BCC1368 (Figure 11). This used the scaffolded replicons from Section 3.2.9.1. Replicon C4 could not be processed via Roary due to the differences in sequences, including short sequence lengths. 2,205, 565 and 115 core genes were found in replicons C1, C2 and C3, respectively. Detailed analysis of the virulence and pathogenicity genes on each replicon can be found in Section 1.1.1.3.

The core gene phylogenies for replicons C1, C2 and C3 were similar when comparing against each other. Groupings for lineage 1 and lineage 2 remained consistent. Replicon 1 showed that the lineage 1

C1 was more closely related to the BCC1368 C1 outgroup. Lineage 2 C1 diverged from lineage 1 C1, with the closest neighbours BCC0865 and BCC0921. CGD1 was the outgroup of the diverged lineage 2, but the rest of the sub-lineage formation remained consistent.

Replicon C3 showed a similar evolution to that of C1, where lineage 2 evolved from lineage 1, where the latter was diverse. BCC0009 and AU11358 were the closest common lineage 1 C3 strains compared to lineage 2 C3. Interestingly, most lineage 1 strains did group together, however, BCC0080 from lineage 1 was observed within the lineage 2 cluster. This indicated that the C3 replicon of BCC0080 had a common association with the lineage 2 C3 replicon. Whilst lineage 2 C3 did split into two groups, it was not consistent in terms of sub-lineage with the whole genome core gene phylogeny (Section 3.2.5.2). Instead, it showed that lineage 2 sub-lineages were mixed amongst the two diverging groups in the C3 phylogeny. This meant that the C3 replicon, in terms of lineage 2, was evolutionarily diverse amongst the sub-lineages observed via phylogenomics on the total genomic content (Figure 11). Replicon 2 showed the potential divergence from a common ancestor for all *B. multivorans* strains. Interestingly, for C3 there was an initial divergence, with lineage 1 branching off from lineage 2, contrasting the other replicons.


C1
BCC0303
FDAARGOS_546
BCC0080
BCC1385
BCC0737
BCC0904
BCC0814
BCC0101
BCC0702
R-20526
BCC0084
BCC1177
BCC0381
AU1185
BCC1190
AU11358
BCC0009
BCC0006
CGD2
HI3534
Lineage 1
BCC0921
BCC0865
AU28442
AU10047
BCC0141
Lineage 1
BCC1368
NKI379
AU28069
BCC0188
BCC0066
AU17545
BCC0074
BCC0225
DWS_421-B
BCC0266
MSMB1640WGS
CF2
AU18096
AU10398
BCC0317
AU15814
MSMB1272WGS
BCC0047
BCC1272
BCC0264
BCC0241
AU4507
BCC0079
CF170.0a
BCC0032
BCC0075
BCC0082
AU22892
AU21015
BCC0497
BCC0043
BCC0493
AU20929
BCC0068
BCC1147
BCC1148
BCC0096
AU24277
BCC0269
BCC1185
BCC0087
BCC0246
BCC0384
BCC0710
BCC0247
BCC0033
D2214
BCC0065
CGD1
Lineage 2
0.008


C2
MSMB1640WGS
BCC0266
BCC0066
DWS 421-B
MSMB1272WGS
AU17545
CF2
BCC0068
AU18096
BCC0317
AU15814
BCC0074
BCC0225
AU28069
ATCC
BCC1272
BCC0047
BCC0188
NKI379
BCC0264
BCC0032
AU10398
AU21015
BCC0075
CGD1
CF170.0a
BCC0493
BCC0043
BCC0497
BCC0079
AU4507
BCC0241
BCC0082
AU22892
AU20929
BCC0096
BCC1148
BCC1147
BCC0384
BCC0710
BCC0033
BCC0065
BCC1185
BCC0247
D2214
AU24277
BCC0269
BCC0087
BCC0246
Lineage 2
BCC0084
BCC1177
BCC0381
AU1185
BCC1190
BCC0375
R-20526
BCC0702
BCC0737
BCC1385
BCC0303
FDAARGOS 546
BCC0080
BCC0101
BCC0814
BCC0904
BCC0141
AU28442
AU10047
BCC0009
AU11358
BCC0921
BCC0865
HI3534
ATCC BAA-247
BCC0006
CGD2
Lineage 1
BCC1368
0.008

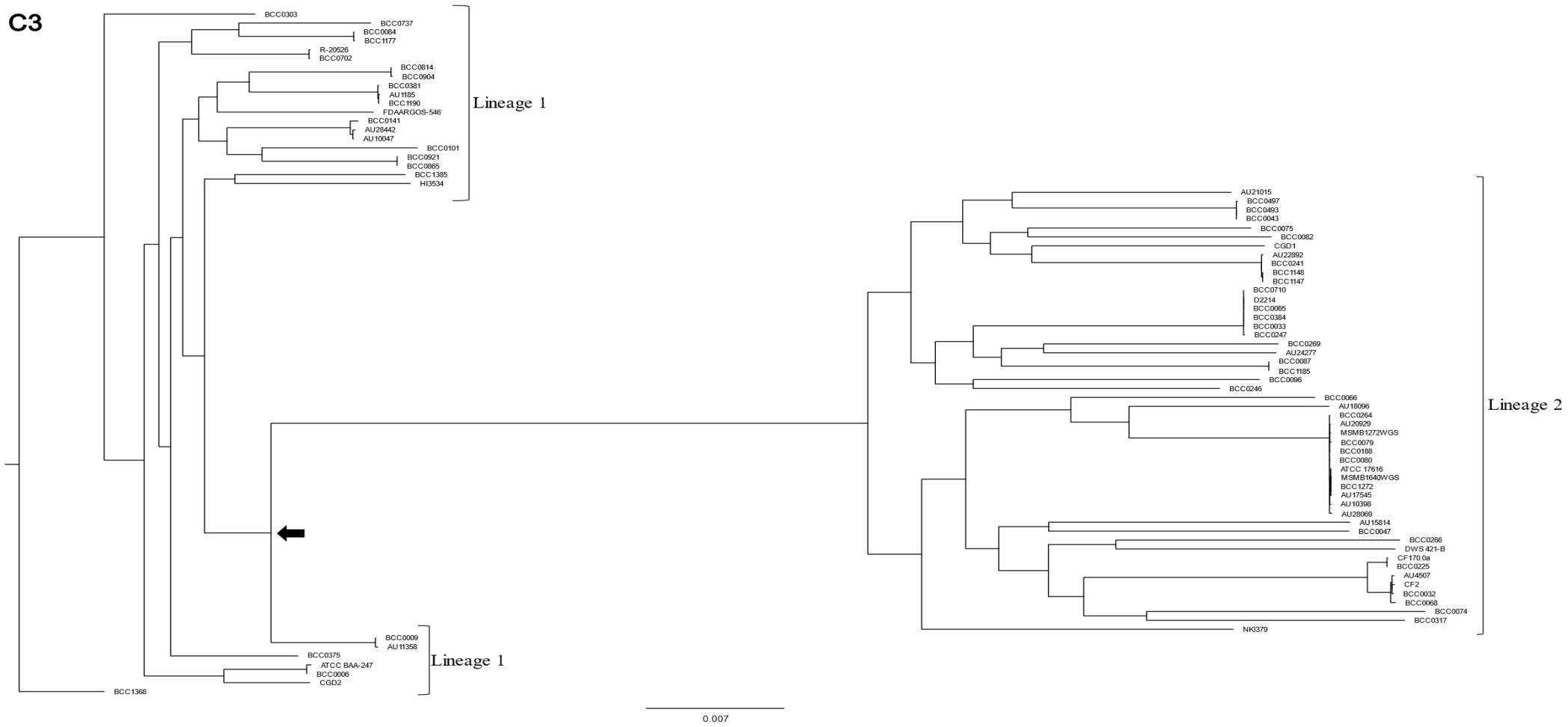


Figure 11 | Core-gene phylogenies of replicons C1, C2 and C3 in the *B. multivorans* strain panel (*n* = 77). RAxML with 100 bootstraps was used to construct the phylogenetic trees. 2,205, 565, and 115 core-genes were used to draw the trees for C1 (top), C2 (middle) and C3 (bottom), respectively. All phylogenetic trees were rooted using outgroup *B. multivorans* BCC1368. The branch divergence of lineage 2 from lineage 1 in the replicon C1 and C3 phylogenies is highlighted by a bold black arrow. Replicon C1 and C2 scale bars represent 0.008 nucleotide substitutions per site and the replicon C3 scale bar represents 0.007 nucleotide substitutions.

3.2.6 Multi-locus sequence typing

After gaining insight on the two-lineage split of *B. multivorans* at whole genome level, the next aim was to see whether this could be observed at 7 gene base MLST level (Baldwin *et al.* 2005). Whilst this is a less powerful phylogenetic analysis tool in comparison to WGS, the greatest diversity of *B. multivorans* isolates (*n* = 4149 isolates; June 2021) are present on the MLST global epidemiological database. Here the MLST-based phylogeny was compared with the high resolution phylogenomic analysis of *B. multivorans.*

3.2.6.1 Identification of STs

An initial ST identification was performed on all the *B. multivorans* strains (*n* = 283), using the pubMLST (Jolley *et al.* 2018) typing scheme to scan contig FASTA files. This checked both unknown STs as well as those previously defined for any errors. There were 63 unique STs found amongst the 283 genomes tested. Strains with unknown STs had the sequences for each housekeeping gene extracted into FASTA files. These were inputted to the pubMLST database (Jolley *et al.* 2018) to cross-check any gaps in the dataset before noting a ST or allele number as 'novel'. From the analyses, there were a total of 7 unique STs after checking. There were also 8 novel alleles within 7 *B. multivorans* strains (Table 16). Four of these alleles were within 3 (BCC0082, BCC0266, and BCC0737) strains in the *B. multivorans* 77 strain panel. STs of the *B. multivorans* strain panel are shown in Table 16. There were 43 unique STs in the strain panel, representing 68.3% of all STs identified in the master dataset. Epidemic STs 16 and 21 (Baldwin *et al.* 2008) were present within both the full master dataset as well as the strain panel.

Table 16 | STs, Isolation source, MLST loci and clonal complexes of the *B. multivorans* strain panel (*n* = 77). Novel alleles are observed by a ~ before the number and the ST is identified as 'novel'.

Strain	Isolation source	MLST loci							ST	Clonal Complex
		atpD	*gltB*	*gyrB*	*recA*	*lepA*	*phaC*	*trpB*		
ATCC 17616	ENV	13	78	100	94	92	96	6	21	-
BCC0006	CF	11	60	251	81	37	96	5	650	-
BCC0009	CGD	9	223	445	81	137	35	215	1530	-
BCC0032	CF	13	151	168	139	142	100	132	191	-
BCC0033	CF	8	5	5	7	7	42	105	16	1
BCC0043	CF	13	9	83	12	7	42	391	806	-
BCC0047	CF	13	62	695	110	45	14	452	1077	-
BCC0065	NON	8	5	5	7	7	42	105	16	1
BCC0066	CF	336	61	97	11	64	96	104	880	
BCC0068	CF	168	220	303	133	7	96	4	329*	
BCC0074	CF	14	8	55	11	46	96	281	618	-
BCC0075	CF	13	7	6	10	224	42	415	899	-
BCC0079	CF	13	150	166	88	7	42	6	1792*	
BCC0080	CF	3	50	4	81	7	35	57	1964	
BCC0082	CF	13	188	~611	165	200	96	~220	Novel	
BCC0084	CF	9	50	53	81	63	96	133	195	5
BCC0087	CF	13	5	172	133	145	96	137	199	7
BCC0096	CF	168	190	259	133	7	96	132	317	8
BCC0101	CF	9	205	285	141	63	35	5	304	-
BCC0141	CF	9	50	84	141	37	96	7	1023	-
BCC0188	CF	125	154	171	140	144	14	136	196	-
BCC0241	NON	14	8	55	11	46	96	281	618	-
BCC0225	CF	13	329	261	7	7	42	132	605	-
BCC0246	CF	13	5	262	188	203	42	132	273	-
BCC0247	CF	8	5	5	7	7	42	105	16	1
BCC0264	CF	13	61	264	184	144	42	6	274	-
BCC0266	CF	13	152	~695	196	143	96	135	Novel	
BCC0269	CF	13	196	265	189	201	96	195	354*	
BCC0303	CF	10	60	4	77	37	35	5	25	-
BCC0317	ENV	13	63	53	80	61	96	56	22	-
BCC0375	CF	76	50	99	93	37	35	111	117	-
BCC0381	NON	9	75	54	93	63	35	66	18	-
BCC0384	CF	8	5	5	7	7	42	5	15	1
BCC0493	CF	13	9	83	12	7	42	391	806	-
BCC0497	CF	13	9	83	12	7	42	7	26	2
BCC0702	CF	9	50	169	81	409	96	133	836	-
BCC0710	CF	8	5	5	7	7	42	5	15	1
BCC0737	CF	123	50	170	81	~37	35	5	Novel	-
BCC0814	CF	118	50	158	6	37	96	5	180	-
BCC0865	CF	9	142	161	81	137	96	66	181	-
BCC0904	NON	118	50	158	6	37	96	5	180	-
BCC0921	CF	9	142	161	81	137	96	66	181	-
BCC1147	CF	168	190	259	133	7	96	132	317	8
BCC1148	CF	168	190	259	133	7	96	132	317	8
BCC1177	CF	9	50	53	81	63	96	133	195	5

BCC1185	CF	13	5	172	133	145	96	137	199	7
BCC1190	CF	9	75	54	93	63	35	66	18	-
BCC1272	CF	13	78	100	94	92	96	6	21	-
BCC1368	CF	211	205	170	93	37	35	251	179	-
BCC1385	CF	7	270	4	81	137	35	5	847	-
ATCC BAA-247	CF	13	236	354	133	231	42	4	650	-
AU1185	NON	9	75	54	93	63	35	66	18	-
AU4507	CF	13	61	620	133	424	42	6	891	-
AU10047	CF	9	50	84	289	37	96	5	564	-
AU10398	CF	13	397	283	135	623	42	340	1512	-
AU11358	CF	9	223	445	81	37	35	215	646	-
AU15814	CF	9	75	54	93	63	35	66	18	-
AU17545	CF	193	234	325	185	239	42	256	623	-
AU18096	CF	13	334	483	309	355	42	340	603	-
AU20929	CF	13	328	475	7	239	96	334	715	-
AU21015	CF	13	329	259	133	46	96	132	622	-
AU22892	CF	13	333	482	133	10	96	4	190	4
AU24277	CF	121	138	167	138	141	42	132	625	-
AU28069	CF	13	9	484	7	64	266	195	630	-
AU28442	CF	13	145	488	135	10	96	104	645	-
BMUL CF170.0a	CF	13	236	354	133	231	42	4	783	-
CF2	CF	193	453	695	207	461	343	4	1079	-
CGD1	CGD	12	6	118	9	63	100	6	1762*	
CGD2	CGD	11	75	251	141	37	35	7	442	-
D2214	CF	8	5	5	7	7	42	105	16	1
DWS 42B-1	ENV	122	373	98	7	230	96	376	809	-
FDAARGOS 546	ND	10	153	315	93	37	96	66	355	-
HI3534	Other[a]	7	332	170	81	63	35	5	620	-
MSMB1272WGS	ENV	122	148	164	80	10	45	302	1088	-
MSMB1640WGS	ENV	158	371	98	11	230	96	251	802	-
NKI379	ENV	13	786	166	11	239	42	715	1771	
R-20526	ENV	9	50	169	81	409	96	133	836	-

3.2.6.2 Identification of clonal complexes

The MLST data also provided insight into the *B. multivorans* clonal complexes which had been observed (Baldwin *et al.* 2008). Within the *B. multivorans* strain panel, 6 clonal complexes (CC1, CC2, CC4, CC5, CC7 and CC8 (Baldwin *et al.* 2008)) were identified (Table 16). To look further into this, the 283 genomes in the master dataset were subjected to BURST analysis on pubMLST (Jolley *et al.* 2018), grouping the strains based on their allelic profile. The default profile match of n-2 loci was used, identifying 11 groups (Appendix Table 46). No central STs were found for any of the groups at this profile match level. 72 of the STs were also identified as singletons, with frequencies ranging from 1 to 6 (Appendix Table 46).

3.2.6.3 MLST phylogeny

Using the *B. multivorans* strain panel (n =77), a phylogeny was created (Figure 12) using the MLST types extracted from the whole genome sequences through the Page *et al.* (2016b) get_sequence_type (Section 2.9.2). This confirmed the robustness of the core-gene phylogenies previously presented. There were no differences regarding two-lineage split of the MLST-tree (Figure 12) compared to the core-gene tree (Figure 8). However, there were differences observed in the tree in terms of phylogenomic places of the strains within the lineages, particularly between lineages 2a and 2b (Figure 12). AU21015, AU20929, BCC0068, BCC0096, BCC1148, BCC1147 were observed in lineage 2a in the MLST phylogeny (Figure 12), whereas they were found to be in lineage 2b of the core-gene phylogeny (Figure 8). Whereas NKI379, AU28069, AU17545, AU10398 and BCC0264 were flipped from lineage 2a in the core-gene phylogeny into lineage 2b of the MLST phylogeny. Another main difference noted was the deep branch lengths in both lineages, with particular attention to a group within lineage 1, which were noted as having shorter branch lengths in the core gene phylogeny (Section 3.2.5.2).

The bootstrap values of the MLST phylogeny (Figure 12) were also comparatively lower overall, with many less than 80% confidence, compared to the core gene phylogeny (Figure 8). There was over an 80% bootstrap value for the splitting of the two lineages at MLST-level (Figure 12). However, many branches within the lineages, particularly in lineage 2, had a bootstrap value of less than 10%, and 0% in certain cases. This demonstrated the lack of resolution using a phylogenetic analysis of DNA from just 7 genes in MLST.

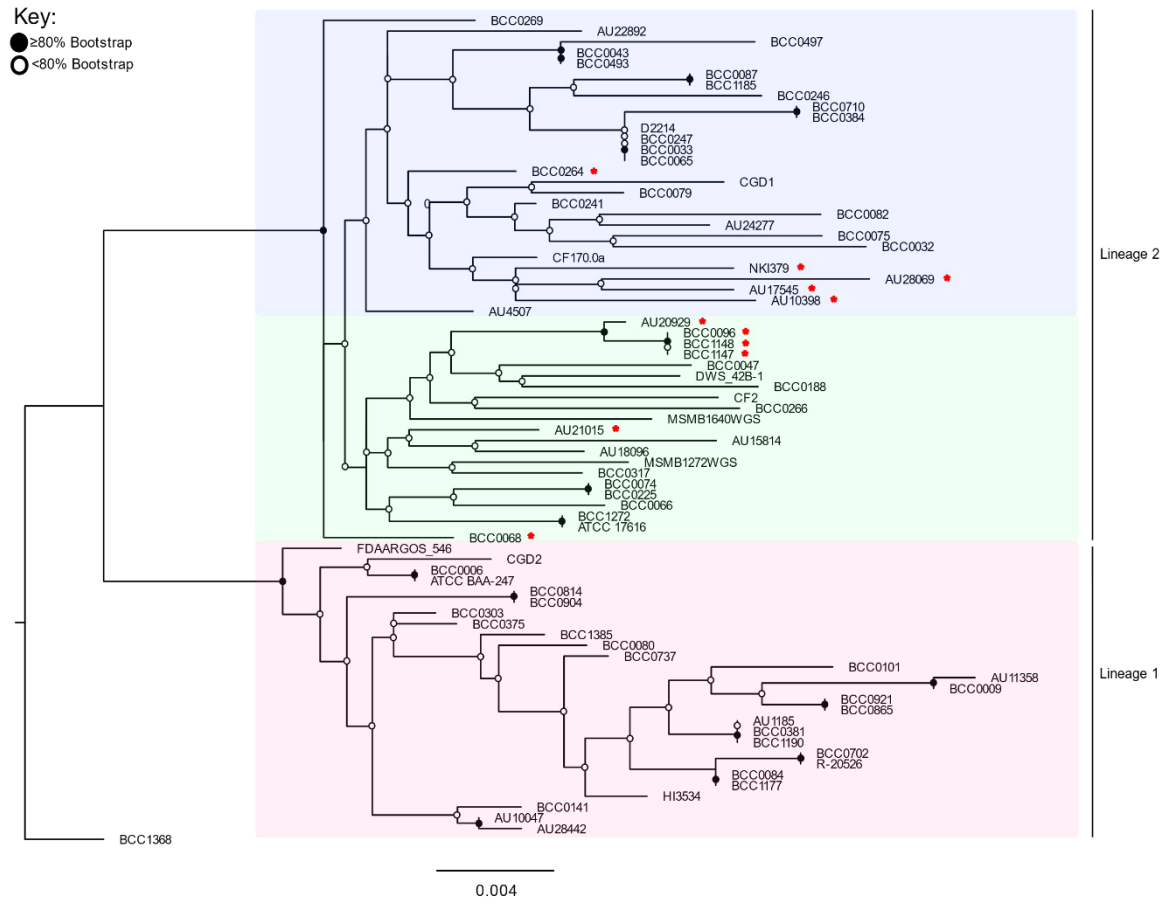


Figure 12 | MLST-gene phylogeny of the 77 genome *B. multivorans* strain panel. The RAxML (100 bootstraps) phylogenetic tree was created by extracting the MLST sequences from the sequenced whole genomes using MLST check (Page *et al.* 2016b). Bootstrap values are indicated using the key. Filled circles represent a bootstrap of ≥80% and a hollow (white) circle means confidence of ≤80%. Lineages are shown on the right hand side of the figure, with colour coding illustrating the lineages and sub-lineages. Lineage 1 = red, lineage 2a = green, lineage 2b = blue. The red asterix next to strain names represents the flipped genomes in lineage 2. Scale bar represents the phylogenetic distance of 0.004 nucleotide substitutions.

3.2.6.4 *B. multivorans* MLST clusters and lineage thresholds

Phyloviz (Francisco *et al.* 2012; Ribeiro-Gonçalves *et al.* 2016) was used as a tool to identify the following features of the *B. multivorans* population structure : 1) whether those of the same isolation source or country grouped together, 2) to see whether those with the same ST clustered together and 3) if ST is lineage-specific. All the available *B. multivorans* MLST types from pubMLST (Jolley *et al.* 2018) were downloaded (October 2020) and combined with the STs identified in the genomes both sequenced in this thesis and downloaded from NCBI (Section 2.1). This provided a net total of 581 *B. multivorans* with 221 unique STs. Any duplicate strains were removed from the analysis, providing a total of 566 unique strains. No correlations or clusters were observed when looking at isolation source, country of isolate origin or ST in the dataset.

Phyloviz (Francisco *et al.* 2012; Ribeiro-Gonçalves *et al.* 2016) was also used to analyse the two defined lineages (Section 3.2.4.1**Error! Reference source not found.**, 3.2.4.2 and 3.2.5.1) in greater detail. The 'tree cut-off' option was utilised to look at the relatedness of the lineages (Figure 13). As shown by the phylogenomic analysis (Figure 13), lineage 1 was more closely related than lineage 2 in all instances where differing distance levels were used, with lineage 1 clusters forming at difference level 4 compared to lineage 2's distance level 5. Initial phylogenetic tree joining was observed at cut-off level 3. There were then 29 strain clusters which formed at cut-off level 5, with lineage 1 and 2 remaining separate, spanning various STs, isolation sources and countries (Figure 13A). This indicates that the groups are due to recombination. The joining of lineage 1 and lineage 2 was then observed at cut-off level 6 (Figure 13B), and the whole phylogenetic tree formed at cut-off level 8. At cut off level 7, 4 isolates were independent of any groups (AU0485, QLD023, CF2 and an unknown isolate).

N Locus Variant (NLV) graphs were used to identify locus variants through link creation between nodes. The most appropriate links included those with differences of up to and including NLV 3 and 4. Like previously, lineage 1 has a tighter grouping at NLV 3, with lineage 2 at NLV 4 (Figure 14). Less connections are observed in lineage 2a at NLV 4 compared to lineage 2b, confirming it is more distantly related.

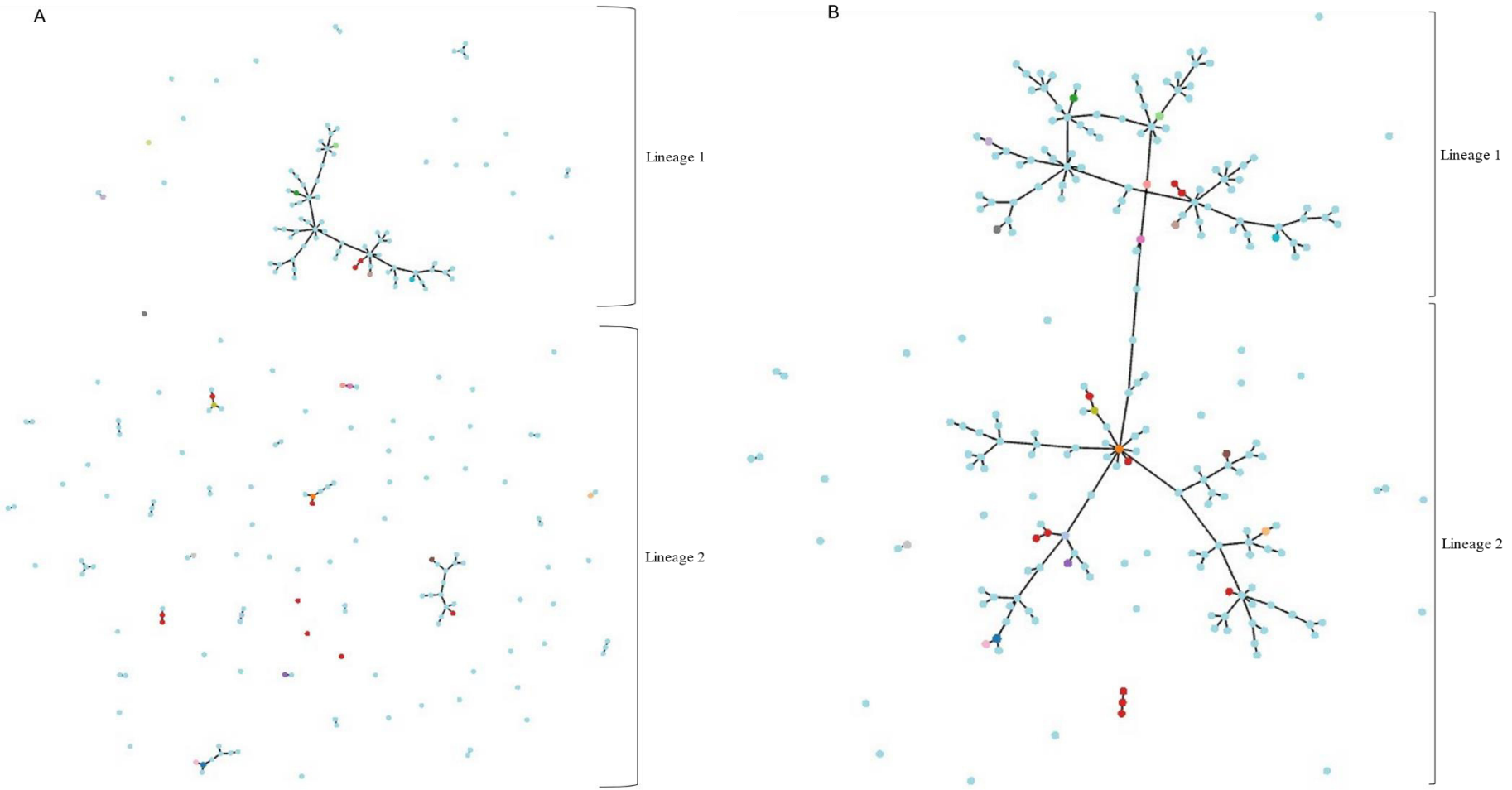


Figure 13 | Tree cut-off graphs for levels 5 and 6. Figures created using Phyloviz online (Ribeiro-Gonçalves *et al.* 2016) using the combined STs from the 77 *B. multivorans* strain panel and the STs available in pubMLST (October 2020). A total of 221 STs from 566 genomes was used for the analysis. **(A)** Tree cut-off level of 5, showing the lineages remaining separate but some attachment of nodes within lineage 1. **(B)** Tree cut-off level 6, with both lineage 1 and lineage 2 attached yet some isolates still independent of any group.

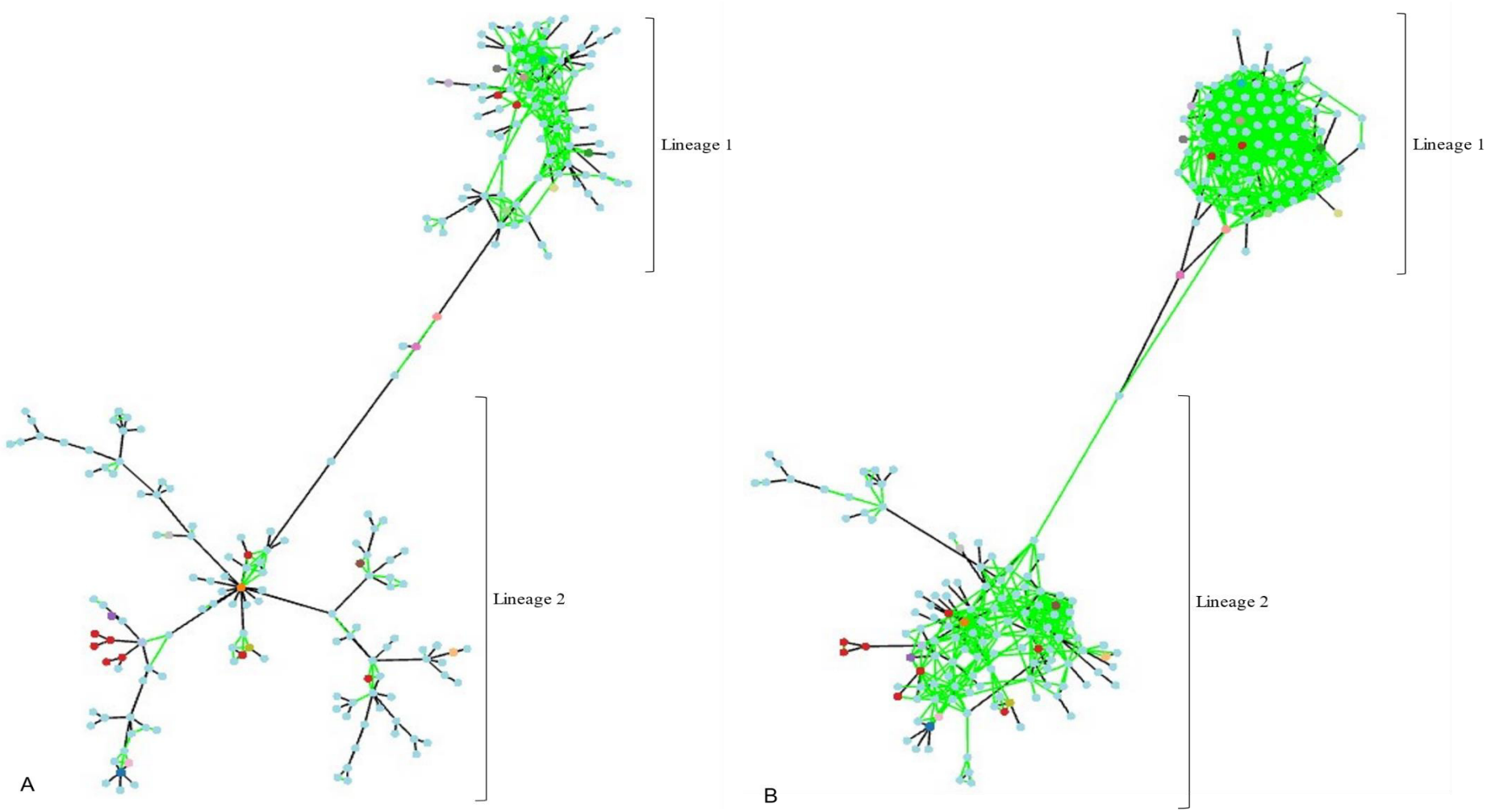


Figure 14 | N Locus Variant (NLV) graphs illustrating tree cut-off thresholds in 581 *B. multivorans* genomes. Analysis combined the 77 *B. multivorans* strain panel with all the available STs from PubMLST (October 2020). This provided a total of 221 unique STs for analysis. **(A)** NLV cut-off of 3 and **(B)** NLV cut-off of 4. The figure illustrates the links between nodes, showing the number of differences up to and including the defined thresholds. This type of graph is used to show clustering of phylogenomic groups. Figure created using Phyloviz online web tool (Ribeiro-Gonçalves *et al.* 2016).

3.2.7 Genomic features and basic gene content

3.2.7.1 Basic genomic content

Roary and R studio was used to provide statistics on predicted gene counts. Overall, 449,846 genes were analysed for the strain panel (n = 77), which had an average gene count of 5,811 genes, of which 112 (1.9%) was the average number of unique genes. There were 22,751 orthologous groups in the dataset and 8,728 unique genes overall. BCC0904 was the strain which harboured the least number of predicted genes (5,329 genes). Of the 77 genomes, BCC0047 has the greatest number of unique genes (~500) (Figure 15). Whilst BCC1147 and BCC1148 appear identical on the phylogenetic tree (Figure 8), BCC1148 has a higher overall gene content (5,842 genes) than BCC1147 (5,836 genes). Whilst it is hard to confirm whether these two strains are 100% identical, they are still very closely related and may have partial genes present without full predictions. Therefore, it is likely that the strains are still isogenic.

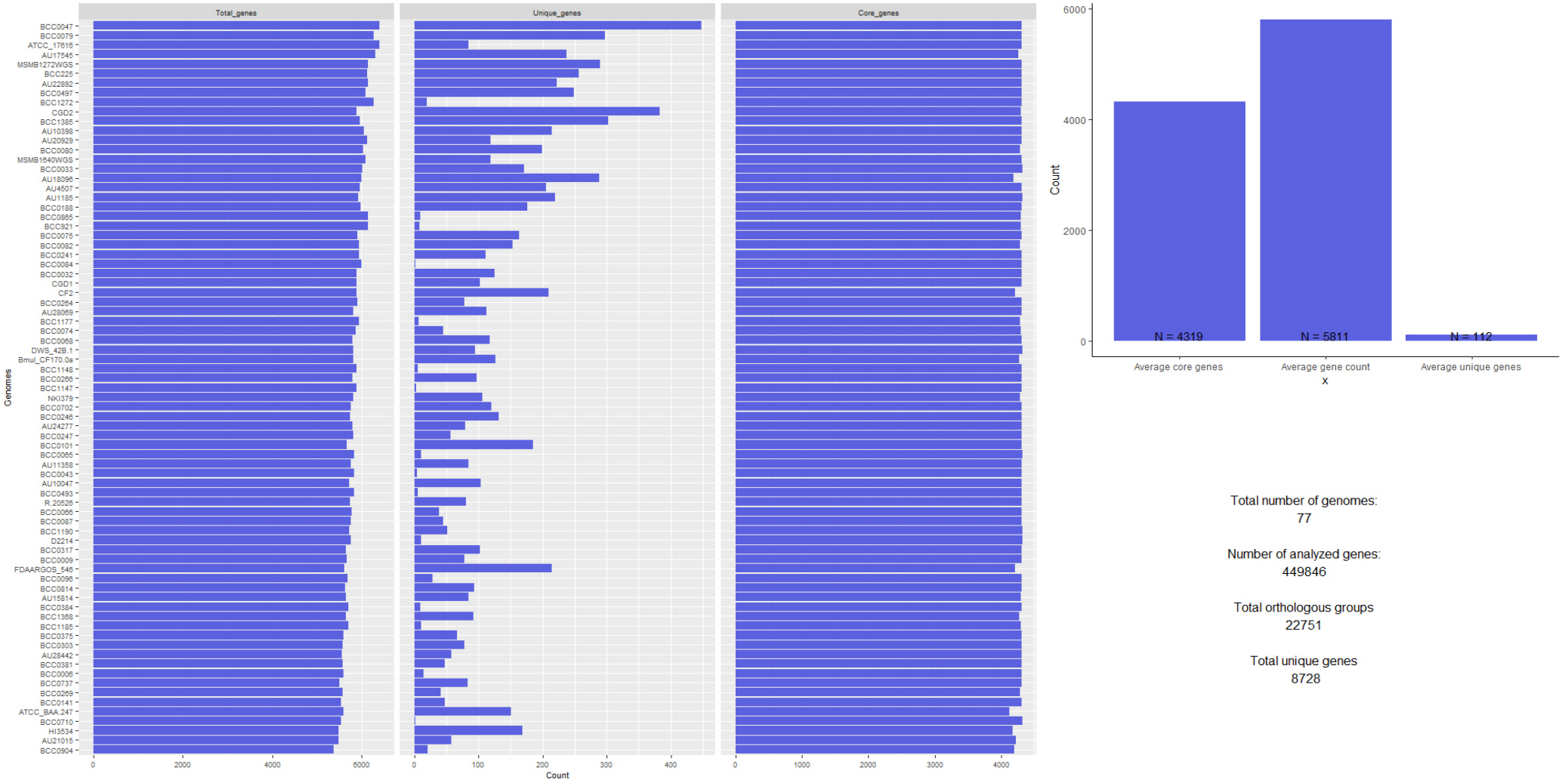


Figure 15 | Basic genomic content for the 77-genome *B. multivorans* strain panel. The horizontal bar plots show the total, unique and core genes present in each of the genomes. The vertical bar plot (right-hand side) shows the overall counts for the average core genes, gene count and unique genes in the pan-genome. Total numbers of genes are noted on the bottom right. R studio tutorial by Thomas Lesley Sitter.

3.2.8 Long-read sequence DNA extraction and assembly

4 model CF *B. multivorans* strains were identified as part of this thesis, based on their genotypic placement (Section 3.2.5.1), ST (Section 3.2.6.1) and phenotypic characterisation (Chapter 5). Strains BCC0033, BCC0084, and BCC1272 were chosen alongside the well-characterized ATCC 17616 strain. Full details for each strain can be found in Table 5. BCC0033, also known as C5568, is an isolate from California, USA. BCC0033 has previously been used in the literature to explore mucoid phenotype (Zlosnik *et al.* 2011) and intracellular survival inside macrophages (Chu *et al.* 2004; Schmerk and Valvano 2013). BCC0084 (commonly known as strain C6398) has been subjected to research on antibiotic susceptibility and the use of targeting bacterial hopanoids as CF treatment (Malott *et al.* 2014). BCC1272, also known as strain HI 2229, is the isogenic CF isolate of ATCC 17616. BCC1272 has been subject to experimentation for antibiotic susceptibility (Veloira *et al.* 2003).

3.2.8.1 Assembling the *B. multivorans* PacBio reads

The PacBio genomes were subject to two long-read assemblers: Flye (Kolmogorov *et al.* 2019) and Trycycler (Wick *et al.* 2021). The results and statistics of the two assemblers can be found in Table 17. Trycycler proved to be the more robust long-read assembler due to the nature of using three independent assembly tools, reducing the risk of misassembles, and providing a more reliable consensus sequence (Wick *et al.* 2021). Both assemblers gave an output of 4 contigs for BCC0084 and BCC1272, with BCC0033 originally having 7 contigs (Table 17). Kraken2 (Wood *et al.* 2019) was used to check for contamination of the assembled contigs (Section 2.6.6), whereby contigs 5-7 from BCC0033 were omitted as non-*Burkholderia* origin, leaving 4 contigs in total. This is in-line with the multi-replicon structure of the Bcc (Mahenthiralingam *et al.* 2005).

Table 17 | Comparing the genome statistics of the complete *B. multivorans* genome bacterial assemblies (*n* = 4). Assemblies include Flye, Trycyler and Trycyler with read polishing.

Strain	Assembly Type	Genome Size (Mb)	Contigs	Longest Contig	Shortest Contig	>1000bp	>500000bp
BCC0033	Flye	6.84	7	3544223	19070	6839546	6798590
	Trycycler	6.83	7	6828606	6798593	66.55	3544238
	Trycycler with Read Polishing	6.74	4	6736607	6736607	67.03	3544234
BCC0084	Flye	6.67	5	3345187	13705	6670087	6656382
	Trycycler	6.66	4	6656391	6656391	67	3345190
	Trycycler with Read Polishing	6.66	4	6656395	6656395	66.55	3345195
BCC1272	Flye	7.01	4	3448462	167422	7008814	7008814
	Trycycler	7.01	4	7008831	7008831	66.69	2473136
	Trycycler with Read Polishing	7.01	4	7008829	7008829	66.69	2473135

3.2.9 Processing of strain panel genomic contigs using the model strains

3.2.9.1 Scaffolding of strain panel contigs

Four complete *B. multivorans* genomes (BCC0033; lineage 2b, BCC0084; lineage 1, BCC1272 and ATCC 17616; lineage 2a) were used for scaffolding of the short-read contig assemblies. ATCC 17616 was sourced and downloaded from NCBI. The BCC numbered genomes were produced in this thesis from PacBio long-read sequencing (Section 2.6.2). The remaining strain panel (n = 73) (Section 3.2.5.2) were scaffolded onto all four complete genomes (BCC0033, BCC0084, BCC1272 and ATCC 17616) one-by-one using CONTIGuator which is a single-reference based scaffolder.

The outputs of CONTIGuator were manually checked for each scaffold to identify the reference which provided the fewest scaffolds and greatest overall alignment (bp). This was then used as the reference replicon structure for that specific *B. multivorans* strain. In a case where multiple structures were identical, the well-documented *B. multivorans* ATCC 17616 strain was used.

On average, 98.2% of each *B. multivorans* whole genome was scaffolded to a reference (Table 18). Two *B. multivorans* (BCC0921 and BCC0865) had 93-94% of the genomes scaffolded at the lowest percentage. 18, 15 and 39 *B. multivorans* genomes scaffolded to 95-97%, 98-99% and >99% of the chosen reference genome, respectively (Table 18). Of this, 4 genomes (FDAARGOS 546, DWS 421-B, AU1185 and MSMB1640WGS) completely scaffolded (100%) to the reference ATCC 17616 genome. Therefore, the scaffolded genomes outputted from the CONTIGuator tool have been used for further analyses in this thesis.

3.2.9.2 Genome size and statistics of the *B. multivorans* scaffolds

The CONTIGuator scaffolded genome size across the *B. multivorans* strain panel (Section 3.2.5.2) was quite conserved, ranging from 5.95 Mb (BCC0904) to 6.96 Mb (BCC0079). BCC0904 remained the smallest *B. multivorans* genome, decreasing by 0.07 Mb after scaffolding. Prior to scaffolding, BCC0047 had the longest *B. multivorans* genome (7.11 Mb), which decreased by 0.17 Mb after scaffolding (Table 18), making it the second largest *B. multivorans* scaffolded genome in the strain panel. It was overtaken by BCC0079 as the largest scaffolded genome at 6.96 Mb, which only decreased by 0.03 Mb during genome scaffolding (Table 18).

All the 77 strain panel genomes had at least 3 genomic replicons, illustrated from the genome scaffolding (Section 3.2.9.1). This is consistent with the multi-replicon structure of *Burkholderia* species (Mahenthiralingam *et al.* 2005). C1 was the largest replicon which has an average length of 3.35 Mb, ranging from 3.70 Mb (BCC0079) to 3.14 Mb (ATCC BAA 247). C2 was next in size with an average length of 2.43 Mb, ranging from 2.17 Mb (FDAARGOS 546) to 2.53 Mb (AU1185). C3 was the smallest replicon present in all 77 genomes which has an average of 0.65 Mb, spanning 0.24 Mb (HI3534) to 0.92 Mb (ATCC 17616).

Several *B. multivorans* strains also possessed a 4th replicon (C4), and it was observed in all the 4 model strains with complete genomic sequences. C4 of each *B. multivorans* panel strain had an average length of 0.01 Mb with an average GC content of 62.1% (Table 18). This suggested that the C4 was a plasmid replicon in terms of its smaller size. The genomic content of the C4 replicon is discussed further in Section 4.2.4.

Table 18 | Genomic statistics of the *B. multivorans* scaffolds (*n* = 77). Table shows the *B. multivorans* strain, reference scaffolded to, the percentage of genome scaffolded (%), total length before and after scaffolding, and Mb and GC content of the independent replicons for each *B. multivorans* genome.

Strain	Reference	Scaffolded percentage (%)	Total length		Scaffold length (Mb)	C1		C2		C3		C4	
			Mb	GC (%)		Mb	GC (%)	Mb	GC (%)	Mb	GC (%)	Mb	GC (%)
ATCC 17616	NA	NA	7.01	66.7	NA	3.45	66.9	2.47	67.1	0.92	65.8	0.08	61.7
ATCC BAA-247	BCC0084	95.99	6.20	67.1	5.95	3.14	66.8	2.29	67.4	0.52	67.6		
AU10047	BCC0084	97.78	6.34	67.2	6.20	3.26	67.2	2.36	67.5	0.57	67.7		
AU10398	ATCC 17616	97.44	6.71	67.0	6.54	3.35	67.1	2.44	67.5	0.65	67.2	0.01	62.1
AU11358	BCC0084	96.50	6.43	67.1	6.20	3.33	67.1	2.34	67.5	0.53	68.0	0.03	61.5
AU1185	ATCC 17616	100.00	6.62	66.9	6.62	3.53	66.6	2.53	67.1	0.56	67.6		
AU15814	BCC0033	99.75	6.33	67.5	6.32	3.27	67.3	2.31	67.8	0.73	67.4		
AU17545	ATCC 17616	95.84	6.98	66.6	6.68	3.38	67.1	2.40	67.0	0.84	66.2	0.61	63.1
AU18096	ATCC 17616	95.89	6.59	66.9	6.31	3.43	67.0	2.18	67.7	0.67	66.7	0.03	65.00
AU20929	ATCC 17616	98.43	6.81	66.9	6.71	3.37	67.1	2.46	67.3	0.74	67.0	0.01	61.7
AU21015	BCC0084	99.64	6.18	67.4	6.16	3.28	67.3	2.20	67.6	0.68	67.6		
AU22892	ATCC 17616	98.48	6.80	67.1	6.69	3.50	67.1	2.37	67.6	0.83	66.5		
AU24277	BCC0033	99.33	6.52	67.3	6.48	3.39	67.2	2.37	67.6	0.71	67.6		
AU28069	ATCC 17616	99.34	6.52	67.1	6.47	3.41	66.9	2.29	67.6	0.67	67.3	0.01	63.6
AU28442	BCC0084	96.79	6.24	67.3	6.04	3.20	67.3	2.29	67.6	0.55	68.0	0.03	63.6
AU4507	BCC0084	99.69	6.58	67.1	6.56	3.45	67.0	2.45	67.4	0.65	66.9		
BCC0006	BCC0084	99.34	6.27	67.3	6.23	3.37	67.1	2.33	67.6	0.54	68.0		
BCC0009	BCC0084	99.10	6.34	67.2	6.29	3.37	67.0	2.38	67.4	0.53	68.0		
BCC0032	BCC0033	96.89	6.51	67.1	6.30	3.46	67.0	2.26	67.8	0.58	67.2		
BCC0033	BCC0033	99.61	6.70	67.1	6.67	3.51	66.8	2.37	67.7	0.72	67.5	0.07	57.3
BCC0043	BCC0084	98.96	6.54	67.1	6.48	3.31	67.1	2.52	67.0	0.65	67.8		
BCC0047	ATCC 17616	97.60	7.12	66.6	6.94	3.63	66.5	2.48	67.3	0.83	66.4		
BCC0065	BCC0033	99.66	6.51	67.4	6.48	3.23	67.3	2.41	67.7	0.72	67.5		
BCC0066	ATCC 17616	98.17	6.51	67.2	6.39	3.36	67.2	2.35	67.6	0.66	67.5	0.01	62.5
BCC0068	BCC0084	96.86	6.44	67.1	6.24	3.23	67.2	2.37	67.6	0.56	67.8	0.08	62.0
BCC0074	ATCC 17616	99.01	6.54	67.3	6.48	3.43	67.1	2.29	67.7	0.75	67.5		
BCC0075	BCC0033	97.85	6.61	67.2	6.47	3.35	67.2	2.33	67.7	0.78	67.1		

BCC0079	ATCC 17616	99.53	6.99	66.8	6.96	3.70	66.4	2.40	67.7	0.68	67.4	0.02	62.2
BCC0080	ATCC 17616	98.09	6.71	66.8	6.58	3.37	66.8	2.29	67.5	0.81	66.6	0.01	62.2
BCC0082	BCC0033	95.49	6.62	67.2	6.32	3.38	67.1	2.29	68.0	0.66	67.7		
BCC0084	BCC0084	98.96	6.60	67.1	6.53	3.28	67.3	2.50	67.3	0.46	66.5	0.03	65.4
BCC0087	BCC0033	99.73	6.45	67.4	6.44	3.27	67.4	2.44	67.6	0.73	67.3		
BCC0096	BCC0084	99.09	6.38	67.2	6.32	3.33	67.1	2.30	67.7	0.68	66.7		
BCC0101	BCC0084	97.81	6.38	67.1	6.24	3.40	67.0	2.31	67.5	0.54	67.9		
BCC0141	ATCC 17616	99.39	6.17	67.4	6.13	3.25	67.1	2.28	67.7	0.60	67.7		
BCC0188	BCC0084	94.97	6.65	67.0	6.32	3.33	67.1	2.35	67.7	0.60	67.7	0.04	60.4
BCC0225	BCC0033	94.76	6.77	67.1	6.42	3.44	67.1	2.44	67.4	0.53	67.8		
BCC0241	ATCC 17616	99.66	6.66	67.1	6.63	3.34	67.1	2.44	67.4	0.67	67.5	0.02	62.2
BCC0246	BCC0033	99.35	6.47	67.3	6.43	3.29	67.2	2.43	67.5	0.71	67.6		
BCC0247	BCC0033	98.29	6.53	67.3	6.42	3.30	67.3	2.40	67.5	0.72	67.5		
BCC0264	ATCC 17616	99.68	6.63	67.1	6.61	3.27	67.3	2.47	67.3	0.88	66.3		
BCC0266	BCC0084	96.87	6.53	67.3	6.32	3.29	67.1	2.47	67.6	0.57	68.2		
BCC0269	BCC0033	99.66	6.28	67.5	6.26	3.24	67.4	2.33	67.7	0.68	67.6		
BCC0303	ATCC 17616	98.43	6.25	67.2	6.15	3.27	67.1	2.31	67.6	0.58	67.4		
BCC0317	BCC0084	99.25	6.36	67.5	6.31	3.25	67.3	2.38	67.7	0.68	67.8	0.03	63.2
BCC0375	ATCC 17616	97.74	6.30	67.3	6.15	3.24	67.3	2.29	67.7	0.61	67.7		
BCC0381	BCC0084	99.50	6.25	67.3	6.22	3.33	67.0	2.35	67.6	0.55	67.8		
BCC0384	ATCC 17616	98.72	6.38	67.4	6.30	3.24	67.4	2.36	67.7	0.58	67.8	0.01	61.8
BCC0493	BCC0033	98.90	6.53	67.1	6.46	3.30	67.1	2.51	67.0	0.65	67.7		
BCC0497	ATCC 17616	96.10	6.79	66.9	6.52	3.56	66.8	2.38	67.6	0.59	68.0		
BCC0702	ATCC 17616	96.98	6.46	67.2	6.26	3.29	67.2	2.38	67.5	0.59	67.6		
BCC0710	BCC0033	99.54	6.23	67.5	6.21	3.25	67.4	2.37	67.7	0.59	67.8		
BCC0737	ATCC 17616	99.57	6.22	67.3	6.20	3.30	67.0	2.33	67.7	0.56	67.9		
BCC0814	ATCC 17616	98.84	6.27	67.3	6.20	3.30	67.2	2.32	67.7	0.58	67.4		
BCC0865	BCC0084	93.44	6.79	66.6	6.35	3.36	66.9	2.34	67.2	0.54	67.9	0.01	62.6
BCC0904	BCC0084	98.63	6.03	67.4	5.95	3.26	67.4	2.31	67.7	0.37	67.2		
BCC0921	BCC0084	93.15	6.79	66.6	6.32	3.33	67.0	2.34	67.2	0.54	67.8	0.01	62.6
BCC1147	ATCC 17616	99.21	6.58	67.1	6.52	3.35	67.0	2.31	67.7	0.84	66.4	0.02	65.4
BCC1148	ATCC 17616	97.58	6.57	67.1	6.41	3.37	66.9	2.30	67.7	0.74	67.1		

BCC1177	BCC0084	98.85	6.52	67.1	6.45	3.28	67.3	2.43	67.3	0.60	67.4	0.01	61.0
BCC1185	BCC0033	99.72	6.38	67.4	6.37	3.22	67.4	2.42	67.6	0.07	67.3		
BCC1190	ATCC 17616	97.82	6.37	67.1	6.23	3.37	66.9	2.31	67.7	0.55	67.7		
BCC1272	BCC1272	99.21	6.88	66.8	6.83	3.37	67.0	2.43	67.3	0.87	66.1	0.02	61.3
BCC1368	BCC0084	98.19	6.27	67.2	6.16	3.19	67.1	2.38	67.6	0.57	67.5		
BCC1385	BCC0084	96.23	6.61	66.9	6.36	3.36	67.1	2.33	67.6	0.59	67.5	0.10	59.7
CF170.0a	BCC0084	99.55	6.41	67.1	6.38	3.36	67.0	2.32	67.4	0.70	66.6		
CF2	ATCC 17616	95.17	6.50	66.9	6.19	3.29	66.7	2.35	67.0	0.54	67.2		
CGD1	ATCC 17616	98.15	6.62	67.0	6.50	3.49	66.9	2.46	67.3	0.56	67.6		
CGD2	BCC0033	99.90	6.55	67.0	6.55	3.42	66.9	2.52	67.2	0.60	66.8		
D2214	BCC0033	99.98	6.47	67.4	6.46	3.28	67.3	2.38	67.7	0.73	67.4	0.07	57.6
DWS_421-B	ATCC 17616	100.00	6.51	67.3	6.51	3.37	67.1	2.38	67.5	0.76	67.4		
FDAARGOS_546	ATCC 17616	100.00	6.19	67.1	6.19	3.42	66.8	2.17	67.5	0.58	67.6	0.02	62.2
HI3534	ATCC 17616	96.90	6.16	67.0	5.97	3.33	67.1	2.39	67.3	0.24	67.0	0.01	63.7
MSMB1272WGS	ATCC 17616	95.91	6.84	67.0	6.56	3.36	67.1	2.29	67.6	0.67	67.4	0.02	61.9
MSMB1640WGS	ATCC 17616	100.00	6.85	67.0	6.85	3.45	66.9	2.49	67.2	0.90	66.1		
NKI379	BCC0033	96.78	6.54	674	6.33	3.30	67.1	2.36	67.6	0.68	67.8		
R-20526	BCC0084	96.84	6.44	67.0	6.24	3.30	67.1	2.41	67.4	0.53	67.9		
Average	NA	98.16	6.52	67.1	6.39	3.35	67.1	2.37	67.5	64.5	67.4	0.01	62.1

3.2.10 Comparison of the *B. multivorans* strains to genomic databases and genetic variation between the strains

3.2.10.1 Pan, core, and accessory genes

Roary (Page *et al.* 2015) was used to identify the pan, core, and accessory genes in the *B. multivorans* genomes, utilising the Roary plots python script which generates a pan genome frequency, gene presence absence matrix and a pie chart of the gene breakdown. This was performed for all 283 genomes, as well as the later established genome strain panel (n = 77) (Section 3.2.5.2) and the two lineages individually. The pan-genome illustrates the overall number of genes which are present in all the population. Core genes are then defined as being present in ≥95% of the population. The master *B. multivorans* pan-genome (n = 283) consists of 37,462 genes, of which 6,724 are part of the core genome (Figure 16). Of this, strain CF170.8b encoded the greatest number of genes at 6,663 genes whereas stain AU16734 encoded the least at 5,293 genes. The pan genome decreased to 22,751 genes, of which 6,873 (30.2%) are core genes, when looking at the strain panel (n = 77). When looking at the two lineages, there were 12,654 pan genes and 6,225 core genes in lineage 1 whereas lineage 2 harboured 17,928 pan genes and 6,554 core genes. The percentage of core genes decreased in both groups alone, by 12.85% for lineage 1 and 4.64% for lineage 2, compared to the 77-genome strain panel. This also showed that the core genome of lineage 2 was around 1.09x larger compared to lineage 1. However, these gene differences correlate with the greater number of strains examined in lineage 2 compared to lineage 1, causing a result bias.

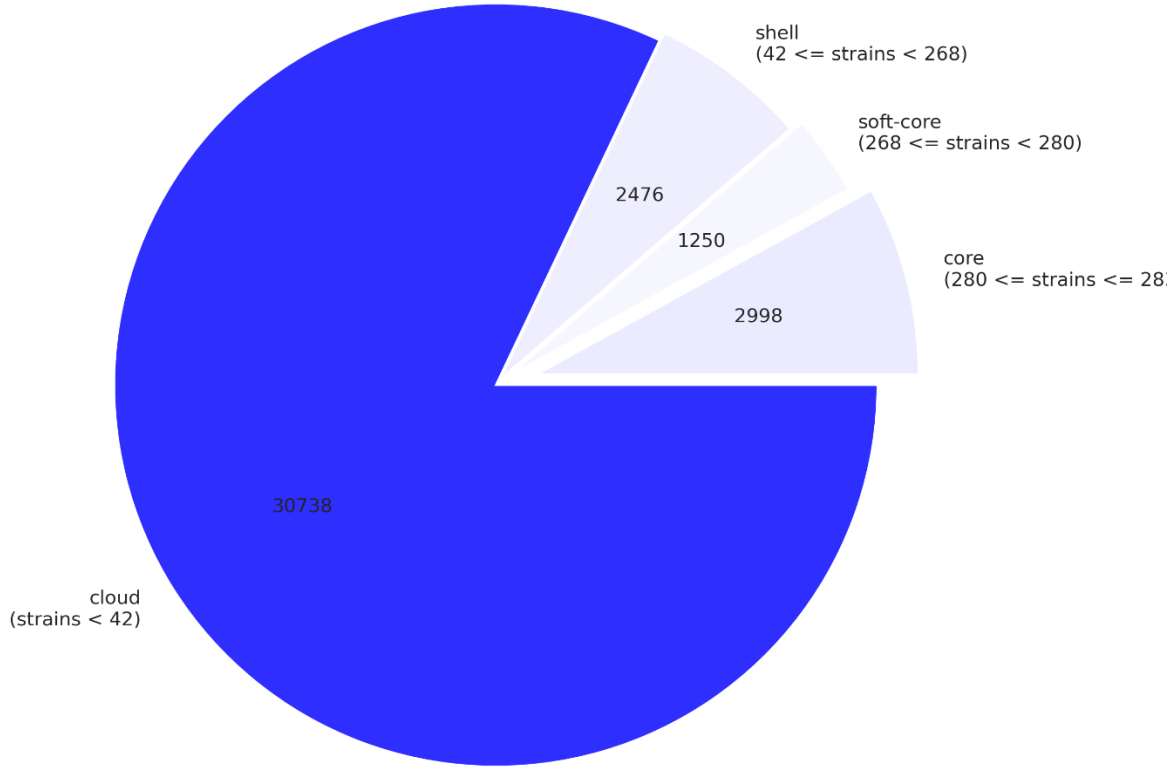


Figure 16 | Pan genome pie chart illustrating the core and accessory genome of the 283 *B. multivorans* master panel. The core genome is comprised of the core, soft-core and shell genes. The cloud genes represent the accessory genome.

The accessory genome, also known as the cloud (Whelan *et al.* 2021), is comprised by the remaining genes, present in ~14.8% of the *B. multivorans* genomes, which is roughly 4.6x the size (30,738 genes) of the core genome when looking at the 283 *B. multivorans* genomes. Comparing the accessory genes of lineage 1 and lineage 2, it was observed that lineage 2 had a greater number of accessory genes (11,374 genes) then lineage 1 (6,429 genes). This correlates to the phylogenomic analysis in Sections 3.2.5.1 and 3.2.5.2, where lineage 2 has longer genomic branches (Figure 7 and Figure 8), indicating greater evolutionary diversity.

3.2.10.2 Genomic context of the scaffolded replicons

Gene prediction and pan-genomic analysis was also performed on the reference and scaffolded replicons (C1, C2 and C3), to provide greater genomic insights into the pan, core, and accessory gene distribution throughout *B. multivorans*. All members of the strain panel (n =77) (Section 3.2.5.2) harboured all three replicons, identified during scaffolding (Section 3.2.9.1). C1, the largest of the three replicons, harboured the greatest number of 11,461 total genes. Of this, 3,204 were core genes and 8,257 were accessory genes. C2 in turn had 6,794 genes (2,445 core and 6,794 accessory). In contrast, the smallest replicon C3 encoded 2,390 total genes, with the smallest amount of both core (901 genes) and accessory genes (1,489 genes). This showed the size of each genomic replicon was relative to the number of genes it encoded.

3.2.10.3 GC content of the genomes and replicons

The total average GC content of the strain panel (n = 77) genomes (Section 3.2.5.2) is 67.13%, ranging from 66.58% (BCC0865) to 67.51% (BCC0710). Overall, the GC content of the whole genomes was relatively similar, with only a 0.55% GC difference. When splitting this into replicons, the mean GC content was relatively consistent at 67.1%, 67.5% and 67.4% for C1, C2 and C3, respectively.

The average GC (%) content for replicon C4, found in around one third of the *B. multivorans* panel strains (Section 3.2.5.2), is 62.1% (Table 18). This is slightly lower than that of replicons C1, C2 and C3 at a difference of, 5.4% and 5.2%, respectively. The GC content of C4 differs between the *B. multivorans* strains more so than the other replicons. BCC0033 had the lowest GC of 57.32% for C4, with ATCC 17616 having the highest at 61.74%, a difference of 4.42%.

3.3 Discussion

This work has enabled us to identify key genotypic features amongst *B. multivorans* as a species. The *Burkholderia* genus has seldom been investigated in terms of in-depth next generation sequencing phylogenetic analysis and population biology, with MLST-level being the most highly understood (Baldwin *et al.* 2008). The work here embraces next generation sequencing techniques to underpin the *B. multivorans* population at whole genome level. Short-read Illumina sequencing was used predominantly in this Chapter, also with complete genomes from public databases, defining the placement of 283 *B. multivorans* genomes before curating a representative strain panel of 77 *B. multivorans* genomes for further analysis. The population biology of *B. multivorans* has been thoroughly investigated in this Chapter, opening the door for exploratory genomic analysis into the two defined genomic lineages.

3.3.1 At least two distinct evolutionary lineages were observed within the *B. multivorans* population

B. multivorans was confirmed as a single genomic species through use of ANI analysis (n = 283). The results showed all strains to have ≥97.24% identity, over the 95% threshold which defines a bacterial species (Goris *et al.* 2007). The population split into at least two lineages, now defined as *B. multivorans* lineage 1 and lineage 2, with the latter splitting further into sub-lineages 2a and 2b (Sections 3.2.4.1**Error! Reference source not found.**, 3.2.4.2 and 3.2.5.1). These similarities mimic that of related CF pathogens *P. aeruginosa* (Ozer *et al.* 2019; Weiser *et al.* 2019) and *B. cenocepacia* (Wallner *et al.* 2019), where phylogenomic analysis has also demonstrated the presence of genomic subgroups.

Interestingly, there is a broad placement of non-CF isolates, but it was observed that lineage 2 held the largest proportion of environmental *B. multivorans* isolates. These particularly clustered within sub-lineage 2a. When there is lack of patient-patient strain transmission, the environment is the most likely source of *B. multivorans.* Combining our knowledge on this and fact that the BCC are opportunistic

pathogens where pathogenicity is highly dependent on host vulnerability, the data strongly suggests that environmental isolates cannot really be split from infection isolates in the context of phylogenomics.

By understanding how the population biology of these related CF bacteria affect phenotype, genotype, outbreaks, and clinical outcome, we can begin to hypothesise that similar factors may occur in *B. multivorans.* There is not enough clinical data at present to create links and identify patterns in the data compared to *B. multivorans* lineage, however, this is certainly a topic for further research. Rapid diagnostics to identify each *B. multivorans* lineage would greatly assist in perspective analysis of virulence traits and clinical outcome for CF infections. The genomic data was used to develop PCR probes to enable this as described in Section 4.2.8.

3.3.2 The two-lineage split is also observed at MLST-level

This work shows clearly that the two-lineage split in the *B. multivorans* population was not only seen at whole genome level but could also be observed at MLST-level. However, some differences in phylogenetic placement were found within lineage 2. The analysis also showed longer, deeper branches within the population at using MLST, indicating a larger genetic divergence than observed with the core gene trees. Another main difference was the confidence in bootstrap values between the two analyses. These "confidence" values were introduced by Felsenstein (1985) to help understand the reliability of the clades and lineages within a phylogenetic tree. This method is reliable for closely-related bacterial species (Maiden *et al.* 2013), but is disadvantageous when looking to analyse multiple species within a genus, for example. Prior to WGS technologies, isolate identification and typing would require multiple approaches for accuracy and reliability (Maiden *et al.* 2013). Therefore, whilst we can see the split using MLST in this dataset, a core gene phylogeny is required for a more accurate identification of bacterial populations.

MLST is also useful for understanding the ST of the bacterial isolate and identifying any clonal complexes in the dataset. Six clonal complexes (CC1, CC2, CC4, CC5, CC7 and CC8) were identified within the *B. multivorans* strain panel. CC1, CC4 and CC5 have been found in varies countries and continents (Baldwin *et al.* 2008) CC1 contains the South Wales outbreak strain (ST-15) (Millar-Jones *et al.* 1998) and the globally disseminated French epidemic clone (ST-16) (Segonds *et al.* 1997; Baldwin *et al.* 2008). Both have caused multiple infections within CF patients. Looking closer into clonal complexes and the types of infection they cause, such as the human infection by ST-16 from CC1, we can hypothesise which *B. multivorans* strains within the dataset could be better adapted to human infection.

We can also underpin epidemiology and identify any STs which may have links to previous *B. multivorans* outbreaks. For instance, ST-16, also known as the French epidemic clone, has been linked to other global outbreaks (Baldwin *et al.* 2008). Understanding these STs means that we can create

probable links to countries and between isolation sources. For instance, Baldwin *et al.* (2008) showed that 50% of the environmental *B. multivorans* isolates examined were indistinguishable to CF isolates.

3.3.3 There are genomic differences between the two lineages

Understanding the core gene phylogeny of the *B. multivorans* strain panel has helped set indicators for population evolution. Lineage 1 have short branches whereas lineage 2 contrastingly have deep evolutionary branches. This presents the idea of lineage 1 being more conserved than lineage 2. This can be investigated through pan genome analysis. ANI analysis is also very useful here, clearly distinguishing genomic relatedness both in the overall *B. multivorans* population as well as within the lineage groupings. Wallner *et al.* (2019) showed that ANIb could be used to discriminate between lineages, as shown by proposing *B. cenocepacia* IIIB as a separate species to IIIA. ANIm was used within this work to look at the *B. multivorans* population. Using ANIm has its limitations compared to ANIb, especially in effectiveness. ANIm often has a larger number of false positives in genome pairs compared to ANIb (Yoon *et al.* 2017). Although, ANIb is very computationally intensive and has a longer running time than ANIm. To overcome this limitation, and ensure the two-lineage split was certain, a secondary ANI analysis was performed using FastANI. This removes the requirement for genome alignment. Instead, orthologous gene pairs are identified when comparing two bacterial genomes. The percentage ANI (%) was highly comparable to that shown using PyANI, with an overall mean of 97.8% amongst the *B. multivorans* strains.

Whilst much is known about evolutionary lineages in bacterial populations, descending from a common ancestor, little is understood around gene gain and loss within lineages as they evolve over time within the CF lung (Gabrielaite *et al.* 2020). SNP diversity analysis could be used to highlight similarities and differences within the lineages. This could indicate towards any isogenic isolates as well as clusters of populations. Insertion sequences (IS) have also been shown to shape bacterial lineage evolution in other *Burkholderia* species, such as in *B. cenocepacia* and the emergence of *B. mallei* from *B. pseudomallei* (Graindorge *et al.* 2012). Niche adaptations occur due to IS elements entering the bacterial genome, making the organism more specialised to human CF infection.

Peeters *et al.* (2017) showed the conserved genomic structure of *B. multivorans,* making links between ST and genomic lineage in a limited analysis of 8 complete genome sequences, comprising pairs of environmental versus CF isolates. They showed that genomic lineage is directly linked to ST with high conservation. Within the thesis genomic dataset, some correlations were observed between ST and lineage. Lineage 1 harboured all ST-15 and 75% of the ST-16 isolates (CC1). ST-18 (Canada and UK) and ST-650 (ATCC BAA-247; Belgian clone) were the other two STs found solely in lineage 1. Lineage 2 held ST-21(Canada and USA), ST-802 and ST-618 not found in lineage 1. ST-21 is composed of both environmental and CF isolates (Baldwin *et al.* 2008). From this, suggestions can be made that lineage 1 harbours strains more likely to cause epidemic outbreaks, corresponding to what is seen within the *B.*

cenocepacia IIIA lineage (Drevinek and Mahenthiralingam 2010). Further investigations are required to either prove or disprove this assumption, using genome analysis to pinpoint key genes different between the lineages.

3.3.4 We have selected at least 3 model CF *B. multivorans* strains for further analysis

Before this work, there were no defined model *B. multivorans* for CF infection. Three model CF *B. multivorans* strains have now been identified in this Chapter (BCC0033, BCC0084 and BCC1272), as well as the well-characterized environmental isolate ATCC 17616. These were chosen based on their consistent phenotypes from Chapter ***5*** as well as genotypic analysis (Chapters ***3*** and ***4***) and phylogenetic placement. The model strains mirror the key genotype and phenotype seen within the *B. multivorans* panel. These model strains will be used later in this thesis to present co-infection modelling with other CF pathogens, for use in murine inhalation models and for genome analysis. Further understanding all these characteristics will help with intervention of chronic infection or interference through specific targets and treatments.

3.4 Conclusions

Following the population biology and pan-genome analysis of the *B. multivorans* strains, the following conclusions can be made:

1) *B. multivorans* has a high genomic diversity between strains, and the novel lineages that have been identified
2) The population biology of *B. multivorans* splits into at least two genomic groups, consistent when using different core-gene phylogenomic analyses as well as MLST phylogeny
3) This work has opened the door for genome analysis of lineage-specific genes and development of PCR probes to help with rapid identification in future
4) At least 4 model *B. multivorans* have been identified from the genomic lineages which can be used for further comparative analysis and genome analysis

4 Identification of Key Genomic Features Within the *B. multivorans* Strain Panel Using Comparative Genomics and Genome Analysis

4.1 Introduction

Complete bacterial genomes provide a foundation for greater analysis depth in terms of genome analysis. Whilst various long-read sequencing methods exist, one of the most common and accurate is Single Molecule, Real-Time (SMRT) Pacific Biosciences (PacBio) which can provide reference-standard assemblies (De Maio *et al.* 2019). Combined with short-read sequences, identification of key genes can be determined. Bioinformatic and phylogenetic tools can be used to highlight genes unique to certain groups of strains or species (Wallner *et al.* 2019), as well as potential virulence and pathogenicity genes within a genome.

There are multiple virulence and pathogenicity genes which may confer advantages to a pathogen in CF infection, particularly for persistence and advancing chronic lung infection. Many virulence determinants have previously been noted for the BCC (Mahenthiralingam *et al.* 2005; Loutet and Valvano 2010). For example, major virulence determinants of *B. cenocepacia* which have been related to CF infection include acquisition of foreign DNA encoding novel virulence factors, quorum sensing, and iron acquisition (Drevinek and Mahenthiralingam 2010). *Burkholderia* species. are also well-known for their intrinsic resistance to antibiotics and mediation of inflammation through polysaccharides on the cell surface and membrane (Drevinek and Mahenthiralingam 2010). Secretion systems have also been associated with *B. cenocepacia* infection, showing survival during murine experimentation (Loutet and Valvano 2010). There are several bioinformatics tools which are useful for identifying potential virulence and pathogenicity genes within bacterial species. One widely used example is Abricate, a tool which performs mass contig screening for various genes based on multiple, compiled virulence and antibiotic resistance databases (Seemann 2017).

Amongst these bioinformatics tools, there are various databases for identifying Clusters of Orthologous Groups (COGs) (Tatusov *et al.* 2000). These COGs are annotated genome coding sequences (CDS) which are conserved in sequence and can be frequently linked to a functional role. Initially, classification of protein families was performed using COG analysis (Tatusov *et al.* 2000). However, it is now used for genome annotations, to identify undetected or missing genes in a bacterial genome, to predict novel functional systems, analyze pathways, predict enzymes and for comparative COG analysis between organisms or bacterial species (Galperin *et al.* 2019). One useful tool COG analysis provides us with is the ability to prioritize targets for both structural and functional characterization (Galperin *et al.* 2019). A useful way of analyzing COGs in genomes is by utilizing the public EggNOG database (Jensen *et al.* 2008). In the most recent update (2019), the eggNOG database harbours over 4,445 bacteria as well as 4.4 M orthologous groups overall (Huerta-Cepas *et al.* 2019).

As mentioned, *Burkholderia* species are well-known for their evolution by the acquisition of foreign DNA and providing considerable genome plasticity. Genomic islands (GIs) are amongst the common DNA elements which have been extensively found in *Burkholderia* species, particularly *B. cenocepacia* strain J2315 (Guo *et al.* 2017). GIs are sets of genes acquired via horizontal gene transfer (HGT) (Gogarten and Townsend 2005), which can be advantageous to bacteria as they can carry genes required for niche adaptation (Dobrindt *et al.* 2004). Those GIs which confer the ability for bacterial fitness and survival are split into categories, including pathogenicity islands (PAIs) (Hacker and Carniel 2001). These PAIs are widespread throughout bacteria, including within the BCC. Baldwin *et al.* (2004) functionally defined the first genomic island in the BCC, found in *B. cenocepacia*, known as *cenocepacia* island (cci). This was an atypical PAI compared to others seen in Gram-negative bacteria (Baldwin *et al.* 2004). Often, PAIs will carry well known pathogenicity genes such as secretion systems, T3SS or T4SS (Hacker and Carniel 2001), but the cci was unusual in carrying quorum sensing genes, cciI and cciR as key virulence determinants, together with multiple metabolism-associated genes (Baldwin *et al.* 2004). This opens the idea that PAIs may contribute to both ecological fitness as well as links to metabolic genes in *Burkholderia* species (Baldwin *et al.* 2004).

In addition to HGT, plasmids are key agents in driving bacterial adaptation and evolution. Not only can they mediate the transfer of genetic material between bacterial species, but plasmids may also contribute to chromosomal evolution and adaptations through the accessory genes that they carry (Rodríguez-Beltrán *et al.* 2021). There are two broad types of plasmids which are found in bacteria. The larger of the two are low copy number plasmids (LCPs) and the smaller being high copy number plasmids (HCPs) (Rodríguez-Beltrán *et al.* 2021). Plasmids are not uncommon in *Burkholderia* species. The *Burkholderia* species are known to have a multi replicon structure (Mahenthiralingam *et al.* 2005), where the 3rd replicon has now been found to act as a large virulence plasmid (which can be forced out), as opposed to an essential replicon (Agnoli *et al.* 2012). Further research has shown a 4th replicon element in several *Burkholderia* species. For example, the *B. cenocepacia* J2315 strain is comprised of three circular genomic replicons and a 0.927 Mb plasmid (pBCJ2315) (Holden *et al.* 2009). *B. multivorans* ATCC 17616 has also been shown to harbour a plasmid, pBMUL01 (Sant'Anna *et al.* 2009; Yagi *et al.* 2009).

Genomic rearrangement by the movement or loss of insertion sequences, GIs, and recombination also represents a rapid way by which *Burkholderia* can gain or lose genes (Mahenthiralingam and Vandamme 2005). On a smaller scale, microbial evolution can also be driven by base mutations in the genome, known as variants or single nucleotide polymorphisms (SNPs). This is where nucleotide substitutions, insertions or deletions occur (Bryant *et al.* 2012). These mutations can be synonymous, meaning they do not affect the resulting amino acid. Contrastingly, nonsynonymous mutations do affect the DNA sequence, altering the final amino acid (Bryant *et al.* 2012). Whilst SNPs and variation occur during bacterial evolution, they may not always remain fixed within a population. They are often

associated with selection pressures, both positive and negative. It is advantageous for a bacterium to only keep genes required for survival. Metabolizing unrequired genes is costly in terms of energy consumption and may catalyze a gene loss effect overall. Therefore, we can often infer 'fixed' SNPs to be an inference to niche adaption. A paper by Hassan *et al.* (2020) described that SNPs observed in *B. multivorans* and *B. cenocepacia* CF co-infection were centred around metabolism, oxidative stress responses and antibiotic resistance. All of these are positive-selection mutations in the CF lung, where oxygen is limited in dense sputum and antibiotics are used to combat infection (Hassan *et al.* 2020).

Looking deeper into bacterial evolution and potential virulence factors, the *B. cenocepacia* J2315 low oxygen-activated (lxa) locus is an interesting cluster of genes. This novel 50 kb locus up-regulates potential virulence factors, like metabolism, transport and regulation, when exposed to stressful low oxygen growth conditions (Sass *et al.* 2013). Interestingly, the lxa locus is not always present as a conserved single locus in the BCC, with either partial or no lxa-associated genes in strains of *B. multivorans* and *B. cenocepacia* (Sass *et al.* 2013). However, Nunvar *et al.* (2016) showed that *B. contaminans* has a consistent presence of the locus. The lxa locus contributes to growth and survival in stressful environments, with the key phenotype associated with its presence being enhanced survival under anaerobic conditions (Sass *et al.* 2013), ideal for adaptation to the chronically infected CF lung.

Use of PCR amplification is a powerful tool for identification of bacteria and has been widely applied to BCC CF infection. For example, improved identification of BCC was driven by have been made for the design of species-specific PCR primers for BCC was driven by the design of species-specific PCR primers based on the *recA* gene (Mahenthiralingam *et al.* 2000). Development of strain-specific PCR assays adds another level to BCC detection in CF patients, such as using the *B. cepacia* epidemic strain marker (BCRSM) PCR to identify epidemic strains with high risk of transmission (Mahenthiralingam *et al.* 1997). However, as BCC taxonomy improved, the BCESM PCR has been superseded for instance, Dedeckova *et al.* (2013) establishing a novel assay for *B. cenocepacia*, focusing on ST-32 epidemic strains which have caused the major outbreak in CF infections seen it Prague, Czech Republic. Using molecular assays like this creates a more rapid and accurate detection of highly transmissible, more threatening BCC strains. With the expansion in genomic analyses, unique genes can be pinpointed through genome analysis to identify key bacterial groups or strains and underpin PCR primer design.

4.1.1 *Aims and objectives*

The overall aim of this Chapter was to identify the key genomic features and expand genomic understanding of the *B. multivorans* strains using the following objectives:

1) Identification of genes predicted to be associated with antimicrobial resistances and pathogenesis (note throughout this Chapter and thesis all genes described and considered as predicted in terms of the putative function).

2) To compare the genomic content of the strains within each genomic lineage, and identify genes which are unique to each

3) To completely sequence and assemble the three identified *B. multivorans* model CF strains, and categorise the functional gene categories within them

4) To identify prophages and plasmids in the strain panel, and SNP variants within a representative *B. multivorans* strain panel

5) To evaluate whether the *B. multivorans* model strains were amenable to genetic manipulation by the introduction of the reporter plasmids harbouring fluorescent-proteins encoding genes

6) To perform infection modelling of the *B. multivorans* model strains using a mouse inhalation model of lung infection and genomic re-sequencing of post-infection isolates

Hypothesis 1: Genomic lineage-specific genes can be identified for *B. multivorans*

Hypothesis 2: There will be differences between the genomic content of the two *B. multivorans* lineages

Hypothesis 3: The model *B. multivorans* strains will be amenable to plasmid uptake and basic genetic manipulation

4.2 Results

4.2.1 Key virulence and pathogenicity genes

1.1.1.1 Virulence and pathogenicity gene categories

The gene matrix from Roary was manually investigated for virulence and pathogenicity genes in the dataset of $n = 77$ *B. multivorans* genomes. This was performed by grouping the genes into categories and totalling the number of gene counts per category for each *B. multivorans* strain (Figure 17). Overall, the quorum sensing gene category had the greatest number of genes (mean = 54.6) in the *B. multivorans* strain panel, closely followed by invasion and flagella (mean = 46.6), then drug resistance (mean = 35.5) genes. Only 8 *B. multivorans* strains genomes harboured toxin-antitoxin genes, which was the least abundant gene group amongst the strains (Figure 17).

The mean and median number of genes in each category for each lineage was similar. However, there were differences in the distributions (standard error of the mean) of gene hits for most categories in both lineages. This could be due to bias in the number of genomes in each lineage (lineage 1: $n = 27$; lineage 2: $n = 50$). Therefore, the results were statistically analysed, grouping by genomic lineage, to observe any patterns. A Dunn-test revealed that there was a significant difference ($p = <0.0001$) in the number of genes related to biocide and chemical resistance, biofilm and adhesion, drug resistance, haemolysins, outer membrane proteins and quorum sensing between the *B. multivorans* lineages. Genes related to biofilm formation and haemolysins were significantly greater in numbers for lineage 1. Lineage 2 had a significantly greater number of invasion and flagella, T6SS genes, stress proteins, and cation efflux genes (Table 19).

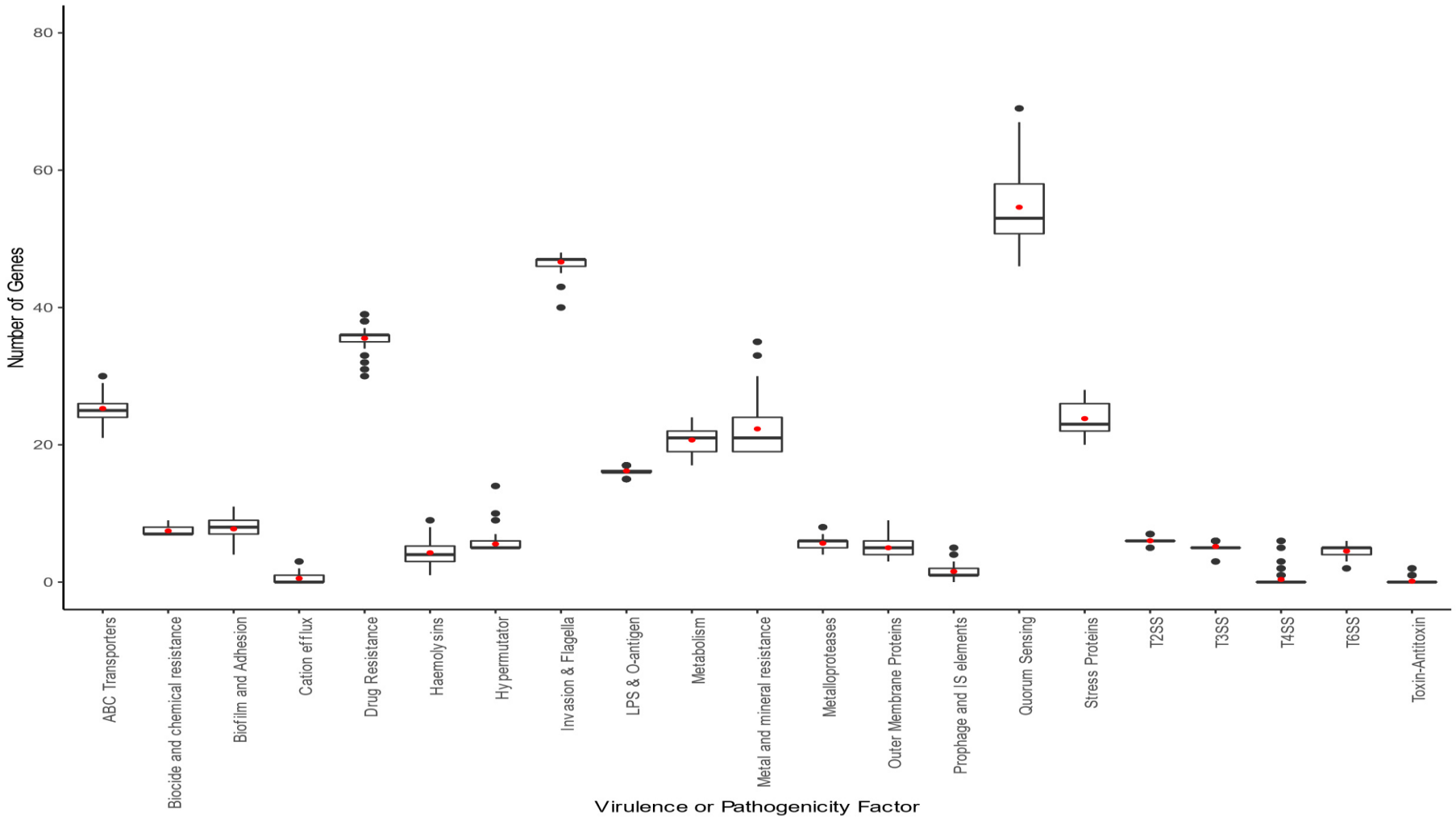


Figure 17 | Virulence and pathogenicity factors within the *B. multivorans* strains genomes. The box plot illustrates the number of predicted genes within each functional grouping per *B. multivorans* strain (n = 77 genomes analysed), with the means are illustrated by the red point.

Table 19 | Virulence and pathogenicity gene groups associated with the two *B. multivorans* genomic lineages (*n* =77)

Virulence or pathogenicity factor	Lineage	Mean number of genes in category	Standard error	Significance between the lineages
ABC Transporters	1	25.3	1.5	ns
	2	25.3	1.5	
Biocide and chemical resistance	1	7.0	0.7	****
	2	7.5	0.7	
Biofilm and Adhesion	1	7.8	1.6	****
	2	7.7	1.6	
Cation efflux	1	0.5	0.7	*
	2	0.5	0.7	
Drug Resistance	1	35.5	1.7	****
	2	35.5	1.7	
Haemolysins	1	4.3	1.7	****
	2	4.2	1.7	
Hypermutator	1	5.5	1.3	ns
	2	5.5	1.3	
Invasion & Flagella	1	46.6	1.2	***
	2	46.6	1.2	
LPS & O-antigen	1	16.2	0.5	ns
	2	16.2	0.5	
Metabolism	1	20.7	1.7	ns
	2	20.7	1.7	
Metal and mineral resistance	1	22.1	3.6	ns
	2	22.4	3.9	
Metalloproteases	1	5.6	0.8	ns
	2	5.7	0.8	
Outer Membrane Proteins	1	5.0	1.4	****
	2	5.0	1.4	
Prophage and IS elements	1	1.6	1.1	ns
	2	1.5	1.1	
Quorum Sensing	1	54.5	5.1	****
	2	54.7	5.2	
Stress Proteins	1	23.7	2.4	**
	2	23.8	2.4	
T2SS	1	6.0	0.2	ns
	2	6.0	0.2	
T3SS	1	5.2	0.5	ns
	2	5.2	0.5	
T4SS	1	0.4	1.2	ns
	2	0.4	1.2	
T6SS	1	4.5	0.7	***
	2	4.5	0.7	
Toxin-Antitoxin	1	0.1	0.4	ns
	2	0.1	0.4	

1.1.1.2 Virulence and pathogenicity genes in the draft genomes

Abricate was used to identify virulence and pathogenicity genes within the *B. multivorans* strain panel ($n = 77$) using the short-read sequences. This included use of the databases ARG-ANNOT, MEGARES, VFDB, CARD, PlasmidFinder and Resfinder (see Methods for each analysis). ARG-ANNOT, PlasmidFinder and Resfinder did not provide any gene his against the *B. multivorans* strain panel. The results of all other databases were then compiled, and any duplicate hits removed. In total, 4190 predicted virulence and pathogenicity genes were identified across the 77 genomes. There were 61 genes identified as part of this analysis that had at least one genome associated with it (Figure 18). *B. multivorans* BAA-247 had the least number of virulence genes (46 genes) detected through Abricate. AU18096 had the most virulence genes (58 genes) amongst the panel strains. Thirty-tree genes were present in 100% of the genomes analysed, encompassing genes associated with efflux pumps, drug resistance, chemotaxis and motility, and cell regulation (Figure 18).

Virulence gene absence was observed in certain *B. multivorans* strain genomes. There were 21 *B. multivorans* strain genomes which did not harbour the *fliC* flagellin gene. Multiple other flagella/chemotaxis genes were absent across the strains. Chemotaxis gene *cheA* was absent in 3 strains (ATCC BAA-247, CF2 and CGD2), with *cheB* also absent in ATCC BAA-247 (Figure 18). These genes have been shown to affect flagellar rotation in *E. coli* (Parkinson 1976). The ATCC BAA-247 strain was missing 6 other flagella/chemotaxis-related genes (*flgG, flgJ, flhF, fliE, fliH, fliS*) and pilin gene *pilB*. *B. multivorans* strain AU10047 was also missing 6 genes associated with flagella and chemotaxis (*cheW, flhB, flhF, flhG, fliA*, and *fliL*). Strains CF2 and AU10398 were missing 3 and 4 chemotaxis/flagella genes, respectively. All other *B. multivorans* strains analysed had either one or no missing genes in this virulence group (Figure 18).

There were interesting differences in the T6SS as follows. The VipB protein forms a complex with VipA in *V. chlolerae* as part of the T6SS in this pathogen (Bönemann *et al.* 2009; Chen *et al.* 2011). Gene *hsiB1/vipA* was absent in 36 of 77 (46.8%) *B. multivorans* strains analysed (Figure 18). This absence did not correlate with genomic lineage. Ten of the strains (27.8%) which were missing *hsiB1/vipA* also showed an absence of *hsiC1/vipB* (Figure 18).

The *pmlI/bspI1* gene was absent in 24 *B. multivorans* panel strains, all of which were lineage 1 strains. However, *pmlI/bspI1* was present in model strain BCC0084 (Figure 18). PmlI is a component of the *B. pseudomallei* quorum sensing system (Valade *et al.* 2004)

Three *B. multivorans* genomes also had a unique gene presence of several genes. The *penA* gene was observed in only two *B. multivorans* strains (BCC1272 and ATCC 17616) in the strain panel. The location of this gene is further discussed in Section 1.1.1.3, focussing on the complete genomes. Strain AU18096 was the only member of the strain panel to harbour the lipid biosynthesis genes *wcbR, wcbS*,

and *wcbT*. All three genes have been implicated in anti-phagocytosis mechanisms of *B. pseudomallei* through lipid biosynthesis (Cuccui *et al.* 2012).

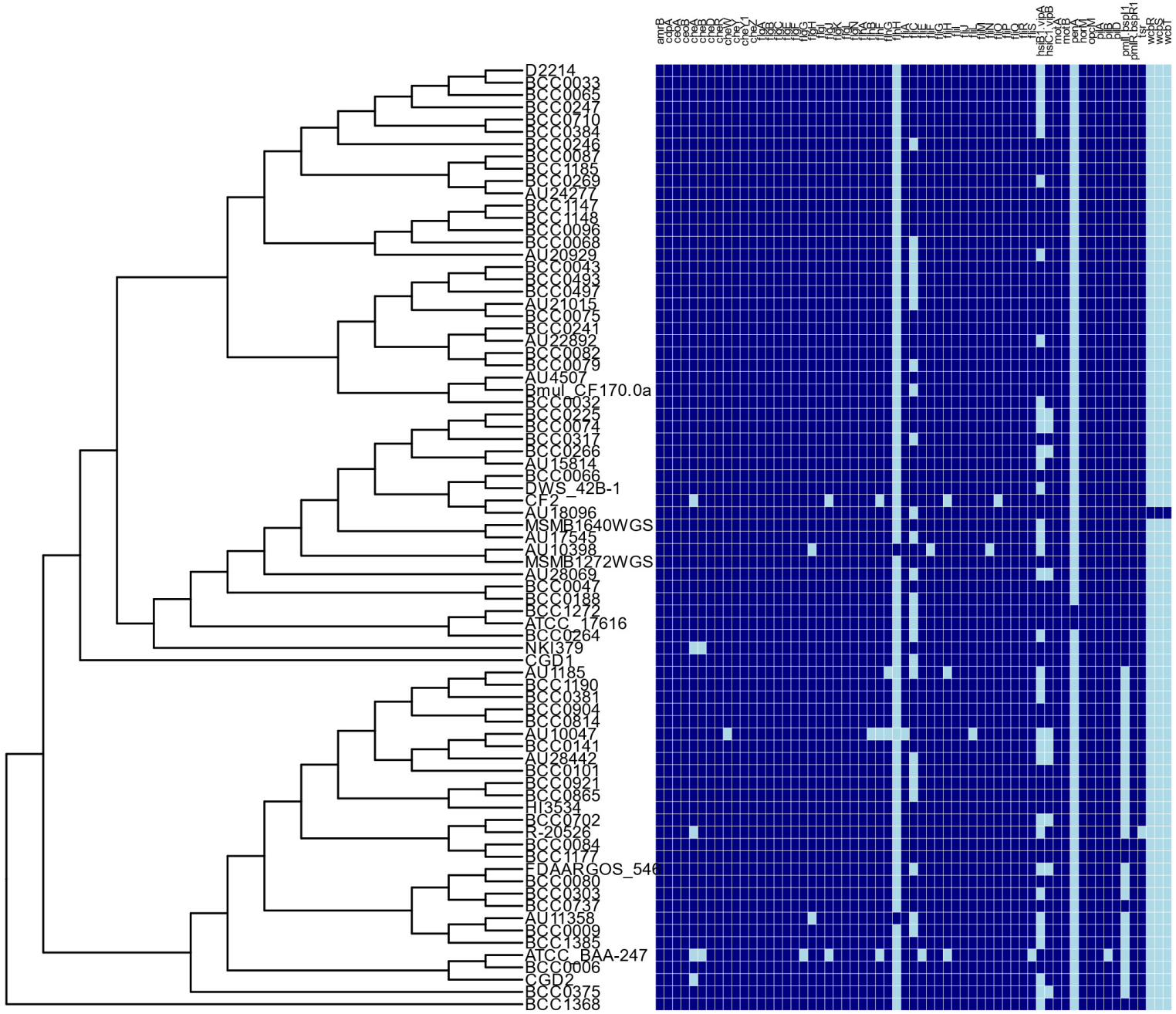


Figure 18 | Presence absence matrix of virulence and pathogenicity genes in the *B. multivorans* strain panel. This analysis used Abricate to identify gene hits in the 77 *B. multivorans* draft genomes. Any duplicate hits from multiple Abricate databases were noted as one overall hit. The dark navy blue indicates gene presence and light blue indicates gene absence.

1.1.1.3 Using the model strains to identify virulence and pathogenicity gene locations on the *B. multivorans* replicons

The complete genomes the model *B. multivorans* strains (n = 4), ATCC 17616 (lineage 2a), BCC0033 (lineage 2b), BCC0084 (lineage 1) and BCC1272 (lineage 2a), were subsequently scanned for virulence and pathogenicity genes of interest. This was performed to identify the genomic location of the genes on the replicons and to see whether there were any differences between the lineages. The same databases were used as above and compiled as before. A total of 58 virulence and pathogenicity genes were identified for each individual *B. multivorans* complete genome. Most virulence genes in each *B. multivorans* model strain were associated with flagella (Table 20).

Table 20 | Putative virulence and pathogenicity genes in the *B. multivorans* complete genomes (n = 4). Gene predictions created using Abricate (Seemann 2017).

***B. multivorans* strain**	**Gene category**	**Total number of genes in that category**
ATCC 17616	AMR	6
	Chemotaxis	9
	Flagella	33
	OMP	2
	QS	2
	Regulator	1
	T4SS	3
	T6SS	2
BCC1272	AMR	6
	Chemotaxis	8
	Flagella	34
	OMP	2
	QS	2
	Regulator	1
	T4SS	2
	T6SS	3
BCC0084	AMR	5
	Chemotaxis	8
	Flagella	35
	OMP	2
	QS	2
	Modulator	1
	T4SS	3
	T6SS	2
BCC0033	AMR	6
	Chemotaxis	8
	Flagella	35
	OMP	2
	QS	2
	Modulator	1
	T4SS	3
	T6SS	1

All putative genes had a percentage identity of at least 75% when analysed and a coverage of between 1.37% and 100%; this was used as an initial analysis to capture a broad number of predicted genes or gene fragments (see below for the restricted screen at 75% coverage). It was observed that most virulence and pathogenicity genes resided on both replicons 1 and 2 of the *complete B. multivorans* strains. Replicon 1 harboured 60 virulence and pathogenicity-associated genes for ATCC 17616 (lineage 2a), BCC0084 (lineage 1) and BCC1272 (lineage 2a) and 61 genes for BCC0033 (lineage 2b). Replicon 2 held a smaller number of pathogenicity and virulence genes amongst the four model strains with 18, 18, 15 and 14 for ATCC 17616, BCC1272, BCC0033 and BCC0084 respectively. Only one virulence gene, *mexK* (predicted inner membrane RND efflux transporter) was observed on replicon 3 of strains ATCC 17616 and BCC1272 (coverage of 10.36% and identity of 75.85% for both strains).

For further analysis, only genes with a percentage coverage of at least 75% were considered. Figure 22 shows the four complete genomes and the genomic locations of the associated virulence and pathogenicity genes. Most virulence and pathogenicity genes are localised to the same genomic locations in each of the *B. multivorans* strains, but two key differences were noted. Replicon 1 harboured chemotaxis, flagella, Type IV pilli, and modulating genes whereas replicon 2 had more AMR, OMP, T6SS and QS genes. Most genes in replicon 1 were associated with chemotaxis and flagella. These were located between ~3.17-2.00 kb and ~27.90-34.06 k*B.* Replicon 1 also harboured genes *norM* and *cdpA*, located in different places in the *B. multivorans* genomes. For ATCC 17616 and BCC1272 (both lineage 2a), the genes are located ~230-245 kb whereas in BCC0084 (lineage 1) and BCC0033 (lineage 2b) the genes are located ~100-120 kb (Figure 22). NorM is a multi-drug efflux protein which provides drug and biocide resistance (Podnecky *et al.* 2015). CdpA has been shown to modulate motility in *B. cenocepacia*, using c-di-GMP (Lee *et al.* 2010). The Type IV pilus genes, genes *pilA, pilD* and *pilB*, found in replicon 1 also had differing genomic locations amongst the strains. PilD and PilB were observed around 50-55 kb in BCC0033 and BCC0084, but the same two genes were observed between 305-310 kb in BCC1272 and ATCC 17616. Interestingly, the other Type IV pilus gene, *pilA*, appeared to almost switch places with the *pilB* and *pilD* genes in each genome. The *pilA* gene was located around 70-75 kb in ATCC 17616 and BCC1272, and ~270-280 kb in BCC0033 and BCC0084 (Figure 22). Overall, this suggested there is not a link between genomic lineage and virulence gene placement in terms of Type IV pilus genes.

When comparing virulence/pathogenicity gene presence and absence of replicon 1, gene *fliC* was missing from the BCC1272 and ATCC 17616 (lineage 2a) genomes yet present in BCC0033 (lineage 2b) and BCC0084 (lineage 1). Gene *fliC* encodes the major flagellin subunit and is associated with initiation of an immune response in other BCC species (Hanuszkiewicz et al., 2014). Deletion of *fliCII* in *B. cenocepacia* has previously been associated with a reduction in mortality compared to the

respective wild types (Urban *et al.* 2004). However, a gene absence of *fliC* was shown not to affect its ability to colonise and persist within murine lungs (Roux *et al.* 2018). Virulence of the 4 *B. multivorans* model strains have been assessed using murine models in Section 4.2.10.

Replicon 2 also had a distinct difference when comparing the presence-absence virulence and pathogenicity genomic content. PenA, encoded by gene *penA*, was present in BCC1272 and ATCC 17616 (lineage 2a) but absent in BCC0033 (lineage 2b) and BCC0084 (lineage 1). PenA, is located on replicon 2 of the two lineage 2a genomes at ~639 k*B.* This gene has been shown to play a role in ceftazidime resistance mediation in *B. pseudomallei* (Randall *et al.* 2015). Other genes encoded on replicon 2 included AMR genes, OMPs, QS proteins and T6SS genes. The CeoAB-OpcM RND efflux pump complex, which confers multidrug resistance in *Burkholderia* species, (Podnecky et al., 2015) was observed ~25-27 kb in each genome. Clustered more tightly together, ~85-95 kb are the T6SSs *hsiB1/vipA* and *hsiC1/vipB*, OMP *omp38*, and QS *pmlI/bspI1* and *pmlR/bspR1* genes.

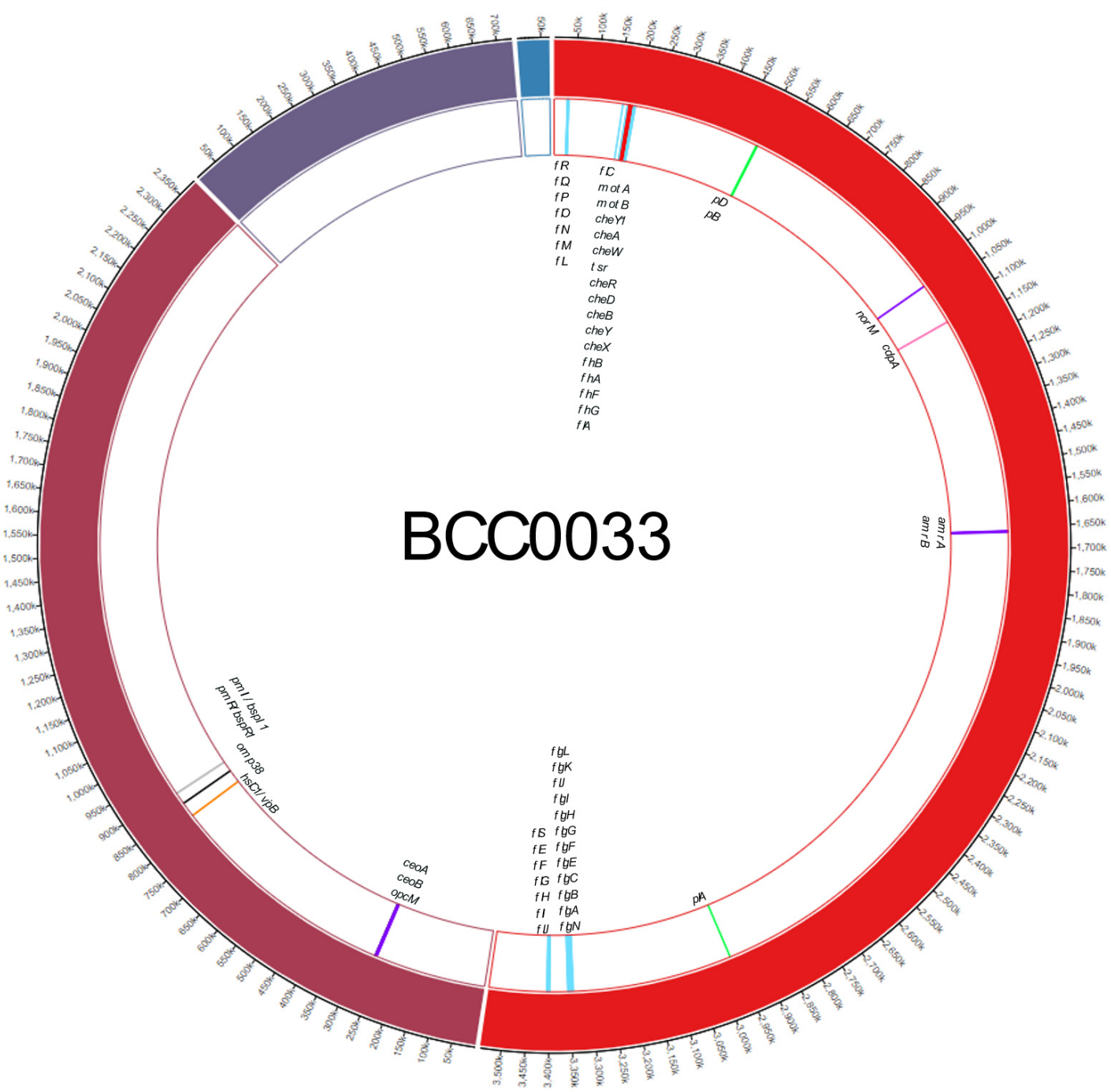


Figure 19 | Location of virulence and pathogenicity genes in the *B. multivorans* strain BCC0033. Key genes were identified using Abricate databases against the complete genome. Figure was drawn using Circa and edited using Inkscape. The gene locations are colour coded by group. Flagella = light blue, chemotaxis genes = red, T4 pili = lime green, T6SS genes = orange, QS genes = grey, OMP = black, AMR genes = purple, modulating proteins = pink.

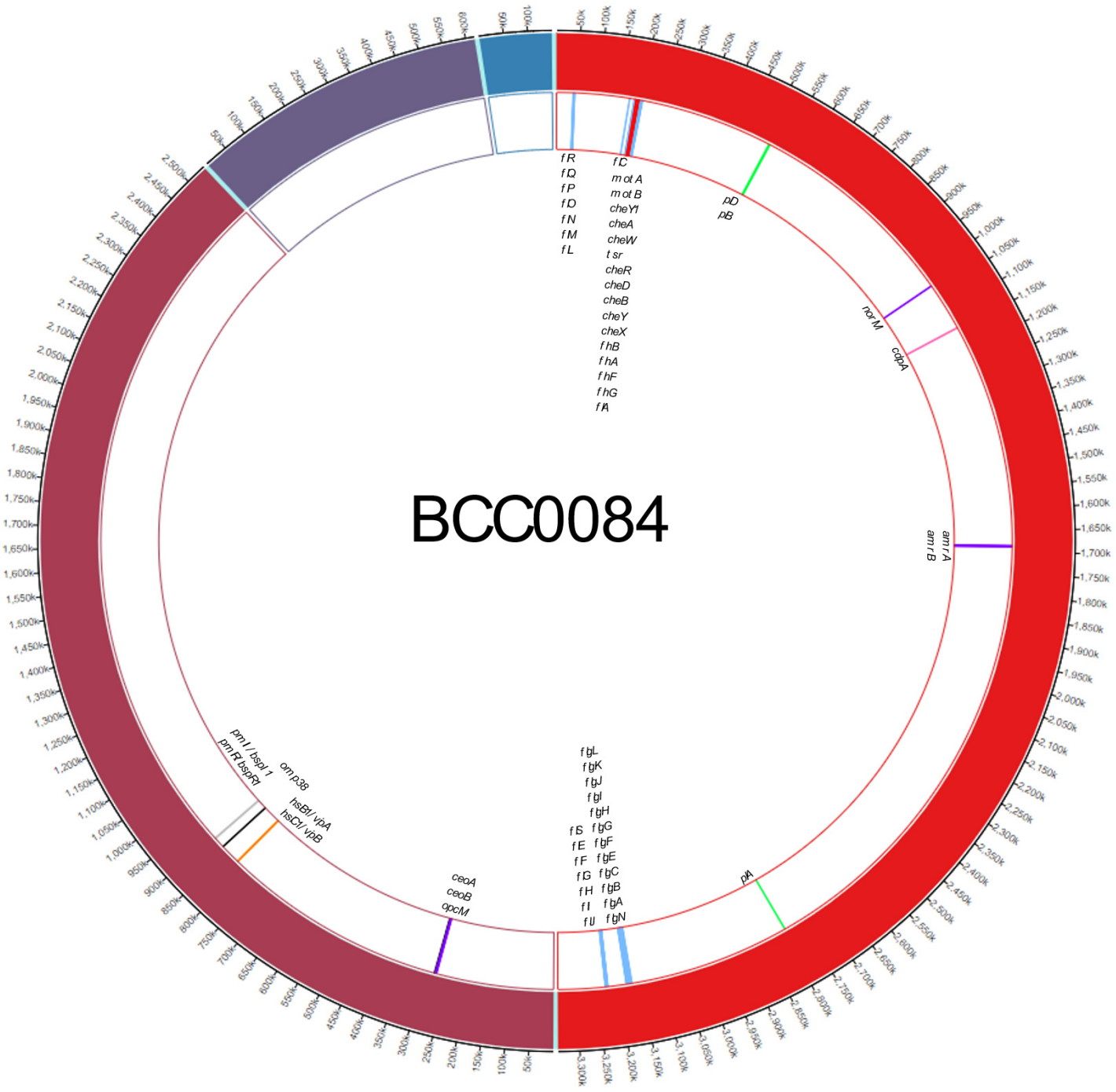


Figure 20 | Location of virulence and pathogenicity genes in the *B. multivorans* strain ATCC 17616. Key genes were identified using Abricate databases against the complete genome. Figure was drawn using Circa and edited using Inkscape. The gene locations are colour coded by group. Flagella = light blue, chemotaxis genes = red, T4 pili = lime green, T6SS genes = orange, QS genes = grey, OMP = black, AMR genes = purple, modulating proteins = pink.

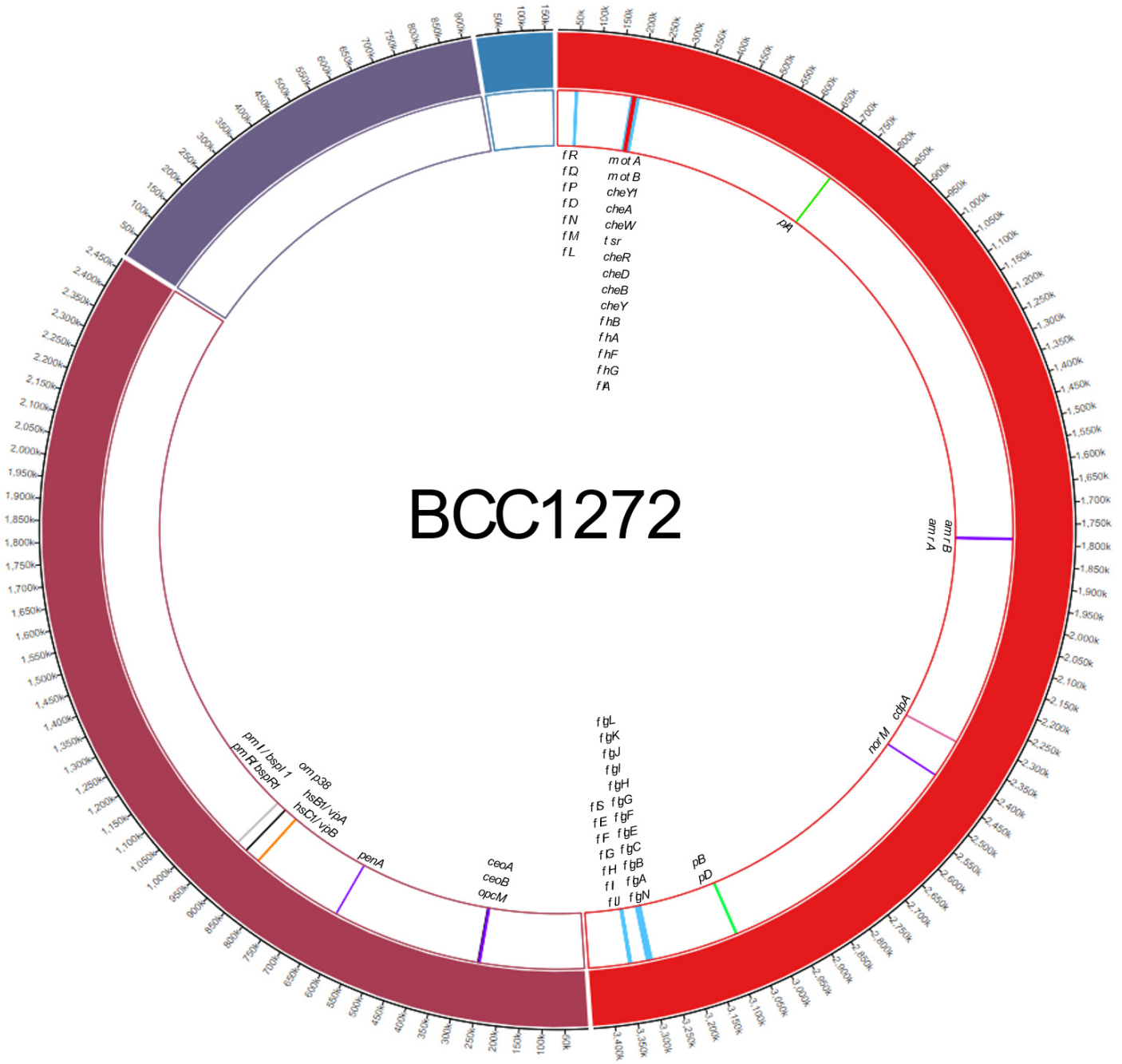


Figure 21 | Location of virulence and pathogenicity genes in the *B. multivorans* strain BCC1272. Key genes were identified using Abricate databases against the complete genome. Figure was drawn using Circa and edited using Inkscape. The gene locations are colour coded by group. Flagella = light blue, chemotaxis genes = red, T4 pili = lime green, T6SS genes = orange, QS genes = grey, OMP = black, AMR genes = purple, modulating proteins = pink.

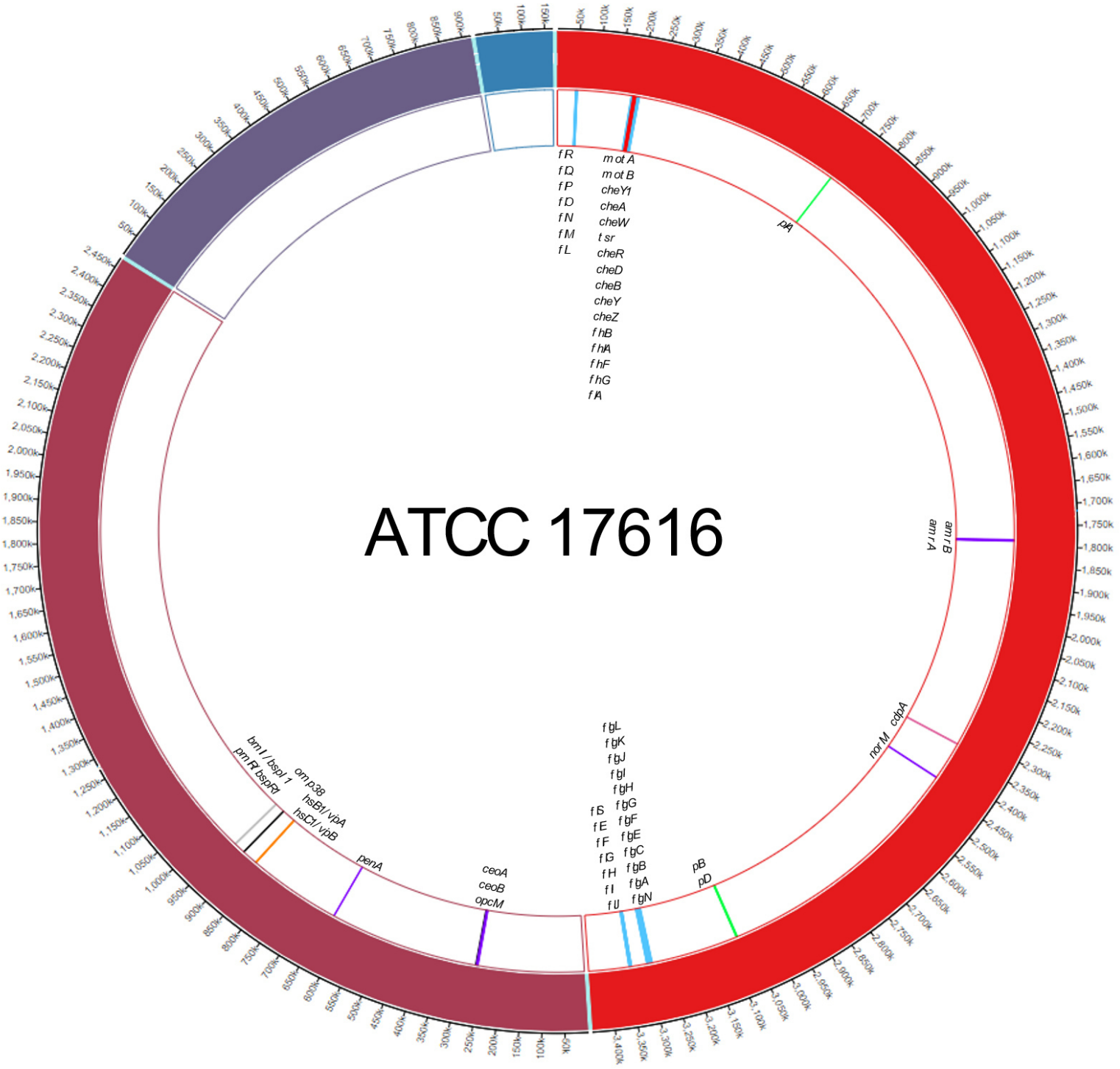


Figure 22 | Location of virulence and pathogenicity genes in the *B. multivorans* strain ATCC 17616. Key genes were identified using Abricate databases against the complete genome. Figure was drawn using Circa and edited using Inkscape. The gene locations are colour coded by group. Flagella = light blue, chemotaxis genes = red, T4 pili = lime green, T6SS genes = orange, QS genes = grey, OMP = black, AMR genes = purple, modulating proteins = pink.

1.1.1.4 Examining the low oxygen locus genes using Abricate

A custom Abricate database was used to identify homologs of the *B. cenocepacia* low oxygen locus genes (Sass *et al.* 2013) in 49 *B. multivorans* strains. This used the low oxygen locus (*lxa*) and associated genes extracted from the *B. cenocepacia* J2315 genome (Holden *et al.* 2009). The panel of strains used in these analyses were the initial 49 strains sequenced as part of this PhD thesis. R studio was used to compile a gene heatmap with a respective phylogenetic tree to illustrate the gene presence absence of both the *lxa* locus and associated genes. There were no clear correlations observed when comparing lineage and *lxa* genes. However, lineage 2 had a greater number of strains (13 of 30; 43.3%) with ≥40 *lxa* genes (80% or more of the locus) than lineage 1 (1 of 18; 5.6%) (Figure 23).

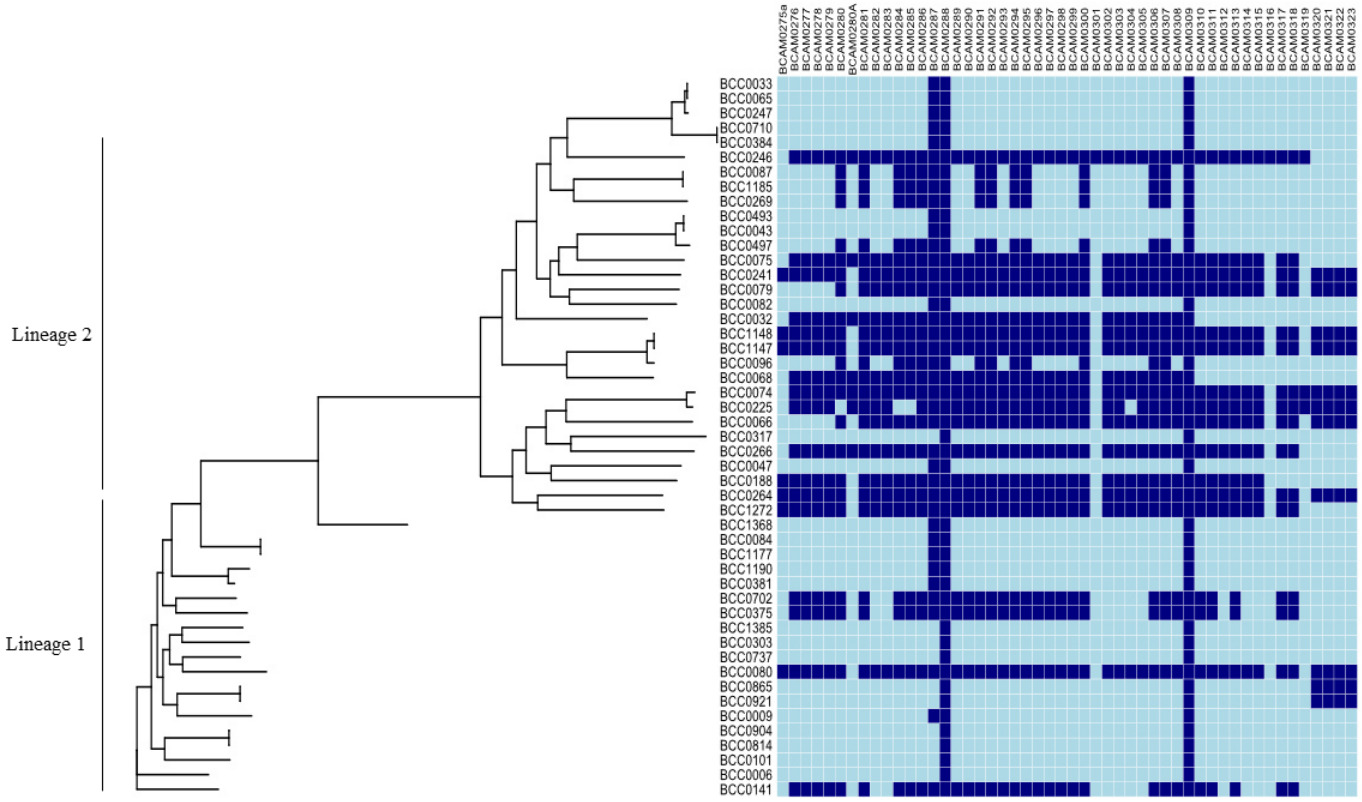


Figure 23 | Gene presence absence matrix for the low oxygen associated (*lxa*) genes in the *B. multivorans* strains (*n* = 49). A custom Abricate database was used to identify the genes, with a percentage coverage and identity cut off at 80%. Navy blue shows gene presence and light blue shows gene absence.

4.2.2 *Identification of prophages in the strain panel*

Screening for phages within bacterial genomes (*n* = 77) provides evolutionary insights into mechanisms bacteria use for fitness in different settings. Phage therapy has also been of interest to treat antibiotic-resistant infections such as those caused by the BCC (Seed and Dennis 2005). For example, research has previously been shown that a mutant phage from *B. cenocepacia* J2315, known as KS4-M, can target other pathogenic *B. cenocepacia* and *B. multivorans* (Lauman and Dennis 2021).

The search tool PHASTER (Arndt *et al.* 2016) was used to provide insights on prophages within the *B. multivorans* strain panel genomes. All 77 *B. multivorans* strains examined harboured prophage material. The number of potential prophage regions within each genome ranged from 1 to 8 (Figure 24). Eighteen of the *B. multivorans* genomes harboured the minimum of 1 prophage whilst two strains (AU22892 and CF2) both had 8 prophages. Overall, 92 intact, 26 questionable, and 124 incomplete prophages were found amongst the 77 genomes (Figure 24).

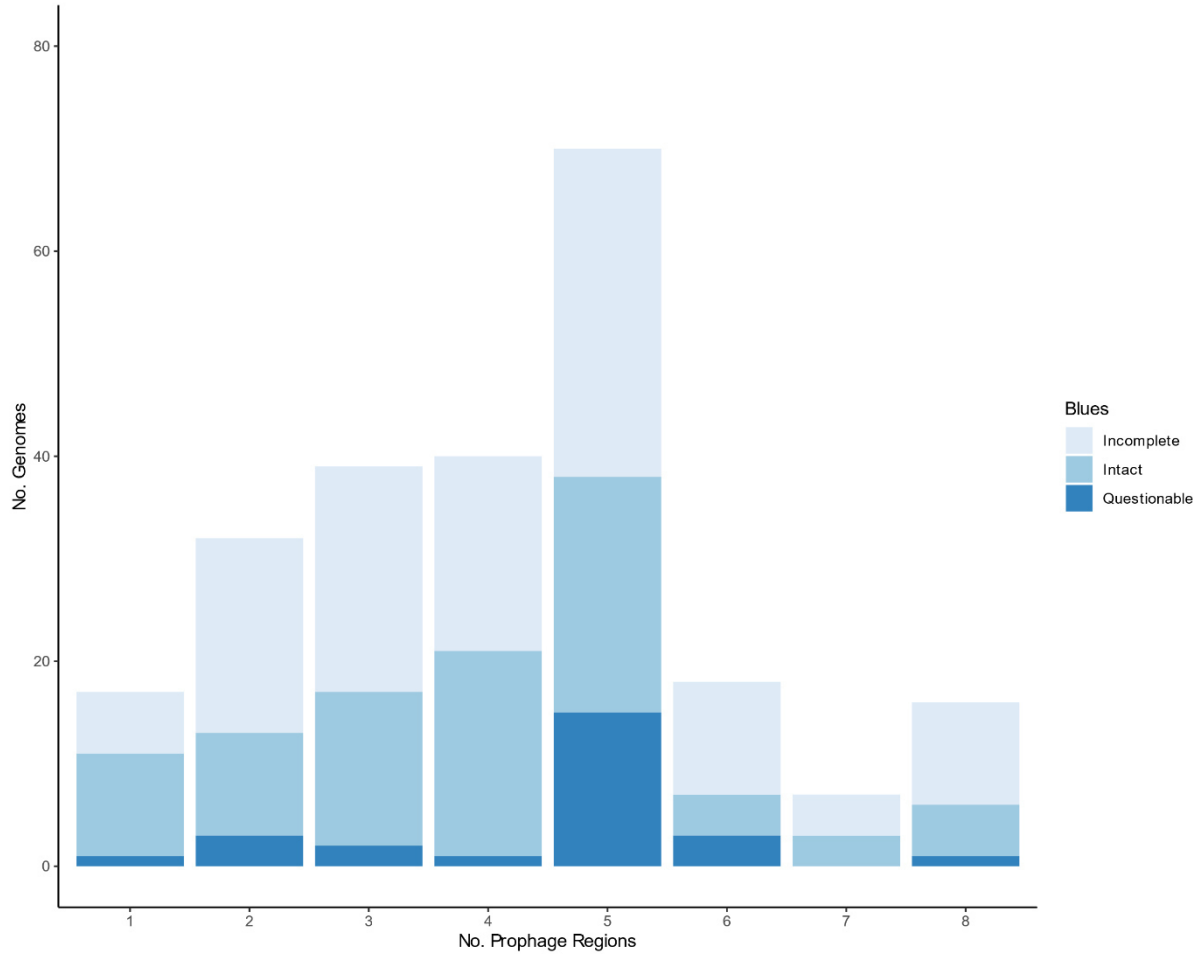


Figure 24 | Distribution of prophages within the *B. multivorans* genomes (*n* =77). Prophages were split into the number of prophage regions per *B. multivorans* genome for the analysis undertaken with PHASTER (Arndt *et al.* 2016). This showed the number of incomplete (*n* = 124), intact (*n* = 92) and questionable (*n* = 26) prophages in the 77 *B. multivorans* genomes examined.

The total sizes of all intact prophage predicted regions found in each *B. multivorans* genome ranged from 15.8 kb to 60.8 kb (mean = 26.1 kb) Incomplete prophage predicted regions ranged from 5.1 kb to 45.8 kb (mean = 26.3 kb). Questionable prophage predicted regions ranged from 8.5 kb to 57.3 kb (mean = 26.9 kb) (Figure 25). The total amount of prophage material in the 77 *B. multivorans* genomes ranged from 9.5 kb (0.15% of total genome content) to 218.2 kb (3.21% of total genome content).

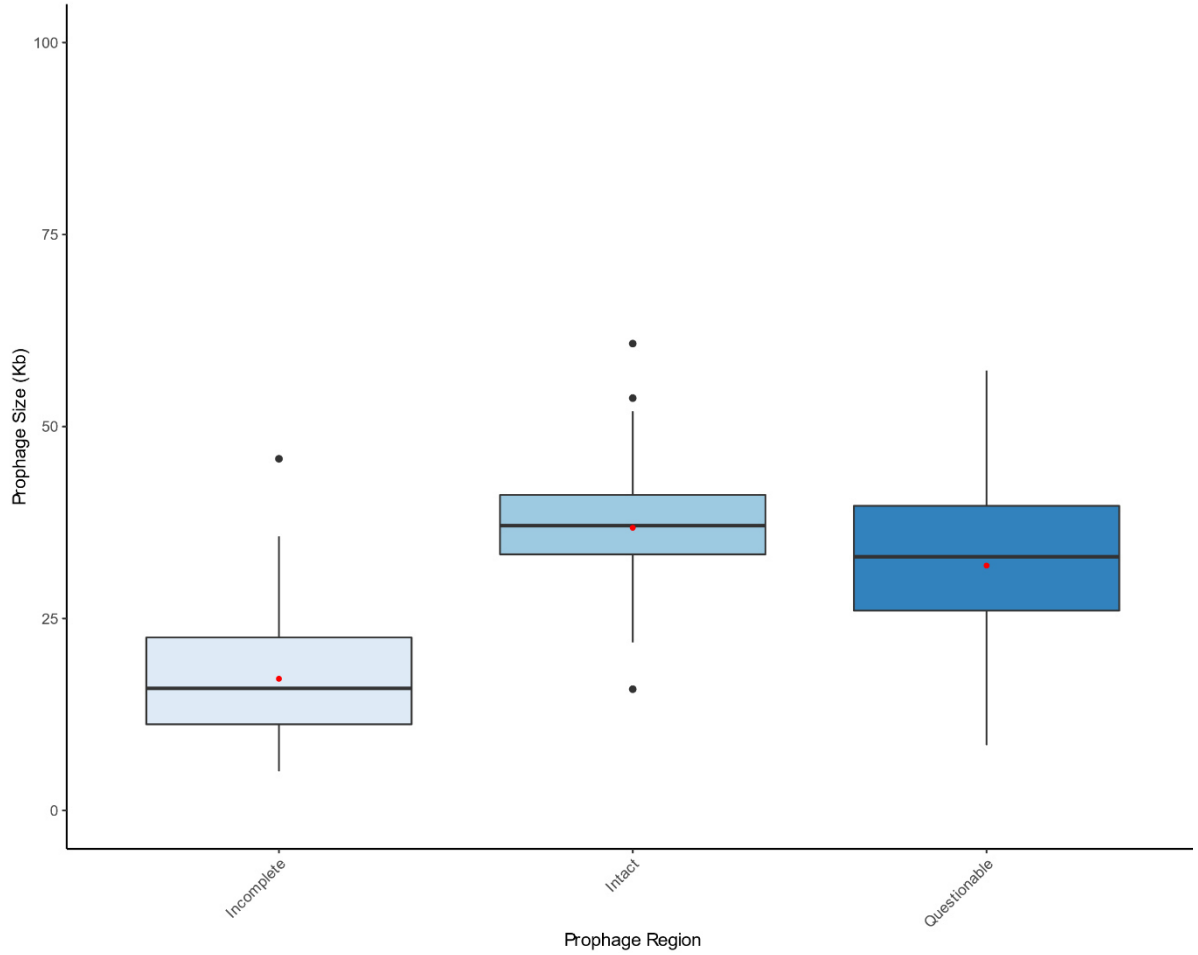


Figure 25 | Types of prophages compared to prophage size in the *B. multivorans* strain panel (*n* = 77). Box plots illustrate the mean, median, upper, and lower quartiles as well as any outliers. Plots have been split into incomplete (*n* = 124), intact (*n* = 92), and questionable (*n* = 26) regions found in 77 *B. multivorans* strains.

Using a linear regression model showed that, on average, *B. multivorans* with a larger genome size (Mb) were observed to harbour a greater number of prophages on average (Figure 26). Although there was considerable variation observed in prophage number, the *p*-value for the regression was 0.053. The low R^2 value can be explained by the residual points which are far from the mean (Figure 26). Firm conclusions cannot be made about the number of prophages and *B. multivorans* genome size (Mb) due to the low number of points (~3%) that are assessed within the regression.

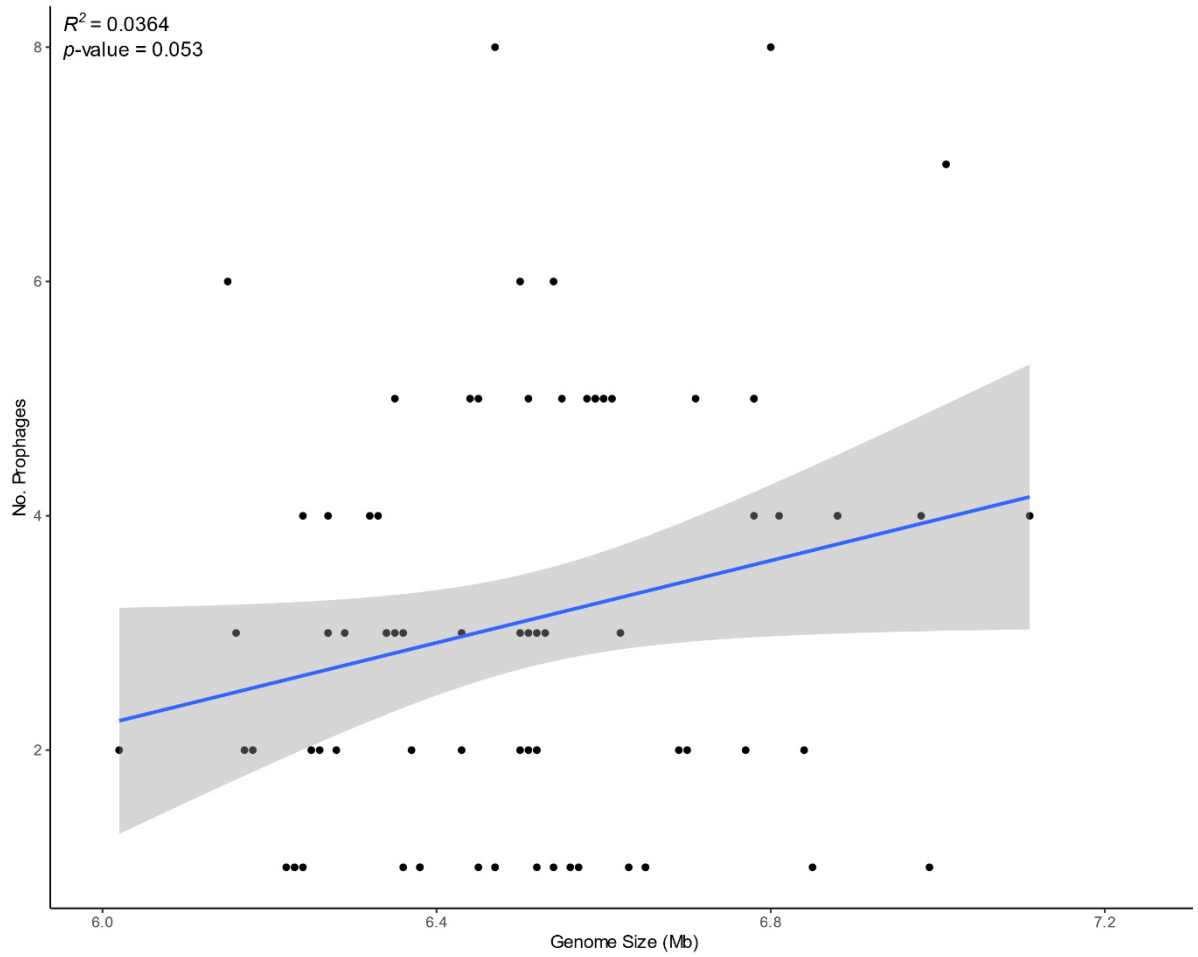


Figure 26 | Correlation between the number of phages and genome size (Mb) of the *B. multivorans* strain panel (*n* = 77). A linear regression line has been drawn for the prophage number compared against genome size, showing a positive distribution. This has estimated the relationship between *B. multivorans* genome size and the total number of putative prophages (irrespective of whether they are incomplete, intact, or questionable) identified in the genome. The blue line represents the regression line of best fit.

4.2.3 *Identification of insertion sequences*

Insertion sequence (IS) elements are important to bacterial genome evolution (Mahillon and Chandler 1998). IS genome rearrangements in the BCC are implicated in their environmental adaption capabilities (Lessie *et al.* 1996). This Section investigated ISs within the *B. multivorans* strain panel (*n* = 77), using ISEscan with default parameters (Xie and Tang 2017). The scaffolded genomes were used (Section 3.2.9.1) for the analysis. In total, there were 592 complete and 760 partial IS elements found in the *B. multivorans* strains genomes (scaffolded against the complete genomes). Replicon C1 had the greatest amount of IS elements (739 IS), followed by C2 (384 IS). Replicon C3 had less IS elements than C1 and C2, harbouring 174 overall. C4 had the lowest amount of 59 IS elements. Plasmid analysis (see Section 4.2.4) demonstrated that the model *B. multivorans* strain genomes also encoded a fourth replicon that was designated C4; this plasmid encoded the fewest number of predicted IS elements (Table 21). The same IS family was not necessarily found on all replicons.

There were 21 IS families identified overall in the scaffolded *B. multivorans* strain panel (*n* =77). Of this, 5 IS families were 'new' (Table 21). The term 'new' is used when an IS cluster cannot be classified against those currently in the IS database (Xie and Tang 2017). The greatest number of hits was against IS5, with 335 in total from all 77 *B. multivorans* strains. The ISs for these were found on all four genomic replicons, with hits for both complete and partial IS element types. Over half (52.8%) of the IS elements were found on C1, followed by 29.6% on C2, 15.2% on C3 and 2.4% on C4 (Table 21). IS3 had the largest amount of complete IS elements overall, with 180 hits identified. In terms of partial IS elements identified, IS5 had 243 as the largest identification of partial elements within the dataset.

Comparing the lineages, IS families new_2 and new_269 were only found in lineage 1. Unique to lineage 2a were IS families new_5, new_298, ISAS1, and IS30. IS1595 and new_269 were only observed in lineage 2b (Table 21). In all, the relative number of lineage hits for IS elements was similar, with 487, 463, and 388 hits for lineage 1, 2a and 2b, respectively.

Table 21 | Insertion Sequence (IS) elements identified in the *B. multivorans* strain panel.

IS Family	Total Hits	Lineage Hits				Length of IS elements (bp)	C1		C2		C3		C4		Total	
		1	2a	2b	Other		complete	partial	complete	partial	complete	partial	complete	partial	complete	partial
IS110	101	34	50	16	1	180-4466	30	20	17	9	11	0	11	3	69	32
IS1182	2	1	1	0	0	1406-1570	1	0	0	0	1	0	0	0	2	0
IS1595	2	0	0	2	0	828-834	0	0	0	2	0	0	0	0	0	2
IS21	231	80	77	70	3	294-4542	38	105	18	29	14	5	14	8	84	147
IS256	150	62	33	54	1	135-2473	18	53	18	61	0	0	0	0	36	114
IS3	291	148	74	66	3	111-2104	105	64	42	33	33	11	0	3	180	111
IS30	1	0	1	0	0	174	0	0	0	0	0	1	0	0	0	1
IS4	45	23	7	15	0	175-1473	24	18	2	1	0	0	0	0	26	19
IS481	14	2	7	5	0	281-3174	3	4	6	1	0	0	0	0	9	5
IS5	335	103	133	95	4	147-2027	22	155	22	77	44	7	3	5	91	244
IS6	7	1	6	0	0	353-1201	0	0	1	1	4	0	1	0	6	1
IS630	47	1	23	23	0	279-1978	0	43	1	0	3	0	0	0	4	43
IS66	37	6	17	14	0	281-2353	8	7	5	10	1	1	2	3	16	21
IS91	3	1	2	0	0	3204	0	0	0	0	0	2	1	0	1	2
ISAS1	2	0	2	0	0	610-1124	0	2	0	0	0	0	0	0	0	2
ISL3	56	22	15	18	1	362-2866	15	1	26	0	3	11	0	0	44	12
new_4	18	1	9	8	0	3452-6455	2	0	0	0	16	0	0	0	18	0
new_298	1	0	1	0	0	909	1	0	0	0	0	0	0	0	1	0
new_2	2	2	0	0	0	3941-3942	0	0	1	0	0	0	1	0	2	0
new_6	5	0	5	0	0	1381-2648	0	0	1	0	0	4	0	0	1	4
new_269	2	0	0	2	0	3298	0	0	0	0	2	0	0	0	2	0

4.2.4 Identification of plasmids

4.2.4.1 Statistics and initial analysis of the distinct C4 regions

The results of the assembly tool Trycycler (Wick *et al.* 2021) provided a polished consensus sequence (Section 3.2.8.1) for each complete genome (n = 4), with BCC0084 (lineage 1) and BCC1272 (lineage 2a) having 4 reconciled replicons and BCC0033 (lineage 2b) being split into 7 replicons (Section 3.2.8.1). The complete genome of *B. multivorans* ATCC 17616 has a 167422 bp (61.3% GC content) replicon previously characterised as plasmid pBMUL01 (Sant'Anna *et al.* 2009; Yagi *et al.* 2009). As the PacBio completed genomes revealed a similar replicon structure, further investigation was done to see whether a homologous plasmid could be found within the 4 model *B. multivorans* genomes.

RFPlasmid (van der Graaf-van Bloois *et al.* 2021) was initially used to predict the replicon type within the *B. multivorans* model strains (n = 4). RFPlasmid is a tool which predicts whether a bacterial contig is of chromosomal or plasmid origin. The output identified the C4 replicon for all 4 genomes to be of plasmid origin, whilst replicons C1-3 were all predicted as chromosomal by this tool. The C4 replicon comprised between 1.0% to 2.4% of the 4 *B. multivorans* genomes (Table 22). BCC0033 (lineage 2b) had a smaller C4 contig length (67 kb) than the other model strains, that comprised 0.01% of the total genome in that strain (Table 22). BCC0033 also had the least percentage of predicted plasmid genes on its replicon C4 (2.4%) compared to the other genomes. ATCC 17616 and BCC1272 (both lineage 2a) had 90.0% and BCC0084 (lineage 1) had 65.6% plasmid genes on replicon C4 (Table 22). This differences in BCC0033 could have been due to the contamination or any misassembly of the BCC0033 genome (see Section 3.2.8.1), meaning plasmid genes were potentially spread across the 3 discarded consensus regions. However, the RFPlasmid "votes" scale for a plasmid replicon in relation to BCC0033 C4 are high, at 0.8 out of a total score of 1 (Table 22).

Table 22 | RFPlasmid predictions for the *B. multivorans* complete genomes (*n* = 4)

Genome	Replicon	Prediction	Votes Chromosomal[c]	Votes Plasmid[c]	Contig Length (bp)	Percentage of genome (%)	Plasmid Genes (%)
	C1	c[a]	0.9824	0.0176	3448421	49.2	6.3
ATCC 17616	C2	c	0.9788	0.0212	2473162	35.3	11.2
(lineage 2a)	C3	c	0.8922	0.1078	919805	13.1	21.2
	C4	p[b]	0.0306	0.9694	167422	2.4	90.3
	C1	c	0.9824	0.0176	3448466	49.2	6.3
BCC1272	C2	c	0.9788	0.0212	2473135	35.3	11.2
(lineage 2a)	C3	c	0.8922	0.1078	919806	13.1	21.3
	C4	p	0.0308	0.9692	167422	2.4	90.4
	C1	c	0.9834	0.0166	3345195	50.3	5.8
BCC0084	C2	c	0.9794	0.0206	2543087	38.2	10.6
(lineage 1)	C3	c	0.976	0.024	618272	9.3	10.4
	C4	p	0.0342	0.9658	149841	2.3	65.6
	C1	c	0.981	0.019	3544234	52.6	5.5
BCC0033	C2	c	0.981	0.019	2391009	35.5	11.0
(lineage 2b)	C3	c	0.9798	0.0202	733512	10.9	16.6
	C4	p	0.207	0.793	67852	1.0	2.4

[a]c = chromosomal prediction, [b]p = plasmid prediction, [c]Votes for chromosomal or plasmid are on a scale of 0-1.

4.2.4.2 Plasmid analysis within the strain panel

To expand the analysis, the un-scaffolded contigs of the *B. multivorans* strain panel (*n* = 77) were used to predict chromosomal and plasmid hits via RFPlasmid (van der Graaf-van Bloois *et al.* 2021) within a representative group of species genomes. The analysis showed that at least one plasmid contig from 75 of the 77 (97%) of the strains in the panel was present. Plasmid contig hits were missing from strains AU1185 and DWS_421-*B.* These are non-CF and environmental isolates, respectively. Therefore, it can be noted that all CF-associated *B. multivorans* strains tested harbour plasmid content in the genome.

BLAST was then used to identify the similarities between the C4 regions of the complete genomes and the scaffolded *B. multivorans* strain panel. A total of 24 C4-like regions were identified in the *B. multivorans* 77 (31% of strains) draft genomes, each with between 80-100% identity matches to the C4 regions in the model strains. *B. multivorans* strain AU28442 genome had the lowest BLAST identity to the model C4 region (in BCC0084; lineage 1) at 80%. BCC1147 and D2214 had the highest similarity at 100% ID to ATCC 17616 (lineage 2a) and BCC0033 (lineage 2b), respectively. The mean percentage BLAST identity between the regions was 96%. The alignment lengths varied amongst the strains, ranging from 90 (strain FDAARGOS 546) to 109412 bp (strain BCC1177).

4.2.5 *COG annotation of the model strains*

4.2.5.1 COG annotation of the whole genome

Categorisation of COGs in the *B. multivorans* model strains (n = 4) were done by using EggNog described in Section 2.14.6. ATCC 17616 (lineage 2a), BCC1272 (lineage 2a), BCC0084 (lineage 1) and BCC0033 (lineage 2b) had 6,559, 6558, 6,256, and 6,285 COG categorised CDS, respectively (Table 23). All genomes also had CDS with at least two COG annotations. BCC0033 scored the highest with 462 duplicate COG categories, followed by BCC0084 at 453, then ATCC 17616 at 440, with BCC1272 having the smallest number of duplicated categories at 435. Each COG category was split into type: 1) information storage and processing, 2) cellular processes and signalling, 3) metabolism, and 4) poorly characterised. Metabolism held the greatest percentage of COGs for each model genome, ranging from 32.8-36.1% of the total COGs. Information storage and processing held the least percentage of COGs across all 4 genomes, between 16.3-17.8%. Around one quarter of the COGs analysed (23.1-24.7%) were poorly categorised. This included the 'S' group where functions were predicted as unknown. When looking at the characterised COGs, there were no hits for RNA processing and modification, nuclear structure, or cytoskeleton. The top 3 categories in all genomes were transcription (K), inorganic ion transport and metabolism (P), and amino acid transport and metabolism (E) (Table 23).

4.2.5.2 COG annotation of the genomic replicons

The COGs within the complete *B. multivorans* model strain genomes (n = 4) were then divided into the 4 replicons. As expected, the number of COGs increased as the size of the replicon increased. This meant that C1 held the largest number of COGs, followed by C2, then C3, and finally C4 in all cases. ATCC 17616 and BCC1272 (both lineage 2a) had 3,311 COGs in C1, BCC0084 (lineage 1) had 3,215, and BCC0033 (lineage 2b) had 3,370. The number of COGs in C2 decreased by roughly 1,000. This was 2,280 for ATCC 17616 and BCC1272, followed by 2,363 for BCC0084 and 2,198 for BCC0033. The COGs found in C3 were much lower, between 558 and 850 depending on the model strain analysed. C4 also harboured predicted COGs. Interestingly, BCC0084 had around double the COGs in C4 (totalling 123) than the other model strains. ATCC 17616 and BCC1272 had 48 COGs in C4, and BCC0033 had 46.

There were similarities between replicon C1 COG distribution and the complete genomes as follows. The top 3 characterised categories remained at K (transcription), P (ion transport and metabolism), and E (amino acid transport and metabolism) for all strains (Table 23). This was also true for C2. C1 and C2 also harboured most of the CDS of unknown function (category S), relative to their replicon sizes. The percentage of unknown CDS in C1 and C2 combined was between 82-88% of the total unknown CDS found in any one genome.

When comparing COG distributions on replicons C3 and C4, there was more variability between the strains. The total number of unknown CDS was 9-15% for replicon C3 and 1-3% for replicon C4 (Table 23). The top three categorised COG categories for replicon C3 were K, L, and C for *B. multivorans* strains BCC1272 and ATCC 17616 (lineage 2a). COG categories K, C, and E were the top 3 for replicon C3 of *B. multivorans* strains BCC0033 (lineage 2b) and BCC0084 (lineage 1) (Table 23). Therefore, energy production and conversion (L) was favoured in *B. multivorans* strains ATCC 17616 and BCC1272. However, CDS associated with amino acid transport and metabolism were more common in BCC0033 and BCC0084 (Table 23).

Replicon C4 did not harbour any CDS categorised into information and storage (J) or extracellular structures (W) COGs. BCC0084 (lineage 1) and BCC0033 (lineage 2b) have a single CDS hit for the defence mechanism (V) COG category. A single hit for secondary metabolite biosynthesis, transport, and catabolism (Q) was observed in the C4 replicon of BCC0084 and BCC001272 (lineage 2a). *B. multivorans* isogenic strains BCC1272 and ATCC 17616 (lineage 2a) had a single hit for coenzyme transport and metabolism (H) (Table 23). Overall, the COG category hits on the C4 regions of the 4 *B. multivorans* model strains were variable.

Table 23 | The COG categories of the *B. multivorans* model strains (*n* = 4). COG categories were grouped based on the CDS in the whole genome and each replicon of *B. multivorans* strains BCC0033, BCC0084, BCC1272 and ATCC 17616. The CDS were assigned COGs categories using the online version of EggNOG mapper.

Type	COG	Description	ATCC 17616 (lineage 2a)					BCC1272 (lineage 2a)					BCC0033 (lineage 2b)					BCC0084 (lineage 1)				
			Whole Genome	C1	C2	C3	C4	Whole Genome	C1	C2	C3	C4	Whole Genome	C1	C2	C3	C4	Whole Genome	C1	C2	C3	C4
Information Storage and Processing	A	RNA processing and modification	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	B	Chromatin structure and dynamics	3	3	0	0	0	3	3	0	0	0	3	3	0	0	0	3	3	0	0	0
	J	Translation, ribosomal structure, and biogenesis	205	176	23	6	0	205	176	23	6	0	202	177	21	4	0	200	175	20	4	0
	K	Transcription	623	255	275	86	3	623	255	275	86	8	629	266	276	84	1	586	237	278	65	6
	L	Replication, recombination, and repair	334	164	81	67	7	334	164	81	67	20	191	146	29	12	5	243	155	52	19	17
Cellular Processes and signalling	D	Cell cycle control, cell division, chromosome partitioning	57	41	12	2	0	57	41	12	2	2	52	41	10	1	0	56	40	13	1	2
	O	Molecular chaperones and related functions	161	106	39	12	0	161	106	39	12	3	147	98	40	9	0	154	98	42	12	2
	M	Cell wall, membrane, and envelope biogenesis	388	211	120	48	5	387	211	120	48	9	370	213	123	32	2	379	210	132	29	8
	N	Cell motility	115	70	35	7	0	115	70	35	7	2	110	70	30	6	4	122	75	35	6	6
	P	Inorganic ion transport and metabolism	472	225	182	56	4	472	225	182	56	8	469	238	183	48	0	454	218	185	42	8
	T	Signal transduction and mechanisms	213	80	99	32	3	213	80	99	32	4	192	78	96	18	0	188	76	91	18	4
	U	Intracellular trafficking, secretion, and vesicular transport	149	92	30	17	2	149	92	30	17	8	132	86	29	4	12	142	85	32	5	20
	V	Defence mechanisms	61	29	17	15	0	61	29	17	15	0	60	37	17	6	1	61	33	21	6	1

	W	Extracellular structures	6	3	3	0	0	6	3	3	0	0	7	3	3	0	0	5	3	2	0	0
	Y	Nuclear structure	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Z	Cytoskeleton	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Metabolism	C	Energy production and conversion	468	209	170	87	0	468	209	170	87	4	467	212	178	77	0	436	203	179	52	2
	G	Carbohydrate transport and metabolism	349	167	146	36	0	349	167	146	36	0	370	173	154	43	0	355	164	150	41	0
	E	Amino acid transport and metabolism	551	264	236	50	2	552	264	236	50	3	580	277	241	62	0	551	262	241	47	2
	F	Nucleotide transport and metabolism	137	104	22	11	0	137	104	22	11	0	132	105	20	7	0	132	104	20	8	0
	H	Coenzyme transport and metabolism	239	163	50	25	1	239	163	50	25	1	234	165	52	17	0	224	159	49	16	0
	I	Lipid transport and metabolism	233	116	87	28	1	233	116	87	28	2	274	115	104	54	1	244	110	105	28	1
	Q	Secondary metabolites biosynthesis, transport, and catabolism	178	79	78	20	0	178	79	78	20	1	214	91	95	28	0	191	73	99	18	1
Poorly characterised	R	General function predictions	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	S	Unknown function	1150	562	408	156	8	1149	561	408	156	22	1094	588	380	122	7	1163	569	465	106	23
	-	Protein dissimilar to any COG	467	192	167	89	12	467	193	167	89	25	356	188	117	37	13	367	163	152	35	20

4.2.6 *Variant analysis of the B. multivorans lineages*

4.2.6.1 Core SNPs

To further investigate the evolution of the two-lineages within the *B. multivorans* strain panel (n = 77), SNP analysis was performed using Snippy (Section 2.14.7). This looked for insertions, deletions within the two lineages. Initially, the core SNPs in the 77-genome strain panel were examined. There were 373,694 core SNPs found amongst the panel, splitting into 229,596 in replicon C1 (61.4%), 128,969 (34.5%) in C2, and 15,129 (4.0%) in C3. This correlates to the results observed in Section 3.2.7.1, showing the distribution of core genes within the *B. multivorans* replicons.

To perform lineage variant analysis within the lineages, the complete genomes created during this thesis were used. BCC0084 was the reference for lineage 1 (n = 27) and BCC1272 was the reference for lineage 2 (n =50). A total of 149,249 core SNPs were identified in lineage 1 and 403,241 in lineage 2. The core SNPs situated on each replicon for lineage 1 were as follows: C1 was comprised of 80,352 (53.8%; 24.0 SNPs/Kb) core SNPs, C2 had 60,198 (40.3%; 23.7 SNPs/Kb) and C3 had 8,699 (5.8%; 14.2 SNPs/Kb). In comparison, lineage 2 was comprised of 230,378 (57.1%) core SNPs on C1, 149,420 (37.1%) on C2, and 23,443 (5.8%) on C3. Despite lineage 2 having 2.7x the number of core SNPs compared to lineage 1, the overall distributions of the core SNPs were relatively consistent between the replicons, decreasing based on replicon size. Replicon C1 in both instances harboured over half the core SNPs, whereas C3 only had 5.8% of the core SNPs in both cases. Neither lineage showed core SNPs on replicon C4.

Phylogenies were drawn for the two lineages using the core SNP alignment produced from Snippy with Gubbins recombination event filtering (Section 2.10.2). For lineage 1, the -filter_percentage flag for Gubbins was adjusted to 27% as the default of 25% cut-off caused missing data. Whilst there were differences within the phylogenomic splits of the core SNP lineages (Figure 27) compared to the core gene trees (Figure 10 and Figure 10; Section 3.2.5.2), the overall placement of strains remained consistent in lineage 1. However, lineage 2 genomes had several differences in terms of the phylogenomic placement of the strains and their closest relative sequences (Figure 27). The two sub-lineage splits remained relatively supported in the SNP phylogeny. Two outliers were observed, BCC0264 and NKI379 in the SNP phylogeny (Figure 27). The BCC0264 strain also presented as an outlier in the lineage 2 core gene tree (Figure 10; Section 3.2.5.2). The SNP phylogenies also had much longer and deeper branch lengths than the core gene phylogenies. This indicated a larger genetic variation in the strains when using all the core SNPs available for comparison, instead of the more limited number of core genes.

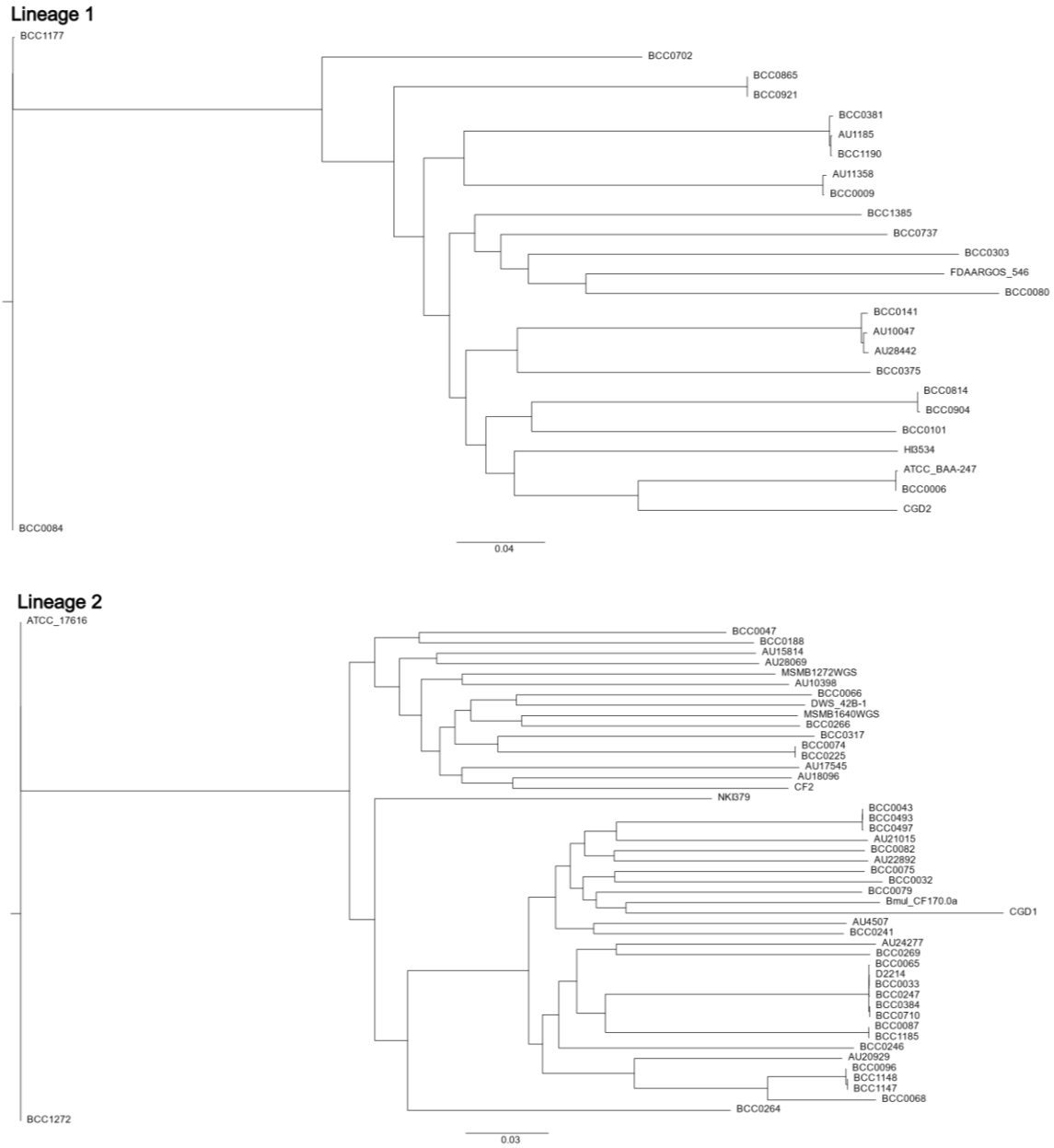


Figure 27 | Gubbins-filtered core SNP phylogenies of the two *B. multivorans* lineages. Top = lineage 1 (*n* =27), bottom = lineage 2 (*n* = 50). Phylogenetic trees drawn using RAxML with 100 bootstraps. Lineage 1 used 149,249 core SNPs and lineage 2 used 403,241 core SNPs (before filtering) for their creation. Scale bar represents the evolutionary distances of the phylogenies per SNP substitution.

4.2.6.2 Variants within the lineages

Variant differences were also analysed within the two *B. multivorans* lineages (lineage 1: *n* =27; lineage 2: *n* = 50). The approach used was to eliminate the core SNPs from the analysis and focus only on the differences between the reference strains (*n* = 4) (BCC0084, lineage 1; or BCC1272, lineage 2a) and the other strains in the panel (*n* =73). In total, there were an average of 41,785.9 differences in lineage 1 and 75,549.5 variants in lineage 2 (Figure 28); this difference was significant (p = <0.0001; Figure 28). The number of variants also varied within the lineages. Lineage 1 ranged from 125-51,234 and

lineage 2 from 5-95,010 differences between the reference and the comparative strain. The most distantly related strains compared to the references were BCC0080 and BCC0266 when compared to the lineages 1 and 2 references, respectively (Figure 28). The strain with the 5-variant difference from BCC1272 was ATCC 17616, shown to be an isogenic strain (Section 3.2.5). This was comprised of 3 variants on replicon C1 (1 SNP, 1 insertion, and 1 deletion) and 2 on replicon C2 (1 SNP and 1 deletion). Therefore, if this was to be removed from the dataset, the next smallest variant total would be 64,790 in lineage 2. This emphasises the argument that lineage 1 is much more closely related than lineage 2.

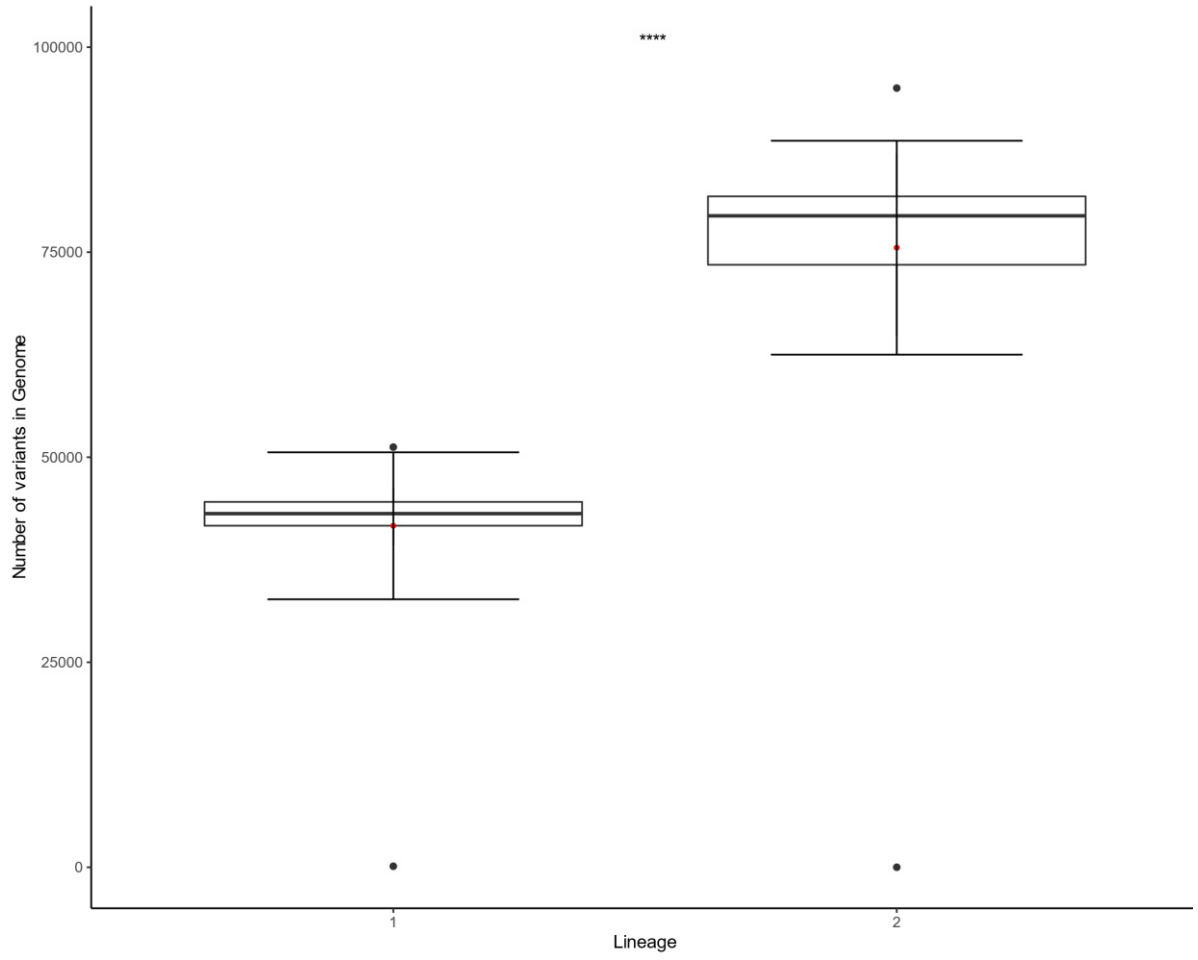


Figure 28 | Number of SNP variants in the *B. multivorans* genomes (*n* = 76) grouped by lineage. Total variants were calculated based on the results output from Snippy. Variants were grouped by lineage (lineage 1 = 27 strains, lineage 2 = 49 strains). Statistical significance compared the number of SNP variants between the two lineages ($p = 0.0001$).

The distributions of variant type are relatively similar between lineage 1 and lineage 2 (Figure 29). SNPs were the most common, between 85-85% of the total number of variants. This was followed by complex SNPs, which made up 12-13%. Deletions formed 0.8-0.9% and insertions 0.6-0.7% of the total variants. Finally, MNPs were rarely seen, comprising only 0.001% of the total variants in both lineages (Figure 29).

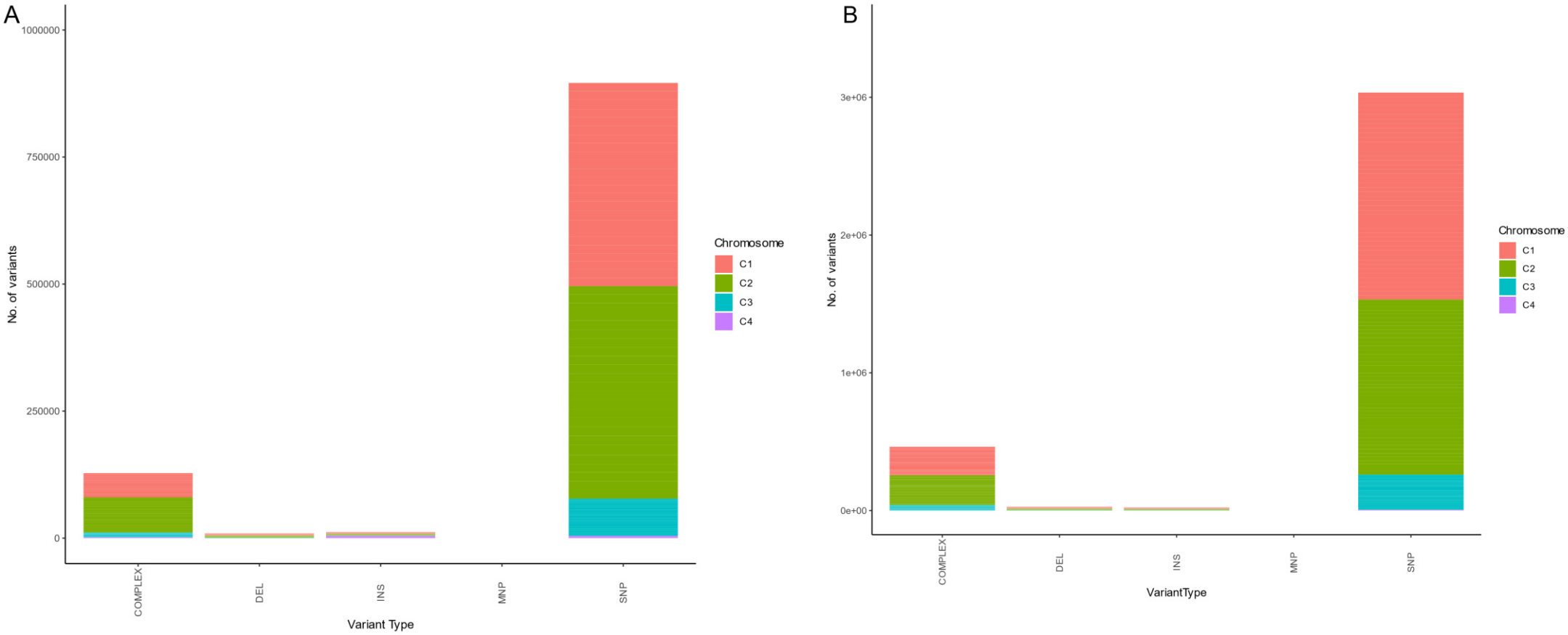


Figure 29 | Number of variants in each *B. multivorans* lineage grouped by variant type. (A) Number of variants in lineage 1 (n = 27), (B) number of variants in lineage 2 (n = 49). Variant type grouped into SNP, complex SNP, deletion, insertion and MNP.

When looking at the replicons individually, most variants were found on C1 and C2 for both lineages (n = 77) (Figure 29). In lineage 1 (n = 27), 43.8% of the variants were found on C1 and 47.5% of variants on C2. In lineage 2 (n =50), this was flipped as most variants were seen on replicon C1 (48.9%), with C2 possessing 42.3%. The percentage of variants on replicons C3 and C4 were much lower for both lineages. C3 harboured 8.1% and 8.4% for lineages 1 and 2 respectively. C4 was even lower, holding 0.6% and 0.2% of lineage 1 and 2 variants respectively.

Statistical analysis was performed to identify the significant differences between number of variants on each replicon. An overall significance of $p = < 2^{-16}$ was recorded for both lineages regarding the total number of variants observed on all four genomic replicons. The only comparison which had no significant difference was between replicons C1 and C2 in lineage 1. The same comparison in the lineage 2 strains had a p-adjusted value of 0.02. All statistical comparisons of variants on the replicons can be found in Table 24.

Table 24 | Statistical comparisons of number of variants found on the *B. multivorans* replicons, grouped by lineage.

Lineage	Replicon Comparison	*p*-adjusted	Significance
1	C1 and C2	0.6093	ns
2		0.0200	*
1	C1 and C3	0.0010	**
2		<0.00001	****
1	C1 and C4	<0.00001	****
2		<0.00001	****
1	C2 and C3	0.0001	***
2		<0.00001	****
1	C2 and C4	<0.00001	****
2		<0.00001	****
1	C3 and C4	<0.00001	****
2		<0.00001	****

Lineage 1: n = 27 and lineage 2: n = 49.

4.2.7 Identification of lineage-specific genes

4.2.7.1 Scoary to identify unique genes

A pan-GWAS approach was used to look for genes unique to each genomic lineage. The *B. multivorans* strain panel was split into its appropriate two-lineages (lineage 1: n = 27 genomes; lineage 2: n = 50 genomes), with the aim of identifying genes in the group based on a presence-absence criteria using Scoary (Section 2.11.1). The output was manually filtered for presence of genes in one lineage and complete absence in the other. The final number of unique genes were 590 in lineage 1 and 745 in lineage 2. After identification, genes of interest were extracted from the pan genome file so that a BLAST search could be performed on the 'unique' genes against the full 283 master *B. multivorans*

panel. This helped to identify any hits which were not previously noted, as well as against the 'other' lineage harbouring the single strain BCC1368 (see Section 3.2.5.2; Chapter 3). The final number of unique genes identified for lineage 1 was 405, and for lineage 2 was 435. These encompassed both genes with annotated predicted functions and hypothetical genes.

4.2.7.2 Shell script creation to sort the data

The unique lineage genes identified in Section 4.2.7 were subjected to sorting using a computational approach. A bash shell script (available: Github) was used to sort the BLAST data and find the number of hits for each gene compared to the number of genomes in that lineage. The genes selected for further analysis in this work were those with ≥99% presence in the corresponding lineage *B. multivorans* strains. This stringent filtering resulted in having 14 unique genes with 100% presence in all lineage 1 genomes ($n = 58$). For lineage 2 ($n = 221$), one gene with a 99.1% presence for the group was identified. The three genomes which were missing the lineage 2 gene of interest were CGD1, DSOPR57 and DSOPR54. As these are CGD and environmental isolates, the gene was deemed to be suitable for use overall, not affecting CF isolates. To further filter the genes for suitability, any hypothetical or putative annotations were discarded, meaning only genes with annotated functions were processed. As lineage 2 only possessed one gene for use, despite being of putative predicted function, it was kept for further analysis. Overall, lineage 1 had 3 suitable genes and lineage 2 had 1 suitable gene for analysis (Table 25).

Table 25 | Lineage-specific genes chosen for further investigation. Table shows gene name, annotation, and percentage of hits in the strain panel genomes.

Lineage	Gene	Annotation	Genome Hits (%)
1	yiaJ_1	DNA-binding transcriptional repressor YiaJ	100.0
	ghrB_1	Glyoxylate/hydroxypyruvate reductase B	100.0
	naiP_3	Putative niacin/nicotinamide transporter NaiP	100.0
2	glnM_2	putative glutamine ABC transporter permease protein GlnM	99.1

4.2.7.3 BLAST and Artemis to check genomic regions

To check the genomic regions of the unique genes of interest, BLAST was used to pinpoint the replicon and location of the gene in the complete genomes for each respective genomic lineage. In lineage 1 ($n = 1$), all three genes (*yiaJ_1, ghrB_1*, and *naiP_3*) were located on replicon 2 in *B. multivorans* BCC0084. In the lineage 2 genomes ($n = 3$) (BCC0033, BCC1272 and ATCC 1761), *glnM_2* was identified on replicon 1. Artemis was used to confirm the genomic context of the genes. This was used to evaluate the adjacent genes and check the overall GC content of the region. YiaJ_1 was found sandwiched between hypothetical/putative genes and an RNA modulator (gene *rraA*). GhrB_1 was encoded between two hypothetical genes, near to nicotinate phosphoribosyl transferase *pncb2* and lactate utilisation protein *lutC* encoding genes. NaiP_3 was also sandwiched between hypothetical

groups as well as MFS transporters *ttuB_4* and *ttuB_3*. *B. multivorans* ATCC 17616 was used in the Artemis searches for the lineage 2 unique gene, *glnM_2,* an identified it between genes *mltF* and *yecS_2*. MltF is a predicted membrane-bound lytic protein and YecS_2 is associated with ABC transporters as an inner membrane amino acid. The GC content of all gene regions with the lineage-specific genes were encoded was equivalent to the overall GC content of *B. multivorans* (67.1%), with no indication atypical GC content association with foreign genetic material.

4.2.8 *Lineage-specific PCR design*

4.2.8.1 *In silico* analysis

To design the lineage-specific PCR primers, the extracted FASTA sequences (n =76) were aligned so that mismatches could be identified in the opposing lineage. This was required because the genes were identified using a complete gene presence-absence cut-off. Therefore, homologous CDS were present in the opposing lineage, at a lower identity and coverage, compared to the target lineage. This manual mismatch search was first performed at the start of the target gene sequence, looking for at least two mismatches, on at least one of the primer pair, on the 3' end of a 18-22 bp nucleotide sequence. The same approach was used downstream for the reverse primer. Once the sequence pairs were identified, they were checked for quality using online tools (Section 2.12). This looked for primer dimers, self-annealing hairpin sequences, and for appropriate annealing temperatures. One primer pair was created for each lineage-specific gene (Table 26).

Table 26 | PCR primer sequences identified from the *B. multivorans* lineage-specific target genes. Associated PCR sequence metrics are also noted.

Target Gene	Primer Name	Primer Sequence	Primer Length (bp)	Position	Annealing Temperature (°C)	Product (bp)	Size
YiaJ_1	YIAJBM1F	**AT**C**C**GG**C**AA**CT**ATTCGCT	18	4007519-4007536[a]	53.3	537	
	YIAJBM1R	**CAACGCTTTCCGTAGATG**	18	4007000-4007017[a]			
ghrB_1	GHRBBM1F	CAAGCAACCGACCGAA**AG**	18	4008677-4008694[a]	53.0	744	
	GHRBBM1R	GGAGACAG**A**ATCACGTT**C**	18	4009403-4009420[a]			
naiP_3	NAIPBM1F	**A**GCCGC**C**GA**AC**AAA**G**A**T**TG**A**	20	4014235-4014254[a]	55.7	981	
	NAIPBM1R	CTGAAGCCGGTCAGAAA**G**	18	4015198-4015215[a]			
glnM_2	GLNMBM2F	T**G**AA**T**G**CCG**GCCACGTA**TG**	19	1792198-1792216[b]	55.5	322	
	GLNMBM2R	GACGCATACGACAG**T**TCC	18	1791895-1791912[b]			

*Mismatches for each primer sequence are highlighted in bold. Red indicates a mismatch in all the opposing lineage and blue indicates mismatches in some strains, but not all. [a]Position found in complete genome BCC0084 (lineage 1). [b]Position found in complete genome ATCC 17616 (lineage 2a).

To visually estimate the amplicon size for each of the genes, SnapGene was used to provide images of the expected banding pattern (Figure 30).

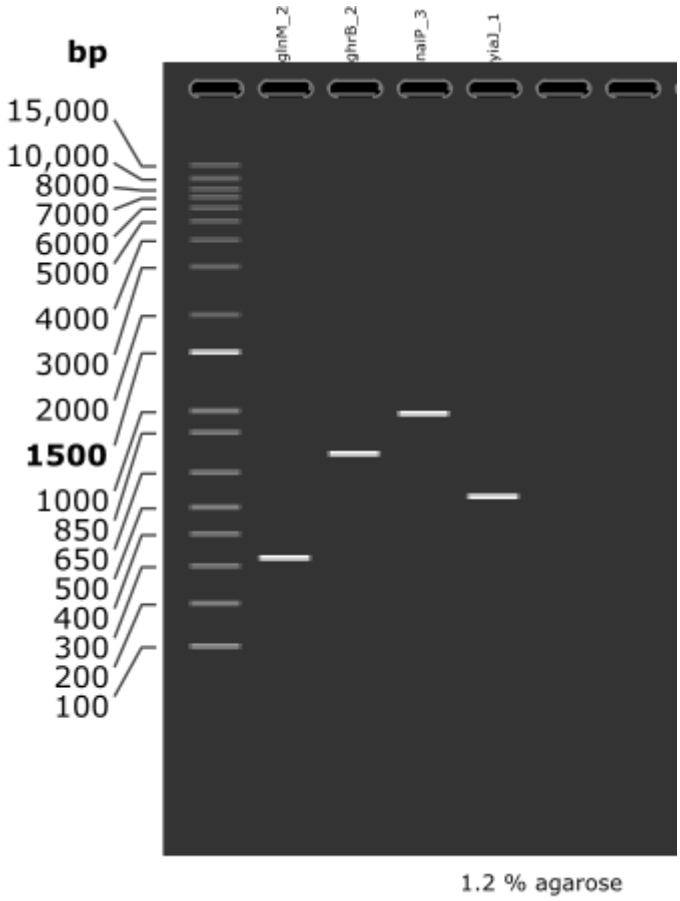


Figure 30 | *In silico* analysis of the lineage-specific PCR products on a 1.2% agarose gel. Figure produced using SnapGene software.

4.2.8.2 Optimisation of PCR protocol

To optimise the PCR protocol, a gradient PCR was performed on the *B. multivorans* model strains. This used a range of annealing temperatures, from 48-62 °C. The PCR setup is found in Table 27, using the standard elongation time for expected products less than 1k*B.* The PCR was performed for 35 cycles.

Table 27 | *B. multivorans* lineage-specific PCR primer programming

PCR Stage	Temperature (°C)	Number of cycles	Time (mins)
Initial heating step	95	1	5
Denaturation	95	30	0.5
Annealing	62[a] or 65[b]		
Extension	72		
Final extension	72	1	10

The results of the gradient PCR showed that the optimum melting temperature (°C) for all lineage 1 primers (YIAJBM1, GHRBBM1, and NAIPBM1) was 62 °C and lineage 2 primer pair (GLNMBM2) was 65 °C. The lineage 1 primer pair with the best banding pattern was GHRBMM1 (*ghrB_1*). This

was chosen alongside the lineage 2 primer pair to take forward for further testing against a larger panel of *B. multivorans* strains (Section 2.1).

4.2.8.3 Testing of lineage-specific PCR primers

The lineage-specific primer pairs (GHRBBM1 and GLNMBM2) were tested on a subset (n =49) of the *B. multivorans* strain panel (Table 5, Section 2.1), which have also been used in phenotypic analysis in this PhD thesis (Chapter 5). Primer pair GHRBBM1 successfully amplified the genomic DNA of all lineage 1 *B. multivorans* strains tested (n =18) (Figure 31). This provided an amplicon of ~744 bp GLNMBM2 primers were successful in the amplification of all lineage 2 *B. multivorans* strains tested (n = 31), with an amplicon ~322 bp. Negative controls were performed using DNAse-free water and DNA from opposing lineage strains for both primer pairs. The results showed that the primer pairs were 100% specific for identifying the target *B. multivorans* lineage in all samples tested. A secondary analysis was also performed to check their specificity against 2 other *Burkholderia* species. The strains tested were *B. ambifaria* AMMD, *B. cenocepacia* K56-2 and *B. cenocepacia* BCC0019. All three were negative when amplified with both sets of primers. This further concluded that not only were the primers *B. multivorans* lineage-specific, but species-specific in this limited comparison of other members of the BCC.

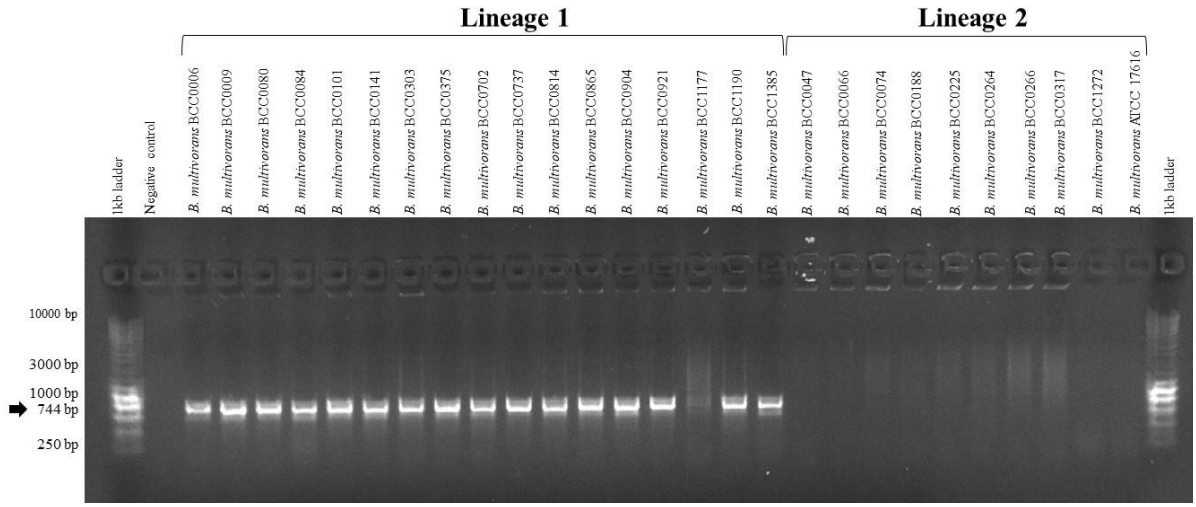


Figure 31 | GHRBM1 PCR gel of *B. multivorans* strains. PCR designed to target lineage 1 (n = 18) *B. multivorans* strains. DNAse free water was used for the negative control. Negative *B. multivorans* controls were those of the opposing lineage 2 (n = 31). Agarose gel was a concentration of 1.2% (w/v) run at 80V.

4.2.9 *Amenability of B. multivorans model strains to genetic manipulation*

4.2.9.1 Triparental mating of pMLBAD

The *B. multivorans* model strains (n = 4) (ATCC 17616, BCC1272, BCC0033, and BCC0084) were initially subjected to triparental mating (Section 2.13.1) using the empty backbone pMLBAD vector (Lefebre and Valvano 2002) to assess their ability to uptake plasmids known to conjugate into BCC

species. The pMLBAD vector provided the *B. multivorans* strains with a resistance to 150 µg/ml trimethoprim from the presence of this selective marker. After incubation, all the WT control plates had no visible colonies, whilst viable counts could be made on the transformed *B. multivorans* plates. This confirmed the successful conjugation into and carriage of pMLBAD in the *B. multivorans* strains.

4.2.9.2 Electroporation of eGFP and mCherry

Once it was shown that the *B. multivorans* model strains could be genetically manipulated via conjugation and triparental mating (Section 4.2.9.1), electroporation using pIN301-eGFP (Mesureur *et al.* 2017) and pIN233-mCherry (Gomes *et al.* 2018) fluorescent gene reporter plasmids was performed. This also included use of strains *B. multivorans* BCC1385 and *P. aeruginosa* LESB58. BCC1385 (total strains: $n = 6$) was later identified as another potential model strain but because it was not successfully genome sequenced to completion (Section 3.2.8), it was not ultimately selected. *P. aeruginosa* LESB58 was also transformed with the same plasmids to yield fluorescent reporter strains for use in co-culture assays (Section 5.2.7). Chloramphenicol selection (25 µg/ml) was used to select for bacterial transformants in all cases. These were re-streaked onto fresh selective agar before imaging to confirm the uptake of each fluorescent plasmid. All *B. multivorans* ($n = 5$) and *P. aeruginosa* experimental strains ($n = 1$) had the ability to receive pIN301-eGFP. However, *B. multivorans* strains ATCC17616 and BCC1272 were unable to uptake the pIN233-mCherry plasmid over the course of 5 attempts. Therefore, only 3 *B. multivorans* strains (60%) were able to uptake the pIN233-mCherry plasmid. An example of the fluorescent vs WT *B. multivorans* is found in Figure 32. To confirm the presence or absence of the plasmids, an M13 PCR (targeting the reporter plasmid backbone) was performed and visualised using gel electrophoresis (Section 2.13.4). This showed a band where a plasmid was present and no band for the respective WTs.

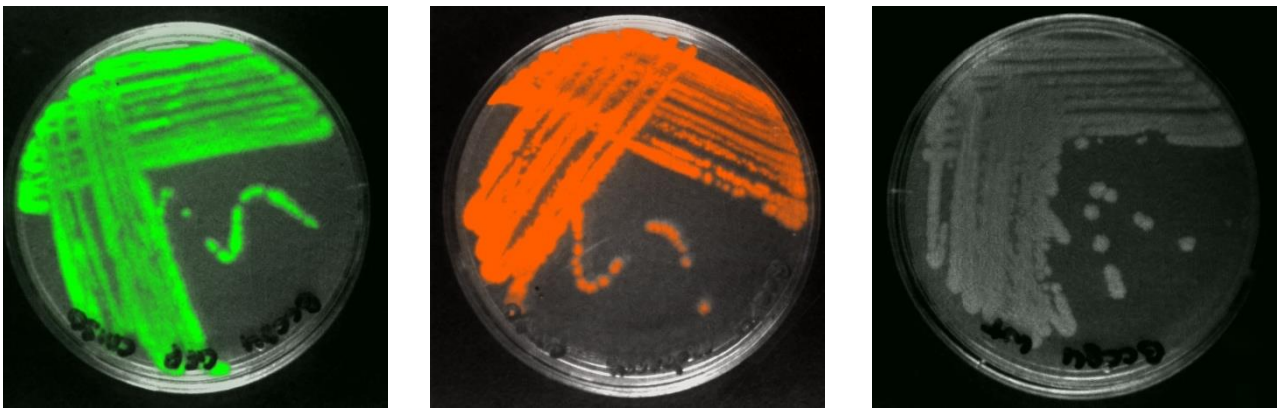


Figure 32 | Successful genetic manipulation of *B. multivorans* using pIN301-eGFP and pIN233-mCherry plasmids. Example shows *B. multivorans* model strain BCC0084. Left: BCC0084::pIN301-eGFP, middle: BCC0084::pIN233-mCherry, right: BCC0084 parent (wild-type). Genetically manipulated strains were grown on TSA supplemented with 50 µg Chloramphenicol. Wild-type strain was grown on TSA without antibiotic selection.

Fluorescence was imaged on a Biospace Lab PhotonIMAGER Optima (pIN301-eGFP excitation = 488 nm and emission = 522 nm; pIN233-mCherry excitation = 580 nm and emission = 610 nm).

4.2.9.3 Genetic stability of the fluorescent reporter plasmids in *B. multivorans* and *P. aeruginosa*

The fluorescent-tagged strains with pIN301-eGFP (n = 6) and pIN233-mCherry (n = 4) were checked for stability of the plasmids over a 5-day period without antibiotic selection (Section 2.13.5). The relative fluorescence units were logged ($\log_{10}$) for each time point for easier comparison of the results. The results differed slightly depending on which fluorescent plasmid was being analysed. The pIN301-eGFP plasmid was stable over the course of the experiment for *P. aeruginosa* and 2 of 4 (50%) *B. multivorans* strains. However, for ATCC 17616 and BCC1272, the $\log_{10}$ RFU remained around 5-5.5 for days 0 and 1, before rapidly declining over the remainder of the experiment (Figure 33).

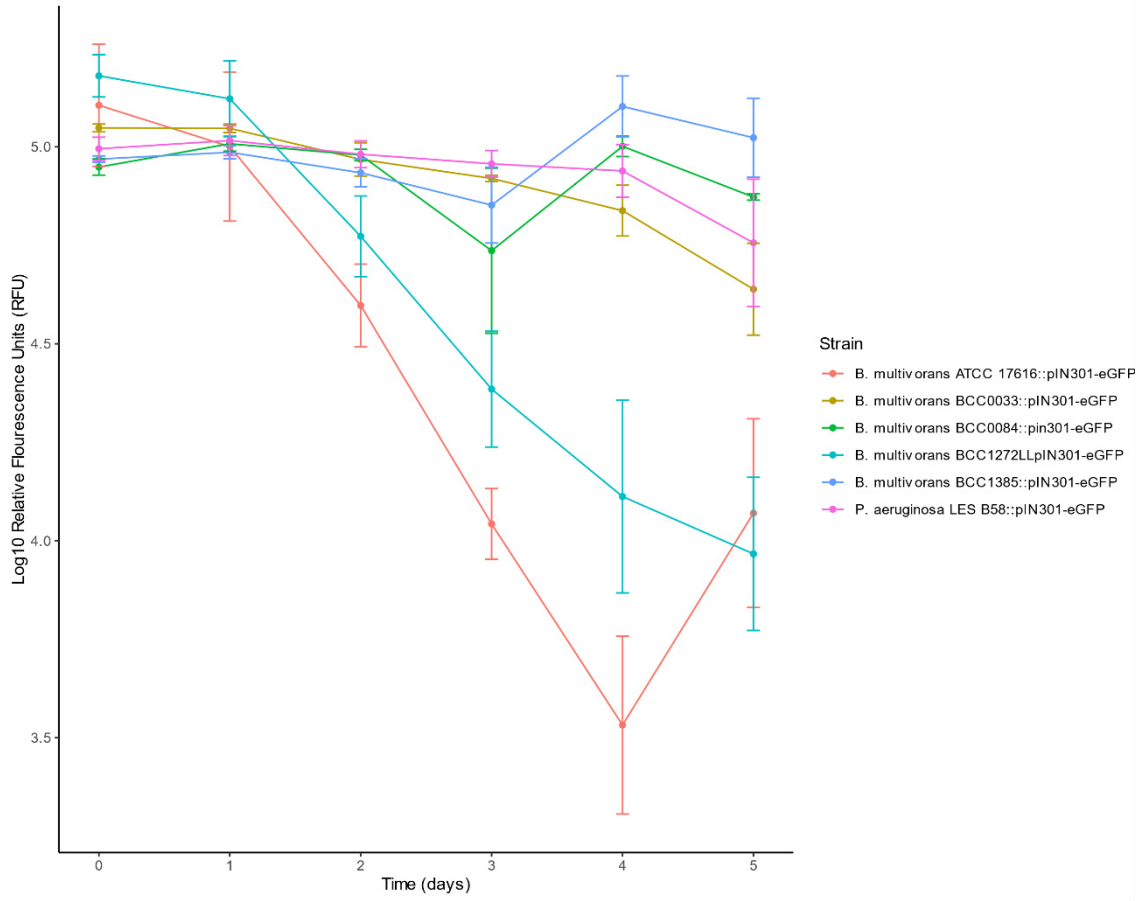


Figure 33 | $\log_{10}$ Relative Fluorescence Units (RFU) of the *B. multivorans* and *P. aeruginosa* LESB58 strains genetically manipulated to harbour the pIN301-eGFP plasmid. Readings were taken in a plate reader every 24-hours over the course of 5 days. Wavelengths used was excitation at 488_{nm} and emission at 522_{nm}. Error bars represent the standard error mean (SEM).

When looking at the stability of pIN233-mCherry, the $\log_{10}$ RFU for all *P. aeruginosa* (n = 1) and *B. multivorans* (n = 3) strains started high (between 4-4.25 RFU) and remained stable until day 4. The

log_{10} RFU then decreased for all strains examined by day 5. BCC0084 (lineage 1) retained the most stability of fluorescent signal at day 5, reading around 3.8 log_{10} RFU (Figure 34).

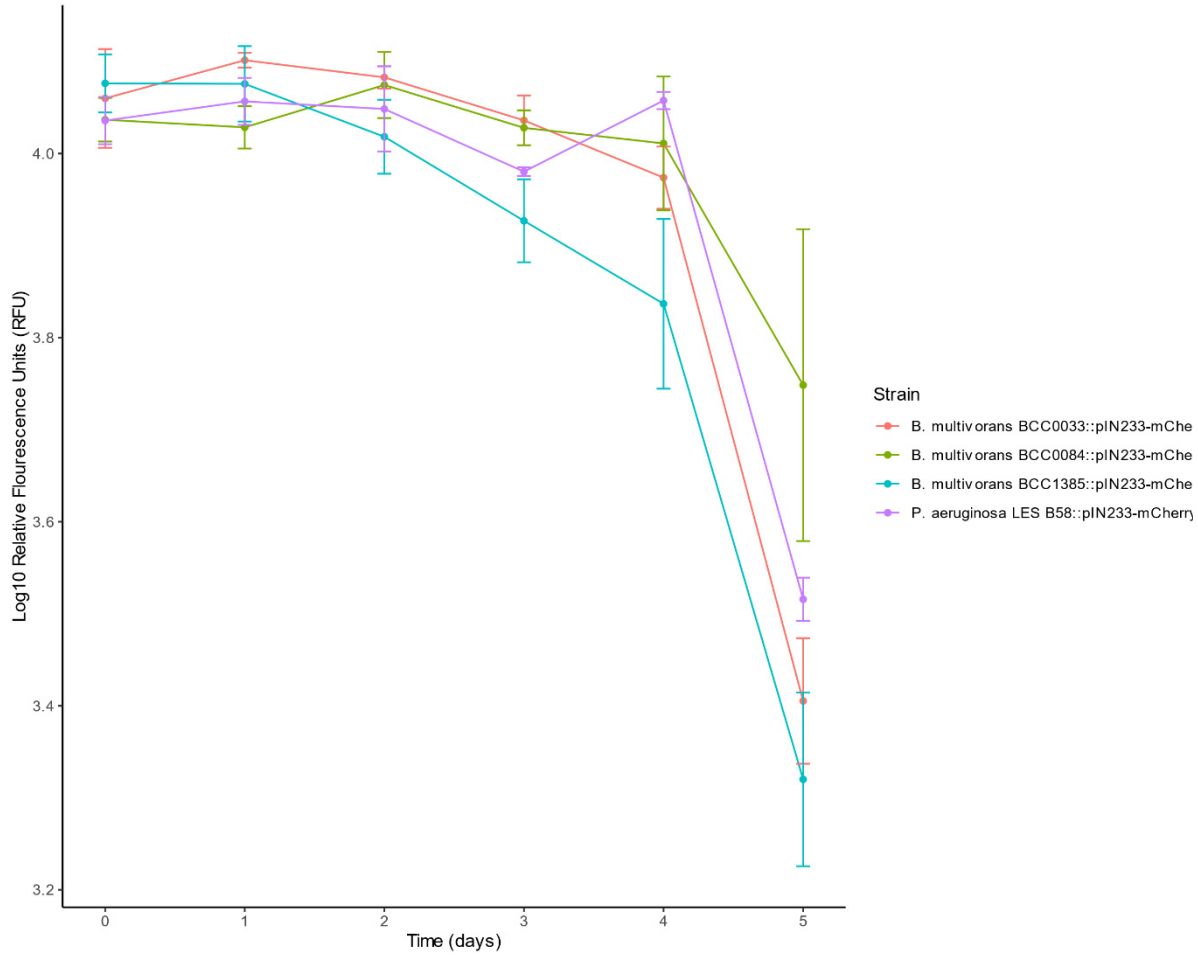


Figure 34 | Log_{10} Relative Fluorescence Units (RFU) of the *B. multivorans* and *P. aeruginosa* LESB58 strains genetically manipulated to harbour the pIN233-mCherry plasmid. Readings were taken in a plate reader every 24-hours over the course of 5 days. Wavelengths used was excitation at 550_{nm} and emission at 610_{nm}. Error bars represent the standard error mean (SEM).

4.2.10 *Lung infection modelling of B. multivorans model strains using in vivo experiments*

4.2.10.1 *B. multivorans* infection model strains and bacterial dissemination

Four *B. multivorans* model strains (BCC0033 WT; lineage 2b, BCC0033::pIN301-eGFP, BCC0084::pIN301-eGFP; lineage 1 and ATCC 17616 WT; lineage 2a) were chosen for *in vivo* experimental analysis in a modified murine lung infection model (Green *et al.* 2021). Previously in this thesis (Section 3.2.8), these model strains were identified as representatives of CF *B. multivorans* model strains, with ATCC 17616 being a well-characterised environmental strain, isogenic with CF strain BCC1272. This research was carried out in collaboration with Dr. Daniel Neill and Dr. Angharad Green

at the University of Liverpool. They performed the creation of infection stocks, initial murine inoculation and monitoring of infection, and murine organ harvesting. There was also their assistance with me for performing organ homogenization and serial dilutions. The analysis of the output data presented herein being my own. Each mouse was intranasally infected under anaesthesia with ~10^7 CFU/50µl of each *B. multivorans* isolate (Section 2.15.2). Bacteria were isolated from both the lung and nasopharynx of each mouse then quantified to identify the bacterial burden on these organs at 1-, 3-, and 5- days post-infection.

When looking at the lung CFU burden, BCC0033 WT had no significant differences between all 3 time points (Figure 35). BCC0033::pIN301-eGFP had a small but significant difference in CFUs between time points 1 and 5 days ($p = 0.032$) (Figure 35). However, the most significant changes in CFUs, where a loss in microbial burden was observed, occurred with strains BCC0084::pIN301-eGFP and ATCC 17616 WT. BCC0084::pIN301-eGFP showed a decrease in significant viable count when comparing day 1 and 3, day 1 and 5 and days 3 and 5 respectively in the murine lung ($p < 0.05$ for all; Figure 35). For *B. multivorans* strain ATCC 17616 WT, the greatest loss in viability and microbial burden occurred between day 1 and 3 ($p = 0.01$), with the CFU counts not altering significantly between day 3 and 5 (Figure 35). In contrast, no significant differences in CFU counts were seen between any of the time points (1, 3 and 5 days) for all strains when looking at the nasopharynx (Figure 36). Here the viable count remained at approximately $\log_{10}$ 3.8 CFU/ml for BCC0033 (and its GFP variant) and $\log_{10}$ 4.5 CFU/ml BCC0084, while the nasopharynx counts of *B. multivorans* strain ATCC 17616 remained at approximately$\log_{10}$ 1.1 CFU/ml.

Overall, BCC0084::pIN301-eGFP (lineage 1) had the highest viability counts, at all three time points, for both the lung (Figure 35) and nasopharynx when compared to the other strains (Figure 36). In contrast, environmental strain ATCC 17616 (lineage 2b) had the lowest viable counts (CFUs) of all the *B. multivorans* strains for both the lung (Figure 35) and nasopharynx at all time points (Figure 36).

Both pIN301-eGFP and WT strains were used for this experiment to determine whether there would be differences in a fluorescent-presenting genetically manipulated *B. multivorans* model strain. BCC0033 (strain C5568; lineage 2b) was chosen as the comparator for this as it has been previously used in pulmonary BALB/c mouse infection models (Chu *et al.* 2004). Chronic infection was recorded by the number of cultivatable, viable *B. multivorans* cells counted on *Burkholderia cepacia* selective agar at each time point (Section 2.15.2). Figure 35 and Figure 36 showed that the viable count of BCC0033 WT had no significant difference at all time points when compared to BCC0033::pIN301-eGFP in both the lung and nasopharynx.

These findings indicate that a short-term chronic infection of both the lung and nasopharynx can occur when infected by any of the four *B. multivorans* strains. Within the nasopharynx a mean of 4.1, 4.0, 4.7 and .80 $\log_{10}$ CFU was recovered after 5-days post-infection for BCC0033 WT, BCC0033::pIN301-

eGFP, BCC0084::pIN301-eGFP and ATCC 17616 WT, respectively. The lung infection viable count (log_{10} CFU) for BCC0033 WT, BCC0033::pIN301-eGFP, BCC0084::pIN301-eGFP and ATCC 17616 WT was 4.1, 3.8, 4.2 and 2.0, respectively.

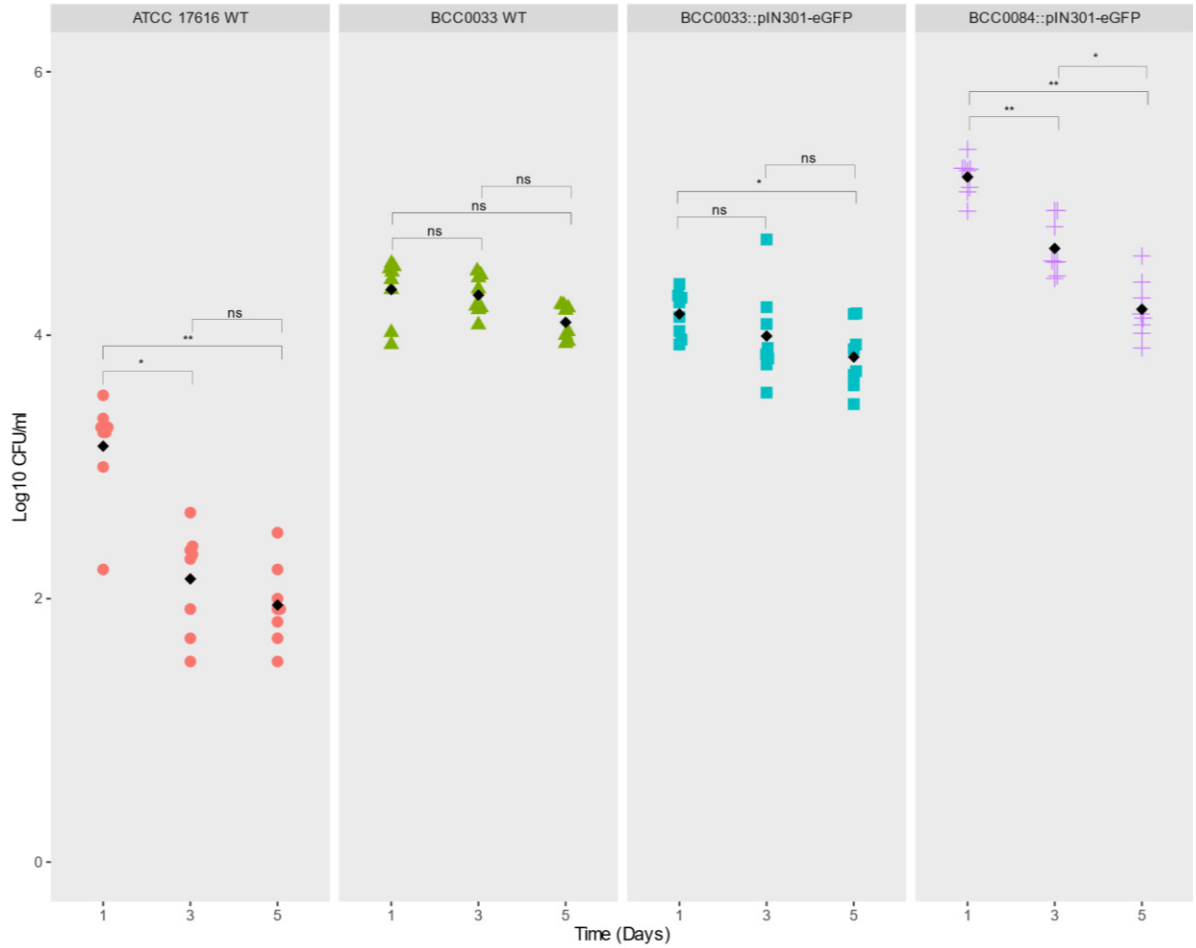


Figure 35 | Log_{10} CFU/ml of the *B. multivorans* strains at 1-, 3- and 5-days post-infection in the murine lung.

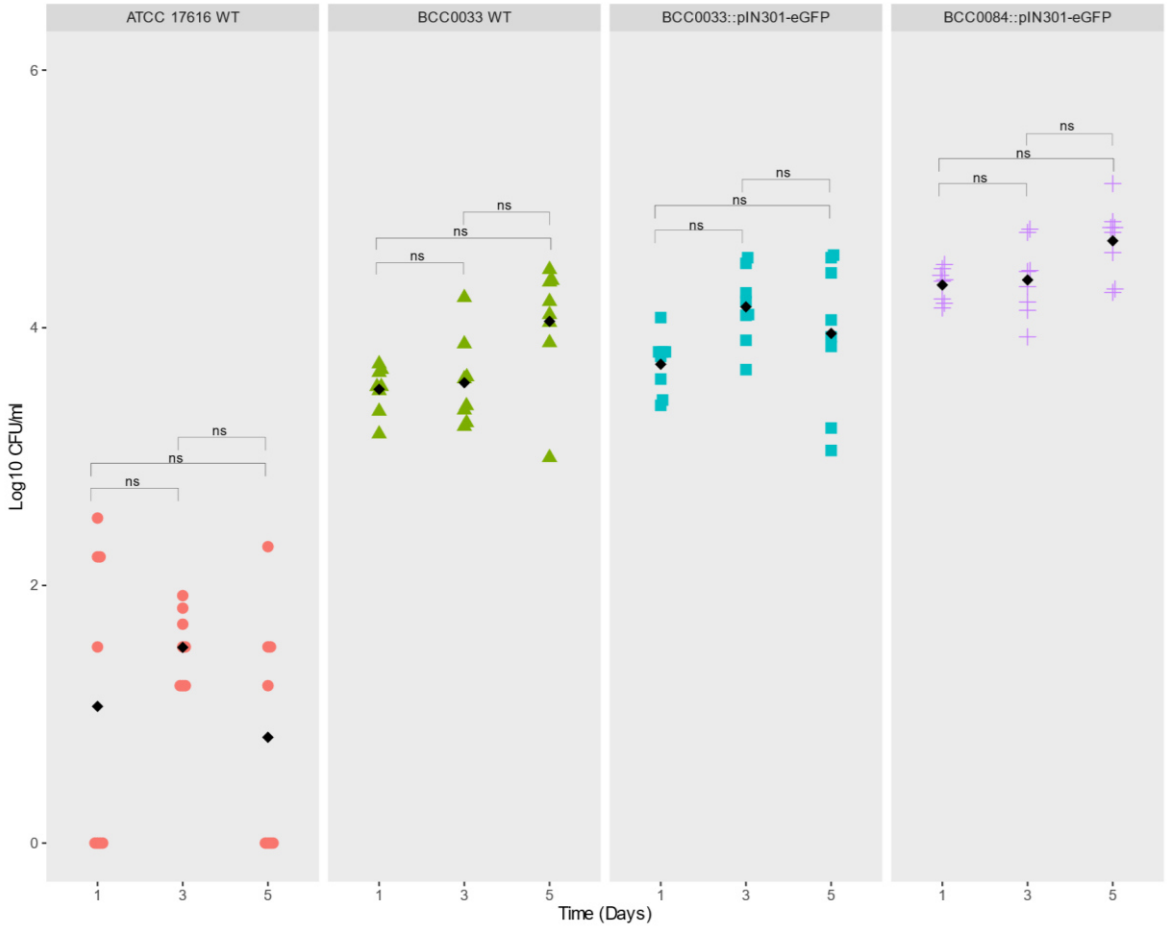


Figure 36 | Log_{10} CFU/ml of the *B. multivorans* strains at 1-, 3- and 5-days post-infection in the murine nasopharynx.

4.2.10.2 Mice pain during chronic infection with *B. multivorans*

Pain scores for each mouse were taken every 24-hours, between 1- and 5-days post-infection, presenting on a scale of 0-5. At the first infection time point (24-hours), all mice (n = 96) presented with pain (scores 2-5), irrespective of the *B. multivorans* strain they were infected with (Figure 37). Two mice (2.08%) reached the highest pain score of 5, both infected with BCC0084::pIN301-eGFP (lineage 1) at the 24-hour time point (Figure 37). Nine mice (9.38%) presented a pain score of 4 at the 24-hour time point, encompassing all four strains. BCC0033::pIN301-eGFP, BCC0033 WT (lineage 2b), BCC0084::pIN301-eGFP and ATCC 17616 WT (lineage 2a) had 2, 1, 4 and 2 mice (n = 24 per *B. multivorans* strain) presenting a score of 4 at 24-hours, respectively (Figure 37).

By 48-hours of infection, pain scores for all mice (n = 64 total; n = 16 per *B. multivorans* strain) decreased to between 0 and 3. BCC0033 WT had three mice with the highest presenting pain (score 3 of 5) at this time point. A pain score of 0 was observed at least once for each *B. multivorans* infection strain. ATCC 17616 WT was the most observed strain to present a pain score of 0 amongst the mice at

48-hours (n = 11 of 16; 68.75%). BCC0084::pIN301-eGFP and BCC0033 WT were seen to have a pain score of 0 in only one mouse per infection strain (6.25%) (Figure 37).

The pain scores for the mice further decreased by 72-hours, with scores of 0 and 1 (Figure 37). BCC0033 WT was the only *B. multivorans* strain to cause pain in the mice at this time point, with 6 of the 16 mice infected by this strain (37.5%) showing a small amount of pain. No pain was observed in mice infected with the ATCC 17616 WT (n = 4), BCC0033::pIN301-eGFP (n = 8) and BCC0084 (n = 8) at 72-hours (Figure 37).

At the final 96-hour post-infection time point, no pain was observed in any of the mice (Figure 37). This suggests either the mice became more resistant to the effects of the infection or that the *B. multivorans* had evaded the immune response, as seen in *B. cenocepacia* (Ganesan and Sajjan 2011), enabling further infection without an immune reaction.

Except for the 96-hour post-infection mark, the mice pain scores largely correlate to the infection burden (Figure 35 and Figure 36).

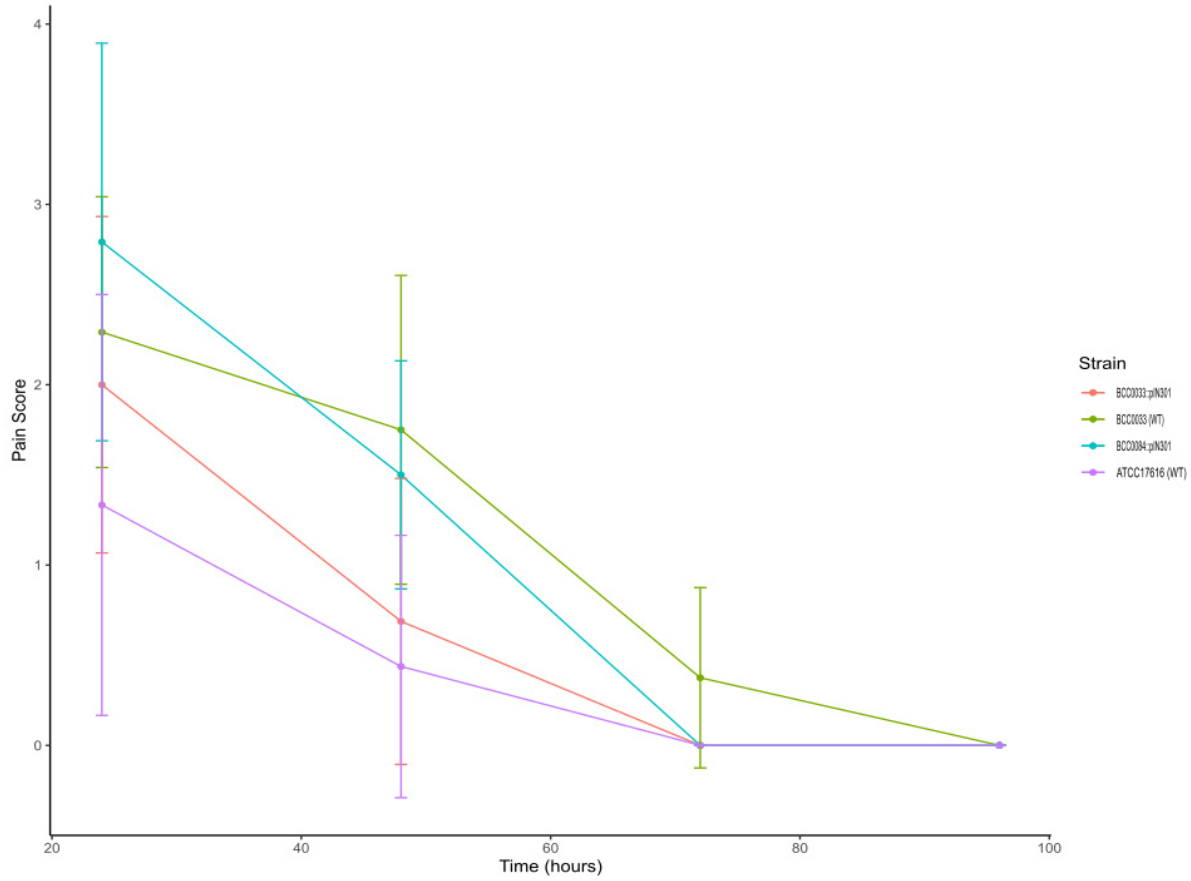


Figure 37 | Pain scores for the mice infected with each *B. multivorans* strain over a 96-hour period. Error bars represent the standard deviation.

4.2.10.3 Mice weight over the course of *B. multivorans* infection

The experimental mice were weighed every 24-hours over the 5-day post-infection period, including records of their starting weight. The average starting weight for the mice were 19.5 g, 18.9 g, 19.1 g and 20.1 g for BCC0033 WT (lineage 2b), BCC0033::pIN301-eGFP, BCC0084::pIN301-eGFP (lineage 1) and ATCC 17616 WT (lineage 2a), respectively. Weights for each infected mouse was recorded as a percentage of the original starting weight at each time point (Figure 38).

BCC0033::pIN301-eGFP did not correlate to the other infection strains, with the mice having an average maintenance or gaining of weight over the 5-day period (Figure 38). Mice infected with BCC0033::pIN301-eGFP lost, on average, 0.3% of their starting weight after 24-hours. At 48-hours, the mice gained 0.8% of their starting weight, further increasing to an average 2.8% gain after 72-hours. The average weight of the mice had increased to 103.1% of the original weight (+3.1% gain) after the full 96-hour infection period (Figure 38).

As for the other three *B. multivorans* strains (BCC0084::pIN301-eGFP, BCC0033 WT and ATCC 17616 WT), the mice in all lost weight in the first 48-hours of infection but regained at 72- and 96-hours. The average weight lost for the mice was 4.3%, 5.4% and 4.1% of the starting weight for strains

BCC0033 WT, BCC0084::pIN301-eGFP and ATCC 17616, respectively. The weights dropped a further 1.2% for BCC0033 WT and 2.5% for ATCC 17616 WT. However, BCC0084::pIN301-eGFP mice averagely gained 0.5% of their starting weight back at 48-hours. By 72-hours, the average weights of the mice increased to 99.6% (BCC0033 WT), 98.8% (BCC0084::pIN301-eGFP) and 97.0% (ATCC 17616 WT) of their original weight. Mice infected with BCC0033 WT and BCC0084::pIN301-eGFP succeeded their starting weights by 0.5% and 0.2% after 72-hours, whereas the ATCC 17616 WT-infected mice didn't regain all their average starting weight, finishing at a 1.5% weight loss.

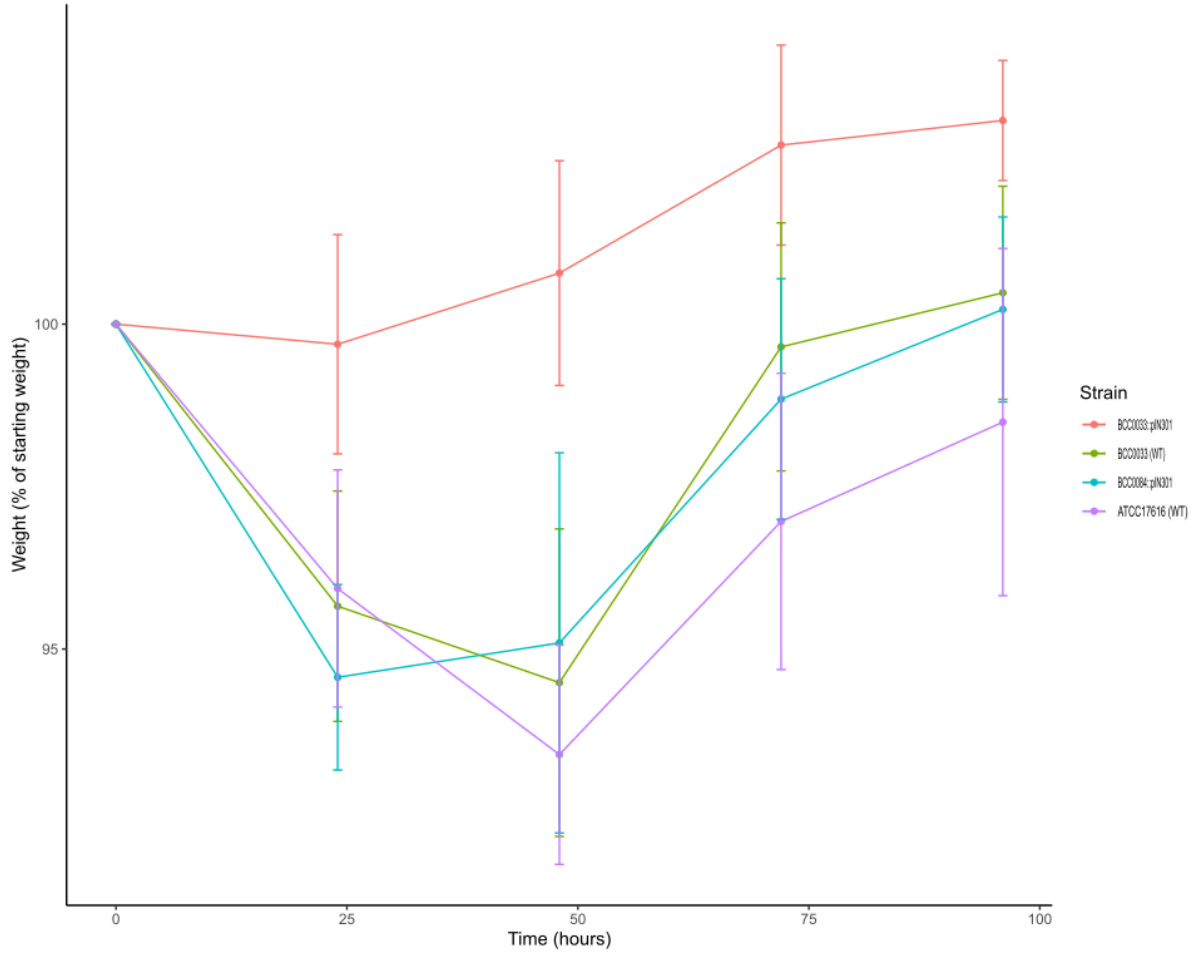


Figure 38 | Weight of the mice infected with each *B. multivorans* strain over a 96-hour period. Error bars represent the standard deviation. Each point is the percentage (%) of starting weight.

4.2.11 *Re-sequencing of B. multivorans infection isolates*

To understand if genetic adaptations occurred in the *B. multivorans* after short-term infection in the mouse models, genome re-sequencing was performed. This used pooled bacterial isolates from each strain, for each time point at day 3 and day 5 post-infection (n = 20 *B. multivorans* strains). Fresh confluent colony growth was obtained on TSA before isolating and performing DNA extraction as previously described (Section 2.3.1). The genome sequences were then processed for quality, contamination adapters trimmed, and *de novo* assembled as per Sections 2.6.1 and 2.6.3. The *B. multivorans* isolates were then scaffolded against the representative complete genome as per Section 2.6.5.

4.2.11.1 Quality metrics of the re-sequenced genomes

The QUAST (Gurevich *et al.* 2013) quality metrics of the re-sequenced genomes (n = 20) are shown in Table 28. The sequence size of the infection stocks were 6.7 Mb, 6.7 Mb, 6.6 Mb, and 6.9 Mb for the BCC0033::pIN301-eGFP, BCC0033 WT (lineage 2b), BCC0084::pIN301-eGFP (lineage 1) and ATCC 17616 WT (lineage 2a) infection stocks, respectively (Table 28). The genome sizes remained consistent

between the infection stocks and the day 3 and 5 lung and nasopharynx isolates for each strain. Number of contigs ranged from 67 to 158 (mean = 106.6) (Table 28). The GC content for all strains was between 66.82% and 67.09%. The average number of CDS, rRNAs and tRNAs were 6038.4, 2.25 and 69.85, respectively (Table 28). To confirm that the re-sequenced genomes were *B. multivorans,* the minikraken database from Kraken2 was used. No evident of contaminating DNA was found in the genome re-sequencing. From the overall genome quality metrics, the initial infection stocks were concluded to be the same as the model strains at this top-level, thus facilitating further SNP variant analysis.

Table 28 | QUAST quality metrics of the *B. multivorans* murine infection model isolates and respective infection stocks (*n* = 20).

Strain	Sequence size (bp)	No. of contigs	GC content (%)	N50 value	L50 value	CDS	rRNAs	tRNAs
BCC33::pIN301-eGFP Infection Stock	6704899	77	67.08	348864	7	5975	2	71
BCC33::pIN301-eGFP Lung Day 3	6705084	67	67.08	361472	7	5981	2	71
BCC33::pIN301-eGFP Lung Day 5	6705126	69	67.08	361472	7	5979	2	71
BCC33::pIN301-eGFP Naspharynx Day 3	6705056	74	67.08	361472	7	5979	2	71
BCC33::pIN301-eGFP Naspharynx Day 5	6701849	84	67.08	339227	6	5980	2	71
BCC33 WT Infection Stock	6700552	73	67.09	348872	7	5975	2	71
BCC33 WT Lung Day 3	6700861	75	67.09	339204	7	5972	2	69
BCC33 WT Lung Day 5	6700749	71	67.09	361472	7	5976	2	71
BCC33 WT Naspharynx Day 3	6700133	81	67.09	461270	5	5972	2	71
BCC33 WT Naspharynx Day 5	6705054	68	67.08	381097	7	6048	2	70
BCC84::pIN301-eGFP Infection Stock	6603701	137	67.09	166195	13	5954	2	69
BCC84::pIN301-eGFP Lung Day 3	6604269	133	67.09	186877	13	5958	2	69
BCC84::pIN301-eGFP Lung Day 5	6606395	123	67.09	188074	13	5961	2	69
BCC84::pIN301-eGFP Naspharynx Day 3	6606810	123	67.09	188074	13	5959	2	69
BCC84::pIN301-eGFP Naspharynx Day 5	6605897	129	67.09	165817	14	5956	2	69
ATCC 17616 WT Infection Stock	6879189	158	66.82	172395	11	6226	3	69
ATCC 17616 WT Lung Day 3	6883471	155	66.82	178859	11	6228	3	69
ATCC 17616 WT Lung Day 5	6881268	148	66.82	172367	12	6226	3	69
ATCC 17616 WT Naspharynx Day 3	6884754	141	66.82	178859	10	6230	3	69
ATCC 17616 WT Naspharynx Day 5	6887053	146	66.82	172388	12	6233	3	69

4.2.11.2 Genome scaffolding and statistics

The re-sequenced murine infection stocks (*n* = 4) and isolates (*n* =16) were scaffolded to the corresponding complete *B. multivorans* model strains (*n* = 4) (Section 3.2.8.1) using CONTIGuator. All isolates had over 99% of the genomes scaffolded to the references (Table 29). Both the infection isolates and the respective lung and nasopharynx isolates all harboured 4 replicons, like that observed in Section 3.2.8.1. The re-sequenced lengths and GC content for the genomic scaffolds and each replicon were consistent with the previously scaffolded *B. multivorans* genomes (Section 3.2.9.2).

Table 29 | Scaffold statistics of the *B. multivorans* murine infection stocks (*n* = 4) and infection isolates (*n* = 16) compared to the *B. multivorans* complete genomes.

		Total length		Scaffolded length		C1		C2		C3		C4	
Strain	**% Scaffolded**	**bp**	**GC (%)**	**bp**	**GC (%)**	**bp**	**GC (%)**	**bp**	**GC (%)**	**bp**	**GC (%)**	**bp**	**GC (%)**
BCC33::pIN301-eGFP Infection Stock	99.70318847	6697853	67.08	6677973	67.1	3512212	66.8	2377802	67.7	722282	67.5	65677	57.34
BCC33::pIN301-eGFP Lung Day 3	99.7032252	6698682	67.08	6678802	67.1	3513180	66.8	2377527	67.69	722418	67.5	65677	57.34
BCC33::pIN301-eGFP Lung Day 5	99.70322706	6698724	67.08	6678844	67.1	3513142	66.8	2377643	67.7	722382	67.5	65677	57.34
BCC33::pIN301-eGFP Nasopharynx Day 3	99.69498204	6698294	67.08	6677863	67.1	3512587	66.8	2377651	67.7	722499	67.5	65126	57.32
BCC33::pIN£01-eGFP Nasopharynx Day 5	99.62140055	6694146	67.08	6668802	67.11	3507075	66.81	2373565	67.69	722485	67.5	65677	57.34
BCC33 WT Infection Stock	99.76430393	6693790	67.09	6678013	67.1	3512271	66.8	2377697	67.7	722368	67.5	65677	57.34
BCC33 WT Lung Day 3	99.74208254	6694002	67.09	6676737	67.1	3510906	66.8	2377789	67.7	722365	67.5	65677	57.34
BCC33 WT Lung Day 5	99.76431087	6693987	67.09	6678210	67.1	3512518	66.8	2377707	67.7	722308	67.5	65677	57.34
BCC33 WT Nasopharynx Day 3	99.743704	6692262	67.09	6675110	67.1	3511063	66.8	2378291	67.69	720630	67.5	65126	57.32
BCC33 WT Nasopharynx Day 5	99.70322387	6698652	67.08	6678772	67.1	3513022	66.8	2377697	67.7	722376	67.5	65677	57.34
BCC84::pIN301-eGFP Infection Stock	99.12143575	6592688	67.09	6534767	67.13	3279521	67.25	2509240	67.27	598462	67.39	147544	60.94
BCC84::pIN301-eGFP Lung Day 3	99.12153546	6593664	67.09	6535741	67.13	3279993	67.25	2509742	67.27	598462	67.39	147544	60.94
BCC84:pIN301-eGFP Lung Day 5	99.15419933	6596235	67.09	6540444	67.13	3282531	67.25	2511785	67.26	598584	67.39	147544	60.94
BCC84::pIN301-eGFP Nasopharynx Day 3	99.15398629	6596347	67.09	6540541	67.13	3282533	67.25	2511880	67.26	598584	67.39	147544	60.94
BCC84::pIN301-eGFP Nasopharynx Day 5	99.14066971	6594554	67.09	6537885	67.13	3281215	67.25	2510542	67.26	598584	67.39	147544	60.94
ATCC 17616 WT Infection Stock	99.40529452	6869787	66.82	6828932	66.85	3377204	67.02	2427305	67.26	872114	66.06	152309	61.32
ATCC 17616 WT Lung Day 3	99.28837504	6873986	66.82	6825069	66.86	3376634	67.02	2424489	67.26	870466	66.07	153480	61.33
ATCC 17616 WT Lung Day 5	99.40554342	6872327	66.82	6831474	66.86	3377503	67.02	2427254	67.26	873231	66.07	153486	61.32
ATCC 17616 WT Nasopharynx Day 3	99.37786518	6876323	66.82	6833543	66.85	3381027	67.01	2425919	67.26	871128	66.07	155456	61.28
ATCC 17616 WT Nasopharynx Day 5	99.32455948	6878622	66.82	6832161	66.85	3381015	67.01	2424499	67.26	871141	66.07	155519	61.29
Average	99.48185563	6715246	67.0195	6680484	67.0465	3420858	66.967	2422801	67.4795	728978.3	67.09421	110419	59.32947

4.2.12 Comparative genome analysis

Gene presence absence of the *B. multivorans* infection isolates (n = 20) were interrogated using the output matrix from Roary. The day 3 and 5 isolates (n = 16) were compared to the initial infection stock (n = 4) to look for any missing genes. This aimed to identify whether there were any key genes which may be lost as part of niche colonisation and adaption in a short time.

4.2.12.1 Potential gene absence in post-infection isolates

The gene presence absence matrix from Roary was used to identify genes which were may have been lost and deleted from the infection isolate genomes compared to the original infection stock genomes. The gene absence predictions may have been influenced by the annotation and assembly of the draft genomes. However, *Burkholderia* species are known to have genome plasticity, where large deletions can occur. This includes deletion of the 3rd replicon (Agnoli *et al.* 2012).

There were 11 genes which were missing in at least one genome of the BCC0033::pIN301-eGFP infection isolates. Of this, 8 (72.7%) absent genes were hypothetical proteins. The annotated missing genes were *rfbB_1, rmlC_1,* and *etfB_1*. All three of these genes were absent only in BCC0033::pIN301-eGFP nasopharynx infection isolate day 5. Genes *rmlC_1* and *rfbB_1* are both associated with polysaccharide biosynthesis, with associations to LPS. RmlC_1 biosynthesises type II O-antigenic polysaccharides whilst RfbB_1 biosynthesises ABC transporter and LPS genes in other *Burkholderia* species (*B. pseudomallei* and *B. thailandensis*) (Kovacs-Simon *et al.* 2019). EtfB is an electron transfer flavoprotein ß-subunit which has previously been shown as essential in *B. cenocepacia* (Bloodworth *et al.* 2015).

When comparing the ATCC 17616 WT (lineage 2a) genomes, there were 22 genes absent in at least one of the day 3 and 5 infection isolates compared to the infection stock. 20 of the 22 genes (90.9%) were hypothetical proteins. Gene *lifO*_1 lipase chaperone was missing in all infection isolates except for the initial infection stock. IS5 family transposase ISBmu2 was also absent in ATCC 17616 WT nasopharynx day 5 isolate pool. ISBmu2 was found as a novel 1,210 bp IS element previously in ATCC 17616 (Ohtsubo *et al.* 2005).

When analysing the BCC0033 WT (lineage 2b) group, there were 10 genes missing from at least one infection isolate. Acs_3 and EftB_1 was missing in the nasopharynx day 3 isolate and *fitB_2* was missing in the lung day 3 isolate. The other 7 genes (70.0%) were hypotheticals. The gene absence correlated with the *eftB* gene missing in the BCC0033::pIN301-eGFP nasopharynx day 5, indicating that this gene may be lost in chronic nasopharynx infection in this *B. multivorans* strain. The *fitB* gene is the toxic part of the type II toxin-antitoxin system (Mattison *et al.* 2006).

For BCC0084::pIN301-eGFP (lineage 1), 17 genes were absent in at least one infection isolate compared to the infection stock. Of this, 14 genes (82.4%) were made up of hypotheticals. Annotated

absent genes were *btuD_8*, *phnY_2* and *phnY_3*. Expression of *btuD* enhances both vitamin B_{12} absorption as well as antibiotic transfer (Wang *et al.* 2019). This gene was absent in 3 of 4 *B. multivorans* BCC0084::pIN301-eGFP infection isolates, with presence in the lung day 3 isolate. The *phnY* genes were absent in lung day 3 and nasopharynx day 5 BCC0084::pIN301-eGFP isolates. The *phnY* gene is a phosphonoacetaldehyde dehydrogenase. There was no specific links to annotated gene loss and isolate type in this *B. multivorans* infection strain.

Overall, by annotating the draft genomes of the murine infection isolates compared to the initial infection stocks, it was clear that there were no major deletions of genes (100s of loci) over the 5-day experimental period. Small numbers of putative gene deletions have been noted above and would require further experimentation to examine whether they had occurred due to the genome sequencing rather than during the infection model.

4.2.13 Variant analysis of *B. multivorans* infection isolates

4.2.13.1 Core SNP phylogenomics of the *B. multivorans* infection models

To provide initial insight into the evolutionary relationships of the *B. multivorans* model strain infection stocks compared to the day 3 and 5 lung and nasopharynx infection isolates, core SNP alignments were used to create FastTree phylogenetic trees. There were no immediate trends within the evolutionary patterns of each model *B. multivorans* strain and the respective infection isolates. BCC00844::pIN301-eGFP (lineage 1) had the most closely-related core SNP strain evolution, reflecting that of the shorter branch distance lengths in the lineage 1 strains (Section 3.2.5.2). In this phylogeny, the lung day 3 isolate was the most distantly related to the other infection isolates, sitting on its own branch from the BCC0084::pIN301-eGFP infection stock reference. The nasopharynx isolates were more closely related to each other, branching off from the BCC0084::pIN301-eGFP lung day 5 isolate (Figure 39).

BCC0033::pIN301-eGFP (lineage 2b) had the least-related core SNPs (Figure 40) compared to all the *B. multivorans* infection models. Interestingly, the lung day 3 and nasopharynx day 3 isolates were the most closely related in this phylogeny. The lung day 5 isolate was most closely related to the BCC0033::pIN301-eGFP infection stock, whilst the nasopharynx day 5 isolate was distant from all other isolates (Figure 40). The BCC0033 WT phylogeny contrasted that of the BCC0033::pIN301-eGFP phylogeny. The most closely related strain to the infection stock reference was the lung day 3 isolate. From this, the lung day 5 and both nasopharynx isolates branched off, with the two nasopharynx isolates showing a higher similarity in the BCC0033 WT infection evolution model (Figure 41) than BCC0033::pIN301-eGFP.

For ATCC 17616 (lineage 2a), the infection isolate evolution was different again. The nasopharynx day 3 isolate was most closely related to the reference infection isolate. From this, the lung day 5 isolate

branched off. A further diversion was then seen for the lung day 3 and nasopharynx day 5 isolates, which shared the most similarity in this phylogeny (Figure 42).

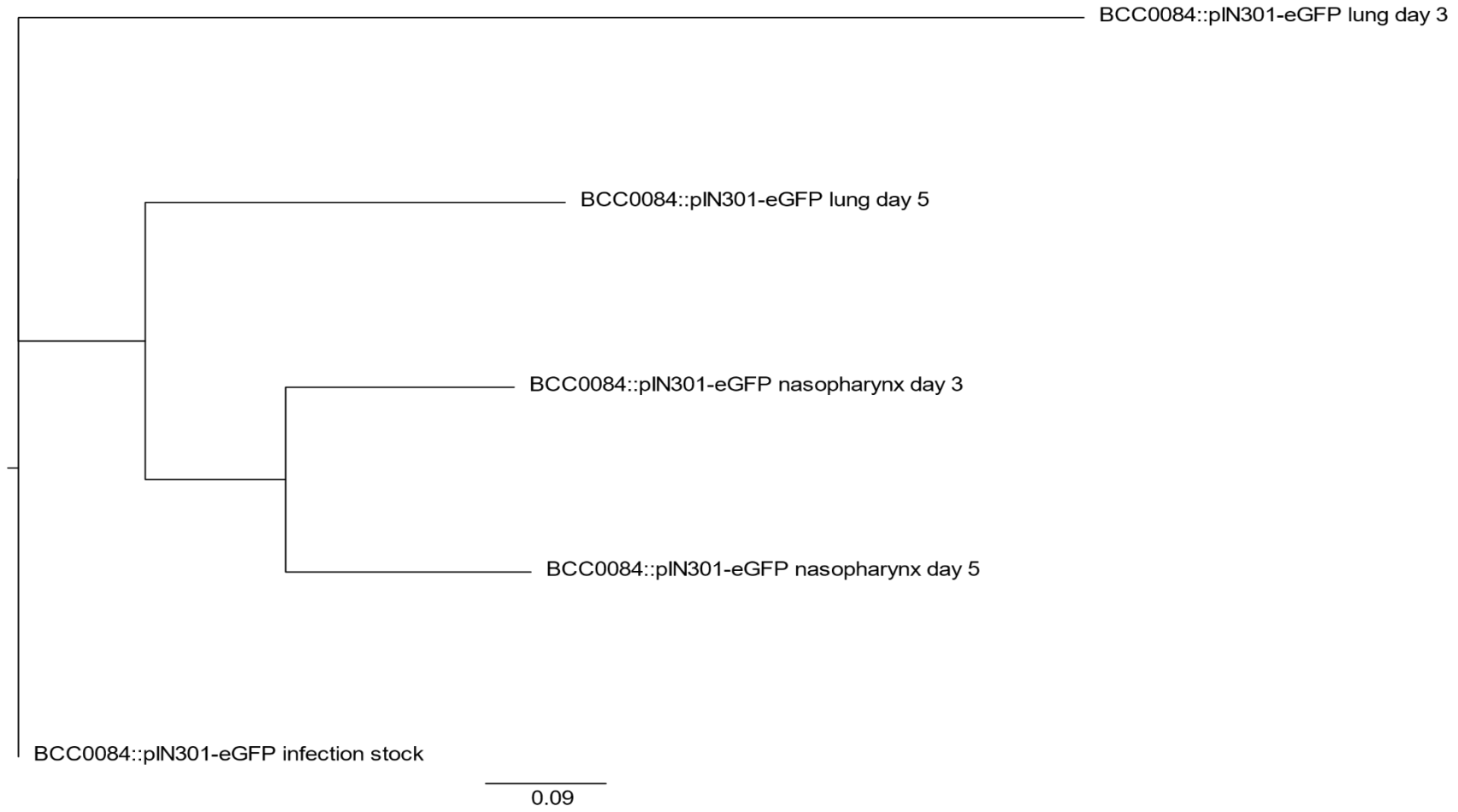


Figure 39 | Core SNP phylogeny of BCC0084::pIN301-eGFP isolates from the murine infection models. FastTree was used to create the phylogeny after filtering for recombination using Gubbins. Phylogeny used 37 core genes for creation.

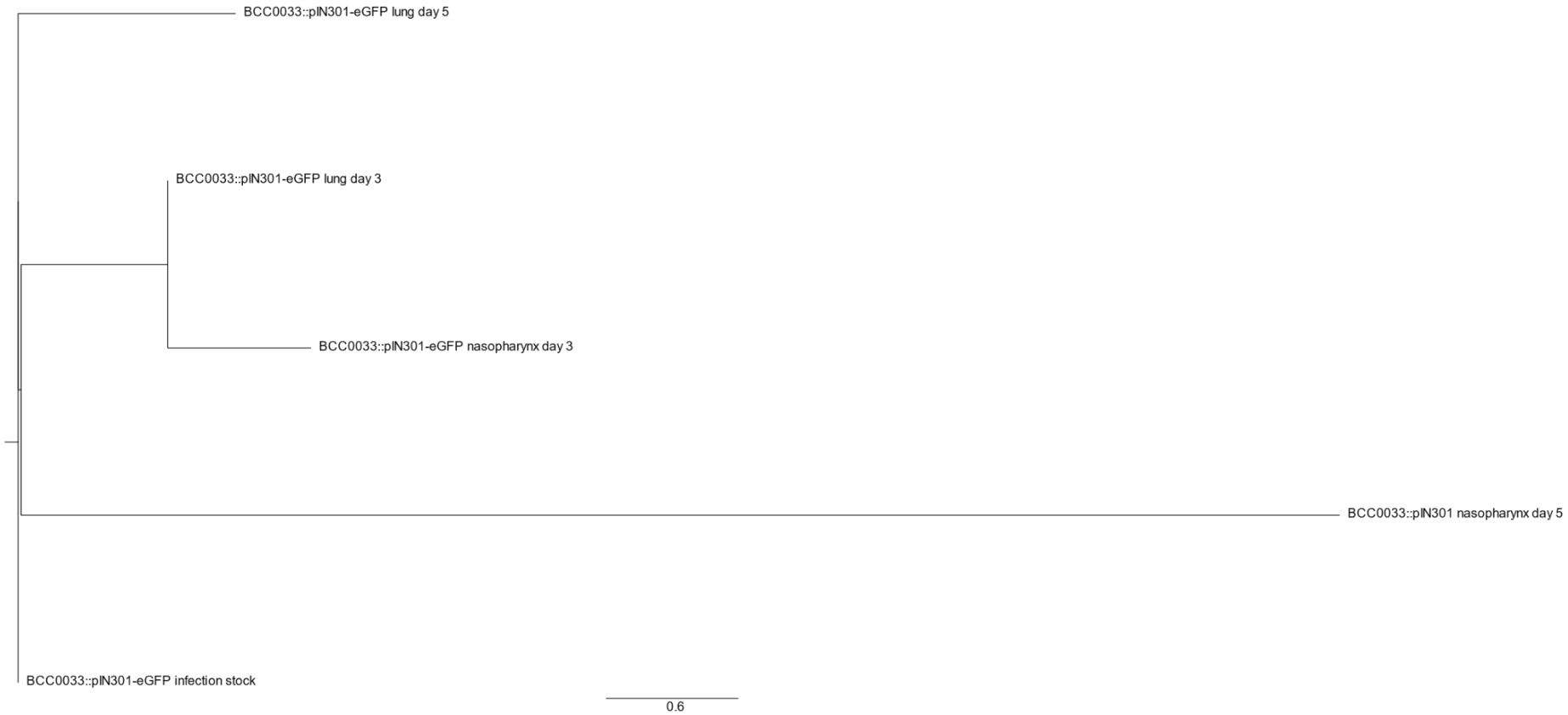


Figure 40 | Core SNP phylogeny of BCC0033::pIN301-eGFP isolates from the murine infection models. FastTree was used to create the phylogeny after filtering for recombination using Gubbins. Phylogeny used 20 core genes for creation.

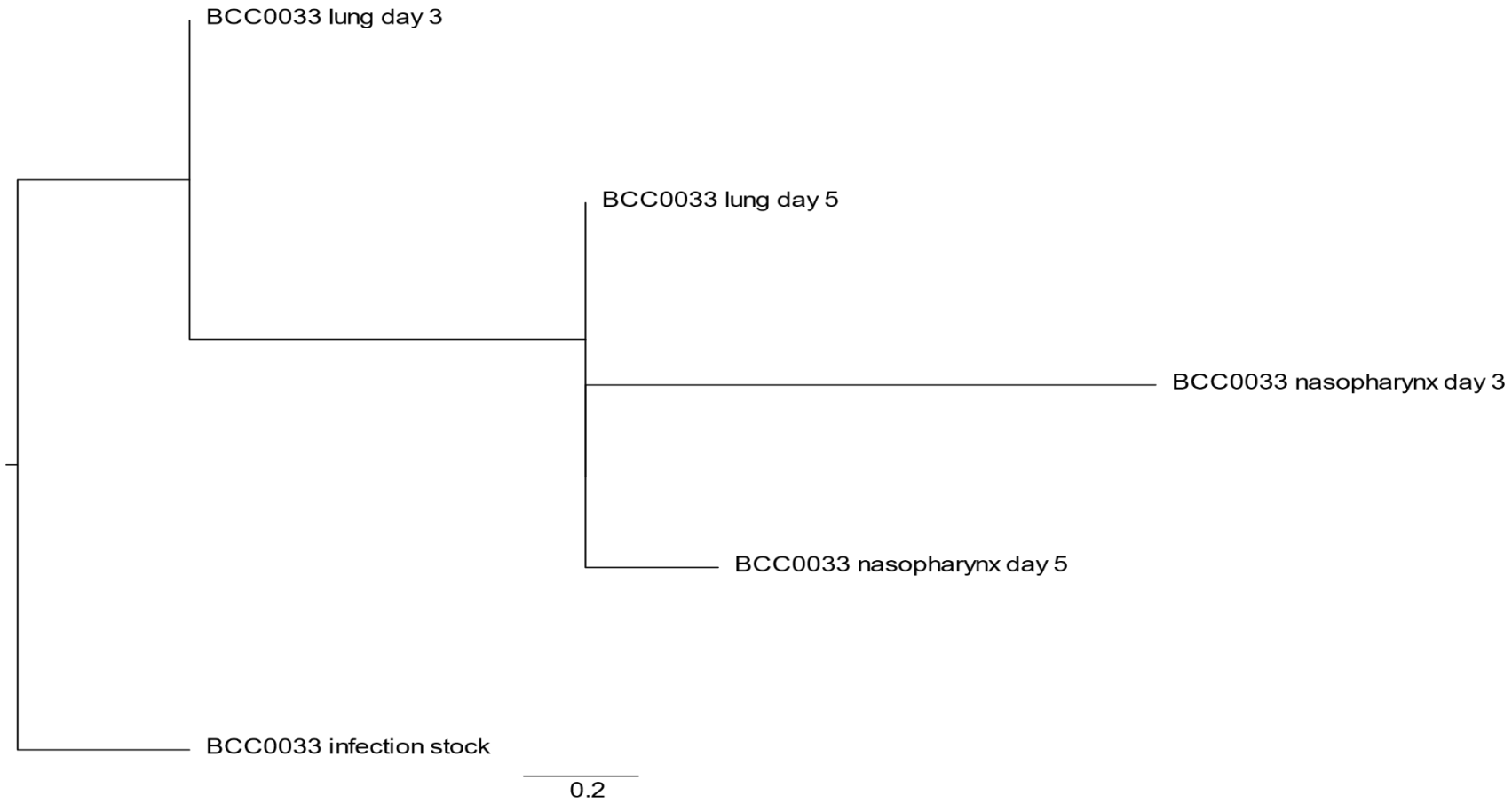


Figure 41 | Core SNP phylogeny of BCC0033 WT (parent) isolates from the murine infection models. FastTree was used to create the phylogeny after filtering for recombination using Gubbins. Phylogeny used 12 core genes for creation.

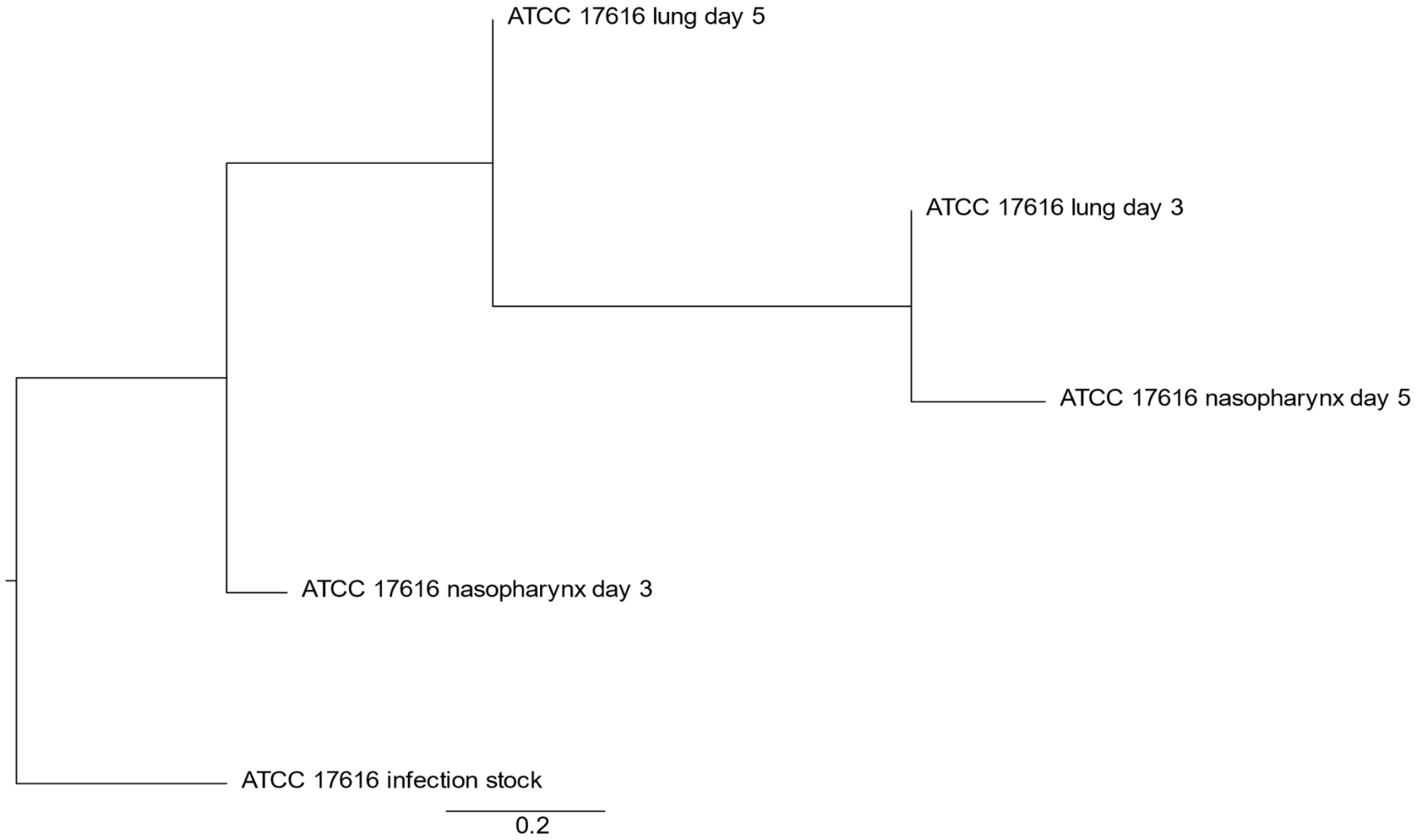


Figure 42 | Core SNP phylogeny of ATCC 17616 WT (parent) isolates from the murine infection models. FastTree was used to create the phylogeny after filtering for recombination using Gubbins. Phylogeny used 57 core genes for creation.

4.2.13.2 Identification of variants between the *B. multivorans* infection stocks and mouse infection isolates

The re-sequenced *B. multivorans* day 3 and 5 lung and nasopharynx isolates (n = 16) were then investigated for variants against the original infection stock (n = 4) using Snippy. This used the un-scaffolded short-read sequences (Section 4.2.11). The number of variants was variable amongst the *B. multivorans* model infection strains. Table 30 shows the full variant details for each strain. In all, there were 242 variants across all the strains examined. These were comprised of 190 SNPs, 35 complex mutations, 11 insertions and 6 deletions. Both BCC0033 WT (lineage 2b) and BCC0033::pIN301-eGFP had the smallest number of SNPs overall, in both infection site and day. ATCC 17616 WT (lineage 2a) had the greatest number of variants when comparing the strains.

There were comparable numbers of synonymous and non-synonymous variants across the board, with 34 and 38 respectively. For the non-synonymous mutations, there were 4 conservative in frame insertions, 4 stop lost and splice region variants, 26 missense variants, and 4 disruptive in frame deletions. Synonymous mutations are also known do not necessarily affect the overall phenotype. However, nonsynonymous affect the protein sequences, and consequently adjust bacterial phenotype. This means that the 38 nonsynonymous variants may affect chronic colonisation.

There was a relatively even split of variants observed overall in the lung and nasopharynx, with totals of 115 and 127 variants, respectively. In the lung, there were 92 SNPs, 15 complex mutations, 6 insertions and 3 deletions overall. Similarly, there were 99 SNPs, 20 complex mutations, 5 insertions and 3 deletions total in the nasopharynx. When looking into days, the split remained even for the total variants in the lung, with 57 variants on day 3 and 58 variants on day 5. The split in variant types reflected a similar pattern of similarity. A larger overall SNP variant difference was observed in the nasopharynx compared to the lung. There was a total of 53 variants on day 3, which increased to 74 on day 5. In the nasopharynx, the larger number of variants on day 5 came from SNPs in the majority. There was a difference of 21 SNPs between days 3 and 5 of nasopharynx infection (Table 30).

Table 30 | Total variants found in the *B. multivorans* strains after 3- and 5- days post-infection of the murine models. Number of variants were grouped by day and location. Details of the variant type (SNP, complex, insertion and deletion) are shown.

Strain	Location	Day	Total variants	SNPs	Complex	Insertions	Deletions
BCC0033::pIN301-eGFP	Lung	3	7	6	1	0	0
		5	14	11	2	1	0
	Nasopharynx	3	6	5	1	0	0
		5	6	5	1	0	0
BCC0033 WT	Lung	3	4	4	0	0	0
		5	7	5	2	0	0
	Nasopharynx	3	8	7	1	0	0
		5	8	6	2	0	0
BCC0084::pIN301-eGFP	Lung	3	19	12	5	1	1
		5	17	12	3	2	0
	Nasopharynx	3	16	13	3	0	0
		5	20	16	4	0	0
ATCC 17616 WT	Lung	3	27	24	1	1	1
		5	20	17	1	1	1
	Nasopharynx	3	23	14	4	3	2
		5	40	33	4	2	1

4.2.13.3 Identification of potential adaptive polymorphisms

To identify variants which may be linked to niche adaptation and chronic infection, the results were screened for mutations which occurred in both days 3 and 5 of the lung or nasopharynx. This was performed separately for each *B. multivorans* infection strain. The number of variants found in each *B. multivorans* strain differed in both overall number and numbers of commonly associated variants. There were 29 overall adaptive variants in ATCC 17616 WT (13 found in the nasopharynx and 16 in the lungs). Of this, there were 6 variants sites (4 SNPs, 1 deletion and 1 insertion) common between both infection (Appendix Table 47). Two of the common variants had unknown effects, but the other 4 were determined. There was 1 disruptive in frame deletion, 1 conservative in frame insertion, 1 synonymous mutation and 1 stop lost and splice mutation (Appendix Table 47). BCC0084::pIN301-eGFP (lineage 1) had less adaptive variants than ATCC 17616 (lineage 2b), with 19 in total (12 in the nasopharynx and 7 in the lungs). There were 7 common variants in BCC0084::pIN301-eGFP, comprised of 2 complex mutations and 5 SNPs, all with unknown effects (Appendix Table 48).

Both BCC0033 WT and BCC0033::pIN301-eGFP strains had less adaptive variants in the infection isolate genomes compared to the infection stocks. There were no common variants in BCC0033::pIN301-eGFP (Appendix Table 49) and 1 common SNP in BCC0033 WT (Appendix Table 50). Both strains had 6 variants overall. BCC0033::pIN301-eGFP (Appendix Table 49) had 5 variants

in the nasopharynx and 1 in the lung, whereas BCC0033 WT had 4 nasopharynx and 2 lung variants (Appendix Table 50).

4.3 Discussion

This work has highlighted key genomic features within the *B. multivorans* strains which may contribute to their chronic colonization and persistence in the CF lung. This work has expanded the understanding of similarities and differences between the two *B. multivorans* lineages, with a focus on the complete model CF strains. A foundation has also been built for comparison of genotype (Chapters ***3*** and ***4***) to the respective phenotypes (Chapter ***5***). The comparative analysis was also able to develop lineage specific probes for *B. multivorans* to aid their identification and future studies on the epidemiological significance of these genomic groupings.

4.3.1 The genomic content of the *B. multivorans* strains is variable

An *in-silico* approach was used to predict the genomic virulence and pathogenicity features of *B. multivorans*, which highlighted many genes associated with both the BCC, in particular *B. cenocepacia*, and *B. pseudomallei.* In future, manual constitution of a BCC virulence and pathogenicity database would be advantageous to identify *B. pseudomallei* gene homologs. This would ensure that any genes with partial matches to the *B. pseudomallei* virulence genes have an increased chance of detection in the developed custom database, reducing the risk of false positives/negatives. Having published virulence databases are useful as it provides a quick and efficient tool for bacterial virulence factor screening and identification. The VFDB is particularly good as it contains gene information for various clinically significant bacteria.

Only one link was made between genomic lineage and presence-absence of a gene. In 24 of the lineage 1 *B. multivorans* genomes, there was an absence of gene *pmlI/bspI1*. Previous research has shown that the absence of this quorum-sensing component can decrease another *B. pseudomallei* virulence in mice (Valade *et al.* 2004). This gene complex has similarities to the *cepI* and *cepR* genes in the BCC (Valade *et al.* 2004), whereby full function has been associated with heightened *B. cenocepacia* infection in CF (Sokol *et al.* 2003b). The BCC CepIR two-component quorum regulatory system is also linked to regulation of biofilm formation, swarming motility (Huber *et al.* 2001), and protease production (Lewenza *et al.* 1999). This suggests that lineage 1 strains may be less virulent than lineage 2 *B. multivorans.* However, all the model CF *B. multivorans,* including BCC0084 from lineage 1, harbour this virulence gene, and all colonised the mouse lung infection model in this study. Therefore, continued investigation would be required to determine the link between *pmlI/bspI1* in *B. multivorans* and virulence with the use of knock-out model strain mutants and virulence assays.

Gene *cdpA*, found in 100% of the *B. multivorans* strains, has previously been shown to regulate both swimming motility and protease activity of *B. cenocepacia* K56-2 in CF sputum (Kumar *et al.* 2018).

Protease production and swimming motility in non-BCC *B. pseudomallei* has also been shown to be affected by *cdpA* (Lee *et al.* 2010). This finding is interesting as the *B. multivorans* strains studied in this thesis were protease negative, and not all which harboured the *cdpA* gene had the ability to swim on agar (Section 5.2.3). The results confirms that phenotype regulation in different BCC species, even associated with the same gene systems, can be highly variable.

Interestingly, all the predicted virulence and pathogenicity genes in the *B. multivorans* model strains were only found on replicons C1 and C2. The 3rd replicon (C3) of the BCC bacteria has previously been described and shown to functionally act as a virulence mega plasmid (Agnoli *et al.* 2014). Weaknesses in the virulence databases are highlighted by the inability to predict putative virulence genes on the C3 replicon, also enhanced by the number of hypothetical genes on this replicon. Further systematic investigation is required for the role of replicon C3 in *B. multivorans* chronic infection. This could be performed through deletion of C3 (Agnoli *et al.* 2014) and studying the model strains in this study using murine infection models.

4.3.2 Low Oxygen Locus

Sass *et al.* (2013) previously described the low oxygen activated (*lxa*) locus in *B. cenocepacia,* whereby its requirement for prolonged anaerobic growth was key. This work has initially pinpointed the *lxa* genes found in the *B. multivorans* strain panel, illustrating how much of the locus is harboured by each strain. Whilst there was no clear-cut finding from the panel, it can be said that a greater number of lineage 2 strains harboured a higher number of *lxa* genes than lineage 1. This included the two model strains BCC1272 and ATCC 17616 from lineage 2a. In contrast, lineage 2b strain BCC0033 and lineage 1 BCC0084 model strains had very little of the locus genes. Therefore, it can be hypothesised that BCC1272 and ATCC 17616 may better survive oxygen-deprived stress conditions in both the changing environment as well as the CF lung. Not only this, but a larger proportion of *lxa* locus genes may also favour attachment to host cells as shown in a recent study by Cullen *et al.* (2018).

From this initial understanding, a collection of *B. multivorans* strains with either low or high amounts of *lxa* genes can be used for further analysis; *B. multivorans lxa* mutants could also be constructed. Such analysis should include viability cultures in an anaerobic chamber or use of oil to create an anaerobic state for growth analysis in the Bioscreen C instrument. This can then be compared to *B. cenocepacia* to draw conclusions around the similarities and differences under hypoxic growth in the two bacterial species.

4.3.3 Lineage-specific genes have been identified for novel PCR probes

There have already been great advances made in PCR design of BCC pathogens, for identification and differentiation in clinical samples (Mahenthiralingam *et al.* 2000; McDowell *et al.* 2001). This work has taken that one step further for *B. multivorans*, developing PCR probes for each genomic lineage to

enable further understanding of their clinical relevance. The primers were designed based on 283 *B. multivorans* genomes, using a gene presence-absence approach. Agnoli *et al.* (2014) showed loss of the 3rd replicon has been observed in BCC bacteria (3.64% studied). Therefore, it was crucial to design the *B. multivorans* lineage-specific PCR primers against either replicon C1 or C2, to ensure genomic stability.

These PCR primers will help with the accurate identification of *B. multivorans* in samples to the lineage level. Differentiation from all members of the BCC remains to be fully determined but were proven to clearly distinguish from *B. cenocepacia* as the other dominant CF BCC pathogen. Identification of other BCC pathogens have been successful directly from sputum (Campbell *et al.* 1995; McDowell *et al.* 2001). Drevínek *et al.* (2002) identified PCR primers which could detect BCC bacteria directly from CF sputum. However, the practical use of these lineage-specific *B. multivorans* primers direct from sputum, in a clinical setting, is yet to be determined.

4.3.4 *B. multivorans* model strains can be genetically manipulated

Due to the natural intrinsic resistance of the BCC to multiple commonly used antibiotics, the use of antibiotic selections for genetic manipulation is limited. This currently encompasses three useful antibiotics (Trimethoprim, Chloramphenicol, and Tetracycline) whereby the modified BCC bacterium has an increased resistance (Shastri *et al.* 2017). Therefore, trimethoprim (Tm) and chloramphenicol (cm) were used as the selection antibiotic for the pMLBAD and the fluorescent reporter plasmids (pIN301-eGFP and pIN233-mCherry), respectively, in this work.

All *B. multivorans* model strains had the ability to uptake pMLBAD through transconjugation and pIN301-eGFP into their genomes via electroporation. However, ATCC 17616 and BCC1272 (both lineage 2a) were incapable of transformation with pIN233-mCherry plasmid via electroporation. It is important for model strains to have the ability to be genetically manipulated. Not only does genetic manipulation allow for understanding gene functions (Xu and Zhang 2016), but also for co-culture and infection modelling. Even though two of the model strains could not uptake the mCherry plasmid, uptake of the other two plasmids into these genomes showed that the ability to genetically modify these bacteria was possible. Therefore, all strains are deemed suitable to be model CF *B. multivorans.* The pIN301-eGFP probes will prove useful in later studies of tissues, looking into the dissemination of bacterial infection.

4.3.5 Functional annotation and classification of the *B. multivorans* model strains in chronic infection

The CDS of the *B. multivorans* model strains were classified into COG categories to help understand genomic functions of the whole genomes and replicons. All four *B. multivorans* model strains consistently had the same top 3 COG groups in replicons C1 and C2: “transcription” (K), “inorganic ion transport and metabolism” (P), and “amino acid transport and metabolism” (E). Transcription CDS

may have both increased synonymous mutation rates (Zhou *et al.* 2020) and increased ability to adapt to varying environments (Peeters *et al.* 2017). However, whilst these 4 *B. multivorans* strains appear to be enriched with (K) COGs, Peeters *et al.* (2017) has previously shown that the abundance is significantly lower than that of related *B. cenocepacia.* Whilst COG comparisons to other members of the BCC have not been made in this work, it poses the question whether there may be differences in adaptation between the *B. multivorans* strains. Overall, as BCC0084 (lineage 1) has a lower number of (K) CDS compared to the other strains, suggesting that this strain may be less adaptable to the CF lung. However, it proved proficient as a short-term mouse lung infection.

4.3.6 Murine modelling of phenotypic and genotypic differences between the model strains

Overall, there were no clear similarities in SNP variants between the model strains and the course of murine infection over the 5-day period. Each infection was unique, presenting different core SNP phylogenies, genomic variants, viability burden, and effects on mouse pain and weight. *B. multivorans* ATCC 17616 (lineage 2a) had the most adaptive polymorphic variants amongst the strains, further strengthening the idea that the strain can readily change under positive selection (Zhou *et al.* 2020).

Using core SNPs, phylogenomic diversity can be measured more accurately for closely related, or sequential, organisms. BCC0084::pIN301-eGFP (lineage 1) had the closest relationship between the infection stock and the respective infection isolates. This correlates with the short branching of lineage 1 *B. multivorans* in the core gene phylogeny when compared to lineage 2 (Section 3.2.5). This could indicate the nature of lineage 1 to have a more closely conserved genome compared to the plasticity of lineage 2 strain genomes. However, SNP rates in the CF lung for diverging BCC lineages, with only around 2.1-2.4 SNPs per year (Lood *et al.* 2021). The limited modelling in this thesis does come with its caveats, due to the short-term nature. Understanding the long-term evolution of the *B. multivorans* model strains in greater detail, with experimentation over months or years, would provide much greater insight into adaptive traits associated with chronic infection.

4.4 Conclusions

Following the genome analysis of the *B. multivorans* strains, the following conclusions can be made:

1) The genomic content of the *B. multivorans* strains is variable, and in terms of major content differences, is not linked to genomic lineage
2) Two novel diagnostic PCR primers have been developed, targeting the specific *B. multivorans* lineage for use in a clinical setting
3) COG analysis showed similarities in the CDS found on the C1 and C2 replicons of the *B. multivorans* strains
4) Murine infection did not show any correlations between the *B. multivorans* strains and evolution of genomic variants during short-term respiratory infection

5 Phenotypic Characterization of a *B. multivorans* strain collection and its interactions with other members of the CF microbiota

5.1 Introduction

The term 'phenotype' is used for characterizing observable and expressed traits in an organism, such as bacterial cell growth (Bochner 2009). Multiple phenotypes of CF pathogens have been previously studied in *P. aeruginosa* (Deligianni *et al.* 2010; Mayer-Hamblett *et al.* 2014; Clark *et al.* 2015), *S. aureus* (Kahl 2010; Bernardy *et al.* 2020), *B. cenocepacia* (Lee *et al.* 2017) and *B. multivorans* (Silva *et al.* 2011). Understanding bacterial phenotypes helps to make observations in the population at single time points and longitudinally over time to potentially identify virulence-associated characteristics which alter during chronic infection. In parallel with bacteria genotype, phenotypic characteristics change and evolve over time. These alterations in pathogen phenotype are often complex, and show heterogeneity with the bacterial genotype, such as in *Stenotrophamonas maltophilia* CF infection (Esposito *et al.* 2017). This is where a wide range of phenotypes are expressed but are not necessarily correlated to genotype, such as growth, biofilm formation, and mutation frequency (Esposito *et al.* 2017). For chronic CF lung infection of *P. aeruginosa*, studies have shown a loss of motility (Mahenthiralingam *et al.* 1994) and decreases in virulence factor production like T3SS and QS, adaptation to reside as a biofilm, and an increase in antimicrobial resistance (Winstanley *et al.* 2016).

All organisms undergo cell replication and growth, although the understanding of how the environment affects bacterial growth continues to expand in knowledge. Bacteria have been found to adjust their metabolism to confer the optimal growth rate for the environment (Bosdriesz *et al.* 2015), making bacterial growth rates highly variable depending on their niche. Often, a fast-growing microorganism would be referred to as possessing a positive fitness trait (La Rosa *et al.* 2021), as there is the ability to outcompete another bacterium in the same ecological niche. However, adaptation to a slow growing phenotype may also be advantageous in certain circumstances, like chronic colonization and infection (Yang *et al.* 2008). The slow growing nature of certain bacteria has been shown to confer other associated phenotypes such as an increase in antimicrobial resistance and biofilm production (Bartell *et al.* 2019; La Rosa *et al.* 2021).

Bacterial biofilms are an important phenotype of many pathogenic bacteria. These are aggregated cells and other matrix material which are formed over time as bacteria grow. They start with a surface attachment, then cells, DNA and substances, such as extracellular polymers (EPS), build upon this to create an autogenous matrix (Vestby *et al.* 2020). Around 40 to 80% of all bacterial cells having biofilm-forming ability (Flemming and Wuertz 2019), contributing to their adaptation and survival in harsh environmental conditions (Vestby *et al.* 2020). Biofilms also heighten the ability for the bacterial colony to respond to stress and antibiotics, specifically increasing their antimicrobial resistance (Aiyer *et al.*

2021). Biofilms frequently enable persistent infections which are hard to eradicate (Gordon *et al.* 2019). It is well understood that biofilms are the driving force for chronic infection, as opposed to planktonic growth (Martin *et al.* 2021). Within CF infection, microbial biofilms are known to form within the thick and sticky mucus in the lungs and airways.

Another pathogenic phenotype associated with bacterial CF pathogens is motility. There are several types of bacterial motility: swimming, swarming, twitching, gliding, and floating (Jarrell and McBride 2008). The former 3 motility types listed are the ones most commonly seen in CF-associated bacteria such as *P. aeruginosa* (O'May and Tufenkji 2011). The driving force of swimming motility is ultimately down to one or more functional flagellar propelling the bacteria (Ha *et al.* 2014). Swarming motility enables solid surface migration of bacteria though use of both a rotational flagella and surfactant combined (Kearns 2010). QS is also important for regulation of bacterial swarming, mediating multiple bacterial cell interactions (Daniels *et al.* 2004). BCC species often possess motility traits, but have been shown to harbour different numbers of flagella (Mahenthiralingam and Vandamme 2005). Whilst swimming motility has been shown to attenuate in *P. aeruginosa* chronic CF infection isolates (Mahenthiralingam *et al.* 1994), BCC infections do not appear to have the same longitudinal loss of motility from current data (Zlosnik *et al.* 2014).

Extracellular polysaccharide (EPS) production in BCC species are also a phenotype associated with chronic CF infection. Not only do these substances enhance bacterial biofilms, but related CF-pathogen *P. aeruginosa* has been shown switch phenotype from nonmucoid to mucoid during infection, contributing to morbidity and mortality (Zlosnik *et al.* 2008). EPS in the BCC is less well understood than *P. aeruginosa*. Knowledge currently extends to the understanding of EPS types excreted by BCC bacteria. The most common is cepacin, whereby many clinically associated BCC have been shown to produce this EPS type (Ferreira *et al.* 2010). A paper by Zlosnik *et al.* (2011) showed that there was a correlation between less mucoidy/EPS-producing BCC strains, including *B. multivorans*, and decreased CF lung function. However, further research into EPS and CF understanding in *B. multivorans* remains open for investigation.

Antimicrobial resistance (AMR) is a huge threat to public health, especially today with resistance continuing to rise. Whilst not everyone agrees with the predictions (de Kraker *et al.* 2016) it has been suggested that 10 million people will die in 2050 due to AMR-related complications (Murray 2022). Not all of these will be attributed to CF infection, but it is still a large cause for concern in associated AMR bacteria. One major issue we face is the way in which CF infections are treated. Novel therapies are a rarity, meaning that widely used clinical antibiotics are given as standard and continuous treatment (Conly and Johnston 2005). Unfortunately, the AMR is a phenotype possessed by multiple CF pathogens beyond *Burkholderia*. This includes bacterial species and strains with both multi- and pan-drug resistances such as *P. aeruginosa* (Clark *et al.* 2018), conferring resistance to the most common

antibiotic therapies. This makes infection eradication exceedingly difficult. Therefore, understanding antibiotic susceptibility patterns of the specific bacterial strain present during infection is crucial for improved patient care.

Proteases are a key pathogenicity phenotype for bacteria, accelerating disease in a number of situations (Culp and Wright 2017). Multiple pathogenic bacteria secrete proteases which have been shown to interact and interfere with host defence mechanisms (Voynow *et al.* 2008). Extracellular proteases have been observed in several CF-associated bacteria. This includes *P. aeruginosa*, which has previously been shown to secrete two major proteases, an alkaline protease (AprA) and an elastase (LasB) (Mateu-Borrás *et al.* 2021). Sun *et al.* (2020) showed that the LasB extracellular protease of *P. aeruginosa* actively affected the pulmonary system in CF by inducing lung inflammation and damage through activation of interleukin-1ß. The BCC bacteria have also been shown to excrete extracellular proteases. For example, metalloprotease ZmpA was found in *B. cenocepacia* and *B. cepacia*, yet the 7 *B. multivorans* strains tested in this study were negative for this zinc metalloprotease (Gingues *et al.* 2005). Whether this lack of protease production is a feature of *B. multivorans* as a species remains to be determined.

Chronic CF infections with only one bacterial species are unusual in the context of multiple diseases. There have been studies which look at the community analysis and co-occurrence of different CF bacteria in the lung, with evidence that bacterial diversity decreases over time (Zhao *et al.* 2012; Narayanamurthy *et al.* 2017). In younger individuals, community diversity is observed to be the greatest, up until around 10 years of age. This then decreases, before hitting a plateau around 25-years of age (Coburn *et al.* 2015). By understanding the lung microbiota diversity at different time points, disease state (Cuthbertson *et al.* 2020) and clinical outcome can be estimated. For instance, *B. cenocepacia* has been shown to escalate rapid lung function decline, significantly greater than when an individual is infected with *P. aeruginosa* or *B. multivorans* (Courtney *et al.* 2004). *B. multivorans* has been shown to co-infect individuals with other common CF bacteria. A longitudinal study of *B. multivorans* has shown its simultaneous presence with 3 other CF pathogens (*S. aureus, H. influenzae* and *P. aeruginosa*) for several years before microbial diversity decreased and *B. multivorans* was reported alone (Silva *et al.* 2016). In adults with CF, *Burkholderia* are frequently the dominant pathogens in the lung microbiome along with *P. aeruginosa, Ralstonia, Achromobacter, Stenotrophamonas,* and other Gram-negative AMR species (Flight *et al.* 2015).

Infections with multiple microorganisms not only makes patient treatment difficult in terms of antibiotic resistance patterns but can also affect pathogen phenotype alterations based on polymicrobial interactions. Previous work showed an increase in biofilm formation and inflammatory response when *P. aeruginosa* and *B. cenocepacia* were modelled together in a murine model (Bragonzi *et al.* 2012). *P. aeruginosa* has often thought to be a dominant pathogen in CF co-infections, inhibiting the growth of

other strains (Costello *et al.* 2014). However, Costello *et al.* (2014) also showed that a CF isolate of *B. multivorans* has also previously been shown to suppress the growth of co-infecting pathogens such as *B. cenocepacia* and *P. apista*. Interesting observations have also been found when modelling co-infection of *B. multivorans* with *P. aeruginosa*. Both bacterial species can inhibit the growth of one another, to different extents, reducing the overall growth in a co-infection (Costello *et al.* 2014). In long-term CF lung infection, where other microorganisms were present, *B. multivorans* has been shown to exhibit traits linked to chronic colonization. These include reductions in motility and O-antigen repeats, but an increase in biofilm formation and antibiotic resistance (Silva *et al.* 2016). However, it remains unclear whether these phenotypic alterations are driven by interactions during *B. multivorans* co-infection (Silva *et al.* 2016).

5.1.1 *Aims and objectives*

The overall aim of this Chapter was to investigate the phenotype of a *B. multivorans* sub-strain collection representative of the diversity within the genome sequenced isolates (see Chapter 3) and identify potential phenotypic differences (Chapter 5) between the genomic lineages. The following objectives were investigated:

1) To determine the growth kinetics of the *B. multivorans* collection in liquid media
2) To evaluate the motility of the *B. multivorans* strain collection, specifically evaluating the swimming and swarming phenotype
3) To examine the antibiotic susceptibility patterns of the *B. multivorans* strains using a disk diffusion-based assay
4) To characterise the biofilm forming ability of *B. multivorans* across the strain collection
5) To screen the *B. multivorans* collection for protease production, exopolysaccharide production and colony morphology type on a range of different growth media
6) To define phenotypic alterations that occur when *B. multivorans* is interacted with a secondary CF pathogen

Hypothesis 1: The *B. multivorans* phenotype will be highly variable and not associated with genomic lineage

Hypothesis 2: The *B. multivorans* strains will change certain phenotypic properties when interacted with a secondary CF lung pathogen

5.2 Results

5.2.1 *Colony morphology of the B. multivorans strains*

The colony morphology was examined for 89 *B. multivorans* strains, including the 73 sequenced in Section 3.2.1. The colonies were assessed for shape, size, and surface texture after 24-hours of growth

on TSA (Table 31). Eleven (12.4%) of the *B. multivorans* strains had an irregular shape, with the other 88% of strains presenting a round morphology. Forty (44.9%) of the *B. multivorans* strains had a medium colony size. The remaining 49 (55.1%) of strains had small colonies. The appearance of all *B. multivorans* strains was shiny (Table 31).

Congo red medium (CRM) was used to look at morphological differences in colonial variants and cell surface differences through dye uptake (n = 89) (Chung *et al.* 2003). Agar plates were examined by eye after 48-hours of growth. Two strains (BCC1367 and BCC0059). Twelve strains (13.5%) had pink colonies upon examination. BCC0006, BCC0225, BCC1368 and BCC0901 all had red colonies (n = 4; 4.5% of strains). It has previously been suggested that an increase in EPS production may reduce the ability to uptake the congo red dye (Chung *et al.* 2003), but that was not the case in this study. Instead, the difference in uptake of congo red dye may indicate other cell surface differences (Chung *et al.* 2003). The strains with red centres all encompassed CF isolates. All other *B. multivorans* strains (79.8%) had pink with orange-red centres. There was no correlation between colony morphology and genomic lineages.

Table 31 | Colony morphology of the *B. multivorans* stains (*n* = 89). The strains and their associated shape, size, and appearance on TSA are noted. TSA with 0.01% (w/v) congo red agar was also used to assess colony morphology of the strains.

Lineage	Strain	Morphology				
		Shape	Size	Appearance	Congo Red Agar	EPS production (YEM agar)
1	BCC0006	Round	Small	Shiny	Red	-
	BCC0009	Irregular	Medium	Shiny	Pink to pink with orange centre	++++
	BCC0010	Round	Small	Shiny	Pink with red centre	++++
	BCC0067	Round	Small	Shiny	Pink with red centre	++++
	BCC0080	Round	Medium	Shiny	Pink with red centre	++
	BCC0084	Irregular	Medium	Shiny	Pink with orange centre	+++
	BCC0093	Round	Small	Shiny	Pink with red centre	++++
	BCC0101	Round	Medium	Shiny	Pink with red centre	++
	BCC0115	Irregular	Medium	Shiny	Pink with orange centre	++++
	BCC0141	Round	Small	Shiny	pink with red centre	++++
	BCC0149	Round	Small	Shiny	Pink with orange-red centre	++++
	BCC0175	Round	Medium	Shiny	Pink with orange-red centre	++
	BCC0292	Round	Medium	Shiny	Pink with red centre	++++
	BCC0293	Round	Small	Shiny	Pink with orange centre	++
	BCC0303	Round	Small	Shiny	Pink	+++
	BCC0321	Round	Small	Shiny	Pink	+
	BCC0375	Round	Medium	Shiny	Pink with red centre	++
	BCC0702	Round	Small	Shiny	Pink with orange centre	+++
	BCC0704	Round	Small	Shiny	Pink with orange centre	+++
	BCC0737	Round	Small	Shiny	Pink with red centre	+++
	BCC0814	Round	Medium	Shiny	Pink to pink with orange centre	+++
	BCC0865	Round	Small	Shiny	Pink with red centre	++
	BCC0904	Round	Small	Shiny	Pink	+
	BCC0915	Round	Small	Shiny	Pink with orange-red centre	+++
	BCC0921	Round	Small	Shiny	Pink with orange centre	+
	BCC0962	Round	Small	Shiny	Pink with red centre	++++
	BCC0968	Round	Small	Shiny	Pink with orange-red centre	+++
	BCC1177	Round	Small	Shiny	Pink with orange centre	+++
	BCC1190	Round	Small	Shiny	Pink with red centre	++++
	BCC1367	Round	Small	Shiny	Orange to red	++
	BCC1384	Round	Small	Shiny	Pink	++++
	BCC1385	Irregular	Medium	Shiny	Pink with red centre	++++
2	ATCC17616	Irregular	Medium	Shiny	Pink with red centre	+++
	BCC0005	Round	Medium	Shiny	Pink with red centre	++++
	BCC0008	Round	Small	Shiny	Pink with orange-red centre	++++
	BCC0031	Round	Medium	Shiny	Pink with orange-red centre	++++
	BCC0032	Round	Medium	Shiny	Pink with red centre	++
	BCC0033	Irregular	Medium	Shiny	Pink with red centre	+++
	BCC0037	Round	Medium	Shiny	Pink with orange-red centre	++++
	BCC0043	Round	Small	Shiny	Pink with orange centre	++
	BCC0047	Round	Medium	Shiny	pink	++++
	BCC0050	Round	Medium	Shiny	Pink with orange-red centre	++++
	BCC0059	Round	Small	Shiny	Orange to red	+++

	BCC0065	Round	Medium	Shiny	Pink to pink with orange centre	++
	BCC0066	Round	Medium	Shiny	Pink with red centre	++
	BCC0068	Round	Small	Shiny	pink	-
	BCC0074	Round	Medium	Shiny	Pink with red centre	++
	BCC0075	Round	Small	Shiny	Pink with orange centre	++++
	BCC0079	Round	Medium	Shiny	Pink	++++
	BCC0082	Irregular	Medium	Shiny	Pink with red centre	++
	BCC0087	Round	Medium	Shiny	pink with orange centre	++
	BCC0089	Round	Medium	Shiny	Pink with orange-red centre	++++
	BCC0096	Round	Medium	Shiny	Pink with orange centre	+++
	BCC0099	Round	Medium	Shiny	Pink with red centre	+++
	BCC0102	Round	Small	Shiny	Pink with red centre	++++
	BCC0134	Round	Small	Shiny	Pink with red centre	++
	BCC0181	Round	Small	Shiny	Pink with red centre	++++
	BCC0188	Round	Small	Shiny	Pink with red centre	-
	BCC0225	Irregular	Medium	Shiny	Red	+++
	BCC0241	Round	Medium	Shiny	pink with red centre	++
	BCC0244	Round	Medium	Shiny	Pink with red centre	+
	BCC0246	Round	Medium	Shiny	Pink with red centre	++++
	BCC0247	Round	Medium	Shiny	Pink with red centre	++
	BCC0255	Round	Small	Shiny	Pink with orange centre	++++
	BCC0264	Round	Medium	Shiny	Pink with red centre	+
	BCC0266	Round	Medium	Shiny	pink with red centre	++
	BCC0269	Round	Medium	Shiny	pink with red centre	+++
	BCC0300	Round	Small	Shiny	Pink with orange centre	+++
	BCC0317	Irregular	Medium	Shiny	Pink with orange centre	++++
	BCC0384	Round	Small	Shiny	pink	+
	BCC0470	Round	Medium	Shiny	Pink with red centre	+++
	BCC0493	Round	Small	Shiny	Pink	-
	BCC0497	Round	Small	Shiny	Pink	-
	BCC0533	Irregular	Medium	Shiny	Pink with red centre	++++
	BCC0553	Round	Small	Shiny	Pink with orange-red centre	+
	BCC0583	Round	Small	Shiny	Pink with red centre	++
	BCC0585	Round	Small	Shiny	Pink with red centre	+++
	BCC0585	Round	Small	Shiny	Pink with orange-red centre	+++
	BCC0710	Round	Small	Shiny	Pink with red centre	++
	BCC0729	Irregular	Medium	Shiny	Pink with red centre	+++
	BCC0901	Round	Small	Shiny	Red	+++
	BCC0907	Round	Medium	Shiny	Pink with red centre	+++
	BCC1147	Round	Small	Shiny	Pink with orange centre	++
	BCC1148	Round	Small	Shiny	pink	++
	BCC1185	Round	Small	Shiny	Pink	+++
	BCC1271	Round	Small	Shiny	Pink with orange centre	+++
	BCC1272	Round	Small	Shiny	Pink with red centre	+++
	BCC1421	Round	Small	Shiny	Pink with orange-red centre	++++
Other	BCC1368	Round	Small	Shiny	Red	+

5.2.2 *B. multivorans growth rate analysis*

5.2.2.1 Growth in TSB

Bacterial growth rates were performed for the *B. multivorans* phenotype panel (*n* = 50) (Figure 43) using a Bioscreen C instrument. Analysis was performed in TSB, incubated at 37ºC, over a 48-hour period. Figure 43 illustrates the growth curve for the 50 *B. multivorans* strains examined. The observed maximum growth rate of the *B. multivorans* strains split into two broad groups (Figure 43). Stationary phase was reached by 29 strains in less than 24-hours, whereas 21 strains took over 24-hours to reach the same growth phase (Figure 43).

To define the growth parameters of the *B. multivorans* strains in greater detail, the R Grofit package was used. This modelled the length of lag phase, maximum growth rate and maximum culture density for each strain up to the 48-hour growth point. Data summaries can be found in the Figure 44 box plots. All 50 *B. multivorans* strains reached a maximum culture density of 0.36 $OD_{480\text{-}520nm}$. This was unaffected by the slower growth rate observed in half the strains. The average lag phase was 5.02 hours. There were 5 strains (BCC0303, BCC0269, BCC0493, BCC0921, and BCC1185) which exhibited prolonged lag phases (mean = 11.16 hours). All these strains were found to be above the upper quartile in terms of the length of the lag phase (Figure 44). As for the growth rate, the average for all strains was 0.032 h^{-1}. To identify very slow growing *B. multivorans* strains, the strains which fell below the lower quartile were noted. BCC0032, BCC006, BCC0075, BCC0188, BCC0225, BCC0247, BCC0375, BCC0497, BCC0702, BCC0814 and BCC0865 were all noted as slow growing (mean = 0.018 h^{-1}), comprising 21% of the total strains. The well characterised reference soil strain ATCC 17616 (0.066 h^{-1}) and its isogenic BCC1272 CF strain (0.059 h^{-1}) were on or within the upper quartile in growth rate (Figure 44). This means that ATCC 17616 and BCC1272 represented two of the fast-growing *B. multivorans* strains.

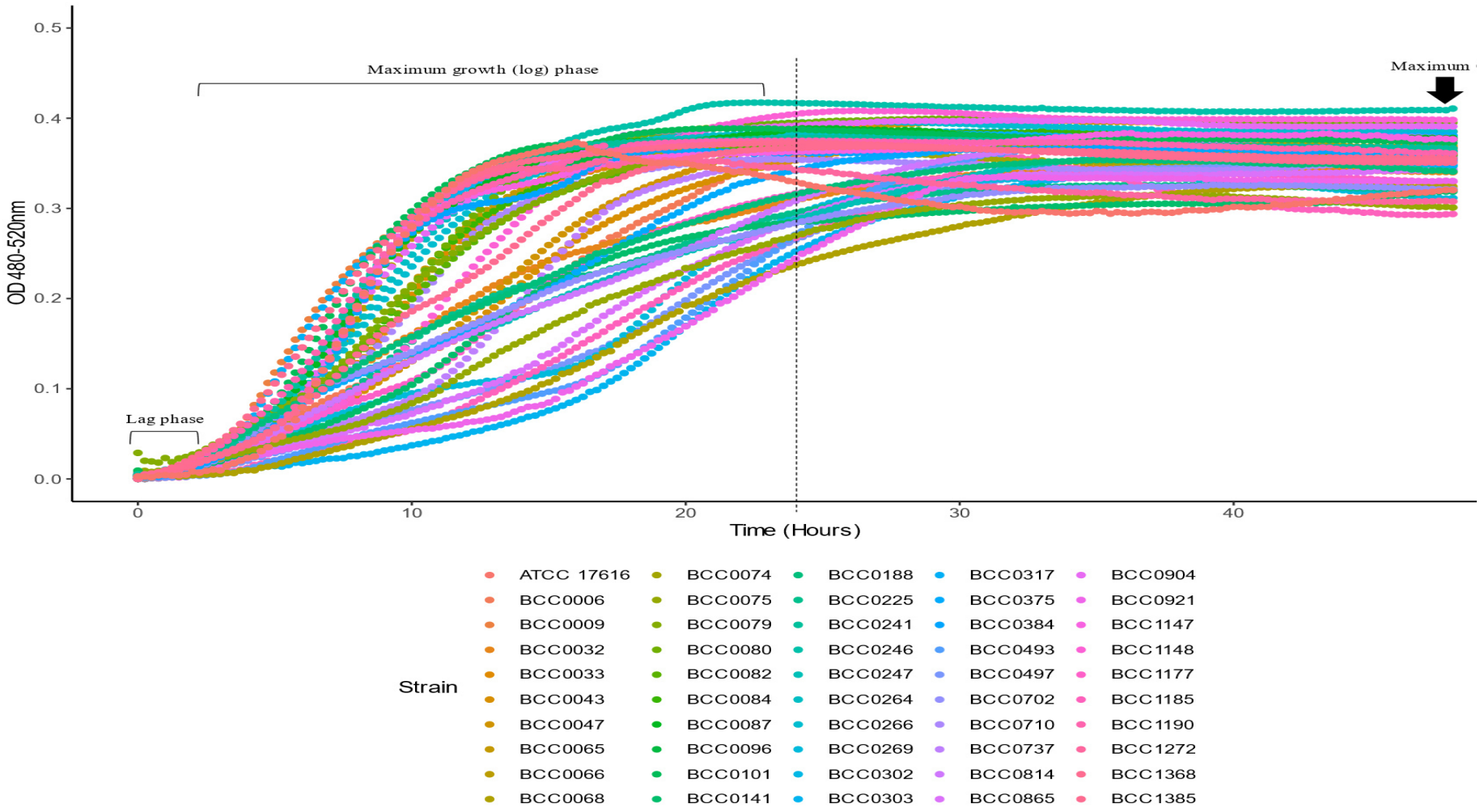
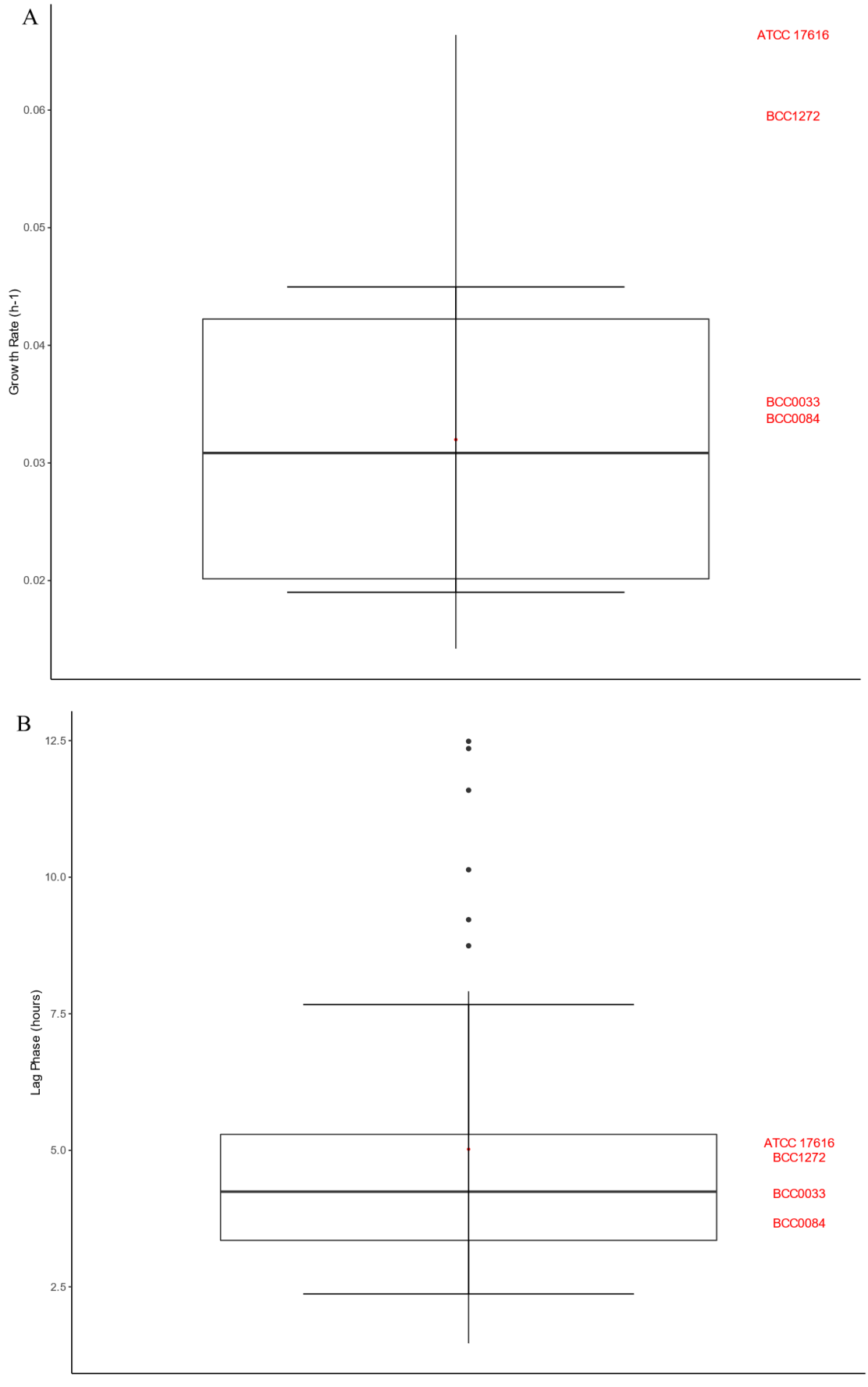


Figure 43 | Growth curves of the *B. multivorans* phenotypic strain panel over a 48-hour period. Growth rates were measured using a Bioscreen C instrument. Cultures were grown in TSB at 37ºC, shaking 10 seconds before each 15-minute reading. ODs for each *B. multivorans* strain (n = 50) were aggregated and the means plotted.


A
Grow th Rate (h-1)
0.06
0.05
0.04
0.03
0.02
ATCC 17616
BCC1272
BCC0033
BCC0084
B
Lag Phase (hours)
12.5
10.0
7.5
5.0
2.5
ATCC 17616
BCC1272
BCC0033
BCC0084

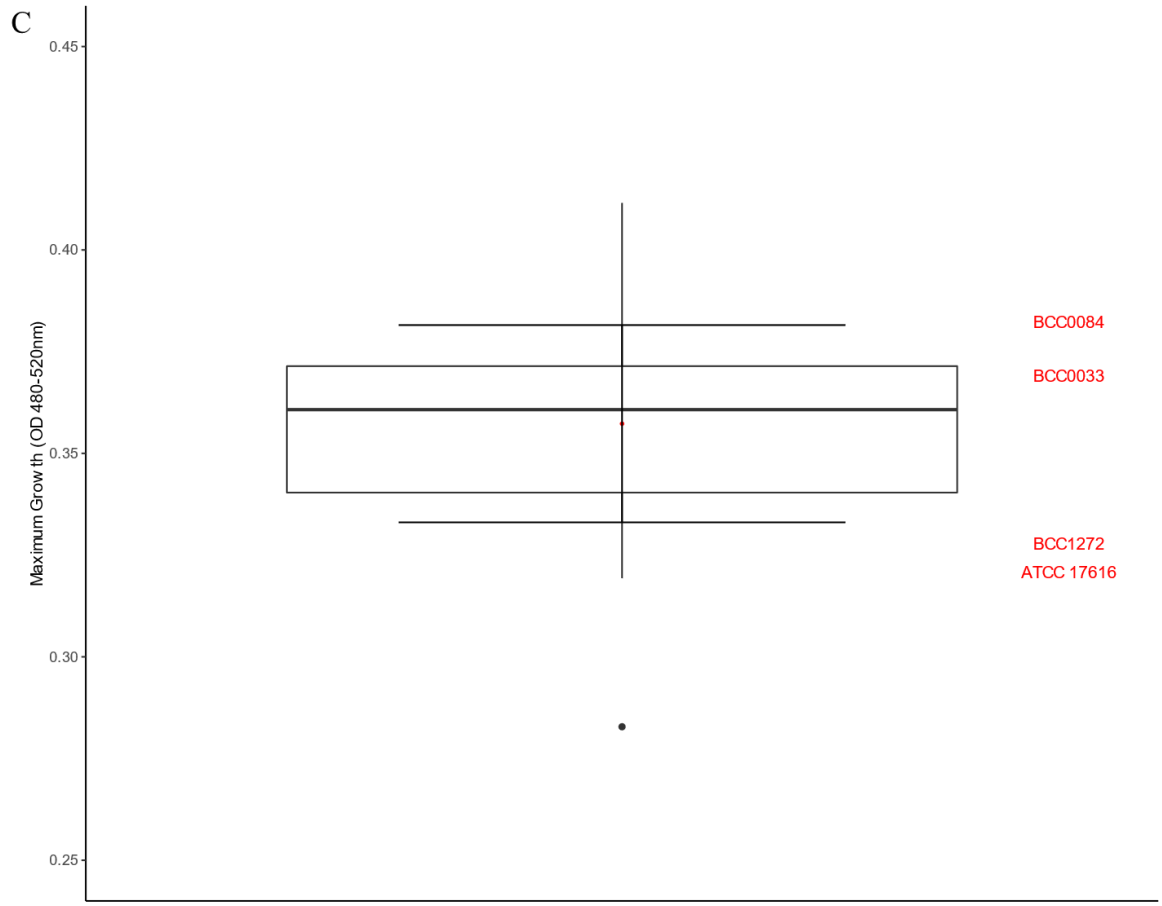


Figure 44 | Growth parameters of the *B. multivorans* strains (*n* = 50) at 48-hours. The box plots have been drawn from data produced in R using the GroFit package: **(A)** growth rate (h^{-1}), **(B)** lag phase (hours), and **(C)** maximum growth ($OD_{480\text{-}520\ nm}$). All box plots are annotated with the model strains on the right-hand side (red) and display the upper quartile, mean, and lower quartile. Outliers have also been added to the plots.

5.2.2.2 Growth in ASM

Artificial sputum medium (ASM) (Kirchner *et al.* 2012) was used in a further experiment as a more biomimetic growth comparison against the conventional nutrient rich TSB. A standardised amount of fresh TSB culture was used as the inoculum for the experiment to ensure the starting point was the same. The ASM growth was only determined for the four selected model strains, ATCC 171616 (lineage 2a), BCC0033 (lineage 2b), BCC0084 (lineage 1), and BCC1272 (lineage 2a) (see Section 3.2.8), as shown in Figure 45. In all four strains, the growth rate (h^{-1}) was faster in TSB than ASM. In *B. multivorans* strains ATCC 17616 ($p = 0.004$) and BCC1272 ($p = 0.005$), the growth rate in ASM was around 0.05 to 0.06 h^{-1} significantly slower than in TSB. For *B. multivorans* strains BCC0033 and BCC0084, there was a smaller difference in growth rate between TSB and ASM. On average, BCC0033 ($p = 0.1$) and BCC0084 ($p = 0.3$) grew 0.015 and 0.012 h^{-1} slower in ASM than TSB, respectively.

The average lag phase (hours) in ASM was 2.0, 2.1, 1.2, and 1.1 slower for *B. multivorans* strains ATCC 17616 ($p = 0.03$), BCC1272 ($p = 0.04$), BCC0033 ($p = 0.07$), and BCC0084 ($p = 0.04$), respectively. The lag phase was between 1.9 and 2.4 hours longer when grown in ASM compared to TSB, depending on the *B. multivorans* strain examined. *B. multivorans* ATCC 17616 had the longest lag phase when grown in ASM, with an average of 4.4 hours. *B. multivorans* BCC0033 had the smallest increase in lag phase growth, with a difference of 1.9 hours when grown in ASM compared to TSB, of all four *B. multivorans* strains tested.

All four strains had a reduction in overall maximum growth (OD) when grown in ASM compared to TSB. All four tested strains reached a maximum $OD_{480\text{-}520nm}$ of between 0.16 and 0.21 when grown in ASM. The greatest difference in overall OD were observed in BCC0033 ($p = 0.05$) and BCC1272 ($p = 0.02$), both with reduction in overall growth by 0.19 $OD_{480\text{-}520nm}$. ATCC 17616 had a reduction of 0.16 $OD_{480\text{-}520nm}$ ($p = 0.06$) and BCC0084 had a reduction of $0.15_{480\text{-}520nm}$ ($p = 0.05$) when grown in ASM.

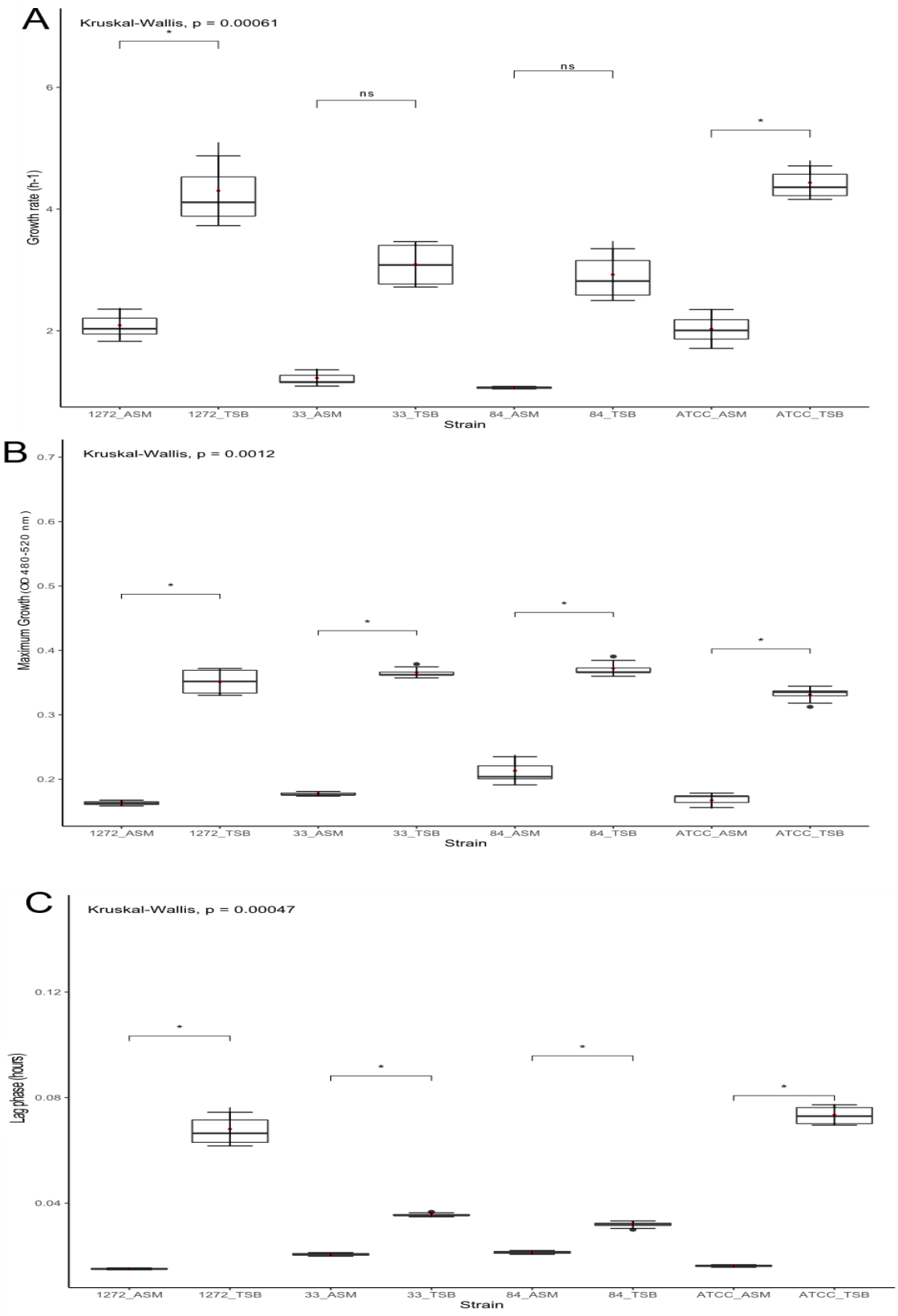


Figure 45 | Growth parameter comparisons of the four *B. multivorans* model strains (*n* = 4) when grown in TSB and ASM. Analysis of results outputted by the R package GroFit. (A) growth rate (h^{-1}), (B) maximum growth ($OD_{480\text{-}520\ nm}$), and (C) lag phase (hours). All box plots are annotated with the model strains on the right-hand side (red) and display the upper quartile, mean, and lower quartile.

Table 32 | Statistical comparison of the *B. multivorans* strains (*n* = 4) grown in TSB compared to ASM. Results are split into growth type from the output of the R GroFit package.

Growth Parameter	Model Strain	Z-Value	*P*-value	*P-adjusted*	Significance
Growth Rate (h -1)	ATCC 17616	-3.3045	0.00048	0.00443962	**
	BCC1272	-3.3842	0.00036	0.00499773	**
	BCC0033	-1.566	0.05867	0.12637448	ns
	BCC0084	-0.6636	0.25349	0.2957333	ns
Lag Phase (hours)	ATCC 17616	-2.5481	0.00542	0.03032946	*
	BCC1272	-2.2296	0.01289	0.03608689	*
	BCC0033	-1.9111	0.028	0.07126948	ns
	BCC0084	-2.2296	0.01289	0.04009655	*
Maximum OD (480-520 nm)	ATCC 17616	-1.9111	0.028	0.07126948	ns
	BCC1272	-2.7073	0.00339	0.01899158	*
	BCC0033	-2.1898	0.01427	0.04994902	*
	BCC0084	-2.1101	0.01742	0.05420753	ns

The viable bacteria were also enumerated prior to and after growth analysis in TSB media. Counts started at 10^6 CFU/ml and increased to 10^7 CFU/ml by 30-hours growth. Viable enumeration was performed for a range of both the slower and faster growing strains. The number of colonies was consistent across the strains by 30-hours, supporting assumption that all strains, irrespective of their growth rate, had reached stationary phase by the 30-hour period.

5.2.2.3 Growth of the genetically manipulated strains

The genetically manipulated pIN301-eGFP and pIN233-mCherry expressing *B. multivorans* strains from Section 4.2.9.2 were grown in TSB. *P. aeruginosa::*pIN233-mCherry was also included in the analysis to ensure its suitability for growth interaction modelling with *B. multivorans* in Section 5.2.9.4. This was to observe whether there were any differences in the growth parameters compared to the wild-type strain. As chloramphenicol (50 µg/ml) antibiotic is used to maintain the pIN301-eGFP and pIN233-mCherry plasmids, the genetically manipulated strains were grown with and without antibiotic present. Overall, no significant difference was observed in the growth parameters of the pIN301-eGFP genetically manipulated strains, grown both with and without antibiotic, compared to the respective wild-type strains (Table 33). BCC0033::pIN233-mCherry grown without antibiotic had a significant difference in the lag phase, being longer by 0.8 hours, compared to BCC0033 wild-type (Table 33). All other pIN233-mCherry strains did not have any significant differences (Table 33).

Table 33 | Statistical comparison of the *B. multivorans* parent strains and the pIN301-eGFP genetically manipulated strains (n = 4).

Growth Parameter	Model Strain (Parent)	Comparison (pIN301-eGFP)	Z-Value	*P*-value	*P-adjusted*	Significance
Growth Rate (h -1)	ATCC 17616	without antibiotic	0.316179	0.375933	0.443064	ns
	ATCC 17616	with antibiotic	1.449154	0.073647	0.151898	ns
	BCC1272	without antibiotic	-0.10539	0.458032	0.472345	ns
	BCC1272	with antibiotic	1.132975	0.128612	0.202105	ns
	BCC0033	without antibiotic	-0.13174	0.447594	0.476472	ns
	BCC0033	with antibiotic	-0.26348	0.396089	0.45863	ns
	BCC0084	without antibiotic	-0.11384	0.454683	0.476335	ns
	BCC0084	with antibiotic	0.081312	0.467597	0.474791	ns
Lag Phase (hours)	ATCC 17616	without antibiotic	-0.97489	0.164809	0.231433	ns
	ATCC 17616	with antibiotic	-1.81803	0.03453	0.094957	ns
	BCC1272	without antibiotic	-1.34376	0.089513	0.151483	ns
	BCC1272	with antibiotic	-2.08151	0.018694	0.056081	ns
	BCC0033	without antibiotic	-1.08028	0.140009	0.200883	ns
	BCC0033	with antibiotic	-1.44915	0.073647	0.138878	ns
	BCC0084	without antibiotic	-0.06505	0.474067	0.488882	ns
	BCC0084	with antibiotic	-0.48787	0.312819	0.382335	ns
Maximum OD (480-520 nm)	ATCC 17616	without antibiotic	0.685054	0.246655	0.33915	ns
	ATCC 17616	with antibiotic	-0.47427	0.317654	0.374378	ns
	BCC1272	without antibiotic	1.923422	0.027214	0.11974	ns
	BCC1272	with antibiotic	0.922189	0.178215	0.280052	ns
	BCC0033	without antibiotic	0.711403	0.238417	0.334799	ns
	BCC0033	with antibiotic	0.843144	0.199574	0.292708	ns
	BCC0084	without antibiotic	-0.58545	0.279123	0.361218	ns
	BCC0084	with antibiotic	0.032525	0.487027	0.494519	ns

Table 34 | Statistical comparison of the *B. multivorans* (*n* = 2) and *P. aeruginosa* LESB58 parent (wild type) (*n* = 1) strains and the pIN233-mCherry genetically manipulated strains.

Growth Type	Model Strain (Parent)	Comparison (pIN233-mCherry)	Z-Value	*P*-value	*P-adjusted*	Significance
Growth Rate (h -1)	LESB58	without antibiotic	0.248214	0.401985	0.438529	ns
	LESB58	with antibiotic	0.347499	0.364108	0.409622	ns
	BCC0033	without antibiotic	-0.49643	0.309796	0.39831	ns
	BCC0033	with antibiotic	-1.58857	0.056079	0.126178	ns
	BCC0084	without antibiotic	-0.58377	0.279686	0.372915	ns
	BCC0084	with antibiotic	-0.21228	0.415944	0.440411	ns
Lag Phase (hours)	LESB58	without antibiotic	1.489283	0.068206	0.175388	ns
	LESB58	with antibiotic	1.191426	0.116743	0.221198	ns
	BCC0033	without antibiotic	2.233924	0.012744	0.045879	*
	BCC0033	with antibiotic	2.084996	0.018535	0.055605	ns
	BCC0084	without antibiotic	-0.27862	0.390269	0.453215	ns
	BCC0084	with antibiotic	-0.4909	0.311748	0.400819	ns
Maximum OD (480-520 nm)	LESB58	without antibiotic	0.496428	0.309796	0.619593	ns
	LESB58	with antibiotic	0.446785	0.327515	0.589527	ns
	BCC0033	without antibiotic	-0.94321	0.172786	0.691145	ns
	BCC0033	with antibiotic	0.049643	0.480204	0.523858	ns
	BCC0084	without antibiotic	-0.87566	0.190607	0.686186	ns
	BCC0084	with antibiotic	-0.39803	0.345305	0.540477	ns

5.2.3 *Swimming and swarming motilities*

P. aeruginosa has been shown to completely 'switch off' its motility as a major pathogenesis factor in CF (Mahenthiralingam *et al.* 1994; Wolfgang *et al.* 2004). The *B. multivorans* strains (n = 49) were assessed for swimming and swarming motility, to identify whether any of these strains exhibit low or non-motility. Swimming motility was tested on 0.3% LB agar and swarming motility was tested on 0.5% LB and BSM-G agars. The two agar types used in the swarming motility assay were to allow for comparison of an enriched and sparse nutrient medium, with the latter encouraging the *B. multivorans* strains to swarm further. Figure 46 illustrates the motility and how it was categorised for the *B. multivorans* strains.

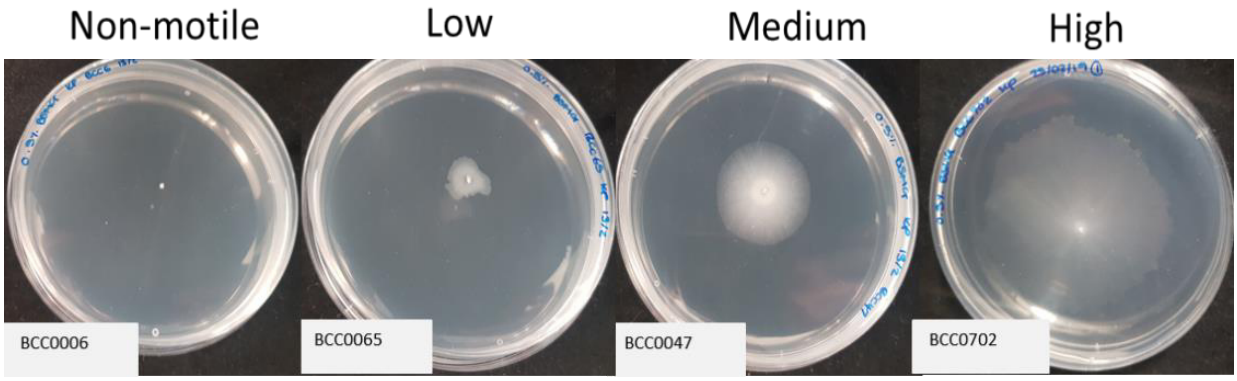


Figure 46 | Representative images for the *B. multivorans* swimming and swarming motility categories. Results were recorded after 24-hours incubation at 37°C. The figure shows a representation of results on BSM-G agar.

The findings showed that both motility types were highly variable amongst the strains. No statistical significance was observed in motility when comparing the *B. multivorans* genomic lineages. LB swimming, LB swarming, and BSM-G swarming motilities presented mean diameters of 25.08 mm, 9.86 mm, and 22.76 mm, respectively. Most *B. multivorans* (n = 47; 96%) showed motility on at least one type of agar. When comparing motility types, 87% of *B. multivorans* strains had the ability to swim whereas only 80% could swarm. The swarming ability varied when tested on LB versus BSM-G agar, however, no statistical difference between the growth media was found (non-parametric Kruskal-Wallis test with Chi-squared value of 1.82; p = 0.18). Full motility measurements can be found in Appendix Table 51. Interestingly, two strains (BCC0068 and BCC0904) were completely nonmotile. BCC0068 was a CF isolate whilst BCC0904 was a non-CF infection isolate.

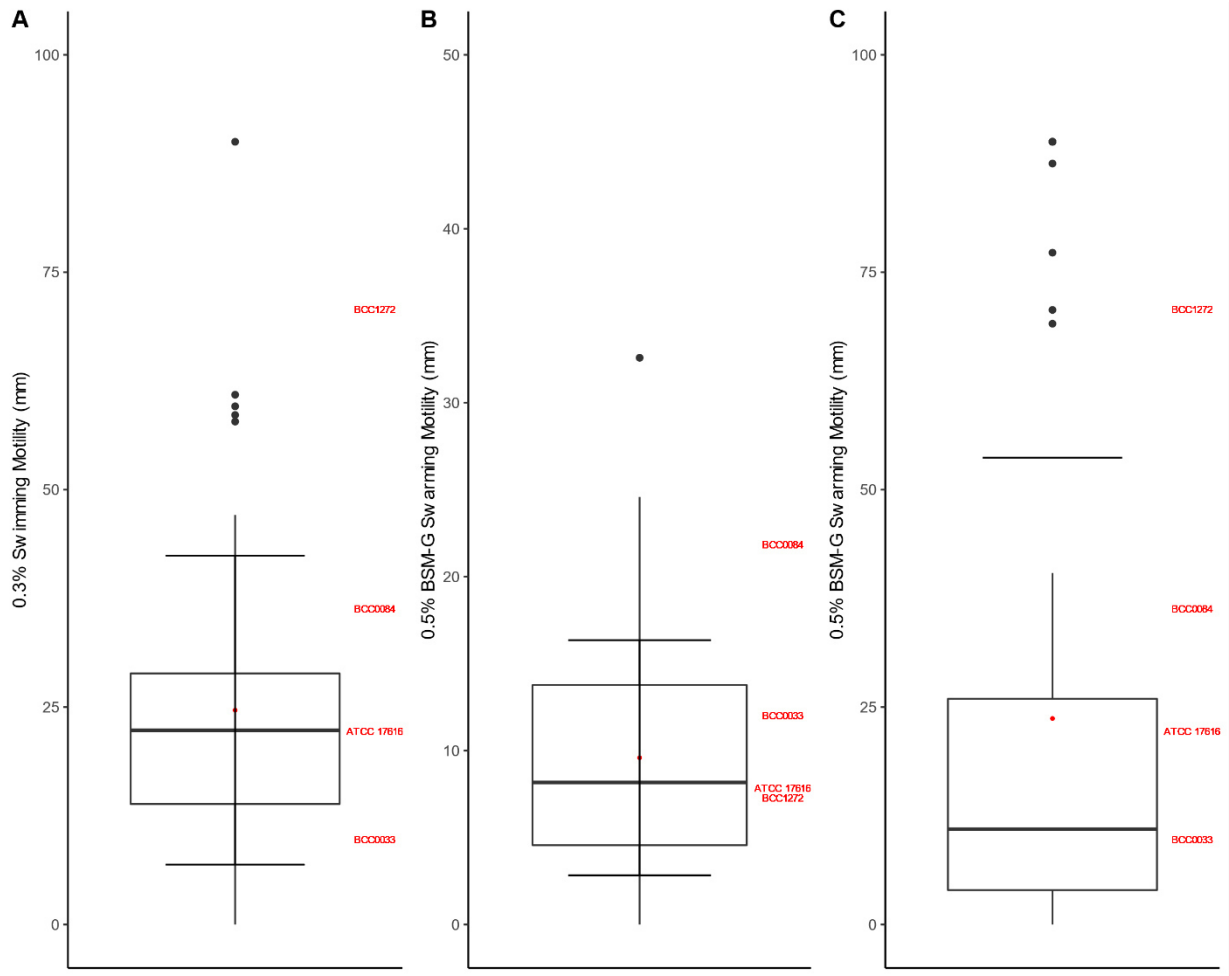


Figure 47 | Swimming and swarming motility in the *B. multivorans* strain panel after 24-hours incubation at 37°C. (A) Swimming motility was performed on 0.3% LB agar (n = 49). Swarming motility was performed on (B) 0.5% LB agar (n = 48) and (C) 0.5% BSM-G agar (n = 44). Model *B. multivorans* strains (n = 4) are annotated in red on the right-hand side of each box plot.

5.2.4 *Antibiotic susceptibility patterns*

Antibiotic screening was performed on the *B. multivorans* strains (n = 45) using the agar disc diffusion assay. This screened colistin sulphate (100 µg), chloramphenicol (50 µg), ampicillin (25 µg), tetracycline (100 µg), streptomycin (25 µg), nitrofurantoin (50 µg), nalidixic acid (30 µg), and kanamycin (30 µg). Figure 48 showed the mean susceptibility diameters for each of the tested strains. There were no EUCAST breakpoints for BCC bacteria, as well as no suitable PK/PD values for the antibiotics tested. Therefore, the mean susceptibilities have been used for analysis.

All *B. multivorans* strains, irrespective of lineage, were resistant to nitrofurantoin but susceptible to tetracycline. On average, the strains had a zone of inhibition of 19.6 mm when in the presence of tetracycline. Whilst resistance to tetracycline has been observed in *B. cenocepacia* and *B. ubonensis*, *B. multivorans* lacks the *tetA* resistance gene (Somprasong *et al.* 2021). Chloramphenicol susceptibility was shown in every lineage 2 strain (n = 29), and in 15 (93.8%) of the lineage 1 (n = 16) strains, with

an average zone of inhibition at 20.4 mm and 20.8 mm, respectively. BCC0375 was the only *B. multivorans* strain to possess resistance to chloramphenicol

All lineage 1 strains ($n = 16$) were resistant to colistin sulphate. In lineage 2, there were 5 strains (17.2%) which showed sensitivity to colistin sulphate. ATCC 17616, BCC0225, BCC1272, BCC0269, and BCC0479 all had between 9 mm and 15 mm zones of inhibition. Resistance to colistin (polymyxin E) was expected as this is an intrinsic phenotypic trait of BCC bacteria, including *B. multivorans* (Henry *et al.* 1999; Malott *et al.* 2012). These five strains which showed some sensitivity to the antibiotic are therefore of interest, where hopanoids and membrane stability may be a factor in increased susceptibility (Malott *et al.* 2012).

All *B. multivorans,* except two strains (BCC1190 and BCC0074), were resistant to ampicillin (Figure 48). As susceptible strains, BCC1190 had a clearing zone of 23.5 mm and BCC0074 of 17 mm when exposed to ampicillin. *B. multivorans*, specifically the ATCC 17616 strain, has previously been shown to harbour the PenA enzyme, conferring resistance to ß-lactam antibiotics (Poirel *et al.* 2009). However, the role of the *penA* gene in *B. multivorans* is not well understood (Rhodes and Schweizer 2016). Whilst the *penA* gene was found through Abricate in strains ATCC 17616 and BCC1271 (Section 1.1.1.3), it was absent in all other strains tested in this thesis. Therefore, the ampicillin resistance seen in most strains is likely to be through another mechanism.

Four strains (8.7% of total strains: BCC0375, BCC0264, BCC0710 and BCC1185) were resistant to nalidixic acid (Figure 48). All remaining *B. multivorans* strains were susceptible to nalidixic acid (mean zone of clearing was 13.1 mm). The only strain susceptible to streptomycin was BCC0384, with a zone of inhibition of 11.7 mm. Kanamycin resistance was observed in 5 strains (BCC0904, BCC0068, BCC0264, BCC0075 and BCC1185), whereas the remaining strains had a mean susceptibility of 14.0 mm (Figure 48).

Three *B. cenocepacia* strains ($n = 3$) (J2315, K56-2, and BCC0019) were subjected to antibiotic susceptibility testing using the same antibiotic rings as the *B. multivorans* strains. Antibiotic resistance was observed in all 3 *B. cenocepacia* strains when exposed to colistin sulphate, ampicillin, streptomycin, nitrofurantoin, and kanamycin. Tetracycline sensitivity was observed in *B. cenocepacia* K56-2 (mean zone of inhibition was 16 mm), but BCC0019 and J2315 were resistant. All 3 *B. cenocepacia* strains were sensitive to nalidixic acid, ranging from 15.7 mm to 20.3 mm zones of inhibition. In all, this correlated to the results observed in the *B. multivorans* strains, with the exception that all *B. multivorans* were sensitive to tetracycline (Figure 48).

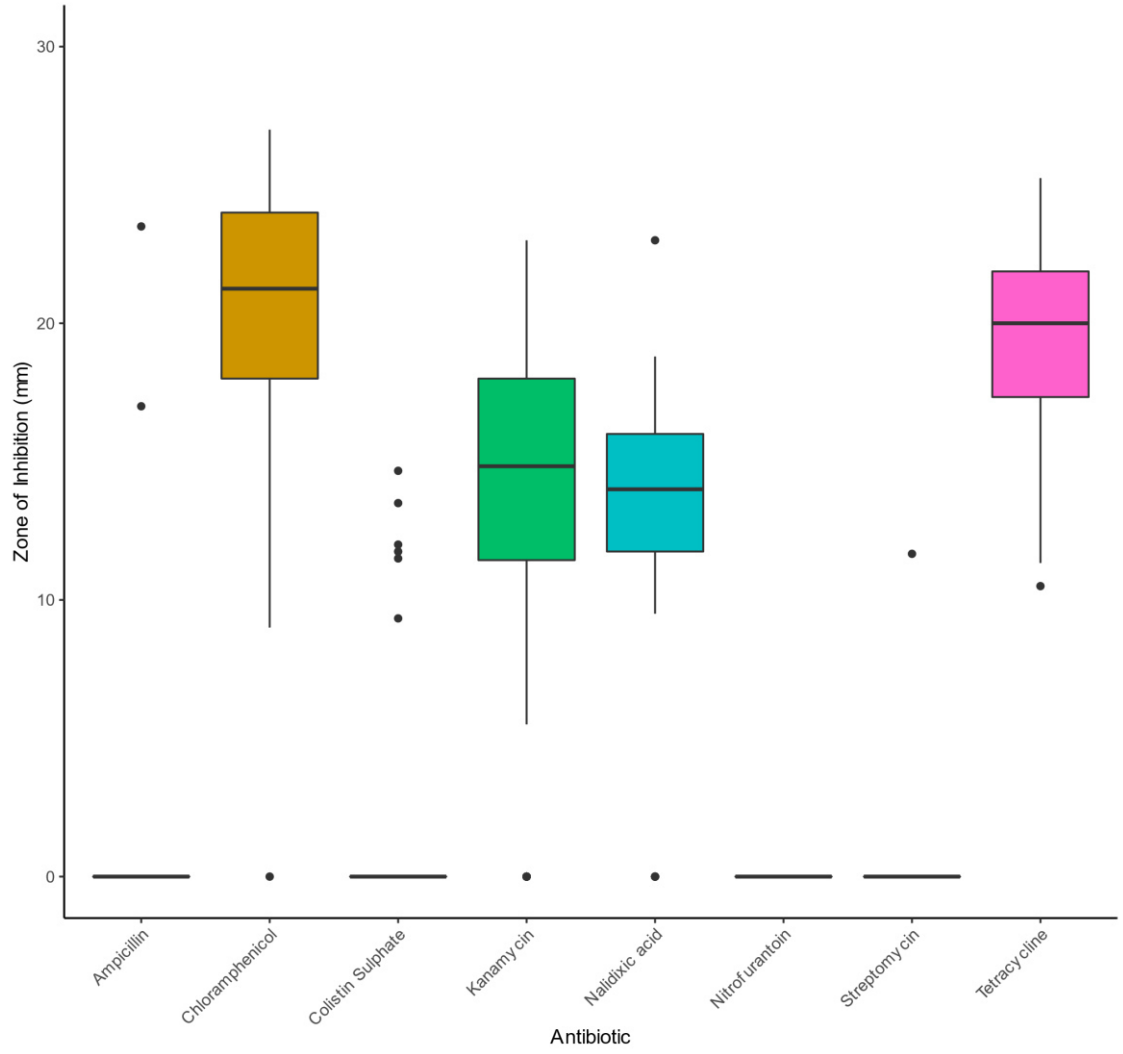


Figure 48 | Antibiotic susceptibility of the *B. multivorans* strain panel (*n* = 45) using the M26-NCE MASTRING disc diffusion assay. Box plots show the mean, lower quartile, and upper quartile zone of clearing (mm) for each antibiotic. Outliers are also shown.

5.2.5 *Biofilm formation of planktonic B. multivorans*

Biofilm production is a pathogenic phenotype of multiple CF pathogens. Crystal violet staining assays (O'Toole 2011) were performed to assess the ability of the *B. multivorans* strains (*n* = 49) to create biofilms in a 96-well plate over the course of 24-hours. *B. multivorans* ATCC 17616 was used as the positive control strain because of its known good biofilm-producing ability (Caraher *et al.* 2007), which presented a mean biofilm of 0.18 OD_{600nm} (Figure 49). Caraher *et al.* (2007) also highlighted a weak biofilm former, BCC0010 (also known as strain number C1962). This was used as the low biofilm control in the experiment, which formed a mean biofilm of 0.017 OD_{600nm} (Figure 49). The overall mean biofilm produced by the *B. multivorans* strains was 0.076 OD_{600nm}. Strains which produced a biofilm greater than ATCC 17616 were noted as high biofilm forming. BCC0047, BCC1147 and BCC1272 (6% of strains tested) all produced a high amount of biofilm (Figure 49). In contrast, 14% of the strains

tested (n = 7) were noted as weak biofilm formers (BCC0068, BCC0075, BCC0264, BCC0493, BCC0814, BCC0865, BCC0921), with an average OD_{600nm} of less than BCC0010. Intriguingly, not only was BCC0068 nonmotile, but it was also unable to form measurable biofilms in this PVC plate binding crystal violet assay.

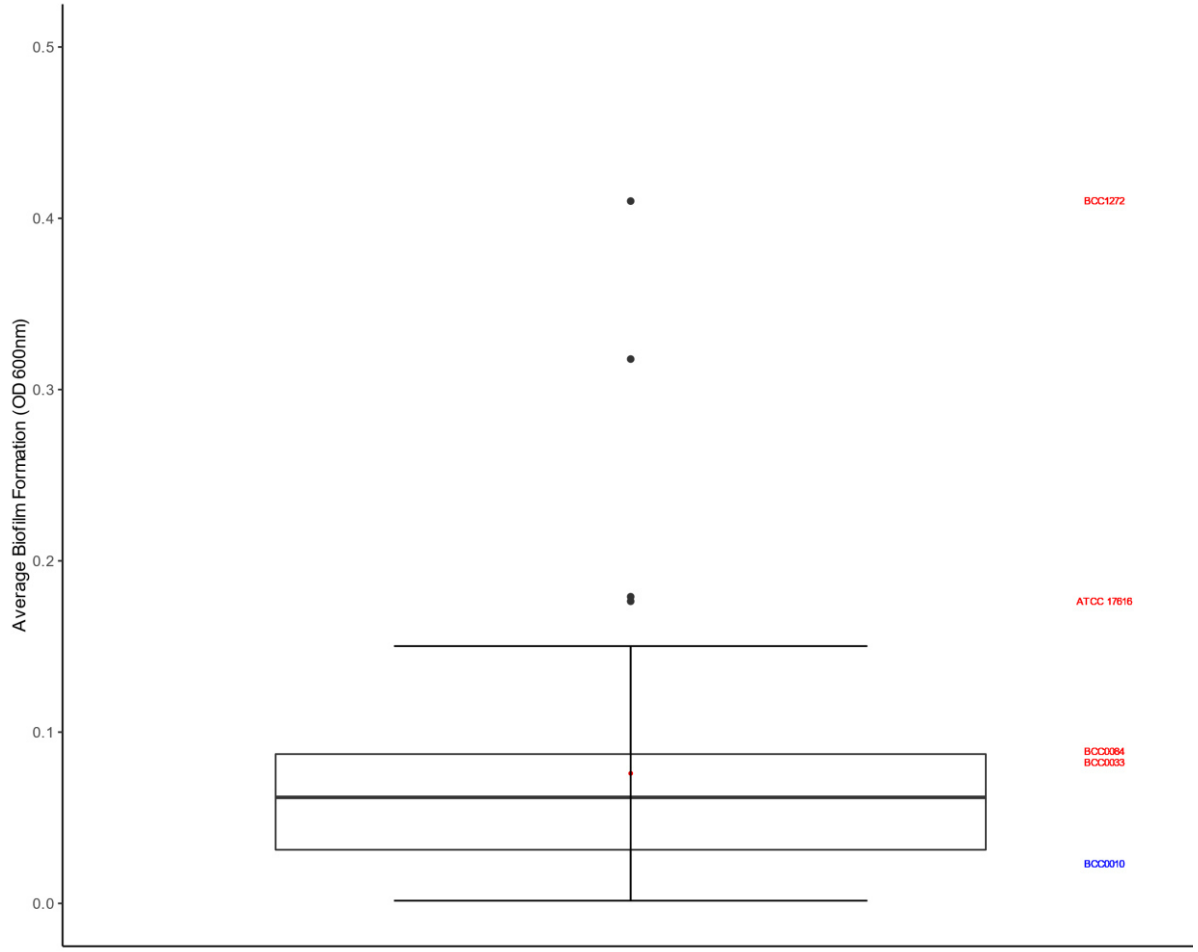


Figure 49 | Biofilm formation of the *B. multivorans* strain panel (n = 49) after 24-hours. Among of biofilm was assessed using the crystal violet assay (O'Toole 2011), reading the results using a plate reader at 600_{nm}. Biofilm controls were *B. multivorans* ATCC 17616 for the 'high' former and BCC0010 (blue) for the 'low' control. Model CF *B. multivorans* strains are noted in red on the right-hand side of the box plot. Box plots show the mean, lower quartile, and upper quartile for biofilm staining. Outliers are also shown.

To mimic the CF lung environment more closely, ASM was used as the growth medium for a follow up experiment. This used a selection of *B. multivorans* strains (n = 11) from the initial TSB experiment chosen to span both high and low biofilm-formers. This included all four model strains (BCC0033, BCC0084 and BCC1272) and both controls (ATCC 17616 and BCC0010). In 6 of 11 *B. multivorans* strains (54.5%), the average biofilm formation in ASM was greater than that in TSB (Figure 50). BCC1272 was one of the strains where TSB media produced more biofilm formation (Figure 50). BCC0010 (p = 0.03) and BCC0033 (p = 0.02) also had statistically greater biofilm production in TSB, with an average greater biofilm of 0.12 and 0.09 $OD_{600nm,}$ respectively. All other strains had no

significance in terms of biofilm formation when grown in TSB or ASM. This suggested that TSB is a suitable medium for mass biofilm screening of *B. multivorans* strains.

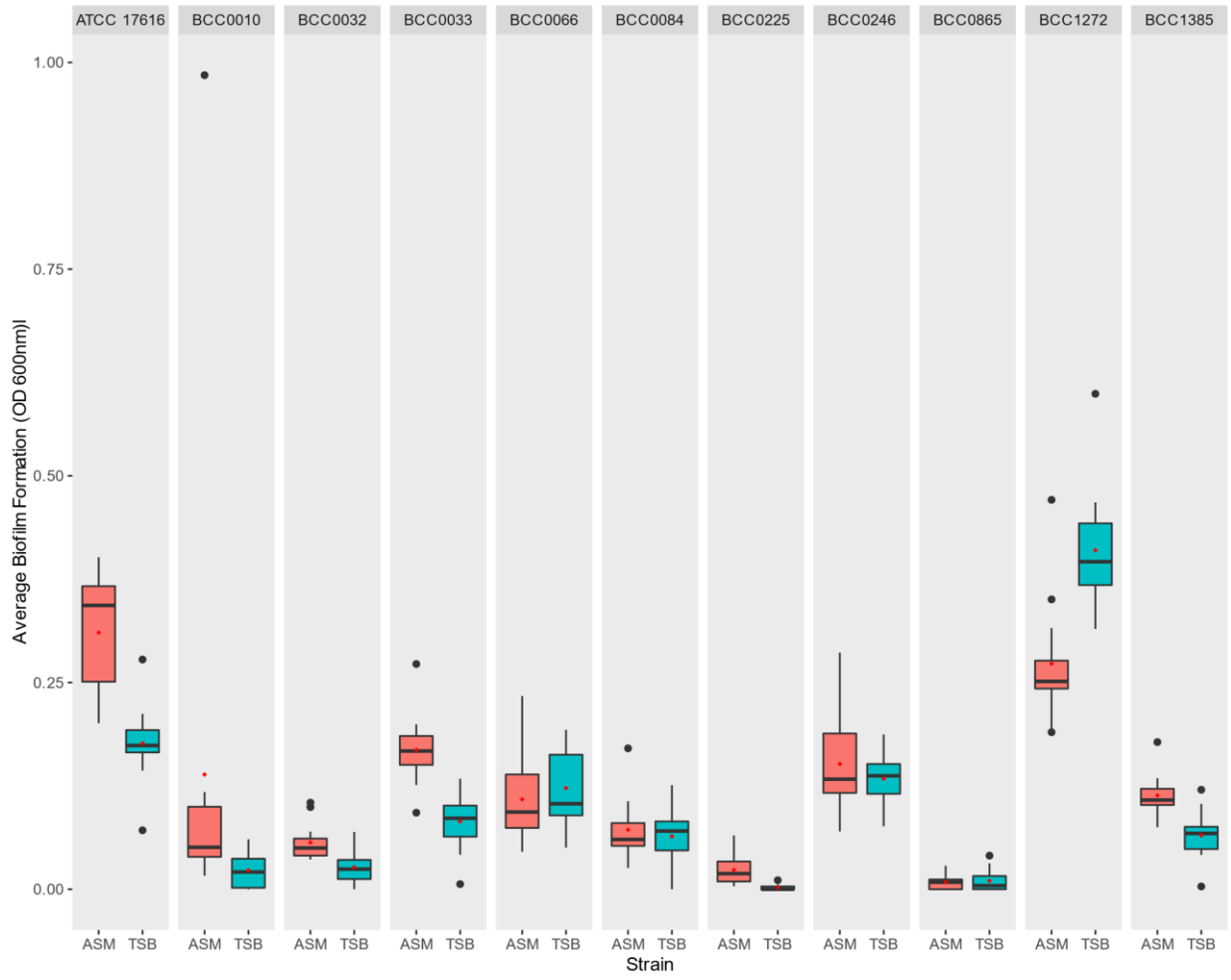


Figure 50 | Comparison of biofilm formation in TSB and ASM amongst the *B. multivorans* strains (*n* = 11). A selection of 11 strains, spanning low and high biofilm formers, were chosen for the comparative analysis. Each plot shows the mean, lower quartile, upper quartile, and any outliers.

5.2.6 *Exopolysaccharide production*

Exopolysaccharides (EPS) are macromolecules excreted from microorganisms, which also compose part of biofilms. They play a crucial role in virulence of the organism, contributing to microbial persistence (Cunha *et al.* 2004). Yeast extract medium (YEM) was used to examine the EPS production of *B. multivorans* strains (*n* = 84) (Figure 52). After 48-hours incubation, agar streak plates were examined by the naked eye, semi-quantitatively categorising the EPS production of each strain from non-mucoid (-) to highly mucoid (++++) (Figure 51). Figure 52 showed the mucoid phenotype categorisations for each of the *B. multivorans* strains. Most *B. multivorans* strains (96%) had the ability to produce anything from partial to high amounts of exopolysaccharide on the YEM agar. Five strains, BCC0006, BCC0068, BCC0188, BCC0493 and BCC0497, were observed to be non-mucoid (Table 31; Figure 52). Interestingly, all these non-mucoid *B. multivorans* strains also possessed a non- or low

motility (both swimming and swarming) phenotype (Section 5.2.3); the exception to this was BCC0188 in which motility was too variable to measure reproducibly.

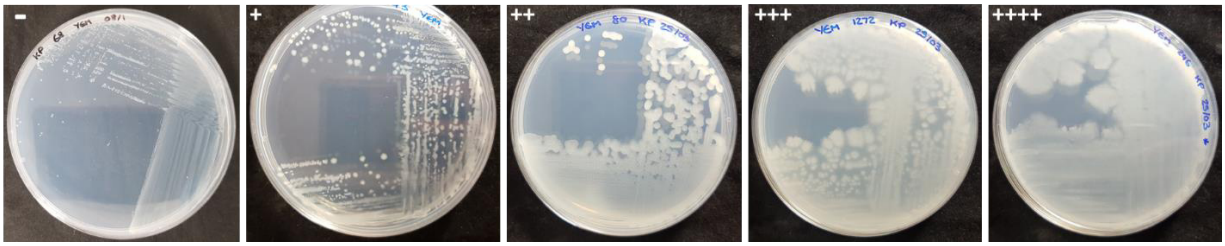


Figure 51 | EPS scoring system of the *B. multivorans* strain panel on YEM agar. The scale represents the amount of exopolysaccharide produced by each *B. multivorans* strain. Categories: '-' = non-mucoid, '+' = partially mucoid, '++' = frankly mucoid, '+++' = mucoid, '++++' = highly mucoid.

The strains were then grouped by isolation source, to identify correlations to EPS production. All 5 strains noted as non-mucoid were CF isolates (Figure 52). This category comprised 6.5% of all CF isolates examined. A similar number of CF isolates (n = 6) were categorised into the partially mucoid (+) category. For the mucoid categories ++ and ++++, a similar split of CF isolates was observed at 24.7% and 29.9%, respectively. Most CF isolates (24.7%; n = 24) were found in the mucoid +++ category. The environmental isolates (n = 3) had 2 strains in the ++++ category, with 1 in the + and +++ categories. A similar split was seen in the CGD (n = 3) isolates, where 2 fell into the highly mucoid category, with the other 1 strain was placed into the relatively ++ category. As for the non-CF isolates (n = 5), there was 1 strain categorised into groups +, +++ and ++++, and two strains in group ++ (Figure 52). Overall, the non-mucoid phenotype was not seen outside of the CF isolates, but it should be noted that the number of strains from other sources was more limited.

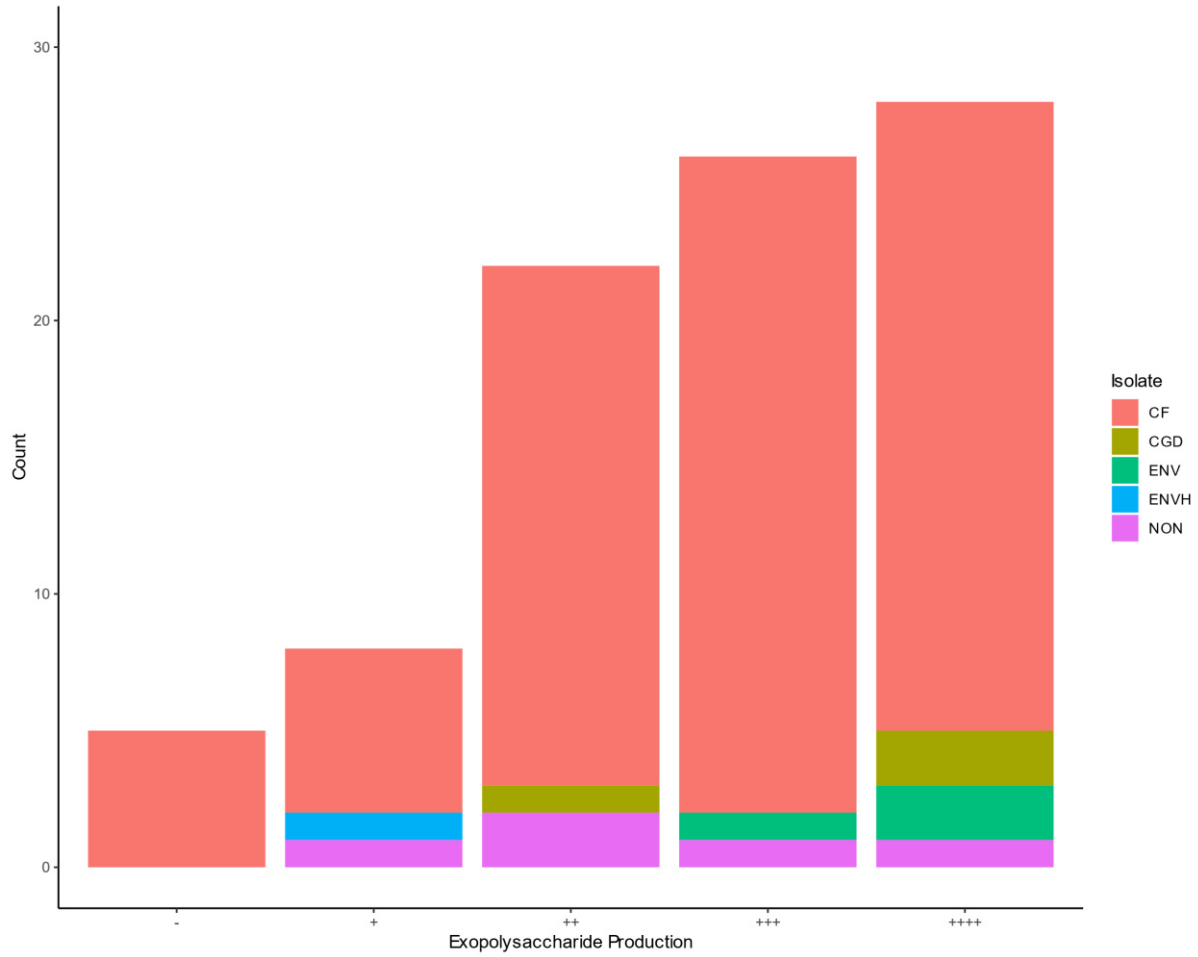


Figure 52 | Exopolysaccharide production of the *B. multivorans* strains (*n* = 84) after 48-hours incubation at 37°C. The bar chart shows the counts for each EPS category. *B. multivorans* strains have been categorised by isolate type.

5.2.7 *Protease production*

Bacterial protease production have been shown as a major virulence factor in CF infection by impacting the innate immune response and enhancing inflammation (Voynow *et al.* 2008). The aim of this experiment was to determine the amount of protease production by *B. multivorans* strains on 1% lactose-free skimmed milk agar (Morris *et al.* 2012). Two agar bases were used for the experiment: TSA and brain-heart infusion (BHI) agar. After 24-hours incubation, all *B. multivorans* strains tested (*n* = 49) were negative for protease production on both agar types (Figure 53). The plates were then incubated for a further 24-hours (48-hours total) and protease production remained negative. This is consistent with the literature where zinc metalloprotease (*zmpA* gene) was absent in *B. multivorans* (Gingues *et al.* 2005). *P. aeruginosa* LESB58 was used as a positive control for the protease assay and shown to elicit a halo around its colony as result of active protease production (see Section 5.2.9.4).

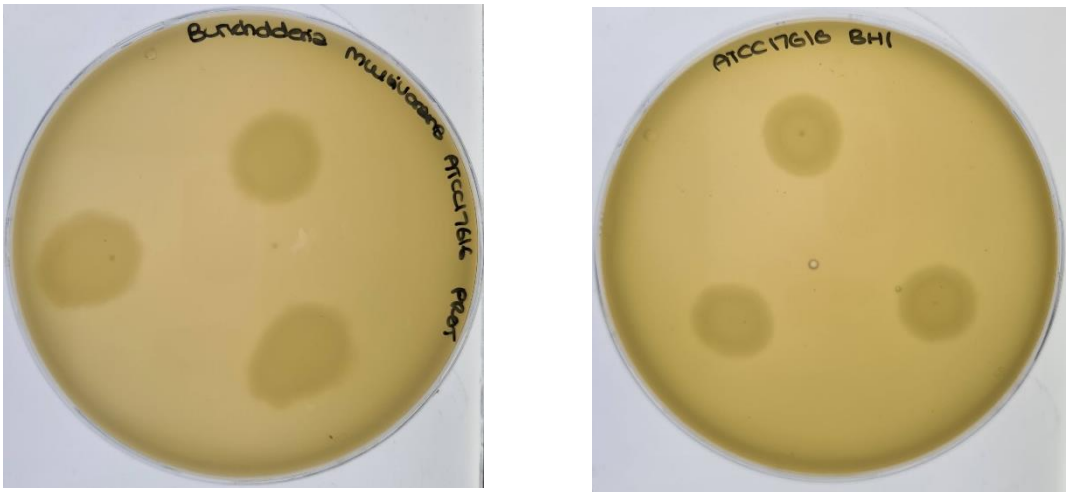


Figure 53 | *B. multivorans* absence of protease production on lactose-free skimmed-milk agar. *B. multivorans* strain ATCC 17616 colonies grown on TSA-base (left) and BHI-agar (right) skimmed milk agar do not show halos due to protease production.

5.2.8 *Quorum sensing*

In 1965, the first recording of quorum sensing (QS) as a means of microbial communication was published (Bassler *et al.* 1993). Whilst QS in *P. aeruginosa* (Barr *et al.* 2015) and *B. cenocepacia* (Deng *et al.* 2009b; Schmid *et al.* 2012) has been observed and well characterised as a virulence phenotype, QS in *B. multivorans* is not well understood. Two *E. coli* bioreporter strains (pSB406 and pSB1142) (Winson *et al.* 1998) were used to measure acyl-homoserine lactone (AHL) quorum sensing systems in the *B. multivorans* model strains (BCC0033; lineage 2b, BCC0084; lineage 1, BCC1272, ATCC 17616; lineage 2a) and strain BCC1385 (lineage 1). Reporter pSB406 was used to measure C_4-HSL and C_6-HSLs whilst pSB1142 responds to long-chain AHLs with sensitivity to 3-oxo-C_{12}-HSL (Winson *et al.* 1998). These are *lux* bioluminescence reporter constructs which detect Gram-negative bacteria AHLs. The reporters are plasmid-based and are activated by LuxR, LasR, and RhlR (Winson *et al.* 1998). By investigating whether AHLs are produced by *B. multivorans* (the four model strains plus BCC1385 were tested) (n = 5), it would provide insights into whether QS driven cell-cell communication is an important driver of virulence as seen in other species (Castillo-Juárez *et al.* 2015). Previous research subtracted the RLU readings of the *E. coli* bioluminescence reporter strains from the experimental test strains (Abdul Malik *et al.* 2020) as the reporters are self-luminescing. However, this was not performed for this study as the RLUs of the test strains were lower than that of the *E. coli* reporter controls.

The RLUs for each reading were log-transformed ($\log_{10}$) for *B. multivorans* QS analysis. Figure 54 showed the $\log_{10}$ C_4-HSL and C_6-HSL signals produced by the *B. multivorans* test strains and *E. coli* reporter control strains. The mean for the *E. coli* pSB406 control strains was 5.7 $\log_{10}$ RLU. The *B. multivorans* strains had a lower $\log_{10}$ RLU than the *E. coli* pSB406 reporter alone, ranging from 4.6 $\log_{10}$ RLU (BCC1272) to 5.6 $\log_{10}$ RLU (BCC0033; lineage 2b). *B. multivorans* ATCC 17616 had a

comparable signal to BCC1272 (both lineage 2a), emitting a mean of 4.8 $\log_{10}$ RLU (Figure 54). *B. multivorans* strains BCC0084 and BCC1385 (both lineage 1) had means of 5.3 and 5.5 $\log_{10}$ RLU, respectively. This showed that lineage 2a strains, BCC1272 and ATCC 17616, produced less C_4-HSL and C_6-HSL than the other *B. multivorans* strains (BCC0033 from lineage 2b; BCC1385 and BCC0084 from lineage 1) tested.

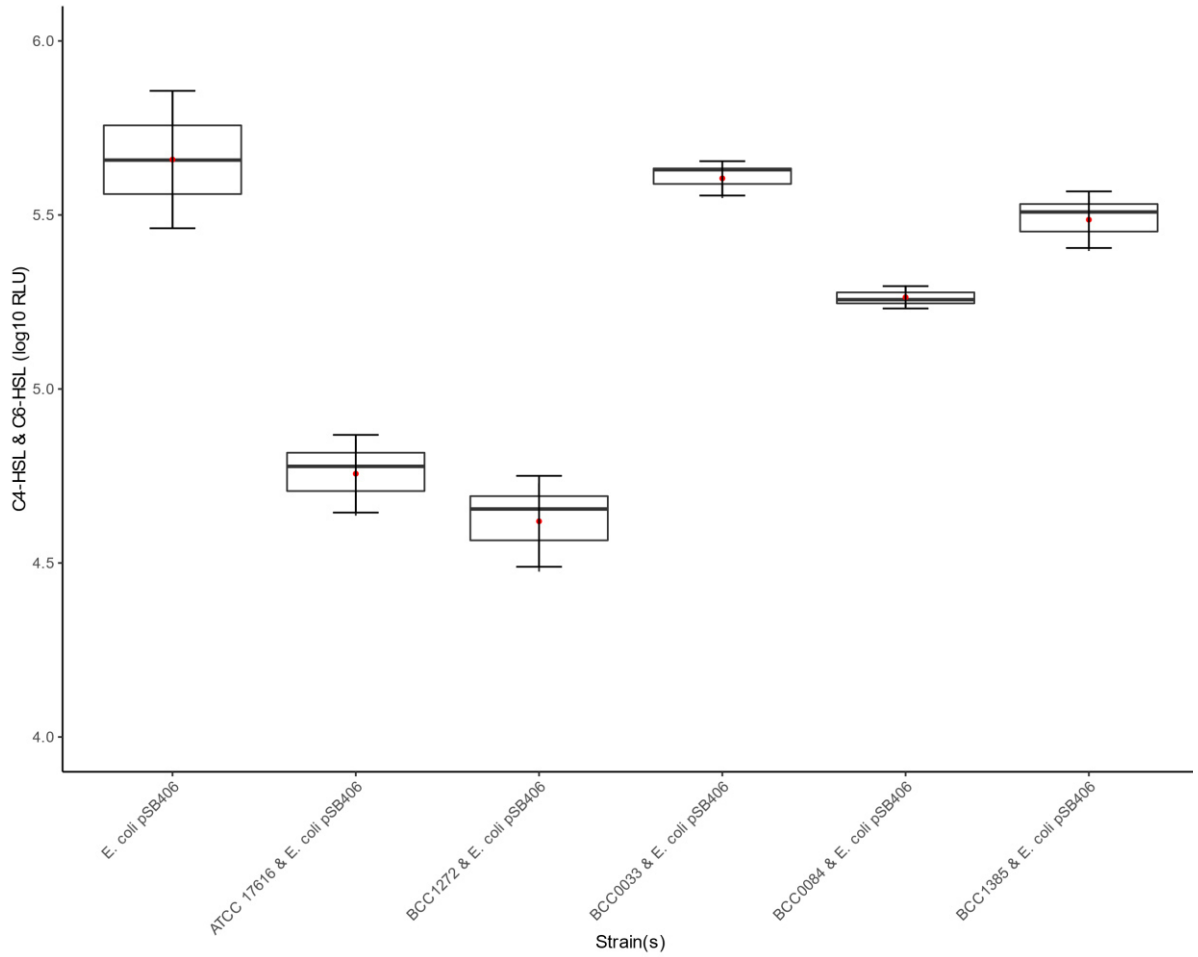


Figure 54 | C_4-HSL & C_6-HSL signals in the *B. multivorans* model strains (n = 5). *E. coli* pSB406 used as a mono-culture control. Box plots represent the $\log_{10}$ RLU.

The analysis also showed that bioreporter signals indicative of long-chain AHLs (3-oxo-C_{12}-HSL) are not produced by the *B. multivorans* strains (Figure 55). As for *E. coli* pSB1142, the control mean signal produced was 4.9 $\log_{10}$ RLU. The 3-oxo-C_{12}-HSL signals in all 5 *B. multivorans* strains were lower than the *E. coli* bioreporter control. All tested *B. multivorans* strains had a 3-oxo-C_{12}-HSL signal between 2.8 and 3.3 $\log_{10}$ RLU. On average for all the *B. multivorans* strains, $\log_{10}$ RLUs was 2.0 units lower than the control *E. coli* pSB1142 reporter control (Figure 55).

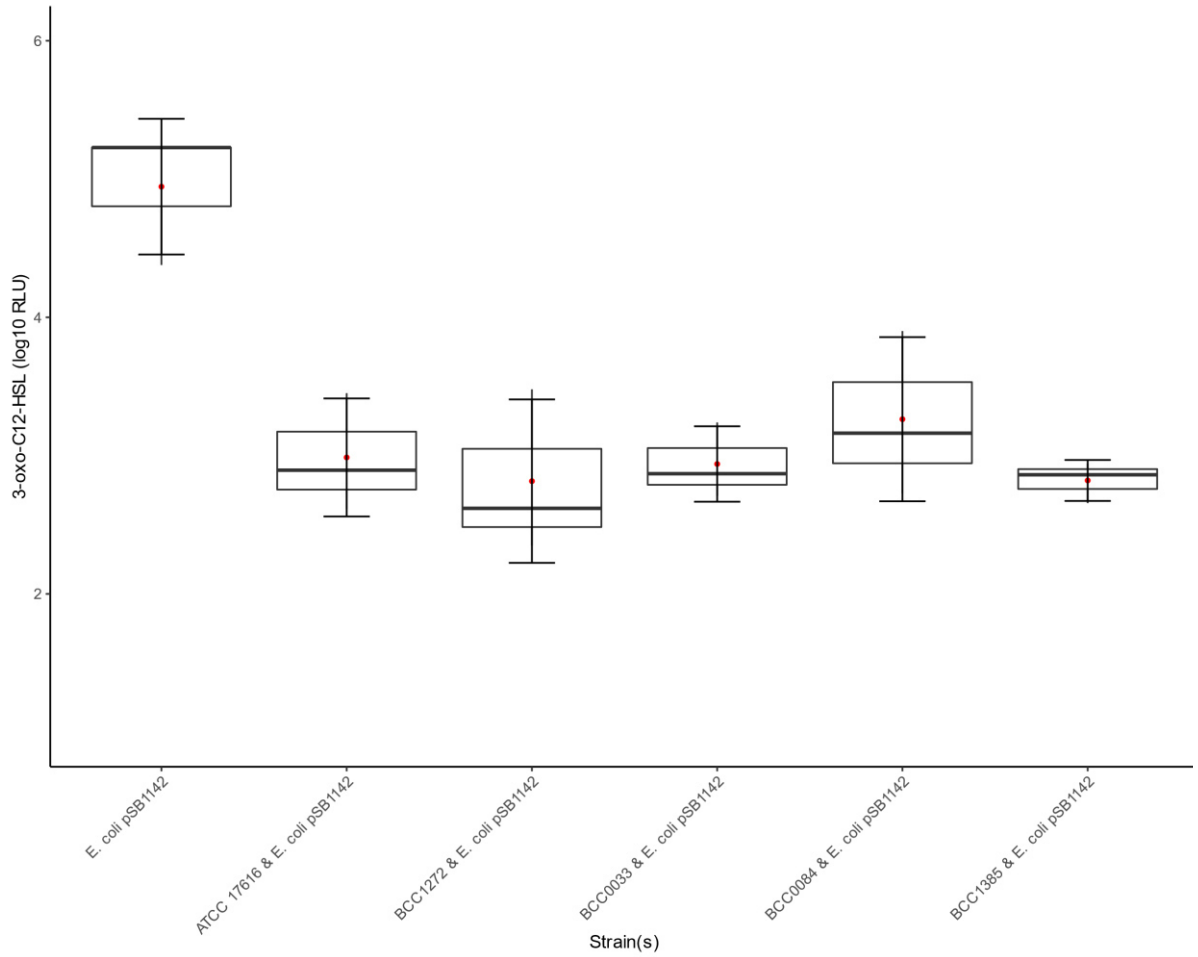


Figure 55 | 3-oxo-C_{12}-HSL signals in the *B. multivorans* model strains (*n* = 5). *E. coli* pSB1142 used as a mono-culture control. Box plots represent the $\log_{10}$ RLU.

5.2.9 *Interactions of B. multivorans with secondary CF pathogens*

Following on from the mono-culture *B. multivorans* phenotype evaluation, modelling was performed in co-culture with another CF pathogen. This pathogen is described as the secondary pathogen in the context of mixing within the assay. All secondary pathogens tested represented microbial species that act as known CF pathogens. The secondary pathogen species examined were *P. aeruginosa, B. cenocepacia, A. xylosoxidans R. mannitolilytica, S. maltophilia, Candida albicans* and *S. aureus* (Table 6). The purpose of these experiments was to examine if any *B. multivorans* phenotypic alterations occur because of microbial interactions. Microbial growth, motility, biofilm formation and protease production were evaluated as follows.

5.2.9.1 Growth rate analysis

Mixed pathogen growth was determined over a 48-hour period at 37 ºC, interacting the *B. multivorans* model strains (*n* = 4) with relevant secondary CF pathogens (*n* = 9) (Figure 56). *C. albicans* was not used in the growth interaction analysis because its behaviour was not suitable for the assays. The aim

was to observe whether the growth curves and growth parameters of the *B. multivorans* strains would be affected by the presence of the secondary pathogen. Figure 56 showed the growth rates of the example *B. multivorans* model strain ATCC 17616 alone and with a secondary pathogen present. Differences were observed between the *B. multivorans* strain grown alone and with a secondary pathogen present as follows. For all *B. multivorans* model strains, interaction with *S. aureus* caused a more rapid exponential growth, before hitting a plateau around the same time point as the other interactions. All other secondary pathogen interactions had similar exponential growth rates to the *B. multivorans* baseline, irrespective of the model strain used. In terms of BCC0033 and BCC0084, all interacting strains reach a similar overall OD $_{480\text{-}520nm}$ and stationary phase growth. For *B. multivorans* ATCC 17616 (Figure 56) and BCC1272 secondary pathogen growth interactions, the stationary growth phase and overall, OD $_{480\text{-}520nm}$ was more variable.

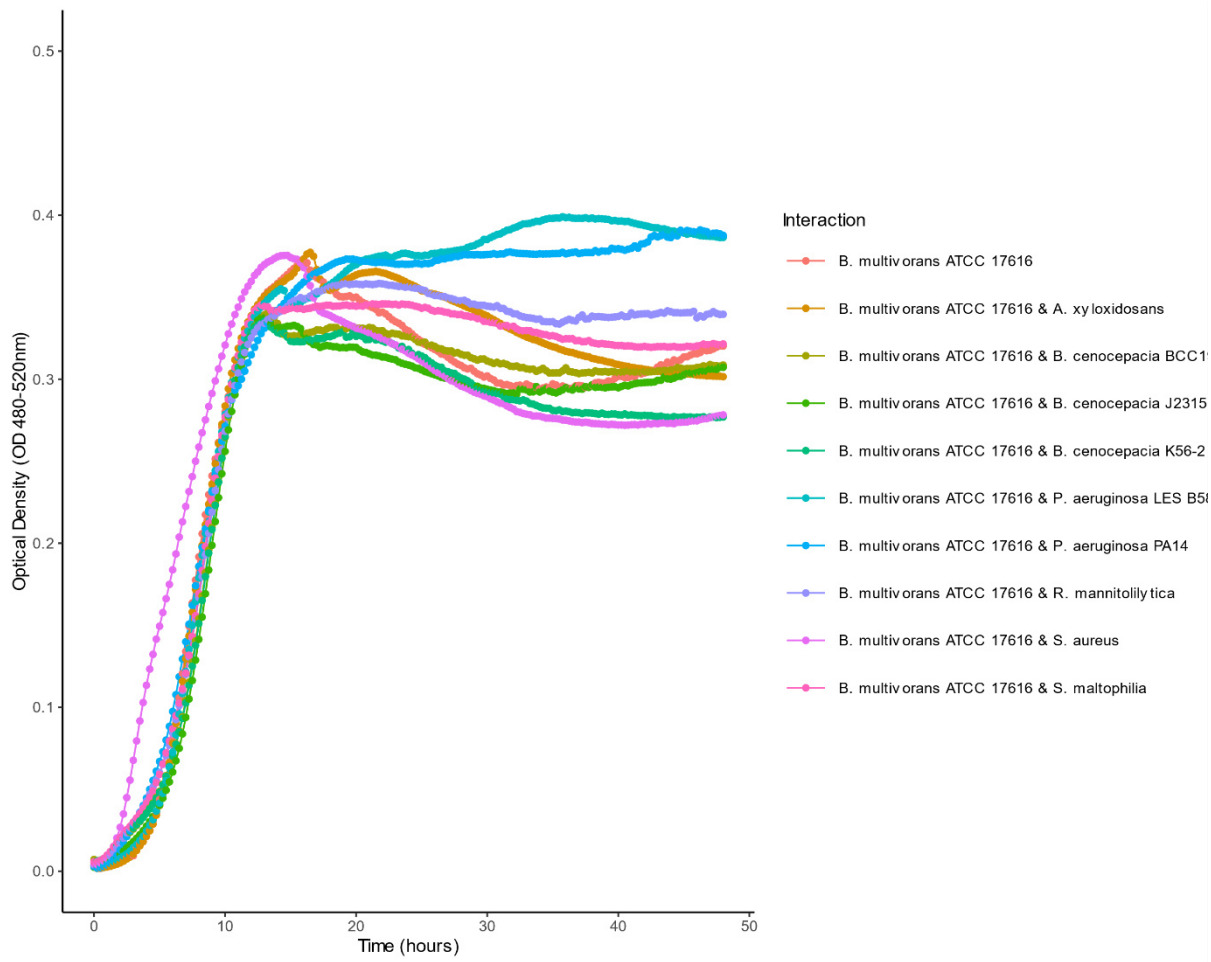


Figure 56 | Growth curve of the representative *B. multivorans* model strain ATCC 17616 with secondary CF pathogens. Growth curves have been created over a 48-hour period at 37 ºC, shaking 10 seconds before each 15-minute reading. Microbial growth was performed in a Bioscreen C instrument. All figures show the *B. multivorans* model strain in monoculture as well as co-culture with each respective secondary pathogen.

To analyse the growth parameters (maximum OD, lag phase, and growth rate) further, the GroFit package in R was used. Overall, there was no significant differences in *B. multivorans* growth with a secondary pathogen presence.

5.2.9.2 Motility assays

In Section 5.2.3, it was determined that the majority of *B. multivorans* strains possess a motile phenotype. Therefore, mixed interactions were performed to see whether the presence of a motile *B. multivorans* altered swimming and swarming motility of the secondary pathogen (n = 10). *B. multivorans* interaction strains were chosen to cover the model strains, as well as a range of low to highly motile strains. Table 35 showed the motility results for all the secondary pathogens alone. Of the secondary pathogens, *P. aeruginosa* PA14 was motile on all agar types (0.3% LB swimming, 0.5% LB swarming, and 0.5% BSM-G swarming). *A.* xylosoxidans, *B. cenocepacia* J2315, *B. cenocepacia* K56-2, *B. cenocepacia* BCC0019, *R. mannitolilytica,* and *S. maltophilia* were motile on 0.3% swimming LB agar but did not possess swarming motility on either agar type. *P. aeruginosa* LESB58 was non-motile on 0.3% LB swimming motility but was motile when assessing 0.5% swarming motility on both agars. *S. aureus* and *C. albicans* were completely non-motile on all agar types (Table 35).

Table 35 | Motility of the secondary CF pathogens. Motility was assessed after 24-hours incubation for each of the agar types. Cells highlighted in light blue indicate that the pathogen was motile.

Secondary pathogen	Motility (mm)		
	0.3% LB (swimming)	0.5% LB (swarming)	0.5% BSM-G (swarming)
P. aeruginosa LESB58	4.0	0.0	3.1
P. aeruginosa PA14	90.0	20.9	35.3
C. albicans	3.3	4.2	6.0
S. aureus	2.5	4.5	8.5
A. xylosoxidans	15.25	4.2	9.6
B. cenocepacia J2315	5.2	1.3	2.2
B. cenocepacia K56-2	12.75	3.3	6.6
B. cenocepacia BCC0019	11.3	3.0	5.2
R. mannitolilytica	8.0	4.6	10.8
S. maltophilia	7.0	3.9	7.9

An increase in swimming motility zones were observed for all secondary pathogens, except for *P. aeruginosa* PA14, when in mixed culture with a *B. multivorans* strain. The swimming motility of *P. aeruginosa* PA14 was decreased in 5 of 7 (71.4%) applicable interactions (Table 36). Interestingly, *B. multivorans* CGD isolate, BCC0009, had the ability to cover the entire agar plate when interacted with a secondary pathogen, covering a larger diameter than observed with the pathogen alone. As *P. aeruginosa* PA14 had an average swimming diameter of 90 mm when cultured alone, *B. multivorans* BCC0009 did not significantly influence its motility. Overall, there were 69 significant swimming motility interactions when *B. multivorans* was cultured with a secondary CF bacterium (Table 36)

Table 36 | Statistical comparisons of mixed pathogen swimming interactions compared to the secondary pathogen baseline. Assays performed using 0.3% LB agar. Green cells represent an increase in secondary pathogen swimming motility and red cells represent a decrease in secondary pathogen swimming motility when *B. multivorans* is present.

Strain	Motility	*C. albicans*	*A. xylosoxidans*	*B. cenocepacia*			*P. aeruginosa*		*R. mannitolilytica*	*S. aureus*	*S. maltophilia*
				BCC0019	J2315	K56-2	LESB58	PA14			
ATCC17616	Low	**	ns	*	***	ns	***	NA	*	NA	****
BCC0006		ns	ns	NA	ns	ns	ns	****	ns	ns	ns
BCC0101		*	**	ns	**	*	ns	NA	**	**	ns
BCC0737		ns	*	NA	ns	ns	ns	ns	ns	ns	ns
BCC0033	Intermediate	***	**	ns	ns	ns	***	NA	*	***	***
BCC0075		NA	NA	***	NA	ns	NA	NA	ns	ns	ns
BCC0080		*	ns	NA	*	**	*	***	**	**	*
BCC0702		*	ns	**	***	*	*	***	***	**	**
BCC1272		**	**	ns	**	ns	**	*	**	NA	**
BCC1385		***	**	***	**	*	***	NA	***	**	***
BCC0009	High	****	****	**	****	****	****	ns	****	****	****
BCC0084		ns	**	**	ns	*	*	*	*	ns	*

A similar pattern was observed when assessing the swarming motility of secondary pathogens when *B. multivorans* was present. Eight out of 10 (80%) *B. multivorans* strains could significantly increase the LB swarming motility of at least one secondary pathogen (Table 37). *B. multivorans* strain BCC0006 was unique in its ability to reduce the swarming motility of *P. aeruginosa* PA14 on LB agar. In contrast, interactions with *B. multivorans* strains BCC0009, BCC0033, BCC0075, and BCC1272 caused a significant increase of *P. aeruginosa* PA14 swarming motility on LB agar.

Table 37 | Statistical comparisons of mixed pathogen LB swarming interactions compared to the secondary pathogen baseline. Assays performed using 0.3% LB agar. Green cells represent an increase in secondary pathogen swimming motility and red cells represent a decrease in secondary pathogen swimming motility when *B. multivorans* is present.

Strain	Motility	*C. albicans*	*A. xylosoxidans*	*B. cenocepacia*			*P. aeruginosa*		*R. mannitolilytica*	*S. aureus*	*S. maltophilia*
				BCC0019	J2315	K56-2	LESB58	PA14			
ATCC17616	Low	**	*	*	**	ns	****	NA	*	ns	***
BCC0006		NA	NA	NA	NA	NA	NA	**	NA	ns	NA
BCC0009		*	****	**	**	ns	****	**	***	**	***
BCC0033		ns	ns	ns	ns	ns	*	*	ns	ns	ns
BCC0075		NA	***	***	****	*	ns	*	ns	ns	ns
BCC0084		ns	***	**	**	*	**	ns	*	ns	**
BCC0101		ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
BCC1272		ns	ns	ns	****	ns	**	***	**	NA	****
BCC1385		***	***	***	***	***	****	ns	***	ns	*
BCC0702	Intermediate	NA	**	**	****	**	ns	ns	***	ns	*

Similarly, 11 (100%) of *B. multivorans* strains tested could significantly increase the swarming motility of a secondary pathogen on BSM-G agar (Table 38). Effects of *B. multivorans* on secondary pathogen swarming ability on BSM-G were like that observed for the LB swimming motility assays (Table 36). All secondary pathogens, except for *P. aeruginosa* PA14, had their swarming motility diameter significantly increased by the presence of at least one *B. multivorans* strain. Three *B. multivorans* strains (BCC0737; lineage 1, BCC0084; lineage 1, and BCC1272; lineage 2a) significantly decreased the swarming motility of *P. aeruginosa* PA14 on BSM-G agar (Table 38). There was a total of 50 statistically significant interactions on LB swarming agar (Table 37) and 52 on BSM-G swarming agar (Table 38).

Table 38 | Statistical comparisons of mixed pathogen BSM-G swarming interactions compared to the secondary pathogen baseline. Assays performed using 0.3% LB agar. Green cells represent an increase in secondary pathogen swimming motility and red cells represent a decrease in secondary pathogen swimming motility when *B. multivorans* is present.

Strain	Motility	*C. albicans*	*A. xylosoxidans*	*B. cenocepacia*			*P. aeruginosa*		*R. mannitolilytica*	*S. aureus*	*S. maltophilia*
				BCC0019	J2315	K56-2	LESB58	PA14			
ATCC17616	Low	***	***	NA	***	NA	*	NA	*	NA	*
BCC0033		ns	ns	**	ns	ns	ns	ns	ns	ns	ns
BCC0101		****	ns	**	ns	ns	ns	ns	ns	*	ns
BCC0303		ns	ns	ns	*	NA	ns	ns	*	****	*
BCC0737		ns	*	ns	ns	ns	NA	*	***	ns	ns
BCC0084	Intermediate	**	*	***	ns	NA	**	***	*	ns	**
BCC1385		NA	**	*	**	NA	****	ns	***	***	***
BCC0009	High	****	****	NA	****	**	****	ns	****	****	****
BCC0075		NA	***	ns	**	NA	*	ns	ns	ns	**
BCC0702		*	NA	**	NA	NA	***	NA	ns	ns	NA
BCC1272		ns	ns	***	****	ns	***	*	***	**	**

The results suggest that the presence of *B. multivorans* may affect the motility of a secondary CF pathogen. When looking at the statistically significant results, every mixture of *B. multivorans* with a secondary pathogen could result in different amounts of motility changes, which is completely dependent on the strain and motility type. In general, the presence of *B. multivorans* was shown to increase the motility of a secondary pathogen. However, it can be said that *P. aeruginosa* PA14 reduces motility in multiple significant interactions with *B. multivorans* (9 of 21; 42.9%). Therefore, the CF lung microbiome would impact CF individuals differently depending on what microbes are present.

To gather further information about the swimming motility interactions, RISA PCR was used on a range of the significant bacterial interactions. RISA PCR uses the bacterial ribosomal RNA operon (Flight *et al.* 2015) for amplification, which meant that *C. albicans* would not produce a signal as a fungal species. Statistically significant interactions, based on the secondary pathogen baseline results, were analysed to see whether there was a presence of both pathogens on the outer edge of the swimming motility zone.

Figure 57 showed the RISA banding pattern of the secondary pathogens and *B. multivorans* ATCC 17616 baselines. The baseline pathogens alone were performed and amplified at the same time as the co-pathogen interactions to act as the positive controls. The strongest signalling PCR amplicon for *B. multivorans* (lane 3) was around the same size (~825 bp) as the band *A. xylosoxidans* (lane 5). This made it hard to differentiate whether two pathogens were present in the interaction studies. *S. maltophilia* had a band around 750 bp, which was also close to the *B. multivorans* pattern (Figure 57). However, the difference in size would remain distinguishable in a mixed interaction. All other banding patterns had a different size (bp) to *B. multivorans* and were suitable for analysis.

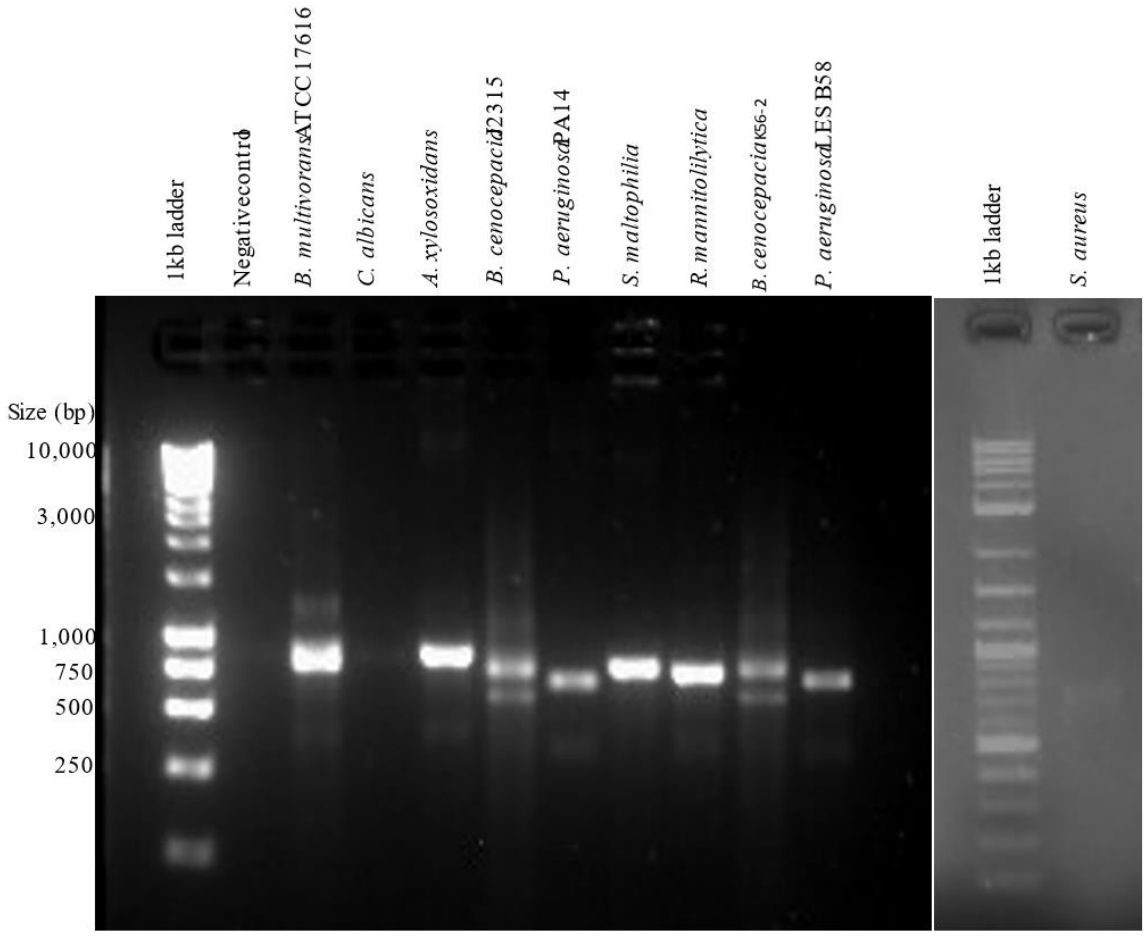


Figure 57 | RISA PCR gel of *B. multivorans* and the secondary CF community pathogen baselines. DNAse free water was used for the negative controls. Agarose gel was a concentration of 1.2% (w/v) run at 80V.

The banding patterns of example co-pathogen interactions (*B. multivorans* and the secondary pathogen) are shown in Figure 58. All secondary pathogen examples shown were non-motile for swimming, interacted with a low-high motile *B. multivorans* (Figure 58). There were two bands for each motility RISA interaction performed in this experiment. One band at around 825 bp indicated the presence of *B. multivorans.* A secondary band (between 500 bp and 750 bp) was then observed for each interaction. This indicated the presence of a secondary pathogen in the mixed motility assay. Therefore, motile *B. multivorans* showed the ability to move a non-motile secondary CF pathogen using swimming motility.

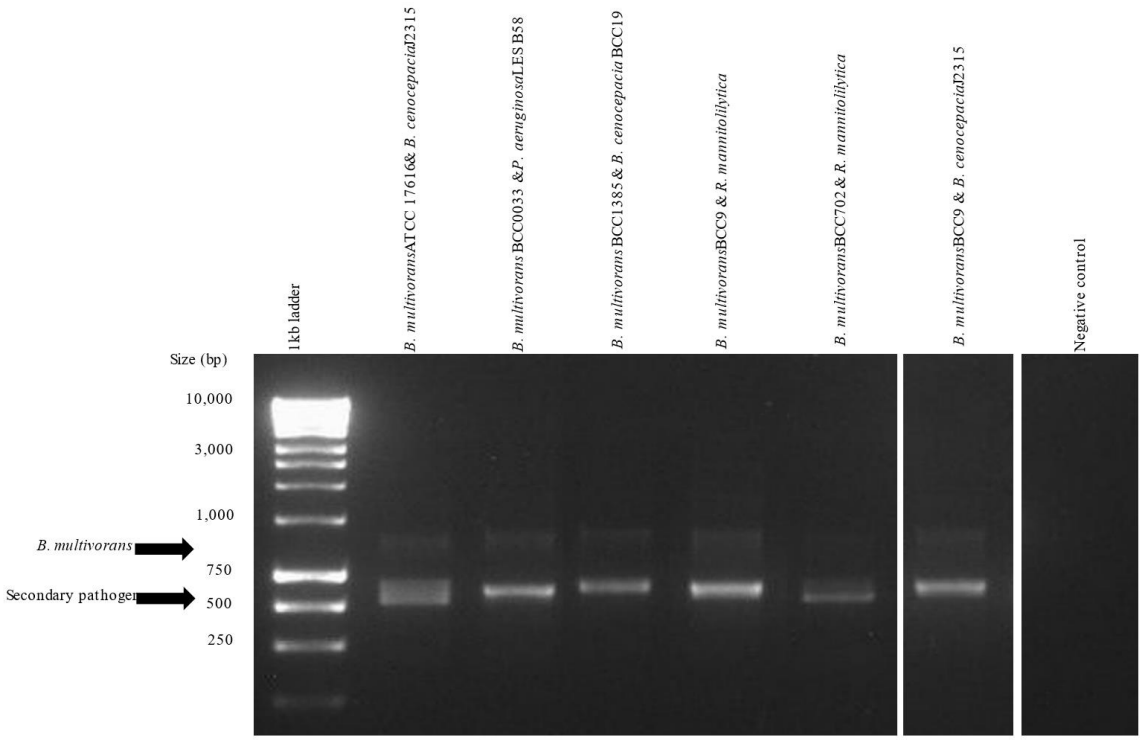


Figure 58 | RISA PCR gel of swimming agar mixed pathogen interactions. Examples are a composite image all taken from one single gel. All secondary CF pathogens were co-cultured with a primary *B. multivorans* panel strain. DNAse free water was used for the negative controls. Agarose gel was a concentration of 1.2% (w/v) run at 80V.

5.2.9.3 Biofilm production

To determine the influence of a secondary pathogen (n = 10) on *B. multivorans* (n = 20) biofilm production, the crystal violet biomass of the mixed biofilm interactions was compared to the *B. multivorans* baseline (Section 5.2.5). There were no significant differences in mean biofilm formation of the *B. multivorans* strains tested alone or with a secondary pathogen present (Appendix Table 52).

5.2.9.4 Protease production

As protease production in the *B. multivorans* strains was universally negative on skimmed milk agar (Section 5.2.7) (Table 39), interest was sparked as to whether these protease negative bacteria could alter the protease production of a protease positive bacterium. There were 9 initial secondary pathogens tested (*P. aeruginosa* PA14, *P. aeruginosa* LESB58, *R. mannitolilytica, S. maltophilia, A.* xylosoxidans*, B. cenocepacia* J2315, *B. cenocepacia* K56-2, *B. cenocepacia* BCC0019, and *S. aureus*) for protease production. Only 3 (25%) were found to be protease positive on TSA with lactose-free skimmed milk (Table 39) (Morris *et al.* 2012): *P. aeruginosa* PA14, *P. aeruginosa* LESB58, and *S. aureus.* The model strains (Section 3.2.8) were used for an initial screening against all the protease positive secondary pathogens. As the average protease production, measured by zone of clearing, was

not significantly altered when *B. multivorans* was co-cultured with *S. aureus,* the focus shifted solely onto *P. aeruginosa* strains.

Table 39 | Protease production of *B. multivorans* and the secondary pathogens on lactose-free skimmed milk agar. Average zone of clearing was determined after 24-hours growth at 37ºC.

Strain	Average zone of protease clearance (mm)
***B. multivorans* (all strains)**	0
***P. aeruginosa* PA14**	5.8
***P. aeruginosa* LESB58**	6.3
R. mannitolilytica	0
S. maltophilia	0
***A.* xylosoxidans**	0
S. aureus	1.8
***P. aeruginosa* PAK**	3.6
***P. aeruginosa* CHA**	4.1
***P. aeruginosa* PAO1**	4.8
***P. aeruginosa* Mil 162**	0
***P. aeruginosa* Les 400**	0.7
***P. aeruginosa* Les 431**	4. 7
***P. aeruginosa* AA43**	3.7
***P. aeruginosa* LMG 14084**	1. 8
***P. aeruginosa* C3719**	1. 6

The initial results showed that *P. aeruginosa* LESB58 significantly decreased protease production in all *B. multivorans* interactions. There was either none or a limited amount of protease-driven clearing zone observed on the plates after co-culture with the *B. multivorans* model strains at 24-hours (Table 40). In contrast, there were differences between the protease zone of clearing created by *P. aeruginosa* PA14 alone compared to when co-cultured with *B. multivorans*¸ the reductions were not significant. Therefore, the experiment was expanded to include other *P. aeruginosa* strains (PAK, CHA, PAO1, Mil 162, LES 400, LES 431, AA43, LMG 14084, and C3719) to cover the different *P. aeruginosa* lineages and clades observed in a previous study (Weiser *et al.* 2019). Mil 162 was protease negative on the TSA with lactose-free skimmed milk so was not used for further analysis. There were 4 significant interactions observed in the additional *P. aeruginosa* strains. *B. multivorans* BCC1272 (lineage 2a) and BCC0084 (lineage 1) both significantly decreased the protease production of *P. aeruginosa* strains LES 431 and C3719. When *P. aeruginosa* these strains were interacted with model strains *B. multivorans* BCC0033 and ATCC 17616, no significant difference was found (Table 40).

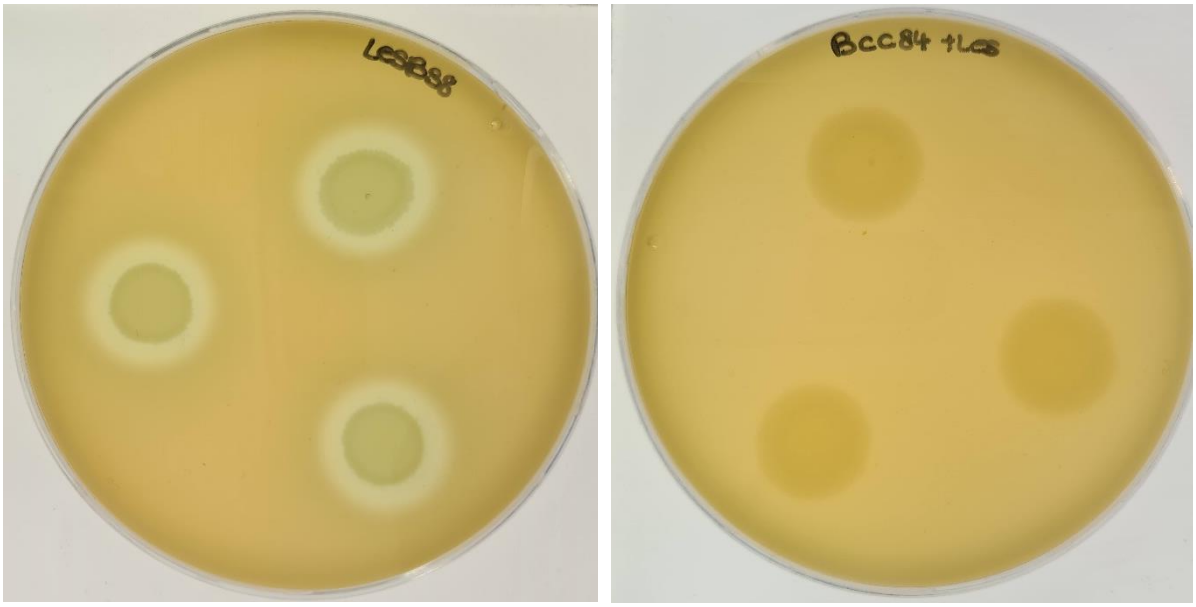


Figure 59 | Protease production of *P. aeruginosa* LESB58 alone and in co-culture with *B. multivorans*. Figure is an example of the protease suppression by *B. multivorans* on lactose-free skimmed-milk agar when grown simultaneously. Growth of inoculation spots after 24-hours at 37ºC. Left: *P. aeruginosa* LESB58 alone. Right: *P. aeruginosa* LESB58 and *B. multivorans* BCC0084 interaction.

Table 40 | Protease zone of clearing when *B. multivorans* is co-cultured with a secondary CF pathogen.

B. multivorans strain	Secondary pathogen	Zone of clearance	p-adjusted value[a]	p-adjusted significance
BCC1272	*S. aureus*	1.4	0.403	ns
	P. aeruginosa PA14	4.8	0.307	ns
	P. aeruginosa LESB58	0.0	<0.00001	****
	P. aeruginosa PAO1	4.0	0.116	ns
	P. aeruginosa CHA	4.7	0.417	ns
	P. aeruginosa PAK	2.9	0.225	ns
	P. aeruginosa Mil 162	0.5	0.277	ns
	P. aeruginosa LES 400	0.0	0.275	ns
	P. aeruginosa LES 431	2.4	0.047	*
	P. aeruginosa AA43	2.7	0.249	ns
	P. aeruginosa LMG 14084	0.7	0.144	ns
	P. aeruginosa C3719	0.0	0.048	*
BCC0033	*S. aureus*	1.8	0.506	ns
	P. aeruginosa PA14	4.6	0.319	ns
	P. aeruginosa LESB58	0.0	<0.00001	****
	P. aeruginosa PAO1	4.4	0.205	ns
	P. aeruginosa CHA	34	0.271	ns
	P. aeruginosa PAK	2.8	0.287	ns
	P. aeruginosa Mil 162	0.0	0.526	ns
	P. aeruginosa LES 400	0.0	0.275	ns
	P. aeruginosa LES 431	2.7	0.100	ns
	P. aeruginosa AA43	4.4	0.365	ns
	P. aeruginosa LMG 14084	1.4	0.392	ns
	P. aeruginosa C3719	0.7	0.183	ns
BCC0084	*S. aureus*	1.8	0.5189	ns
	P. aeruginosa PA14	4.1	0.202	ns
	P. aeruginosa LESB58	0.0	<0.00001	****
	P. aeruginosa PAO1	4.0	0.151	ns
	P. aeruginosa CHA	3.8	0.257	ns
	P. aeruginosa PAK	3.4	0.432	ns
	P. aeruginosa Mil 162	0.0	0.527	ns
	P. aeruginosa LES 400	0.0	0.276	ns
	P. aeruginosa LES 431	0.0	<0.00001	****
	P. aeruginosa AA43	5.0	0.301	ns
	P. aeruginosa C3719	0.0	0.047	*
	P. aeruginosa LMG 14084	0.3	0.068	ns
ATCC	*S. aureus*	1.6	0.470	ns
	P. aeruginosa PA14	7.3	0.350	ns
	P. aeruginosa LESB58	0.5	<0.00001	****
	P. aeruginosa PAO1	4.1	0.192	ns
	P. aeruginosa CHA	3.2	0.240	ns
	P. aeruginosa PAK	3.5	0.521	ns
	P. aeruginosa Mil 162	0.0	0.528	ns
	P. aeruginosa LES 400	0.0	0.277	ns
	P. aeruginosa LES 431	3.2	0.198	ns
	P. aeruginosa AA43	46	0.261	ns
	P. aeruginosa LMG 14084	1.6	0.445	ns

	P. aeruginosa C3719	0.4	0.171	ns
B. multivorans BCC0009	*P. aeruginosa* PA14	5.	0.303	ns
	P. aeruginosa LESB58	0.3	<0.00001	****
B. multivorans BCC0006	*P. aeruginosa* PA14	4.4	0.160	ns
	P. aeruginosa LESB58	2.9	<0.001	***
B. multivorans BCC0101	*P. aeruginosa* PA14	4.6	0.171	ns
	P. aeruginosa LESB58	0.0	<0.00001	****
B. multivorans BCC1385	*P. aeruginosa* PA14	4.4	0.141	ns
	P. aeruginosa LESB58	0.0	<0.00001	****
B. multivorans BCC0080	*P. aeruginosa* PA14	4.1	0.082	ns
	P. aeruginosa LESB58	1.3	<0.00001	****
B. multivorans BCC0737	*P. aeruginosa* PA14	4.4	0.133	ns
	P. aeruginosa LESB58	2.9	<0.001	***
B. multivorans BCC0303	*P. aeruginosa* PA14	4.4	0.070	ns
	P. aeruginosa LESB58	3.7	<0.00001	****

[a]*P*-values were calculated using the protease production of the secondary pathogens alone as the baseline comparisons.

To confirm whether this *P. aeruginosa* LESB58 interaction result was unique to *B. multivorans*, protease negative *B. cenocepacia* strain (n = 3) were used in the interaction. After 24-hours on the lactose-free skimmed milk TSA, zone of clearing was observed in the co-culture (Table 41). Results showed that there were significant differences between *P. aeruginosa* LESB58 alone and with *B. cenocepacia* present. The smallest reduction of protease secretion by *P. aeruginosa* LESB58 was in mixture with *B. cenocepacia* strain J2315 (p = 0.03) (Table 41). However, *B. cenocepacia* strains BCC0019 and K56-2 had a greater effect on reducing the protease-producing ability of *P. aeruginosa* LESB58, with a p-value of 0.0006 and 0.0001, respectively (Table 41).

Table 41 | Protease zone of clearing of *B. cenocepacia* alone and when co-cultured with *P. aeruginosa* LESB58. *P-values* were calculated using the protease production of the secondary pathogens alone as the baseline comparisons.

Strain	Average zone of clearing	*p*-value	*p*-adjusted	Significance
B. cenocepacia J2315	0			
B. cenocepacia BCC0019	0			
B. cenocepacia K56-2	0			
P. aeruginosa LESB58	6.3			
B. cenocepacia J2315 & *P. aeruginosa* LESB58	4.666667	0.01645	0.03289	*
B. cenocepacia BCC0019 & *P. aeruginosa* LESB58	3.5	0.0002	0.0006	***
B. cenocepacia K56-2 & *P. aeruginosa* LESB58	3	1.81^{-5}	0.0001	***

To further assess the effect of *B. multivorans* on *P. aeruginosa* LESB58 protease production, the lactose-free skimmed milk assay was continued to a total of 48-hours. The longer incubation was used

to observe whether there were any changes in zone of clearing at the later time point. All the *P. aeruginosa* strains and *B. multivorans* model from the 24-hour interaction experiment were used. The most significant interactions were observed when all four model strains were interacted with *P. aeruginosa* LESB58. Interaction with *B. multivorans* strains BCC0033 ($p = 0.0003$), BCC0084 ($p = <0.00001$), BCC1272 ($p = <0.00001$), and ATCC 17616 ($p = <0.00001$) significantly reduced *P. aeruginosa* LESB58 protease production. One other significant reduction in protease was observed when *P. aeruginosa* LES 431 was mixed with BCC0084 ($p = 0.01$). All other *P. aeruginosa* and *B. multivorans* protease interactions were not statistically significant at 48-hours incubation.

Table 42 | Protease fold change (FC) of the secondary pathogens when interacted with *B. multivorans* model strains.

***B. multivorans* strain**	**Secondary pathogen**	**Protease Fold Change (FC)**	
		24-hours	**48-hours**
BCC1272	*S. aureus*	/	/
	P. aeruginosa PA14	-1.4	2.1
	P. aeruginosa LESB58	-7.1	-11.5
	P. aeruginosa PAO1	-2.1	1.8
	P. aeruginosa CHA	0.6	1.7
	P. aeruginosa PAK	-1.1	0.6
	P. aeruginosa Mil 162	1.0	-1.4
	P. aeruginosa LES 400	-0.7	-3.6
	P. aeruginosa LES 431	-1.7	-1.2
	P. aeruginosa AA43	0.0	-1.7
	P. aeruginosa LMG 14084	-0.6	-0.9
	P. aeruginosa C3719	-1.8	-1.6
BCC0033	*S. aureus*	/	/
	P. aeruginosa PA14	0.4	3.3
	P. aeruginosa LESB58	-7.1	-10.8
	P. aeruginosa PAO1	-0.3	5.9
	P. aeruginosa CHA	-0.7	3.2
	P. aeruginosa PAK	-0.1	1.7
	P. aeruginosa Mil 162	0.0	-0.9
	P. aeruginosa LES 400	-0.7	-3.6
	P. aeruginosa LES 431	-1.2	-3.2
	P. aeruginosa AA43	1.8	4.0
	P. aeruginosa LMG 14084	-0.6	0.9
	P. aeruginosa C3719	0.2	-1.6
BCC0084	*S. aureus*	/	/
	P. aeruginosa PA14	0.4	2.1
	P. aeruginosa LESB58	-7.1	-14.2
	P. aeruginosa PAO1	-1.1	3.4
	P. aeruginosa CHA	-1.1	1.7
	P. aeruginosa PAK	-0.3	0.7
	P. aeruginosa Mil 162	0.0	-2.8
	P. aeruginosa LES 400	-0.7	-3.6
	P. aeruginosa LES 431	-4.7	-6.9
	P. aeruginosa AA43	2.6	5.5
	P. aeruginosa C3719	-1.1	-3.6
	P. aeruginosa LMG 14084	-1.7	-0.4
ATCC	*S. aureus*	/	/
	P. aeruginosa PA14	0.6	1.0
	P. aeruginosa LESB58	-6.6	-11.4
	P. aeruginosa PAO1	-1.1	7.8
	P. aeruginosa CHA	-0.1	2.6
	P. aeruginosa PAK	0.1	0.6
	P. aeruginosa Mil 162	0.0	1.9
	P. aeruginosa LES 400	-0.7	-3.6
	P. aeruginosa LES 431	-1.3	-1.4

P. aeruginosa AA43	1.4	1.6
P. aeruginosa LMG 14084	0.0	-0.4
P. aeruginosa C3719	-1.2	-0.4

Fold change (FC) was calculated for the *B. multivorans* model strains (n = 4) protease interactions with both *S. aureus* (n = 1) and *P. aeruginosa* (n = 11) strains. Presence of the *B. multivorans* resulted in a negative protease fold change when interacted with most *P. aeruginosa* strains (n = 30; 88.6%) after 24-hours (Table 42). *P. aeruginosa* strain LESB58 had the greatest negative FC in all *B. multivorans* interactions (Figure 60). At 24-hours, there was a -7.1 FC when *P. aeruginosa* LESB58 was interacted with *B. multivorans* BCC0033, BCC0084, and BCC1272. A smaller FC (-6.6) was observed at 24-hours when *P. aeruginosa* LESB58 was interacted with *B. multivorans* strain ATCC 17616 (Figure 60). There was also a negative FC (-4.7) when *P. aeruginosa* LES431 was interacted with *B. multivorans* BCC0084. All other interactions had a FC difference of less than 2 (Table 42).

By 48 hours, the negative FC of *P. aeruginosa* LESB58 decreased further to between -10.8 and -14.2, dependant on the interacting *B. multivorans* strain. The biggest FC (-14.2) was observed when *P. aeruginosa* LESB58 was interacted with *B. multivorans* BCC0084 (Table 42; Figure 60). There were no other consistent FCs across all the *B. multivorans* model strains and a specific *P. aeruginosa* strain. Therefore, *B. multivorans* protease interaction with *P. aeruginosa* LESB58 was deemed the most interesting and subject to further investigation.

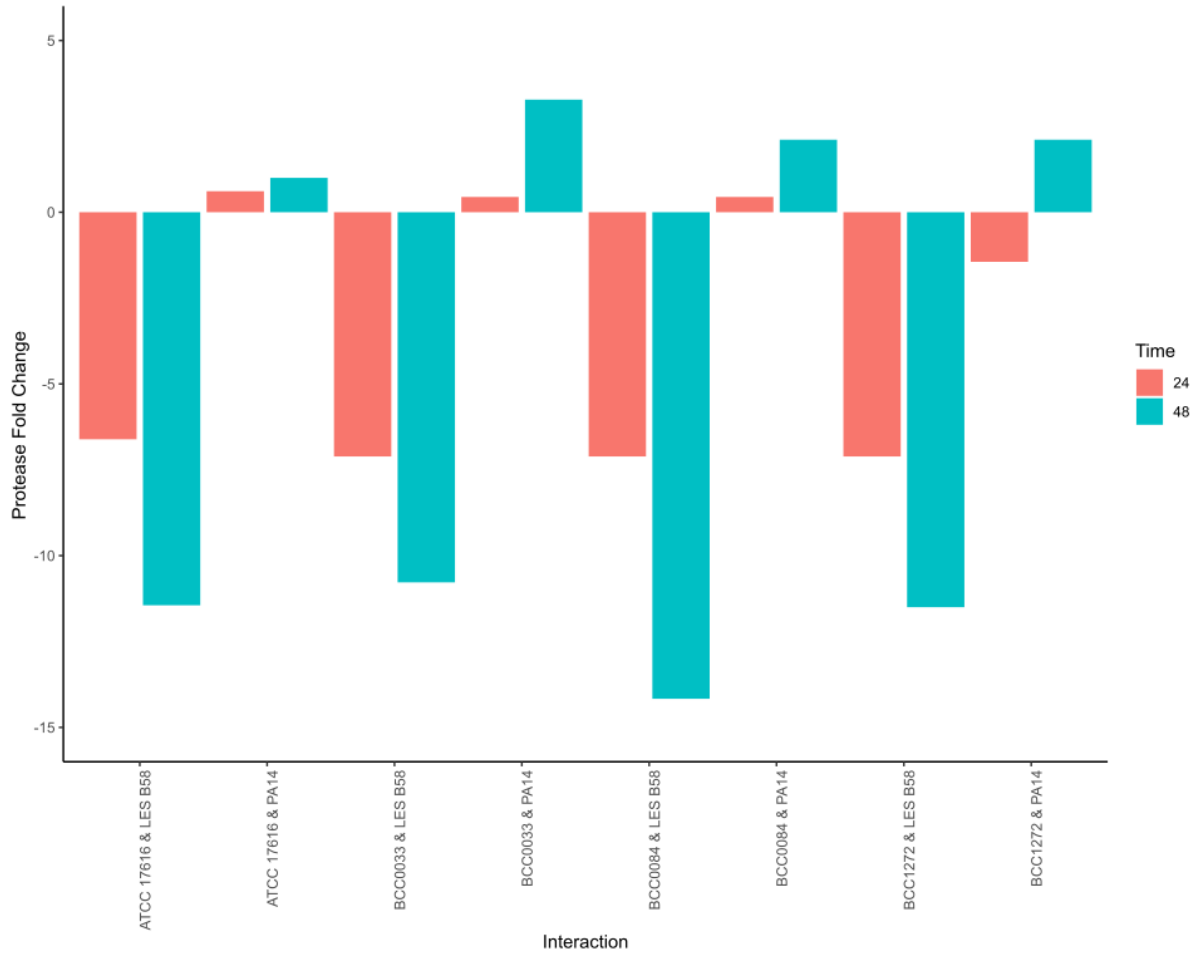


Figure 60 | Fold change (FC) of protease production in the *P. aeruginosa* PA14 and LESB58 strains when interacted with *B. multivorans* model strains after 24-hours and 48-hours incubation. Assay was performed on lactose-free skimmed-milk agar and incubated at 37 ºC. FC was calculated using the protease production of the *P. aeruginosa* strains alone as the baseline. This was compared to the protease production when *B. multivorans* was present.

The mechanism behind why *B. multivorans* was reducing protease production in *P. aeruginosa* LESB58 was investigated as follows. Initially, there was thought that the *P. aeruginosa* LESB58 strain was being killed, through a previously described contact-dependant T6SS method (Spiewak *et al.* 2019), by the *B. multivorans* strains, thus not producing a protease clear zone on the lactose-free skimmed milk agar. To test this, viability assays (Section 2.1.3) were performed using antibiotic selection agars. *P. aeruginosa* LESB58 was sensitive to 600 units polymyxin B and *B. multivorans* was sensitive to 50 µg/ml trimethoprim when grown on TSA. End-point interaction and mono-culture spots, from the 24-hour plates, were scraped from the protease agar plates and re-suspended. These were plated onto the appropriate antibiotic selections for bacterial enumeration. The results of the monocultures were compared to the co-culture interactions.

Overall, there was no indication that *B. multivorans* killed *P. aeruginosa* in this protease assay interaction. The mono-culture comparisons showed that the *P. aeruginosa* LESB58 strain had a mean

viability of 10^8 CFU/ml. When mixed in the protease assay interaction, the 24-hour viability of *B. multivorans* and *P. aeruginosa* LESB58, on average, was 10^9 CFU/ml and 10^6 CFU/ml, respectively. Whilst the *P. aeruginosa* LESB58 viability depleted in co-culture by 2 logs, there was evidence that the two bacteria were surviving as a mixture. Therefore, the assumption of contact-dependant killing of *P. aeruginosa* was rejected. The decrease in viability of *P. aeruginosa* LESB58 in the culture could instead be due to its slow-growing nature in comparison to the *B. multivorans* strain.

To ensure the *P. aeruginosa* LESB58 strains could still produce protease in monoculture at lower bacterial concentrations, such as the 10^6 CFU/ml seen after interaction with *B. multivorans*, serial dilutions were used. The results showed that protease production could be observed at 24-hours with starting inoculum as low as at 10^3 CFU/ml. After 48-hours incubation, protease production by *P. aeruginosa* LESB58 was seen in the 10^2 inoculated cultures. Therefore, even if the interaction with *B. multivorans* suppressed *P. aeruginosa* growth, protease production should have still been seen at the lower bacterial concentrations.

To confirm that *B. multivorans* BCC0084 was not killing *P. aeruginosa* LESB58, genetically manipulated strains (*B. multivorans* BCC0084::pIN301-eGDP and *P. aeruginosa* LESB58::pIN233-mCherry) were used as fluorescent reporters in a planktonic growth assay. Interval measurements were recorded for the fluorescent reporters in mono- and co-culture. The signal from the monocultures were used to normalise the signals produced in co-culture, providing insight into the bacterial growth using a fluorescent signal intensity mechanism. Figure 61 showed the $\log_{10}$ relative fluorescence units (RFU) observed over a 24-hour period. The *P. aeruginosa* LESB58::pIN233-mCherry strain has a lower $\log_{10}$ RFU compared to *B. multivorans* BCC0084::pIN301-eGFP alone and the two strains combined. The RFU of *P. aeruginosa* LESB58::pIN233-mCherry remains consistent over the 24-hour period, between 3.5 and 4 RFU. BCC0084::pIN233-mCherry has a higher starting $\log_{10}$ RFU at ~4.75 before dropping to ~4.25 by 5 hours. The $\log_{10}$ RFU BCC0084::pIN233-mCherry then increased over the remaining 24-hour period to ~4.6 (Figure 61). To analyse the combined strain data, readings were normalised against both pIN301-eGFP and pIN233-mCherry and plotted separately. This enabled visualisation of the fluorescence emitted by both strains individually in the co-culture. The pIN301-eGFP normalised co-culture followed the same $\log_{10}$ RFU pattern over the 24-hours which the BCC0084::pIN301-eGFP strain did alone. In contrast, the pIN233-mCherry co-culture showed a continual reduction in $\log_{10}$ RFU over 24-hours. This indicated that the *B. multivorans* BCC0084::pIN301-eGFP culture was able to continually grow in the co-culture, whilst *P. aeruginosa* LESB58::pIN233-mCherry depleted over time. This coincides with the reduction, but not complete loss, in *P. aeruginosa* LESB58 viability after 24-hours in co-culture with *B. multivorans*.

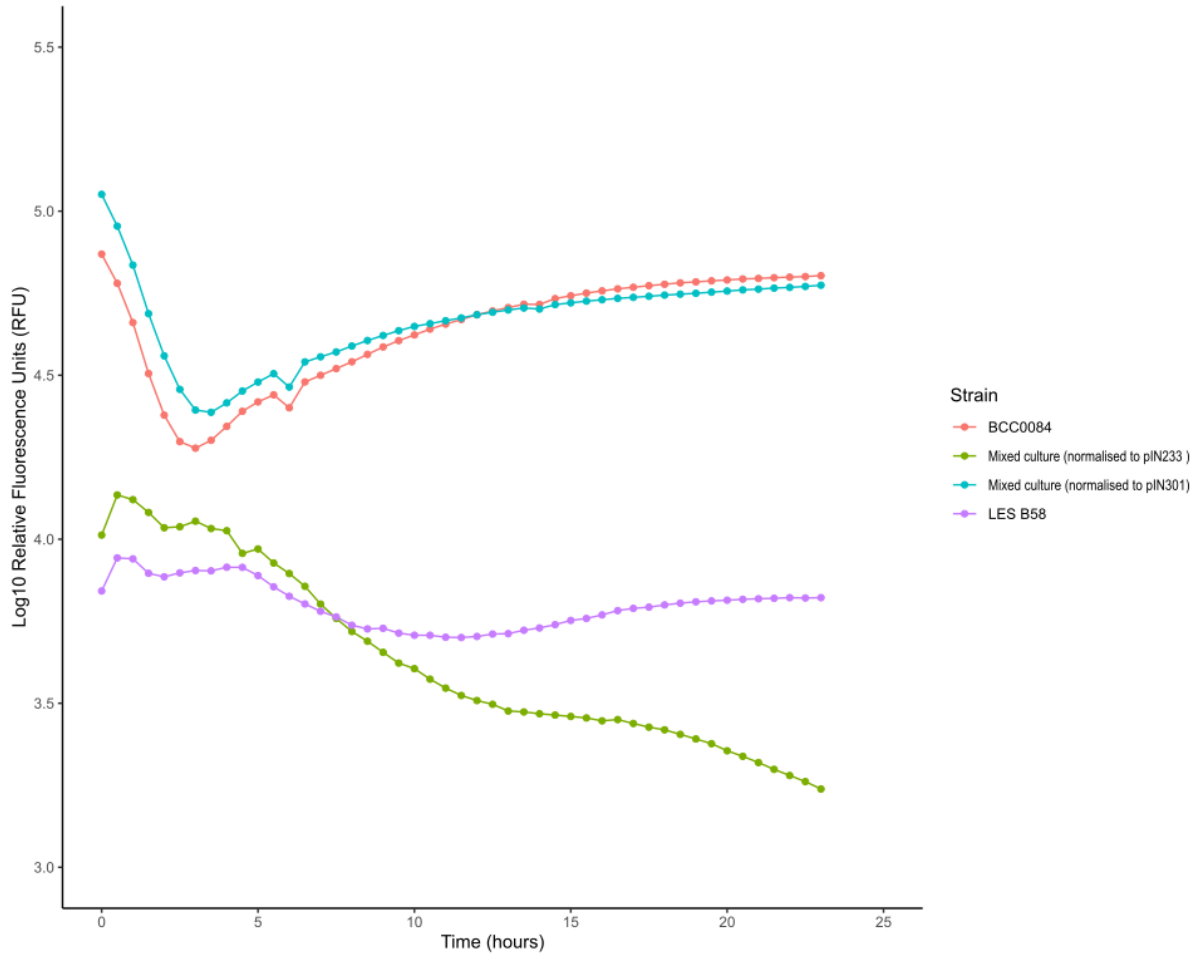


Figure 61 | Log_{10} relative fluorescence units (RFU) of *B. multivorans* BCC0084::pIN301-eGFP and *P. aeruginosa* LESB58::pIN233-mCherry both in mono- and co-culture. The mixed culture readings were normalised against each of the fluorescent strains individually. These were plotted separately to illustrate the RFU of each strain in the co-culture.

The growth curve ($OD_{480\text{-}520nm}$) showed that the *P. aeruginosa* LESB58::pIN233-mCherry strain was much slower growing than both *B. multivorans* BCC0084::pIN301-eGFP alone and the two strains in mixed culture (Figure 62). This correlates to the results shown in Section 5.2.9.1, where *B. multivorans* had a faster growth rate compared to *P. aeruginosa* LESB58 (Figure 56). Overall, this strengthens the idea that the two bacterial strains can grow simultaneously, and that protease production of *P. aeruginosa* LESB58 is suppressed in an alternative mechanism when driven by the interaction with *B. multivorans.*

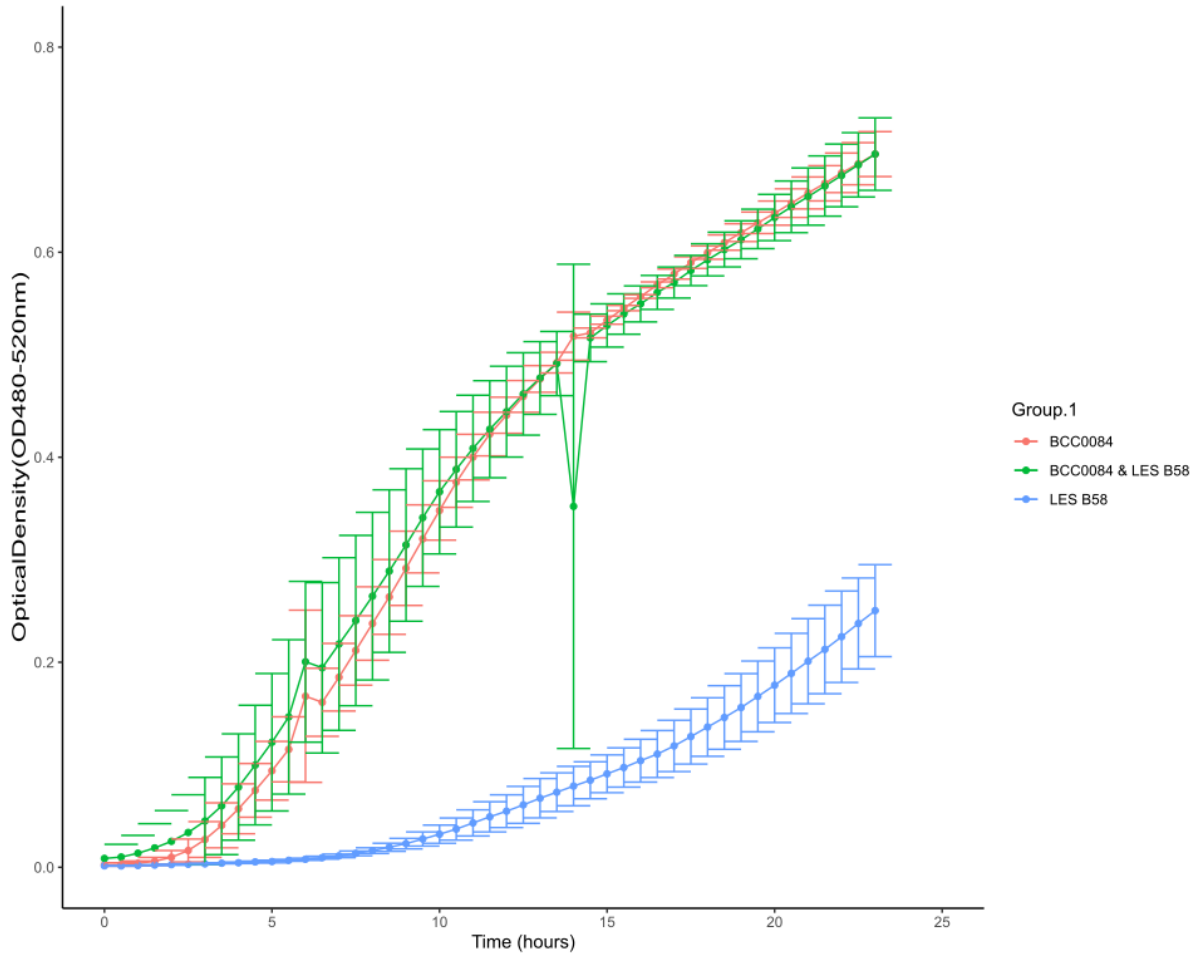


Figure 62 | Log_{10} optical density ($OD_{480\text{-}520nm}$) of *B. multivorans* BCC0084::pIN301-eGFP and *P. aeruginosa* LESB58::pIN233-mCherry both in mono- and co-culture. The mixed culture OD readings were normalised against each strain individually. These were plotted separately to illustrate the $OD_{480\text{-}520nm}$ of each strain in the co-culture.

5.2.9.5 Quorum sensing

It is known that *P. aeruginosa* secretes proteases during CF infection, often regulated by QS (Li and Lee 2019). To see whether QS of *P. aeruginosa* LESB58 was affected when *B. multivorans* was present, a mixed interaction QS assay was performed using *E. coli* bioreporter strains as previously mentioned (Section 5.2.8). It was thought that a decrease in *P. aeruginosa* LESB58 QS, when *B. multivorans* was present, could correlate to a reduction of protease production. C_4-HSL and C_6-HSL RLUs for the *P. aeruginosa* LESB58 strain alone and in co-culture with the *B. multivorans* model strains can be seen in Figure 63. All the interactions had a lower log_{10} RFU compared to the control *E. coli* pSB406 reporter strain alone, indicating the lack of AHL production in all test samples. The results showed that the *P. aeruginosa* LESB58 strain had a similar RLU when cultured alone and with *B. multivorans* strains (Figure 63). This result showed a lower C_4-HSL and C_6-HSL RLU expression compared to that observed of the *B. multivorans* strains alone in Section 5.2.8. This gave the idea that *B. multivorans* can decrease *P. aeruginosa* LESB58 C_4-HSL and C_6-HSL production in co-culture interactions.

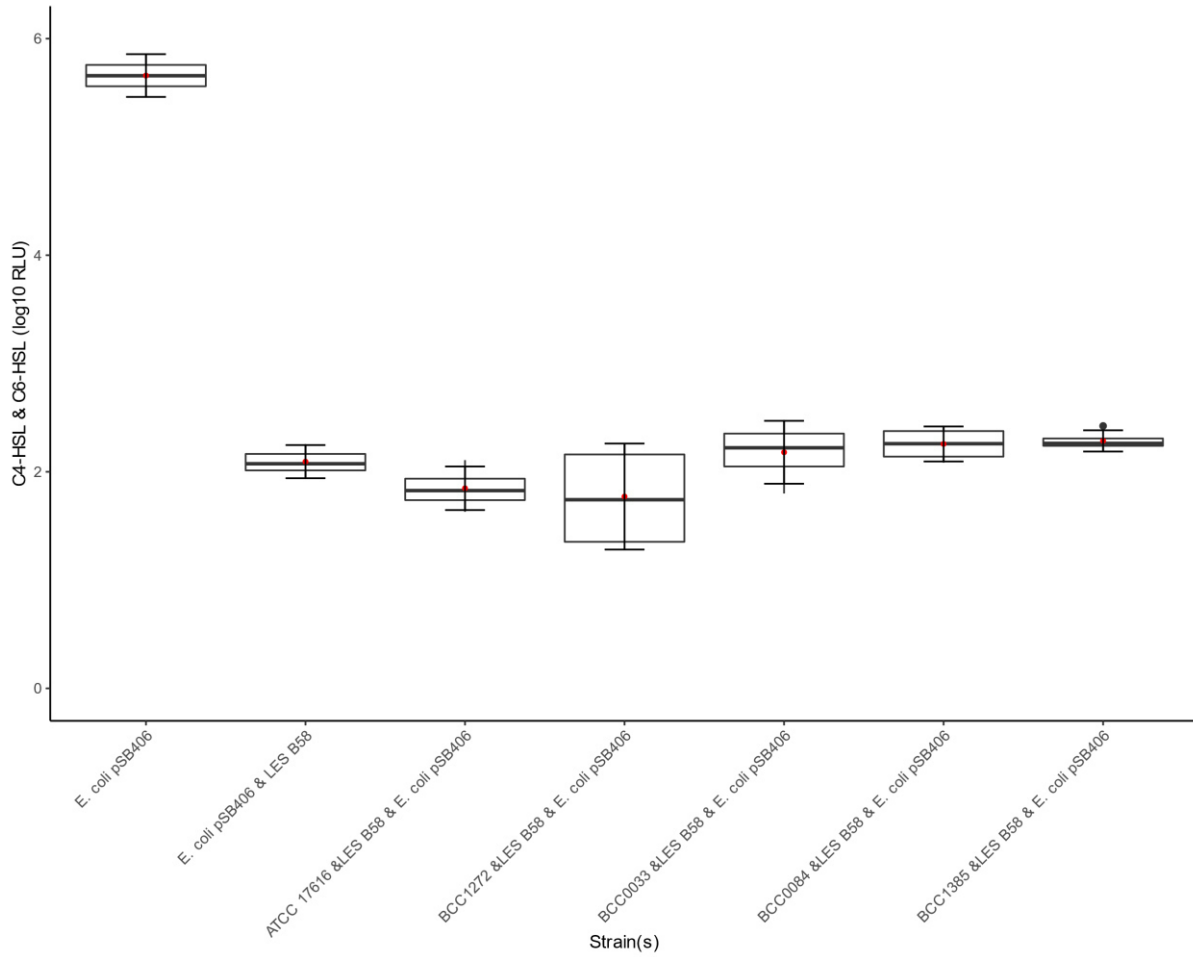


Figure 63 | C_4-HSL and C_6-HSL signals in the *B. multivorans* model strains co-cultured with *P. aeruginosa* LESB58 after 24-hours incubation. *E. coli* pSB406 used as a mono-culture control.

Figure 64 showed the 3-oxo-C_{12}-HSL RLUs produced by *P. aeruginosa* LESB58 and the *B. multivorans* interaction strains. Like that of the C_4-HSL and C_6-HSL, all the experimental strains had a lower $\log_{10}$ RLU than the *E. coli* pSB1142 reporter alone. All the *B. multivorans* co-culture interactions with *P. aeruginosa* LESB58 were lower in 3-oxo-C_{12}-HSL production than the *B. multivorans* strains alone (Section 5.2.8). Interestingly, the *P. aeruginosa* LESB58 strain alone had a lower 3-oxo-C_{12}-HSL RLU production than when interacted with *B. multivorans* (Figure 64). Overall, there were no significant differences when *P. aeruginosa* LESB58 was cultured alone or with any *B. multivorans* model strain for all QS HSL types.

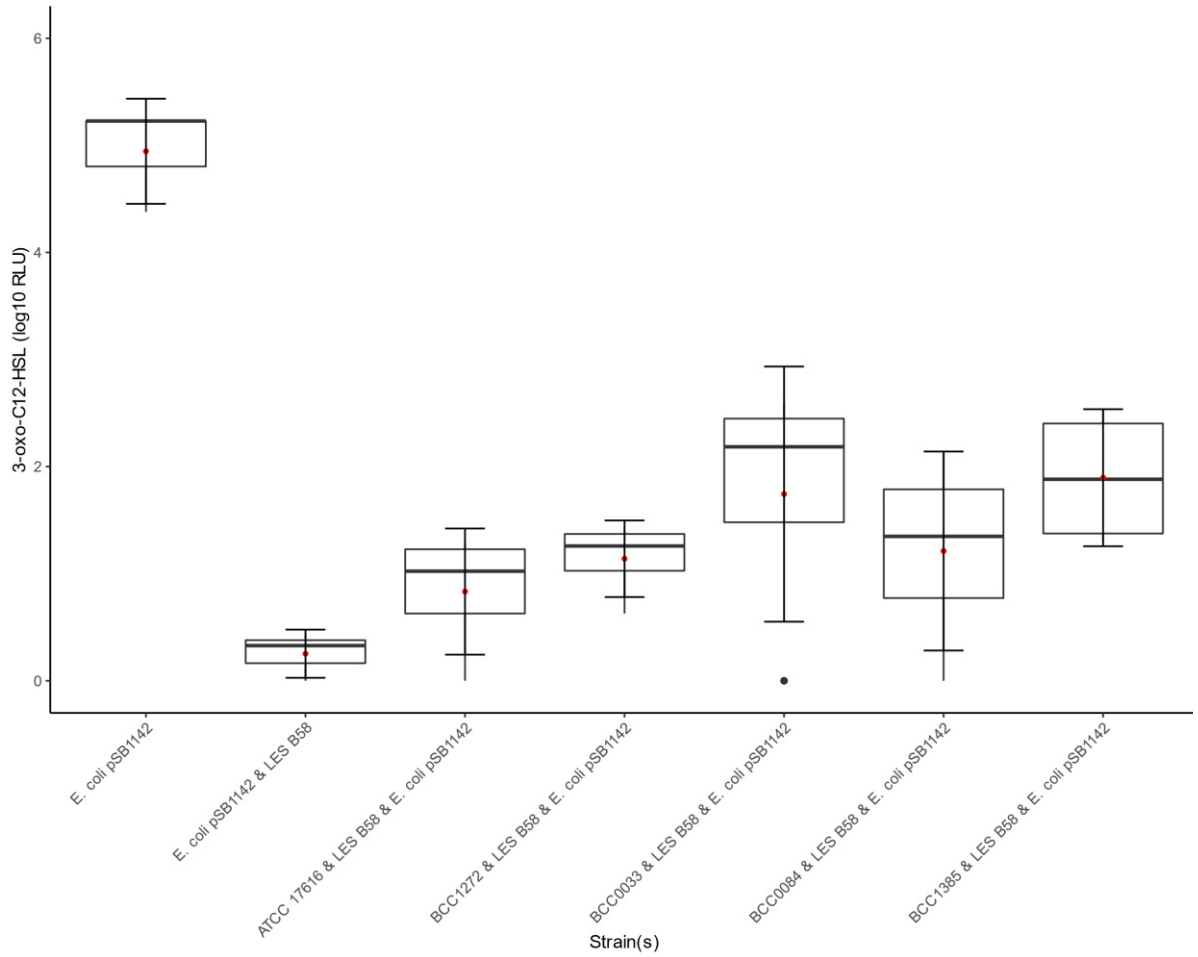


Figure 64 | 3-oxo-$_{C12}$-HSL signals in the *B. multivorans* model strains co-cultured with *P. aeruginosa* LESB58 after 24-hours incubation. *E. coli* pSB1142 used as a mono-culture control. Box plots represent the $\log_{10}$ RLU.

5.2.10 *Genes associated with phenotypic traits*

It is clear from the experiments in this thesis Chapter that the *B. multivorans* phenotype is highly variable overall. However, the phenotypes can be grouped into those which are motile (swarming motility was scored based on motility on at least one agar type), biofilm forming, have good growth rates and can produce exopolysaccharide (positive trait group) compared to those which do not present these phenotypes and have a slow growth rate (negative trait group). The totals for each of these assigned groups were as follows: swimming motility = 45 positive and 4 negatives; swarming motility = 39 positive and 10 negatives; growth rate = 38 positive and 11 negatives; biofilm formation = 46 positive and 3 negatives; exopolysaccharide production = 39 positive and 10 negatives. Scoary (Brynildsrud *et al.* 2016) was implemented to perform a pan-genome genome wide association study (GWAS) on the dataset. Those in the positive phenotype group were given the same binary trait, likewise for the negative phenotype group. This was to enable identification of genes which are

associated with that grouped phenotype. The output for each trait were manually filtered and combined to identify genes which were present in one group but not in the other.

The exploratory analysis showed that there were 73 unique genes present in the 'positive trait' phenotype group compared to 456 unique genes in the 'negative trait' phenotype group. Of these, 35 of 73 (48%) genes in the positive group and 309 of 456 (68%) in the negative group were hypothetical. Genes in the positive trait group which had the highest number of hits recorded were *actP_2* (cation/acetate symporter), group_3113 (*pbuR_3*) (purine efflux pump) and ttuB_12 (putative tartrate transporter), all associated with biofilm formation, which were present in 35 of 46 (76%) of the genomes. As for the negative trait group, replicative DNA helicase *dnaB_2* was the most common unique gene, present in 5 genomes, and is associated with swarming motility.

BCC0068 presented as a unique strain in the sense that it was completely non-motile (Section 5.2.3), non-biofilm forming (Section 5.2.55.2.9.3), did not produce exopolysaccharide (Section 5.2.6) and had a slow growth rate (Section 5.2.2.1). It was the only *B. multivorans* in the examined collection to portray this phenotype. Therefore, it was used to identify key genes which are both unique and absent from the strain when compared against the other examined 49 *B. multivorans* phenotype strains. From the Scoary output, there were 179 genes unique to BCC0068, of which 119 (66%) were hypothetical. There were 20 genes absent in BCC0068 which were present in 95-97% of the other strains. Of these genes, 7 (35%) were hypothetical and others spanned transcription factors, DNA repair, transporters, chemotaxis, secretion system genes, binding genes, and homeostasis genes.

5.3 Discussion

Phenotypic characterization of *B. multivorans* has been thoroughly investigated in this thesis Chapter. This encompassed biofilm formation, motility assays, exopolysaccharide and protease production, and antibiotic resistance. Whilst there have been some longitudinal phenotypic studies on *B. multivorans* in CF (Zlosnik *et al.* 2014; Silva *et al.* 2016), the overall phenotype for a representative *B. multivorans* strain collection is undefined. Therefore, this work is important in building the foundation for future work on the important *B. multivorans* pathogenicity phenotypes. This work showed that the *B. multivorans* strains proved to have highly variable phenotypes, with no specific correlation to genomic lineage, as certain traits like motility and biofilm formation were positive for most strains within the species.

Interaction of *B. multivorans* with secondary CF pathogens was also investigated, focusing on whether phenotypic shifts occurred during co-infection. There is currently limited knowledge on microbial interplay in the CF lung, particularly when *B. multivorans* is involved. The aim of this work was to examine the difference in growth, motility, biofilm formation and protease production during CF polymicrobial infection, highlighting pathogenesis. Polymicrobial interactions were very strain-

specific, with no clear-cut correlations overall. However, understanding the interplay between the different pathogens may enable predictions for disease progression and outcomes in future.

5.3.1 The growth rates of the *B. multivorans* collection splits into two broad groups

Over the 48-hour period, the *B. multivorans* strains split into two groups regarding their growth. The population was divided into those which reached stationary phase by 24-hours, and those which had a slower growth rate, reaching stationary growth by 30-hours (Figure 43). All the slow growing *B. multivorans* strains in this experiment were of CF origin. Reduced growth rates of *B. multivorans* strains has been previously shown by Silva *et al.* (2016), which were linked to chronic CF infection. Similar patterns were also observed in *P. aeruginosa* chronic CF strains (Rau *et al.* 2010). Clinical data would be required for determination of chronic infection isolates to growth rates. Following growth of *B. multivorans* sequential isolates may also be of interest for future studies.

Mixed microbial populations can grow in the same nutrient-rich media, where competition for resources occurs (Hibbing *et al.* 2010). In conditions where growth nutrients are sparse, there is an antagonistic effect on the microbial community (Harrison 2007). Modelling co-infections of *B. multivorans* and a secondary CF pathogen in TSB showed no difference in the growth rate parameters when a secondary pathogen was present. This contrasts results for *B. cenocepacia* and modelling interaction with *P. aeruginosa,* whereby *B. cenocepacia* growth was negatively affected (Bragonzi *et al.* 2012).

5.3.2 *B. multivorans* generally retains motility

Reduction and loss of motility in related CF pathogen *P. aeruginosa* has been well demonstrated in chronic infection (Mahenthiralingam *et al.* 1994). This phenomenon is not necessarily observed in BCC bacteria, though (Zlosnik *et al.* 2014). Swimming motility in BCC is driven by a single flagellum (Hales *et al.* 1998; Mahenthiralingam *et al.* 2005). The results in this thesis observed that *B. multivorans* was motile in most cases. This linked to work by Zlosnik *et al.* (2014) where only few non-motile *B. multivorans* were observed in a longitudinal study. Zlosnik *et al.* (2014) also demonstrated the ability of 14% of *B. multivorans* strains tested to swim across an entire agar plate. This was observed in one *B. multivorans* strain BCC0317 in this thesis. The clinical significance of the *B. multivorans* swimming motility phenotype being motile overall remains to be determined. However, in retaining motility, *B. multivorans* would be able to cause invasive infections as seen with the unique link of CF septicaemia ("*cepacia*" syndrome) (Isles *et al.* 1984) with all BCC species.

Swarming motility requires movement co-ordination, via methods such as quorum sensing, as well as surfactant production (Kearns 2010; Nickzad *et al.* 2015). Most of the *B. multivorans* strains in this work showed swarming capabilities on nutrient-rich TSA, like that observed in *B. cenocepacia* (Bernier *et al.* 2017). Swarming motility of *B. multivorans* was also altered by the presence of several secondary pathogens in this study. However, the mechanisms of each interaction remain unknown. A previously

researched swarming motility interaction was that between *P. aeruginosa* and *B. cenocepacia* (Bernier *et al.* 2017)which linked to the findings with *B. multivorans* in this work. Promotion of swarming in *B. cenocepacia* can be affected by rhamnolipids secreted by *P. aeruginosa* (Bernier *et al.* 2017). Rhamnolipids are produced by multiple bacteria, predominantly *Burkholderia* and *Pseudomonas species* (Abdel-Mawgoud *et al.* 2010; Irorere *et al.* 2017). Six of 11 (54.5%) of *B. multivorans* swarming motility interactions with motile *P. aeruginosa* PA14 had a significant increase in LB agar swarming diameter (Table 37). Assessment of the effect of *P. aeruginosa* rhamnolipids on *B. multivorans* is yet to be investigated. Product secretion, like rhamnolipids, of the other secondary CF pathogens and their effects on *B. multivorans* swarming abilities could also be investigated.

5.3.3 *B. multivorans* is an inherently good biofilm former

Another aim of this study was to understand biofilm formation across the *B. multivorans* strains both in mono- and co-culture with a secondary CF pathogen. Overall, the *B. multivorans* strains proved to have biofilm-forming capabilities in the 96-well microtitre plates, with only 7 of the strains forming weak biofilms. This corroborates previous *in vitro* research on BCC biofilm formation, helped by AHL communication (Conway *et al.* 2002). Bacterial biofilms are hard to eradicate and cause increased resistance to antibiotic therapies (Costerton *et al.* 1999; de la Fuente-Núñez *et al.* 2013). As *B. multivorans* are shown to form strong biofilms, therapies targeting their biofilms directly may be advantageous for future novel therapies, with the aim to prevent chronic CF lung colonisation.

In contrast, BCC bacteria may not always co-exist as biofilms in the same way as other CF pathogens, like *P. aeruginosa*. Schwab *et al.* (2014) proposed the idea that BCC bacteria *B. cenocepacia* and *B. multivorans* were more likely to reside in macrophages as single cells or clusters rather than as biofilms. This may reflect the different niche adaptations of the two pathogens. Whilst BCC bacteria can form *in vitro* biofilms, much information is still unknown about *in vivo* formation (Schwab *et al.* 2014). This poses the idea of building more *in vivo* BCC biofilm models, potentially by using *ex vivo* systems such as the porcine lung model (Harrington *et al.* 2020), to better mimic the CF lung.

In terms of mixed biofilms, only clinically relevant CF pathogens have been the focus of modelling. This includes BCC bacteria and *P. aeruginosa* (Riedel *et al.* 2001; Bragonzi *et al.* 2012). This work expands the current knowledge by interacting both clinical and environmental *B. multivorans* strains with other CF pathogens. Other work has shown that the presence of *B. cenocepacia* causes formation of a better biofilm when interacted with *P. aeruginosa* (Bragonzi *et al.* 2012). The results in this thesis, however, showed that there was no significant difference in biofilm formation when a secondary pathogen was present with *B. multivorans.* The difference in the experiments was that Bragonzi *et al.* (2012) utilised a shaking incubation whereas the work in this thesis was static. It has been shown in *Pseudomonas fluorescens* that static culture for biofilms reduces early adhesion, but more effective growth overall (Jara *et al.* 2020). The use of shaking would disrupt the bacterial dispersal, potentially

affecting biofilm formation at varying time points, as previously observed in *P. aeruginosa* PA14 (Quinn *et al.* 2021). Therefore, single, and mixed culture biofilm experiments could be extended to 48 and 72-hours, comparing static and shaking growth to see if this has an effect.

5.3.4 The antibiotic susceptibility patterns of *B. multivorans* have slight variation

The intrinsic antibiotic resistance nature of BCC bacteria is well known, providing limited treatment options in clinical settings (Rhodes and Schweizer 2016). *De novo* resistance also occurs in the BCC through efflux, beta-lactamases, and outer membrane permeation (Gautam *et al.* 2020). By understanding the antibiotic susceptibility patterns of *B. multivorans* and other BCC bacteria, there is more opportunity to make effective choices for treatment. This is important as people with CF will have to continuously take antibiotics, predisposing them to antibiotic resistant bacterial strains (Coutinho *et al.* 2008). Based on the findings, tetracycline (100 µg) is an effective antibiotic against the *B. multivorans* panel, as all strains were susceptible. Literature has shown tetracycline to be less effective against other species of BCC bacteria (Nzula *et al.* 2002; Somprasong *et al.* 2021) compared to this work, with only few susceptible strains (Lu *et al.* 1997; Nzula *et al.* 2002). The *B. multivorans* ATCC 17616 strain has been shown to have a lower tetracycline MIC (greater susceptibility) compared to other BCC species and *P. aeruginosa* strains (Nzula *et al.* 2002; Malott *et al.* 2014). The reason for this may be the difference in concentration used between the two studies. In this thesis, 100 µg tetracycline was used, compared to 30 µg used in other studies against *Burkholderia* species (O'Carroll *et al.* 2003; Omar *et al.* 2015).

Other potentially effective antibiotics would be chloramphenicol (50 µg), kanamycin (30 µg), and nalidixic acid (30 µg). Only one strain (BCC0375) (2.2 %) was completely resistant to chloramphenicol. There were 6 (13%) and 4 (8.9 %) *B. multivorans* strains completely resistant to kanamycin and nalidixic acid, respectively. Chloramphenicol susceptibility in the *B. multivorans* strains was interesting, as *Burkholderia* species, such as *B. cenocepacia*, have previously been shown to harbour efflux pumps which confer resistance to this antibiotic (Tseng *et al.* 2014; Podnecky *et al.* 2015). Efflux pumps in *B. multivorans,* though, are less understood. Chapter 4; Section 1.1.1.3 showed that genes *ceoA* and *ceoB* was present in all the draft *B. multivorans* genomes tested. This gene is part of the CeoAB efflux system which confers resistance to chloramphenicol (50 µg) in *B. cenocepacia* (Burns *et al.* 1996; Nair *et al.* 2004). Despite the presence of these two genes, the results suggest that there is an alternative reason for antibiotic susceptibility, potentially through collateral sensitivity (CS) (Flanagan *et al.* 2020). Kanamycin is a broad-spectrum aminoglycoside antibiotic which targets protein synthesis, specifically binding to the A-site 16S rRNA on the 30S ribosome (Kotra *et al.* 2000). The mode of action of kanamycin is like that of streptomycin, interfering with protein synthesis by directly binding to the bacterial ribosomal 30S subunit. Interestingly, streptomycin was deemed not to be suitable for use against *B. multivorans* due to the high levels of resistance displayed amongst the examined strains.

Nalidixic acid, another proposed antimicrobial for BCC treatment from this study, is an antibiotic often used to treat urinary tract infection (UTI). The *bcrA* gene, comprising part of an MDR efflux pump, in related *B. cepacia* has previously been shown to increase resistance by four-fold to nalidixic acid in four UTI strains tested (Wigfield *et al.* 2002).

In contrast, ampicillin, colistin sulphate, and nitrofurantoin would be ineffective in treating *B. multivorans* infection due to the high levels of resistance displayed in most strains. It must also be considered that the disc diffusion assay in this work used antibiotics targeted against clinical urinary tract infections (UTIs). For instance, including additional antibiotics such as ceftazidime plus an appropriate inhibitor (Tamma *et al.* 2018; Van Dalem *et al.* 2018), minocycline (Zhou *et al.* 2007), carbapenems, meropenem, and trimethoprim (Tseng *et al.* 2014) should be an expansion of the current study to mirror clinically used antibiotics.

There are no current EUCAST breakpoints for the BCC, meaning that determination of strains into the resistant, intermediate, and susceptible categories for each antibiotic were not performed. The alternative is to use the pharmacokinetic/pharmacodynamics (PK/PD) breakpoints instead (Van Dalem *et al.* 2018). However, these were also deemed unsuitable for this work due to the antibiotics used. Whilst disc diffusion assays are ideal for mass antibiotic screening of bacterial strains, MICs using microbroth dilution assays, should be used to determine the exact concentration of antibiotic for effectiveness against the strain. Whilst it is a more laborious task, it is currently the 'gold standard' parameter for understanding antibiotic effectiveness, and thus improving therapy success for an individual (Kowalska-Krochmal and Dudek-Wicher 2021). MICs may also pinpoint collateral sensitivity (CS) and cross-resistance (CR) in *B. multivorans.* CS occurs when the bacterium increases sensitivity to an unrelated antibiotic due to gaining resistance to another. In contrast, CR is when resistance increases to other drugs which are not used for treatment (Kavanaugh *et al.* 2020). Understanding CS and CR increases the chance of effective treatment combinations.

This work solely investigated the antibiogram of *B. multivorans* strains. As we know, CF communities are polymicrobial. Therefore, it would be advantageous in future to analyse the effect of a secondary pathogen on the antibiotic resistance and susceptibility patterns of *B. multivorans.* Modelling polymicrobial communities with *B. multivorans* could pinpoint certain antibiotics which increase or decrease in susceptibility when a specific secondary pathogen is present. One issue is that there is currently little understanding on the response of polymicrobial CF infection to antibiotic treatment (Rogers *et al.* 2010).

5.3.5 *B. multivorans* reduces *P. aeruginosa* LESB58 protease activity

All *B. multivorans* strains were protease negative on lactose-free skimmed-milk agar with both BHI agar and TSA base. This correlated with previous research which found *B. multivorans* did not harbour the *zmpA* zinc metalloprotease gene (Gingues *et al.* 2005). An alternative protease agar test utilises a

similar approach to the motility assays, where bacterial inoculation is performed using surface inoculation by a toothpick. Protease is assessed by measuring the zone of clearing around the inoculum (Burtnick *et al.* 2014). This may allow for more accurate measurement of protease production, as it would omit the need to subtract the size of the inoculation spot from the overall zone of clearing. Both assays are only semi-quantitative so there are limitations. Whilst both agar assays are good for mass screening, the use of the proteinase K assay may provide a more accurate result for overall total protease production for the bacterial strains (Harrison *et al.* 2014).

This work determined that protease negative *B. multivorans* could suppress the protease production of *P. aeruginosa* strain LESB58 when interacted at equal starting viability. In the CF lung, it is unlikely that the same ratio of *B. multivorans* and *P. aeruginosa* will be present at any one time (Bragonzi *et al.* 2012), although they do co-occur in CF lung infection (Flight *et al.* 2015). Further experiments could include examining the protease production of *P. aeruginosa* LESB58 when mixed at different starting concentrations with *B. multivorans.* For instance, Bragonzi *et al.* (2012) showed that *P. aeruginosa* was dominant in early CF infection, which was overtaken by BCC bacteria by the end-stage. It would also be interesting to study whether *B. multivorans* influences an already established chronic *P. aeruginosa* LESB58 community using animal models. This could be performed using both chronically established, transient and epidemic *B. multivorans* CF strains for comparison. However, further clinical data would be required for this.

5.3.6 *Preliminary data for quorum sensing in B. multivorans and P. aeruginosa*

AHL-dependant communication, using QS, has previously been described in the BCC (Sokol *et al.* 2007). This study included experimentation of *B. multivorans* and *P. aeruginosa* LESB58 AHL secretion using two *E. coli* bioluminescence reporter plasmids (pSB406 and pSB1142). Evidence showed that the normal *B. multivorans* phenotype lacked AHL production, linking to previously published observations (Gotschlich *et al.* 2001). This finding extended to *P. aeruginosa* LESB58, which also was shown to be AHL-negative in the assay. It has previously been shown in *P. aeruginosa* that a loss of QS 3-oxo-C_{12}-HSL may also cause a reduction in protease production (Harrison *et al.* 2014). Testing for QS interactions between *B. multivorans* and *P. aeruginosa* LESB58 were also performed in this study. However, it does not appear that *B. multivorans* negatively effects QS of *P. aeruginosa*, and thus the reduction in protease production must be due to an alternative mechanism.

The assay performed in this study used overnight cultures for the experimentation. However, this may have caused interference of signals and interactions between the bacterial cells. Therefore, cell-free supernatant should be used in future, as described previously (Harrison *et al.* 2014; Abdul Malik *et al.* 2020). This would be straight-forward for the single cultures, but for the co-culture could be performed in two ways: 1) growing the overnight culture as a mixture before use or 2) growing the cultures separately and then mixing for the assay. The first option would be more useful if wanting to check

AHL production at one time point as the bacteria would have been interacting for the given time frame. Option two would be better for a growth assay where cultures would be then mixed and AHL production checked at given intervals in a plate reader.

The experiment was also performed using a 30-minute incubation of the test strains with the *E. coli* reporter, providing a reading for secreted AHLs at a single time point. For more accurate analysis, the *B. multivorans* and *P. aeruginosa* strains could be grown both alone and in a mixed interaction and monitored for luminescence over a 24-hour period, for example. A similar approach has been performed by Abdul Malik *et al.* (2020) and Fletcher *et al.* (2007). This would enable visualisation of AHL production levels at different time points.

Ultimately, the assay would also need a true positive control. The *E. coli* reporter strains have been used as negative controls as they do not produce AHLs but have a background luminescence. Purified AHLs, which would be interactive with the reporter, would be advantageous in future to show the activation of the reporters. Therefore, no thorough conclusions can be made about QS and AHLs in *B. multivorans* and the *P. aeruginosa* interaction assay from this data output. However, this could be overcome through optimisation of the assay, which was a time constraint in this study.

5.4 Conclusions

Following the population biology and pan-genome analysis of the *B. multivorans* strains, the following conclusions can be made about phenotypic traits of the species:

1) The *B. multivorans* phenotype does not directly correlate to its genomic lineage
2) As a species, the majority of *B. multivorans* strains are motile and good biofilm formers
3) *B. multivorans* does not produce proteases on lactose-free skimmed-milk agar
4) *B. multivorans* influences the protease production of *P. aeruginosa* LESB58 in co-culture on a lactose-free skimmed-milk agar assay
5) Swimming motility of secondary CF pathogens can be influenced by the presence of motile *B. multivorans* strains

6 Conclusions, general discussion, and future research

6.1 Conclusions

This work presents a novel and comprehensive understanding of the population biology and pathogenic phenotype of the CF lung pathogen *B. multivorans.* A collection of *B. multivorans* strains with limited previous characterization but known genetic diversity (at MLST-level), was subjected to whole genome sequencing, and analyzed at the population-level through SNP variant and core-gene phylogenomic analysis. This identified the diversification of *B. multivorans* as a species into two major genomic lineages. These were designated lineage 1 and lineage 2, with lineage 2 having at least two further sub-groupings and greater genomic diversity. The *B. multivorans* strain panel examined multiple isolation sources, but with a strong focus on CF isolates (241 of 283 strains: 85.2%). Four model *B. multivorans* strains were identified, representative of genomic and phenotypic diversity within the collection. The *B. multivorans* model strains were subjected to complete genome sequencing using PacBio long-read analysis to enable further genome analysis and comparative analysis. Interrogation of all available *B. multivorans* genome sequences, including the complete genomes from the model strains, provided a foundation for diagnostic gene identification and *B. multivorans* lineage-specific PCR primer design. Phenotypic characterization of *B. multivorans* included growth, motility, biofilm formation, cell surface differences, and protease production. This phenotypic analysis demonstrated key *B. multivorans* species traits, but diversity was strain dependent. This Section of work included experimental co-infection modelling of the change in *B. multivorans* phenotype when an additional secondary CF pathogen is present. The *B. multivorans* model strains were able to infect and persist at high levels in the murine respiratory tract over the course of at least 5 days. This presented a basis for future infection modelling and therapeutic testing of *B. multivorans.*

The main conclusions from this work were:

6.2 The phylogenetic analysis of the B. multivorans strains highlighted a clear divide in the population, with at least two distinct genomic lineages (Chapter 3)

This work is one of very few examples where the *B. multivorans* population has been studied. It is the first time the two-lineage split in the *B. multivorans* population has been identified at whole-genome level. A representative *B. multivorans* strain panel was also identified in this work, encompassing both CF and non-CF isolates from an array of different countries worldwide. Initially, ANI was used to check that all strains used in the analysis were of the same genomic species. This was the first instance where the two-lineage divide was observed. The variance of ANI of lineage 1 around the mean (98.90%) was less than that of lineage 2 (mean = 98.43%). This provided evidence that *B. multivorans* strains which resided in lineage 1 were more closely related than that of lineage 2. This was further backed by phylogenomic analysis. A RAxML phylogeny was built on 2,998 core genes from 283 genomes which

showed longer and deeper branches in lineage 2 genomes compared to those in lineage 1. Within lineage 2, there were a further 2 sub-lineages identified. There was also a potential 3rd lineage, harbouring CF strain BCC1368 sequenced in this thesis, as well as 3 strains downloaded from public databases. The ANI of this 3rd group (97.5%), compared to *B. multivorans* reference strain ATCC 17616, still supported its conclusion within the *B. multivorans* species. This was not defined as a separate lineage in this work and was instead used as an outgroup for rooting phylogenies due to its relatedness to *B. dolosa*. Variant analysis devolved deeper into the core-SNP phylogenies of the two lineages. The clear split of the two sub-lineages in lineage 2 remained during this analysis. Lineage 1 in turn presented a phylogeny showing 3 distinct clades at the SNP level. This *B. multivorans* population biology will set the benchmark for future systematic testing of phenotypic and clinical association between genomic lineage.

6.3 There are lineage-specific genes in the B. multivorans population which can be used to design clinically relevant PCR primers (Chapter 3)

Comparative genomic analysis was used to identify genes unique to each *B. multivorans* lineage. Four *Candida*te genes were identified for PCR primer design. These genes were based on a gene presence-absence approach and checked for genomic stability. All three genes identified within lineage 1 (yiaJ_1, ghrB_1, naiP_3) were present in 100% of the *B. multivorans* lineage 1 genomes. The glnM_2 gene target identified within lineage 2 was only present in 99.1% of the lineage 2 genomes. However, glnM_2 was deemed fit for PCR design as it was a functionally annotated gene with the most hits overall across the entire available genomic collection, and the two genomes the gene was missing in were not of CF isolate origin. Testing of the PCR primers in the laboratory showed that they were species-specific within a limited panel of BCC species and had the ability to discriminate between the two genomic *B. multivorans* lineages. Tests were performed using all the *B. multivorans* phenotypic strain panel which showed 100% specificity. These are the first primers designed to specifically target one of the two *B. multivorans* genomic lineages.

6.4 Four model CF B. multivorans strains have been identified and analyzed for pathogenesis using murine models (Chapters 3 and 4)

To date, there have been no model *B. multivorans* strains identified for studying CF lung infection. The work identified 4 model strains, based on their phylogenetic placement and genetic content, 3 of which were CF strains and sequenced as part of this thesis. BCC0084 was chosen from lineage 1, BCC1272 and ATCC 17616 from lineage 2a, and BCC0033 from lineage 2b. Despite being an environmental isolate, ATCC 17616 was selected as a relevant model strain. This was because of its isogenic relationship with CF isolate BCC1272 as well as its well-studied nature in the context of the published

literature. All model strains were also subject to basic genetic manipulation, to see if they could be transformed with fluorescent gene reporter plasmids. Uptake and stability were observed in all strains when using pIN301-eGFP. For the pIN233-mCherry construct, *B. multivorans* strains ATCC 17616 and BCC1272 were unable to be transformed with this plasmid, but the remaining strains demonstrated good amenability.

The model strains were also used to identify the location of virulence genes in the complete genomes *in silico*. Pathogenic mechanisms of *B. multivorans in vivo* remains elusive. Therefore, a murine lung inhalation model was used to study the model strains over a period of up to 5-days. This employed both the wild type strains and genetically manipulated fluorescent reporter tagged strains. All the 3 *B. multivorans* strains (BCC0033, BCC0084, and ATCC 17616) had the ability to colonize both the murine lung and nasopharynx at all time points. However, varying degrees of infection burden were observed between the strains. BCC0084::pIN301-eGFP had the greatest colonization at each time point, compared to ATCC 17616 WT which was the weakest colonizer in terms of overall murine respiratory tract microbial burden. Whether this is lineage-specific or attributed to isolation source remains inconclusive due to the limited strain sample size. Variant analysis was performed to identify key genes which may indicate niche adaptation. No genes were identified consistently throughout the short-term experiment. However, this successful infection modelling has created the possibility of studying *B. multivorans* evolution during chronic CF infection. This includes testing of novel therapeutics in *B. multivorans* chronic lung infection.

6.5 The phenotype of the B. multivorans strains does not necessarily reflect genomic lineage (Chapters 4 and 5)

Comparing strain phenotype to the population biology of *B. multivorans* had not previously been performed before this work. A key study hypothesis was the assumption that genomic lineage may affect the phenotype of the *B. multivorans* strains. However, this was proven not to be the case. The overall phenotype was highly variable between the strains. As a species, *B. multivorans* is motile in terms of swimming and swarming, can form biofilms, produces exopolysaccharide, and is non-protease producing. *B. multivorans* strain BCC0068 was noted as an outlier in the analysis. This strain was slow growing, completely non-motile, non-biofilm forming, and non-mucoid. Using a pan-GWAS approach, 20 genes were identified as missing from BCC0068 compared to the rest of the *B. multivorans* phenotype strain panel. These genes were associated with DNA repair, transporters, chemotaxis, secretion systems, binding genes, and hemostasis. Comparison of isolation source to phenotype was limited due to the small numbers of non-CF and environmental isolates used in this work.

Exopolysaccharide production was the only phenotype where isolation source was linked to a suggestion of an altered phenotype. Interestingly, lack of exopolysaccharide production (nonmucoid phenotype) was only observed in CF isolates (n = 5), meaning that this could be associated with

pathogenicity. Previous research by Zlosnik *et al.* (2011) showed that a more rapid clinical decline in persons with CF was linked to a nonmucoid BCC bacterial phenotype. This BCC phenotypic switch from mucoid to nonmucoid (Zlosnik *et al.* 2011) in chronic CF infection is inverse to that observed in *P. aeruginosa* (Govan and Deretic 1996). Virulent *B. cenocepacia* has been shown to mostly be nonmucoid when assessing clinical isolates (Zlosnik *et al.* 2008). Interestingly, mucoid variants of *B. cenocepacia* have also been shown to persist in the murine CF lung longer than their nonmucoid counterparts (Chung *et al.* 2003). As all the *B. multivorans* model strains produce EPS on YEM agar, further comparison of these strains in murine models to nonmucoid *B. multivorans* strains could be performed. This would help to identify if differences in EPS production links to chronic colonization in *B. multivorans*.

6.6 B. multivorans strains can suppress the protease production of epidemic P. aeruginosa LESB58 when using the skimmed-milk agar assay (Chapter 5)

To date, there have been a couple of studies looking at co-culture modelling of two CF lung microbiota pathogens. However, these are limited to more common interactions such as *P. aeruginosa* with *S. aureus* (Baldan *et al.* 2014; Biswas and Götz 2021), or *P. aeruginosa* and *B. cenocepacia* (Bragonzi *et al.* 2012). Phenotypic modelling of *B. multivorans* with other CF pathogens is an area wide-open for further research. This work employed the phenotypes used to characterize *B. multivorans* alone to see what would happen when a secondary pathogen was introduced. There was no significant difference in biofilm production, and differences in motility were negligible. This work provided evidence that motile *B. multivorans* strains could "carry" non-motile secondary CF pathogens across an agar surface. Protease interactions had a more clear-cut conclusion. Protease negative *B. multivorans* had the ability to suppress the protease production of *P. aeruginosa* LESB58 on 1% lactose-free skimmed-milk agar. The statistically significant reduction in protease production was found to be unique to the epidemic *P. aeruginosa* LESB58 CF strain (Cheng *et al.* 1996; Winstanley *et al.* 2009), no matter which *B. multivorans* strain was used for the interaction. It was thought that *B. multivorans* was either killing or outgrowing the *P. aeruginosa* LESB58 strain in the co-culture. However, this theory was dismissed by using the fluorescent gene reporter probe strains over a 24-hour growth period. This means that *B. multivorans* can suppress *P. aeruginosa* LESB58 protease production via another mechanism other than killing which is worthy of further study.

6.7 Discussion and Future Research

Studies regarding population biology and phylogenomic analysis of the BCC are limited. Currently, *B. multivorans* epidemiology has only been subject to MLST-level analysis (Baldwin *et al.* 2008). Genomic studies of single (Silva *et al.* 2016; Diaz Caballero *et al.* 2018) and limited *B. multivorans* strain collections (Peeters *et al.* 2017; Lood *et al.* 2021) have also been performed, but not as large-scale such as performed in this thesis. There remains little knowledge on the virulence factors and

pathogenic mechanisms of *B. multivorans* as well as the understanding what may cause chronic CF lung colonization. There was a clear divide identified in the population, whereby *B. multivorans* strains were noted as lineage 1 and lineage 2. These two genomic lineages became the focus for understanding any phenotypic and genotypic differences in the *B. multivorans* population. This Section will discuss the relevance of these findings and their utility for future *B. multivorans* research in the context of CF lung infections.

6.8 Comparing population biology and associated virulence

With the constant increase in data availability of bacterial genomes, it has been much easier to perform in-depth analysis of bacterial population structures. This has been attributed to more cost-effective and accessible genome sequencing technologies in the past few years. This work aimed to understand virulence and pathogenicity gene traits in the *B. multivorans* genomes and their association with genomic lineage. Initially, genes were grouped from the Roary gene presence-absence matrix. Total group counts were performed for each *B. multivorans* genomic lineage and statistically compared. Lineage 1 had a greater gene count for biofilm-associated genes and hemolysins. Lineage 2 had a greater number of genes associated with biocide/chemical resistance, cation efflux, drug resistance, invasion, and flagella, OMPs, QS genes, stress proteins and T6SS genes. Virulence genes are often associated with regulation by QS systems. Studies showing the absence of two QS systems in *B. cenocepacia*, located within the *cenocepacia* genomic island 1, caused reduction in strain virulence (Baldwin *et al.* 2004; Malott *et al.* 2005). Genomic islands (GIs) have not been investigated within this work, however, *B. multivorans* has previously been shown to harbour multiple GIs within their genomes, located on multiple replicons (Lood *et al.* 2021). Further investigation of *B. multivorans* GIs and their content may highlight further abilities for adaptation to the CF lung.

Several virulence studies have also been performed on related *B. cenocepacia*, where some of the gene groups observed here have been associated with overall pathogenicity. Increased transcription of flagella (Drevinek *et al.* 2008b) and increased expression of adhesin AdhA (Sajjan *et al.* 2003), have both been observed to have key roles in *B. cenocepacia* CF pathogenesis, such as both cell attachment and invasion. T6SS have also been found in many *Burkholderia species*, but its contribution to virulence is not necessarily associated with host infection. Instead, T6SS are implicated in competition with other bacterial species, using antibacterial properties to outcompete others in that niche (Spiewak *et al.* 2019). The greater presence of T6SS genes in lineage 2 supports the idea that this *B. multivorans* lineage may be better at out-competing other CF lung pathogens than lineage 1. Overall, this would require further analysis both *in vivo* and *in silico* to rule out any analysis bias, due to the sheer proportion of *B. multivorans* strains in lineage 2 compared to lineage 1.

A study by Wallner *et al.* (2019) attempted to link *B. cenocepacia* isolation source to phylogenetic placement and virulence gene content. The results showed that the clinical isolates had a higher number of putative virulence genes than the environmental strains (Wallner *et al.* 2019). However, this work was deeply flawed because the collection included both strains which were previously shown to cause problems in CF individuals. The work also included isogenic environmental isolates of these problematic CF strains (Wallner *et al.* 2019). *B. cenocepacia* may also lose virulence determinants, but this does not necessarily mean that the strain could not cause infection in a vulnerable host. For instance, CF isolate *B. cenocepacia* J2315 has been shown as unable to produce EPS but is still representative of a problematic epidemic BCC strain linked to mortality in people with CF (Drevinek and Mahenthiralingam 2010). Different responses to the CF lung environment may also play a role in the virulence of each *B. cenocepacia* strain, despite residing within the same genomic lineage (Drevinek *et al.* 2010; Drevinek and Mahenthiralingam 2010). This continues the idea that each *Burkholderia* infection outcome is highly strain specific and dependent on host vulnerability in terms of opportunistic infection.

A second analysis was performed using Abricate (Seemann 2017) databases to highlight key genes which were then mapped to the complete *B. multivorans* model strains. The overall virulence genes identified from Abricate were relatively conserved across the *B. multivorans* strains tested, and places of presence or absence were not always related to genomic lineage. One limitation we currently face is the lack of clinical outcome data behind the *B. multivorans* strains examined in this study. Further understanding the clinical outcomes of each *B. multivorans* lineage would help in the endeavor to understand whether the virulence genes identified would affect overall pathogenicity. This work also did not have access to whether the strains used were transient or chronic in the context of long-term CF lung infection. This means that links could not be made between presence and absence of specific genes and chronic infection, except for what was observed in the 5-day murine experiment. In future, knock-out *B. multivorans* mutants could be used to identify the effect of key virulence genes on organism pathogenicity.

6.9 Lineage-specific PCR primers for identification of B. multivorans in clinical samples

There has now been much molecular PCR development for *Burkholderia* species identification (Mahenthiralingam *et al.* 2000; McDowell *et al.* 2001; Cesarini *et al.* 2009; Dedeckova *et al.* 2013). Many of the current PCR primers are used to target phylogenetic markers such as 16S rRNA (LiPuma *et al.* 1999) and *recA* (Mahenthiralingam *et al.* 2000). This work has taken that one step further, identifying unique lineage-specific genes as PCR targets. These are PCR primers which could be used for routine testing in clinical microbiology laboratories and start to allow linkage of lineage to clinical disease. The primers were initially designed for use on the assumption that isolated and purified *B. multivorans* strains would be tested, and that only knowledge of the specific lineage was required.

However, the analysis herein proved their ability to differentiate from *B. cenocepacia* and *B. ambifaria,* as well as the opposing *B. multivorans* lineage. With more than 22 species in the BCC (Mullins and Mahenthiralingam 2021), further testing will be required for species specificity of the PCR primers. The laboratory analysis showed that there was 100% specificity for the targeted lineage, with a single 744 bp band for lineage 1 and 322 bp for lineage 2. This means there is also the ability to combine the two primers into a multiplex PCR due to the differentiation between the banding patterns.

The efficacy of these lineage-specific primers still requires testing in routine clinical *B. multivorans* surveillance. Even though the study was limited to the *B. multivorans* strain collection at Cardiff University, whereby the isolates have already been purified into stocks, it is expected that the primers will still be highly specific to other circulating CF isolates. One limitation of this work is that the primers could not be tested directly on sputum samples, and therefore would need to be checked for sensitivity in this manner. Development of this molecular detection method into a nested PCR would also enhance the ability to increase sensitivity. A similar PCR design, for BCC detection direct from patient CF sputum, was created by Drevínek *et al.* (2002) utilizing the *recA* gene. In all, these primers will reduce costs and diagnostic time. With the later addition of clinical data for the associated lineage, there may also be better understanding of clinical outcome and infection trajectory.

6.10 Use of complete genome sequences and CF model strains to extensively characterize the genomic content and pathogenicity of B. multivorans

By obtaining complete genomes of the 4 *B. multivorans* CF model strains (ATCC 17616; lineage 2a, BCC0033; lineage 2b, BCC0084; lineage 1, and BCC1272; lineage 2a), comprehensive functional analysis and comparative analysis between the lineages could be performed. These 4 *B. multivorans* strains now provide characterized references for future studies, spanning both genome sequencing and phenotypic work. These complete genomes were determined using PacBio sequencing and polished with Illumina short DNA reads. All four complete genomes provided 4 genomic replicons: C1, C2, C3, and C4. The BCC are known to have a multi-replicon structure (Mahenthiralingam *et al.* 2005). However, the 4th replicon does not appear to be in all BCC strains, and not present in all *B. multivorans*. The results in this work provided evidence that the C4 replicon is a plasmid, which varies between strains. Future research should examine the advantages of harbouring this plasmid, and whether the loss of this plasmid has any effect on strain virulence. For example, replicon C3 is known as a non-essential virulence plasmid (Agnoli *et al.* 2014).

Study of microbial evolution is a powerful tool in understanding changes in genetic content based on niche selection pressures (McDonald 2019). As part of this work, murine inhalation models were used to determine the infection burden and pathogenicity of the model CF *B. multivorans* over a 5-day period *in vivo*. As noted in the introduction, there are a wide variety of models which have previously been used to assess BCC pathogenicity. A limitation of using murine models for infection is that they require

ethical consideration in the context of appropriate use and are very costly. This meant that the work could only be repeated once, for a limited number of strains. However, this is one of the more advanced and well-documented models for use in CF (Bragonzi *et al.* 2012; Scoffone *et al.* 2017) and other pathogen (Speert *et al.* 1999; Green *et al.* 2021) respiratory infection modelling. Using this murine model for the first time on *B. multivorans* represents a key advance in this thesis. *B. multivorans* as a species demonstrated a much greater infection burden than BCC bacteria *B. ambifaria* (Mullins *et al.* 2019). All the 4 *B. multivorans* strains colonized the murine respiratory tract, however, full understanding of pathogenicity between the strains and the lineages requires further study. A good alternative to murine models is *the G. mellonella* wax moth larvae model. Use of this invertebrate model is much cheaper than mice and the pathogenicity results of the BCC are in line with those observed in mice (Seed and Dennis 2008). There is also no requirement for animal handling training, and an LD_{50} value can be ethically identified using this model, but not in murine infections (Seed and Dennis 2008). For a larger, high-throughput pathogenicity screen of the *B. multivorans* strains, this model should be considered in future experiments.

6.11 Co-infection modelling to understand pathogen interactions in the CF lung microbiome

It is not clear that the chronically infected CF lung has multiple inhabiting microbial species within it. The CF microbiota is highly complex, where multiple CF microbes create communities which interact with one another via an array of mechanisms These microbial interactions remain poorly understood. During this work, there was one interaction of interest, the suppression of *P. aeruginosa* LESB58 protease production when *B. multivorans* was present. The initial hypothesis that *B. multivorans* was killing *P. aeruginosa* was rejected. The mechanism of how protease production is suppressed is open for experimentation. Optimization of an RNA sequencing (RNA-seq) protocol was evaluated during this study but was not fully implemented due to time and costing constraints. The two CF pathogens were grown in both mono and co-culture on a filter membrane placed on the protease agar for specified time-points (16, 18, and 24-hours). Good quality RNA could be extracted from all time points. This provides a basis for future experimentation through transcriptomics to evaluate the global gene expression in *B. multivorans* and *P. aeruginosa*. This strategy could have revealed gene expression linked to the suppression of protease production in *P. aeruginosa*, as well as that due to the interaction of the two bacteria in co-culture.

Interestingly, *P. aeruginosa* has previously been shown to enhance BCC virulence factor expression through AHL signals (Riedel *et al.* 2001). This proposes the idea that RT-qPCR may be useful for experimentation on the *B. multivorans* and *P. aeruginosa* co-cultures. This could look at specific QS gene expression changes when two CF pathogens are grown together and relate this to possible virulence gene expression. Similar work has been performed when *P. aeruginosa* is interacted with *S.*

aureus. Different strain combinations either co-exist, altering *S. aureus* virulence, or *P. aeruginosa* inhibits the growth of *S. aureus* through competition (Briaud *et al.* 2019). Use of appropriate animal models could also be used for co-infection. This would help overcome the limitation of using *in vitro* techniques which may not necessarily reflect the nature of the CF lung microenvironment. An example of this has been performed by Bragonzi *et al.* (2012) where *B. cenocepacia* and *P. aeruginosa* biofilms were modelled in a murine co-infection (both inbred and CFTR mice) over a 13-day period.

6.12 Final perspectives

Overall, this PhD thesis research used a multi-disciplinary approach, using both computational and laboratory-based techniques. By defining the population biology, pathogenic phenotype, and potential *in silico* virulence factors, this work has opened the door for further in-depth research into differences that may be present between the two *B. multivorans* genomic lineages. The work performed in this thesis provided a highly data-rich output. Finally, this work has provided a strong foundation for rapidly identifying *B. multivorans* lineage in a clinical setting using lineage specific diagnostic PCR.

7 References

Aaron, S. D., Ferris, W., Henry, D. A., Speert, D. P. and Macdonald, N. E. 2000. Multiple combination bactericidal antibiotic testing for patients with cystic fibrosis infected with Burkholderia cepacia. *Am J Respir Crit Care Med* 161(4 Pt 1), tt. 1206-1212. doi: 10.1164/ajrccm.161.4.9907147

Abdel-Mawgoud, A. M., Lépine, F. and Déziel, E. 2010. Rhamnolipids: diversity of structures, microbial origins and roles. *Appl Microbiol Biotechnol* 86(5), tt. 1323-1336. doi: 10.1007/s00253-010-2498-2

Abdul Malik, S. A., Bazire, A., Gamboa-Muñoz, A., Bedoux, G., Robledo, D., García-Maldonado, J. Q. and Bourgougnon, N. 2020. Screening of Surface-associated Bacteria from the Mexican Red Alga Halymenia floresii for Quorum Sensing Activity. *Microbiology* 89(6), tt. 778-788. doi: 10.1134/S0026261720060132

Agnoli, K. *et al.* 2014. The third replicon of members of the Burkholderia cepacia Complex, plasmid pC3, plays a role in stress tolerance. *Appl Environ Microbiol* 80(4), tt. 1340-1348. doi: 10.1128/aem.03330-13

Agnoli, K. *et al.* 2012. Exposing the third chromosome of Burkholderia cepacia complex strains as a virulence plasmid. *Mol Microbiol* 83(2), tt. 362-378. doi: 10.1111/j.1365-2958.2011.07937.x

Aiyer, A. *et al.* 2021. Disruption of biofilms and killing of Burkholderia cenocepacia from cystic fibrosis lung using an antioxidant-antibiotic combination therapy. *Int J Antimicrob Agents* 58(2), t. 106372. doi: 10.1016/j.ijantimicag.2021.106372

Allan, N. D., Kooi, C., Sokol, P. A. and Beveridge, T. J. 2003. Putative virulence factors are released in association with membrane vesicles from Burkholderia cepacia. *Can J Microbiol* 49(10), tt. 613-624. doi: 10.1139/w03-078

Allen, B., Drake, M., Harris, N. and Sullivan, T. 2017. Using KBase to Assemble and Annotate Prokaryotic Genomes. *Curr Protoc Microbiol* 46, tt. 1e.13.11-11e.13.18. doi: 10.1002/cpmc.37

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. 1990. Basic local alignment search tool. *J Mol Biol* 215(3), tt. 403-410. doi: 10.1016/s0022-2836(05)80360-2

Andam, C. P. 2019. Clonal yet Different: Understanding the Causes of Genomic Heterogeneity in Microbial Species and Impacts on Public Health. *mSystems* 4(3). doi: 10.1128/mSystems.00097-19

Andersen, D. H. 1938. Cystic fibrosis of the pancreas and its relation to celiac disease - a clinical and pathological study. *American journal of Diseases of Children* 56, tt. 344-399.

Andrews, S. 2009. *FastQC: A quality control tool for high throughput sequence data*. Babraham Bioinformatics. Ar gael: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ [Gwelwyd.

Arkin, A. P. *et al.* 2018. KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nat Biotechnol* 36(7), tt. 566-569. doi: 10.1038/nbt.4163

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y. and Wishart, D. S. 2016. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 44(W1), tt. W16-21. doi: 10.1093/nar/gkw387

Aziz, R. K. *et al.* 2008. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9, t. 75. doi: 10.1186/1471-2164-9-75

Baldan, R. *et al.* 2014. Adaptation of Pseudomonas aeruginosa in Cystic Fibrosis airways influences virulence of Staphylococcus aureus in vitro and murine models of co-infection. *PLoS One* 9(3), t. e89614. doi: 10.1371/journal.pone.0089614

Baldwin, A. *et al.* 2008. Elucidating global epidemiology of Burkholderia multivorans in cases of cystic fibrosis by multilocus sequence typing. *J Clin Microbiol* 46(1), tt. 290-295. doi: 10.1128/jcm.01818-07

Baldwin, A. *et al.* 2007. Environmental Burkholderia cepacia complex isolates in human infections. *Emerg Infect Dis* 13(3), tt. 458-461. doi: 10.3201/eid1303.060403

Baldwin, A. *et al.* 2005. Multilocus sequence typing scheme that provides both species and strain differentiation for the Burkholderia cepacia complex. *J Clin Microbiol* 43(9), tt. 4665-4673. doi: 10.1128/jcm.43.9.4665-4673.2005

Baldwin, A., Sokol, P. A., Parkhill, J. and Mahenthiralingam, E. 2004. The Burkholderia cepacia epidemic strain marker is part of a novel genomic island encoding both virulence and metabolism-associated genes in Burkholderia cenocepacia. *Infect Immun* 72(3), tt. 1537-1547. doi: 10.1128/iai.72.3.1537-1547.2004

Bankevich, A. *et al.* 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19(5), tt. 455-477. doi: 10.1089/cmb.2012.0021

Barr, H. L. *et al.* 2015. Pseudomonas aeruginosa quorum sensing molecules correlate with clinical status in cystic fibrosis. *Eur Respir J* 46(4), tt. 1046-1054. doi: 10.1183/09031936.00225214

Bartell, J. A., Sommer, L. M., Haagensen, J. A. J., Loch, A., Espinosa, R., Molin, S. and Johansen, H. K. 2019. Evolutionary highways to persistent bacterial infection. *Nat Commun* 10(1), t. 629. doi: 10.1038/s41467-019-08504-7

Bassler, B. L., Wright, M., Showalter, R. E. and Silverman, M. R. 1993. Intercellular signalling in Vibrio harveyi: sequence and function of genes regulating expression of luminescence. *Mol Microbiol* 9(4), tt. 773-786. doi: 10.1111/j.1365-2958.1993.tb01737.x

Bauer, A. W., Kirby, W. M., Sherris, J. C. and Turck, M. 1966. Antibiotic susceptibility testing by a standardized single disk method. *Am J Clin Pathol* 45(4), tt. 493-496.

Becka, S. A. *et al.* 2018. Sequence heterogeneity of the PenA carbapenemase in clinical isolates of Burkholderia multivorans. *Diagn Microbiol Infect Dis* 92(3), tt. 253-258. doi: 10.1016/j.diagmicrobio.2018.06.005

Bernardy, E. E., Petit, R. A., 3rd, Raghuram, V., Alexander, A. M., Read, T. D. and Goldberg, J. B. 2020. Genotypic and Phenotypic Diversity of Staphylococcus aureus Isolates from Cystic Fibrosis Patient Lung Infections and Their Interactions with Pseudomonas aeruginosa. *mBio* 11(3). doi: 10.1128/mBio.00735-20

Bernier, S. P., Hum, C., Li, X., O'Toole, G. A., Magarvey, N. A. and Surette, M. G. 2017. Pseudomonas aeruginosa-Derived Rhamnolipids and Other Detergents Modulate Colony Morphotype and Motility in the Burkholderia cepacia Complex. *J Bacteriol* 199(13). doi: 10.1128/jb.00171-17

Bernier, S. P., Silo-Suh, L., Woods, D. E., Ohman, D. E. and Sokol, P. A. 2003. Comparative analysis of plant and animal models for characterization of Burkholderia cepacia virulence. *Infect Immun* 71(9), tt. 5306-5313. doi: 10.1128/iai.71.9.5306-5313.2003

Berriman, M. and Rutherford, K. 2003. Viewing and annotating sequence data with Artemis. *Brief Bioinform* 4(2), tt. 124-132. doi: 10.1093/bib/4.2.124

Biddick, R., Spilker, T., Martin, A. and LiPuma, J. J. 2003. Evidence of transmission of Burkholderia cepacia, Burkholderia multivorans and Burkholderia dolosa among persons with cystic fibrosis. *FEMS Microbiol Lett* 228(1), tt. 57-62. doi: 10.1016/s0378-1097(03)00724-9

Biswas, L. and Götz, F. 2021. Molecular Mechanisms of Staphylococcus and Pseudomonas Interactions in Cystic Fibrosis. *Front Cell Infect Microbiol* 11, t. 824042. doi: 10.3389/fcimb.2021.824042

Bloodworth, R. A. M., Zlitni, S., Brown, E. D. and Cardona, S. T. 2015. An electron transfer flavoprotein is essential for viability and its depletion causes a rod-to-sphere change in

Burkholderia cenocepacia. *Microbiology (Reading)* 161(10), tt. 1909-1920. doi: 10.1099/mic.0.000156

Bochner, B. R. 2009. Global phenotypic characterization of bacteria. *FEMS Microbiol Rev* 33(1), tt. 191-205. doi: 10.1111/j.1574-6976.2008.00149.x

Bodilis, J., Denet, E., Brothier, E., Graindorge, A., Favre-Bonté, S. and Nazaret, S. 2018. Comparative Genomics of Environmental and Clinical Burkholderia cenocepacia Strains Closely Related to the Highly Transmissible Epidemic ET12 Lineage. *Front Microbiol* 9, t. 383. doi: 10.3389/fmicb.2018.00383

Bönemann, G., Pietrosiuk, A., Diemand, A., Zentgraf, H. and Mogk, A. 2009. Remodelling of VipA/VipB tubules by ClpV-mediated threading is crucial for type VI protein secretion. *Embo j* 28(4), tt. 315-325. doi: 10.1038/emboj.2008.269

Borneman, J. and Triplett, E. W. 1997. Molecular microbial diversity in soils from eastern Amazonia: evidence for unusual microorganisms and microbial population shifts associated with deforestation. *Appl Environ Microbiol* 63(7), tt. 2647-2653. doi: 10.1128/aem.63.7.2647-2653.1997

Bosdriesz, E., Molenaar, D., Teusink, B. and Bruggeman, F. J. 2015. How fast-growing bacteria robustly tune their ribosome concentration to approximate growth-rate maximization. *Febs j* 282(10), tt. 2029-2044. doi: 10.1111/febs.13258

Böttger, E. C. 1989. Rapid determination of bacterial ribosomal RNA sequences by direct sequencing of enzymatically amplified DNA. *FEMS Microbiol Lett* 53(1-2), tt. 171-176. doi: 10.1016/0378-1097(89)90386-8

Boucher, R. 2007. Evidence for airway surface dehydration as the initiating event in CF airway disease. *Journal of internal medicine* 261(1), tt. 5-16.

Bowers, R. M. *et al.* 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* 35(8), tt. 725-731. doi: 10.1038/nbt.3893

Bragonzi, A. *et al.* 2012. Modelling co-infection of the cystic fibrosis lung by Pseudomonas aeruginosa and Burkholderia cenocepacia reveals influences on biofilm formation and host response. *PLoS One* 7(12), t. e52330. doi: 10.1371/journal.pone.0052330

Bragonzi, A., Wiehlmann, L., Klockgether, J., Cramer, N., Worlitzsch, D., Döring, G. and Tümmler, B. 2006. Sequence diversity of the mucABD locus in Pseudomonas aeruginosa isolates from patients with cystic fibrosis. *Microbiology (Reading)* 152(Pt 11), tt. 3261-3269. doi: 10.1099/mic.0.29175-0

Brenner, D. J., Fanning, G. R., Johnson, K. E., Citarella, R. V. and Falkow, S. 1969. Polynucleotide sequence relationships among members of Enterobacteriaceae. *J Bacteriol* 98(2), tt. 637-650. doi: 10.1128/jb.98.2.637-650.1969

Briaud, P., Camus, L., Bastien, S., Doléans-Jordheim, A., Vandenesch, F. and Moreau, K. 2019. Coexistence with Pseudomonas aeruginosa alters Staphylococcus aureus transcriptome, antibiotic resistance and internalization into epithelial cells. *Scientific Reports* 9(1), t. 16564. doi: 10.1038/s41598-019-52975-z

Bryant, J., Chewapreecha, C. and Bentley, S. D. 2012. Developing insights into the mechanisms of evolution of bacterial pathogens from whole-genome sequences. *Future Microbiol* 7(11), tt. 1283-1296. doi: 10.2217/fmb.12.108

Brynildsrud, O., Bohlin, J., Scheffer, L. and Eldholm, V. 2016. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biol* 17(1), t. 238. doi: 10.1186/s13059-016-1108-8

Burkholder, W. H. 1950. Sour skin, a bacterial rot. *Phytopathology* 40, tt. 115-117.

Burns, J. L., Wadsworth, C. D., Barry, J. J. and Goodall, C. P. 1996. Nucleotide sequence analysis of a gene from Burkholderia (Pseudomonas) cepacia encoding an outer membrane lipoprotein involved in multiple antibiotic resistance. *Antimicrob Agents Chemother* 40(2), tt. 307-313. doi: 10.1128/aac.40.2.307

Buroni, S. *et al.* 2009. Assessment of three Resistance-Nodulation-Cell Division drug efflux transporters of Burkholderia cenocepaciain intrinsic antibiotic resistance. *BMC microbiology* 9(1), tt. 1-11.

Burtnick, M. N., Brett, P. J. and DeShazer, D. 2014. Proteomic analysis of the Burkholderia pseudomallei type II secretome reveals hydrolytic enzymes, novel proteins, and the deubiquitinase TssM. *Infect Immun* 82(8), tt. 3214-3226. doi: 10.1128/iai.01739-14

Butt, A. T. and Thomas, M. S. 2017. Iron Acquisition Mechanisms and Their Role in the Virulence of Burkholderia Species. *Front Cell Infect Microbiol* 7, t. 460. doi: 10.3389/fcimb.2017.00460

Cain, A. K., Nolan, L. M., Sullivan, G. J., Whitchurch, C. B., Filloux, A. and Parkhill, J. 2019. Complete Genome Sequence of Pseudomonas aeruginosa Reference Strain PAK. *Microbiol Resour Announc* 8(41). doi: 10.1128/mra.00865-19

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T. L. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10(1), t. 421. doi: 10.1186/1471-2105-10-421

Campbell, P. W., 3rd, Phillips, J. A., 3rd, Heidecker, G. J., Krishnamani, M. R., Zahorchak, R. and Stull, T. L. 1995. Detection of Pseudomonas (Burkholderia) cepacia using PCR. *Pediatr Pulmonol* 20(1), tt. 44-49. doi: 10.1002/ppul.1950200109

Caraher, E., Duff, C., Mullen, T., Mc Keon, S., Murphy, P., Callaghan, M. and McClean, S. 2007. Invasion and biofilm formation of Burkholderia dolosa is comparable with Burkholderia cenocepacia and Burkholderia multivorans. *J Cyst Fibros* 6(1), tt. 49-56. doi: 10.1016/j.jcf.2006.05.007

Carattoli, A. *et al.* 2014. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* 58(7), tt. 3895-3903. doi: 10.1128/aac.02412-14

Cardona, S. T., Wopperer, J., Eberl, L. and Valvano, M. A. 2005. Diverse pathogenicity of Burkholderia cepacia complex strains in the Caenorhabditis elegans host model. *FEMS Microbiol Lett* 250(1), tt. 97-104. doi: 10.1016/j.femsle.2005.06.050

Cash, H. A., Woods, D. E., McCullough, B., Johanson, W. G., Jr. and Bass, J. A. 1979. A rat model of chronic respiratory infection with Pseudomonas aeruginosa. *Am Rev Respir Dis* 119(3), tt. 453-459. doi: 10.1164/arrd.1979.119.3.453

Castellani, C. *et al.* 2018. ECFS best practice guidelines: the 2018 revision. *Journal of cystic fibrosis* 17(2), tt. 153-178.

Castillo-Juárez, I. *et al.* 2015. Role of quorum sensing in bacterial infections. *World J Clin Cases* 3(7), tt. 575-598. doi: 10.12998/wjcc.v3.i7.575

Cesarini, S., Bevivino, A., Tabacchioni, S., Chiarini, L. and Dalmastri, C. 2009. RecA gene sequence and Multilocus Sequence Typing for species-level resolution of Burkholderia cepacia complex isolates. *Lett Appl Microbiol* 49(5), tt. 580-588. doi: 10.1111/j.1472-765X.2009.02709.x

Chen, J. S., Witzmann, K. A., Spilker, T., Fink, R. J. and LiPuma, J. J. 2001. Endemicity and inter-city spread of Burkholderia cepacia genomovar III in cystic fibrosis. *J Pediatr* 139(5), tt. 643-649. doi: 10.1067/mpd.2001.118430

Chen, L., Zheng, D., Liu, B., Yang, J. and Jin, Q. 2016. VFDB 2016: hierarchical and refined dataset for big data analysis--10 years on. *Nucleic Acids Res* 44(D1), tt. D694-697. doi: 10.1093/nar/gkv1239

Chen, Y., Wong, J., Sun, G. W., Liu, Y., Tan, G. Y. and Gan, Y. H. 2011. Regulation of type VI secretion system during Burkholderia pseudomallei infection. *Infect Immun* 79(8), tt. 3064-3073. doi: 10.1128/iai.05148-11

Cheng, H. P. and Lessie, T. G. 1994. Multiple replicons constituting the genome of Pseudomonas cepacia 17616. *J Bacteriol* 176(13), tt. 4034-4042. doi: 10.1128/jb.176.13.4034-4042.1994

Cheng, K. *et al.* 1996. Spread of beta-lactam-resistant Pseudomonas aeruginosa in a cystic fibrosis clinic. *Lancet* 348(9028), tt. 639-642. doi: 10.1016/s0140-6736(96)05169-0

Chu, K. K., Davidson, D. J., Halsey, T. K., Chung, J. W. and Speert, D. P. 2002. Differential persistence among genomovars of the Burkholderia cepacia complex in a murine model of pulmonary infection. *Infection and immunity* 70(5), tt. 2715-2720.

Chu, K. K., MacDonald, K. L., Davidson, D. J. and Speert, D. P. 2004. Persistence of Burkholderia multivorans within the pulmonary macrophage in the murine lung. *Infect Immun* 72(10), tt. 6142-6147. doi: 10.1128/iai.72.10.6142-6147.2004

Chung, J. W., Altman, E., Beveridge, T. J. and Speert, D. P. 2003. Colonial morphology of Burkholderia cepacia complex genomovar III: implications in exopolysaccharide production, pilus expression, and persistence in the mouse. *Infect Immun* 71(2), tt. 904-909. doi: 10.1128/iai.71.2.904-909.2003

Cieri, M. V., Mayer-Hamblett, N., Griffith, A. and Burns, J. L. 2002. Correlation between an in vitro invasion assay and a murine model of Burkholderia cepacia lung infection. *Infect Immun* 70(3), tt. 1081-1086. doi: 10.1128/iai.70.3.1081-1086.2002

Clark, S. T. *et al.* 2015. Phenotypic diversity within a Pseudomonas aeruginosa population infecting an adult with cystic fibrosis. *Sci Rep* 5, t. 10932. doi: 10.1038/srep10932

Clark, S. T., Guttman, D. S. and Hwang, D. M. 2018. Diversification of Pseudomonas aeruginosa within the cystic fibrosis lung and its effects on antibiotic resistance. *FEMS Microbiol Lett* 365(6). doi: 10.1093/femsle/fny026

Coburn, B. *et al.* 2015. Lung microbiota across age and disease stage in cystic fibrosis. *Sci Rep* 5, t. 10241. doi: 10.1038/srep10241

Coenye, T. and LiPuma, J. J. 2002. Multilocus restriction typing: a novel tool for studying global epidemiology of Burkholderia cepacia complex infection in cystic fibrosis. *J Infect Dis* 185(10), tt. 1454-1462. doi: 10.1086/340279

Coenye, T. *et al.* 2001a. Burkholderia ambifaria sp. nov., a novel member of the Burkholderia cepacia complex including biocontrol and cystic fibrosis-related isolates. *Int J Syst Evol Microbiol* 51(Pt 4), tt. 1481-1490. doi: 10.1099/00207713-51-4-1481

Coenye, T. and Vandamme, P. 2007. *Burkholderia: molecular microbiology and genomics*. Horizon Bioscience.

Coenye, T., Vandamme, P., Govan, J. R. and LiPuma, J. J. 2001b. Taxonomy and identification of the Burkholderia cepacia complex. *J Clin Microbiol* 39(10), tt. 3427-3436. doi: 10.1128/jcm.39.10.3427-3436.2001

Conly, J. and Johnston, B. 2005. Where are all the new antibiotics? The new antibiotic paradox. *Can J Infect Dis Med Microbiol* 16(3), tt. 159-160. doi: 10.1155/2005/892058

Connor, T. R. *et al.* 2016. CLIMB (the Cloud Infrastructure for Microbial Bioinformatics): an online resource for the medical microbiology community. *Microb Genom* 2(9), t. e000086. doi: 10.1099/mgen.0.000086

Conway, B. A., Venu, V. and Speert, D. P. 2002. Biofilm formation and acyl homoserine lactone production in the Burkholderia cepacia complex. *J Bacteriol* 184(20), tt. 5678-5685. doi: 10.1128/jb.184.20.5678-5685.2002

Costa, S. S., Guimarães, L. C., Silva, A., Soares, S. C. and Baraúna, R. A. 2020. First Steps in the Analysis of Prokaryotic Pan-Genomes. *Bioinform Biol Insights* 14, t. 1177932220938064. doi: 10.1177/1177932220938064

Costello, A., Reen, F. J., O'Gara, F., Callaghan, M. and McClean, S. 2014. Inhibition of co-colonizing cystic fibrosis-associated pathogens by Pseudomonas aeruginosa and Burkholderia multivorans. *Microbiology (Reading)* 160(Pt 7), tt. 1474-1487. doi: 10.1099/mic.0.074203-0

Costerton, J. W., Stewart, P. S. and Greenberg, E. P. 1999. Bacterial biofilms: a common cause of persistent infections. *Science* 284(5418), tt. 1318-1322. doi: 10.1126/science.284.5418.1318

Courtney, J. M., Dunbar, K. E., McDowell, A., Moore, J. E., Warke, T. J., Stevenson, M. and Elborn, J. S. 2004. Clinical outcome of Burkholderia cepacia complex infection in cystic fibrosis adults. *J Cyst Fibros* 3(2), tt. 93-98. doi: 10.1016/j.jcf.2004.01.005

Coutinho, C. P., de Carvalho, C. C., Madeira, A., Pinto-de-Oliveira, A. and Sá-Correia, I. 2011a. Burkholderia cenocepacia phenotypic clonal variation during a 3.5-year colonization in the lungs of a cystic fibrosis patient. *Infect Immun* 79(7), tt. 2950-2960. doi: 10.1128/iai.01366-10

Coutinho, C. P., Dos Santos, S. C., Madeira, A., Mira, N. P., Moreira, A. S. and Sá-Correia, I. 2011b. Long-term colonization of the cystic fibrosis lung by Burkholderia cepacia complex bacteria: epidemiology, clonal variation, and genome-wide expression alterations. *Front Cell Infect Microbiol* 1, t. 12. doi: 10.3389/fcimb.2011.00012

Coutinho, H. D., Falcão-Silva, V. S. and Gonçalves, G. F. 2008. Pulmonary bacterial pathogens in cystic fibrosis patients and antibiotic therapy: a tool for the health workers. *Int Arch Med* 1(1), t. 24. doi: 10.1186/1755-7682-1-24

Craig, F. F., Coote, J. G., Parton, R., Freer, J. H. and Gilmour, N. J. 1989. A plasmid which can be transferred between Escherichia coli and Pasteurella haemolytica by electroporation and conjugation. *J Gen Microbiol* 135(11), tt. 2885-2890. doi: 10.1099/00221287-135-11-2885

Croucher, N. J., Coupland, P. G., Stevenson, A. E., Callendrello, A., Bentley, S. D. and Hanage, W. P. 2014. Diversification of bacterial genome content through distinct mechanisms over different timescales. *Nat Commun* 5, t. 5471. doi: 10.1038/ncomms6471

Croucher, N. J. *et al.* 2015. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* 43(3), t. e15. doi: 10.1093/nar/gku1196

Cuccui, J. *et al.* 2012. Characterization of the Burkholderia pseudomallei K96243 capsular polysaccharide I coding region. *Infect Immun* 80(3), tt. 1209-1221. doi: 10.1128/iai.05805-11

Cullen, L. *et al.* 2018. The involvement of the low-oxygen-activated locus of Burkholderia cenocepacia in adaptation during cystic fibrosis infection. *Sci Rep* 8(1), t. 13386. doi: 10.1038/s41598-018-31556-6

Culp, E. and Wright, G. D. 2017. Bacterial proteases, untapped antimicrobial drug targets. *J Antibiot (Tokyo)* 70(4), tt. 366-377. doi: 10.1038/ja.2016.138

Cunha, M. V., Sousa, S. A., Leitão, J. H., Moreira, L. M., Videira, P. A. and Sá-Correia, I. 2004. Studies on the involvement of the exopolysaccharide produced by cystic fibrosis-associated isolates of the Burkholderia cepacia complex in biofilm formation and in persistence of respiratory infections. *J Clin Microbiol* 42(7), tt. 3052-3058. doi: 10.1128/jcm.42.7.3052-3058.2004

Cuthbertson, L. *et al.* 2020. Lung function and microbiota diversity in cystic fibrosis. *Microbiome* 8(1), t. 45. doi: 10.1186/s40168-020-00810-3

Cystic-Fibrosis-Canada. 2022. *The Canadian Cystic Fibrosis Registry 2020 Annual Data Report*. Toronto, Canada: Ar gael: https://www.cysticfibrosis.ca/registry/2020AnnualDataReport.pdf [Gwelwyd.

Daigneault, J., Aubin, G., Simard, F. and De Braekeleer, M. 1991. Genetic epidemiology of cystic fibrosis in Saguenay-Lac-St-Jean (Quebec, Canada). *Clin Genet* 40(4), tt. 298-303. doi: 10.1111/j.1399-0004.1991.tb03099.x

Daniels, R., Vanderleyden, J. and Michiels, J. 2004. Quorum sensing and swarming migration in bacteria. *FEMS Microbiol Rev* 28(3), tt. 261-289. doi: 10.1016/j.femsre.2003.09.004

Darch, S. E. *et al.* 2015. Recombination is a key driver of genomic and phenotypic diversity in a Pseudomonas aeruginosa population during cystic fibrosis infection. *Sci Rep* 5, t. 7649. doi: 10.1038/srep07649

Davidson, D. J. *et al.* 1995. Lung disease in the cystic fibrosis mouse exposed to bacterial pathogens. *Nature genetics* 9(4), tt. 351-357.

Davis, P. B. 2006. Cystic fibrosis since 1938. *Am J Respir Crit Care Med* 173(5), tt. 475-482. doi: 10.1164/rccm.200505-840OE

de Kraker, M. E., Stewardson, A. J. and Harbarth, S. 2016. Will 10 Million People Die a Year due to Antimicrobial Resistance by 2050? *PLoS Med* 13(11), t. e1002184. doi: 10.1371/journal.pmed.1002184

de la Fuente-Núñez, C., Reffuveille, F., Fernández, L. and Hancock, R. E. 2013. Bacterial biofilm development as a multicellular adaptation: antibiotic resistance and new therapeutic strategies. *Curr Opin Microbiol* 16(5), tt. 580-589. doi: 10.1016/j.mib.2013.06.013

De Maio, N. *et al.* 2019. Comparison of long-read sequencing technologies in the hybrid assembly of complex bacterial genomes. *Microb Genom* 5(9). doi: 10.1099/mgen.0.000294

De Soyza, A., Ellis, C. D., Khan, C. M., Corris, P. A. and Demarco de Hormaeche, R. 2004. Burkholderia cenocepacia lipopolysaccharide, lipid A, and proinflammatory activity. *Am J Respir Crit Care Med* 170(1), tt. 70-77. doi: 10.1164/rccm.200304-592OC

Dedeckova, K., Kalferstova, L., Strnad, H., Vavrova, J. and Drevinek, P. 2013. Novel diagnostic PCR assay for Burkholderia cenocepacia epidemic strain ST32 and its utility in monitoring infection in cystic fibrosis patients. *J Cyst Fibros* 12(5), tt. 475-481. doi: 10.1016/j.jcf.2012.12.007

DeLeon, S., Clinton, A., Fowler, H., Everett, J., Horswill, A. R. and Rumbaugh, K. P. 2014. Synergistic interactions of Pseudomonas aeruginosa and Staphylococcus aureus in an in vitro wound model. *Infect Immun* 82(11), tt. 4718-4728. doi: 10.1128/iai.02198-14

Deligianni, E. *et al.* 2010. Pseudomonas aeruginosa cystic fibrosis isolates of similar RAPD genotype exhibit diversity in biofilm forming ability in vitro. *BMC Microbiol* 10, t. 38. doi: 10.1186/1471-2180-10-38

Deng, Y., Boon, C., Eberl, L. and Zhang, L.-H. 2009a. Differential modulation of Burkholderia cenocepacia virulence and energy metabolism by the quorum-sensing signal BDSF and its synthase. *Journal of Bacteriology* 191(23), tt. 7270-7278.

Deng, Y., Boon, C., Eberl, L. and Zhang, L. H. 2009b. Differential modulation of Burkholderia cenocepacia virulence and energy metabolism by the quorum-sensing signal BDSF and its synthase. *J Bacteriol* 191(23), tt. 7270-7278. doi: 10.1128/jb.00681-09

Denman, C. C. and Brown, A. R. 2013. Mannitol promotes adherence of an outbreak strain of Burkholderia multivorans via an exopolysaccharide-independent mechanism that is associated with upregulation of newly identified fimbrial and afimbrial adhesins. *Microbiology (Reading)* 159(Pt 4), tt. 771-781. doi: 10.1099/mic.0.064832-0

Depoorter, E., Bull, M. J., Peeters, C., Coenye, T., Vandamme, P. and Mahenthiralingam, E. 2016. Burkholderia: an update on taxonomy and biotechnological potential as antibiotic producers. *Appl Microbiol Biotechnol* 100(12), tt. 5215-5229. doi: 10.1007/s00253-016-7520-x

Devanga Ragupathi, N. K. and Veeraraghavan, B. 2019. Accurate identification and epidemiological characterization of Burkholderia cepacia complex: an update. *Annals of clinical microbiology and antimicrobials* 18(1), tt. 1-10.

Di Lorenzo, F. *et al.* 2015. Activation of Human Toll-like Receptor 4 (TLR4)·Myeloid Differentiation Factor 2 (MD-2) by Hypoacylated Lipopolysaccharide from a Clinical Isolate of Burkholderia cenocepacia. *J Biol Chem* 290(35), tt. 21305-21319. doi: 10.1074/jbc.M115.649087

Diaz Caballero, J. *et al.* 2018. A genome-wide association analysis reveals a potential role for recombination in the evolution of antimicrobial resistance in Burkholderia multivorans. *PLoS Pathog* 14(12), t. e1007453. doi: 10.1371/journal.ppat.1007453

Ditta, G., Stanfield, S., Corbin, D. and Helinski, D. R. 1980. Broad host range DNA cloning system for gram-negative bacteria: construction of a gene bank of Rhizobium meliloti. *Proc Natl Acad Sci U S A* 77(12), tt. 7347-7351. doi: 10.1073/pnas.77.12.7347

Dobrindt, U., Hochhut, B., Hentschel, U. and Hacker, J. 2004. Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol* 2(5), tt. 414-424. doi: 10.1038/nrmicro884

Dobritsa, A. P. and Samadpour, M. 2016. Transfer of eleven species of the genus Burkholderia to the genus Paraburkholderia and proposal of Caballeronia gen. nov. to accommodate twelve species of the genera Burkholderia and Paraburkholderia. *Int J Syst Evol Microbiol* 66(8), tt. 2836-2846. doi: 10.1099/ijsem.0.001065

Döring, G., Parameswaran, I. G. and Murphy, T. F. 2011. Differential adaptation of microbial pathogens to airways of patients with cystic fibrosis and chronic obstructive pulmonary disease. *FEMS Microbiol Rev* 35(1), tt. 124-146. doi: 10.1111/j.1574-6976.2010.00237.x

Doster, E. *et al.* 2020. MEGARes 2.0: a database for classification of antimicrobial drug, biocide and metal resistance determinants in metagenomic sequence data. *Nucleic Acids Res* 48(D1), tt. D561-d569. doi: 10.1093/nar/gkz1010

Drevinek, P., Baldwin, A., Dowson, C. G. and Mahenthiralingam, E. 2008a. Diversity of the parB and repA genes of the Burkholderia cepacia complex and their utility for rapid identification of Burkholderia cenocepacia. *BMC Microbiol* 8, t. 44. doi: 10.1186/1471-2180-8-44

Drevinek, P. *et al.* 2010. Oxidative stress of Burkholderia cenocepacia induces insertion sequence-mediated genomic rearrangements that interfere with macrorestriction-based genotyping. *J Clin Microbiol* 48(1), tt. 34-40. doi: 10.1128/jcm.01433-09

Drevinek, P., Holden, M. T., Ge, Z., Jones, A. M., Ketchell, I., Gill, R. T. and Mahenthiralingam, E. 2008b. Gene expression changes linked to antimicrobial resistance, oxidative stress, iron depletion and retained motility are observed when Burkholderia cenocepaciagrows in cystic fibrosis sputum. *BMC infectious diseases* 8(1), tt. 1-16.

Drevínek, P., Hrbácková, H., Cinek, O., Bartosová, J., Nyc, O., Nemec, A. and Pohunek, P. 2002. Direct PCR detection of Burkholderia cepacia complex and identification of its genomovars by using sputum as source of DNA. *J Clin Microbiol* 40(9), tt. 3485-3488. doi: 10.1128/jcm.40.9.3485-3488.2002

Drevinek, P. and Mahenthiralingam, E. 2010. Burkholderia cenocepacia in cystic fibrosis: epidemiology and molecular mechanisms of virulence. *Clin Microbiol Infect* 16(7), tt. 821-830. doi: 10.1111/j.1469-0691.2010.03237.x

Drevinek, P., Vosahlikova, S., Cinek, O., Vavrova, V., Bartosova, J., Pohunek, P. and Mahenthiralingam, E. 2005. Widespread clone of Burkholderia cenocepacia in cystic fibrosis patients in the Czech Republic. *J Med Microbiol* 54(Pt 7), tt. 655-659. doi: 10.1099/jmm.0.46025-0

Eisen, J. A. 1995. The RecA protein as a model molecule for molecular systematic studies of bacteria: comparison of trees of RecAs and 16S rRNAs from the same species. *J Mol Evol* 41(6), tt. 1105-1123. doi: 10.1007/bf00173192

Elborn, J. S. 2016. Cystic fibrosis. *Lancet* 388(10059), tt. 2519-2531. doi: 10.1016/s0140-6736(16)00576-6

Esposito, A. *et al.* 2017. Evolution of Stenotrophomonas maltophilia in Cystic Fibrosis Lung over Chronic Infection: A Genomic and Phenotypic Population Study. *Front Microbiol* 8, t. 1590. doi: 10.3389/fmicb.2017.01590

EUCAST. 2017. Antimicrobial susceptibility testing of Burkholderia cepacia complex (BCC).

Ewels, P., Magnusson, M., Lundin, S. and Käller, M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32(19), tt. 3047-3048. doi: 10.1093/bioinformatics/btw354

Farrell, P. M. 2008. The prevalence of cystic fibrosis in the European Union. *J Cyst Fibros* 7(5), tt. 450-453. doi: 10.1016/j.jcf.2008.03.007

Faure, E., Kwong, K. and Nguyen, D. 2018. Pseudomonas aeruginosa in Chronic Lung Infections: How to Adapt Within the Host? *Front Immunol* 9, t. 2416. doi: 10.3389/fimmu.2018.02416

Fauroux, B. *et al.* 2004. Burkholderia cepacia is associated with pulmonary hypertension and increased mortality among cystic fibrosis patients. *J Clin Microbiol* 42(12), tt. 5537-5541. doi: 10.1128/jcm.42.12.5537-5541.2004

Feldgarden, M. *et al.* 2019. Validating the AMRFinder Tool and Resistance Gene Database by Using Antimicrobial Resistance Genotype-Phenotype Correlations in a Collection of Isolates. *Antimicrob Agents Chemother* 63(11). doi: 10.1128/aac.00483-19

Felsenstein, J. 1985. CONFIDENCE LIMITS ON PHYLOGENIES: AN APPROACH USING THE BOOTSTRAP. *Evolution* 39(4), tt. 783-791. doi: 10.1111/j.1558-5646.1985.tb00420.x

Ferreira, A. S., Leitão, J. H., Silva, I. N., Pinheiro, P. F., Sousa, S. A., Ramos, C. G. and Moreira, L. M. 2010. Distribution of cepacian biosynthesis genes among environmental and clinical Burkholderia strains and role of cepacian exopolysaccharide in resistance to stress conditions. *Appl Environ Microbiol* 76(2), tt. 441-450. doi: 10.1128/aem.01828-09

Figurski, D. H. and Helinski, D. R. 1979. Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. *Proc Natl Acad Sci U S A* 76(4), tt. 1648-1652. doi: 10.1073/pnas.76.4.1648

FitzSimmons, S. C. 1993. The changing epidemiology of cystic fibrosis. *J Pediatr* 122(1), tt. 1-9. doi: 10.1016/s0022-3476(05)83478-x

Flanagan, J. N., Kavanaugh, L. and Steck, T. R. 2020. Burkholderia multivorans Exhibits Antibiotic Collateral Sensitivity. *Microb Drug Resist* 26(1), tt. 1-8. doi: 10.1089/mdr.2019.0202

Flemming, H. C. and Wuertz, S. 2019. Bacteria and archaea on Earth and their abundance in biofilms. *Nat Rev Microbiol* 17(4), tt. 247-260. doi: 10.1038/s41579-019-0158-9

Fletcher, M. P., Diggle, S. P., Cámara, M. and Williams, P. 2007. Biosensor-based assays for PQS, HHQ and related 2-alkyl-4-quinolone quorum sensing signal molecules. *Nature protocols* 2(5), tt. 1254-1262.

Flight, W. G. *et al.* 2015. Rapid Detection of Emerging Pathogens and Loss of Microbial Diversity Associated with Severe Lung Disease in Cystic Fibrosis. *J Clin Microbiol* 53(7), tt. 2022-2029. doi: 10.1128/jcm.00432-15

France, M. W., Dodd, M. E., Govan, J. R., Doherty, C. J., Webb, A. K. and Jones, A. M. 2008. The changing epidemiology of Burkholderia species infection at an adult cystic fibrosis centre. *J Cyst Fibros* 7(5), tt. 368-372. doi: 10.1016/j.jcf.2008.01.002

Francisco, A. P., Vaz, C., Monteiro, P. T., Melo-Cristino, J., Ramirez, M. and Carriço, J. A. 2012. PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. *BMC Bioinformatics* 13, t. 87. doi: 10.1186/1471-2105-13-87

Freschi, L. *et al.* 2019. The Pseudomonas aeruginosa Pan-Genome Provides New Insights on Its Population Structure, Horizontal Gene Transfer, and Pathogenicity. *Genome Biol Evol* 11(1), tt. 109-120. doi: 10.1093/gbe/evy259

Furlan, J. P. R., Pitondo-Silva, A., Braz, V. S., Gallo, I. F. L. and Stehling, E. G. 2019. Evaluation of different molecular and phenotypic methods for identification of environmental Burkholderia cepacia complex. *World J Microbiol Biotechnol* 35(3), t. 39. doi: 10.1007/s11274-019-2614-0

Gabrielaite, M., Johansen, H. K., Molin, S., Nielsen, F. C. and Marvig, R. L. 2020. Gene Loss and Acquisition in Lineages of Pseudomonas aeruginosa Evolving in Cystic Fibrosis Patient Airways. *mBio* 11(5). doi: 10.1128/mBio.02359-20

Galardini, M., Biondi, E. G., Bazzicalupo, M. and Mengoni, A. 2011. CONTIGuator: a bacterial genomes finishing tool for structural insights on draft genomes. *Source Code Biol Med* 6, t. 11. doi: 10.1186/1751-0473-6-11

Galperin, M. Y., Kristensen, D. M., Makarova, K. S., Wolf, Y. I. and Koonin, E. V. 2019. Microbial genome analysis: the COG approach. *Brief Bioinform* 20(4), tt. 1063-1070. doi: 10.1093/bib/bbx117

Ganesan, S. and Sajjan, U. S. 2011. Host evasion by Burkholderia cenocepacia. *Front Cell Infect Microbiol* 1, t. 25. doi: 10.3389/fcimb.2011.00025

García-Romero, I. and Valvano, M. A. 2020. Complete Genome Sequence of Burkholderia cenocepacia K56-2, an Opportunistic Pathogen. *Microbiol Resour Announc* 9(43). doi: 10.1128/mra.01015-20

Gautam, V. *et al.* 2020. Exploring the Interplay of Resistance Nodulation Division Efflux Pumps, AmpC and OprD in Antimicrobial Resistance of Burkholderia cepacia Complex in Clinical Isolates. *Microb Drug Resist* 26(10), tt. 1144-1152. doi: 10.1089/mdr.2019.0102

Gillum, A. M., Tsay, E. Y. and Kirsch, D. R. 1984. Isolation of the Candida albicans gene for orotidine-5'-phosphate decarboxylase by complementation of S. cerevisiae ura3 and E. coli pyrF mutations. *Mol Gen Genet* 198(2), tt. 179-182. doi: 10.1007/bf00328721

Gingues, S., Kooi, C., Visser, M. B., Subsin, B. and Sokol, P. A. 2005. Distribution and expression of the ZmpA metalloprotease in the Burkholderia cepacia complex. *J Bacteriol* 187(24), tt. 8247-8255. doi: 10.1128/jb.187.24.8247-8255.2005

Gogarten, J. P. and Townsend, J. P. 2005. Horizontal gene transfer, genome innovation and evolution. *Nat Rev Microbiol* 3(9), tt. 679-687. doi: 10.1038/nrmicro1204

Gomes, M. C. *et al.* 2018. The afc antifungal activity cluster, which is under tight regulatory control of ShvR, is essential for transition from intracellular persistence of Burkholderia cenocepacia to acute pro-inflammatory infection. *PLoS Pathog* 14(12), t. e1007473. doi: 10.1371/journal.ppat.1007473

Gordon, V., Bakhtiari, L. and Kovach, K. 2019. From molecules to multispecies ecosystems: the roles of structure in bacterial biofilms. *Phys Biol* 16(4), t. 041001. doi: 10.1088/1478-3975/ab1384

Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P. and Tiedje, J. M. 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* 57(Pt 1), tt. 81-91. doi: 10.1099/ijs.0.64483-0

Gotschlich, A. *et al.* 2001. Synthesis of multiple N-acylhomoserine lactones is wide-spread among the members of the Burkholderia cepacia complex. *Syst Appl Microbiol* 24(1), tt. 1-14. doi: 10.1078/0723-2020-00013

Govan, J. R. and Deretic, V. 1996. Microbial pathogenesis in cystic fibrosis: mucoid Pseudomonas aeruginosa and Burkholderia cepacia. *Microbiological reviews* 60(3), tt. 539-574.

Graindorge, A., Menard, A., Monnez, C. and Cournoyer, B. 2012. Insertion sequence evolutionary patterns highlight convergent genetic inactivations and recent genomic island acquisitions among epidemic Burkholderia cenocepacia. *J Med Microbiol* 61(Pt 3), tt. 394-409. doi: 10.1099/jmm.0.036822-0

Grant, S. G., Jessee, J., Bloom, F. R. and Hanahan, D. 1990. Differential plasmid rescue from transgenic mouse DNAs into Escherichia coli methylation-restriction mutants. *Proc Natl Acad Sci U S A* 87(12), tt. 4645-4649. doi: 10.1073/pnas.87.12.4645

Green, A. E. *et al.* 2021. Pneumococcal Colonization and Virulence Factors Identified Via Experimental Evolution in Infection Models. *Mol Biol Evol* 38(6), tt. 2209-2226. doi: 10.1093/molbev/msab018

Guglierame, P., Pasca, M. R., De Rossi, E., Buroni, S., Arrigo, P., Manina, G. and Riccardi, G. 2006. Efflux pump genes of the resistance-nodulation-division family in Burkholderia cenocepacia genome. *BMC microbiology* 6(1), tt. 1-14.

Guo, F. B., Xiong, L., Zhang, K. Y., Dong, C., Zhang, F. Z. and Woo, P. C. 2017. Identification and analysis of genomic islands in Burkholderia cenocepacia AU 1054 with emphasis on pathogenicity islands. *BMC Microbiol* 17(1), t. 73. doi: 10.1186/s12866-017-0986-6

Gupta, S. K., Padmanabhan, B. R., Diene, S. M., Lopez-Rojas, R., Kempf, M., Landraud, L. and Rolain, J. M. 2014. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob Agents Chemother* 58(1), tt. 212-220. doi: 10.1128/aac.01310-13

Gurevich, A., Saveliev, V., Vyahhi, N. and Tesler, G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29(8), tt. 1072-1075. doi: 10.1093/bioinformatics/btt086

Ha, D. G., Kuchma, S. L. and O'Toole, G. A. 2014. Plate-based assay for swimming motility in Pseudomonas aeruginosa. *Methods Mol Biol* 1149, tt. 59-65. doi: 10.1007/978-1-4939-0473-0_7

Hacker, J. and Carniel, E. 2001. Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes. *EMBO Rep* 2(5), tt. 376-381. doi: 10.1093/embo-reports/kve097

Hales, B. A., Morgan, J. A., Hart, C. A. and Winstanley, C. 1998. Variation in flagellin genes and proteins of Burkholderia cepacia. *J Bacteriol* 180(5), tt. 1110-1118. doi: 10.1128/jb.180.5.1110-1118.1998

Harrington, N. E., Sweeney, E. and Harrison, F. 2020. Building a better biofilm - Formation of in vivo-like biofilm structures by Pseudomonas aeruginosa in a porcine model of cystic fibrosis lung infection. *Biofilm* 2, t. 100024. doi: 10.1016/j.bioflm.2020.100024

Harriott, M. M. and Noverr, M. C. 2009. Candida albicans and Staphylococcus aureus form polymicrobial biofilms: effects on antimicrobial resistance. *Antimicrob Agents Chemother* 53(9), tt. 3914-3922. doi: 10.1128/aac.00657-09

Harrison, F. 2007. Microbial ecology of the cystic fibrosis lung. *Microbiology (Reading)* 153(Pt 4), tt. 917-923. doi: 10.1099/mic.0.2006/004077-0

Harrison, F., Muruli, A., Higgins, S. and Diggle, S. P. 2014. Development of an ex vivo porcine lung model for studying growth, virulence, and signaling of Pseudomonas aeruginosa. *Infect Immun* 82(8), tt. 3312-3323. doi: 10.1128/iai.01554-14

Hassan, A. A., Coutinho, C. P. and Sá-Correia, I. 2019. Burkholderia cepacia Complex Species Differ in the Frequency of Variation of the Lipopolysaccharide O-Antigen Expression During Cystic Fibrosis Chronic Respiratory Infection. *Front Cell Infect Microbiol* 9, t. 273. doi: 10.3389/fcimb.2019.00273

Hassan, A. A., Dos Santos, S. C., Cooper, V. S. and Sá-Correia, I. 2020. Comparative Evolutionary Patterns of Burkholderia cenocepacia and B. multivorans During Chronic Co-infection of a Cystic Fibrosis Patient Lung. *Front Microbiol* 11, t. 574626. doi: 10.3389/fmicb.2020.574626

Henry, D. *et al.* 1999. Comparison of isolation media for recovery of Burkholderia cepacia complex from respiratory secretions of patients with cystic fibrosis. *J Clin Microbiol* 37(4), tt. 1004-1007. doi: 10.1128/jcm.37.4.1004-1007.1999

Henry, D. A., Mahenthiralingam, E., Vandamme, P., Coenye, T. and Speert, D. P. 2001. Phenotypic methods for determining genomovar status of the Burkholderia cepacia complex. *J Clin Microbiol* 39(3), tt. 1073-1078. doi: 10.1128/jcm.39.3.1073-1078.2001

Herasimenka, Y. *et al.* 2007. Exopolysaccharides produced by clinical strains belonging to the Burkholderia cepacia complex. *J Cyst Fibros* 6(2), tt. 145-152. doi: 10.1016/j.jcf.2006.06.004

Hibbing, M. E., Fuqua, C., Parsek, M. R. and Peterson, S. B. 2010. Bacterial competition: surviving and thriving in the microbial jungle. *Nat Rev Microbiol* 8(1), tt. 15-25. doi: 10.1038/nrmicro2259

Hogardt, M., Ulrich, J., Riehn-Kopp, H. and Tümmler, B. 2009. EuroCareCF quality assessment of diagnostic microbiology of cystic fibrosis isolates. *J Clin Microbiol* 47(11), tt. 3435-3438. doi: 10.1128/jcm.01182-09

Holden, M. T. *et al.* 2009. The genome of Burkholderia cenocepacia J2315, an epidemic pathogen of cystic fibrosis patients. *J Bacteriol* 191(1), tt. 261-277. doi: 10.1128/jb.01230-08

Horsley, A., Jones, A. M. and Lord, R. 2016. Antibiotic treatment for Burkholderia cepacia complex in people with cystic fibrosis experiencing a pulmonary exacerbation. *Cochrane Database Syst Rev* 2016(1), t. Cd009529. doi: 10.1002/14651858.CD009529.pub3

Huber, B. *et al.* 2001. The cep quorum-sensing system of Burkholderia cepacia H111 controls biofilm formation and swarming motility. *Microbiology (Reading)* 147(Pt 9), tt. 2517-2528. doi: 10.1099/00221287-147-9-2517

Huerta-Cepas, J. *et al.* 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* 47(D1), tt. D309-d314. doi: 10.1093/nar/gky1085

Hunt, T. A., Kooi, C., Sokol, P. A. and Valvano, M. A. 2004. Identification of Burkholderia cenocepacia genes required for bacterial survival in vivo. *Infect Immun* 72(7), tt. 4010-4022. doi: 10.1128/iai.72.7.4010-4022.2004

Insightful-Science. SnapGene software.

Irorere, V. U., Tripathi, L., Marchant, R., McClean, S. and Banat, I. M. 2017. Microbial rhamnolipid production: a critical re-evaluation of published data and suggested future publication criteria. *Appl Microbiol Biotechnol* 101(10), tt. 3941-3951. doi: 10.1007/s00253-017-8262-0

Isles, A., Maclusky, I., Corey, M., Gold, R., Prober, C., Fleming, P. and Levison, H. 1984. Pseudomonas cepacia infection in cystic fibrosis: an emerging problem. *J Pediatr* 104(2), tt. 206-210. doi: 10.1016/s0022-3476(84)80993-2

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. and Aluru, S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications* 9. doi: 10.1038/s41467-018-07641-9

Jara, J. *et al.* 2020. Self-Adaptation of Pseudomonas fluorescens Biofilms to Hydrodynamic Stress. *Front Microbiol* 11, t. 588884. doi: 10.3389/fmicb.2020.588884

Jarrell, K. F. and McBride, M. J. 2008. The surprisingly diverse ways that prokaryotes move. *Nat Rev Microbiol* 6(6), tt. 466-476. doi: 10.1038/nrmicro1900

Jean-Pierre, F., Vyas, A., Hampton, T. H., Henson, M. A. and O'Toole, G. A. 2021. One versus Many: Polymicrobial Communities and the Cystic Fibrosis Airway. *mBio* 12(2). doi: 10.1128/mBio.00006-21

Jensen, L. J., Julien, P., Kuhn, M., von Mering, C., Muller, J., Doerks, T. and Bork, P. 2008. eggNOG: automated construction and annotation of orthologous groups of genes. *Nucleic Acids Res* 36(Database issue), tt. D250-254. doi: 10.1093/nar/gkm796

Jia, B. *et al.* 2017. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res* 45(D1), tt. D566-d573. doi: 10.1093/nar/gkw1004

Jin, Y. *et al.* 2020. Genome-based classification of Burkholderia cepacia complex provides new insight into its taxonomic status. *Biol Direct* 15(1), t. 6. doi: 10.1186/s13062-020-0258-5

Jolley, K. A. *et al.* 2012. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. *Microbiology (Reading)* 158(Pt 4), tt. 1005-1015. doi: 10.1099/mic.0.055459-0

Jolley, K. A., Bray, J. E. and Maiden, M. C. J. 2018. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res* 3, t. 124. doi: 10.12688/wellcomeopenres.14826.1

Jolley, K. A. and Maiden, M. C. 2010. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* 11, t. 595. doi: 10.1186/1471-2105-11-595

Jones, A. M., Dodd, M. E., Govan, J. R., Barcus, V., Doherty, C. J., Morris, J. and Webb, A. K. 2004. Burkholderia cenocepacia and Burkholderia multivorans: influence on survival in cystic fibrosis. *Thorax* 59(11), tt. 948-951. doi: 10.1136/thx.2003.017210

Kahl, B. C. 2010. Impact of Staphylococcus aureus on the pathogenesis of chronic cystic fibrosis lung disease. *Int J Med Microbiol* 300(8), tt. 514-519. doi: 10.1016/j.ijmm.2010.08.002

Kahm, M., Hasenbrink, G., Lichtenberg-Frate, H., Ludwig, J. and Kschischo, M. 2010. Grofit: Fitting biological growth curves. *Nature Precedings*. doi: 10.1038/npre.2010.4508.1

Karlin, S., Weinstock, G. M. and Brendel, V. 1995. Bacterial classifications derived from recA protein sequence comparisons. *J Bacteriol* 177(23), tt. 6881-6893. doi: 10.1128/jb.177.23.6881-6893.1995

Karpati, F. and Jonasson, J. 1996. Polymerase chain reaction for the detection of Pseudomonas aeruginosa, Stenotrophomonas maltophilia and Burkholderia cepacia in sputum of patients with cystic fibrosis. *Mol Cell Probes* 10(6), tt. 397-403. doi: 10.1006/mcpr.1996.0055

Katoh, K., Misawa, K., Kuma, K. and Miyata, T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30(14), tt. 3059-3066. doi: 10.1093/nar/gkf436

Kavanaugh, L. G., Flanagan, J. N. and Steck, T. R. 2020. Reciprocal antibiotic collateral sensitivity in Burkholderia multivorans. *Int J Antimicrob Agents* 56(1), t. 105994. doi: 10.1016/j.ijantimicag.2020.105994

Kearns, D. B. 2010. A field guide to bacterial swarming motility. *Nat Rev Microbiol* 8(9), tt. 634-644. doi: 10.1038/nrmicro2405

Kenna, D. T. D. *et al.* 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), tt. 490-501. doi: https://doi.org/10.1099/jmm.0.000458

Kim, S. J. and Skach, W. R. 2012. Mechanisms of CFTR Folding at the Endoplasmic Reticulum. *Front Pharmacol* 3, t. 201. doi: 10.3389/fphar.2012.00201

Kirchner, S., Fothergill, J. L., Wright, E. A., James, C. E., Mowat, E. and Winstanley, C. 2012. Use of artificial sputum medium to test antibiotic efficacy against Pseudomonas aeruginosa in conditions more relevant to the cystic fibrosis lung. *J Vis Exp* (64), t. e3857. doi: 10.3791/3857

Klinger, K. W. 1983. Cystic fibrosis in the Ohio Amish: gene frequency and founder effect. *Hum Genet* 65(2), tt. 94-98. doi: 10.1007/bf00286641

Kolmogorov, M., Yuan, J., Lin, Y. and Pevzner, P. A. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 37(5), tt. 540-546. doi: 10.1038/s41587-019-0072-8

Köthe, M., Antl, M., Huber, B., Stoecker, K., Ebrecht, D., Steinmetz, I. and Eberl, L. 2003. Killing of Caenorhabditis elegans by Burkholderia cepacia is controlled by the cep quorum-sensing system. *Cell Microbiol* 5(5), tt. 343-351. doi: 10.1046/j.1462-5822.2003.00280.x

Kotra, L. P., Haddad, J. and Mobashery, S. 2000. Aminoglycosides: perspectives on mechanisms of action and resistance and strategies to counter resistance. *Antimicrob Agents Chemother* 44(12), tt. 3249-3256. doi: 10.1128/aac.44.12.3249-3256.2000

Kovacs-Simon, A., Hemsley, C. M., Scott, A. E., Prior, J. L. and Titball, R. W. 2019. Burkholderia thailandensis strain E555 is a surrogate for the investigation of Burkholderia pseudomallei replication and survival in macrophages. *BMC Microbiol* 19(1), t. 97. doi: 10.1186/s12866-019-1469-8

Kowalska-Krochmal, B. and Dudek-Wicher, R. 2021. The Minimum Inhibitory Concentration of Antibiotics: Methods, Interpretation, Clinical Relevance. *Pathogens* 10(2). doi: 10.3390/pathogens10020165

Krueger, F. 2017. Trim Galore!

Krzywinski, M. *et al.* 2009. Circos: an information aesthetic for comparative genomics. *Genome Res* 19(9), tt. 1639-1645. doi: 10.1101/gr.092759.109

Kumar, B., Sorensen, J. L. and Cardona, S. T. 2018. A c-di-GMP-Modulating Protein Regulates Swimming Motility of Burkholderia cenocepacia in Response to Arginine and Glutamate. *Front Cell Infect Microbiol* 8, t. 56. doi: 10.3389/fcimb.2018.00056

La Rosa, R., Rossi, E., Feist, A. M., Johansen, H. K. and Molin, S. 2021. Compensatory evolution of Pseudomonas aeruginosa's slow growth phenotype suggests mechanisms of adaptation in cystic fibrosis. *Nat Commun* 12(1), t. 3186. doi: 10.1038/s41467-021-23451-y

Lamothe, J., Thyssen, S. and Valvano, M. A. 2004. Burkholderia cepacia complex isolates survive intracellularly without replication within acidic vacuoles of Acanthamoeba polyphaga. *Cellular microbiology* 6(12), tt. 1127-1138.

Lanave, C., Preparata, G., Saccone, C. and Serio, G. 1984. A new method for calculating evolutionary substitution rates. *J Mol Evol* 20(1), tt. 86-93. doi: 10.1007/bf02101990

Land, M. *et al.* 2015. Insights from 20 years of bacterial genome sequencing. *Funct Integr Genomics* 15(2), tt. 141-161. doi: 10.1007/s10142-015-0433-4

Lane, D. J., Pace, B., Olsen, G. J., Stahl, D. A., Sogin, M. L. and Pace, N. R. 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc Natl Acad Sci U S A* 82(20), tt. 6955-6959. doi: 10.1073/pnas.82.20.6955

Lauman, P. and Dennis, J. J. 2021. Advances in Phage Therapy: Targeting the Burkholderia cepacia Complex. *Viruses* 13(7). doi: 10.3390/v13071331

Laws, T. R., Smith, S. A., Smith, M. P., Harding, S. V., Atkins, T. P. and Titball, R. W. 2005. The nematode Panagrellus redivivus is susceptible to killing by human pathogens at 37 C. *FEMS microbiology letters* 250(1), tt. 77-83.

Ledson, M., Gallagher, M., Corkill, J., Hart, C. and Walshaw, M. 1998. Cross infection between cystic fibrosis patients colonised with Burkholderia cepacia. *Thorax* 53(5), tt. 432-436.

Lee, A. H. *et al.* 2017. Phenotypic diversity and genotypic flexibility of Burkholderia cenocepacia during long-term chronic infection of cystic fibrosis lungs. *Genome Res* 27(4), tt. 650-662. doi: 10.1101/gr.213363.116

Lee, D. G. *et al.* 2006. Genomic analysis reveals that Pseudomonas aeruginosa virulence is combinatorial. *Genome Biol* 7(10), t. R90. doi: 10.1186/gb-2006-7-10-r90

Lee, H. S., Gu, F., Ching, S. M., Lam, Y. and Chua, K. L. 2010. CdpA is a Burkholderia pseudomallei cyclic di-GMP phosphodiesterase involved in autoaggregation, flagellum synthesis, motility, biofilm formation, cell invasion, and cytotoxicity. *Infect Immun* 78(5), tt. 1832-1840. doi: 10.1128/iai.00446-09

Lefebre, M. D. and Valvano, M. A. 2002. Construction and evaluation of plasmid vectors optimized for constitutive and regulated gene expression in Burkholderia cepacia complex isolates. *Appl Environ Microbiol* 68(12), tt. 5956-5964. doi: 10.1128/aem.68.12.5956-5964.2002

Leitão, J. H., Sousa, S. A., Ferreira, A. S., Ramos, C. G., Silva, I. N. and Moreira, L. M. 2010. Pathogenicity, virulence factors, and strategies to fight against Burkholderia cepacia complex pathogens and related species. *Applied Microbiology and Biotechnology* 87(1), tt. 31-40. doi: 10.1007/s00253-010-2528-0

Lessie, T. G., Hendrickson, W., Manning, B. D. and Devereux, R. 1996. Genomic complexity and plasticity of Burkholderia cepacia. *FEMS Microbiol Lett* 144(2-3), tt. 117-128. doi: 10.1111/j.1574-6968.1996.tb08517.x

Lewenza, S., Conway, B., Greenberg, E. P. and Sokol, P. A. 1999. Quorum sensing in Burkholderia cepacia: identification of the LuxRI homologs CepRI. *J Bacteriol* 181(3), tt. 748-756. doi: 10.1128/jb.181.3.748-756.1999

Li, H. 2016. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* 32(14), tt. 2103-2110. doi: 10.1093/bioinformatics/btw152

Li, H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34(18), tt. 3094-3100. doi: 10.1093/bioinformatics/bty191

Li, H. *et al.* 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16), tt. 2078-2079. doi: 10.1093/bioinformatics/btp352

Li, L. and Somerset, S. 2014. Digestive system dysfunction in cystic fibrosis: challenges for nutrition therapy. *Dig Liver Dis* 46(10), tt. 865-874. doi: 10.1016/j.dld.2014.06.011

Li, X. H. and Lee, J. H. 2019. Quorum sensing-dependent post-secretional activation of extracellular proteases in Pseudomonas aeruginosa. *J Biol Chem* 294(51), tt. 19635-19644. doi: 10.1074/jbc.RA119.011047

Lieberman, T. D. *et al.* 2011. Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. *Nat Genet* 43(12), tt. 1275-1280. doi: 10.1038/ng.997

Lipuma, J. J. 2010. The changing microbial epidemiology in cystic fibrosis. *Clin Microbiol Rev* 23(2), tt. 299-323. doi: 10.1128/cmr.00068-09

LiPuma, J. J., Dulaney, B. J., McMenamin, J. D., Whitby, P. W., Stull, T. L., Coenye, T. and Vandamme, P. 1999. Development of rRNA-based PCR assays for identification of Burkholderia cepacia complex isolates recovered from cystic fibrosis patients. *J Clin Microbiol* 37(10), tt. 3167-3170. doi: 10.1128/jcm.37.10.3167-3170.1999

LiPuma, J. J., Mortensen, J. E., Dasen, S. E., Edlind, T. D., Schidlow, D. V., Burns, J. L. and Stull, T. L. 1988. Ribotype analysis of Pseudomonas cepacia from cystic fibrosis treatment centers. *J Pediatr* 113(5), tt. 859-862. doi: 10.1016/s0022-3476(88)80018-0

Livermore, D. M. and Woodford, N. 2006. The beta-lactamase threat in Enterobacteriaceae, Pseudomonas and Acinetobacter. *Trends Microbiol* 14(9), tt. 413-420. doi: 10.1016/j.tim.2006.07.008

Lood, C. *et al.* 2021. Genomics of an endemic cystic fibrosis Burkholderia multivorans strain reveals low within-patient evolution but high between-patient diversity. *PLoS Pathog* 17(3), t. e1009418. doi: 10.1371/journal.ppat.1009418

Lord, R., Jones, A. M. and Horsley, A. 2020. Antibiotic treatment for Burkholderia cepacia complex in people with cystic fibrosis experiencing a pulmonary exacerbation. *Cochrane Database Syst Rev* 4(4), t. Cd009529. doi: 10.1002/14651858.CD009529.pub4

Loutet, S. A. and Valvano, M. A. 2010. A decade of Burkholderia cenocepacia virulence determinant research. *Infect Immun* 78(10), tt. 4088-4100. doi: 10.1128/iai.00212-10

Loutet, S. A. and Valvano, M. A. 2011. Extreme antimicrobial Peptide and polymyxin B resistance in the genus burkholderia. *Front Microbiol* 2, t. 159. doi: 10.3389/fmicb.2011.00159

Lu, D. C., Chang, S. C., Chen, Y. C., Luh, K. T., Lee, C. Y. and Hsieh, W. C. 1997. Burkholderia cepacia bacteremia: a retrospective analysis of 70 episodes. *J Formos Med Assoc* 96(12), tt. 972-978.

Lyczak, J. B., Cannon, C. L. and Pier, G. B. 2002. Lung infections associated with cystic fibrosis. *Clin Microbiol Rev* 15(2), tt. 194-222. doi: 10.1128/cmr.15.2.194-222.2002

Madeira, A. *et al.* 2013. Proteomic profiling of Burkholderia cenocepacia clonal isolates with different virulence potential retrieved from a cystic fibrosis patient during chronic lung infection. *PLoS One* 8(12), t. e83065. doi: 10.1371/journal.pone.0083065

Mahenthiralingam, E., Bischof, J., Byrne, S. K., Radomski, C., Davies, J. E., Av-Gay, Y. and Vandamme, P. 2000. DNA-Based diagnostic approaches for identification of Burkholderia cepacia complex, Burkholderia vietnamiensis, Burkholderia multivorans, Burkholderia stabilis, and Burkholderia cepacia genomovars I and III. *J Clin Microbiol* 38(9), tt. 3165-3173. doi: 10.1128/jcm.38.9.3165-3173.2000

Mahenthiralingam, E., Campbell, M. E., Henry, D. A. and Speert, D. P. 1996. Epidemiology of Burkholderia cepacia infection in patients with cystic fibrosis: analysis by randomly amplified polymorphic DNA fingerprinting. *J Clin Microbiol* 34(12), tt. 2914-2920. doi: 10.1128/jcm.34.12.2914-2920.1996

Mahenthiralingam, E., Campbell, M. E. and Speert, D. P. 1994. Nonmotility and phagocytic resistance of Pseudomonas aeruginosa isolates from chronically colonized patients with cystic fibrosis. *Infect Immun* 62(2), tt. 596-605. doi: 10.1128/iai.62.2.596-605.1994

Mahenthiralingam, E., Marchbank, A., Drevinek, P., Garaiova, I. and Plummer, S. 2009. Use of colony-based bacterial strain typing for tracking the fate of Lactobacillusstrains during human consumption. *BMC Microbiology* 9(1), t. 251. doi: 10.1186/1471-2180-9-251

Mahenthiralingam, E., Simpson, D. A. and Speert, D. P. 1997. Identification and characterization of a novel DNA marker associated with epidemic Burkholderia cepacia strains recovered from patients with cystic fibrosis. *J Clin Microbiol* 35(4), tt. 808-816. doi: 10.1128/jcm.35.4.808-816.1997

Mahenthiralingam, E., Urban, T. A. and Goldberg, J. B. 2005. The multifarious, multireplicon Burkholderia cepacia complex. *Nat Rev Microbiol* 3(2), tt. 144-156. doi: 10.1038/nrmicro1085

Mahenthiralingam, E. and Vandamme, P. 2005. Taxonomy and pathogenesis of the Burkholderia cepacia complex. *Chron Respir Dis* 2(4), tt. 209-217. doi: 10.1191/1479972305cd053ra

Mahenthiralingam, E. *et al.* 2001. Infection with Burkholderia cepacia complex genomovars in patients with cystic fibrosis: virulent transmissible strains of genomovar III can replace Burkholderia multivorans. *Clin Infect Dis* 33(9), tt. 1469-1475. doi: 10.1086/322684

Mahillon, J. and Chandler, M. 1998. Insertion sequences. *Microbiol Mol Biol Rev* 62(3), tt. 725-774. doi: 10.1128/mmbr.62.3.725-774.1998

Maiden, M. C., Jansen van Rensburg, M. J., Bray, J. E., Earle, S. G., Ford, S. A., Jolley, K. A. and McCarthy, N. D. 2013. MLST revisited: the gene-by-gene approach to bacterial genomics. *Nat Rev Microbiol* 11(10), tt. 728-736. doi: 10.1038/nrmicro3093

Maldonado, R. F., Sá-Correia, I. and Valvano, M. A. 2016. Lipopolysaccharide modification in Gram-negative bacteria during chronic infection. *FEMS Microbiol Rev* 40(4), tt. 480-493. doi: 10.1093/femsre/fuw007

Mall, M. A. and Hartl, D. 2014. CFTR: cystic fibrosis and beyond. *Eur Respir J* 44(4), tt. 1042-1054. doi: 10.1183/09031936.00228013

Malott, R. J., Baldwin, A., Mahenthiralingam, E. and Sokol, P. A. 2005. Characterization of the cciIR quorum-sensing system in Burkholderia cenocepacia. *Infection and immunity* 73(8), tt. 4982-4992.

Malott, R. J., Steen-Kinnaird, B. R., Lee, T. D. and Speert, D. P. 2012. Identification of hopanoid biosynthesis genes involved in polymyxin resistance in Burkholderia multivorans. *Antimicrob Agents Chemother* 56(1), tt. 464-471. doi: 10.1128/aac.00602-11

Malott, R. J. *et al.* 2014. Fosmidomycin decreases membrane hopanoids and potentiates the effects of colistin on Burkholderia multivorans clinical isolates. *Antimicrob Agents Chemother* 58(9), tt. 5211-5219. doi: 10.1128/aac.02705-14

Mardis, E., McPherson, J., Martienssen, R., Wilson, R. K. and McCombie, W. R. 2002. What is finished, and why does it matter. *Genome Res* 12(5), tt. 669-671. doi: 10.1101/gr.032102

Marier, J.-F., Lavigne, J. and Ducharme, M. P. 2002. Pharmacokinetics and efficacies of liposomal and conventional formulations of tobramycin after intratracheal administration in rats with pulmonary Burkholderia cepacia infection. *Antimicrobial agents and chemotherapy* 46(12), tt. 3776-3781.

Martin, I., Waters, V. and Grasemann, H. 2021. Approaches to Targeting Bacterial Biofilms in Cystic Fibrosis Airways. *Int J Mol Sci* 22(4). doi: 10.3390/ijms22042155

Martin, M. 2009. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal* 17(1), t. 10.

Mateu-Borrás, M., Zamorano, L., González-Alsina, A., Sánchez-Diener, I., Doménech-Sánchez, A., Oliver, A. and Albertí, S. 2021. Molecular Analysis of the Contribution of Alkaline Protease A and Elastase B to the Virulence of Pseudomonas aeruginosa Bloodstream Infections. *Front Cell Infect Microbiol* 11, t. 816356. doi: 10.3389/fcimb.2021.816356

Mathee, K. *et al.* 2008. Dynamics of Pseudomonas aeruginosa genome evolution. *Proc Natl Acad Sci U S A* 105(8), tt. 3100-3105. doi: 10.1073/pnas.0711982105

Mattison, K., Wilbur, J. S., So, M. and Brennan, R. G. 2006. Structure of FitAB from Neisseria gonorrhoeae bound to DNA reveals a tetramer of toxin-antitoxin heterodimers containing pin domains and ribbon-helix-helix motifs. *J Biol Chem* 281(49), tt. 37942-37951. doi: 10.1074/jbc.M605198200

Mayer-Hamblett, N. *et al.* 2014. Pseudomonas aeruginosa phenotypes associated with eradication failure in children with cystic fibrosis. *Clin Infect Dis* 59(5), tt. 624-631. doi: 10.1093/cid/ciu385

McClean, S. and Callaghan, M. 2009. Burkholderia cepacia complex: epithelial cell–pathogen confrontations and potential for therapeutic intervention. *Journal of medical microbiology* 58(1), tt. 1-12.

McClean, S. *et al.* 2016. Linocin and OmpW Are Involved in Attachment of the Cystic Fibrosis-Associated Pathogen Burkholderia cepacia Complex to Lung Epithelial Cells and Protect Mice against Infection. *Infect Immun* 84(5), tt. 1424-1437. doi: 10.1128/iai.01248-15

McDonald, M. J. 2019. Microbial Experimental Evolution - a proving ground for evolutionary theory and a tool for discovery. *EMBO Rep* 20(8), t. e46992. doi: 10.15252/embr.201846992

McDowell, A. *et al.* 2001. PCR-based detection and identification of Burkholderia cepacia complex pathogens in sputum from cystic fibrosis patients. *J Clin Microbiol* 39(12), tt. 4247-4255. doi: 10.1128/jcm.39.12.4247-4255.2001

McKeon, S. A., Nguyen, D. T., Viteri, D. F., Zlosnik, J. E. and Sokol, P. A. 2011. Functional quorum sensing systems are maintained during chronic Burkholderia cepacia complex infections in patients with cystic fibrosis. *J Infect Dis* 203(3), tt. 383-392. doi: 10.1093/infdis/jiq054

Mesureur, J. *et al.* 2017. Macrophages, but not neutrophils, are critical for proliferation of Burkholderia cenocepacia and ensuing host-damaging inflammation. *PLoS Pathog* 13(6), t. e1006437. doi: 10.1371/journal.ppat.1006437

Miles, A. A., Misra, S. S. and Irwin, J. O. 1938. The estimation of the bactericidal power of the blood. *J Hyg (Lond)* 38(6), tt. 732-749. doi: 10.1017/s002217240001158x

Millar-Jones, L., Ryley, H. C., Paull, A. and Goodchild, M. C. 1998. Transmission and prevalence of Burkholderia cepacia in Welsh cystic fibrosis patients. *Respir Med* 92(2), tt. 178-183. doi: 10.1016/s0954-6111(98)90092-0

Morris, L. S., Evans, J. and Marchesi, J. R. 2012. A robust plate assay for detection of extracellular microbial protease activity in metagenomic screens and pure cultures. *J Microbiol Methods* 91(1), tt. 144-146. doi: 10.1016/j.mimet.2012.08.006

Mullen, T., Markey, K., Murphy, P., McClean, S. and Callaghan, M. 2007. Role of lipase in Burkholderia cepacia complex (Bcc) invasion of lung epithelial cells. *European Journal of Clinical Microbiology & Infectious Diseases* 26(12), tt. 869-877. doi: 10.1007/s10096-007-0385-2

Mullins, A. J. and Mahenthiralingam, E. 2021. The Hidden Genomic Diversity, Specialized Metabolite Capacity, and Revised Taxonomy of Burkholderia Sensu Lato. *Front Microbiol* 12, t. 726847. doi: 10.3389/fmicb.2021.726847

Mullins, A. J. *et al.* 2019. Genome mining identifies cepacin as a plant-protective metabolite of the biopesticidal bacterium Burkholderia ambifaria. *Nature Microbiology* 4(6), tt. 996-1005. doi: 10.1038/s41564-019-0383-z

Murray, C. 2022. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *Lancet* 399(10325), tt. 629-655. doi: 10.1016/s0140-6736(21)02724-0

Murray, G. L., Attridge, S. R. and Morona, R. 2006. Altering the length of the lipopolysaccharide O antigen has an impact on the interaction of Salmonella enterica serovar Typhimurium with macrophages and complement. *J Bacteriol* 188(7), tt. 2735-2739. doi: 10.1128/jb.188.7.2735-2739.2006

Nair, B. M., Cheung, K. J., Jr., Griffith, A. and Burns, J. L. 2004. Salicylate induces an antibiotic efflux pump in Burkholderia cepacia complex genomovar III (B. cenocepacia). *J Clin Invest* 113(3), tt. 464-473. doi: 10.1172/jci19710

Narayanamurthy, V., Sweetnam, J. M., Denner, D. R., Chen, L. W., Naureckas, E. T., Laxman, B. and White, S. R. 2017. The metabolic footprint of the airway bacterial community in cystic fibrosis. *Microbiome* 5(1), t. 67. doi: 10.1186/s40168-017-0289-z

Needham, B. D. and Trent, M. S. 2013. Fortifying the barrier: the impact of lipid A remodelling on bacterial pathogenesis. *Nat Rev Microbiol* 11(7), tt. 467-481. doi: 10.1038/nrmicro3047

Newmark, P. 1985. Testing for cystic fibrosis. *Nature* 318(6044), t. 309. doi: 10.1038/318309a0

Nickzad, A., Lépine, F. and Déziel, E. 2015. Quorum Sensing Controls Swarming Motility of Burkholderia glumae through Regulation of Rhamnolipids. *PLoS One* 10(6), t. e0128509. doi: 10.1371/journal.pone.0128509

Nishiyama, E., Ohtsubo, Y., Nagata, Y. and Tsuda, M. 2010. Identification of Burkholderia multivorans ATCC 17616 genes induced in soil environment by in vivo expression technology. *Environ Microbiol* 12(9), tt. 2539-2558. doi: 10.1111/j.1462-2920.2010.02227.x

Nordmann, P. and Poirel, L. 2002. Emerging carbapenemases in Gram-negative aerobes. *Clin Microbiol Infect* 8(6), tt. 321-331. doi: 10.1046/j.1469-0691.2002.00401.x

Nunvar, J. *et al.* 2016. Understanding the Pathogenicity of Burkholderia contaminans, an Emerging Pathogen in Cystic Fibrosis. *PLoS One* 11(8), t. e0160975. doi: 10.1371/journal.pone.0160975

Nzula, S., Vandamme, P. and Govan, J. R. 2002. Influence of taxonomic status on the in vitro antimicrobial susceptibility of the Burkholderia cepacia complex. *J Antimicrob Chemother* 50(2), tt. 265-269. doi: 10.1093/jac/dkf137

O'Callaghan, E. M., Tanner, M. S. and Boulnois, G. J. 1994. Development of a PCR probe test for identifying Pseudomonas aeruginosa and Pseudomonas (Burkholderia) cepacia. *J Clin Pathol* 47(3), tt. 222-226. doi: 10.1136/jcp.47.3.222

O'May, C. and Tufenkji, N. 2011. The swarming motility of Pseudomonas aeruginosa is blocked by cranberry proanthocyanidins and other tannin-containing materials. *Appl Environ Microbiol* 77(9), tt. 3061-3067. doi: 10.1128/aem.02677-10

O'Toole, G. A. 2011. Microtiter dish biofilm formation assay. *J Vis Exp* (47). doi: 10.3791/2437

O’Carroll, M. R., Kidd, T. J., Coulter, C., Smith, H. V., Rose, B. R., Harbour, C. and Bell, S. C. 2003. Burkholderia pseudomallei: another emerging pathogen in cystic fibrosis. *Thorax* 58(12), tt. 1087-1091. doi: 10.1136/thorax.58.12.1087

Ohtsubo, Y., Genka, H., Komatsu, H., Nagata, Y. and Tsuda, M. 2005. High-temperature-induced transposition of insertion elements in burkholderia multivorans ATCC 17616. *Appl Environ Microbiol* 71(4), tt. 1822-1828. doi: 10.1128/aem.71.4.1822-1828.2005

Olaitan, A. O., Morand, S. and Rolain, J. M. 2014. Mechanisms of polymyxin resistance: acquired and intrinsic resistance in bacteria. *Front Microbiol* 5, t. 643. doi: 10.3389/fmicb.2014.00643

Omar, N., Raouf, H. A. E., Okasha, H. and Nabil, N. 2015. Microbiological assessment of Burkholderia cepacia complex (Bcc) isolates in Alexandria Main University Hospital. *Alexandria Journal of Medicine* 51(1), tt. 41-46. doi: 10.1016/j.ajme.2014.08.005

Ortega, X., Silipo, A., Saldías, M. S., Bates, C. C., Molinaro, A. and Valvano, M. A. 2009. Biosynthesis and structure of the Burkholderia cenocepacia K56-2 lipopolysaccharide core

oligosaccharide: truncation of the core oligosaccharide leads to increased binding and sensitivity to polymyxin B. *J Biol Chem* 284(32), tt. 21738-21751. doi: 10.1074/jbc.M109.008532

Overbeek, R. *et al.* 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* 33(17), tt. 5691-5702. doi: 10.1093/nar/gki866

Overbeek, R. *et al.* 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res* 42(Database issue), tt. D206-214. doi: 10.1093/nar/gkt1226

Ozer, E. A., Nnah, E., Didelot, X., Whitaker, R. J. and Hauser, A. R. 2019. The Population Structure of Pseudomonas aeruginosa Is Characterized by Genetic Isolation of exoU+ and exoS+ Lineages. *Genome Biol Evol* 11(1), tt. 1780-1796. doi: 10.1093/gbe/evz119

Page, A. J. *et al.* 2015. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31(22), tt. 3691-3693. doi: 10.1093/bioinformatics/btv421

Page, A. J., Taylor, B., Delaney, A. J., Soares, J., Seemann, T., Keane, J. A. and Harris, S. R. 2016a. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb Genom* 2(4), t. e000056. doi: 10.1099/mgen.0.000056

Page, A. J., Taylot, B. and and Keane, J. A. 2016b. Multilocus sequence typing by blast from de novo assemblies against PubMLST. *The Journal of Open Source Software* 1(8), t. 118.

Palleroni, N. and Holmes, B. 1981. Pseudomonas cepacia sp. nov., nom. rev. *International Journal of Systematic and Evolutionary Microbiology* 31(4), tt. 479-481.

Parkinson, J. S. 1976. cheA, cheB, and cheC genes of Escherichia coli and their role in chemotaxis. *J Bacteriol* 126(2), tt. 758-770. doi: 10.1128/jb.126.2.758-770.1976

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. and Tyson, G. W. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25(7), tt. 1043-1055. doi: 10.1101/gr.186072.114

Parmar, J. S. and Nasser, S. 2005. Antibiotic allergy in cystic fibrosis. *Thorax* 60(6), tt. 517-520. doi: 10.1136/thx.2004.027953

Peeters, C., Cooper, V. S., Hatcher, P. J., Verheyde, B., Carlier, A. and Vandamme, P. 2017. Comparative genomics of Burkholderia multivorans, a ubiquitous pathogen with a highly conserved genomic structure. *PLoS One* 12(4), t. e0176191. doi: 10.1371/journal.pone.0176191

Pier, G. B. 2007. Pseudomonas aeruginosa lipopolysaccharide: a major virulence factor, initiator of inflammation and target for effective immunity. *Int J Med Microbiol* 297(5), tt. 277-295. doi: 10.1016/j.ijmm.2007.03.012

Pirnay, J. P. *et al.* 2009. Pseudomonas aeruginosa population structure revisited. *PLoS One* 4(11), t. e7740. doi: 10.1371/journal.pone.0007740

Pirone, L. *et al.* 2008. Burkholderia cenocepacia strains isolated from cystic fibrosis patients are apparently more invasive and more virulent than rhizosphere strains. *Environ Microbiol* 10(10), tt. 2773-2784. doi: 10.1111/j.1462-2920.2008.01697.x

Plesa, M., Hernalsteens, J. P., Vandenbussche, G., Ruysschaert, J. M. and Cornelis, P. 2006. The SlyB outer membrane lipoprotein of Burkholderia multivorans contributes to membrane integrity. *Res Microbiol* 157(6), tt. 582-592. doi: 10.1016/j.resmic.2005.11.015

Podnecky, N. L., Rhodes, K. A. and Schweizer, H. P. 2015. Efflux pump-mediated drug resistance in Burkholderia. *Front Microbiol* 6, t. 305. doi: 10.3389/fmicb.2015.00305

Poirel, L., Rodriguez-Martinez, J. M., Plésiat, P. and Nordmann, P. 2009. Naturally occurring Class A ss-lactamases from the Burkholderia cepacia complex. *Antimicrob Agents Chemother* 53(3), tt. 876-882. doi: 10.1128/aac.00946-08

Prayle, A. and Smyth, A. R. 2010. Aminoglycoside use in cystic fibrosis: therapeutic strategies and toxicity. *Curr Opin Pulm Med* 16(6), tt. 604-610. doi: 10.1097/MCP.0b013e32833eebfd

Price, K. E. *et al.* 2013. Unique microbial communities persist in individual cystic fibrosis patients throughout a clinical exacerbation. *Microbiome* 1(1), t. 27. doi: 10.1186/2049-2618-1-27

Price, M. N., Dehal, P. S. and Arkin, A. P. 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* 26(7), tt. 1641-1650. doi: 10.1093/molbev/msp077

Prince, A., Wood, M. S., Cacalano, G. S. and Chin, N. X. 1988. Isolation and characterization of a penicillinase from Pseudomonas cepacia 249. *Antimicrob Agents Chemother* 32(6), tt. 838-843. doi: 10.1128/aac.32.6.838

Pritchard, L. 2014. Pyani: python module for average nucleotide identity analyses.

Pritchard, L., Glover, R. H., Humphris, S., Elphinstone, J. G. and Toth, I. K. 2016. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Analytical Methods* 8(1), tt. 12-24. doi: 10.1039/c5ay02550h

Project, I. 2020. Inkscape.

Pruitt, K. D., Tatusova, T. and Maglott, D. R. 2006. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research* 35(suppl_1), tt. D61-D65. doi: 10.1093/nar/gkl842

Quinlan, A. R. and Hall, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6), tt. 841-842. doi: 10.1093/bioinformatics/btq033

Quinn, R. J. *et al.* 2021. Periodically Disturbing the Spatial Structure of Biofilms Can Affect the Production of an Essential Virulence Factor in Pseudomonas aeruginosa. *mSystems* 6(5), t. e0096121. doi: 10.1128/mSystems.00961-21

R-Core-Team. 2013. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Radlinski, L. *et al.* 2017. Pseudomonas aeruginosa exoproducts determine antibiotic efficacy against Staphylococcus aureus. *PLoS Biol* 15(11), t. e2003981. doi: 10.1371/journal.pbio.2003981

Rambaut, A. 2009. FigTree v1. 3.1. *http://tree. bio. ed. ac. uk/software/figtree/.*

Ramsay, K. A., Butler, C. A., Paynter, S., Ware, R. S., Kidd, T. J., Wainwright, C. E. and Bell, S. C. 2013. Factors influencing acquisition of Burkholderia cepacia complex organisms in patients with cystic fibrosis. *J Clin Microbiol* 51(12), tt. 3975-3980. doi: 10.1128/jcm.01360-13

Randall, L. B., Dobos, K., Papp-Wallace, K. M., Bonomo, R. A. and Schweizer, H. P. 2015. Membrane-Bound PenA β-Lactamase of Burkholderia pseudomallei. *Antimicrob Agents Chemother* 60(3), tt. 1509-1514. doi: 10.1128/aac.02444-15

Rashid, M. H. and Kornberg, A. 2000. Inorganic polyphosphate is needed for swimming, swarming, and twitching motilities of Pseudomonas aeruginosa. *Proc Natl Acad Sci U S A* 97(9), tt. 4885-4890. doi: 10.1073/pnas.060030097

Rau, M. H. *et al.* 2010. Early adaptive developments of Pseudomonas aeruginosa after the transition from life in the environment to persistent colonization in the airways of human cystic fibrosis hosts. *Environ Microbiol* 12(6), tt. 1643-1658. doi: 10.1111/j.1462-2920.2010.02211.x

Ravi, R. K., Walton, K. and Khosroheidari, M. 2018. MiSeq: A Next Generation Sequencing Platform for Genomic Analysis. *Methods Mol Biol* 1706, tt. 223-232. doi: 10.1007/978-1-4939-7471-9_12

Razvi, S., Quittell, L., Sewall, A., Quinton, H., Marshall, B. and Saiman, L. 2009. Respiratory microbiology of patients with cystic fibrosis in the United States, 1995 to 2005. *Chest* 136(6), tt. 1554-1560. doi: 10.1378/chest.09-0132

Regan, K. H. and Bhatt, J. 2016. Eradication therapy for Burkholderia cepacia complex in people with cystic fibrosis. *Cochrane Database Syst Rev* 11, t. Cd009876. doi: 10.1002/14651858.CD009876.pub3

Rhodes, K. A. and Schweizer, H. P. 2016. Antibiotic resistance in Burkholderia species. *Drug Resist Updat* 28, tt. 82-90. doi: 10.1016/j.drup.2016.07.003

Ribeiro-Gonçalves, B., Francisco, A. P., Vaz, C., Ramirez, M. and Carriço, J. A. 2016. PHYLOViZ Online: web-based tool for visualization, phylogenetic inference, analysis and sharing of minimum spanning trees. *Nucleic Acids Res* 44(W1), tt. W246-251. doi: 10.1093/nar/gkw359

Ridderberg, W., Nielsen, S. M. and Nørskov-Lauritsen, N. 2015. Genetic Adaptation of Achromobacter sp. during Persistence in the Lungs of Cystic Fibrosis Patients. *PLoS One* 10(8), t. e0136790. doi: 10.1371/journal.pone.0136790

Riedel, K. *et al.* 2001. N-acylhomoserine-lactone-mediated communication between Pseudomonas aeruginosa and Burkholderia cepacia in mixed biofilms. *Microbiology (Reading)* 147(Pt 12), tt. 3249-3262. doi: 10.1099/00221287-147-12-3249

Rodríguez-Beltrán, J., DelaFuente, J., León-Sampedro, R., MacLean, R. C. and San Millán, Á. 2021. Beyond horizontal gene transfer: the role of plasmids in bacterial evolution. *Nat Rev Microbiol* 19(6), tt. 347-359. doi: 10.1038/s41579-020-00497-1

Rogers, G. B., Hoffman, L. R., Whiteley, M., Daniels, T. W., Carroll, M. P. and Bruce, K. D. 2010. Revealing the dynamics of polymicrobial infections: implications for antibiotic therapy. *Trends Microbiol* 18(8), tt. 357-364. doi: 10.1016/j.tim.2010.04.005

Rosselló-Mora, R. 2005. Updating prokaryotic taxonomy. *J Bacteriol* 187(18), tt. 6255-6257. doi: 10.1128/jb.187.18.6255-6257.2005

Roux, D. *et al.* 2018. A putative lateral flagella of the cystic fibrosis pathogen Burkholderia dolosa regulates swimming motility and host cytokine production. *PLoS One* 13(1), t. e0189810. doi: 10.1371/journal.pone.0189810

RStudio-Team. 2020. RStudio: Integrated Development for R. PBC, Boston, MA.

Rushton, L., Khodr, A., Menard-Szczebara, F., Maillard, J. Y., Cupferman, S. and Mahenthiralingam, E. 2020. Mapping the Efficacy and Mode of Action of Ethylzingerone [4-

(3-Ethoxy-4-Hydroxyphenyl) Butan-2-One] as an Active Agent against Burkholderia Bacteria. *Appl Environ Microbiol* 86(19). doi: 10.1128/aem.01808-20

Sabharwal, S. 2016. Gastrointestinal Manifestations of Cystic Fibrosis. *Gastroenterol Hepatol (N Y)* 12(1), tt. 43-47.

Sage, A., Linker, A., Evans, R. L., and Lessie, T. G. 1990. Hexose phosphate metabolism and exopolysaccharide formation inPseudomonas cepacia. *Current Microbiology* 20, tt. 191-198. doi: https://doi.org/10.1007/BF02091996

Saiman, L. *et al.* 2014. Infection prevention and control guideline for cystic fibrosis: 2013 update. *Infect Control Hosp Epidemiol* 35 Suppl 1, tt. S1-s67. doi: 10.1086/676882

Sajjan, U., Liu, L., Lu, A., Spilker, T., Forstner, J. and LiPuma, J. J. 2002. Lack of cable pili expression by cblA-containing Burkholderia cepacia complex. *Microbiology (Reading)* 148(Pt 11), tt. 3477-3484. doi: 10.1099/00221287-148-11-3477

Sajjan, U. *et al.* 2001. Enhanced susceptibility to pulmonary infection with Burkholderia cepacia in Cftr−/− mice. *Infection and immunity* 69(8), tt. 5138-5150.

Sajjan, U. S., Sun, L., Goldstein, R. and Forstner, J. F. 1995. Cable (cbl) type II pili of cystic fibrosis-associated Burkholderia (Pseudomonas) cepacia: nucleotide sequence of the cblA major subunit pilin gene and novel morphology of the assembled appendage fibers. *J Bacteriol* 177(4), tt. 1030-1038. doi: 10.1128/jb.177.4.1030-1038.1995

Sajjan, U. S., Xie, H., Lefebre, M. D., Valvano, M. A. and Forstner, J. F. 2003. Identification and molecular analysis of cable pilus biosynthesis genes in Burkholderia cepacia. *Microbiology* 149(4), tt. 961-971.

Saldías, M. S., Ortega, X. and Valvano, M. A. 2009. Burkholderia cenocepacia O antigen lipopolysaccharide prevents phagocytosis by macrophages and adhesion to epithelial cells. *Journal of medical microbiology* 58(12), tt. 1542-1548.

Salloum, T., Nassour, E., Araj, G. F., Abboud, E. and Tokajian, S. 2018. Insights into the genome diversity and virulence of two clinical isolates of Burkholderia cenocepacia. *J Med Microbiol* 67(8), tt. 1157-1167. doi: 10.1099/jmm.0.000759

Salsgiver, E. L., Fink, A. K., Knapp, E. A., LiPuma, J. J., Olivier, K. N., Marshall, B. C. and Saiman, L. 2016. Changing Epidemiology of the Respiratory Bacteriology of Patients With Cystic Fibrosis. *Chest* 149(2), tt. 390-400. doi: 10.1378/chest.15-0676

Salunkhe, P. *et al.* 2005. A cystic fibrosis epidemic strain of Pseudomonas aeruginosa displays enhanced virulence and antimicrobial resistance. *J Bacteriol* 187(14), tt. 4908-4920. doi: 10.1128/jb.187.14.4908-4920.2005

Sangiovanni, M., Granata, I., Thind, A. S. and Guarracino, M. R. 2019. From trash to treasure: detecting unexpected contamination in unmapped NGS data. *BMC Bioinformatics* 20(Suppl 4), t. 168. doi: 10.1186/s12859-019-2684-x

Sant'Anna, F. H., Trentini, D. B., de Souto Weber, S., Cecagno, R., da Silva, S. C. and Schrank, I. S. 2009. The PII superfamily revised: a novel group and evolutionary insights. *J Mol Evol* 68(4), tt. 322-336. doi: 10.1007/s00239-009-9209-6

Sass, A. M., Schmerk, C., Agnoli, K., Norville, P. J., Eberl, L., Valvano, M. A. and Mahenthiralingam, E. 2013. The unexpected discovery of a novel low-oxygen-activated locus for the anoxic persistence of Burkholderia cenocepacia. *Isme j* 7(8), tt. 1568-1581. doi: 10.1038/ismej.2013.36

Sawana, A., Adeolu, M. and Gupta, R. S. 2014. Molecular signatures and phylogenomic analysis of the genus Burkholderia: proposal for division of this genus into the emended genus Burkholderia containing pathogenic organisms and a new genus Paraburkholderia gen. nov. harboring environmental species. *Front Genet* 5, t. 429. doi: 10.3389/fgene.2014.00429

Schbath, S., Martin, V., Zytnicki, M., Fayolle, J., Loux, V. and Gibrat, J. F. 2012. Mapping reads on a genomic sequence: an algorithmic overview and a practical comparative analysis. *J Comput Biol* 19(6), tt. 796-813. doi: 10.1089/cmb.2012.0022

Schmerk, C. L. and Valvano, M. A. 2013. Burkholderia multivorans survival and trafficking within macrophages. *J Med Microbiol* 62(Pt 2), tt. 173-184. doi: 10.1099/jmm.0.051243-0

Schmid, N. *et al.* 2012. The AHL- and BDSF-dependent quorum sensing systems control specific and overlapping sets of genes in Burkholderia cenocepacia H111. *PLoS One* 7(11), t. e49966. doi: 10.1371/journal.pone.0049966

Schwab, U. *et al.* 2014. Localization of Burkholderia cepacia complex bacteria in cystic fibrosis lungs and interactions with Pseudomonas aeruginosa in hypoxic mucus. *Infect Immun* 82(11), tt. 4729-4745. doi: 10.1128/iai.01876-14

Scoffone, V. C., Chiarelli, L. R., Trespidi, G., Mentasti, M., Riccardi, G. and Buroni, S. 2017. Burkholderia cenocepacia Infections in Cystic Fibrosis Patients: Drug Resistance and Therapeutic Approaches. *Front Microbiol* 8, t. 1592. doi: 10.3389/fmicb.2017.01592

Seed, K. D. and Dennis, J. J. 2005. Isolation and characterization of bacteriophages of the Burkholderia cepacia complex. *FEMS Microbiol Lett* 251(2), tt. 273-280. doi: 10.1016/j.femsle.2005.08.011

Seed, K. D. and Dennis, J. J. 2008. Development of Galleria mellonella as an alternative infection model for the Burkholderia cepacia complex. *Infect Immun* 76(3), tt. 1267-1275. doi: 10.1128/iai.01249-07

Seed, K. D. and Dennis, J. J. 2009. Experimental bacteriophage therapy increases survival of Galleria mellonella larvae infected with clinically relevant strains of the Burkholderia cepacia complex. *Antimicrobial agents and chemotherapy* 53(5), tt. 2205-2208.

Seemann, T. MLST. Github.

Seemann, T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30(14), tt. 2068-2069. doi: 10.1093/bioinformatics/btu153

Seemann, T. 2017. Abricate.

Seemann, T. 2018. Snippy: fast bacterial variant calling from NGS reads [Internet]

Segonds, C. *et al.* 1997. Genotypic analysis of Burkholderia cepacia isolates from 13 French cystic fibrosis centers. *J Clin Microbiol* 35(8), tt. 2055-2060. doi: 10.1128/jcm.35.8.2055-2060.1997

Segonds, C., Heulin, T., Marty, N. and Chabanon, G. 1999. Differentiation of Burkholderia species by PCR-restriction fragment length polymorphism analysis of the 16S rRNA gene and application to cystic fibrosis isolates. *J Clin Microbiol* 37(7), tt. 2201-2208. doi: 10.1128/jcm.37.7.2201-2208.1999

Sengyee, S., Yoon, S. H., West, T. E., Ernst, R. K. and Chantratita, N. 2019. Lipopolysaccharides from Different Burkholderia Species with Different Lipid A Structures Induce Toll-Like Receptor 4 Activation and React with Melioidosis Patient Sera. *Infect Immun* 87(12). doi: 10.1128/iai.00692-19

Sharma, P., Gupta, S. K. and Rolain, J. M. 2014. Whole genome sequencing of bacteria in cystic fibrosis as a model for bacterial genome adaptation and evolution. *Expert Rev Anti Infect Ther* 12(3), tt. 343-355. doi: 10.1586/14787210.2014.887441

Shastri, S. *et al.* 2017. An efficient system for the generation of marked genetic mutants in members of the genus Burkholderia. *Plasmid* 89, tt. 49-56. doi: 10.1016/j.plasmid.2016.11.002

Silipo, A. *et al.* 2007. The complete structure and pro-inflammatory activity of the lipooligosaccharide of the highly epidemic and virulent gram-negative bacterium Burkholderia cenocepacia ET-12 (strain J2315). *Chemistry* 13(12), tt. 3501-3511. doi: 10.1002/chem.200601406

Silva, I. N. *et al.* 2011. Mucoid morphotype variation of Burkholderia multivorans during chronic cystic fibrosis lung infection is correlated with changes in metabolism, motility, biofilm formation and virulence. *Microbiology (Reading)* 157(Pt 11), tt. 3124-3137. doi: 10.1099/mic.0.050989-0

Silva, I. N., Pessoa, F. D., Ramires, M. J., Santos, M. R., Becker, J. D., Cooper, V. S. and Moreira, L. M. 2018. The OmpR Regulator of Burkholderia multivorans Controls Mucoid-to-Nonmucoid Transition and Other Cell Envelope Properties Associated with Persistence in the Cystic Fibrosis Lung. *J Bacteriol* 200(17). doi: 10.1128/jb.00216-18

Silva, I. N. *et al.* 2016. Long-Term Evolution of Burkholderia multivorans during a Chronic Cystic Fibrosis Infection Reveals Shifting Forces of Selection. *mSystems* 1(3). doi: 10.1128/mSystems.00029-16

Skerman, V. B. D., McGowan, V. and Sneath, P. H. A. 1989. Approved lists of bacterial names (amended).

Smith, E. E. *et al.* 2006. Genetic adaptation by Pseudomonas aeruginosa to the airways of cystic fibrosis patients. *Proc Natl Acad Sci U S A* 103(22), tt. 8487-8492. doi: 10.1073/pnas.0602138103

Sokol, P., Sajjan, U., Visser, M., Gingues, S., Forstner, J. and Kooi, C. 2003a. The CepIR quorum-sensing system contributes to the virulence of Burkholderia cenocepacia respiratory infections. *Microbiology* 149(12), tt. 3649-3658.

Sokol, P. A., Malott, R. J., Riedel, K. and Eberl, L. 2007. Communication systems in the genus Burkholderia: global regulators and targets for novel antipathogenic drugs. *Future Microbiol* 2(5), tt. 555-563. doi: 10.2217/17460913.2.5.555

Sokol, P. A., Sajjan, U., Visser, M. B., Gingues, S., Forstner, J. and Kooi, C. 2003b. The CepIR quorum-sensing system contributes to the virulence of Burkholderia cenocepacia respiratory infections. *Microbiology (Reading)* 149(Pt 12), tt. 3649-3658. doi: 10.1099/mic.0.26540-0

Somprasong, N. *et al.* 2021. Burkholderia ubonensis High-Level Tetracycline Resistance Is Due to Efflux Pump Synergy Involving a Novel TetA(64) Resistance Determinant. *Antimicrob Agents Chemother* 65(3). doi: 10.1128/aac.01767-20

Sousa, S. A., Feliciano, J. R., Pita, T., Guerreiro, S. I. and Leitão, J. H. 2017. Burkholderia cepacia Complex Regulation of Virulence Gene Expression: A Review. *Genes (Basel)* 8(1). doi: 10.3390/genes8010043

Sousa, S. A. *et al.* 2007. Virulence of Burkholderia cepacia complex strains in gp91phox−/− mice. *Cellular microbiology* 9(12), tt. 2817-2825.

Southern, K. W., Munck, A., Pollitt, R., Travert, G., Zanolla, L., Dankert-Roelse, J. and Castellani, C. 2007. A survey of newborn screening for cystic fibrosis in Europe. *J Cyst Fibros* 6(1), tt. 57-65. doi: 10.1016/j.jcf.2006.05.008

Speert, D. P., Henry, D., Vandamme, P., Corey, M. and Mahenthiralingam, E. 2002. Epidemiology of Burkholderia cepacia complex in patients with cystic fibrosis, Canada. *Emerg Infect Dis* 8(2), tt. 181-187. doi: 10.3201/eid0802.010163

Speert, D. P., Steen, B., Halsey, K. and Kwan, E. 1999. A murine model for infection with Burkholderia cepacia with sustained persistence in the spleen. *Infect Immun* 67(8), tt. 4027-4032. doi: 10.1128/iai.67.8.4027-4032.1999

Spiewak, H. L., Shastri, S., Zhang, L., Schwager, S., Eberl, L., Vergunst, A. C. and Thomas, M. S. 2019. Burkholderia cenocepacia utilizes a type VI secretion system for bacterial competition. *Microbiologyopen* 8(7), t. e774. doi: 10.1002/mbo3.774

Spilker, T., Baldwin, A., Bumford, A., Dowson, C. G., Mahenthiralingam, E. and LiPuma, J. J. 2009. Expanded multilocus sequence typing for burkholderia species. *J Clin Microbiol* 47(8), tt. 2607-2610. doi: 10.1128/jcm.00770-09

Spratt, B. G. and Maiden, M. C. 1999. Bacterial population genetics, evolution and epidemiology. *Philos Trans R Soc Lond B Biol Sci* 354(1384), tt. 701-710. doi: 10.1098/rstb.1999.0423

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9), tt. 1312-1313. doi: 10.1093/bioinformatics/btu033

Stanier, R. Y., Palleroni, N. J. and Doudoroff, M. 1966. The aerobic pseudomonads: a taxonomic study. *J Gen Microbiol* 43(2), tt. 159-271. doi: 10.1099/00221287-43-2-159

Starke, J. R., Edwards, M. S., Langston, C. and Baker, C. J. 1987. A mouse model of chronic pulmonary infection with Pseudomonas aeruginosa and Pseudomonas cepacia. *Pediatric research* 22(6), tt. 698-702.

Stover, C. K. *et al.* 2000. Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen. *Nature* 406(6799), tt. 959-964. doi: 10.1038/35023079

Straus, D. C., Lonon, M. K. and Hutson, J. C. 1992. Inhibition of rat alveolar macrophage phagocytic function by a Pseudomonas cepacia lipase. *J Med Microbiol* 37(5), tt. 335-340. doi: 10.1099/00222615-37-5-335

Sun, J. *et al.* 2020. The Pseudomonas aeruginosa protease LasB directly activates IL-1β. *EBioMedicine* 60, t. 102984. doi: 10.1016/j.ebiom.2020.102984

Sun, L. *et al.* 1995a. The emergence of a highly transmissible lineage of cbl+ Pseudomonas (Burkholderia) cepacia causing CF centre epidemics in North America and Britain. *Nature medicine* 1(7), tt. 661-666.

Sun, L. *et al.* 1995b. The emergence of a highly transmissible lineage of cbl+ Pseudomonas (Burkholderia) cepacia causing CF centre epidemics in North America and Britain. *Nat Med* 1(7), tt. 661-666. doi: 10.1038/nm0795-661

Suppiger, A., Schmid, N., Aguilar, C., Pessi, G. and Eberl, L. 2013. Two quorum sensing systems control biofilm formation and virulence in members of the Burkholderia cepacia complex. *Virulence* 4(5), tt. 400-409. doi: 10.4161/viru.25338

Tamma, P. D. *et al.* 2018. Successful Treatment of Persistent Burkholderia cepacia Complex Bacteremia with Ceftazidime-Avibactam. *Antimicrob Agents Chemother* 62(4). doi: 10.1128/aac.02213-17

Tange, O. 2018. GNU Parallel.

Tatusov, R. L., Galperin, M. Y., Natale, D. A. and Koonin, E. V. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* 28(1), tt. 33-36. doi: 10.1093/nar/28.1.33

Tettelin, H. *et al.* 2005. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". *Proc Natl Acad Sci U S A* 102(39), tt. 13950-13955. doi: 10.1073/pnas.0506758102

Tomich, M., Griffith, A., Herfst, C. A., Burns, J. L. and Mohr, C. D. 2003. Attenuated virulence of a Burkholderia cepacia type III secretion mutant in a murine model of infection. *Infection and immunity* 71(3), tt. 1405-1415.

Tomich, M., Herfst, C. A., Golden, J. W. and Mohr, C. D. 2002. Role of flagella in host cell invasion by Burkholderia cepacia. *Infect Immun* 70(4), tt. 1799-1806. doi: 10.1128/iai.70.4.1799-1806.2002

Toussaint, B., Delic-Attree, I. and Vignais, P. M. 1993. Pseudomonas aeruginosa contains an IHF-like protein that binds to the algD promoter. *Biochem Biophys Res Commun* 196(1), tt. 416-421. doi: 10.1006/bbrc.1993.2265

Tseng, S. P. *et al.* 2014. The contribution of antibiotic resistance mechanisms in clinical Burkholderia cepacia complex isolates: an emphasis on efflux pump activity. *PLoS One* 9(8), t. e104986. doi: 10.1371/journal.pone.0104986

Turton, J. F., Kaufmann, M. E., Mustafa, N., Kawa, S., Clode, F. E. and Pitt, T. L. 2003. Molecular comparison of isolates of Burkholderia multivorans from patients with cystic fibrosis in the United Kingdom. *J Clin Microbiol* 41(12), tt. 5750-5754. doi: 10.1128/jcm.41.12.5750-5754.2003

Uehlinger, S., Schwager, S., Bernier, S. P., Riedel, K., Nguyen, D. T., Sokol, P. A. and Eberl, L. 2009. Identification of specific and universal virulence factors in Burkholderia cenocepacia strains by using multiple infection hosts. *Infect Immun* 77(9), tt. 4102-4110. doi: 10.1128/iai.00398-09

UK-Cystic-Fibrosis-Registry. 2021. *UK Cystic Fibrosis Registry 2020 Annual Data Report*. Ar gael: https://www.cysticfibrosis.org.uk/sites/default/files/2022-05/2020%20Annual%20data%20report%20-%20Version%204.pdf [Gwelwyd.

Urban, T. A., Griffith, A., Torok, A. M., Smolkin, M. E., Burns, J. L. and Goldberg, J. B. 2004. Contribution of Burkholderia cenocepacia flagella to infectivity and inflammation. *Infect Immun* 72(9), tt. 5126-5134. doi: 10.1128/iai.72.9.5126-5134.2004

Valade, E., Thibault, F. M., Gauthier, Y. P., Palencia, M., Popoff, M. Y. and Vidal, D. R. 2004. The PmlI-PmlR quorum-sensing system in Burkholderia pseudomallei plays a key role in virulence and modulates production of the MprA protease. *J Bacteriol* 186(8), tt. 2288-2294. doi: 10.1128/jb.186.8.2288-2294.2004

Van Dalem, A. *et al.* 2018. In Vitro Susceptibility of Burkholderia cepacia Complex Isolated from Cystic Fibrosis Patients to Ceftazidime-Avibactam and Ceftolozane-Tazobactam. *Antimicrob Agents Chemother* 62(9). doi: 10.1128/aac.00590-18

Van de Kerkhove, C. *et al.* 2016. Continuous alternating inhaled antibiotic therapy in CF: A single center retrospective analysis. *J Cyst Fibros* 15(6), tt. 802-808. doi: 10.1016/j.jcf.2016.09.002

van der Graaf-van Bloois, L., Wagenaar, J. A. and Zomer, A. L. 2021. RFPlasmid: predicting plasmid sequences from short-read assembly data using machine learning. *Microb Genom* 7(11). doi: 10.1099/mgen.0.000683

Vandamme, P., Holmes, B., Coenye, T., Goris, J., Mahenthiralingam, E., LiPuma, J. J. and Govan, J. R. 2003. Burkholderia cenocepacia sp. nov.--a new twist to an old story. *Res Microbiol* 154(2), tt. 91-96. doi: 10.1016/s0923-2508(03)00026-3

Vandamme, P. *et al.* 1997. Occurrence of multiple genomovars of Burkholderia cepacia in cystic fibrosis patients and proposal of Burkholderia multivorans sp. nov. *Int J Syst Bacteriol* 47(4), tt. 1188-1200. doi: 10.1099/00207713-47-4-1188

Vandeplassche, E., Tavernier, S., Coenye, T. and Crabbé, A. 2019. Influence of the lung microbiome on antibiotic susceptibility of cystic fibrosis pathogens. *Eur Respir Rev* 28(152). doi: 10.1183/16000617.0041-2019

Veloira, W. G., Domenico, P., LiPuma, J. J., Davis, J. M., Gurzenda, E. and Kazzaz, J. A. 2003. In vitro activity and synergy of bismuth thiols and tobramycin against Burkholderia cepacia complex. *Journal of Antimicrobial Chemotherapy* 52(6), tt. 915-919.

Venturi, V., Friscina, A., Bertani, I., Devescovi, G. and Aguilar, C. 2004. Quorum sensing in the Burkholderia cepacia complex. *Research in microbiology* 155(4), tt. 238-244.

Vestby, L. K., Grønseth, T., Simm, R. and Nesse, L. L. 2020. Bacterial Biofilm and its Role in the Pathogenesis of Disease. *Antibiotics (Basel)* 9(2). doi: 10.3390/antibiotics9020059

Voynow, J. A., Fischer, B. M. and Zheng, S. 2008. Proteases and cystic fibrosis. *Int J Biochem Cell Biol* 40(6-7), tt. 1238-1245. doi: 10.1016/j.biocel.2008.03.003

Waine, D. J., Henry, D. A., Baldwin, A., Speert, D. P., Honeybourne, D., Mahenthiralingam, E. and Dowson, C. G. 2007. Reliability of multilocus sequence typing of the Burkholderia cepacia complex in cystic fibrosis. *J Cyst Fibros* 6(3), tt. 215-219. doi: 10.1016/j.jcf.2006.10.003

Wallner, A., King, E., Ngonkeu, E. L. M., Moulin, L. and Béna, G. 2019. Genomic analyses of Burkholderia cenocepacia reveal multiple species with differential host-adaptation to plants and humans. *BMC Genomics* 20(1), t. 803. doi: 10.1186/s12864-019-6186-z

Wang, H. *et al.* 2018. RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on Streptomyces coelicolor. *PLoS Comput Biol* 14(10), t. e1006541. doi: 10.1371/journal.pcbi.1006541

Wang, Y. *et al.* 2019. Indole Reverses Intrinsic Antibiotic Resistance by Activating a Novel Dual-Function Importer. *mBio* 10(3). doi: 10.1128/mBio.00676-19

Weiser, R. *et al.* 2019. Not all Pseudomonas aeruginosa are equal: strains from industrial sources possess uniquely large multireplicon genomes. *Microb Genom* 5(7). doi: 10.1099/mgen.0.000276

Whelan, F. J., Hall, R. J. and McInerney, J. O. 2021. Evidence for Selection in the Abundant Accessory Gene Content of a Prokaryote Pangenome. *Mol Biol Evol* 38(9), tt. 3697-3708. doi: 10.1093/molbev/msab139

Whiteford, M. L., Wilkinson, J. D., McColl, J. H., Conlon, F. M., Michie, J. R., Evans, T. J. and Paton, J. Y. 1995. Outcome of Burkholderia (Pseudomonas) cepacia colonisation in children with cystic fibrosis following a hospital outbreak. *Thorax* 50(11), tt. 1194-1198. doi: 10.1136/thx.50.11.1194

Wick, R. 2017. Filtlong. Github.

Wick, R. R. and Holt, K. E. 2019. Benchmarking of long-read assemblers for prokaryote whole genome sequencing. *F1000Res* 8, t. 2138. doi: 10.12688/f1000research.21782.4

Wick, R. R. *et al.* 2021. Trycycler: consensus long-read assemblies for bacterial genomes. *Genome Biol* 22(1), t. 266. doi: 10.1186/s13059-021-02483-z

Wick, R. R., Judd, L. M., Gorrie, C. L. and Holt, K. E. 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 13(6), t. e1005595. doi: 10.1371/journal.pcbi.1005595

Wigfield, S. M., Rigg, G. P., Kavari, M., Webb, A. K., Matthews, R. C. and Burnie, J. P. 2002. Identification of an immunodominant drug efflux pump in Burkholderia cepacia. *J Antimicrob Chemother* 49(4), tt. 619-624. doi: 10.1093/jac/49.4.619

Winson, M. K. *et al.* 1998. Construction and analysis of luxCDABE-based plasmid sensors for investigating N-acyl homoserine lactone-mediated quorum sensing. *FEMS Microbiol Lett* 163(2), tt. 185-192. doi: 10.1111/j.1574-6968.1998.tb13044.x

Winstanley, C. *et al.* 2009. Newly introduced genomic prophage islands are critical determinants of in vivo competitiveness in the Liverpool Epidemic Strain of Pseudomonas aeruginosa. *Genome Res* 19(1), tt. 12-23. doi: 10.1101/gr.086082.108

Winstanley, C., O'Brien, S. and Brockhurst, M. A. 2016. Pseudomonas aeruginosa Evolutionary Adaptation and Diversification in Cystic Fibrosis Chronic Lung Infections. *Trends Microbiol* 24(5), tt. 327-337. doi: 10.1016/j.tim.2016.01.008

Wolfgang, M. C., Jyot, J., Goodman, A. L., Ramphal, R. and Lory, S. 2004. Pseudomonas aeruginosa regulates flagellin expression as part of a global response to airway fluid from cystic fibrosis patients. *Proc Natl Acad Sci U S A* 101(17), tt. 6664-6668. doi: 10.1073/pnas.0307553101

Wood, D. E., Lu, J. and Langmead, B. 2019. Improved metagenomic analysis with Kraken 2. *Genome Biol* 20(1), t. 257. doi: 10.1186/s13059-019-1891-0

Xie, Z. and Tang, H. 2017. ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics* 33(21), tt. 3340-3347. doi: 10.1093/bioinformatics/btx433

Xu, J. Z. and Zhang, W. G. 2016. Strategies used for genetically modifying bacterial genome: site-directed mutagenesis, gene inactivation, and gene over-expression. *J Zhejiang Univ Sci B* 17(2), tt. 83-99. doi: 10.1631/jzus.B1500187

Yabuuchi, E. *et al.* 1992. Proposal of Burkholderia gen. nov. and transfer of seven species of the genus Pseudomonas homology group II to the new genus, with the type species Burkholderia cepacia (Palleroni and Holmes 1981) comb. nov. *Microbiol Immunol* 36(12), tt. 1251-1275. doi: 10.1111/j.1348-0421.1992.tb02129.x

Yabuuchi, E., Kosako, Y., Yano, I., Hotta, H. and Nishiuchi, Y. 1995. Transfer of two Burkholderia and an Alcaligenes species to Ralstonia gen. nov.: proposal of Ralstonia pickettii (Ralston, Palleroni and Doudoroff 1973) comb. nov., Ralstonia solanacearum (Smith 1896) comb. nov. and Ralstonia eutropha (Davis 1969) comb. nov. *Microbiology and immunology* 39(11), tt. 897-904.

Yagi, J. M., Sims, D., Brettin, T., Bruce, D. and Madsen, E. L. 2009. The genome of Polaromonas naphthalenivorans strain CJ2, isolated from coal tar-contaminated sediment, reveals physiological and metabolic versatility and evolution through extensive horizontal gene transfer. *Environ Microbiol* 11(9), tt. 2253-2270. doi: 10.1111/j.1462-2920.2009.01947.x

Yang, L., Haagensen, J. A., Jelsbak, L., Johansen, H. K., Sternberg, C., Høiby, N. and Molin, S. 2008. In situ growth rates and biofilm development of Pseudomonas aeruginosa populations in chronic lung infections. *J Bacteriol* 190(8), tt. 2767-2776. doi: 10.1128/jb.01581-07

Yoder-Himes, D. R., Chain, P. S., Zhu, Y., Wurtzel, O., Rubin, E. M., Tiedje, J. M. and Sorek, R. 2009. Mapping the Burkholderia cenocepacia niche response via high-throughput sequencing. *Proc Natl Acad Sci U S A* 106(10), tt. 3976-3981. doi: 10.1073/pnas.0813403106

Yoon, S. H., Ha, S. M., Lim, J., Kwon, S. and Chun, J. 2017. A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie Van Leeuwenhoek* 110(10), tt. 1281-1286. doi: 10.1007/s10482-017-0844-4

Zankari, E. *et al.* 2012. Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* 67(11), tt. 2640-2644. doi: 10.1093/jac/dks261

Zhao, J. *et al.* 2012. Decade-long bacterial community dynamics in cystic fibrosis airways. *Proc Natl Acad Sci U S A* 109(15), tt. 5809-5814. doi: 10.1073/pnas.1120577109

Zhou, J., Chen, Y., Tabibi, S., Alba, L., Garber, E. and Saiman, L. 2007. Antimicrobial susceptibility and synergy studies of Burkholderia cepacia complex isolated from patients with cystic fibrosis. *Antimicrob Agents Chemother* 51(3), tt. 1085-1088. doi: 10.1128/aac.00954-06

Zhou, J. *et al.* 2020. Characterization of Burkholderia cepacia Complex Core Genome and the Underlying Recombination and Positive Selection. *Front Genet* 11, t. 506. doi: 10.3389/fgene.2020.00506

Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J. and Wishart, D. S. 2011. PHAST: a fast phage search tool. *Nucleic Acids Res* 39(Web Server issue), tt. W347-352. doi: 10.1093/nar/gkr485

Zlosnik, J. E. *et al.* 2011. Mucoid and nonmucoid Burkholderia cepacia complex bacteria in cystic fibrosis infections. *Am J Respir Crit Care Med* 183(1), tt. 67-72. doi: 10.1164/rccm.201002-0203OC

Zlosnik, J. E., Hird, T. J., Fraenkel, M. C., Moreira, L. M., Henry, D. A. and Speert, D. P. 2008. Differential mucoid exopolysaccharide production by members of the Burkholderia cepacia complex. *J Clin Microbiol* 46(4), tt. 1470-1473. doi: 10.1128/jcm.02273-07

Zlosnik, J. E., Mori, P. Y., To, D., Leung, J., Hird, T. J. and Speert, D. P. 2014. Swimming motility in a longitudinal collection of clinical isolates of Burkholderia cepacia complex bacteria from people with cystic fibrosis. *PLoS One* 9(9), t. e106428. doi: 10.1371/journal.pone.0106428

Zlosnik, J. E. *et al.* 2015. Burkholderia species infections in patients with cystic fibrosis in British Columbia, Canada. 30 years' experience. *Ann Am Thorac Soc* 12(1), tt. 70-78. doi: 10.1513/AnnalsATS.201408-395OC

8 Appendices

Table 43 | Quality Metrics of the *B. multivorans* strains sequenced in this thesis (*n* = 73).

Strain	Lineage	Sequence size	Number of contigs	GC content (%)	Shortest contig size	Median sequence size	Mean sequence size	Longest contig size	N50 value	L50 value	Number of subsystems	Number of coding sequences	Number of RNAs	Number of rRNAs	number of tRNAs	number of tmRNAs
BCC6	1	6272301	107	67.3	119	1564	58619.6	923050	239280	9	511	5560	60	3	65	1
BCC9	1	6342332	83	67.2	130	2945	76413.6	1042576	226583	9	515	5618	66	4	73	1
BCC80	1	6706461	113	66.8	101	1294	59349.2	686101	289154	10	514	5999	64	3	72	1
BCC84	1	6599540	137	67.1	101	1857	48171.8	377138	165313	10	517	5962	58	3	67	1
BCC101	1	6384309	113	67.1	120	2162	56498.3	377138	219505	10	512	5636	56	3	66	1
BCC141	1	6171198	83	67.4	110	16803	74351.8	689872	216985	10	511	5498	62	3	70	1
BCC303	1	6252321	96	67.2	18	8940	63799.2	707653	192319	10	512	6098	61	3	69	1
BCC375	1	6289743	100	67.3	103	3795	62897.4	893830	253362	10	512	5561	61	3	71	1
BCC381	1	6248359	82	67.3	134	3188	76199.5	725497	229512	10	510	5531	56	3	65	1
BCC702	1	6455034	110	67.1	108	3289	58682.1	515667	241164	10	514	5723	62	3	69	1
BCC737	1	6223066	76	67.3	102	18360	81882.4	599007	239689	10	509	5472	61	3	67	1
BCC814	1	6273978	110	67.3	101	2755	57036.2	427585	238427	10	512	5591	56	3	66	1
BCC865	1	6791523	134	66.6	106	2148	50683	557856	262358	10	516	6110	61	3	71	1
BCC904	1	6029426	163	67.4	100	5437	36990.3	380474	106352	10	505	5338	51	3	60	1
BCC921	1	6786237	158	66.6	106	2274	42950.9	557856	197853	10	516	6098	59	3	69	1
BCC1177	1	6520943	160	67.1	101	2151	40755.9	470646	148928	10	516	5908	53	3	61	1
BCC1190	1	6366933	112	67.1	109	6177	56847.6	378394	176650	10	513	5694	58	3	68	1
BCC1385	1	6606699	112	66.9	101	1752	58988.4	544646	236600	10	513	5918	57	3	63	1
BCC1367	1	6246910	99	67.2	102	14809	94461.5	707677	207005	10	362	6228	61	3	69	1
BCC1384	1	6246946	99	67.2	108	7439	64458.8	707668	220614	10	362	5530	61	3	69	1
BCC0292	1	6360854	74	67.14	103	1675	57904.6	771986	207640	10	362	5682	58	3	66	1
BCC0293	1	6508922	143	66.99	102	1540	32956.4	537897	193063	10	362	5835	58	3	70	1
BCC0321	1	6244833	84	67.32	102	2870	58975.8	659695	210918	10	364	5579	66	3	74	1

BCC0704	1	6449118	85	67.15	108	6604	61471.5	463159	217445	10	367	5724	62	3	69	1
BCC0915	1	6277422	61	67.32	121	2755	72227.8	594460	243519	10	368	5598	65	3	75	1
BCC0093	1	6278118	64	67.23	312	40751	143203.1	739761	275222	10	366	5618	63	3	69	1
BCC0962	1	6277315	62	67.32	121	2972	71405.8	594459	243519	10	368	5598	65	3	75	1
BCC0968	1	6785185	113	66.58	107	2303	47500.3	557857	231462	10	372	6113	61	3	71	1
Lineage 1 average	N/A	6392715	104.75	67.12679	112.7143	6093.179	63060.1	613914.5	218944.3	9.928571	460.0357	5743.571	60.07143	3.035714	68.53571	1
BCC32	2a	6506850	70	67.1	129	1739	92955	597776	396722	7	515	5848	60	3	66	1
BCC47	2a	7115961	80	66.6	122	2624	88949.5	1060012	360758	6	523	6365	59	2	68	1
BCC66	2a	6510316	101	67.2	100	1564	64458.6	471548	225974	10	518	5726	67	4	74	1
BCC74	2a	6541253	52	67.3	117	28069	125793.3	741944	401179	10	513	5839	61	3	69	1
BCC188	2a	6652537	144	67	114	9369	46198.2	572214	164596	10	513	5939	63	2	71	1
BCC225	2a	6773647	59	67.1	127	1882	114807.6	1416856	6405560	10	515	6096	62	3	68	1
BCC264	2a	6629658	60	67.1	114	27268	110494.3	519753	365645	10	510	5860	62	3	71	1
BCC266	2a	6527276	92	67.3	115	12924	70948.7	447105	196973	10	514	5758	59	3	70	1
BCC317	2a	6355099	81	67.5	106	9257	78458	758055	235405	10	515	5606	61	3	69	1
BCC1272	2a	6882314	163	66.8	117	3934	42222.8	772166	178859	10	516	6229	60	3	69	1
BCC1271	2a	6997886	158	66.8	101	2320	32022.5	434582	151267	10	373	6361	53	3	62	1
BCC0134	2a	6414252	66	67.35	129	1574	66217.5	460055	308606	10	373	7246	53	3	69	1
BCC0583	2a	7040902	200	66.73	101	2597	24693.6	426567	131205	10	371	6405	51	3	60	1
BCC0729	2a	6316285	102	67.31	105	1312	45512.6	468355	188298	10	368	5589	60	3	69	1
BCC0907	2a	6658191	66	66.93	124	1949	67349.9	839825	265179	10	362	5889	61	3	70	1
Lineage 2a average	N/A	6661495	99.6	67.07467	114.7333	7225.467	71405.47	665787.5	665081.7	9.533333	466.6	6050.4	59.46667	2.933333	68.33333	1
BCC33	2b	6697341	82	67.1	101	1488	81674.9	873117	361065	7	518	5973	58	3	69	1
BCC43	2b	6544285	95	67.1	121	1132	68887.2	946841	383582	6	514	5795	62	3	73	1
BCC65	2b	6507164	64	67.4	101	1505	101674.4	731639	407234	6	518	5793	61	4	71	1
BCC68	2b	6440630	197	67.1	101	1519	32693.6	377625	154108	10	518	5773	50	4	60	1
BCC75	2b	6612635	78	67.2	103	11573	84777.4	938953	268072	10	515	5869	61	3	71	1
BCC79	2b	6991234	82	66.8	108	2155	85259	947587	337014	10	522	6229	61	3	71	1
BCC82	2b	6621312	305	67.2	101	954	21709.2	304664	92055	10	516	5913	43	3	56	1
BCC87	2b	6452471	62	67.4	105	4359	104072.1	859151	493463	10	521	5722	62	3	69	1
BCC96	2b	6375796	91	67.2	109	2853	70063.7	1058392	233122	10	514	5648	62	3	71	1

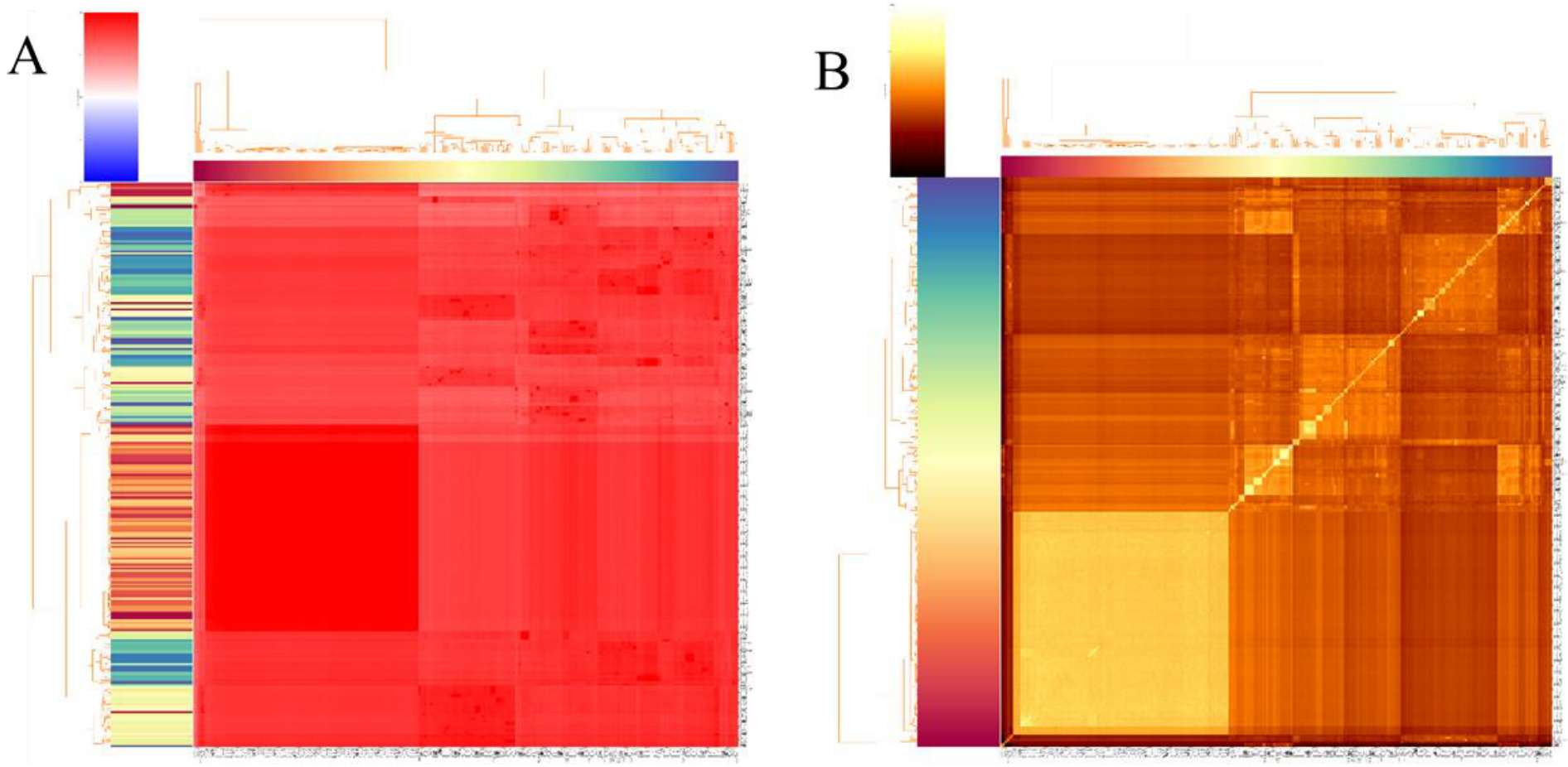
BCC241	2b	6655513	94	67.1	125	7260	70803.3	631291	200202	10	515	5898	57	3	72	1
BCC246	2b	6471839	65	67.3	122	16255	99566.8	678041	271451	10	516	5706	59	2	71	1
BCC247	2b	6531380	92	67.3	111	2137	70993.3	731592	286611	10	520	5773	60	3	70	1
BCC0269	2b	6281070	44	67.5	105	7560	114201.3	1145816	321515	6		5529		3	71	1
BCC384	2b	6378847	159	67.4	100	449	40118.5	485421	244802	10	517	5677	49	3	59	1
BCC493	2b	6530376	126	67.1	109	722	51828.4	450290	261403	10	512	5792	51	1	62	3
BCC497	2b	6789721	97	66.9	121	2101	69997.1	592166	259787	10	521	6040	60	3	71	1
BCC710	2b	6234591	89	67.5	100	887	70051.6	684558	310618	10	517	5504	56	3	66	1
BCC1147	2b	6576406	109	67.1	105	5295	60334	565382	173107	10	517	5842	61	3	70	1
BCC1148	2b	6573791	130	67.1	104	1652	50567.6	365455	171919	10	517	5850	59	3	68	1
BCC1185	2b	6383426	65	67.4	105	4145	98206.6	692368	368878	10	520	5658	61	3	68	1
BCC1421	2b	6609469	60	67.2	101	2320	32022.5	940166	352409	10	370	5871	62	3	72	1
BCC0181	2b	6537738	62	67.3	101	893	105554.1	835514	408687	10	374	5809	63	4	72	1
BCC0255	2b	6734567	87	67.1	108	1413	77497.6	510982	327645	10	368	6755	61	2	71	1
BCC0300	2b	6582551	44	67.24	108	1428	89061.5	729256	386587	10	375	5915	62	4	69	1
BCC0470	2b	6984344	60	66.8	108	1413	76838.7	947619	337014	10	371	6236	61	2	71	1
BCC0533	2b	6410978	47	67.35	124	1058	82311.4	1043364	421975	10	367	5678	62	3	71	1
BCC0553	2b	6694432	43	67.09	101	1625	101527.8	930655	405205	10	373	5975	60	3	71	1
BCC0585	2b	6447635	46	67.4	105	4145	105765.1	694113	337699	10	373	5728	62	3	69	1
BCC0901	2b	6497146	43	67.17	103	1752	106579.9	866533	529217	10	371	5830	63	2	75	1
Lineage 2b average	N/A	6556851	90.27586	67.20172	107.4483	3174.069	76711.68	743398.3	314015.4	9.482759	470.3571	5854.517	58.89286	2.931034	68.96552	1.068966
BCC1368	Other	6273724	109	67.2	102	2542	57557.1	796635	218602	10	509	5601	60	3	74	1
Overall average	N/A	6511519	98	67.14685	110.8904	5117.548	70122.75	678515.1	348379.7	9.671233	466.0972	5848.74	59.48611	2.972603	68.73973	1.027397

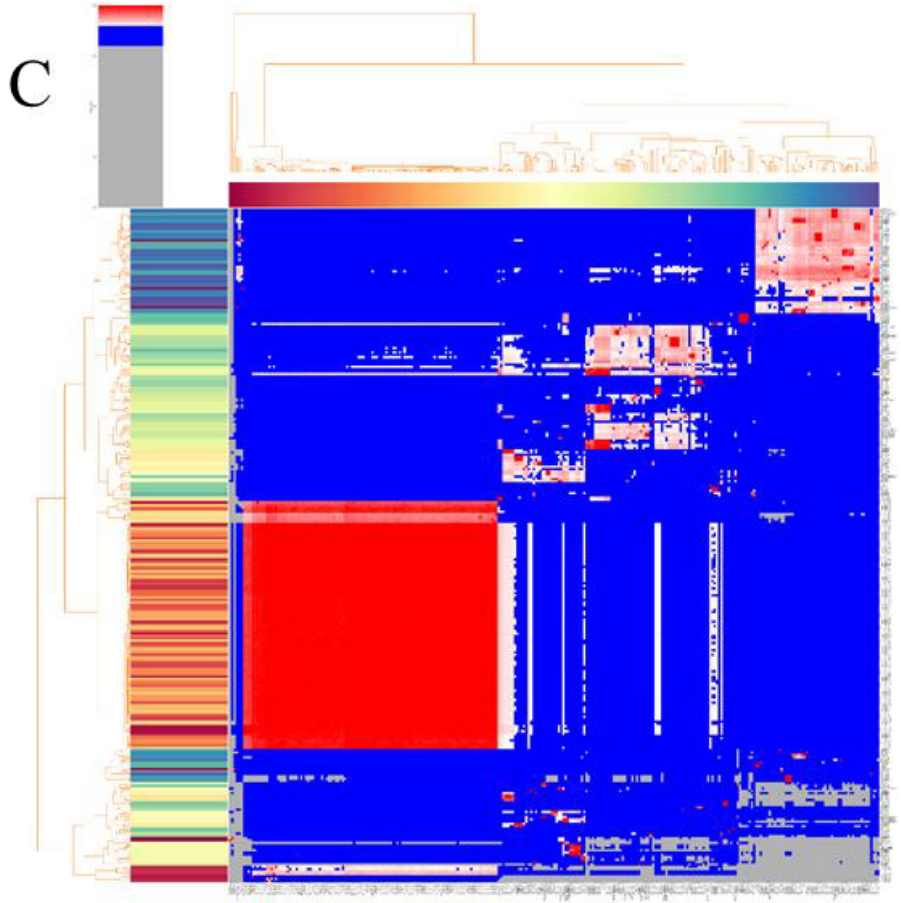
Table 44 |Statistical analysis (BAM files) of the mapped *B. multivorans* reads against the reference genomes

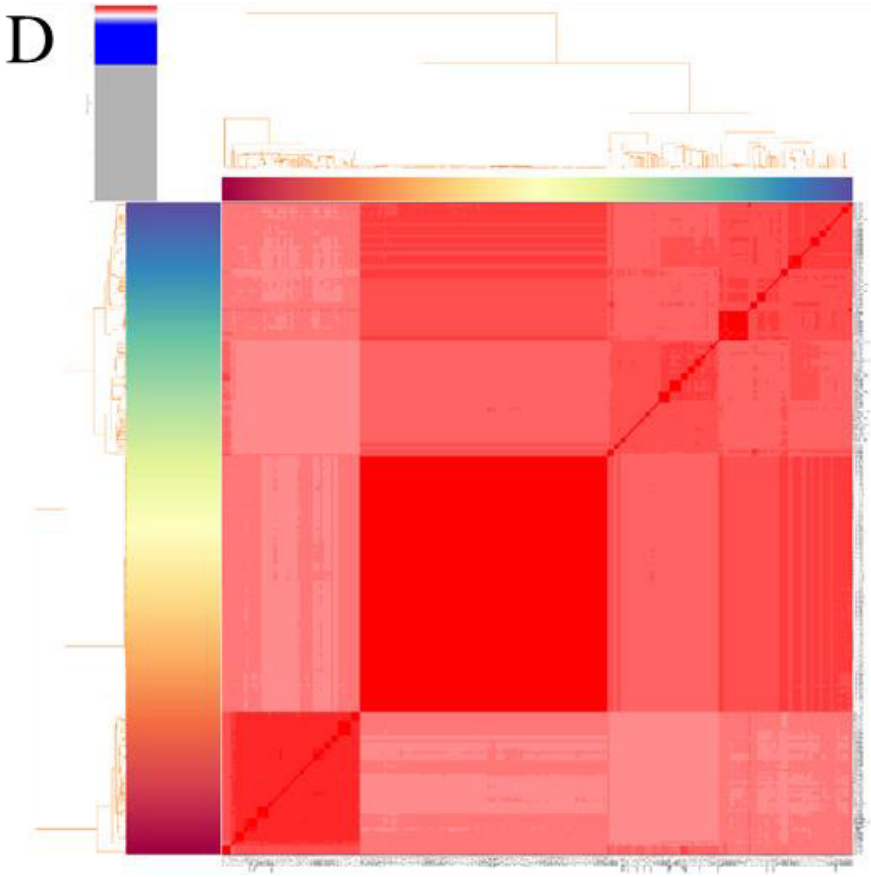
	Total Reads	Mapped Reads		Forward Strand		Reverse Strand		Paired end Reads		Proper pairs'		Both Pairs Mapped		Read 1	Read 2	Singletons	
Strain	Number	Number	%	Number	%	Number	%	Number	%	Number	%	Number	%	Number	Number	Number	%
BCC0006	6999937	5929577	84.71%	4055702	57.94%	2944235	42.06%	6999937	100.00%	5553108	79.33%	5771602	82.45%	3499962	3499975	157975	2.26%
BCC0009	7147747	5976506	83.61%	4179652	58.48%	2968095	41.52%	7147747	100.00%	5608002	78.46%	5814210	81.34%	3573970	3573777	162296	2.27%
BCC0032	7563566	6630209	87.66%	4270151	56.46%	3293415	43.54%	7563566	100.00%	6243046	82.54%	6492802	85.84%	3781832	3781734	137407	1.82%
BCC0033	6576276	5407673	82.23%	3890391	59.16%	2685885	40.84%	6576276	100.00%	5097854	77.52%	5294500	80.51%	3288121	3288155	113173	1.72%
BCC0043	7879679	6662612	84.55%	4570646	58.01%	3309033	41.99%	7879679	100.00%	6282480	79.73%	6527169	82.84%	3939863	3939816	135443	1.72%
BCC0047	7618920	6138217	80.57%	4569078	59.97%	3049842	40.03%	7618920	100.00%	5803956	76.18%	6023591	79.06%	3809496	3809424	114626	1.50%
BCC0065	6268462	5370549	85.68%	3598815	57.41%	2669647	42.59%	6268462	100.00%	5070698	80.89%	5259864	83.91%	3134193	3134269	110685	1.77%
BCC0066	4911001	4469990	91.02%	2688818	54.75%	2222183	45.25%	4911001	100.00%	4242292	86.38%	4396396	89.52%	2455474	2455527	73594	1.50%
BCC0068	6998383	6180131	88.31%	3928991	56.14%	3069392	43.86%	6998383	100.00%	5814752	83.09%	6047342	86.41%	3499240	3499143	132789	1.90%
BCC0074	4369650	3862646	88.40%	2446549	55.99%	1923101	44.01%	4369650	100.00%	3645328	83.42%	3770772	86.29%	2184875	2184775	91874	2.10%
BCC0075	3957376	3332338	84.21%	2297231	58.05%	1660145	41.95%	3957376	100.00%	3149088	79.58%	3254220	82.23%	1978688	1978688	78118	1.97%
BCC0079	5564716	4732597	85.05%	3210631	57.70%	2354085	42.30%	5564716	100.00%	4476652	80.45%	4630658	83.21%	2782404	2782312	101939	1.83%
BCC0080	4418318	3772729	85.39%	2540120	57.49%	1878198	42.51%	4418318	100.00%	3548648	80.32%	3664087	82.93%	2209149	2209169	108642	2.46%
BCC0082	5710214	4895729	85.74%	3273419	57.33%	2436795	42.67%	5710214	100.00%	4624594	80.99%	4784917	83.80%	2855120	2855094	110812	1.94%
BCC0084	6132940	5043622	82.24%	3626831	59.14%	2506109	40.86%	6132940	100.00%	4728028	77.09%	4899728	79.89%	3066484	3066456	143894	2.35%
BCC0087	5591820	4856586	86.85%	3177202	56.82%	2414618	43.18%	5591820	100.00%	4596400	82.20%	4760569	85.13%	2795945	2795875	96017	1.72%
BCC0093	9422526	7907101	83.92%	5469369	58.05%	3953157	41.95%	9422526	100.00%	7661134	81.31%	7754546	82.30%	4711200	4711326	152555	1.62%
BCC0096	4451467	3863475	86.79%	2528857	56.81%	1922610	43.19%	4451467	100.00%	3647408	81.94%	3777494	84.86%	2225723	2225744	85981	1.93%
BCC0101	5729484	4805944	83.88%	3340542	58.30%	2388942	41.70%	5729484	100.00%	4501760	78.57%	4659826	81.31%	2864741	2864743	146118	2.55%
BCC0134	10806449	9691156	89.68%	5961605	55.17%	4844844	44.83%	10806449	100.00%	9489466	87.81%	9598174	88.82%	5403273	5403176	92982	0.86%
BCC0141	6376726	5498661	86.23%	3643672	57.14%	2733054	42.86%	6376726	100	5164250	80.99%	5349688	83.89%	3188411	3188315	148973	2.34%
BCC0181	9590393	8270812	86.24%	5456045	56.89%	4134348	43.11%	9590393	100.00%	8078140	84.23%	8171720	85.21%	4795209	4795184	99092	1.03%
BCC0188	5253107	4540520	86.43%	2997699	57.07%	2255408	42.93%	5253107	100.00%	4276204	81.40%	4450602	84.72%	2626616	2626491	89918	1.71%
BCC0225	5668870	4919612	86.78%	3224987	56.89%	2443883	43.11%	5668870	100.00%	4638756	81.83%	4825123	85.12%	2834417	2834453	94489	1.67%
BCC0241	7602204	6661805	87.63%	4294379	56.49%	3307825	43.51%	7602204	100.00%	6254506	82.27%	6522908	85.80%	3801188	3801016	138897	1.83%
BCC0246	4801009	4159865	86.65%	2730752	56.88%	2070257	43.12%	4801009	100.00%	3935692	81.98%	4068173	84.74%	2400494	2400515	91692	1.91%
BCC0247	7123525	6042465	84.82%	4122019	57.86%	3001506	42.14%	7123525	100.00%	5690728	79.89%	5909793	82.96%	3561799	3561726	132672	1.86%
BCC0255	9645373	8532850	88.47%	5379651	55.77%	4265722	44.23%	9645373	100.00%	8319992	86.26%	8430664	87.41%	4822649	4822724	102186	1.06%

BCC0264	5415599	4922956	90.90%	2968719	54.82%	2446880	45.18%	5415599	100.00%	4658076	86.01%	4838523	89.34%	2707771	2707828	84433	1.56%
BCC0266	5829237	5147764	88.31%	3268615	56.07%	2560622	43.93%	5829237	100.00%	4875078	83.63%	5055254	86.72%	2914573	2914664	92510	1.59%
BCC0269	5213968	4560372	87.46%	2945907	56.50%	2268061	43.50%	5213968	100.00%	4306658	82.60%	4467159	85.68%	2606965	2607003	93213	1.79%
BCC0292	10628184	9140705	86.00%	6057795	57.00%	4570389	43.00%	10628184	100.00%	8790546	82.71%	8942567	84.14%	5313966	5314218	198138	1.86%
BCC0293	9446406	8062656	85.35%	5415677	57.33%	4030729	42.67%	9446406	100.00%	7797160	82.54%	7888471	83.51%	4723215	4723191	174185	1.84%
BCC0300	9579920	8168582	85.27%	5496727	57.38%	4083193	42.62%	9579920	100.00%	8000162	83.51%	8076869	84.31%	4789945	4789975	91713	0.96%
BCC0303	4263472	3531521	82.83%	2506893	58.80%	1756579	41.20%	4263472	100.00%	3274334	76.80%	3407888	79.93%	2131780	2131692	123633	2.90%
BCC0317	5634003	4980592	88.40%	3156292	56.02%	2477711	43.98%	5634003	100.00%	4718552	83.75%	4886139	86.73%	2816938	2817065	94453	1.68%
BCC0321	9035614	7839492	86.76%	5115596	56.62%	3920018	43.38%	9035614	100.00%	7614302	84.27%	7674798	84.94%	4517810	4517804	164694	1.82%
BCC0375	7094184	6023096	84.90%	4101745	57.82%	2992439	42.18%	7094184	100.00%	5652798	79.68%	5867330	82.71%	3547033	3547151	155766	2.20%
BCC0381	6528225	5558702	85.15%	3765798	57.68%	2762427	42.32%	6528225	100.00%	5215296	79.89%	5409863	82.87%	3264033	3264192	148839	2.28%
BCC0384	5725748	5009957	87.50%	3234701	56.49%	2491047	43.51%	5725748	100.00%	4736604	82.72%	4912387	85.79%	2862888	2862860	97570	1.70%
BCC0470	9831620	8421141	85.65%	5621660	57.18%	4209960	42.82%	9831620	100.00%	8234520	83.76%	8318647	84.61%	4915829	4915791	102494	1.04%
BCC0493	4483285	3811153	85.01%	2585957	57.68%	1897328	42.32%	4483285	100.00%	3598632	80.27%	3726277	83.11%	2241594	2241691	84876	1.89%
BCC0497	6064845	4954866	81.70%	3601014	59.38%	2463831	40.62%	6064845	100.00%	4667050	76.95%	4835953	79.74%	3032454	3032391	118913	1.96%
BCC0533	9285616	8130863	87.56%	5220685	56.22%	4064931	43.78%	9285616	100.00%	7944076	85.55%	8010079	86.26%	4642628	4642988	120784	1.30%
BCC0553	10611049	8969165	84.53%	6127010	57.74%	4484039	42.26%	10611049	100.00%	8799930	82.93%	8861146	83.51%	5305443	5305606	108019	1.02%
BCC0583	10734684	10197679	95.00%	5635570	52.50%	5099114	47.50%	10734684	100.00%	10028568	93.42%	10150630	94.56%	5367121	5367563	47049	0.44%
BCC0585	10118964	8824591	87.21%	5707083	56.40%	4411881	43.60%	10118964	100.00%	8625794	85.24%	8719404	86.17%	5059368	5059596	105187	1.04%
BCC0702	6202081	5185196	83.60%	3622343	58.41%	2579738	41.59%	6202081	100.00%	4870522	78.53%	5035796	81.20%	3101001	3101080	149400	2.41%
BCC0704	9126843	7734092	84.74%	5259946	57.63%	3866897	42.37%	9126843	100.00%	7492126	82.09%	7556169	82.79%	4563518	4563325	177923	1.95%
BCC0710	6530797	5788645	88.64%	3652730	55.93%	2878067	44.07%	6530797	100.00%	5468470	83.73%	5660819	86.68%	3265451	3265346	127826	1.96%
BCC0729	9388658	8168487	87.00%	5304503	56.50%	4084155	43.50%	9388658	100.00%	7950730	84.68%	8022682	85.45%	4694276	4694382	145805	1.55%
BCC0737	5340865	4522122	84.67%	3091540	57.88%	2249325	42.12%	5340865	100.00%	4245224	79.49%	4391162	82.22%	2670451	2670414	130960	2.45%
BCC0814	5148849	4358393	84.65%	2980471	57.89%	2168378	42.11%	5148849	100.00%	4095498	79.54%	4235735	82.27%	2574439	2574410	122658	2.38%
BCC0865	4413536	3558317	80.62%	2643276	59.89%	1770260	40.11%	4413536	100.00%	3317038	75.16%	3443633	78.02%	2206766	2206770	114684	2.60%
BCC0901	9583186	8350723	87.14%	5408301	56.44%	4174885	43.56%	9583186	100.00%	8160678	85.16%	8228892	85.87%	4791614	4791572	121831	1.27%
BCC0904	3220714	2773024	86.10%	1840095	57.13%	1380619	42.87%	3220714	100.00%	2596548	80.62%	2693393	83.63%	1610393	1610321	79631	2.47%
BCC0907	9922921	9064490	91.35%	5391228	54.33%	4531693	45.67%	9922921	100.00%	8874960	89.44%	8994358	90.64%	4961253	4961668	70132	0.71%
BCC0915	10856895	9410339	86.68%	6151736	56.66%	4705159	43.34%	10856895	100.00%	9131796	84.11%	9241055	85.12%	5428426	5428469	169284	1.56%
BCC0921	4670273	3743567	80.16%	2810274	60.17%	1859999	39.83%	4670273	100	3490240	74.73%	3621959	77.55%	2335189	2335084	121608	2.60%
BCC0962	10228209	8669042	84.76%	5894506	57.63%	4333703	42.37%	10228209	100.00%	8377116	81.90%	8493601	83.04%	5114116	5114093	175441	1.72%

BCC0968	10005675	8254510	82.50%	5877846	58.75%	4127829	41.25%	10005675	100.00%	7987638	79.83%	8075016	80.70%	5002838	5002837	179494	1.79%
BCC1147	3840978	3424407	89.15%	2135282	55.59%	1705696	44.41%	3840978	100.00%	3239766	84.35%	3350379	87.23%	1920482	1920496	74028	1.93%
BCC1148	4810701	4289660	89.17%	2674303	55.59%	2136398	44.41%	4810701	100.00%	4044398	84.07%	4191688	87.13%	2405395	2405306	97972	2.04%
BCC1177	5848961	4814820	82.32%	3454901	59.07%	2394060	40.93%	5848961	100.00%	4525300	77.37%	4682482	80.06%	2924484	2924477	132338	2.26%
BCC1185	4149985	3611985	87.04%	2352667	56.69%	1797318	43.31%	4149985	100.00%	3420804	82.43%	3533268	85.14%	2074966	2075019	78717	1.90%
BCC1190	8957376	7442738	83.09%	5267346	58.80%	3690030	41.20%	8957376	100.00%	6932876	77.40%	7236077	80.78%	4478701	4478675	206661	2.31%
BCC1271	9041417	8729361	96.55%	4676696	51.73%	4364721	48.27%	9041417	100.00%	8600610	95.12%	8693476	96.15%	4520566	4520851	35885	0.40%
BCC1272	8237658	8116573	98.53%	4207756	51.08%	4029902	48.92%	8237658	100.00%	7660340	92.99%	8017771	97.33%	4118815	4118843	98802	1.20%
BCC1367	9593198	8359824	87.14%	5413773	56.43%	4179425	43.57%	9593198	100.00%	8089650	84.33%	8187390	85.35%	4796474	4796724	172434	1.80%
BCC1368	5173199	4486858	86.73%	2941301	56.86%	2231898	43.14%	5173199	100.00%	4226060	81.69%	4377469	84.62%	2586646	2586553	109389	2.11%
BCC1384	10890007	9503000	87.26%	6137986	56.36%	4752021	43.64%	10890007	100.00%	9217038	84.64%	9324193	85.62%	5444962	5445045	178807	1.64%
BCC1385	3901993	3170433	81.25%	2322838	59.53%	1579155	40.47%	3901993	100.00%	2958532	75.82%	3071824	78.72%	1950974	1951019	98609	2.53%
BCC1421	8909407	7554470	84.79%	5132220	57.60%	3777187	42.40%	8909407	100.00%	7363762	82.65%	7447717	83.59%	4454600	4454807	106753	1.20%
Average	7091769.08	6130095	86.23%	4036344	0.570458	3055425	0.429542	7091769	3.7123288	5863271	82.12%	6010528	84.44%	3545872	3545896.671	119566.8	1.79%


A
B


C


D

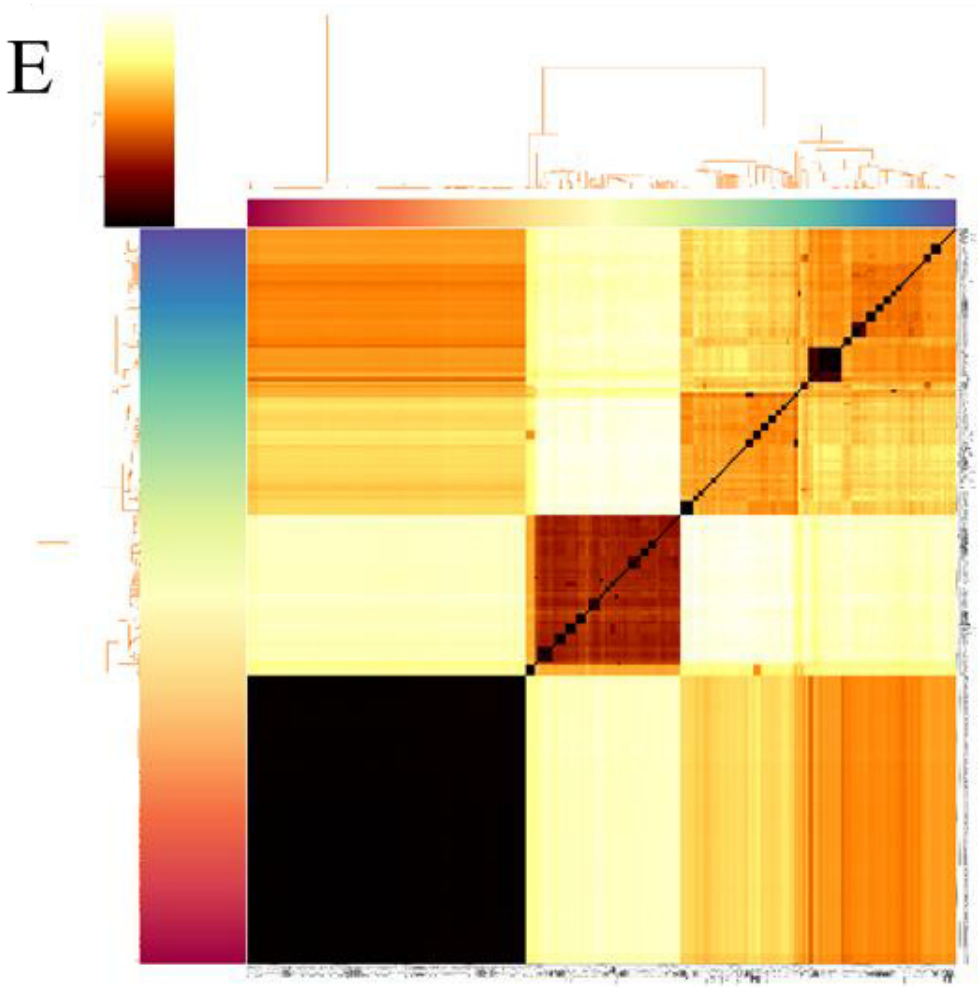


Figure 65 | PyANI outputs of the *B. multivorans* master strain panel (*n* = 283 genomes). (A) Alignment coverage, (B) Total alignment lengths, (C) Hadamard matrix, (D) Percentage (%) ID, (E) Similarity Errors (mismatches but not including indels).

Table 45 | FastANI comparisons of the *B. multivorans* draft genomes (*n* = 283) against reference strain ATCC 17616.

Draft Genome	ANI (%)	Orthologous Matches	Total Sequence Fragments	Percentage of Ortholog Matches (%)
ATCC 17616	100	2332	2334	99.91
701 BMUL	97.75	1572	1738	90.45
800 BMUL	97.72	1763	1978	89.13
ATCC BAA-247	96.82	1612	1786	90.26
AU10047	96.81	1864	2059	90.53
AU10086	97.84	1844	2043	90.26
AU10398	98.35	1953	2183	89.46
AU10897	97.90	1888	2133	88.51
AU11204	96.83	1837	2055	89.39
AU11233	98.12	1892	2101	90.05
AU11358	96.91	1861	2096	88.79
AU11772	97.56	1857	2037	91.16
AU1185	97.02	1897	2206	85.99
AU12481	98.14	1902	2043	93.10
AU13919	97.88	1938	2222	87.219
AU14328	97.88	1931	2222	86.90
AU14364	98.16	1869	2081	89.81
AU14371	98.07	1989	2224	89.43
AU14786	97.56	1859	2033	91.44
AU15814	98.15	1880	2083	90.25
AU15954	97.94	1987	2245	88.51
AU16734	96.95	1818	1979	91.86
AU17135	97.02	1852	2010	92.14
AU17534	98.34	1979	2131	92.87
AU17545	98.31	2019	2251	89.69
AU18096	98.30	1892	2112	89.58
AU19518	98.24	1993	2174	91.67
AU19564	98.13	2016	2248	89.68
AU19659	96.91	1832	2008	91.24
AU19729	98.39	2050	2295	89.32
AU20929	98.09	2009	2222	90.41
AU21015	97.93	1820	2030	89.66
AU21596	97.76	1855	2028	91.47
AU21747	100.00	2224	2228	99.82
AU22436	97.03	1825	2030	89.90
AU22892	97.94	1996	2234	89.35
AU23365	97.01	1840	2060	89.32
AU23668	98.17	1903	2063	92.24
AU23690	97.10	1734	1915	90.55
AU23919	97.51	1843	2013	91.55
AU23995	97.92	1878	2100	89.43
AU24277	97.84	1867	2129	87.69
AU25057	96.94	1859	2041	91.08
AU25543	98.35	1972	2188	90.13
AU26250	98.00	1971	2188	90.08
AU27706	96.98	1835	2136	85.91

AU28069	98.47	1988	2127	93.46
AU28442	96.90	1847	2013	91.75
AU29198	97.83	1880	2066	91.00
AU30050	96.82	1848	2070	89.28
AU30438	96.92	1850	2025	91.36
AU30441	96.89	1828	2088	87.54
AU30760	98.08	1957	2162	90.518
AU4507	97.77	1909	2157	88.50
BCC0005	97.83	1911	2126	89.89
BCC0006	96.92	1864	2052	90.84
BCC0008	97.85	1928	2196	87.80
BCC0009	97.04	1854	2082	89.05
BCC0010	96.97	1867	2076	89.93
BCC0031	97.82	1990	2165	91.92
BCC0032	97.82	1976	2145	92.12
BCC0033	97.83	1910	2206	86.58
BCC0037	98.08	1915	2165	88.45
BCC0043	97.90	1887	2151	87.73
BCC0047	97.87	1990	2341	85.01
BCC0050	98.10	2032	2308	88.04
BCC0059	97.73	1945	2235	87.02
BCC0065	97.91	1906	2148	88.73
BCC0066	98.33	1994	2138	93.26
BCC0067	96.98	1852	2022	91.59
BCC0068	97.91	1926	2080	92.60
BCC0074	98.20	1958	2159	90.69
BCC0075	97.83	1917	2171	88.30
BCC0079	98.04	2008	2299	87.34
BCC0080	97.33	1965	2197	89.44
BCC0082	97.89	1855	2109	87.96
BCC0084	97.08	1903	2152	88.43
BCC0087	97.87	1911	2127	89.84
BCC0089	97.87	1949	2118	92.02
BCC0093	96.91	1876	2080	90.19
BCC0096	97.78	1897	2092	90.68
BCC0099	98.11	2018	2282	88.43
BCC0101	96.99	1856	2086	88.97
BCC0102	98.04	2000	2203	90.79
BCC0115	97.02	1881	2057	91.44
BCC0134	98.10	1920	2106	91.17
BCC0141	96.98	1851	2024	91.45
BCC0149	96.96	1882	2097	89.75
BCC0175	96.89	1879	2215	84.83
BCC0181	97.78	1920	2164	88.72
BCC0188	98.33	1964	2157	91.05
BCC0225	98.11	1989	2235	88.99
BCC0241	97.98	1985	2183	90.93
BCC0244	98.43	2039	2176	93.70
BCC0246	97.82	1910	2130	89.67
BCC0247	97.89	1910	2144	89.09

BCC0255	98.08	2003	2218	90.31
BCC0264	98.44	2033	2188	92.92
BCC0266	98.16	1937	2140	90.51
BCC0269	97.90	1872	2073	90.30
BCC0292	96.94	1844	2087	88.36
BCC0293	97.02	1885	2107	89.46
BCC0300	97.80	1897	2173	87.30
BCC0303	96.95	1846	2047	90.18
BCC0317	98.15	1883	2084	90.36
BCC0321	96.99	1856	2044	90.80
BCC0375	96.97	1873	2062	90.83
BCC0381	97.02	1844	2052	89.86
BCC0384	97.82	1880	2086	90.12
BCC0470	98.05	2004	2299	87.17
BCC0493	97.90	1879	2141	87.76
BCC0497	97.76	1912	2225	85.93
BCC0533	97.84	1902	2116	89.89
BCC0553	97.82	1920	2212	86.80
BCC0583	99.94	2154	2253	95.61
BCC0585	97.92	1904	2127	89.52
BCC0702	96.81	1892	2112	89.58
BCC0704	96.89	1878	2114	88.84
BCC0710	97.82	1889	2052	92.057
BCC0729	97.13	1868	2059	90.72
BCC0737	97.00	1837	2045	89.83
BCC0814	97.05	1844	2051	89.91
BCC0865	96.92	1908	2211	86.30
BCC0901	97.88	1926	2144	89.83
BCC0904	97.09	1764	1944	90.74
BCC0907	98.49	2010	2187	91.91
BCC0915	97.03	1858	2066	89.93
BCC0921	96.90	1899	2203	86.20
BCC0962	97.03	1861	2065	90.12
BCC0968	96.91	1905	2208	86.28
BCC1147	97.97	1982	2155	91.97
BCC1148	97.97	1966	2148	91.53
BCC1177	97.09	1874	2114	88.65
BCC1185	97.91	1887	2104	89.69
BCC1190	97.00	1855	2080	89.18
BCC1271	99.96	2188	2263	96.69
BCC1272	100.00	2230	2233	99.87
BCC1367	96.97	1839	2046	89.88
BCC1368	97.56	1862	2047	90.96
BCC1384	96.96	1844	2047	90.08
BCC1385	96.90	1870	2164	86.41
BCC1421	97.84	1910	2172	87.94
CF170.0a	97.91	1895	2094	90.50
CF170.10a	97.88	1929	2115	91.21
CF170.10b	97.90	1909	2101	90.86
CF170.10c	97.90	1910	2100	90.95

CF170.10d	97.92	1888	2077	90.90
CF170.10e	97.92	1900	2097	90.61
CF170.10f	97.912	1919	2108	91.03
CF170.10g	97.92	1864	2046	91.10
CF170.10h	97.95	1866	2051	90.98
CF170.10i	97.92	1866	2043	91.34
CF170.10j	97.93	1881	2057	91.44
CF170.11a	97.89	1922	2115	90.87
CF170.11b	97.88	1910	2104	90.78
CF170.11c	97.91	1909	2101	90.86
CF170.11d	97.91	1912	2098	91.13
CF170.11e	97.90	1915	2110	90.76
CF170.11f	97.86	1911	2093	91.30
CF170.11g	97.87	1898	2085	91.03
CF170.11h	97.89	1882	2068	91.01
CF170.11i	97.87	1905	2099	90.76
CF170.11j	97.90	1915	2107	90.89
CF170.1a	97.85	1914	2096	91.32
CF170.1b	97.89	1907	2094	91.07
CF170.1c	97.86	1914	2101	91.10
CF170.1d	97.94	1914	2103	91.01
CF170.1e	97.90	1894	2085	90.84
CF170.1f	97.90	1912	2103	90.92
CF170.1g	97.89	1888	2077	90.90
CF170.1h	97.90	1897	2089	90.809
CF170.1i	97.88	1915	2099	91.23
CF170.1j	97.90	1914	2106	90.88
CF170.2a	97.87	1915	2106	90.93
CF170.2b	97.91	1894	2095	90.41
CF170.2c	97.90	1912	2110	90.61
CF170.2d	97.90	1887	2085	90.50
CF170.2e	97.88	1892	2076	91.14
CF170.2f	97.92	1898	2092	90.73
CF170.2g	97.91	1888	2088	90.42
CF170.2h	97.88	1877	2069	90.72
CF170.2i	97.86	1906	2095	90.98
CF170.2j	97.90	1911	2102	90.91
CF170.3a	97.89	1888	2081	90.73
CF170.3b	97.89	1943	2136	90.96
CF170.3c	97.87	1915	2106	90.93
CF170.3d	97.91	1902	2097	90.70
CF170.3e	97.87	1900	2092	90.82
CF170.3f	97.88	1891	2080	90.91
CF170.3g	97.91	1899	2091	90.82
CF170.3h	97.89	1902	2094	90.83
CF170.3i	97.89	1914	2104	90.97
CF170.3j	97.89	1891	2082	90.83
CF170.4a	97.94	1876	2061	91.02
CF170.4b	97.88	1908	2098	90.94
CF170.4c	97.90	1913	2104	90.92

CF170.4d	97.88	1913	2100	91.10
CF170.4e	97.91	1891	2087	90.61
CF170.4f	97.89	1912	2106	90.79
CF170.4g	97.88	1917	2109	90.90
CF170.4h	97.87	1910	2097	91.08
CF170.4i	97.89	1908	2101	90.81
CF170.4j	97.88	1900	2093	90.78
CF170.5a	97.93	1902	2101	90.53
CF170.5b	97.88	1902	2094	90.83
CF170.5c	97.89	1907	2097	90.94
CF170.5d	97.89	1921	2113	90.91
CF170.5e	97.92	1915	2105	90.97
CF170.5f	97.88	1901	2093	90.83
CF170.5g	97.89	1912	2101	91.00
CF170.5h	97.91	1903	2094	90.88
CF170.5i	97.89	1905	2097	90.84
CF170.5j	97.89	1871	2065	90.61
CF170.6a	97.90	1904	2098	90.75
CF170.6b	97.87	1945	2137	91.02
CF170.6c	97.90	1885	2079	90.67
CF170.6d	97.90	1911	2104	90.83
CF170.6e	97.92	1904	2100	90.67
CF170.6f	97.88	1916	2105	91.02
CF170.6g	97.89	1924	2110	91.18
CF170.6h	97.91	1913	2102	91.01
CF170.6i	97.91	1918	2109	90.94
CF170.6j	97.91	1917	2103	91.16
CF170.7a	97.89	1903	2098	90.71
CF170.7b	97.89	1925	2117	90.93
CF170.7c	97.91	1912	2102	90.96
CF170.7d	97.90	1924	2113	91.056
CF170.7e	97.93	1900	2102	90.39
CF170.7f	97.91	1923	2113	91.01
CF170.7g	97.88	1912	2101	91.00
CF170.7h	97.87	1896	2084	90.98
CF170.7i	97.90	1925	2113	91.10
CF170.7j	97.93	1910	2098	91.04
CF170.8a	97.92	1911	2111	90.53
CF170.8b	97.90	1888	2082	90.68
CF170.8c	97.89	1922	2110	91.09
CF170.8d	97.89	1924	2111	91.14
CF170.8e	97.86	1923	2104	91.40
CF170.8f	97.88	1905	2088	91.24
CF170.8g	97.87	1911	2102	90.91
CF170.8h	97.92	1904	2089	91.14
CF170.8i	97.88	1906	2092	91.11
CF170.8j	97.88	1931	2117	91.21
CF170.9a	97.90	1898	2087	90.94
CF170.9b	97.90	1912	2101	91.00
CF170.9c	97.89	1912	2100	91.05

CF170.9d	97.88	1899	2089	90.90
CF170.9e	97.92	1898	2096	90.55
CF170.9f	97.87	1919	2104	91.21
CF170.9g	97.88	1918	2110	90.90
CF170.9h	97.89	1909	2097	91.03
CF170.9i	97.88	1917	2105	91.07
CF170.9j	97.88	1928	2116	91.12
CF2	98.05	1759	1935	90.90
CGD1	97.65	1909	2191	87.13
CGD2	96.91	1880	2171	86.60
CGD2M	96.93	1880	2173	86.52
D2095	97.84	1931	2215	87.18
D2214	97.86	1913	2142	89.31
DDS 15A-1	98.49	2149	2426	88.58
DSOPR54	97.84	782	939	83.28
DSOPR57	97.82	894	1052	84.98
DWS 42B-1	98.15	1942	2167	89.62
FDAARGOS 246	96.98	1886	2106	89.55
FDAARGOS 546	96.93	1836	2062	89.04
FDAARGOS 547	98.20	1947	2130	91.41
FDAARGOS 548	98.35	1999	2216	90.21
HI3534	96.91	1803	2008	89.79
MSMB1128WGS	97.93	1883	2080	90.53
MSMB1272WGS	98.36	2015	2238	90.04
MSMB1535WGS	97.96	1889	2081	90.77
MSMB1640WGS	98.41	2094	2280	91.84
MSMB1641WGS	98.39	2052	2213	92.72
MSMB1794WGS	97.94	1886	2093	90.11
MSMB1916WGS	98.24	1952	2110	92.51
MSMB2008WGS	98.11	1892	2112	89.58
MSMB2021WGS	98.17	1906	2101	90.72
MSMB575WGS	98.15	1890	2101	89.96
MSMB576WGS	98.17	1895	2102	90.15
MSMB612WGS	98.18	1938	2171	89.27
NCTC13007	96.98	1885	2099	89.80
NKI379	98.12	1818	2023	89.87
R-20526	96.95	1833	2054	89.24

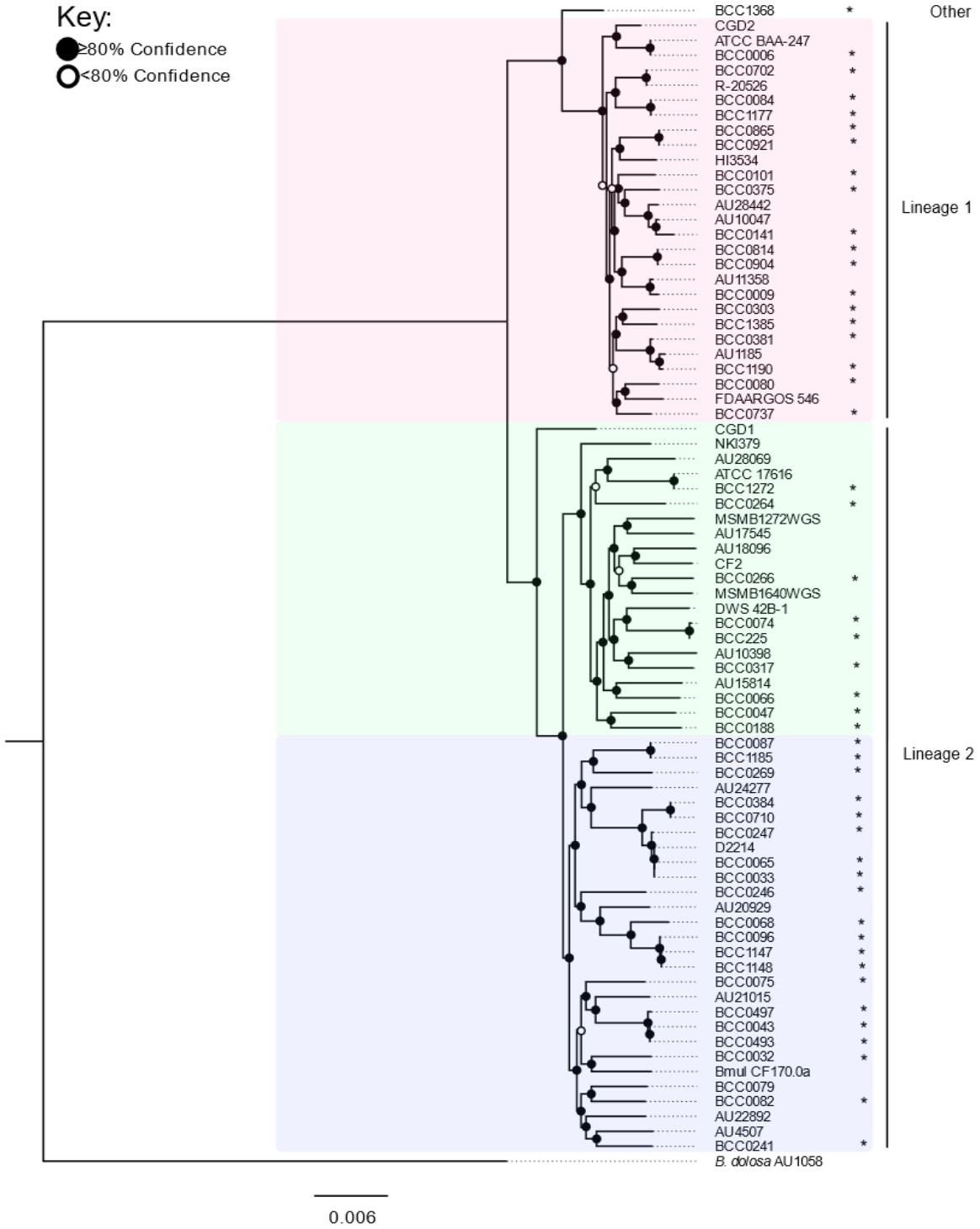


Figure 66 | *B. dolosa* AU1058 rooted phylogenomic analysis of the *B. multivorans* strain panel (*n* = 77). RAxML tree built using the alignment of 3,251 core-genes and 100 bootstraps. Strain genomic relatedness is indicated by the scale bar (0.008). Nodes have been allocated either a filled or hollow circle to represent the bootstrap confidence levels. Filled circles represent a bootstrap of ≥80% and a hollow (white) circle means confidence of ≤80%. Lineages have been marked on the right-hand side and coloured behind the branches. This shows Lineages 1 and 2 as well as 'other'.

Table 46 | Clonal complexes by group for the *B. multivorans* strains (*n* = 566). Output created using PubMLST (Jolley and Maiden 2010; Jolley *et al.* 2018).

ST	Frequency	SLV	DLV	SAT
Group 1				
16	11	1		
15	2	1		
Group 2				
564	1	1	1	
1023	3		1	1
645	1	1		1
Group 3				
1794	1	1		
439	1	1		
Group 4				
646	1	1		
1530	1	1		
Group 5				
666	1		1	2
1786	2		1	2
1785	1		1	2
25	3		3	
Group 6				
1790	1		1	
199	3		1	
Group 7				
19	2	1		
1762	1	1		
Group 8				
806	2	1		
26	1	1		
Group 9				
195	4		1	
836	3		1	
Group 10				
1088	3	1		
1325	1	1		
Group 11				
783	107	1		
1810	1	1		

Table 47 | Consistent SNP variants observed in *B. multivorans* strain ATCC 17616 WT after 3 and 5-days post infection in the murine models

Infection area	Polymorphism type	Strand	Position	Contig	Reference	Alternative	Effect	Gene	Product
Nasopharynx	Complex	-	561	109	GTGT	CTGC	Unknown	Unknown	Hypothetical Protein
	Deletion	+	91152	20	AGACGACGACG	AGACGACG	disruptive inframe deletion	minD	Septum site-determining protein MinD
	Insertion	-	199041	4	CGGCGCGGGCGCGG GCGCGGGCGCCGCA	CGGCGCGGGCGCGGGCGC GGGCGCGGGCGCCGCA	conservative inframe insertion	Unknown	Hypothetical Protein
	Insertion	-	675	109	AGGAA	ACGGAA	Unknown	Unknown	Hypothetical Protein
	SNP	+	318325	4	A	C	stop_lost&splice region variant	Unknown	Hypothetical Protein
	SNP	-	652	109	A	G	Unknown	Unknown	Hypothetical Protein
	SNP	-	6709	14	C	A	Synonymous mutation	Unknown	Hypothetical Protein
	SNP	Unknown	7931	14	G	C	Unknown	Unknown	Unknown
	SNP	Unknown	582	112	G	A	Unknown	Unknown	Unknown
	SNP	Unknown	639	112	G	A	Unknown	Unknown	Unknown
	SNP	Unknown	10788	17	T	A	Unknown	Unknown	Unknown
	SNP	-	741	109	T	C	Unknown	Unknown	Hypothetical Protein
	SNP	Unknown	498	112	T	C	Unknown	Unknown	Unknown
Lung	Deletion	+	91152	20	AGACGACGACG	AGACGACG	disruptive inframe deletion	minD	Septum site-determining protein MinD
	Insertion	-	199041	4	CGGCGCGGGCGCGG GCGCGGGCGCCGCA	CGGCGCGGGCGCGGGCGC GGGCGCGGGCGCCGCA	conservative inframe insertion	Unknown	Hypothetical Protein
	SNP	+	318325	4	A	C	stop_lost&splice region variant	Unknown	Hypothetical Protein

SNP	Unknown	1026	95	A	G	Unknown	Unknown	Unknown
SNP	-	6709	14	C	A	Synonymous mutation	Unknown	Hypothetical Protein
SNP	+	234	109	C	T	Unknown	potH_3	Putrescine transport system permease protein PotH
SNP	Unknown	243	114	C	T	Unknown	Unknown	Unknown
SNP	Unknown	292	129	C	T	Unknown	Unknown	Unknown
SNP	Unknown	7931	14	G	C	Unknown	Unknown	Unknown
SNP	Unknown	135	95	G	A	Unknown	Unknown	Unknown
SNP	Unknown	201	95	G	T	Unknown	Unknown	Unknown
SNP	Unknown	1188	95	G	A	Unknown	Unknown	Unknown
SNP	+	309	109	G	C	Unknown	potH_3	Putrescine transport system permease protein PotH
SNP	Unknown	303	112	G	A	Unknown	Unknown	Unknown
SNP	Unknown	10788	17	T	A	Unknown	Unknown	Unknown
SNP	Unknown	254	129	T	G	Unknown	Unknown	Unknown

Green colured SNP variants indicate that the adaptive mutations are found in both the nasopharynx and lungs. SNP variants without a coloured background are unique to that infection area.

Table 48 | Consistent SNP variants observed in *B. multivorans* strain BCC0084::pIN301-eGFP after 3 and 5-days post infection in the murine models

Infection area	Polymorphism type	Strand	Position	Contig	Reference	Alternative	Effect	Gene	Product
					TGGGGGGGGGGGG GGGGGGGGGGCCC	TGGGGGGGG GGGCGGGGG			
	Complex	Unknown	228888	5	CCCC	CGGCCCC	Unknown	Unknown	Unknown
	Complex	Unknown	801	80	TTTG	CTTC	Unknown	Unknown	Unknown
	Complex	Unknown	855	80	AATT	GATC	Unknown	Unknown	Unknown
	SNP	-	148	72	T	C	ydcV_5	Synonymous mutation	Inner membrane ABC transporter permease protein YdcV
Nasopharynx	SNP	-	1248	72	A	G	potH_3	Putrescine transport system permease protein PotH	
	SNP	Unknown	729	80	C	T	Unknown	Unknown	Unknown
	SNP	Unknown	792	80	G	C	Unknown	Unknown	Unknown
	SNP	Unknown	849	80	G	C	Unknown	Unknown	Unknown
	SNP	Unknown	873	80	T	C	Unknown	Unknown	Unknown
	SNP	Unknown	888	80	C	G	Unknown	Unknown	Unknown
	SNP	Unknown	348	92	C	T	Unknown	Unknown	Unknown
	SNP	Unknown	132	107	G	A	Unknown	Unknown	Unknown
	Complex	Unknown	801	80	TTTG	CTTC	Unknown	Unknown	Unknown
	Complex	Unknown	855	80	AATT	GATC	Unknown	Unknown	Unknown
	SNP	Unknown	729	80	C	T	Unknown	Unknown	Unknown
Lung	SNP	Unknown	792	80	G	C	Unknown	Unknown	Unknown
	SNP	Unknown	849	80	G	C	Unknown	Unknown	Unknown
	SNP	Unknown	873	80	T	C	Unknown	Unknown	Unknown
	SNP	Unknown	888	80	C	G	Unknown	Unknown	Unknown

Green colured SNP variants indicate that the adaptive mutations are found in both the nasopharynx and lungs. SNP variants without a coloured background are unique to that infection area.

Table 49 | Consistent SNP variants observed in *B. multivorans* strain BCC0033::pIN301-eGFP after 3 and 5-days post infection in the murine models

Infection area	Polymorphism type	Strand	Position	Contig	Reference	Alternative	Effect	Gene	Product
Nasopharynx	Complex	Unknown	450	45	GTGC	CTGT	Unknown	Unknown	Unknown
	SNP	-	541	45	G	A	Missense mutation	Unknown	Hypothetical protein
	SNP	-	687	45	C	T	Synonymous mutation	Unknown	Hypothetical protein
	SNP	-	741	45	G	C	Synonymous mutation	Unknown	Hypothetical protein
	SNP	Unknown	345	51	C	T	Unknown	Unknown	Unknown
Lung	SNP	Unknown	174	49	G	C	Unknown	Unknown	Unknown

Table 50 | Consistent SNP variants observed in *B. multivorans* strain BCC0033 WT after 3 and 5-days post infection in the murine models

Infection area	Polymorphism type	Strand	Position	Contig	Reference	Alternative	Effect	Gene	Product
Nasopharynx	Complex	Unknown	41	7	CCGG	ACGT	Unknown	Unknown	Unknown
	SNP	+	27307	3	G	C	Missense mutation	Unknown	Hypothetical protein
	SNP	+	29778	5	C	G	Synonymous mutation	Unknown	Hypothetical protein
	SNP	Unknown	2144	33	G	A	Unknown	Unknown	Unknown
Lung	SNP	+	29778	5	C	G	Synonymous mutation	Unknown	Hypothetical protein
	SNP	Unknown	150	49	A	G	Unknown	Unknown	Unknown

Green colured SNP variants indicate that the adaptive mutations are found in both the nasopharynx and lungs. SNP variants without a coloured background are unique to that infection area.

Table 51 | Swimming and swarming motility zones of the *B. multivorans* strains (n = 50).

Strain	Swimming	Swarming	
	0.3% LB	0.5% LB	0.5% BSM-G
ATCC 17616	21.8	7.8	22.2
BCC0006	10.3	0	0
BCC0009	57.8	14.4	90
BCC0032	18.8	2.6	3.8
BCC0033	25.2	12	9.8
BCC0043	15.3	2.8	0
BCC0047	25.3	10	30.1
BCC0065	24.1	11.8	8.8
BCC0066	22.3	10.4	ND
BCC0068	0	0	0
BCC0074	20.1	3.8	3.1
BCC0075	41.2	17.3	77.3
BCC0079	29.4	18.8	ND
BCC0080	26.8	3.5	0
BCC0082	31.8	9.9	ND
BCC0084	58.6	24.3	36.3
BCC0087	27.3	17.4	24.6
BCC0096	38.8	12.3	18.1
BCC0101	23.7	5.3	8.8
BCC0141	13.9	8.3	13.6
BCC0188	ND	ND	ND
BCC0225	30.1	11.7	ND
BCC0241	29.5	16.4	18
BCC0246	18.7	5.4	5.8
BCC0247	13.8	5.3	6.5
BCC0264	21.3	4.7	3.5
BCC0266	25.1	10	11.1
BCC0269	8.5	5.4	4
BCC0302	46.1	14.3	ND
BCC0303	2.3	2.7	14.8
BCC0317	90	19.5	87.5
BCC0375	19.8	11.8	9
BCC0384	17.3	5.1	10.8
BCC0493	6.8	4.5	0
BCC0497	5.3	2.8	16.2
BCC0702	47.1	32.6	69.1
BCC0710	22.3	8	6.3
BCC0737	9.9	3.8	11.8
BCC0814	24.4	14.3	16.8
BCC0865	7.8	5.7	4.3
BCC0904	3.2	2	0
BCC0921	10.7	7.5	3.3
BCC1147	24.8	8.3	26.6
BCC1148	14.7	5.3	14.3
BCC1177	60.9	24.6	90
BCC1185	17.3	9.5	1.2

BCC1190	59.6	17.8	90
BCC1272	25.8	7.3	70.7
BCC1368	7.7	ND	ND
BCC1385	25.6	14.3	40.4

Blue shaded cells represent *B. multivorans* strains with high motility phenotypes

Table 52 | Biofilm formation of the *B. multivorans* primary pathogens when interacted with a secondary CF pathogen.

Primary Pathogen	Secondary Pathogen	Average Biofilm Production	Standard Deviation	P-value	P-adjusted	Significance
BCC0010	*S. maltophilia*	0.706116	-1.91881	0.027504	0.756371	ns
	R. *mannitolilytica*	0.186722	-1.49241	0.067797	0.466101	ns
	A. *xylosoxidans*	0.168127	-1.2792	0.100413	0.368179	ns
	P. aeruginosa PA14	1.09872	-2.13201	0.016503	0.907672	ns
	P. aeruginosa LES B58	0.025163	-0.4264	0.334908	0.43857	ns
	B. cenocepacia J2315	0.030121	-0.8528	0.196884	0.386737	ns
	C. albicans	0.589953	-1.70561	0.044041	0.403707	ns
	S. aureus	0.082642	-1.066	0.143211	0.437589	ns
	B. cenocepacia K56-2	0.022366	-0.6396	0.261216	0.410482	ns
	B. cenocepacia BCC0019	0.028092	-0.2132	0.415585	0.415585	ns
BCC0255	*S. maltophilia*	0.183859	-1.72133	0.042596	0.390462	ns
	R. *mannitolilytica*	0	0	0.5	0.5	ns
	A. *xylosoxidans*	0.068363	-0.86066	0.194712	0.38247	ns
	P. aeruginosa PA14	0.741174	-1.93649	0.026404	0.484069	ns
	P. aeruginosa LES B58	0.028348	-0.43033	0.333477	0.436696	ns
	B. cenocepacia J2315	0.045874	-0.6455	0.259303	0.407475	ns
	C. albicans	0.097701	-1.50616	0.066013	0.363072	ns
	S. aureus	0	0	0.5	0.509259	ns
	B. cenocepacia K56-2	0.073553	-1.07583	0.141002	0.369291	ns
	B. cenocepacia BCC0019	0.074827	-1.29099	0.098353	0.360627	ns
BCC0080	*S. maltophilia*	0.730443	-1.066	0.143211	0.492288	ns
	R. *mannitolilytica*	-0.01297	1.066004	0.143211	0.437589	ns
	A. *xylosoxidans*	0.142339	-0.4264	0.334908	0.409332	ns
	P. aeruginosa PA14	0.602868	-0.8528	0.196884	0.433145	ns
	P. aeruginosa LES B58	0.111193	-0.2132	0.415585	0.44818	ns
	B. cenocepacia J2315	0.060839	0.213201	0.415585	0.423281	ns
	C. albicans	0.168413	-0.6396	0.261216	0.410482	ns
	S. aureus	0.008592	0.852803	0.196884	0.47081	ns
	B. cenocepacia K56-2	0.058064	0.426401	0.334908	0.42837	ns

	B. cenocepacia BCC0019	0.0357	0.639602	0.261216	0.399079	ns
BCC0009	*S. maltophilia*	0.441023	1.279204	0.100413	0.502063	ns
	R. *mannitolilytica*	-0.00933	-0.8528	0.196884	0.492211	ns
	A. *xylosoxidans*	0.091623	0.639602	0.261216	0.495409	ns
	P. aeruginosa PA14	0.396286	1.066004	0.143211	0.492288	ns
	P. aeruginosa LES B58	0.08078	0.426401	0.334908	0.484735	ns
	B. cenocepacia J2315	0.07157	0.213201	0.415585	0.486323	ns
	C. albicans	0.237283	0.852803	0.196884	0.47081	ns
	S. aureus	0.036143	-0.4264	0.334908	0.497836	ns
	B. cenocepacia K56-2	0.049793	-0.2132	0.415585	0.496895	ns
	B. cenocepacia BCC0019	0.022105	-0.6396	0.261216	0.478895	ns
BCC1190	*S. maltophilia*	0.364575	-0.2132	0.415585	0.486323	ns
	R. *mannitolilytica*	0.038428	1.492405	0.067797	0.466101	ns
	A. *xylosoxidans*	0.079405	1.705606	0.044041	0.403707	ns
	P. aeruginosa PA14	0.533459	-0.4264	0.334908	0.449266	ns
	P. aeruginosa LES B58	0.054989	0.426401	0.334908	0.42837	ns
	B. cenocepacia J2315	0.040283	1.066004	0.143211	0.39383	ns
	C. albicans	0.152492	0.213201	0.415585	0.415585	ns
	S. aureus	0.018532	1.279204	0.100413	0.460224	ns
	B. cenocepacia K56-2	0.044083	0.852803	0.196884	0.401061	ns
	B. cenocepacia BCC0019	0.039369	0.639602	0.261216	0.399079	ns
BCC0033	*S. maltophilia*	0.128441	-0.70065	0.241761	0.494511	ns
	R. *mannitolilytica*	0.031475	0.23355	0.407667	0.44744	ns
	A. *xylosoxidans*	0.02731	0.934199	0.175101	0.375216	ns
	P. aeruginosa PA14	0.094697	-0.23355	0.407667	0.426629	ns
	P. aeruginosa LES B58	0.027828	0.700649	0.241761	0.388545	ns
	B. cenocepacia J2315	0.025673	1.401298	0.080562	0.402812	ns
	C. albicans	0.095451	-0.4671	0.320214	0.423813	ns
	S. aureus	0.025735	1.167748	0.121454	0.455453	ns
	B. cenocepacia BCC0019	0.030899	0.467099	0.320214	0.400268	ns
BCC0043	*S. maltophilia*	0.657933	-1.29099	0.098353	0.505814	ns
	R. *mannitolilytica*	0.024666	0.258199	0.398127	0.447893	ns

	A. *xylosoxidans*	0.142594	-0.5164	0.302788	0.389299	ns
	P. aeruginosa PA14	0.690533	-1.54919	0.060668	0.364006	ns
	P. aeruginosa LES B58	0.144586	-0.7746	0.219289	0.39472	ns
	B. cenocepacia J2315	0.083879	-0.2582	0.398127	0.398127	ns
	C. albicans	0.188517	-1.0328	0.15085	0.362039	ns
	S. aureus	0.00432	0.516398	0.302788	0.47393	ns
BCC1147	*S. maltophilia*	0.54911	-1.4013	0.080562	0.517902	ns
	R. *mannitolilytica*	0.200518	-0.23355	0.407667	0.436786	ns
	A. *xylosoxidans*	0.326163	-0.9342	0.175101	0.375216	ns
	P. aeruginosa PA14	0.745608	-1.63485	0.05104	0.459364	ns
	P. aeruginosa LES B58	0.269873	-0.70065	0.241761	0.402935	ns
	B. cenocepacia J2315	0.23105	-0.4671	0.320214	0.400268	ns
	C. albicans	0.499218	-1.16775	0.121454	0.390388	ns
	S. aureus	2.083573	-1.8684	0.030853	0.6942	ns
	B. cenocepacia BCC0019	2.525446	-2.10195	0.017779	0.800053	ns
BCC1177	*S. maltophilia*	0.619393	-1.0328	0.15085	0.49369	ns
	R. *mannitolilytica*	0.058144	1.032796	0.15085	0.417738	ns
	A. *xylosoxidans*	0.461039	-0.5164	0.302788	0.389299	ns
	P. aeruginosa PA14	0.603623	-0.7746	0.219289	0.415495	ns
	P. aeruginosa LES B58	0.071143	0.774597	0.219289	0.39472	ns
	B. cenocepacia J2315	0.100823	0.258199	0.398127	0.421546	ns
	C. albicans	0.293732	-0.2582	0.398127	0.398127	ns
	S. aureus	0.088503	0.516398	0.302788	0.454182	ns
BCC0375	*S. maltophilia*	0.797633	-1.91881	0.027504	0.756371	ns
	R. *mannitolilytica*	-0.01265	0.213201	0.415585	0.466473	ns
	A. *xylosoxidans*	0.031453	-0.2132	0.415585	0.415585	ns
	P. aeruginosa PA14	0.644445	-1.70561	0.044041	0.403707	ns
	P. aeruginosa LES B58	0.238349	-1.2792	0.100413	0.394478	ns
	B. cenocepacia J2315	0.066503	-0.4264	0.334908	0.418635	ns
	C. albicans	0.269278	-1.49241	0.067797	0.372881	ns
	S. aureus	0.072053	-0.6396	0.261216	0.463447	ns
	B. cenocepacia K56-2	0.080998	-0.8528	0.196884	0.401061	ns

	B. cenocepacia BCC0019	0.083023	-1.066	0.143211	0.375076	ns
BCC0065	*S. maltophilia*	0.17638	-0.8528	0.196884	0.492211	ns
	R. *mannitolilytica*	0.018393	0.639602	0.261216	0.463447	ns
	A. *xylosoxidans*	0.028174	0.426401	0.334908	0.409332	ns
	P. aeruginosa PA14	0.175879	-0.6396	0.261216	0.435359	ns
	P. aeruginosa LES B58	0.039095	0.213201	0.415585	0.439561	ns
	B. cenocepacia J2315	0.161454	-0.4264	0.334908	0.418635	ns
	C. albicans	0.071401	-0.2132	0.415585	0.415585	ns
	S. aureus	0.013069	0.852803	0.196884	0.47081	ns
	B. cenocepacia K56-2	0.249269	-1.2792	0.100413	0.368179	ns
	B. cenocepacia BCC0019	0.224475	-1.066	0.143211	0.375076	ns
BCC0066	*S. maltophilia*	0.631173	-0.4264	0.334908	0.484735	ns
	R. *mannitolilytica*	-0.01649	1.279204	0.100413	0.424822	ns
	A. *xylosoxidans*	0.096171	0.213201	0.415585	0.415585	ns
	P. aeruginosa PA14	0.858004	-0.6396	0.261216	0.448964	ns
	P. aeruginosa LES B58	0.084233	0.426401	0.334908	0.418635	ns
	B. cenocepacia J2315	0.052818	1.066004	0.143211	0.39383	ns
	C. albicans	0.272056	-0.2132	0.415585	0.423281	ns
	S. aureus	0.062799	0.639602	0.261216	0.463447	ns
	B. cenocepacia K56-2	1.614744	-0.8528	0.196884	0.416486	ns
	B. cenocepacia BCC0019	0.061816	0.852803	0.196884	0.386737	ns
BCC0032	*S. maltophilia*	0.589429	-1.70561	0.044041	0.60556	ns
	R. *mannitolilytica*	0.018327	0.213201	0.415585	0.457144	ns
	A. *xylosoxidans*	0.118131	-1.066	0.143211	0.375076	ns
	P. aeruginosa PA14	0.576258	-1.49241	0.067797	0.414312	ns
	P. aeruginosa LES B58	0.015478	0.426401	0.334908	0.43857	ns
	B. cenocepacia J2315	0.05664	-0.6396	0.261216	0.399079	ns
	C. albicans	0.27258	-1.2792	0.100413	0.368179	ns
	S. aureus	0.031953	-0.4264	0.334908	0.472306	ns
	B. cenocepacia K56-2	0.061694	-0.8528	0.196884	0.401061	ns
	B. cenocepacia BCC0019	0.031768	-0.2132	0.415585	0.415585	ns
BCC0101	*S. maltophilia*	0.457348	-1.70561	0.044041	0.60556	ns

	R. *mannitolilytica*	0.074416	-0.4264	0.334908	0.460498	ns
	A. *xylosoxidans*	0.219017	-1.49241	0.067797	0.372881	ns
	P. aeruginosa PA14	0.575471	-1.91881	0.027504	0.504248	ns
	P. aeruginosa LES B58	0.180338	-1.2792	0.100413	0.394478	ns
	B. cenocepacia J2315	0.08782	-0.8528	0.196884	0.401061	ns
	C. albicans	0.164374	-1.066	0.143211	0.375076	ns
	S. aureus	0.028658	0.213201	0.415585	0.476191	ns
	B. cenocepacia K56-2	0.071718	-0.2132	0.415585	0.431268	ns
	B. cenocepacia BCC0019	0.081956	-0.6396	0.261216	0.399079	ns
BCC0384	*S. maltophilia*	0.49682	-1.49581	0.067352	0.463044	ns
	R. *mannitolilytica*	0	0.32053	0.374283	0.447512	ns
	A. *xylosoxidans*	0.164745	-1.28212	0.0999	0.392464	ns
	P. aeruginosa PA14	0.591804	-1.7095	0.04368	0.480476	ns
	P. aeruginosa LES B58	0.154193	-1.06843	0.142662	0.41297	ns
	B. cenocepacia J2315	0.116136	-0.42737	0.334554	0.427917	ns
	C. albicans	0.086286	-0.21369	0.415396	0.423088	ns
	S. aureus	0	0.32053	0.374283	0.457457	ns
	B. cenocepacia K56-2	0.116827	-0.64106	0.260742	0.421788	ns
	B. cenocepacia BCC0019	0.124045	-0.85475	0.196345	0.399963	ns
BCC0246	*S. maltophilia*	0.074617	0.641061	0.260742	0.462606	ns
	R. *mannitolilytica*	0.036593	1.068435	0.142662	0.461554	ns
	A. *xylosoxidans*	0.096321	0.213687	0.415396	0.423088	ns
	P. aeruginosa PA14	0.095845	0.427374	0.334554	0.438106	ns
	P. aeruginosa LES B58	0.030033	1.282122	0.0999	0.422654	ns
	B. cenocepacia J2315	0.030024	1.495808	0.067352	0.411594	ns
	C. albicans	0.201379	-0.21369	0.415396	0.431071	ns
	S. aureus	0.013824	1.816339	0.034659	0.635419	ns
	B. cenocepacia K56-2	0.013824	1.816339	0.034659	0.476564	ns
	B. cenocepacia BCC0019	0.050428	0.854748	0.196345	0.399963	ns
BCC0247	*S. maltophilia*	0.122385	-0.8528	0.196884	0.492211	ns
	R. *mannitolilytica*	0.042768	-0.2132	0.415585	0.44818	ns
	A. *xylosoxidans*	0.028457	0.639602	0.261216	0.399079	ns

	P. aeruginosa PA14	0.094881	-0.4264	0.334908	0.449266	ns
	P. aeruginosa LES B58	0.038657	0.213201	0.415585	0.423281	ns
	B. cenocepacia J2315	0.017503	1.066004	0.143211	0.39383	ns
	C. albicans	0.107961	-0.6396	0.261216	0.410482	ns
	S. aureus	0.019342	0.852803	0.196884	0.47081	ns
	B. cenocepacia K56-2	0.133401	-1.066	0.143211	0.414558	ns
	B. cenocepacia BCC0019	0.033929	0.426401	0.334908	0.409332	ns
ATCC 17616	*S. maltophilia*	0.088363	0.639602	0.261216	0.478895	ns
	R. *mannitolilytica*	0.107142	0.426401	0.334908	0.449266	ns
	A. *xylosoxidans*	0.046784	1.492405	0.067797	0.372881	ns
	P. aeruginosa PA14	0.066855	0.852803	0.196884	0.416486	ns
	P. aeruginosa LES B58	0.054227	1.279204	0.100413	0.394478	ns
	B. cenocepacia J2315	0.107526	0.213201	0.415585	0.415585	ns
	C. albicans	0.057276	1.066004	0.143211	0.375076	ns
	S. aureus	0.026403	1.918806	0.027504	0.756371	ns
	B. cenocepacia K56-2	0.03512	1.705606	0.044041	0.403707	ns
	B. cenocepacia BCC0019	0.01978	2.132007	0.016503	0.907672	ns
BCC1272	*S. maltophilia*	0.707168	-0.4264	0.334908	0.497836	ns
	R. *mannitolilytica*	0.27777	0.426401	0.334908	0.460498	ns
	A. *xylosoxidans*	0.303485	0.213201	0.415585	0.415585	ns
	P. aeruginosa PA14	0.533209	-0.2132	0.415585	0.44818	ns
	P. aeruginosa LES B58	0.201597	1.066004	0.143211	0.414558	ns
	B. cenocepacia J2315	0.217019	0.639602	0.261216	0.399079	ns
	C. albicans	0.131413	1.705606	0.044041	0.403707	ns
	S. aureus	0.202095	0.852803	0.196884	0.451193	ns
	B. cenocepacia K56-2	0.166102	1.492405	0.067797	0.414312	ns
	B. cenocepacia BCC0019	0.182187	1.279204	0.100413	0.368179	ns
BCC0084	*S. maltophilia*	0.219508	-0.8528	0.196884	0.492211	ns
	R. *mannitolilytica*	0.032114	0.426401	0.334908	0.449266	ns
	A. *xylosoxidans*	0.025136	0.639602	0.261216	0.399079	ns
	P. aeruginosa PA14	0.082841	-0.4264	0.334908	0.43857	ns
	P. aeruginosa LES B58	0.064239	-0.2132	0.415585	0.431268	ns

B. cenocepacia J2315	0.23537	-1.066	0.143211	0.375076	ns
C. albicans	0.217108	-0.6396	0.261216	0.410482	ns
S. aureus	0.033777	0.213201	0.415585	0.466473	ns
B. cenocepacia K56-2	0.277243	-1.2792	0.100413	0.394478	ns
B. cenocepacia BCC0019	0.27858	-1.49241	0.067797	0.372881	ns