# Identifying and characterizing antimicrobial producing *Burkholderia* from medicinal plant rhizospheres and soil in the Western Ghats, India.

**Thesis presented by**

Nicolaas Dominique Bruyniks

BSc (Hons) Biotechnology

In candidature for the degree of Master of Philosophy

Microbiomes, Microbes and Informatics Group

School of Biosciences

Cardiff University

July 2020

# Summary

The *Burkholderia* genus is a diverse group of Gram-negative bacteria found in soil, water, man-made products, plants, animals and as opportunistic pathogens in multiple hosts. The *Burkholderia cepacia* complex (Bcc) is a notorious group of *Burkholderia* species which can cause devastating infections in individuals with cystic fibrosis (CF). The biosynthetic capacity of *Burkholderia* species is substantial and has made the genus an interesting source for the discovery of specialized metabolites. This research sought to expand the understanding of *Burkholderia* diversity in the natural environment and the capability of this group to produce antimicrobial specialized metabolites. Bacteria were isolated from the rhizosphere of medicinal plants in the Western Ghats, India, and purified into a collection of 73 environmental isolates. Sequencing of *recA* and 16S rRNA genes identified 35 *Burkholderia* strains, which all grouped in the *Burkholderia cepacia* complex (Bcc). Seven *B. cenocepacia* strains were from the *recA* IIIA lineage, a group of virulent strains for which no clear environmental source has been defined. Draft genome sequences were obtained for 34 *Burkholderia* strains and used to carry out multiple phylogenomic analyses. Four known Bcc species groups and 5 putative novel Bcc taxa were found within the collection. Comparative genomic analysis of the 7 environmental *B. cenocepacia* IIIA strains against 32 well characterized reference genomes revealed they were closely related to a globally spread and virulent *B. cenocepacia* sequence type, ST-32. Antimicrobial activity screening revealed that 26 of the 35 *Burkholderia* strains were antagonistic towards other human and plant pathogenic microbes. Genome mining and chemical analysis of bioactive strains led to the discovery of the first collimonin producing *Burkholderia*, and the first enacyloxin-producing *B. ubonensis* strains, providing further novel insights into the biosynthetic potential of the *Burkholderia* genus.

# Contents

# List of Tables

# List of Figures

# Abbreviations

| | |
|---|---|
| ANI | Average Nucleotide Identity |
| AntiSMASH | Antibiotics and Secondary Metabolite Analysis Shell |
| ATCC | American Type Culture Collection |
| Bcc | *Burkholderia cepacia* complex |
| BCESM | *Burkholderia cepacia* epidemic strain marker |
| BGC | Biosynthetic Gene Cluster |
| BLAST | Basic Local Alignment Search Tool |
| BLASTN | Nucleotide BLAST |
| BLASTP | Protein BLAST |
| bp | base pairs |
| BSM | Basal Salts Media |
| BSMG | Basal Salts Media supplemented with Glycerol |
| *cci* | *Bukholderia cenocepacia* island |
| CF | Cystic Fibrosis |
| CLIMB | Cloud Infrastructure for Microbial Bioinformatics |
| DDH | DNA-DNA hybridisation |
| DMSO | Dimethylsulfoxide |
| DNA | Deoxyribonucleic acid |
| dNTPs | Deoxyribonucleotide triphosphates |
| ENA | European Nucleotide Archive |
| ESI-Q-TOF-MS | Electrospray ionization-quadrupole-time of flight-mass spectrometry |
| *esmR* | epidemic strain marker Regulator |
| HPLC | High Performance Liquid Chromatography |
| kbp | kilo base pairs |
| LMG | Belgian Co-ordinated Collections of Micro-organisms, Gent |
| mbp | mega base pairs |
| MiBig | Minimum Information on Biosynthetic Gene Clusters |
| MIC | Minimum Inhibitory Concentration |
| MLSA | Multi-locus Sequence Analysis |
| MLST | Multi-locus Sequence Typing |
| MS | Mass Spectrometry |
| m/z | Mass-to-charge (ratio) |
| NCBI | National Centre of Biotechnology and Information |
| NCTC | National Collection of Type Cultures |
| NRPS | Non-Ribosomal Peptide Synthetase |
| OD | Optical Density |
| PAB | *Pseudomonas* agar base |
| PacBio | Pacific Biosciences |
| PBS | Phosphate buffer saline |
| PCAT | *Pseudomonas cepacia* azelaic acid tryptamine |
| PCR | Polymerase chain reaction |
| PFGE | Pulse Field Gel electrophoresis |
| RFLP | Random Fragment Length Polymorphism |
| PKS | Polyketide Synthase |
| QUAST | Quality Assessment Tool for Genome Assemblies |
| RAPD | Randomly Amplified Polymorphic DNA |
| *recA* | recombinase A |
| rMLST | Ribosomal Multi Locus Sequence Typing |
| rRNA | Ribosomal Ribonucleic Acid |
| RNA | Ribonucleic Acid |
| rpm | Revolutions per minute |
| TOF-MS | Time of Flight-Mass spectrometry |
| TSA | Tryptone Soya Agar |
| TSB | Tryptone Soya Broth |
| UHPLC | Ultra high performance liquid chromatography |

# Chapter 1 – Introduction

### 1.1.1. Project overview

This project set out to characterise *Burkholderia* bacteria recovered from the soil and rhizosphere in Mysore, South India, and evaluate their diversity and potential for specialised metabolite production. The environmental samples were obtained in collaboration with Prof. Ravishankar Rai at the University of Mysore, Mysore, India, who is interested in endophytic bacteria and the microbial communities that grow with medicinal plants. Basic bacterial enrichments and isolate purifications from the environmental samples were carried out at the University of Mysore, before the bacterial cultures were sent over to Cardiff University for collaborative analysis. This is where this Master's project began, carrying out further enrichments using *Burkholderia* selective agar and ensuring pure bacterial strains were isolated for a systematic collection. *Burkholderia* bacteria are difficult to identify and therefore molecular approaches using single gene sequencing were initially employed to identify putative species. Whole genome sequencing was then used to obtain accurate species identification, as well as characterising the diversity of different strains within these taxa. Antimicrobial metabolites were characterised within the collection of environmental *Burkholderia* strains using phenotypic antagonism assays and genome mining. The background to the project is introduced below and discusses methods for the accurate identification of *Burkholderia* species and strains, the biotechnological applications of this bacterial genus, and specifically their ability to produce specialised metabolites.

### 1.1.2. Species identification within the *Burkholderia* genus

*Burkholderia* bacteria are a highly diverse group of Gram-negatives first described as a genus by Yabuuchi *et al.* (1992) following taxonomic reclassification of isolates formerly identified by William Burkholder as *Pseudomonas cepacia* (Burkholder, 1950). Among the *Burkholderia* genus is a tight monophyletic cluster, the *Burkholderia cepacia* complex (Bcc), that are a group of medically and environmentally relevant species, which currently consists of at least 22 validly named taxa (Jin *et al.,* 2020). The Bcc group of opportunistic pathogens have been isolated from a wide variety of environmental sources and can infect plants, animals and humans (Ramette *et al.*, 2005). Notably, Bcc species pose a significant threat to individuals with cystic fibrosis (CF) where they can establish chronic infections within the lung, leading to increased morbidity and mortality among CF patients (Zlosnik *et al.,* 2015).

The clinical significance of *B. cepacia* complex species as human opportunistic pathogens, and their genetic diversity has made bacteria identification and classification of the *Burkholderia* a topic of

extensive investigation. The first reports of CF lung colonization by *Burkholderia* strains, published in the 1970s and 80s, described isolated strains as the single species *B. cepacia* (Coenye *et al.* 2001). In the mid-1990s, researchers reported suspected *B. cepacia* strains to show great heterogeneity for various phenotypic and biochemical identification techniques. Eventually, Vandamme *et al.* (1997) adopted a polyphasic classification approach by conducting phenotypic tests, biochemical tests and DNA-DNA and DNA-rRNA hybridization assays to demonstrate that *B. cepacia*-like strains, isolated from CF patients, represented at least five different species. These five species came to be known as the *B. cepacia* complex. Species within the complex were initially described as genomovars, which refers to strains as being phylogenetically differentiable but difficult to distinguish phenotypically (Vandamme *et al.,* 1997). The term genomovar is no longer used as the species concerned have now been assigned specific names.

Nucleic acid sequence-based identification techniques have now surpassed biochemical and phenotypic tests as a more accurate means of bacterial identification. The earliest studies using PCR-based identification of Bcc species were based on the diversity of the 16S and 23S rRNA genes (Coeyne *et al.,* 2001). Random fragment length polymorphism (RFLP) analysis of the 16S and 23S rRNA gene successfully differentiated some species of the Bcc (Coeyne *et al.,* 2001). PCR-amplified 16S rDNA RFLP analysis revealed sequence polymorphisms capable of identifying *B. multivorans* and *B. vietnamiensis but* was insufficient to discriminate strains of *B. cepacia* genomovars I and III and *B. stabilis (*Mahenthiralingam *et al.,* 2000). The limited utility of 16S rRNA and 23S rRNA gene RFLP analysis for Bcc species differentiation encouraged the development of PCR-based identification assays with greater discriminatory power. Mahenthiralingam *et al.* (2000) developed a PCR-based identification assay targeting the *recA* gene, a house keeping gene involved in homologous recombination. RFLP analysis of the PCR-amplified *recA* gene demonstrated a clear split of all 5 genomovars, as identified by Vandamme *et al.* (1997), into distinct species (Mahenthiralingam *et al.,* 2000).

Furthermore, the development of genotyping techniques played an important role in understanding the epidemiology of the Bcc (LiPuma, 2003). Studies in the early 1990s using various strain genotyping techniques showed that certain Bcc species could spread from patient-to-patient and helped identify the most problematic species within the complex (LiPuma, 2003). The most predominant species within the complex to colonise CF patients were identified as *B. cenocepacia* and *B. multivorans* formerly known as genomovars II and III, respectively. *B. multivorans* has now become the most prevalent CF species in multiple countries (Turton *et al.,* 2007; LiPuma, 2010; Silva *et al*., 2016) although individuals with *B. cenocepacia* infections have a significantly poorer clinical outcome (Zlosnik *et al.,* 2015).

### 1.1.3. Gene sequence based phylogenetic analysis of *Burkholderia*

Advances in gene sequencing provided greater resolution than RFLP analysis for phylogenetic analysis, improving species delineation of the *Burkholderia* genus. Databases such as Greengenes (http://greengenes.secondgenome.com/), Silva (https://www.arb-silva.de/), EzBioCloud (https://www.ezbiocloud.net/) and NCBI 16S RefSeq (https://www.ncbi.nlm.nih.gov/refseq/) allow comparison of a strain's 16S rRNA gene sequence to those of well characterised species, and can also determine the strain's evolutionary relationship to them. Although, the 16S rRNA gene has been widely adopted for the classification of multiple bacterial species, distinguishing between closely related *Burkholderia* species can still be problematic due to its low sequence variation in this taxonomic group (Mahenthiralingam *et al.,* 2000b). Distinguishing between different members of the *B. cepacia* complex using 16S rRNA gene analysis has been reported to be particularly problematic (Payne *et al.,* 2005). In contrast, the *recA* gene sequence has been reported to discriminate between *B. cepacia* complex species with greater efficacy than 16S rRNA gene analysis (Mahenthiralingam *et al.,* 2000b; Payne *et al.,* 2005; Drevinek *et al*. 2008). Payne *et al.* (2005) also reported the design of PCR primers (BUR1 and BUR2) which could amplify an 869 bp sequence of the *recA* gene beyond species within the Bcc and across multiple species within the *Burkholderia* genus. In the original publication, these PCR primers and the derived *recA* gene sequence was able to successfully separate 28 different *Burkholderia* species.

### 1.1.4. Multi-locus sequence typing provided Bcc discrimination at the species and strain level

Although *recA* gene analysis has been successful in discriminating closely related species of the Bcc, methods such as multi-locus sequence typing were developed to have greater discriminatory power, distinguishing both species and the strains within them (Baldwin *et al.,* 2005; Drevinek, 2008; Cesarini *et al.*, 2009). Baldwin *et al.* (2005) developed the first Bcc multi-locus sequence typing (MLST) scheme, targeting seven house-keeping genes: *atpD*, *gltB*, *gyrB*, *recA*, *lepA*, *phaC*, and *trpB,* which has subsequently been widely adopted for the identification of *B. cepacia* complex species (Drevinek, 2008). MLST assigns individual isolates with an arbitrary number that represents the unique sequences of the 7 MSLT alleles for that strain. Additionally, multi-locus sequence analysis (MLSA) which utilizes the nucleotide sequences of the 7 alleles can be employed for phylogenetic analysis. The MLST scheme designed by Baldwin *et al.* (2005) improved identification at both the species and strain level and aided in determining the evolutionary relationships between members of *Burkholderia cepacia* complex. Additionally, the low mutation rate in the MLST housekeeping genes and a large publicly available web database (https://pubmlst.org/bcc) has made this scheme useful for tracing the global epidemiology of Bcc bacteria (Waine *et al.,* 2007; Drevinek and Mahenthiralingam, 2010; Fila and Drevinek, 2017).

For instance, *B. cenocepacia* strains which were initially identified as belonging to the ET-12 lineage, a lineage determined by genotyping using macrorestriction followed by pulse field gel electrophoresis, were found to belong to 5 different MLST's (Drevinek and Mahenthiralingam, 2010). Interestingly, only the clone representing ST-28 spread intercontinentally between Canada and the UK. Furthermore, MLST has identified other epidemic *B. cenocepacia* clones such as ST-32 which spread epidemically through CF populations in Canada (strain type RAPD01) and the Czech Republic (strain CZ1) (Drevinek and Mahenthiralingam, 2010).

The development of sequence-based phylogenetic analysis has expanded our knowledge concerning the relationships of bacteria isolates within species, and across intraspecies groupings. Characterisation of *B. cenocepacia* using *recA* gene sequence analysis demonstrated this species could be split into distinct phylogenetic clusters based on polymorphisms in the *recA* gene (Mahenthiralingam *et al.,* 2000b; Vandamme *et al.,* 2003). Development of the Bcc MLST scheme improved the resolution of these subgroups and offered further support for *B. cenocepacia* subgroups III-A, III-B, III-C and III-D (Baldwin *et al.*, 2005). An additional subgroup IIIE was reported by Baldwin *et al.* (2005) based on MLSA of several *B. cenocepacia* isolates but was later reported as a miss-assignment (Drevinek *et al.,* 2008). Similarly, *B. cepacia* was proposed to split into two distinct lineages based on phylogenetic analysis of the *recA* gene (Payne *et al.*, 2005). One lineage formed a cluster with the *B. cepacia* type strain (ATCC 25416), and the second lineage was designated group K (Payne *et al.*, 2005). However, MLSA identified two unique species lineages in group K with the members of each lineage eventually classified as *Burkholderia contaminans* (LMG 23361) and *Burkholderia lata* (LMG 22485) (Vanlaere *et al.,* 2009).

MLST has given microbiologists a robust and portable method for identifying bacteria, determining evolutionary relationships and epidemiological monitoring (Maiden, 2006). However, the MLST scheme designed by Baldwin *et al.* (2005) was initially limited to species within the Bcc. Subsequently, Spilker *et al.* (2009) redesigned the PCR primers for each of the housekeeping genes to enable amplification of sequences from multiple *Burkholderia* species across the genus. Furthermore, the expanding number of genomic sequences in microbiology has enabled the development of MLST schemes such as the 53-gene ribosomal MLST (rMLST) (Jolley *et al.,* 2012). This rMLST scheme is a useful tool for comparing core genes across the entire *Burkholderia* genus and sister clades (Depoorter *et al.*, 2016). Genomic methods of gene sequence and whole genome sequence comparisons are now being used as the ultimate means of differentiating bacterial taxa and the strains within them (Maiden *et al.*, 2013). However, the number of *Burkholderia* currently characterized using the 7 gene MLST scheme comprises over 3600 isolates and 1800 sequence types (https://pubmlst.org/bcc), and hence

profiles the greatest current diversity resource available from which to evaluate an uncharacterised strain.

### 1.1.5. Whole genome sequencing and comparative genomics of *Burkholderia*

Since the early 1990s, because of advances in molecular biology, criteria used for the classification of bacteria have been constantly changing. Advances in high-throughput whole genome sequencing, bioinformatic software and sequence database availability has revolutionised the study of bacteria ushering in the era of phylogenomics i.e. the use of whole-genome sequences in bacteria taxonomy (Caputo *et al.,* 2019). Genome-based taxonomy has also significantly altered the classification of *Burkholderia* species, improving species delineation and reducing incongruities observed for phylogenetic analysis utilizing a single gene or MLSA. In recent years, multiple species from the *Burkholderia* genus have been reclassified to several newly described genera including *Paraburkholderia* (Sawana *et al.,* 2014), *Caballeronia* (Dobritsa and Samadpour, 2016), *Robbsia* (Lopes-Santos *et al.,* 2017), *Mycetohabitans* (Estrada-de los Santos *et al.,* 2018) and *Trinickia* (Estrada-de los Santos *et al.,* 2018). All these studies utilised whole genome sequences to some extent in their analysis.

Genome sequence analysis permits researchers to identify pan-genomes, specialised metabolite biosynthetic potential and risk factors for virulence in studied groups of bacteria. Furthermore, the pan-genome can be used to identify the core and accessory genome of a study group. The first definition of the pan-genome was proposed by Tettelin *et al.* (2005) and is defined as being the entire gene content belonging to a study group. The core genome is defined as the genes shared by all the organisms in a study group, whereas the accessory genome includes the genes which are not present in all the organisms under investigation. The degree of sequence similarity within the core genome is considered one of the best phylogenomic measures for comparing microbial genomes (Rokas *et al.,* 2003).

*In silico* whole genome sequence comparison has led to the development of several genome-wide similarity statistics to determine species boundaries and confirm species identification, such as average nucleotide identity (ANI), average amino acid identity (AAI) and maximal unique matches index (MUMi) among others (Chun and Rainey, 2014). Average nucleotide identity is a similarity index generated by comparing the coding regions of a given pair of genomes. ANI has been proposed as a whole genome sequence-based alternative to the classical DNA-DNA hybridisation (DDH) for the differentiation of bacterial species. (Richter and Rosello, 2009). DDH was developed by Brenner *et al.* (1969) and involves the annealing of complementary DNA strands between genomes to form hybrid

molecules. DDH had been considered the gold standard for species delineation as it offers genome wide comparisons between organisms. Hybridization percentages greater than or equal to 70% has been widely adopted as the standard for species delineation (Goris *et al.*, 2007). With the availability of high through-put sequencing, public genomic databases and the laborious nature of DDH the method is seen as an outdated approach to species demarcation (Richter and Rosello, 2009). Richter and Rosello (2009) reports an ANI of 95% to correspond closely to a DDH of 70% and is a suitable alternative to DDH for the classification of bacterial species.

The majority of prokaryote taxonomy to date has relied on polyphasic combinations of biochemical and phenotypic characteristics and genotypic properties including DNA-DNA hybridization, genomic G + C content composition and 16S rRNA sequence similarity (Caputo *et al.,* 2019). The development of tools such as *recA* gene sequencing and MLSA has helped us to gain a better understanding of the taxonomic complexity, diversity and distribution of the *Burkholderia* genus. Furthermore, MLST has been essential in understanding the epidemiology of Bcc bacteria and has played a key role in the development of infection control measures to halt strain transmission in CF populations (Drevinek and Mahenthiralingam, 2010). Advances in high-throughput sequencing, publicly available genome sequence data and advances in bioinformatics has revolutionised bacteria taxonomy, and now allows researchers to carry out genome level analysis to delineate strains and species. Phylogenomic analysis has significantly altered our description of *Burkholderia* taxonomy in recent years and will continue to be essential in understanding the diversity of this genus.

### 1.1.6. The environmental roles and prevalence of *Burkholderia* in different habitats

Members of the *Burkholderia cepacia* complex have a widespread geographical distribution and have been isolated from a diverse range of environmental habitats. For example, Draghi *et al.* (2018) reports isolating 24 different species of *Burkholderia* from Argentinean agricultural soil*,* based on *recA* gene sequence analysis and MLSA, identifying at least 4 Bcc like species. Maravić *et al.* (2012) reports isolating 47 Bcc strains from seawater and mussel (*Mytilus galloprovincialis*) samples collected along the Adriatic coast. In addition, Bcc species have also been isolated from artificial surfaces such as hospital equipment (Moore *et al.,* 2002) and have been reported as a common industrial contaminant (Cunningham-Oakes *et al*., 2019). Other *Burkholderia* species outside of the Bcc also inhabit a wide array of ecological niches (Depoorter *et al.*, 2016). For example, *B. pseudomallei* has been recovered from soil, bodies of water and the plant rhizosphere, as well as being a problematic pathogen and cause of melioidosis (Seng *et al.,* 2019).

Interactions with organisms in the environment can result in contrasting relationships. Symbiotic relationships formed by *Burkholderia* can be beneficial and/or harmful depending on the organism with which they interact. Most studies have focussed on the harmful microbe-human interactions by members of *Burkholderia* pathogenic groups, such as the Bcc and *Pseudomallei* group (Suarez-Moreno *et al.,* 2012). *Burkholderia* species outside of those groups are also considered to be phytopathogens such as *B. glumae*, *B. plantarii* and *B. gladioli* originally identified as the etiological agents of seedling blight, grain rot and sheath rot in rice, respectively (Seo *et al.,* 2015). Nevertheless, certain phytopathogenic species have also been reported to cause opportunistic infections in humans. *B. gladioli for* example has been reported to cause acute respiratory tract infections in CF patients and immune compromised individuals (LiPuma, 2010), highlighting the adaptability of *Burkholderia* bacteria to a variety of lifestyles.

### 1.1.7. *Burkholderia* as plant associated bacteria

*Burkholderia* species are widely reported as plant-associated bacteria either free-living in the rhizosphere or as epiphytic or endophytic bacteria (Suarez-Moreno *et al.,* 2012). *B. cepacia* was initially identified as a plant pathogen responsible for sour skin rot disease of onion (Yabuuchi *et al.*, 1992). Additionally, *Burkholderia* and the closely related, recently defined *Paraburkholderia*, have been reported to form beneficial relationships with insects and fungi (Compant *et al.,* 2008a). For example, Kikuchi *et al.* (2007) reports *Burkholderia* symbionts, isolated from the gut of the alydid stink bug *Riptortus clavatus,* to positively enhance host growth and the insect's fecundity.

Attempts have been made at dividing the *Burkholderia* genus*,* using single gene sequence analysis, multi locus sequence typing and whole genome analysis, into broadly pathogenic and non-pathogenic groups (Sawana *et al.,* 2014). The initial description of *Paraburkholderia* genus was proposed to separate the *Burkholderia* genus into two genera. The *Burkholderia* group contained the clinically important and phytopathogenic members of the genus, and the *Paraburkholderia* group encompassing environmental species with plant-associated beneficial traits. However, this division has been scrutinised as the *Burkholderia* genus consist of species which can be beneficial to plants and animals (Eberl and Vandamme, 2016), and the *Paraburkholderia* contains species such as *Paraburkholderia fungorum* which can occasionally cause severe infections (Gerrits *et al.,* 2005). Divisions within the Bcc based on plant and animal interactions have also been suggested. Wallner *et al.* (2019) proposes a division of *B. cenocepacia* into two separate species lineages based on core genome analysis and the distribution of key virulence factors and plant associated genes. Overall, it is difficult to absolutely state that certain species of *Burkholderia* or *Paraburkholderia* are pathogenic or

not, and nearly all can be found in the natural environment, so each species and the strains within it should also be discussed in the context of both traits.

### 1.1.8. *Burkholderia* as plant beneficial bacteria

Although many of the plant beneficial bacteria previously described as *Burkholderia* have been reclassified as *Paraburkholderia* (Sawanna *et al.*, 2014), or other newly described genera, several strains including those in the Bcc have been reported to have plant beneficial properties (Parke and Gurian-Sherman, 2001). Plant beneficial characteristics of *Burkholderia* can include pathogen protection, promoting plant growth and plant nutrient acquisition (Suarez-Moreno *et al.,* 2012). Plant growth promotion through atmospheric nitrogen fixation and phytohormone synthesis have been reported for *Burkholderia* (Compant *et al.,* 2008a). Tran Van *et al.* (2000) report a nitrogen fixing Bcc strain *B. vietnamiensis* (TVV74) to significantly enhance grain yield in inoculated rice plants. Ho *et al.* (2015) identified several metabolic pathways in *B. cenocepacia* (869T2) responsible for plant growth promoting factors such as 1-aminocyclopropane-1-carboxylate (ACC) deaminase and pyrroloquinoline quinone production. Furthermore, several enzymes and metabolic pathways have been identified in *Burkholderia* allowing them to catabolize plant hydrocarbon derivatives, highlighting this genus' close relationship with plants (Wallner *et al.*, 2019).

### 1.1.9. *Burkholderia* and their use in agriculture

*Burkholderia* bacteria have been recovered from a wide variety of plants, including the rhizosphere of major crop species such as *B. vietnamiensis* from rice in Vietnam (Van Tran *et al.,* 1996), *B. cepacia* from wheat in the UK (Richardson *et al.,* 2002) and up to 7 different Bcc species from maize in the US (Ramette *et al.,* 2005). The plant beneficial properties of *Burkholderia* and the ability to colonise major plant crop species has generated significant commercial interest for their use in agriculture as biocontrol, bioremediation, and plant-growth promoting agents (Compant *et al.,* 2008a). For example, members of the Bcc group have been used in agricultural products such as biopesticides i.e. products with living organisms or their natural products which supress the growth or spread of crop pathogens or pest (Parke and Gurian-Sherman, 2001). Biopesticide products such as DENY (*B. ambifaria*) have been used to control fungi and nematodes, and Blue Circle (*B. cenocepacia*) developed to suppress fungi (Eberl and Vandamme, 2016). However, due to the human pathogenic nature of Bcc species a moratorium has been placed on the use of Bcc bacteria for biopesticides (US Environmental Protection Agency 2004) (Parke and Gurian-Sherman, 2001). The recovery of the transatlantic epidemic strain *B. cenocepacia* (IIIB) PHDC from CF patients and from agricultural soils in the US further highlights the risk associated with their use as biopesticides (LiPuma *et al*., 2002).

The efficacy of Bcc species as biocontrol strains has been well established, however, a greater understanding of the mechanisms of biocontrol is needed in order to develop safer and more effective biopesticides (Parke and Gurian-Sherman, 2001). Nevertheless, concerns over the toxicity and bioaccumulation of man-made pesticides, and emerging resistance in pest populations, has renewed research interest into biopesticides as natural alternatives (Chandler *et al.,* 2011). In order to exploit the biopesticidal properties of bacteria, concerns over efficacy, strain pathogenicity and persistence in the environment need to be addressed. For example, Mullins *et al.* (2019) demonstrated that by curing the *B. ambifaria* strain BCC0191 of its pC3 megaplasmid, a non-essential plasmid linked to virulence in Bcc species, the mutant strain had reduced persistence in a murine respiratory infection model but maintained biocontrol of *Pythium*. Additionally, Mullins *et al.* (2019) demonstrated the polyyne cepacin in *B. ambifaria* to be responsible for the protection of germinating crops against *Pythium* damping-off disease. By identifying the compound responsible for the biopesticidal properties of *B. ambifaria* and its associated biosynthetic gene cluster (BGC)*,* expression of the cepacin BGC in a heterologous host, with no associated pathogenicity, could provide another route to developing safer and more effective biopesticides. Nevertheless, the application of *Burkholderia* based biopesticides remains under review until we gain a better understanding of their safety and fate in the natural environment (Parke and Gurian-Sherman, 2001).

### 1.1.10. The *Burkholderia* genome and third replicon

*Burkholderia* have an average genome size of approximately 7.5 Mb placing them in the top 5% of large bacterial genomes (Kunakom and Eustáquio, 2019). Their large and complex genomes have been credited for the versatility and diversity of the members of this genus. In addition, *Burkholderia* species can have up to 6 plasmids and contain a large number of insertion sequences giving this species considerable genomic plasticity (Kunakom and Eustáquio, 2019). Most Bcc bacteria feature two circular chromosomes and, what was until recently considered a chromosome, a highly conserved megaplasmid (Agnoli *et al.,* 2012). The third replicon or megaplasmid, designated pC3, has been identified as a non-essential plasmid that encodes virulence, specialised metabolism and other accessory functions. Species which have lost their pC3 replicon can occur naturally in the environment highlighting the nonessential nature of this plasmid (Price *et al.,* 2017). In addition, Agnoli *et al.* (2012) reports a protocol to cure Bcc bacteria of the pC3 megaplasmid. Strains which had lost their pC3 megaplasmid were reported to have reduced virulence in several infection models (rat, zebra fish, *C. elegans*, *Galleria mellonella* and *Drosophila melanogaster*) (Agnoli *et al.*, 2012), further highlighting the reductive genome engineering strategy adopted by Mullins *et al.* (2019) could be a potential route to developing safer *Burkholderia* based biopesticides.

### 1.1.11. *Burkholderia* as a source of specialised metabolites

In addition to potential agricultural applications, various other biotechnological applications have been proposed for members of the *Burkholderia* genus. For example, the *B. vietnamiensis* strain G4 has been reported as a useful bioremediation strain due to its ability to degrade trichloroethylene, a common ground water aquifer contaminant in the US, and various other recalcitrant compounds (O'Sullivan *et al.,* 2007). The ability of *Burkholderia* bacteria to degrade complex compounds and xenobiotics, inhabit a wide range of ecosystems, their large plastic genomes, diverse lifestyles and relationships with eukaryotes highlights the diversity of metabolic capabilities in this genus. Furthermore, their ability to synthesise specialised metabolites has made this genus an interesting source for the discovery of novel metabolites.

The production of antagonistic compounds plays a central role in the biocontrol properties of *Burkholderia.* One of the first antifungal and antibacterial compounds discovered from *Burkholderia* was the compound pyrrolnitrin (Arima *et al.* 1964). Pyrrolnitrin (Arima *et al.* 1964) and phenazine (Cartwright *et al.,* 1995) production by *Burkholderia* has been widely reported (Depoorter *et al.,* 2016) and their involvement in the biological control of plant diseases has been well studied (Doornbos *et al.,* 2011). Other metabolites produced by *Burkholderia* involved in plant pathogen suppression have also been identified. For instance, the broad spectrum antifungal and cytotoxic compound occidiofungin, encoded by a non-ribosomal peptide synthetase (NRPS) gene cluster, has been suggested to be responsible for the biocontrol properties of *B. contaminans* (MS14) (Lu *et al.,* 2009) and *B. pyrrocinia* (Lyc2) (Wang *et al.,* 2016), with gene clusters of close homologies to those identified in *B. ambifaria* (AMMD$^{T}$) and *B. vietnamiensis* (DBO1) (Depoorter *et al.,* 2016). Furthermore, Mullins *et al.* (2019) demonstrated the production of the polyyne metabolite cepacin by *B. ambifaria* strains to mediate protection of germinating crops against *Pythium* damping-off diseases.

Specialised metabolites by *Burkholderia* are also being explored for medical applications. These metabolites are products of biosynthetic gene clusters (BGCs) that are frequently not involved in primary metabolism and play a role in a variety of functions involved in lifestyle and/or survival such as virulence, antagonism, cell-to-cell signalling and stress tolerance (Davies, 2013). The Gram-positive bacteria *Actinobacteria* have been a major source of specialised metabolites and clinical drugs, since they have large genomes which have evolved to acquire and express multiple pathways (Genilloud, 2018). However, the need for new antibiotics has encouraged the exploration of new sources for clinical drugs such as unique ecological niches and unusual bacteria genera. A large portion of the *Burkholderia* multireplicon genome is dedicated to specialised metabolism and exploration of the biosynthetic potential of this genus has yielded a variety of structurally diverse compounds with

therapeutic potential (Kunakom and Eustáquio, 2019). A recent review by Kunakom and Eustáquio (2019) reports 66 structural classes of compounds, with a structural class known or expected to be encoded in the same (or very similar) gene cluster, across the *Burkholderia* genus and closely related genera (*Paraburkholderia, Caballeronia and Robbsia*). Thus, *Burkholderia* bacteria with their large genomes, ecological diversity and diverse metabolic capabilities are key bacteria that are now being explored as a source of novel metabolites.

### 1.1.12. Strategies for specialised and antimicrobial metabolite discovery

Various compounds have been identified from *Burkholderia* with bioactivity as antifungal, antibacterial, herbicidal, insecticidal, cytotoxic and anticancer agents (Depoorter *et al.*, 2016; Kunakom and Eustáquio, 2019). Kunakom and Eustáquio (2019) reports non-ribosomal peptides (NRP) as the biosynthetic class with the largest representation in *Burkholderia,* and the closely related *Paraburkholderia, Caballeronia* and *Robbsia.* Additionally, large polyketide synthases (PKSs) or hybrid pathways (NRPS-PKS) are also common. *Trans*-AT type I polyketides appear more often in *Burkholderia* than *cis*-AT which is contrasting to the extensively studied *Actinobacteria* in which *trans*-AT is relatively rare (Kunakom and Eustáquio, 2019). With most antibiotic classes currently in use discovered in *Actinobacteria* (Genilloud, 2018) this unique biochemistry offered by *Burkholderia* BGCs could provide novel therapeutic compounds (Kunakom and Eustáquio, 2019).

Up until the early 2000s the search for new natural products has almost exclusively relied on bioactivity-guided isolation (Zelrikly and Challis, 2009). Bioactivity-guided isolation involves identifying bacteria or culture extracts capable of inhibiting the growth of fungi, nematodes, cancer cells or other bacteria. Analysis of culture extracts using nuclear magnetic resonance (NMR), mass-spectrometry (MS) and high-pressure liquid chromatography (HPLC) allows quantification and structure elucidation of interesting metabolites. Although numerous compounds have been discovered through bioactivity-guided isolation, technical challenges associated with compound identification, purification and structure elucidation can make bioactivity-guided isolation a long and laborious process (Zelrikly and Challis, 2009). Additionally, some metabolites might go undetected through bioactivity-guided isolation if the BGCs are not actively transcribed to detectable levels in the culture conditions of the bioassays (Kunakom and Eustáquio, 2019).

Advances in DNA sequencing has resulted in the deposition of vast quantities of DNA sequence data in publicly available databases, which could be exploited for the discovery of novel natural products. Large quantities of genome sequence data, combined with the growing body of knowledge on the genetics and enzymology of natural product biosynthesis, has permitted the development of

automated genome mining tools (Zelrikly and Challis, 2009). Genome mining involves analysis of an organism's genome sequence for the identification of previously uncharacterised BGCs. Several genome mining tools are available, with antiSMASH as currently one of the most advanced and well curated database and prediction tools (Blin *et al.,* 2017). Genome mining software such as antiSMASH (Blin *et al.,* 2019), SMURF (Khaldi *et al.,* 2010) and PRISM (Skinnider *et al.,* 2017) predict BGCs based on rules developed from existing knowledge about key biosynthetic steps/principles. One drawback of the rule-based approach to genome mining is that certain compounds may be missed if their biosynthetic class is currently unknown (Blin *et al.,* 2017). Additionally, this rule-based approach does not always detect BGCs which may utilise known synthetic genes, but in an unusual combination.

Although genome mining allows the identification of silent or cryptic BGCs, the structure of the encoded compound is often difficult to predict. Conventional chemical analyses such as LC-MS, NMR and X-ray crystallography are still needed to predict the chemical structure of a BGC product. Genetic engineering can be used to facilitate the expression of silent or cryptic BGCs through promoter exchange or heterologous expression so the derived BGC product can be purified and analysed (Kunakom and Eustáquio, 2019). For example, heterologous expression of a biosynthetic operon from *B. pseudomallei* under an IPTG-inducible promoter in *P. aeruginosa* led to the discovery of a type 4 phosphodiesterase inhibitor (Biggins *et al*., 2011). In addition, by linking the predicted BGC to the chemical structure of a metabolite the exact biosynthesis can be determined (Zelrikly and Challis, 2009). Therefore, conventional chemical analysis and genome guided methods should be used complementary to each other. Furthermore, genetic engineering techniques can be used for compound identification and characterisation through comparative metabolite analysis of gene knockout mutants and wild-type strains (Kunakom and Eustáquio, 2019). For example, Mullins *et al.* (2019) reported the loss of antimicrobial activity against *S. aureus* and the plant pathogenic oomycete *Pythium ultimum,* when the cepacin-encoding BGC of *B. ambifaria* (BCC0191) was disrupted through insertional mutagenesis. Mullins *et al.* (2019) thereby demonstrated that the antimicrobial metabolite cepacin was responsible for the antagonism of *Pythium* by *B. ambifaria.*

In addition to genetic engineering seemingly silent BGCs can be activated by modifying lab growth conditions. For example, Lincke *et al.* (2010) reports the discovery of a unique antibiotic, closthioamide, from *Clostridium cellulolyticum* whose production was induced through the addition of aqueous soil extracts to fermentation media. Alterations to growth conditions such as media additives, pH, temperature and incubation period are all relatively easy modifications to make and can have a profound effect on the expression of BGCs. For example, the addition of glycerol to culture

media enhanced the production of pyrrolnitrin in strains *B. cepacia* NB-1 and *Burkholderia sp.* 033 (Depoorter *et al.,* 2016). Moreover, Mahenthiralingam *et al.* (2011) reported the potent bioactivity of *B. ambifaria* (AMMD) against several Gram-negative bacteria when grown on a medium with glycerol as its sole carbon source. The compounds responsible for this activity were subsequently identified as enacyloxin IIa and its novel isomer cis-enacyloxin IIa. Although the relative simplicity of antagonism-based screening assays allows us to evaluate microbial strains under a variety of conditions in a relatively short space of time, this method can be quite labour intensive, especially as the variables which influence BGC expression are often difficult to predict.

Overall, the metabolically and ecologically diverse *Burkholderia* genus is a relatively underexploited resource of bacteria with multiple biotechnological applications. Several strains have been characterised for applications in biocontrol and bioremediation, as plant growth promoting agents and as sources of useful metabolites (Depoorter *et al.,* 2016). Advances in the bioactivity-guided isolation and genome mining has provided researchers with the tools to explore the biosynthetic potential of *Burkholderia*, and multiple compounds have already been identified in *Burkholderia* bacteria with some unique biochemistries (Kunakom and Eustáquio, 2019). Furthermore, developments in genetic manipulation and genome sequencing allows researchers to tap into the hidden biosynthetic potential of *Burkholderia* bacteria and has aided in understanding the regulation and distribution of metabolic pathways within this genus.

### 1.1.13. Project aims and objectives

The overall aim of this research was to establish the diversity of naturally occurring *Burkholderia* isolated from the rhizosphere and surrounding soil of medicinal plants in the Western Ghats, India and characterise their antimicrobial producing potential. The study sought to answer the following hypothesis:

**"The rhizosphere and surrounding soil of medicinal plants in the Western Ghats, India is a rich source of antimicrobial producing *Burkholderia* bacteria."**

The following objectives were completed for this research:

## 1. Objective 1 - Determining the diversity of *Burkholderia* isolates from the soil and rhizosphere

Multiple strategies were applied within this objective to identify and classify the bacteria that were recovered from the surrounding soil and rhizosphere of medicinal plants.

**1a. (Chapter 3) Single gene phylogenetic analysis of the bacterial collection**
A total of 73 bacteria were isolated from rhizosphere and soil samples using selective isolation media. The resulting collection of 35 *Burkholderia* bacteria were identified by gene sequence analysis of the *recA* and 16S rRNA genes. Phylogenetic analysis of the *recA* gene putatively identified all 35 *Burkholderia* to belong to the *Burkholderia cepacia* complex.

**1b. (Chapter 4) Phylogenomic analysis of isolated *Burkholderia***
From the 35 putative *Burkholderia* bacteria, 34 isolates were whole genome sequenced using short read sequences and the Illumina platform. The core, pan and accessory genome of the collection was identified and phylogenomic analysis was performed to establish the final identification of species and strain diversity within the collection.

**1c. (Chapter 4) Detailed characterisation of environmental *Burkholderia cenocepacia***
A total of 10 *Burkholderia cenocepacia* isolates were identified from the collection, of which 7 isolates were determined to belong to the *recA* IIIA lineage. An environmental source of *B. cenocepacia* IIIA strains has not been defined, despite their widespread occurrence as virulent and transmissible CF pathogens (Drevinek and Mahenthiralingam, 2010). A collection of previously described *B. cenocepacia* strains representing a broad diversity of ST types was assembled for whole genome sequencing. Phylogenomic analysis was carried out to determine the evolutionary relationship of the collection. Whole genome sequences were used to extract the Bcc scheme MLST genes and determine the ST types of the environmental isolates to better understand the diversity of the environmental

isolates and relationship to epidemic ST lineages. Additionally, the virulence potential of the *B. cenocepacia* collection was investigated by screening for genes from a pathogenicity island (*cci*) (Baldwin et al. 2004) which has been linked to increased virulence and transmissibility in *B. cenocepacia*.

## 2. Objective 2 - Investigating the biosynthetic potential of the Mysore *Burkholderia* collection

**2a. (Chapter 4) Testing the Mysore *Burkholderia* collection for bioactivity**

The Mysore *Burkholderia* collection was screened for antimicrobial activity against a panel of plant and animal pathogens representing Gram-negative bacteria, Gram-positive bacteria and fungi. A sub-collection of 4 Mysore isolates with interesting bioactivity was further characterised to identify antimicrobial compounds using bioactivity guided isolation and genome mining techniques.

**2b. (Chapter 4) Investigating the effects of different growth conditions on *Burkholderia* metabolite production**

A collection of reference *Burkholderia* strains with known bioactivity and the 4 selected environmental *Burkholderia* with interesting bioactivity, was assembled to investigate if different growth conditions could elicit novel or enhanced antimicrobial activity. The effect of different incubation length, incubation temperatures, seed exudate-based growth media and minimal media containing common carbon sources found in root exudates were investigated.

**2c. (Chapter 4) Investigating the role of the pC3 replicon in selected environmental *Burkholderia***

The 4 selected environmental *Burkholderia* were cured of their pC3 replicon to determine the effect on antimicrobial production and gain an insight into the distribution of biosynthetic pathways within these *Burkholderia.*

**2d. (Chapter 4) Bioactivity guided isolation of interesting compounds from the environmental *Burkholderia***

Culture extracts of the 4 selected environmental *Burkholderia* were analysed by HPLC to identify antimicrobial compounds. The compound enacyloxin IVa was purified from culture extracts to compare its bioactivity against enacyloxin IIa. Further LC-MS was performed by collaborators at the University of Warwick to putatively identify compounds based on their mass ion profiles.

**2e. (Chapter 4) Genome mining of selected environmental *Burkholderia***

Using the genomes assembled for this study, the genome mining tool antiSMASH was used to predict BGCs in environmental *Burkholderia* strains which exhibited interesting bioactivity and to identify the associated BGCs of compounds characterized via bioactivity guided isolation.

# Chapter 2 – Materials and Methods

## 2.1. Sample collection and strain isolation

### 2.1.1. Sample collection

A total of 66 samples were collected from the rhizosphere and surrounding soil of 12 different medicinal plants located across 6 different sites in the Western Ghats, Karnataka region, India on the 28th October 2017. The samples were collected by collaborators at the University of Mysore and from herein the collection of strains isolated from these samples are referred to as the Mysore collection throughout this thesis. Approximately 10 g of soil or root material was collected from 10-15 cm below the surface and transported to the lab in a sterile container. Soil samples of 1 g were added to 9 ml of isotonic saline and vortexed vigorously. Similarly, 1 g of root material was cut into small pieces and vortexed in isotonic saline. For each sample 0.1 mL of diluent was spread plated onto either *Pseudomonas* agar base (PAB) supplemented with cetrimide (300 mg/L) (Lowbury and Collins, 1955), or on basal salt medium (BSM) supplemented with 4 g/L of glycerol as a carbon source (BSMG), Polymyxin B (50 mg/L) (Hareland *et al.*, 1975) and crystal violet (5mg/L). Spread plates were incubated at 37 °C for 2 days. Distinct colonies were selected for the inoculation of PAB and BSMG by quadrant streaking in order to obtain pure colonies. Putative single *Burkholderia* colonies were swabbed onto Amies transport medium supplemented with charcoal (Amies, 1967) and subsequently transported via air transport to Cardiff University, Wales.

### 2.1.2. Strain isolation

Initial examination and frozen storage of the samples from the collection was performed by Lucile Hubert as part of her Erasmus project (January to June 2018), however, full strain purification was repeated and completed for the whole collection as follows. Bacteria were swabbed from Amies transport medium onto tryptic soy agar (TSA) by dilution streaking and incubated at 30°C until growth was observed (Atlas, 2010). Single colonies were picked and used to inoculate 5 mL of BSMG for an overnight culture at 30°C under shaking. Overnight BSMG cultures were used to inoculate 5 mL of *Pseudomonas cepacia* azelaic acid tryptamine (PCAT), supplemented with cycloheximide (100 mg/L) and polymyxin B (50 mg/L), by adding 100 µL of the inoculum (Burbage *et al.,* 1982; Pallud *et al.,* 2001). All samples were left for 24 hours at 30°C under shaking and subsequently stored at -80°C in 8% (v/v) dimethyl sulfoxide (DMSO). Samples were revived by incubation on TSA at 30°C for 24 hours and single colonies were picked and propagated a minimum of 3 times to ensure purity. Putative pure isolates were used to inoculate 5 mL of tryptic soy broth (TSB) for an overnight culture at 30°C under shaking. Overnight cultures were stored at -80°C in 8% DMSO to allow for subsequent follow-up work.

# 2.2. Single gene sequencing and phylogenetic analysis

### 2.2.1. DNA preparation

PCR amplification was carried out on DNA extracted by either Chelex 100 (Bio-Rad Laboratories, USA) or the Maxwell® 16 Tissue DNA Purification Kit with the Maxwell® 16 Instrument (Promega, USA). DNA extractions were carried out as follows.

Chelex-100

A small quantity of solid-surface bacterial growth was mixed with 50 µL of 10% w/v sterile Chelex-100 resin (Bio-Rad) solution. Samples were exposed to 2 alternating cycles of 5 min at 95°C in a thermocycler (MJ Research, Canada) and 5 min at -20°C in the freezer. Samples were permitted to reach room temperature before being used as DNA templates in PCR amplification (Walsh *et al.,* 2013).

Maxwell

Bacterial cultures grown in 5 mL of TSB for 24 hours at 30°C under shaking were centrifuged at 4000 rpm for 10 min. The supernatant was discarded, and the pellet was resuspended in 300 µL guanidinium isothiocyanate. DNA was extracted using the Maxwell® 16 Tissue DNA Purification Kit with the Maxwell® 16 Instrument (Promega, USA) according to the manufacturer's instructions. The DNA was eluted with 300 µL of molecular grade water. For DNA fragment size analysis, guanidinium isothiocyanate was substituted with TNE buffer (10 mM Tris-HCL, 200 mM NaCl, 100 mM EDTA, pH 8) and the sample incubated on ice for 20-30 min prior to extraction. DNA samples were stored at -20°C post-extraction.

### 2.2.2. DNA quantification and qualification

DNA extractions were quantified using the Qubit dsDNA BR (broad range, 100 pg/µL – 1 µg/µL) Assay Kit (Invitrogen, US) and the Qubit 3.0 fluorometer (Invitrogen, US) according to the manufacturer's protocols; a sample volume of 1 µl was added to 199 µl of a Qubit working solution.

DNA fragment size analysis was carried out by capillary electrophoresis using the gDNA ScreenTape with the Agilent TapeStation 4200 instrument (Agilent Technologies, US) according to the manufacturer's instructions.

### 2.2.3. PCR amplification

The universal primers 27F and 907R were used to amplify an 880-bp fragment of the V1 to V4 hypervariable regions on the 16S rRNA gene (Lane, 1991). The *recA* gene was targeted with the primers BUR1 and BUR2 to amplify an 869-bp fragment as reported by Payne *et al.* (2005). The 1.4-kb

BCESM fragment was amplified using the BCESM1 and BCESM2 probes as described by Mahenthiralingam *et al.* (1996). Primers designed by Mullins *et al.* (2019) were used to detect the pMiniC3 plasmid. Primer sequences are shown in Table 2.1, primer specific thermocycler settings in Table 2.2 and PCR mixture components in Table 2.3.

PCR products were analysed using gel electrophoresis with 1.5% agarose gel which was ran at 90 V for 45 min and was visualized using an UV Transilluminator (Bio Rad, US). The size of fragments was determined using the 1 kb DNA Ladder from New England Biolabs, US, or the 1kb HyperLadder™ from Bioline, CA.

### 2.2.4.Sequencing of PCR products and single gene phylogenetic analysis

PCR products were purified using the Monarch® PCR & DNA Cleanup Kit (New England Biolabs, US) and sequenced by Eurofins Genomics (UK).

*recA*

Preliminary identifications of *Burkholderia* isolates were made using the *recA* gene sequences by searching the NCBI "Nucleotide collection (nr/nt)" database using the basic local alignment search tool (*BLAST*; Altschul *et al*., 1990), with the MEGABLAST algorithm. Species names were assigned based on the highest sequence identity match. Furthermore, *recA* sequences from *Burkholderia* isolates were aligned with partial *recA* gene sequences of representative species of the *Burkholderia cepacia* complex obtained from the GenBank database (http://www.ncbi.nlm.nih.gov). Sequences were aligned using ClustalX2 (Larkin *et al*., 2007), trimmed to 412 base positions using MEGA version (v7.0.26) (Kumar *et al.*, 2016) and a phylogenetic tree was constructed using the neighbour-joining method with Jukes-Cantor algorithm (Kumar *et al.,* 2016). Bootstrap analysis was based on 500 replications.

16S rRNA gene

The 16S rRNA gene sequences of the isolates were compared with sequences in the NCBI "16S ribosomal RNA sequences (Bacteria and Archaea)" database using the basic local alignment search tool (*BLAST*; Altschul *et al.,* 1990), with the MEGABLAST algorithm. Preliminary identifications were made of non-*Burkholderia* isolates using 16s rRNA gene sequences, with species names assigned based on the highest sequence identity match.

**Table 2.1. Sequencing primers used in this study.**

| Primer | Sequence | Reference |
|---|---|---|
| 27F | 5' – AGAGTTTGATCMTGGCTCAG – 3' | Lane, 1991 |
| 907R | 5' – CCGTCAATTCMTTTRAGTTT – 3' | Lane, 1991 |
| BUR1 | 5' – GATCGARAAGCAGTTCGGCAA – 3' | Payne *et al.*, 2005 |
| BUR2 | 5' – TTGTCCTTGCCCTGRCCGAT – 3' | Payne *et al.*, 2005 |
| BCESM1 | 5' – CCACGGACGTGACTAACA – 3' | Baldwin *et al.*, 2004 |
| BCESM2 | 5' – CGTCCATCCGAACACGAT – 3' | Baldwin *et al.*, 2004 |
| pMiniC3 – Fwd | 5'– AAG AAA TCT GCT GCC GCT TG – 3' | Mullins *et al.,* 2019 |
| pMiniC3 – Rev | 5' – CAC TTC GCT GTA CCT CAA GC – 3' | Mullins *et al.,* 2019 |

Sequencing primers 27F and 907R amplify the 16S rRNA gene sequence and BUR1 and BUR2 amplify the recA gene. The BCESM1 and BCESM2 primers were used to amplify the Burkholderia epidemic strain marker. IUPAC codes are used for mixed bases; M, adenine or cytosine; R, adenine or guanine.

**Table 2.2. Primer specific thermocycler protocols used for PCR.**

| | recA | | | 16s rRNA | | |
|---|---|---|---|---|---|---|
| **PCR Step** | **Temperature (°C)** | **Time (s)** | **Number of cycles** | **Temperature (°C)** | **Time (s)** | **Number of cycles** |
| Initial denaturation | 95 | 300 | 1 | 95 | 120 | 1 |
| Denaturation | 95 | 30 | 30 | 94 | 30 | 36 |
| Primer annealing | 60 | 30 | 30 | 52 | 30 | 36 |
| Primer extension | 72 | 45 | 30 | 72 | 90 | 36 |
| Final extension | 72 | 300 | 1 | 72 | 300 | 1 |

| | BCESM | | | pMiniC3 | | |
|---|---|---|---|---|---|---|
| **PCR Step** | **Temperature (°C)** | **Time (s)** | **Number of cycles** | **Temperature (°C)** | **Time (s)** | **Number of cycles** |
| Initial denaturation | 95 | 300 | 1 | 95 | 180 | 1 |
| Denaturation | 95 | 60 | 30 | 95 | 30 | 30 |
| Primer annealing | 63 | 60 | 30 | 53 | 30 | 30 |
| Primer extension | 72 | 120 | 30 | 72 | 50 | 30 |
| Final extension | 72 | 600 | 1 | 72 | 300 | 1 |

**Table 2.3. Components of a 25 µL PCR reaction using Taq polymerase**

| Components | Final Volume (µL) | Final Concentration |
|---|---|---|
| ThermoScientific DreamTaq Green PCR Master Mix (2X) | 12.5 | 1X |
| 10 µM Forward Primer | 0.5 | 0.2 µM |
| 10 µM Reverse Primer | 0.5 | 0.2 µM |
| Template | 1 | Variable |
| Nuclease-Free Water | 10.5 | / |

# 2.3. Antimicrobial discovery

### 2.3.1. In vitro antimicrobial screening against pathogens

The entire Mysore collection including non-*Burkholderia* were screened for antimicrobial activity using an overlay assay modified from Mahenthiralingam *et al.* (2011), as follows. All 73 isolates were grown overnight in 3 mL of tryptic soy broth (TSB) at 37 °C under shaking. Five strains were distributed across a 96-well plate (wells B2, B10, D6, G2 and G11); using 184 µL of overnight culture and 16 µL of DMSO (8% final concentration). A 96-pin replicator was used to transfer cultures onto Petri plates of 120x120 mm with 40 mL of basal salts medium (Hareland *et al.,* 1975) supplemented with 4 g/l glycerol (BSMG) (Mahenthiralingam *et al.,* 2011). Each pin transferred roughly 2 µL of culture. Plates were left to dry and subsequently incubated for 3 days at 30°C. Following incubation, bacteria were killed by exposing the Petri dishes to chloroform vapour for 3 minutes.

Chloroform treated plates were overlaid with 20 mL of half-strength iso-sensitest agar (Oxoid, UK) cooled to 42°C and seeded with 200 µL, per 50 mL of agar, of overnight culture of the susceptibility organism. All isolates were tested against three susceptibility organisms including a Gram-negative bacterium, Gram-positive bacterium and a fungal species (Table 2.4.). Susceptibility organisms were grown overnight on TSB at their optimum growth temperature before being used to inoculate iso-sensitest agar (Table 2.4).

Overlaid plates were incubated overnight at the optimum growth temperature for the susceptibility organism. Following incubation, the diameter of the zone of inhibition was measured for each natural isolate. The non-Burkholderia isolates exhibited no antimicrobial activity against the susceptibility organism and were subsequently abandoned for further antimicrobial activity investigation. For the remaining 35 *Burkholderia* isolates the overlay assay was repeated in triplicate. A heatmap of antagonistic activity based on zones of inhibition was generated using Microsoft Excel 365.

### 2.3.2. Choosing a panel of strains for further antimicrobial screening and genome mining

Following preliminary screening of all 35 *Burkholderia* isolates, 4 strains were chosen for further investigation based on interesting antimicrobial activity. A panel of strains (Table 2.5.) with known activity, along with the 4 chosen environmental isolates and their pC3-cured derivatives, were investigated using modified versions of the overlay assay as previously reported. The panel of strains included; *B. ambifaria* strain BCC0191 a known cepacin producer (Mullins *et al.*, 2019), *B. ambifaria* strain BCC0207 a known enacyloxin producer (Mullins *et al.*, 2019), *B. gladioli* strain BCC0238 a known producer of gladiolin (Song *et al.,* 2017) and the *B. anthina* and *B. ubonensis* type strains (Yabuuchi *et al.*, 2000; Vandamme *et al.*, 2002) (Table 2.6.). The *B. anthina* and *B. ubonensis* type strains were

included to investigate if antimicrobial activity was shared amongst other members of the species or unique to the environmental strains. The strain BCC1977 was initially identified as *B. anthina* following *recA* and 16s rRNA gene sequence analysis, hence the type strain for this species was included. To investigate if silent biosynthetic gene clusters (BGCs) could be activated *in vitro* the overlay assay was modified by using various biomimetic media and altered growth conditions (Figure 2.1.).

### 2.3.3.Preparing biomimetic and root exudate media

Various media types were developed to investigate if silent BGCs could be activated *in vitro*. Additionally, the effect of altering growth conditions such as incubation temperature, incubation period and the pH of growth media were also evaluated.

Solid agar media was developed based on the exudates that are passively released during seed imbibition and germination to emulate the rhizospheric environment in which *Burkholderia* can be found. Additionally, glycerol in BSMG was replaced with sugars or amino acids (Table 2.6.) reported to be exuded by crop plants such as sugarbeet, barley and maize in iron deficient soils (Carvalhais *et al.*, 2010). The diameter of the zones of inhibition was used as a measure of antimicrobial activity and considered novel or altered in relation to the zones of inhibition observed after growth on BSMG as a control. Statistical significance was determined by two-sample (unpaired) t-test with the assumption of equal variances (Barlett test) and normally distributed data (Shapiro-Wilk test).

Pea-exudate and wheat-exudate media

Approximately 100 g of Early Onward variety pea (*Pisum sativum*) seeds or Freiston variety wheat (*Triticum aestivum)* seeds were washed 3 times with 400 mL of de-ionised water, and subsequently suspended in 400 mL of de-ionised water at room temperature (~21°C) under agitation. Following 48 hours of incubation the exudate solutions were filtered twice through a Whatman grade GF/D filter (47mm, diameter) after which it was filtered once through a Whatman grade GF/A filter. Exudate solutions were diluted 50% v/v prior to adding Oxoid purified agar (Thermo Scientific, US) at a concentration of 1.5% w/v. The solutions were standardised to a pH of 5.5 (±0.1), autoclaved and poured into plates. Plates were stored away from light at room temperature until further use.

Substituting carbon sources

A total of 4 different carbon sources were investigated as a substitute for glycerol in BSMG which are reported in Table 2.5. For each carbon source 4 g was dissolved in 40mL of de-ionised water and filter sterilised using a 0.2µM filter. Basalt salt media was standardised to a pH of 7 (±0.2), autoclaved and cooled to 50°C prior to adding the filter sterilised carbon source at 0.4% w/v.

### 2.3.4. Mutagenesis and the 3rd replicon

From the *Burkholderia* Mysore collection the 4 strains BCC1970 (*Burkholderia sp.*), BCC1973 (*B. ubonensis*), BCC1977 (*Burkholderia sp.*) and BCC1990 (*B. ubonensis*) were selected to generate 3c-null strains according to the methods described by Agnoli *et al.* (2012).

Agnoli *et al*. (2012) reports the construction of the 12.6-kb plasmid vector pMiniC3 from the *B. cenocepacia* H111 strain. The vector was constructed from the third replicon's origin of replication and the *repA, parB* and *parA* genes, fused to a modified pEX18 backbone. The pMiniC3 also carried the following genes a pBBR322 origin of replication for use in *Escherichia coli*; an origin for conjugal transfer and trimethoprim resistance and sucrose counter selection cassettes. To cure the c3 replicon from selected strains the pMiniC3 was introduced into the wildtype *Burkholderia* strains through triparental mating. This involved the donor *E. coli* MC1601 pMiniC3, the wild type *Burkholderia* strains *E. coli* HB101 carrying the helper plasmid pRK2013. By exploiting plasmid incompatibility, introduction of the pMiniC3 can result in the native plasmid being forced out. Strains carrying the pMini3C were confirmed through trimethoprim resistance and PCR using pMiniC3-specific primers (Table 2.1.). Strains carrying the pMiniC3 were subsequently cured of pMiniC3 through sucrose counter selection. Loss of pMiniC3 was confirmed using the pMiniC3-specific primers.

### 2.3.5. Rapid screening for antimicrobial metabolites

Bacteria under investigation were grown overnight on TSB at 30°C under shaking. Overnight cultures were streaked on a 9 cm diameter petri-dish on BSMG or BSM with the previously reported carbon sources. Agar plates were incubated at the various growth conditions reported in Figure 2.1. prior to extraction. Following incubation, microbial growth was removed from the surface, and a 22 mm diameter circular plug (approximately 1 g) removed from the centre of the plate and added to 500 µl of dichloromethane. The agar plug was incubated in the solvent, under agitation, for 2 hours at room temperature (~21°C). Extracts were transferred to HPLC vials and were screened for metabolites using a Waters® AutoPurification™ HPLC system fitted with a reverse-phase analytical column (Waters® XSelect CSH C18, 4.6 x 100 mm, 5 µm) and a C18 SecurityGuard™ cartridge (Phenomenex) in series. A photo-diode array detector (PDA) scanning between 210-400 nm was used to detect metabolites. Mobile phases consisted of (A) water with 0.1% formic acid and (B) acetonitrile with 0.1% formic acid with a flow rate of 1.5 ml $min^{-1}$. Elution conditions were as follows: 0 to 1 minute, 95% phase A & 5% phase B; 1 to 9 minutes, gradient of phase A from 95 to 5% and gradient of phase B from 5% to 95%; 10 to 11 minutes, 5% phase A & 95% phase B; 11 to 15 minutes, 95% phase A & 5% phase B.

### 2.3.6. Extraction of metabolites and structure elucidation using high-resolution mass spectrometry

Strains BCC1973 and BCC1990 were grown on BSMG for 3 days at 30˚C and strains BCC1970 and BCC1977 were grown on BSMG for 2 days at 22˚C prior to analysis. Single plates were extracted by addition of 4 ml acetonitrile for 2 h, followed by centrifugation to remove debris. Crude extracts were directly analysed by UHPLCESI-Q-TOF-MS by collaborators at the University of Warwick. UHPLC-ESI-Q-TOF-MS analyses were performed using a Dionex UltiMate 3000 UHPLC connected to a Zorbax Eclipse Plus C-18 column (100 × 2.1 mm, 1.8 µm) coupled to a Bruker MaXis II mass spectrometer. Mobile phases consisted of water (A) and acetonitrile (B), each supplemented with 0.1% (v/v) formic acid. A gradient of 5% B to 100% B over 30 min was used at a flow rate of 0.2 ml $min^{-1}$. The mass spectrometer was operated in positive-ion mode with a scan range of 50–3,000 m/z. Source conditions were: end-plate offset at −500 V, capillary at −4,500 V, nebulizer gas (N2) at 1.6 bar, dry gas (N2) at 8 l $min^{-1}$ and dry temperature at 180 °C. Ion transfer conditions were: ion funnel radio frequency (RF) at 200 Vpp, multiple RF at 200 Vpp, quadrupole low mass at 55 m/z, collision energy at 5.0 eV, collision RF at 600 Vpp, ion cooler RF at 50–350 Vpp, transfer time at 121 µs and pre-pulse storage time at 1 µs. Calibration was performed with 1 mM sodium formate through a loop injection of 20 µl at the start of each run.

### 2.3.7. Purifying enacyloxin-like metabolites

Strain BCC1990 was grown overnight on TSB at 30°C with shaking. The overnight culture was then used to streak 40x 9.0 cm diameter petri-dishes containing 20 mL of solid agar BSMG. After 3 days growth at 30°C, the bacterial biomass was removed with a sterile cell scraper from all plates and the agar containing bacterial metabolites (enacyloxins) cut into quarters and extracted with 500 mL of dichloromethane. Extraction was carried out by shaking on a platform rocker set to 40 rpm for 2 hours at room temperature.

To concentrate metabolites, solvent was evaporated using a Buchi Rotavapor R-3 system (Buchi, Switzerland). The dried crude extract was dissolved in 3 ml 60% (v/v) acetonitrile and molecular grade water. Suspected enacyloxin compounds were purified by preparative HPLC using the Waters AutoPurification™ HPLC system with a XSelect CSH C18, 5 µm OBD, 19 x 100x mm column with in line guard column (Waters, US). Mobile phases consisted of water (A) and acetonitrile (B), each supplemented with 0.1% (v/v) formic acid. Gradients were programmed as reported in Table 2.7. with a flow rate of 20 ml $min^{-1}$. The Waters 2998 photodiode array detector was set to scan at 210-400 nm and to selectively monitor at 360 nm for enacyloxins. Fractions of interest were collected in a series of clean glass test tubes using the Waters 2757 sample manager-fraction collector. Fractions were

subsequently pooled and non-aqueous solvent removed by vacuum centrifugation using the Barnstead Genevac miVac Quattro (Genevac, UK). Following centrifugation, the samples were dried using the SP scientific VirTis BenchTop Pro freeze dryer (SP scientific, US) and stored at -20°C until further analysis. Purified compounds were dissolved in acetonitrile and analysed using UHPLCESI-Q-TOF-MS as previously described.

### 2.3.8. Minimum inhibitory concentration (MIC) assay

The MIC assay as previously described by Rushton *et al.* (2013) was used to investigate the antagonistic activity of enacyloxin IVa against the susceptibility organisms used in the previously described overlay assay and two medically relevant bacteria (Table 2.5.). Additionally, the concentration required to cause an 80% knockdown in optical density, compared to the control, was determined to compare the MIC of enacyloxin IVa to enacyloxin IIa. MICs were performed in a microbroth dilution assay using TSB broth, with doubling-dilutions of enacyloxin IVa between 100 and 0.098 µg/ml. Bacteria were grown for 18-24 hrs at 30°C and optical density measurements taken at 600 nm. Each microbroth dilution was performed in triplicate.

**Table 2.4. Antimicrobial susceptibility organisms: source of isolation and incubation temperature.**

| Species | Isolate ID | Source | Incubation temperature (°C) | Reference |
|---|---|---|---|---|
| *Staphylococcus aureus* | NCTC 12981 | Human | 37 | (Mullins *et al.,* 2019) |
| *Pectobacterium carotovorum* | LMG 2464 | Several crop species | 30 | (Mullins *et al.,* 2019) |
| *Candida albicans* | SC 5314 | Human | 37 | (Mullins *et al.,* 2019) |

**Table 2.5. Panel of strains used for this project from Cardiff University collection.**

| Organism | Isolate ID | Source details | Used for | Reference |
|---|---|---|---|---|
| *B. ambifaria* | HI 2345 (J82); ATCC 51993; BCC0191 | Soil | Positive control *recA* and 16s rRNA gene PCR<br>Control strain with known activity in overlay assay | (Mao *et al.,* 1997; Mullins *et al.,* 2019) |
| *B. ambifaria* | AMMD; LMG $19182_T$; BCC0207 | Pea rhizosphere | Control strain with known activity in overlay assay<br>Known enacyloxin producer | (Mullins *et al.,* 2019) |
| *B. gladioli* | BCC0238 | Sputum | Control strain with known activity in overlay assay | (Song *et al.,* 2017) |
| *B. anthina* | LMG $20980_T$ | Soil | Strain with similar ID to strains being tested for bioactivity | (Vandamme *et al.,* 2002) |
| *B. ubonensis* | LMG $20358_T$ | Soil | Strain with similar ID to strains being tested for bioactivity | (Yabuuchi *et al.,* 2000) |
| *B. multivorans* | ATCC17616 | Human | Susceptibility organism for MIC assay | (Mahenthiralingam *et al.,* 2011) |
| *Acinetobacter baumanii* | OXA-23 clone 2 | Human | Susceptibility organism for MIC assay | (Mahenthiralingam *et al.,* 2011) |
| *Staphylococcus aureus* | NCTC 12981 | Human | Susceptibility organism for overlay and MIC assay | (Mullins *et al.,* 2019) |
| *Pectobacterium carotovorum* | LMG 2464 | Several crop species | Susceptibility organism for overlay and MIC assay | (Mullins *et al.,* 2019) |
| *Candida albicans* | SC 5314 | Human | Susceptibility organism for overlay and MIC assay | (Mullins *et al.,* 2019) |
| *B. ubonensis* | ATCC 31433 | Soil | To compare the metabolic profiles of *B. ubonensis* strains | (Loveridge *et al.,* 2017) |

**Figure 2.1. Flow chart: research progression of activity testing.**

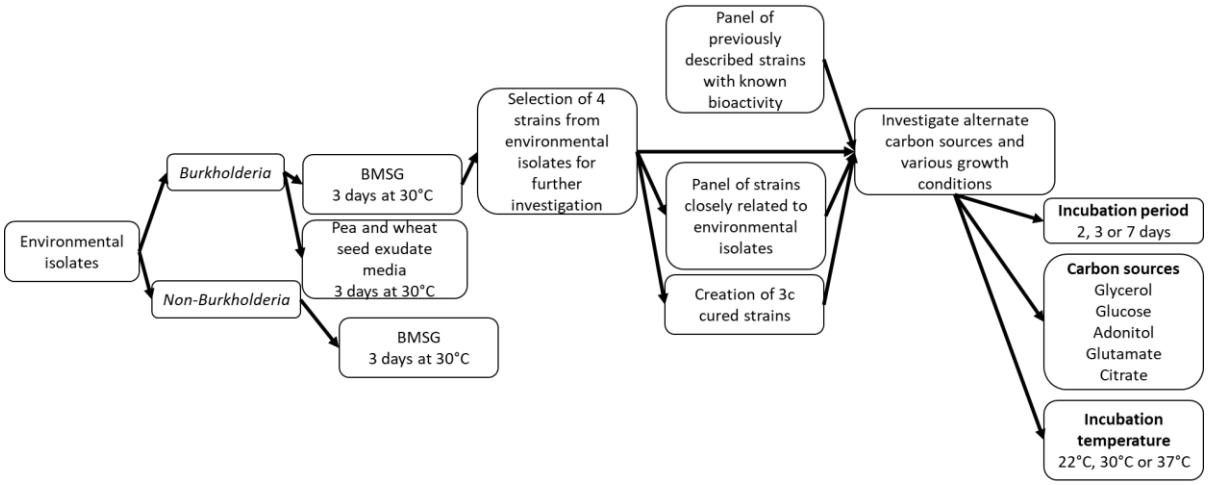



**Table 2.6. Carbon sources used for biomimetic media.**

| Carbon source | Supplier | Chemical formula |
|---|---|---|
| Glycerol | Sigma, US | $C_3H_8O_3$ |
| D-Glucose anhydrous | Formedium, UK | $C_6H_{12}O_6$ |
| Adonitol | Alfa Aesar, US | $C_5H_{12}O_5$ |
| tri-Sodium citrate | Fisher Scientific, US | $Na_3C_6H_5O_7$ |
| Monosodium glutamate | Sigma, US | $C_5H_8NNaO_4$ |

**Table 2.7. HPLC gradient programme – enacyloxin IVa purification**

| Time (mins) | % mobile phase A | % mobile phase B | Curve Number |
|---|---|---|---|
| initial | 95 | 5 | initial |
| 1 | 95 | 5 | 6 |
| 2 | 60 | 40 | 6 |
| 11 | 30 | 70 | 6 |
| 12 | 5 | 95 | 6 |
| 13 | 5 | 95 | 6 |
| 13.5 | 95 | 5 | 6 |
| 25 | 95 | 5 | 11 |

# 2.4. Whole genome sequencing for comparative genomic analysis and genome mining

### 2.4.1. Whole genome sequencing

The *B. cenocepacia* strains obtained from the Cardiff in-house collection and the Mysore *Burkholderia* collection were sequenced at the Cardiff University Genomics Research Hub and Novogene, Cambridge. The Illumina NextSeq500 Sequencer used at Cardiff university and the Illumina HiSeq 4000 at Novogene each generated 150 nucleotide paired-end reads. The Illumina platform was used to sequence 34 environmental isolates and 32 *B. cenocepacia* strains and were later *de novo* assembled into contigs. The 32 *B. cenocepacia* strains from the Cardiff university in-house collection were chosen based on ST types to assemble a collection of genomes with different ST types, including several ST types which correspond to those of epidemic strains.

### 2.4.2. Cloud infrastructure for remote computing

Bioinformatics analysis was performed using the MCR-funded Cloud Infrastructure for Microbial Bioinformatics (CLIMB) (Connor *et al.,* 2016). The CLIMB system is accessed via the internet through an SSH connection. Using the MobaXterm terminal commands are fed to a preconfigured virtual machine with software and a set of predefined pipelines or tools, including the Genomics Virtual Laboratory. Bash scripts were parallelised using the shell tool GNU parallel (Tange, 2018). Bioinformatics analyses that required scripts downloaded from GitHub are indicated as necessary.

### 2.4.3. Short read-(Illumina)-based assembly

Illumina adaptors were trimmed from 150-nucleotide paired-end reads and the read quality assessed using the wrapper script Trim Galore v0.6.5 (Krueger, 2016). Trim Galore combines Cutadapt (v2.8) (Martin, 2011) and FastQC (v0.11.9) (Andrews, 2009) to automate quality and adapter trimming as well as quality control. Genomes were assembled using Unicycler (v0.4.7) (Wick *et al.,* 2017) pipeline which assembles genomes using SPAdes (v3.13.0) (Bankevich *et al.,* 2012), and polishes contigs with Pilon (v1.22) (Walker *et al.*, 2014). The contigs were screened for contamination using Kraken2 (v2.0.8-beta) and the Minikraken database (Wood and Salzberg, 2014). Krona charts were generated using the ktImportTaxonomy function from KronaTools (Ondov *et al.,* 2011) to establish if any contigs were classified outside the family *Burkholderiaceae*. Contigs that were classified outside the family *Burkholderiaceae* were discarded from the sequence data. Genome sequence quality was assessed using Quast (v5.0.2) (Gurevich *et al.,* 2013) and the N50 values, and further statistics have been reported (Table 4.1. and 4.3.).

### 2.4.4. Multi-locus sequence typing and MLST sequence based phylogenetic trees

The software application MLSTcheck was employed to determine the ST types of the Mysore *Burkholderia* collection and the *B. cenocepacia* strains. The MLST profile of each strain was searched against the *Burkholderia cepacia* complex MLST profile database on PubMLST (https://pubmlst.org/bcc/). ST types were determined by running get_sequence_type script with an additional output function that reproduced all MLST genes sequences from each isolate concatenated in FASTA format.

The nucleotide sequences of the alleles utilized in MLST, from each reference Bcc strain used in the *recA* gene phylogeny, was downloaded manually from the PubMLST database for multi locus sequence analysis (MLSA). Almost all strains used in the *recA* phylogeny had their MLST alleles submitted to PubMLST, except for *B. cenocepacia* (C6433, CEP511), *B. stabilis* (HI-2462) and *B. vietnamiensis* (ATCC 29424) which were therefore excluded from the analysis. The concatenated MLST genes from the Mysore collection were aligned with the concatenated MLST genes of the representative species of the Bcc. In addition, the Bcc MLST scheme can be applied to *B. gladioli* and the MLST genes from *B. gladioli* LMG 2216 were downloaded from PubMLST to use in the alignment. Sequences were aligned using ClustalX2 (Larkin *et al*., 2007), with 2773 base positions in the final dataset, using MEGA (v7.0.26) (Kumar *et al.,* 2016) and a phylogenetic tree was constructed using the neighbour-joining method with Jukes-Cantor algorithm. Bootstrap analysis was based on 500 replications and the tree was rooted using *B. gladioli* LMG2216.

### 2.4.5. Average nucleotide identity (ANI)

To confirm the species designation made with the *recA* gene and NCBI database, the Mysore collection was subjected to average nucleotide identity analysis (ANI). The bioinformatic tool fastANI (v1.3) (Jain, 2017) was used for k-mer based pairwise comparison of the genome assemblies from the Mysore collection against all the *Burkholderia* genomes in the European Nucleotide Database (http://www.ebi.ac.uk/ena). Mysore collection strains were named after the strain with which they shared the closest ANI. Strains which shared an ANI less than 95%, a threshold value previously suggested by Richter and Rossello-Mora (2009), with any of the strains in the ENA database were designated *Burkholderia* sp..

The Python script PyANI (v0.2.9) (Pritchard *et al*., 2016) was used to calculate alignment-based pairwise ANI values when strains from the *Burkholderia* Mysore collection were compared amongst each other. In addition, PyANI was also used to compare all *B. cenocepacia* strains, including the

Mysore collection *B. cenocepacia* strains, amongst each other. All nucleotide similarity values are reported to three significant figures.

### 2.4.6. Gene prediction, sequence annotation and pan-genomics

Non-scaffolded contigs were annotated using Prokka (v1.14.0) (Seemann, 2014) with the 'Bacteria' annotation mode (default). Prokka is a prokaryotic genome annotation pipeline that combines multiple scripts to generate an annotation file. The core, accessory and pan-genomes of the entire *Burkholderia* Mysore collection was predicted using Roary (v3.13.0) (Page *et al.*, 2015) and applied to the non-scaffolded contigs to capture the complete genetic diversity of each strain. Additionally, the core, accessory and pan-genomes of the *B. cenocepacia* strains including the *B. cenocepacia* from the Mysore collection were also predicted using Roary (v3.13.0) (Page *et al.*, 2015). A minimum percentage identity of 95% for BLASTP comparisons, and a 99% core gene threshold (all strains) was applied to each dataset. The script roary_plot.py (https://sanger-pathogens.github.io/Roary/) was used to reproduce a piechart of the pan genome, breaking down the core, soft core, shell and cloud.

### 2.4.7. Sequence-based phylogenetic trees

Core gene-alignments were generated using Roary (v3.13.0) (Page *et al.*, 2015) which incorporates MAFFT (Katoh and Standley, 2013) as the alignment algorithm. The script RAxML (v8.2.12) (Stamatakis, 2014) was employed to generate high-accuracy phylogenies using a maximum-likelihood with General Time Reversible (GTR) substitution and a GAMMA model of rate heterogeneity supported by 100 bootstraps. The RAxML executable was compiled with multi-thread functionality: raxmlHPC-PTHREADS. Phylogenies constructed with RAxML were visualised and edited with MEGA (v7.0.26) (Kumar *et al.*, 2016) and Powerpoint (v1912).

### 2.4.8. Identifying *B. cenocepacia* island (*cci*) genes

To investigate if the *cci* gene were present in the collection of 42 *B. cenocepacia* strains, the collection was screened with Abricate (v.0.9.8) (https://github.com/tseemann/abricate) against a database containing all the *cci* genes. Abricate is a bioinformatic tool that can screen contigs for resistance or virulence genes against a designated database. The *cci* was initially identified on *B. cenocepacia* strain J2315 as a 31.7-kb low-GC-content island that encoded 35 predicted coding sequences. The predicted coding sequences of the *cci* were designated numerically (240-275) with the prefix BCAM, with the fully annotated genome (GenBank Accession: AM747721) deposited in the GenBank database (https://www.ncbi.nlm.nih.gov/genbank/). The genes BCAM230-275 were downloaded in fasta format and assembled into a database. Finally, the 42 genomes were screened against the assembled database to determine gene coverage when the minimum nucleotide identity was set at ≥75%.

### 2.4.9. Genome mining of specialised metabolite BGCs

The bioinformatics tool AntiSMASH (v5.1.1) (Blin *et al.,* 2019) was used to detect and annotate secondary metabolite biosynthesis gene clusters. The genomes of strains BCC1970, BCC1973, BCC1977 and BCC1990 were analysed using the online version of antiSMASH (available at: https://antismash.secondarymetabolites.org/#!/start) with the default settings. The default settings included; KnownClusterBlast, compares predicted BGCs against the MIBiG database of characterised BGCs (Medema *et al.,* 2015), SubClusterBlast, identifies operons/sets of genes that encode enzymes involved in the biosynthesis of common precursors for specialised metabolites and ActiveSiteFinder which identifies conserved amino acid motifs of known biosynthetic enzymes (Blin *et al.,* 2019).

# Chapter 3 – Single gene sequence analysis of the Mysore collection and characterisation of *B. cenocepacia* isolates.

## 3.1. Introduction

The following chapter addresses the isolation and identification of bacteria from the surrounding soil and rhizosphere of medicinal plant in the Western Ghats, India. Additionally, a more in-depth analysis of the *B. cenocepacia* isolates following the identification of several *B. cenocepacia* IIIA lineage strains, a lineage which is rarely isolated outside of clinical infection, has been provided. The background to the *B. cenocepacia* species, the isolation methods and single gene sequence analysis used in this chapter are introduced as follows:

### 3.1.1. Isolating *Burkholderia* from the natural environment

The development of isolation media for the *Burkholderia* genus has thus far mainly focussed on species with clinical relevance. The isolation of clinically relevant bacteria such as *Burkholderia cepacia* complex species, *B. pseudomallei* and *B. mallei* has been extensively covered in the literature (Henry *et al*., 1999; Peacock *et al.,* 2005; Glass *et al.,* 2009). Additionally, the isolation of plant pathogenic and plant beneficial species has also received considerable interest, in particular those with crop management applications (Suárez-Moreno *et al.* 2011; Mullins *et al.* 2019). The development of bioinformatic tools and the ever growing publicly available genome databases offer new possibilities for the design of selective isolation media. For example, Haeckle *et al.* (2018) reported using bioinformatic tools to identify suitable media additives for the development of *Burkholderia* selective isolation media. Nevertheless, several isolation media for the *Burkholderia* genus already exists, such as the semi-selective growth medium PCAT which has been reported to successfully recover *Burkholderia* species from soil (Pallud *et al.,* 2001; Vermis *et al.,* 2003; Salles *et al.,* 2006). Other media such as basal salt media with selective additives have also been reported to support the growth of *Burkholderia* species and can be used as enrichment or as semi-selective growth media (Zhang *et al.,* 2000; O'Sullivan *et al.,* 2007).

### 3.1.2. Single genetic markers for species identification

Successful isolation of bacteria from soil samples is often followed up by molecular identification. The 16S rRNA gene has been widely adopted for the classification of bacterial species. However, distinguishing between different members of the Bcc, using 16S rRNA gene sequence analysis, can be problematic due to the low sequence variation in this group (Mahenthiralingam *et al.,* 2000). To improve species delineation of the Bcc, genetic markers such as the housekeeping *recA* gene have

been adopted for phylogenetic analysis (Mahenthiralingam *et al.,* 2000; Payne *et al.,* 2005). Species of the Bcc share 94 to 95% sequence similarity in their *recA* gene and 98 to 100% sequence similarity in their 16S rRNA gene (Suárez-Moreno *et al.,* 2011). The high 16S rRNA gene sequence similarity between members of the Bcc highlights the limited utility of 16S rRNA gene sequence analysis for Bcc species differentiation. Thus, *recA* gene sequence analysis can discriminate between *Burkholderia cepacia* complex species with greater efficacy than 16S rRNA gene analysis (Mahenthiralingam *et al.,* 2000; Payne *et al.,* 2005; Drevinek *et al*. 2008). Nonetheless, the 16S rRNA gene is still useful at determining higher taxonomic rankings above species level (LiPuma, 1999).

### 3.1.3. Description of the *B. cenocepacia* species and *recA* lineages

Characterisation of *B. cenocepacia* using *recA* gene sequence analysis demonstrated this species could be split into distinct phylogenetic clusters based on polymorphisms in the *recA* gene (Mahenthiralingam *et al.,* 2000b; Vandamme *et al.,* 2003). All clinically isolated *B. cenocepacia* strains belonged to subgroups IIIA, IIIB, and IIID, whereas strains from subgroup IIIC have only been isolated from environmental sources in the original description of the species (Vandamme *et al.,* 2003). Development of the Bcc MLST scheme improved the resolution of these subgroups and offered further support for *B. cenocepacia* subgroups IIIA, IIIB, IIIC and IIID (Baldwin *et al.*, 2005). Limited information is available on the distribution of isolates within subgroup IIID, but subgroup IIIA strains have been almost exclusively isolated from clinical sources, whereas strains from subgroup IIIB have been isolated from both clinical and environmental sources (Baldwin *et al.,* 2007; Drevinek *et al.,* 2008).

*B. cenocepacia* infection is frequently associated with reduced survival and an increased risk of developing a form of necrotising pneumonia also known as *cepacia* syndrome (Drevinek and Mahenthiralingam, 2010). Additionally, *B. cenocepacia* is capable of patient-to-patient transmission and has been shown to replace other Bcc species in CF infections (Bernhardt *et al.,* 2003). Various epidemic strains such as the ET-12 and several other strains dominant in Canada and Europe (CZ1 Czech epidemic clone) belong to the *recA* IIIA lineage (Drevinek and Mahenthiralingam, 2010). Fortunately, the implementation of stringent infection control measures in CF clinics in the 90s helped control the spread of transmissible Bcc strains (Zlosnik *et al.,* 2015). However, individuals with CF can acquire *B. cenocepacia* infections from non-patient sources, and despite their widespread occurrence as virulent and transmissible CF pathogens, gaps in our knowledge remain about the acquisition of *B. cenocepacia* infections as an environmental source of *B. cenocepacia* IIIA strains has not been defined (Drevinek and Mahenthiralingam, 2010).

### 3.1.4. BCESM a clinical risk marker for *B. cenocepacia*

The *Burkholderia cepacia* epidemic strain marker (BCESM) is a genetic marker which was initially identified through RAPD analysis as an apparently conserved 1.4 kbp fragment among epidemic strains of *B. cenocepacia.* The BCESM fragment encodes a putative negative transcriptional regulator aptly named the epidemic strain marker regulator (*esmR*), which was discovered to be located on a genomic island (*cci)* with genes linked to both virulence and metabolism (Baldwin *et al.* 2004). Mahenthiralingam *et al.* (1996) designed the PCR primers BCESM1 and BCESM2 to target this marker in order to identify potentially virulent strains of *B. cenocepacia*. Although the BCESM has not been associated with all *B. cenocepacia* strains that can infect multiple patients, it has been widely adopted in infection control as a clinical risk marker for strains capable of patient-to-patient spread (Baldwin *et al.,* 2004).

### 3.1.5. Aims and objectives

This chapter aims to establish the diversity of *Burkholderia* associated with medicinal plant rhizospheres and the natural environment in the Western Ghats, India by addressing the following objectives:

1. Assemble a collection of *Burkholderia* bacteria by isolating and purifying bacteria using semi-selective media.
2. Extract DNA from bacteria isolates for PCR amplification of the 16S rRNA and *recA* genes for gene sequencing.
3. Perform phylogenetic analysis and determine the evolutionary relationship of the environmental *Burkholderia* isolates to previously characterised strains.

# 3.2. Results

### 3.2.1. Species identifications with the *recA* and 16S rRNA gene sequences

A total of 66 samples were collected from the rhizosphere and surrounding soil of 10 different plants located across 6 different sites in the Karnataka region, India (Table 3.1.). From the 66 samples, 73 pure isolates were recovered from 54 of the samples using growth agar enrichments. Using *recA* and 16S rRNA gene sequence analysis preliminary identifications were made grouping 35 *Burkholderia* isolates within the Bcc (Table 3.1.). Furthermore, *recA* gene sequence analysis identified 10 isolates to be closely related to *Burkholderia cenocepacia* with 7 isolates being closely related to the *recA* IIIA lineage. Searching the NCBI "Nucleotide collection (nr/nt)" database using BLASTN other *Burkholderia* isolates were identified as *Burkholderia cepacia* (n=9), *Burkholderia arboris* (n= 9), *Burkholderia ubonensis* (n= 2), *Burkholderia anthina* (n= 1), *Burkholderia pyrrocinia* (n= 1) (Table 3.1.). The *recA* sequences of 3 isolates shared homology with unnamed species within the NCBI database and were given the designation *Burkholderia* sp. Sequence lengths for the *recA* gene analysis ranged between 671-888 bp.

The universal primers 27F and 907R were used to amplify an 880-bp fragment spanning hypervariable regions 1 to 5 from the 16S rRNA gene (Yang *et al.,* 2016). The amplicon sequences of all 73 isolates were used to search the NCBI "16S ribosomal RNA sequences (Bacteria and Archaea)" database using BLASTN. Using this taxonomic marker, the *Burkholderia* isolates represented the following species *Burkholderia ambifaria* (n=1), *Burkholderia arboris* (n=2), *Burkholderia cepacia* (n=9), *Burkholderia lata* (n=18), *Burkholderia metallica* (n=1), *Burkholderia stagnalis* (n=1) and *Burkholderia ubonensis* (n=1) (Table 3.1.). Amplicon sequence lengths of the 16S rRNA gene used for analysis of the *Burkholderia* isolates ranged between 635-852 bp. The 38 non-*Burkholderia* isolates represented the following genera from 16S rRNA gene BLASTN alignment: *Chryseobacterium* (n=2), *Enterobacter* (n=5), *Pantoea* (n=1), *Pseudomonas* (n=11), *Serratia* (n=18) and *Yokenella* (n=1) with one isolate currently unidentified (Table 3.2.). Amplicon sequence lengths of the non-*Burkholderia* isolates ranged between 394-858 bp.

**Table 3.1. Preliminary identifications of the *Burkholderia* isolates using database alignments of the 16S rRNA and recA gene sequences.**

| BCC no. | ID | Source | Sampling location | *recA* BLASTN results | Identities | Identity | 16S BLASTN results | Identity |
|---|---|---|---|---|---|---|---|---|
| BCC1993 | 26 | Soil | Mercara (Location 1) | *Burkholderia anthina* | 718/727 | 98.76% | *Burkholderia arboris* | 99.52% |
| BCC1970 | 3 | *Garcinia gummi-gutta* | Pilikula | *Burkholderia arboris* | 795/812 | 97.91% | *Burkholderia lata* | 99.74% |
| BCC1972 | 5 | *Hydnocarpus pentandra* | Pilikula | *Burkholderia arboris* | 799/805 | 99.25% | *Burkholderia lata* | 99.63% |
| BCC1974 | 7 | *Pterocarpus marsupium* | Pilikula | *Burkholderia arboris* | 751/763 | 98.43% | *Burkholderia lata* | 99.30% |
| BCC1981 | 14 | *Pterocarpus marsupium* | Pilikula | *Burkholderia arboris* | 746/759 | 98.29% | *Burkholderia lata* | 99.28% |
| BCC1985 | 18 | *Mapia foetida* | Pilikula | *Burkholderia arboris* | 750/763 | 98.30% | *Burkholderia lata* | 99.76% |
| BCC1988 | 21 | *Salacia chinensis* | Pilikula | *Burkholderia arboris* | 792/805 | 98.39% | *Burkholderia lata* | 99.06% |
| BCC1995 | 28 | Soil | Mani | *Burkholderia arboris* | 710/721 | 98.47% | *Burkholderia lata* | 99.49% |
| BCC1996 | 29 | Soil | Mani | *Burkholderia arboris* | 760/774 | 98.19% | *Burkholderia lata* | 100% |
| BCC1998 | 31 | Soil | Mercara (Location 2) | *Burkholderia arboris* | 714/726 | 98.35% | *Burkholderia latens* | 99.49% |
| BCC1968 | 1 | *Anthocephalus kadamba* | Pilikula | *Burkholderia cenocepacia* | 816/818 | 99.76% | *Burkholderia cepacia* | 99.88% |
| BCC1969 | 2 | *Garcinia gummi-gutta* | Pilikula | *Burkholderia cenocepacia* | 802/802 | 100% | *Burkholderia cepacia* | 99.58% |
| BCC1971 | 4 | *Salacia chinensis* | Pilikula | *Burkholderia cenocepacia* | 801/801 | 100% | *Burkholderia cepacia* | 99.73% |
| BCC1976 | 9 | *Alstonia scholaris* | Pilikula | *Burkholderia cenocepacia* | 717/720 | 99.58% | *Burkholderia cepacia* | 99.72% |
| BCC1979 | 12 | *Calophyllum inophyllum* | Pilikula | *Burkholderia cenocepacia* | 726/729 | 99.59% | *Burkholderia cepacia* | 99.74% |
| BCC1983 | 16 | *Saraca asoca* | Pilikula | *Burkholderia cenocepacia* | 814/816 | 99.75% | *Burkholderia cepacia* | 99.50% |
| BCC1987 | 20 | Soil | Mani | *Burkholderia cenocepacia* | 793/803 | 98.75% | *Burkholderia lata* | 99.63% |
| BCC1989 | 22 | Soil | Mani | *Burkholderia cenocepacia* | 777/786 | 98.85% | *Burkholderia lata* | 99.23% |
| BCC1997 | 30 | Soil | Mercara (Location 2) | *Burkholderia cenocepacia* | 808/809 | 99.88% | *Burkholderia cepacia* | 99.64% |
| BCC2001 | 36 | Soil | Kukke Subramanya | *Burkholderia cenocepacia* | 804/809 | 99.38% | *Burkholderia lata* | 99.52% |
| BCC1977 | 10 | *Mapia foetida* | Pilikula | *Burkholderia cepacia* | 780/789 | 98.86% | *Burkholderia ambifaria* | 99.25% |
| BCC1978 | 11 | *Mapia foetida* | Pilikula | *Burkholderia cepacia* | 719/720 | 99.86% | *Burkholderia lata* | 99.37% |
| BCC1980 | 13 | *Pterocarpus marsupium* | Pilikula | *Burkholderia cepacia* | 710/713 | 99.58% | *Burkholderia cepacia* | 99.46% |
| BCC1982 | 15 | *Garcinia indica* | Pilikula | *Burkholderia cepacia* | 718/719 | 99.58% | *Burkholderia lata* | 99.64% |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| BCC1984 | 17 | *Pterocarpus marsupium* | Pilikula | *Burkholderia cepacia* | 710/713 | 99.58% | *Burkholderia cenocepacia* | 99.41% |
| BCC1991 | 24 | *Hydnocarpus pentandra* | Pilikula | *Burkholderia cepacia* | 797/801 | 99.50% | *Burkholderia lata* | 99.64% |
| BCC1994 | 27 | Soil | Mani | *Burkholderia cepacia* | 799/805 | 99.25% | *Burkholderia cepacia* | 99.46% |
| BCC2002 | 37 | Unknown | Unknown | *Burkholderia cepacia* | 803/806 | 99.63% | *Burkholderia lata* | 99.76% |
| BCC1986 | 19 | *Mapia foetida* | Pilikula | *Burkholderia cepacia* | 725/726 | 99.86% | *Burkholderia lata* | 99.29% |
| BCC1992 | 25 | Soil | Mercara (Location 1) | *Burkholderia pyrrocinia* | 701/713 | 98.32% | *Burkholderia metallica* | 99.30% |
| BCC1975 | 8 | *Pterocarpus marsupium* | Pilikula | *Burkholderia* sp. CATux-161 | 666/669 | 99.55% | *Burkholderia lata* | 99.50% |
| BCC2000 | 35 | Soil | Western Ghats | *Burkholderia* sp. KBC-3 | 797/803 | 99.25% | *Burkholderia arboris* | 98.82% |
| BCC1999 | 34 | Soil | Kukke Subramanya | *Burkholderia* sp. BCCUG10_1 | 802/804 | 99.75% | *Burkholderia lata* | 99.63% |
| BCC1973 | 6 | *Anthocephalus kadamba* | Pilikula | *Burkholderia ubonensis* | 786/787 | 99.87% | *Burkholderia ubonensis* | 99.01% |
| BCC1990 | 23 | Soil | Mani | *Burkholderia ubonensis* | 792/795 | 99.62% | *Burkholderia stagnalis* | 98.24% |

**Table 3.2. Preliminary identification of the non-*Burkholderia* isolates using the 16S rRNA gene sequence alignments.**

| Reference ID | BLASTN results | Query cover | Identity | Sequence length |
|---|---|---|---|---|
| 45 | *Chryseobacterium artocarpi* | 97.00% | 98.81% | 854 |
| 74 | *Chryseobacterium cucumeris* | 99.00% | 97.95% | 834 |
| 63 | *Enterobacter bugandensis* | 100.00% | 98.14% | 484 |
| 73 | *Enterobacter bugandensis* | 99.00% | 97.29% | 556 |
| 72 | Unidentified | Poor sequence | | 46 |
| 70 | *Enterobacter tabaci* | 100.00% | 97.72% | 525 |
| 56 | *Enterobacter tabaci* | 100.00% | 98.86% | 788 |
| 43 | *Pantoea dispersa* | 99.00% | 99.01% | 807 |
| 60 | *Pseudomonas entomophila* | 99.00% | 99.51% | 613 |
| 48 | *Pseudomonas monteilii* | 99.00% | 99.51% | 617 |
| 71 | *Pseudomonas monteilii* | 100.00% | 99.27% | 825 |
| 55 | *Pseudomonas nitroreducens* | 99.00% | 100.00% | 394 |
| 49 | *Pseudomonas nitroreducens* | 99.00% | 99.80% | 491 |
| 58 | *Pseudomonas nitroreducens* | 98.00% | 99.81% | 545 |
| 62 | *Pseudomonas nitroreducens* | 99.00% | 99.17% | 725 |
| 51 | *Pseudomonas nitroreducens* | 99.00% | 98.76% | 811 |
| 41 | *Pseudomonas nitroreducens* | 99.00% | 98.80% | 834 |
| 57 | *Pseudomonas nitroreducens* | 99.00% | 98.60% | 858 |
| 47 | *Pseudomonas taiwanensis* | 99.00% | 99.38% | 812 |
| 53 | *Serratia marcescens* | 100.00% | 98.71% | 542 |
| 39 | *Serratia marcescens* | 100.00% | 99.00% | 600 |
| 61 | *Serratia marcescens* | 100.00% | 99.28% | 693 |
| 50 | *Serratia marcescens* | 99.00% | 99.44% | 718 |
| 38 | *Serratia marcescens* | 99.00% | 99.17% | 722 |
| 52 | *Serratia marcescens* | 99.00% | 99.31% | 729 |
| 64 | *Serratia marcescens* | 100.00% | 99.46% | 741 |
| 46 | *Serratia marcescens* | 100.00% | 99.20% | 752 |

| | | | | |
|---|---|---|---|---|
| 59 | *Serratia marcescens* | 100.00% | 99.14% | 815 |
| 44 | *Serratia marcescens* | 99.00% | 99.02% | 819 |
| 40 | *Serratia marcescens* | 99.00% | 99.64% | 846 |
| 42 | *Serratia marcescens* | 99.00% | 99.64% | 847 |
| 75 | *Serratia nematodiphila* | 99.00% | 99.55% | 447 |
| 67 | *Serratia nematodiphila* | 99.00% | 99.84% | 645 |
| 66 | *Serratia nematodiphila* | 100.00% | 98.99% | 691 |
| 65 | *Serratia nematodiphila* | 99.00% | 99.57% | 696 |
| 69 | *Serratia nematodiphila* | 99.00% | 99.59% | 736 |
| 68 | *Serratia nematodiphila* | 99.00% | 99.34% | 757 |
| 54 | *Yokenella regensburgei* | 99.00% | 99.87% | 779 |

### 3.2.2. Phylogenetic analysis of the Mysore *Burkholderia* collection

To evaluate the evolutionary relationship of the environmental isolates to members of the *B. cepacia* complex a phylogenetic tree was constructed, utilising the *recA* gene sequence (Figure 3.1.). The BUR1 and BUR2 primers used to amplify the *recA* gene were designed to amplify an 869-bp fragment (Payne *et al.,* 2005). However, only 412-bp were used for constructing the phylogenetic tree due to the limited sequence availability of representative species of the Bcc strains in the GenBank database.

The *recA* gene phylogeny (Figure 3.1.) does not corroborate the preliminary identifications made with the *recA* gene BLAST alignments for several environmental isolates (Table 3.1.). Strain BCC1977, which shared close homology to *B. cepacia* in the NCBI database (Table 3.1.), was more closely related to *B. anthina* in the *recA* phylogeny (Figure 3.1.). Strains BCC1975 and BCC2000 which share close homology to uncharacterised species in the NCBI database (Table 3.1.) are closely related to *B. cepacia* in the *recA* gene phylogeny (Figure 3.1.).

The *recA* gene phylogeny (Figure 3.1.; red highlight) revealed a cluster of 8 closely related isolates (BCC1970, BCC1974, BCC1981, BCC1985, BCC1988, BCC1995, BCC1996, BCC1998). Preliminary identifications using the *recA* gene and BLASTN (Table 3.1.) identified the cluster of 8 isolates as *B. arboris,* however the isolates are positioned separate (Figure 3.1.) from the closest relatives identified by BLASTN indicating potentially novel Bcc species groupings. Similarly, isolate BCC1993 phylogenetically clustered separate from *B. anthina* (Figure 3.1.) which was its closest relative by *recA* gene BLASTN alignment (Table 3.1.). Strains BCC1975 and BCC2000 shared their closest identity match with unnamed *Burkholderia* strains following *recA* gene BLASTN alignment. However, their closest identity matches, following 16S rRNA gene BLASTN alignment, were with *B. arboris* (BCC2000) and *B. lata* (BCC1975) confirming their identification as Bcc species. Additionally, their position in the clade formed by the *B. cepacia* type strain ATCC 25416$^{T}$ suggests they are *B. cepacia* strains (Figure 3.1.; yellow highlight). A cluster of 10 isolates (Figure 3.1.; yellow highlight), belong to the same clade as the *B. cepacia* strains used for this phylogenetic analysis.

Strain BCC1999 appeared to share a common ancestor with the *B. cenocepacia* IIIA lineage (Figure 3.1.), however 16S rRNA gene sequence analysis suggests strain BCC1999 to be more closely related to *B. arboris* (Table 3.1.). To provide further resolution for the *B. cenocepacia* isolates, and strain BCC1999, *recA* sequences were aligned with partial *recA* gene sequences from characterized *B. cenocepacia* strains obtained from the GenBank database (Figure 3.2.). High bootstrap value (71%) at the node observed in separating the *recA* IIIA subgroup and strain BCC1999 further suggested strain BCC1999 may not be *B. cenocepacia* (Figure 3.2.).

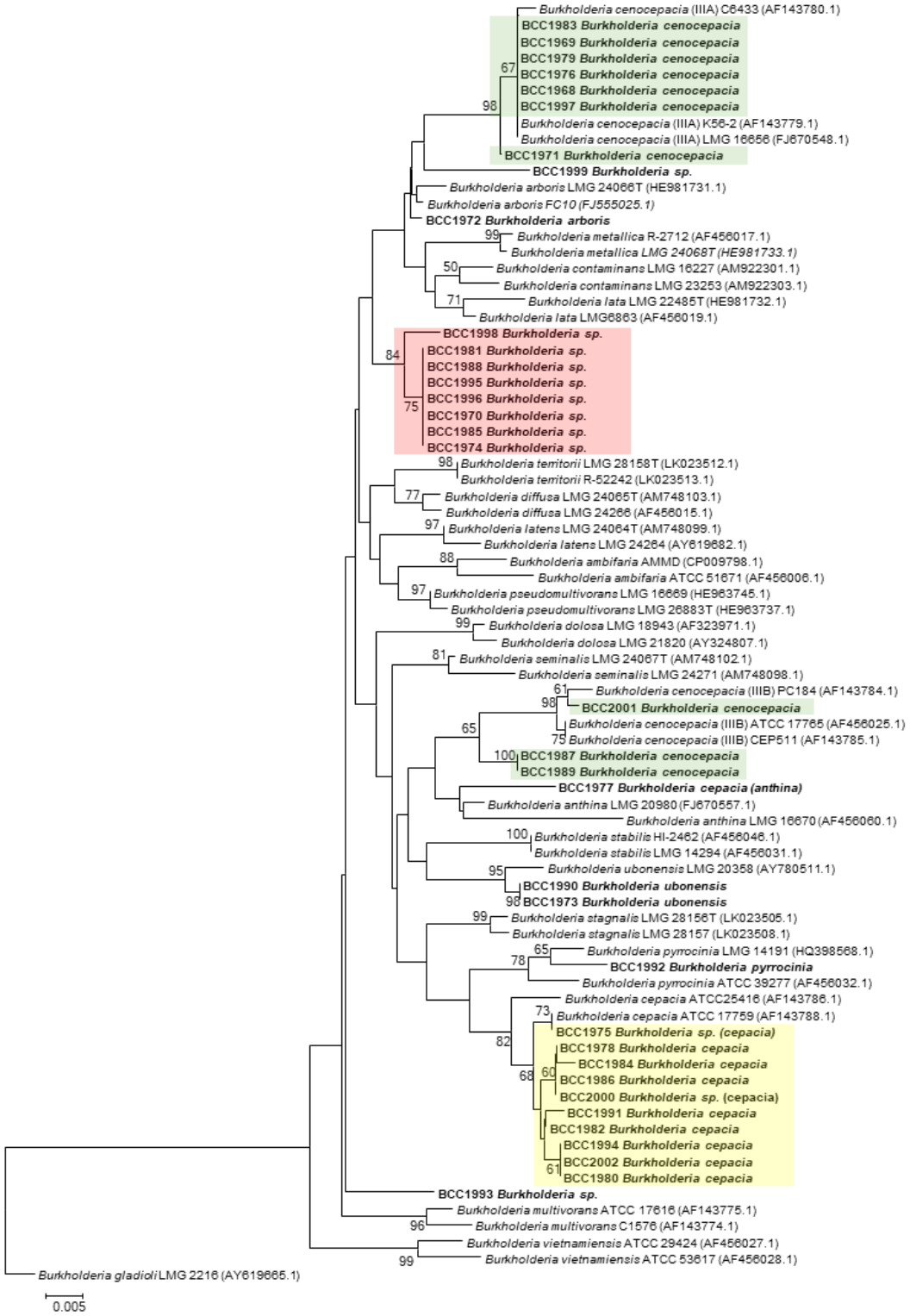



**Figure 3.1. Phylogenetic tree showing the *recA* gene diversity of environmental *Burkholderia* isolates.**

Accession numbers and strain IDs are listed after the species name and environmental isolates are in bold lettering. Alternative names based on taxonomic position are in brackets. Potentially novel environmental isolates are labelled *Burkholderia sp.*. Bootstrap values over 50% and genetic distance scale are indicated. All positions containing gaps and missing data were eliminated. There was a total of 412 positions used in the final dataset. The tree was rooted with *B. gladioli* as representative species outside of the *Burkholderia cepacia* complex.

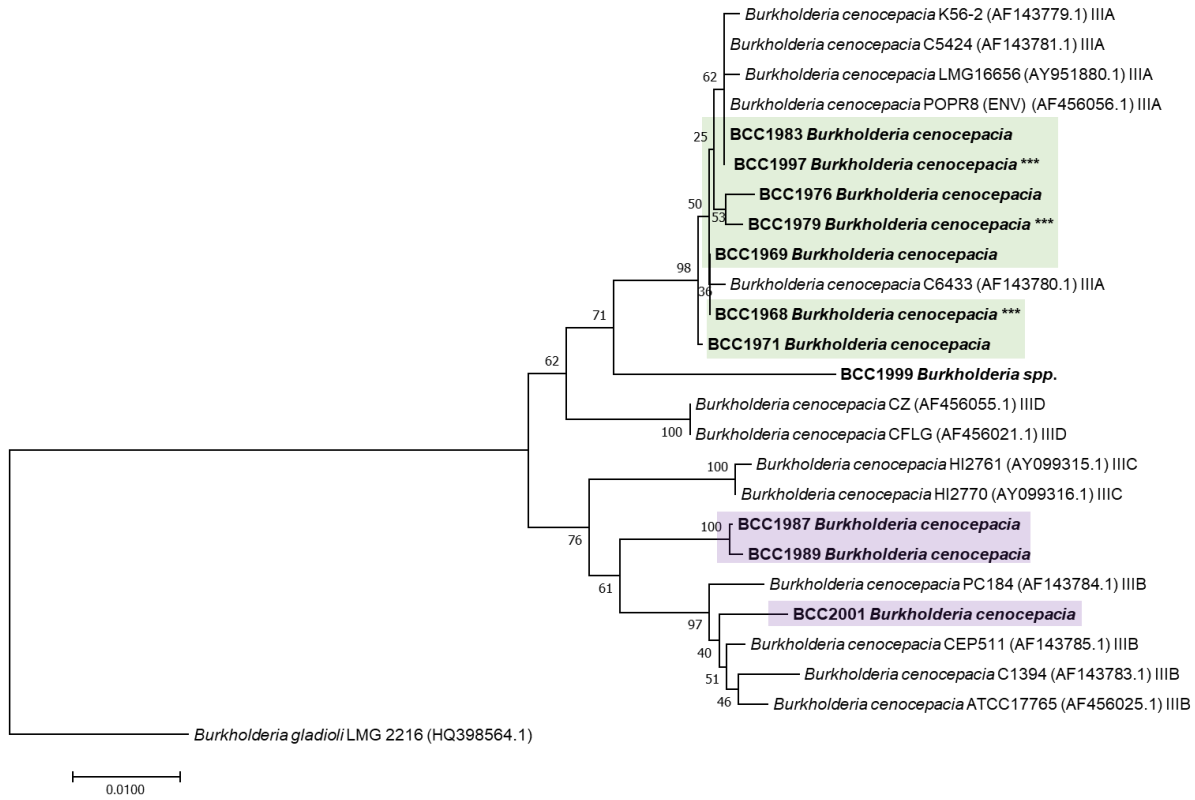



**Figure 3.2. Phylogenetic tree showing the phylogenetic relationship of environmental isolates to the *B. cenocepacia* subgroups.**

Strain IDs, accession numbers and subgroup lineage are listed after the species name and environmental isolates are in bold lettering. Strains originally isolated from environmental samples as opposed to clinical samples are indicated with ENV in brackets. Genetic distance scale is indicated. All positions containing gaps and missing data were eliminated. There was a total of 671 positions used in the final dataset. Strains which were BCESM positive are indicated with asterisks.

### 3.2.3. Phylogenetic analysis of characterized *B. cenocepacia* strains and the Mysore *B. cenocepacia* collection

From the *recA* gene sequence alignment to the NCBI database, 10 isolates (Table 3.3) were found to be closely related to *Burkholderia cenocepacia.* Phylogenetic analysis identified 7 isolates to clade within the *recA* IIIA lineage (Figure 3.2; green highlight) and a further 3 isolates were observed to clade within the *recA* IIIB lineage (Figure 3.2; purple highlight). Two isolates belonging to the *recA* IIIB group were derived from Mani, Karnataka and 1 isolate from Subramanya, Karnataka (Table 3.3). A total of 6 isolates from the *recA* IIIA group were isolated from samples collected in the Pilikula Biological Park, Mangalore, Karnataka and 1 isolate about 140 km south-east of Mangalore in Mercara, Karnataka, indicating a potentially localised environmental source of *B. cenocepacia* belonging to the *recA* group IIIA (Table 3.3). Interestingly, all 6 Pilikula isolates were derived from the rhizosphere of plants whereas the 4 remaining *B. cenocepacia* isolates came from soil samples (Table 3.3).

The *recA* phylogeny (Figure 3.2) revealed strains BCC1983 and BCC1997 to be closely related to the *B. cenocepacia* strains K56-2, C5424 and LMG 16656$^{T}$, which belong to the ET-12 epidemic lineage, a lineage which has spread epidemically within CF populations in Europe and Canada (Holden *et al*. 2008; Drevinek and Mahenthiralingam, 2010). Isolates BCC1976 and BCC1979 share a common ancestor with the strains K56-2, C5424 and LMG16656$^{T}$ (Figure 3.2). Isolates BCC1968 and BCC1969 are closely related to strain C6433 (Figure 3.2), a representative of strains that were reported to have spread among CF patients in Vancouver (Mahenthiralingam *et al.,* 2000). Isolates BCC1987, BCC1989 and BCC2001 all grouped within the same clade as the *recA* IIIB lineage (Figure 3.2; purple highlight). Isolate BCC2001 shares a common ancestor with strains PC184 and CEP511. PC184 was one of the earliest strains reported to be capable of patient-to-patient spread and CEP511 was reported to have spread epidemically among CF patients in Australia (Mahenthiralingam *et al.*, 2000). All sequences obtained from GenBank, to construct the phylogenetic tree in Figure 3.2, are from clinical samples except for POPR8 which was isolated from an environmental source (Baldwin *et al.,* 2007).

**Table 3.3. NCBI database search results from the putative *Burkholderia cenocepacia* isolates using the *recA* gene sequences including sampling location and source.**

| Bcc no. | Isolate | Source | Sampling location | *recA* BLASTN | Genomovar[a] |
|---|---|---|---|---|---|
| BCC1968 | 1 | *Anthocephalus kadamba* | Pilikula | *B. cenocepacia* | IIIA |
| BCC1969 | 2 | *Garcinia gummi-gutta* | Pilikula | *B. cenocepacia* | IIIA |
| BCC1971 | 4 | *Salacia chinensis* | Pilikula | *B. cenocepacia* | IIIA |
| BCC1976 | 9 | *Alstonia scholaris* | Pilikula | *B. cenocepacia* | IIIA |
| BCC1979 | 12 | *Calophyllum inophyllum* | Pilikula | *B. cenocepacia* | IIIA |
| BCC1983 | 16 | *Saraca asoca* | Pilikula | *B. cenocepacia* | IIIA |
| BCC1987 | 20 | Soil | Mani | *B. cenocepacia* | IIIB |
| BCC1989 | 22 | Soil | Mani | *B. cenocepacia* | IIIB |
| BCC1997 | 30 | Soil | Mercara | *B. cenocepacia* | IIIA |
| BCC2001 | 36 | Soil | Kukke Subramanya | *B. cenocepacia* | IIIB |

[a] Based on the phylogenetic analysis in Figure 3.2

### 3.2.4. Investigating the presence of the *Burkholderia* epidemic strain marker (BCESM)

*B. cenocepacia recA* lineages IIIA and IIIB have been reported to be responsible for most infection cases (Vandamme *et al*, 2003). Strains belonging to the IIIA lineage are seemingly characterised by a higher transmissibility and mortality rate than strains belonging to subgroup IIIB (Mahenthiralingam *et al.,* 2001). Although, the *B. cenocepacia* isolates group within IIIA and IIIB, their phylogenetic relationship does not necessarily dictate their ability to cause infection. To further evaluate the isolates' virulence the isolates were investigated for the BCESM (Baldwin *et al.,* 2004). The BCESM has been used as an epidemiological marker for virulent *B. cenocepacia* strains which infect individuals with CF. Isolates BCC1968, BCC1979 and BCC1997 exhibited the right size amplicon when targeted using the BCESM primers (Figure 3.3.). The presence of the BCESM in strains BCC1968, BCC1979 and BCC1997 suggest those strains have a propensity for causing virulent infections in patients with CF.

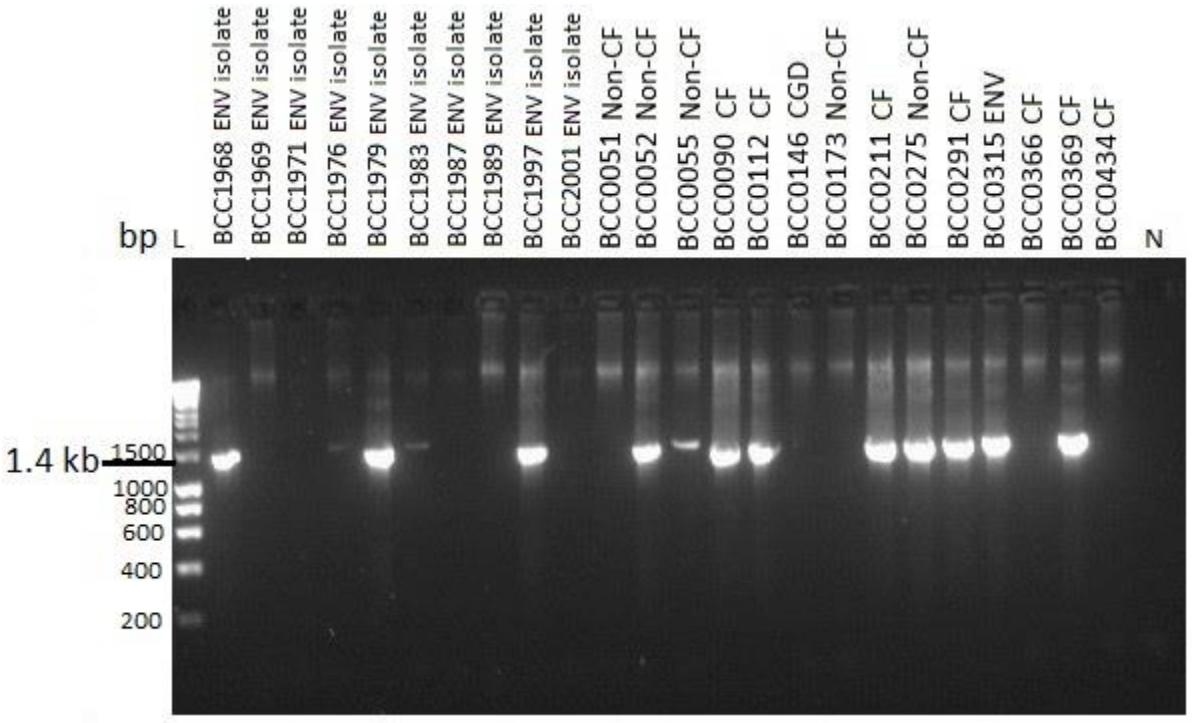



**Figure 3.3. PCR results of isolates and several known *B. cenocepacia* strains when targeted with the BCESM1/2 primers.**
Each well was loaded with 5 µL of sample. N represents the negative control which is the reagent mixture without DNA. All environmental isolates are indicated by ENV isolate. Strains isolated from cystic fibrosis patients are labelled CF and strains isolated from patients with chronic granulomatous disease are labelled CGD. All strains from the Cardiff university in-house collection which tested positive for the BCESM have previously been confirmed BCESM positive, correspondingly those which tested negative had also previously been confirmed BCESM negative. The size of fragments was determined using the 1kb HyperLadder™ from Bioline, CA (lane L). The expected BCESM amplicon size of 1.4-kb is indicated.

## 3.3. Discussion

The following section discusses the isolation methods and species identifications of the 35 *Burkholderia* strains recovered from rhizosphere and soil samples provided by collaborators at Mysore university, Mysore, India. Species identifications were made by obtaining gene sequences of the 16S rRNA and *recA* genes and aligning them to the NCBI database using BLAST. All *Burkholderia* strains were determined to be closely related to the Bcc and phylogenetic analysis using the *recA* gene was conducted to reveal their taxonomic relationship to characterized species of the Bcc. The identification of several *B. cenocepacia* strains which belong to the IIIA lineage, a lineage which has rarely been isolated outside of clinical infection (Baldwin *et al.*, 2007), prompted further investigation into the taxonomic relationship of the environmental *B. cenocepacia* to representative strains of all *B. cenocepacia* lineages. Finally, to assess if environmental *B. cenocepacia* isolates have a capacity for virulent human infection, they were investigated for a clinical risk masker (BCESM) linked to a pathogenicity island that has been shown to enhance virulence and transmissibility in *B. cenocepacia* strains (Baldwin *et al.,* 2004).

### 3.3.1. Isolating *Burkholderia* from the rhizosphere of medicinal plants and the natural environment

To obtain *Burkholderia* isolates the rhizosphere and soil samples were mixed with 10% w/v physiological saline solution and 0.1mL of the diluent was spread plated onto BMSG and PAB. BSMG was supplemented with polymyxin B to inhibit the growth of non-*Burkholderia* Gram-negative bacteria (Hareland *et al.*, 1975; Poirel *et al.*, 2017). Although most *Burkholderia* species have been reported to be intrinsically resistant to polymyxin B, those which are susceptible could have been excluded during isolation (Loutet *et al.,* 2011). PAB is a modification of King's A Medium to which magnesium chloride and potassium sulphate are added to enhance pigment production (King *et al.,* 1954). PAB allows the growth of a broad spectrum of *Pseudomonas* species and has been shown to support growth of *Burkholderia cepacia*. PAB was supplemented with cetrimide which inhibits growth of a broad range of Gram-positive and some Gram-negative bacteria, further improving its selectivity (Lowbury and Collins 1955). However, the applicability of PAB for isolating *Burkholderia* outside of the Bcc has not been well established and was therefore abandoned after initial isolation by the collaborators in Mysore, India, who had provided the sample set.

Colonies which looked visually distinct were swabbed and transported using charcoal medium and subsequently streaked onto TSA. Tryptic Soy Agar is a universal medium which supports the growth of wide variety of microorganisms including Gram-positive and Gram-negative bacteria. Plates were incubated at 30°C until visible growth was observed. Distinct colonies were used to inoculate BSMG

broth. BSMG's use as a minimal growth medium for *Burkholderia* has been well established (Zhang *et al.,* 2000; O'Sullivan *et al.,* 2007). Furthermore, the addition of glycerol can promote antibiotic production which would favour the growth of antibiotic producing bacteria (Keum, 2009). However, BSMG can facilitate the growth of other bacteria genera and overgrowth or inhibition of *Burkholderia* could obstruct their isolation (Hareland *et al.*, 1975; O'Sullivan *et al.,* 2007).

Bacteria grown on BSMG broth were used to inoculate semi-selective PCAT medium plates, supplemented with cycloheximide, to inhibit fungal growth, and polymyxin B, to further isolate *Burkholderia* species (Hareland *et al.,* 1975; Viallard *et al.,* 1998). PCAT has been reported to successfully isolate all members of the Bcc, however the selectivity of PCAT as a medium for the isolation of newly described *Burkholderia,* and the closely related genus *Paraburkholderia,* has not been well established (Pallud *et al.,* 2001; Salles *et al.,* 2006). Viallard *et al.* (1998) reported *Burkholderia vandii* to grow poorly on PCAT and the type strains of *Trinickia caryophylli* (formerly *Paraburkholderia carophylli*), *Burkholderia mallei* and *Burkholderia pseudomallei* to not grow on PCAT. These strains would likely have been excluded during sample screening.

Although the isolation methods used for the current study successfully recovered multiple *Burkholderia* strains, PCAT and BSM have limited utility for exploring the diversity of the entire *Burkholderia* genus within the soil and rhizosphere environment. Haeckle *et al.* (2018) reports combining computational and empirical data to design several selective isolation media more effective in the recovery of *Burkholderia* from rhizosphere samples than PCAT. To validate the efficacy of their design method, Haeckle *et al.* (2018) investigated 6 different media types in an environmental sample screening study with PCAT for comparison. Interestingly, 4 of the media types designed by Haeckle *et al.* (2018) contained $K_2HPO_4$ and yeast extract. Both $K_2HPO_4$ and yeast extract are used in BSM (Hareland *et al.*, 1975) and PCAT (Burbage and Sasser, 1982) recipes suggesting that the basal media components initially selected for these media were the most suitable for the isolation of *Burkholderia*.

The isolation methods used in the current study recovered 73 bacteria isolates including 35 *Burkholderia* strains (47.9%). The remaining bacteria isolates were mostly *Serratia* and *Pseudomonas* species and made up 24.7% and 15.1% of strains, respectively. Interestingly, a parallel study utilising essentially the same BSM-polymyxin enrichment broths isolated predominantly *Paraburkholderia* species from environmental samples from the Borneo jungle (A. Alswat, Cardiff University; unpublished data). Overall, advances in DNA sequencing has presented an opportunity for exploiting genomic information for rational media design (Haeckle *et al.,* 2018) and the efficacy of culture media designed through empirical methods should be re-evaluated when screening environmental samples for a genus or species. Thus, future culture dependent screening should employ media which can

support growth of a wider array of *Burkholderia* species, so that we can continue to investigate the biosynthetic potential of this relatively unexplored resource of natural products (Kunakom and Eustáquio; 2019). Nevertheless, the sampling method used for this study still provided a sufficiently large sample set for follow on investigations.

### 3.3.2. Identifying *Burkholderia* species using *recA* and 16S rRNA gene sequence analysis.

The universal primers 27F and 907R were selected to amplify an 880-bp fragment, which spans hypervariable regions 1 to 5 of the 16S rRNA gene (Lane, 1991; Kim *et al.,* 2011). Although longer amplicon sequences provide greater resolution for phylogenetic analysis, PCR success rate is correlated with amplicon fragment size, with shorter fragments reported to amplify more readily (Webster *et al.*, 2006). Furthermore, the primer set includes degenerate bases to account for variability in the primer target sequences across bacterial taxa improving coverage of the 16S rRNA gene (Table 2.1). The performance of the hypervariable regions of the 16S rRNA gene for phylogenetic analyses has been extensively covered in the literature, with hypervariable regions 1 to 5 reported to provide good phylogenetic resolution (Kim *et al.,* 2011; Yang *et al.,* 2016). Overall, the experience in our group is that the amplicon fragment generated by the 27F and 907R primer set can successfully identify *Burkholderia* and members of sub-taxa such as the Bcc.

Preliminary identifications using the *recA* and 16S rRNA genes identified 35 *Burkholderia* based on closest homology match and best coverage (Table 3.1.). The differences seen between the *recA* gene and the 16S rRNA gene BLAST results can be attributed to factors such as: sequence availability in the database and limitations in the degree with which each gene can be used to discriminate between closely related species. For example, Jin *et al.* (2020) reports identical 16S rRNA gene sequences for the type strains of *B. stabilis* ATCC BAA-67$^{T}$ and *B. pyrrocinia* DSM 10685$^{T}$. Ultimately, the combined strength of *recA* and 16S rRNA gene sequence analysis is greater than each gene individually and has been used here to demonstrate that all *Burkholderia* isolates belong to the Bcc (Table 3.1.).

The cluster of 8 isolates (BCC1970, BCC1974, BCC1981, BCC1985, BCC1988, BCC1995, BCC1996, BCC1998) identified as *B. arboris* using the *recA* gene alignment (Figure 3.1.; red highlight), are located separate from the closest relatives identified by BLASTN, indicating potentially novel Bcc species groupings. It is important to note that other species within the Bcc, such as *B. cenocepacia* and *B. cepacia,* have been observed to split into phylogenetic sub lineages based on *recA* gene sequence analysis (Payne *et al.* 2005). Other less well-characterized species from the Bcc might exhibit similar taxonomic divisions when investigated further i.e. if more *recA* sequences become available for those species. Future taxonomic studies of the Bcc should utilize whole genome sequence-based approaches

to species classification as this provides greater resolution than single gene sequence analysis for discriminating closely related bacteria (Jin *et al.,* 2020).

Single gene sequence analysis of the Mysore collection identified 35 isolates as Bcc species. The Bcc is a group of closely related bacteria species and single gene sequence analysis has limited utility in demarcating closely related species (Baldwin *et al.*, 2005; Richter and Rossello, 2009). MLSA and phylogenomic analysis improves resolution at the species level and should be utilised for further taxonomic classification, especially as several *Burkholderia* strains from the Mysore collection remain unclassified at the species level following *recA* and 16S rRNA gene sequence analysis (Table 3.1.). Additionally, *in silico* genome comparison can generate genomic similarity statistics, such as average nucleotide identity, to verify taxonomic identities and can determine whether strains in the *B. cepacia* cluster and the *Burkholderia* sp. cluster (Figure 3.1.) belong to the same species (Richter and Rossello, 2009).

### 3.3.3. Sequence analysis of the *recA* gene from *B. cenocepacia* isolates

Following *recA* gene sequence alignment with the NCBI database, 10 isolates (Table 3.3.) were found to be closely related to *Burkholderia cenocepacia*. *RecA* gene phylogeny (Figure 3.2.) positioned 7 isolates within the *recA* IIIA lineage and 3 isolates in the *recA* IIIB lineage (Mahenthiralingam, 2000). In the original description of the species, all clinically isolated *B. cenocepacia* strains belonged to subgroups IIIA, IIIB, and IIID, whereas strains from subgroup IIIC had only been isolated from environmental sources (Vandamme *et al.,* 2003). Limited information is available on the distribution of isolates within subgroup IIID, but subgroup IIIA strains has been almost exclusively isolated from clinical sources, whereas subgroup IIIB has been isolated from both clinical and environmental sources (Vandamme *et al.*,2003; Drevinek *et al.,* 2008). In addition to these publications, examination of the Mahenthiralingam group *Burkholderia* collection at Cardiff University which contains over 2000 isolates, shows that only one strain (POPR8, Figure 3.2.) has been identified as an environmental *B. cenocepacia* IIIA strain, and this radish isolated strain comes from a collection in the 1980s with limited provenance and available data. Hence, it is significant that 20% of the entire environmental *Burkholderia* isolates collection belonged to the *B. cenocepacia* IIIA lineage, since clear environmental sources for this lineage have not yet been defined.

All *recA* sequences obtained from GenBank, to construct the phylogenetic tree (Figure 3.2.), are from clinical samples except for POPR8 (Baldwin *et al.,* 2007). The dataset used to construct the phylogenetic tree (Figure 3.2.) includes only a small sample set of *recA* gene sequences compared to the overall diversity of the *B. cenocepacia* species and consisted of mostly sequences from known epidemic strains. *B. cenocepacia recA* subgroups IIIC and IIID were included in the phylogenetic tree

to show *recA* sequence analysis identified isolates sequences to be closely related to IIIA and IIIB lineages, as opposed to these other proposed lineages of the species. It's important to note that although *recA* sequence analysis subdivides *B. cenocepacia* into 4 separate subgroups (Mahenthiralingam *et al.,* 2003), limited genomic data is available on strains in subgroups IIIC and IIID (Jin *et al.,* 2020). Whole-genome sequence analysis would add further resolution to the differences between the *B. cenocepacia* subgroups and improve classification of the *B. cenocepacia* strains from the Mysore collection. Nevertheless, it was clear from the close phylogenetic locations that the *B. cenocepacia* Mysore collection strains grouped within the IIIA and IIIB subgroups based on *recA* gene sequence analysis (Figure 3.1. and 3.2.). Further investigation of the *B. cenocepacia* Mysore collection should involve comparison of their MLST profile to that of the PubMLST database. The PubMLST database has a curated collection of over 1,842 MLST profiles from 3,697 isolates to which MLST profiles can be compared (https://pubmlst.org/bcc/). Multi-locus sequence typing has replaced methods such as RFLP pattern matching for clinical epidemiology and sequence typing of epidemic strains has been successful in tracing the global distribution of problematic CF pathogens (Drevinik *et al.,*2010). Comparing MLST profiles to the PubMLST database can determine if the *B. cenocepacia* strains are related to the proposed *recA* lineage, and if they belong to known ST lineages, offering some indication of their virulence potential.

### 3.3.4. Investigating the presence of the *Burkholderia* epidemic strain marker (BCESM)

*B. cenocepacia* subgroups IIIA and IIIB have been reported to be responsible for most infection cases (Vandamme *et al*, 2003). Strains belonging to the IIIA lineage are characterised by higher transmissibility, disease severity and mortality rate than strains belonging to subgroup IIIB (Mahenthiralingam *et al.,* 2001; Jones *et al.,* 2004; Zlosnik *et al.,* 2015). However, the position of the *B. cenocepacia* environmental isolates within the IIIA and IIIB lineages (Figure 3.2.) does not necessarily determine their pathogenic potential (Pirone *et al.*; 2008).

To further evaluate the virulence potential of the *B. cenocepacia* strains, they were investigated for a clinical risk masker (BCESM) which resides on a pathogenicity island (*cci*) which encodes genes linked to enhanced virulence and transmissibility in *B. cenocepacia* strains (Baldwin *et al.,* 2004; Malott *et al.,* 2005). Epidemiological studies have shown that the BCESM is predominantly associated with IIIA lineage strains (LiPuma *et al.,* 2001; Baldwin *et al.*, 2004). Furthermore, in the original description of the *cci* by Baldwin *et al.* (2004) a complete version of the *cci* was predominantly found in *B. cenocepacia* IIIA lineage strains. Interestingly, the Mysore *B. cenocepacia* IIIA lineage strains BCC1968, BCC1979 and BCC1997 exhibited the right size amplicon (Figure 3.3.) when targeted using the BCESM primers indicating the presence of the BCESM and a capacity for virulent human infection. However,

the BCESM is not an absolute marker for the ability of *B. cenocepacia* strains to cause infection as IIIB lineage strains have been found to encode the marker (Baldwin *et al.*, 2004; Uehlinger *et al.,* 2009). Nevertheless, during the outbreaks of transmissible *B. cenocepacia* strains in the 1990s through to early 2000s, the BCESM was adopted as a clinical risk marker which could rapidly identify problematic strains in individuals with cystic fibrosis (Mahenthiralingam *et al.,* 1997; Baldwin *et al.,* 2004). In conclusion, 3 of the 10 *B. cenocepacia* isolates carry the BCESM and sit within the *recA* IIIA lineage, which suggests environmental strains from the Karnataka region have a high capacity for infecting human hosts. Further research is needed to determine if strains with the BCESM carry the entirety of the *cci,* and what the prevalence and role of the *cci* is in the natural environment.

## 3.4. Conclusions

The main conclusions following the isolation and phylogenetic analysis of bacteria recovered from medicinal plants in the Western Ghats India were as follows:

Using semi-selective media and basic bacterial enrichments 73 pure-bacteria isolates were recovered from samples collected by collaborators at Mysore university, Mysore, India. Sequence analysis of the *recA* and 16S rRNA genes identified 35 isolates as *Burkholderia,* and the remaining isolates as mostly *Pseudomonas* and *Serratia* species. The *recA* gene phylogeny of the environmental *Burkholderia* strains identified all *Burkholderia* strains as Bcc species, including several strains which represent putative novel species lineages. Single gene sequence analysis has limited utility in circumscribing closely related species, and because several strains remain unclassified at the species level, further analysis using MLSA, genomic ANI and phylogenomics should be carried out to resolve their classification. Additionally, the identification of 7 *B. cenocepacia* IIIA strains is the first time a major natural source of IIIA lineage strains has been identified. The detection of a clinical risk marker linked to a pathogenicity island (*cci*) in 3 *B. cenocepacia* IIIA lineage strains suggests that environmental IIIA lineage strains have a high capacity for virulent human infection. Future work should utilise whole genome-based analysis to determine the pathogenicity of the environmental *B. cenocepacia* by investigating their relationship to clinically problematic strains and establishing whether a complete *cci* is present.

# Chapter 4 – Comparative genomic analysis

## 4.1. Introduction

Following isolation and purification of bacteria from rhizosphere and soil samples, and preliminary identification of the bacteria by *recA* and 16S rRNA gene sequence analysis, several strains remained unclassified at the species level. The following chapter addresses whole genome-based analysis of the Mysore *Burkholderia* collection in order to enhance resolution at the species level and provide accurate identifications of the bacteria. A further focus is provided on the *B. cenocepacia* isolates due its clinical significance as a pathogen in individuals with CF. The background to the genomic analyses used in this chapter are introduced as follows:

### 4.1.1. Species identification within the *Burkholderia* genus

The MLST scheme developed by Baldwin *et al.* (2005), targeting seven house-keeping genes: *atpD*, *gltB*, *gyrB*, *recA*, *lepA*, *phaC*, and *trpB,* has been widely adopted for the identification of *B. cepacia* complex species (Drevinek, 2008). The low mutation rate in the MLST housekeeping genes and a large publicly available web database (https://pubmlst.org/bcc) has made this scheme useful for tracing the global epidemiology of Bcc bacteria (Waine *et al.,* 2007; Drevinek and Mahenthiralingam, 2010; Fila and Drevinek, 2017). Furthermore, the nucleotide sequences of the MLST profiles can be concatenated and aligned for phylogenetic analyses and used to distinguish Bcc species based on sequence divergence.

Genome based taxonomy has significantly altered the classification of *Burkholderia* species, improving resolution at the species level and thereby reducing incongruities observed for single gene and MLST genes analysis (MLSA) (Jin *et al.,* 2020). *In silico* whole genome sequence comparison has led to the development of several genome-wide similarity such as average nucleotide identity (ANI) (Chun and Rainey, 2014). ANI is a similarity index generated by calculating the average identity shared among 1 kbp sections of the genome which have >60% overall sequence identity over >70% of their length. ANI has been proposed as a whole genome sequence-based alternative to the classical DNA-DNA hybridisation (DDH) (Richter and Rosello, 2009). DDH percentages greater than or equal to 70% has been widely adopted as the standard for species delineation (Goris *et al.*, 2007). Richter and Rosello (2009) reports an ANI of 95% to correspond closely to a DDH of 70% and is a suitable alternative to DDH for the classification of bacterial species.

### 4.1.2. Aims and objectives

This chapter aims to identify the species from the Mysore *Burkholderia* collection through phylogenomic analyses and provides a focus on the *B. cenocepacia* strains from the collection and their relationship to characterised *B. cenocepacia* strains. The following objectives will be addressed in this chapter:

1. To establish the species and strain diversity of the Mysore *Burkholderia* collection using genomics.
2. Use genomics to establish the relationship of the Mysore *B. cenocepacia* collection to characterised *B. cenocepacia* strains and identify an environmental source of IIIA lineage *B. cenocepacia*.
3. Investigate the pathogenesis of Mysore *B. cenocepacia* collection by investigating the presence of a genomic island with genes linked to virulence and metabolism.

# 4.2. Results

### 4.2.1. Genome statistics of the environmental isolates collection

Of the 35 environmental *B. cepacia* complex species recovered from the Western Ghats, 34 strains were sequenced using the Illumina platform. Strain BCC2002 was removed from the genomic analysis as attempts to extract DNA suitable for sequencing were unsuccessful. Contigs were assembled from the 150 bp nucleotide paired-end sequences and the final contig assembly number and related N50 values demonstrated draft genomes of good quality (Table 4.1).

The genomes of 34 Mysore *Burkholderia* collection ranged between 41-190 total contigs, with a mean of 109 contigs (Table 4.1). The N50 values ranged from 101-608 kbp with a mean of 224 kbp for the entire collection (Table 4.1). The highest N50 value (608 kbp) from strain BCC1989 varied considerably when compared to all other sequenced genomes. The second highest N50 value was from strain BCC1975 with 413 kbp and the assembly had almost twice the number of contigs (n = 76) when compared to strain BCC1989 (n = 41) (Table 4.1). All genomes were of suitable quality for subsequent bioinformatics analysis.

### 4.2.2. Genome size and GC content of the environmental isolates collection

The total sequence length of the 34 Mysore *Burkholderia* collection ranged between 7.47-9.14 Mbp with an average sequence length of 8.03 Mbp. The largest genome sequences belonged to the *B. cepacia* isolates (Table 4.1) and ranged 8.21-9.14 Mbp with an average sequence length of 8.5 Mbp. The GC content of the entire collection ranged between 66.08-67.42% with an average GC content of 66.8%. *B. ubonensis s*train BCC1973 had the smallest genome with 7.47 Mbp but had the second highest GC content of the entire collection with 67.34%.

**Table 4.1. Mysore *Burkholderia* collection genomic statistics.**

| fastANI ID | Strain | Total length (Mbp) | GC% | Total Contig | No. contigs >1000 bp | N50 (bp) |
|---|---|---|---|---|---|---|
| *Burkholderia ubonensis* | BCC1973 | 7.47 | 67.34 | 94 | 55 | 230299 |
| *Burkholderia cenocepacia* | BCC1997 | 7.50 | 67.42 | 74 | 49 | 271063 |
| *Burkholderia* sp. | BCC1999 | 7.58 | 66.61 | 70 | 44 | 273401 |
| *Burkholderia cenocepacia* | BCC1971 | 7.67 | 67.26 | 94 | 58 | 197210 |
| *Burkholderia* sp. | BCC1974 | 7.68 | 66.89 | 152 | 89 | 133844 |
| *Burkholderia* sp. | BCC1981 | 7.71 | 66.88 | 130 | 77 | 162175 |
| *Burkholderia cenocepacia* | BCC1987 | 7.74 | 66.98 | 59 | 47 | 276828 |
| *Burkholderia cenocepacia* | BCC1969 | 7.74 | 67.17 | 87 | 54 | 277448 |
| *Burkholderia cenocepacia* | BCC1979 | 7.76 | 67.23 | 103 | 62 | 185537 |
| *Burkholderia* sp. | BCC1988 | 7.77 | 66.84 | 82 | 47 | 238847 |
| *Burkholderia cenocepacia* | BCC2001 | 7.78 | 66.67 | 81 | 51 | 259956 |
| *Burkholderia* sp. | BCC1977 | 7.79 | 66.83 | 171 | 82 | 132177 |
| *Burkholderia ubonensis* | BCC1990 | 7.82 | 67.00 | 104 | 59 | 212221 |
| *Burkholderia cenocepacia* | BCC1983 | 7.90 | 67.12 | 103 | 65 | 162830 |
| *Burkholderia cenocepacia* | BCC1968 | 7.91 | 67.18 | 115 | 68 | 195436 |
| *Burkholderia cenocepacia* | BCC1989 | 7.99 | 66.94 | 41 | 25 | 608302 |
| *Burkholderia* sp. | BCC1985 | 7.99 | 66.64 | 130 | 85 | 189238 |
| *Burkholderia cenocepacia* | BCC1976 | 8.00 | 67.06 | 128 | 73 | 156475 |
| *Burkholderia* sp. | BCC1993 | 8.01 | 66.75 | 69 | 37 | 348624 |
| *Burkholderia* sp. | BCC1972 | 8.03 | 66.61 | 88 | 51 | 212392 |
| *Burkholderia* sp. | BCC1998 | 8.04 | 66.65 | 101 | 48 | 255935 |
| *Burkholderia* sp. | BCC1970 | 8.11 | 66.63 | 190 | 104 | 100938 |
| *Burkholderia* sp. | BCC1996 | 8.18 | 66.59 | 99 | 56 | 238910 |
| *Burkholderia* sp. | BCC1995 | 8.18 | 66.59 | 103 | 58 | 237645 |
| *Burkholderia pyrrocinia* | BCC1992 | 8.19 | 66.08 | 141 | 80 | 182820 |
| *Burkholderia cepacia* | BCC1984 | 8.21 | 66.82 | 181 | 94 | 125522 |
| *Burkholderia cepacia* | BCC1986 | 8.29 | 66.80 | 181 | 99 | 117708 |
| *Burkholderia cepacia* | BCC1982 | 8.30 | 66.89 | 132 | 77 | 166142 |
| *Burkholderia cepacia* | BCC1980 | 8.32 | 66.85 | 129 | 71 | 176994 |
| *Burkholderia cepacia* | BCC1994 | 8.33 | 66.85 | 91 | 50 | 232754 |
| *Burkholderia cepacia* | BCC1978 | 8.35 | 66.80 | 110 | 68 | 171646 |
| *Burkholderia cepacia* | BCC2000 | 8.65 | 66.59 | 90 | 51 | 236907 |
| *Burkholderia cepacia* | BCC1991 | 8.75 | 66.43 | 119 | 61 | 220760 |
| *Burkholderia cepacia* | BCC1975 | 9.14 | 66.36 | 76 | 35 | 413287 |

### 4.2.3. Species identification using average nucleotide identity analysis

Utilising the fastANI tool (v1.3) (Jain, 2017) a k-mer based pairwise comparison of the assembled genomes to all the *Burkholderia* genomes in the European Nucleotide Archive (http://www.ebi.ac.uk/ena) was conducted. At the time of analysis, the ENA held 3,478 assembled *Burkholderia* genomes and using this dataset a 95% threshold value for species delineation was applied to the average nucleotide identity analysis.

A total of 10 strains were designated as *B. cenocepacia,* strains BCC1968, BCC1969, BCC1971, BCC1976, BCC1979, BCC1983 and BCC1997 shared ANI values ranging 98.8% to 99% with the *B. cenocepacia* type strain J2315. Strain BCC2001, identified as belonging to the *recA* IIIB lineage (Figure 3.1.) shared an ANI of 95.1% with J2315. Strains BCC1987 and BCC1989 which were also identified as belonging to the *B. cenocepacia recA* IIIB lineage (Figure 3.1.) shared an ANI of 94.5% with J2315. The 3 closest identity matches in the ENA database of the strains identified as *B. cenocepacia* were all *B. cenocepacia* strains, demonstrating limited variation in the species identity being called by the analysis. The *B. cenocepacia* strains, except for BCC1987 and BCC1989, shared ANI values over 98% with their 3 closest identity matches. Additionally, strains BCC1987 and BCC1989 shared an ANI of 94.9% with *B. cenocepacia* strain Tatl-371, the type strain for a new species proposed by Wallner *et al.* (2019). Strain BCC1987 shared ANI values with its 3 closest identity matches ranging between 95.4-96.6% and strain BCC1989 the same with 95.4-96.6% and were therefore considered *B. cenocepacia*-like and designated *B. cenocepacia* for the purposes of this study. The 9 strains identified as *B. cepacia* (Table 4.1) had ANI values ranging 97.3% to 98.5% with the *B. cepacia* type strain ATCC 25416. Strain BCC1975, initially identified as *Burkholderia* sp. based on *recA* gene sequence analysis (Figure 3.1) shared an ANI of 97.3% with the *Burkholderia cepacia* type strain and was designated *B. cepacia.* The two environmental isolates BCC1973 and BCC1990 were identified as *B. ubonensis* as they shared ANI values of 97.3% and 97.5% with the *B. ubonensis* reference strain MSMB22, respectively (Price *et al.,* 2017). Strain BCC1992 was designated *B. pyrrocinia* as it had an ANI of 95% with the *B. pyrrocinia* type strain DSM 10685 (Storms *et al.,* 2004).

The remaining 12 strains were all members of the *B. cepacia* complex, and designated *Burkholderia* sp.*,* because they occupied novel taxonomic groups, not matched to any named species. The strains were designated *Burkholderia* sp.*,* if ANI values ≥95% were with exclusively uncharacterised strains in the ENA database, or if the ANI fell below the 95% threshold with the closest identity matches. The closest identity match of the strain BCC1999 had an ANI value of 94.3% falling below the 95% threshold value proposed by Richter and Rosello-Mora for species delineation (2009). Strain BCC1972 had an ANI of 98.9% with an unnamed *Burkholderia* strain but dropped off to 93.9% with the second and

93.8% with the third closest identity match. Strain BCC1977, initially identified as *B. cepacia* based on *recA* gene sequence analysis (Table 3.1), was designated *Burkholderia* sp. as the 3 closest relatives were all *Burkholderia* sp. with ANI values ranging 94.9% to 95.1%. Additionally, strain BCC1977 shared an ANI of only 91.1% with the *B. cepacia* type strain ATCC 25416. The remaining 9 strains (BCCBCC1970, BCC1974, BCC1981, BCC1985, BCC1988, BCC1993, BCC1995, BCC1996, BCC1998) (Figure 4.1; red highlight) did not share ANI values ≥95% with any characterised strains in the ENA database, and were therefore designated *Burkholderia* sp.

### 4.2.4. Multi-locus sequence analysis and ANI analysis to reveal taxonomic clusters in the Mysore *Burkholderia* collection

To gain an initial understanding of the strain-level diversity of the 34 Mysore *Burkholderia* collection, their MLST profiles were compared to all the *Burkholderia cepacia* complex MLST profiles on PubMLST (https://pubmlst.org/bcc/). A total of 19 unique MLST profiles were identified including 5 strains with novel ST types and 29 strains with unknown ST types. Novel and unknown MLST profiles are defined as follows: if one of the alleles is not contained in the database, i.e. if an allele has at least one base pair difference compared to any other sequence already in the database, the MLST profile will be considered unknown. An MLST profile is considered novel if the combination of alleles of each MLST gene has never been seen before in the PubMLST database. For strains with novel or unknown sequence type (ST) (Table 4.2) the nearest ST match found in the PubMLST database has been reported.

The sequence analysis of the *recA* and 16S rRNA genes identified all Mysore *Burkholderia* collection to belong collectively to the *Burkholderia cepacia* complex. To gain a better understanding of the strain-level diversity of the Mysore *Burkholderia* collection, the MLST profile of the strains evaluated for the *recA* gene phylogenetic tree (Figure 3.1) were extracted from the PubMLST database. Strains *Burkholderia cenocepaci*a C6433, *Burkholderia cenocepacia* CEP511, *Burkholderia stabilis* HI-2462 and *Burkholderia vietnamiensis* ATCC 29424 were omitted from the analysis as no MLST profiles were available for those strains. The MLST genes for each representative strain of the Bcc were downloaded as a concatenated file and aligned with the MLST genes of the Mysore *Burkholderia* collection. The *recA* tree in Figure 3.1 was constructed from only 412 bp, whereas the MLSA phylogeny in Figure 4.1 was constructed from a total of 2773 bp, providing greater resolution.

MLSA (Figure 4.1) revealed the evolutionary clustering of environmental *B. cepacia* complex isolates. To investigate the interrelatedness of the different clusters of the Mysore *Burkholderia* collection revealed by the MLSA phylogeny, the Python script PyANI (v0.2.9) (Pritchard *et al*., 2016) was used to calculate alignment-based pairwise ANI values for each strain when compared against the entire

collection (Figure 4.2). Each cell represents the pairwise comparison between the genomes of the strains on the rows and columns. Genomes which share >95% of similarity in the aligned regions used in the comparison are coloured as red, and those with < 95% as blue. The dendrograms are single-linkage clustering trees generated from the matrix of pairwise coverage results.

The cluster of 9 novel *Burkholderia* sp. isolates formed their own clade based on MLSA (Figure 4.1; red highlight) and shared a sequence identity of 97.3-100% to one another. Additionally, the nearest ST match for each of the 9 isolates was ST-1410 (Table 4.2). The isolate BCC1977 (ST-343) appeared to be closely related to this novel *Burkholderia* sp. cluster of 9 isolates (Figure 4.1) sharing a sequence identity with strain BCC1993 of 95%, however with the remaining 8 strains of the cluster BCC1977 shared sequence identity of only 94.8-94.9%. Strains BCC1972 and BCC1999 did not share ANI values ≥95% with any of the strains in the environmental isolates collection (Figure 4.2). Hence, appeared to be novel, single isolate *Burkholderia* sp. at the time of analysis.

The *B. cenocepacia* isolates (BCC1968, BCC1969, BCC1971, BCC1976, BCC1979, BCC1983 and BCC1997) all clustered together in the *B. cenocepacia* IIIA clade with two representative strains of the *B. cenocepacia* IIIA *recA* lineage (Figure 3.1.; red highlight). The Mysore *B. cenocepacia* collection in this cluster possessed a 99.2-100% sequence identity to one another (Figure 4.1; green highlight). Strain BCC2001 clades together with the two *B. cenocepacia* IIIB *recA* lineage (Figure 4.1; purple highlight) and shared only 95.1-95.2% sequence identity to the cluster of Mysore collection strains which sit in the *B. cenocepacia* IIIA clade. BCC2001 (Figure 4.1; purple highlight) clustered adjacent to *B. cenocepacia* strain PC 184, a US Midwest epidemic strain, which is a well characterised member of the *recA* IIIB lineage (Mahenthiralingam *et al.,* 2000). Strains BCC1987 and BCC1989 are positioned in a clade together and share an ANI >99.9% to one another, with *B. arboris, B stabilis* and *B. pyrrocinia* as neighbouring species (Figure 4.1). Interestingly, strains BCC1987 and BCC1989 had clustered with other IIIB lineage strains when examined using the single *recA* gene phylogeny (Figure 3.1; green highlight) but were positioned outside of the *B. cenocepacia* clade in the MLSA phylogeny (4.1; purple highlight). BCC1987 and BCC1989 shared an ANI of 95% with BCC2001 but this dropped to 94.6% with the remaining Mysore *B. cenocepacia* collection. Given that both BCC1987 and BCC1989 shared an ANI of 95% with BCC2001 they were designated as *B. cenocepacia* for the purposes of this study.

The two *B. ubonensis* strains (BCC1973 and BCC1990) possessed a 99.3% nucleotide identity to one another and clustered together next to the *B. ubonensis* species type strain (Figure 4.1; blue highlight). The 9 Mysore collection *B. cepacia* evaluated in the genomic taxonomy analysis, possessed an ANI of 97.5-100% amongst one another (Figure 4.1; yellow highlight) and formed a clade with the

representative *B. cepacia* strains, including the *B. cepacia* (ATCC 25416) type strain (Figure 4.1; yellow highlight) (Yabuuchi *et al.*, 1992).

**Table 4.2. Identity of the Mysore *Burkholderia* collection and associated sequence type.**

| BCC code | fastANI[a] | ST[d] | ST category[b, c] |
|---|---|---|---|
| BCC1987 | *Burkholderia cenocepacia* | 1654~ | Unknown |
| BCC1989 | *Burkholderia cenocepacia* | 1654~ | Unknown |
| BCC1968 | *Burkholderia cenocepacia* | 259~ | Novel ST |
| BCC1976 | *Burkholderia cenocepacia* | 31~ | Novel ST |
| BCC1983 | *Burkholderia cenocepacia* | 31~ | Novel ST |
| BCC1969 | *Burkholderia cenocepacia* | 32~ | Unknown |
| BCC1979 | *Burkholderia cenocepacia* | 32~ | Novel ST |
| BCC1997 | *Burkholderia cenocepacia* | 32~ | Novel ST |
| BCC2001 | *Burkholderia cenocepacia* | 330~ | Unknown |
| BCC1971 | *Burkholderia cenocepacia* | 964~ | Unknown |
| BCC1984 | *Burkholderia cepacia* | 14~ | Unknown |
| BCC1982 | *Burkholderia cepacia* | 1439~ | Unknown |
| BCC2000 | *Burkholderia cepacia* | 1595~ | Unknown |
| BCC1978 | *Burkholderia cepacia* | 266~ | Unknown |
| BCC1986 | *Burkholderia cepacia* | 266~ | Unknown |
| BCC1975 | *Burkholderia cepacia* | 268~ | Unknown |
| BCC1980 | *Burkholderia cepacia* | 494~ | Unknown |
| BCC1994 | *Burkholderia cepacia* | 494~ | Unknown |
| BCC1991 | *Burkholderia cepacia* | 939~ | Unknown |
| BCC1992 | *Burkholderia pyrrocinia* | 707~ | Unknown |
| BCC1970 | *Burkholderia* sp. | 1410~ | Unknown |
| BCC1974 | *Burkholderia* sp. | 1410~ | Unknown |
| BCC1981 | *Burkholderia* sp. | 1410~ | Unknown |
| BCC1985 | *Burkholderia* sp. | 1410~ | Unknown |
| BCC1988 | *Burkholderia* sp. | 1410~ | Unknown |
| BCC1993 | *Burkholderia* sp. | 1410~ | Unknown |
| BCC1995 | *Burkholderia* sp. | 1410~ | Unknown |
| BCC1996 | *Burkholderia* sp. | 1410~ | Unknown |
| BCC1998 | *Burkholderia* sp. | 1410~ | Unknown |
| BCC1972 | *Burkholderia* sp. | 1413~ | Unknown |
| BCC1977 | *Burkholderia* sp. | 343~ | Unknown |
| BCC1999 | *Burkholderia* sp. | 870~ | Unknown |
| BCC1973 | *Burkholderia ubonensis* | 1189~ | Unknown |
| BCC1990 | *Burkholderia ubonensis* | 1189~ | Unknown |

[a] If an environmental isolate does not share an ANI ≥95% with any of type strains in the ENA database the strain was designated *Burkholderia* spp.

[b] If an allele has at least one base pair difference compared to any other sequence already in the PubMLST database the MLST profile is considered unknown.

[c] An MLST profile is considered novel if the combination of alleles of each MLST gene has never been seen before in the PubMLST database.

[d] The tilde indicates which strains differ by at least one polymorphism from the matched ST.

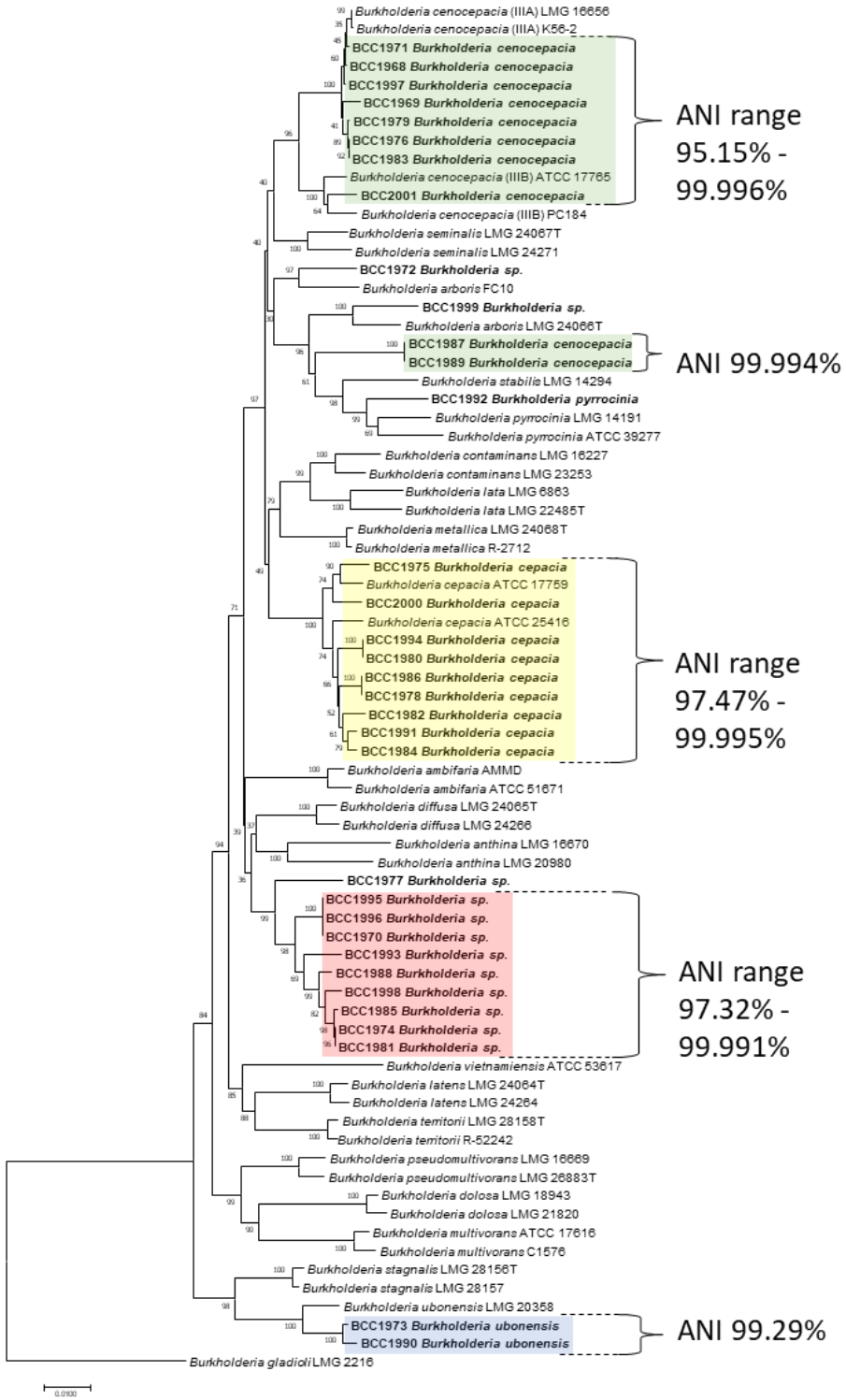



**Figure 4.1. Phylogenetic tree showing the MLST gene-based diversity of Mysore *Burkholderia* collection.**

Strains from the Mysore *Burkholderia* collection are in bold lettering with the in-house collection ID listed in front of the species ID. Representative strains of the *Burkholderia cepacia* complex have their strain ID listed after the species name. Bootstrap values over 50% and genetic distance scale (0.0100) are indicated. The concatenated nucleotide sequences (2773bp) from the seven housekeeping gene fragments *atpD* (443), *gltB* (400), *gyrB* (454), *recA* (393), *lepA* (397), *phaC* (385), *trpB* (301), were used to construct the phylogenetic tree. The tree was rooted with *B. gladioli* as representative species outside of the *Burkholderia cepacia* complex.

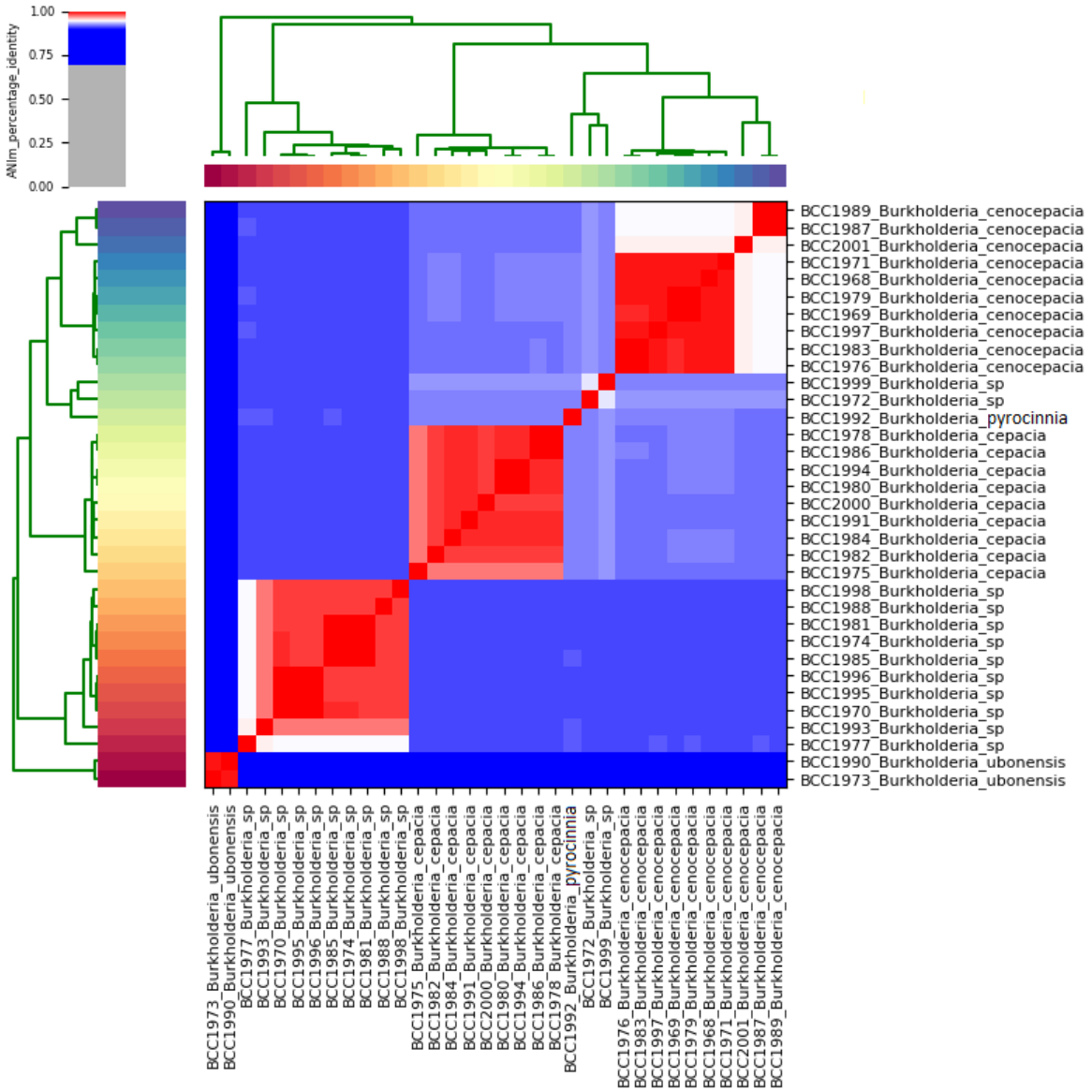



**Figure 4.2. Graphical representation of ANI analysis of the Mysore *Burkholderia* collection.**
The ANI values generated by the PyANI script were plotted as a heatmap to display the degree of nucleotide similarity between the 34 environmental *Burkholderia* isolates. The red indicates strains that possess >95% nucleotide similarity; the darker the red, the greater the similarity. The white highlights strains with <95% similarity. The heatmap was ordered by row and column dendrograms constructed based on the nucleotide similarity value.

### 4.2.5. Pan, core and accessory genome analysis of the environmental isolate collection

Following annotation of the 34 Mysore collection *Burkholderia* using the Prokka (v1.14.0) annotation pipeline (Seemann, 2014), the pan, core and accessory genome of the collection was determined by Roary-based analysis (v3.13.0) (Page *et al*., 2015). Following coding sequence prediction, a BLASTP sequence identity threshold of 95% was imposed to identify homologous proteins across species. Based on this threshold, the pan genome of the entire collection constituted 44,438 genes, 985 genes were identified across all 34 genomes and represented the core genome for the Mysore *Burkholderia* collection. The accessory genome consisted of 43,453 genes and most genes (31,341) were encoded by less than 15% of the 34 genomes (Figure 4.3). That is, for a given accessory gene out of the 31,341 genes, the gene was encoded by less than 15% of the genomes. The *B. ubonensis* strain BCC1973 was predicted to encode the smallest number of genes (6568 genes) while the *B. cepacia* strain BCC1975 was predicted to have the greatest coding sequence potential (8156 genes).

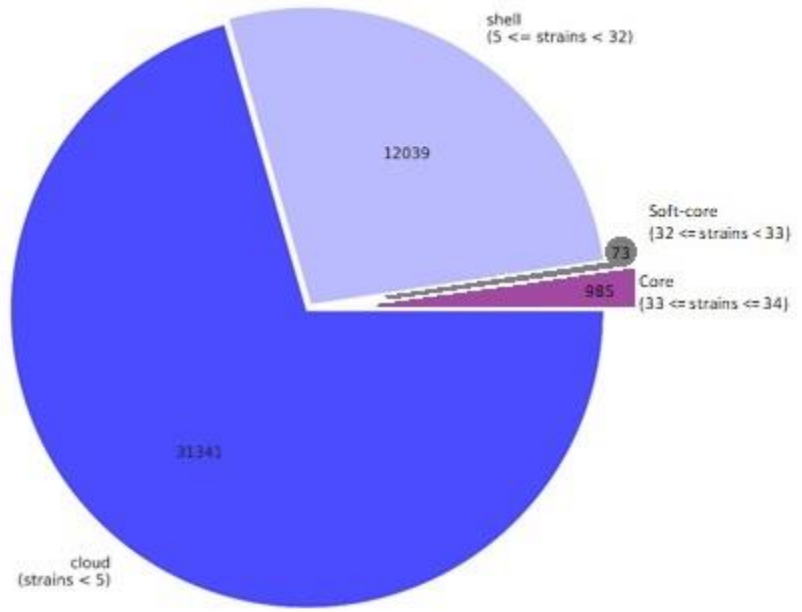



**Figure 4.3. The number of pan, core and accessory genes identified in the Mysore *Burkholderia* collection genomes.**
The program Roary was used to predict the pan, core, cloud, shell and soft-core genomes of the Mysore *Burkholderia* collection. The pan-genome is the total number of genes encoded by the entire collection which includes the cloud, shell, soft-core and core genes (44,438 genes). The core genome are all the genes encoded by each member of the entire collection (985 genes). The cloud, shell and soft-core genes constitute the accessory genome, which are genes not encoded by all the strain in the collection (43,453 genes).

### 4.2.6. Core genome phylogenomics of the Mysore *Burkholderia* collection

The MLSA phylogeny revealed distinct clustering of the Mysore *Burkholderia* collection (Figure 4.1). To determine the evolutionary relationship of the clustered isolates and strain to strain relatedness of the entire collection, the 985 core genes identified by Roary (Page *et al*., 2015) were aligned to construct a maximum likelihood phylogeny (Figure 4.4). The two *B. ubonensis* strains BCC1990 and BCC1973 formed their own clade and were the most distantly related to the entire collection and gave a natural outgroup of the phylogeny (data not shown) which was used to root the tree (Figure 4.4; blue highlight). The strains of *Burkholderia* sp., *B. cenocepacia* and *B. cepacia* which form clusters as observed in the MLSA phylogeny (Figure 4.1) all form distinct clades (Figure 4.4). Strain BCC1992, categorized as *B. pyrrocinia,* shared a common ancestor with the *B. cepacia* and *B. cenocepacia* clades (Figure 4.4). *Burkholderia* sp. strains BCC1972 and BCC1999 formed a clade together and shared a common ancestor with the *B. cenocepacia* strains (Figure 4.4). Strains BCC1987 and BCC1989 were in the same clade with a common ancestor as all the other Mysore collection *B. cenocepacia*, highlighting the evolutionary relationship of the isolates to other *B. cenocepacia* species (Figure 4.4).

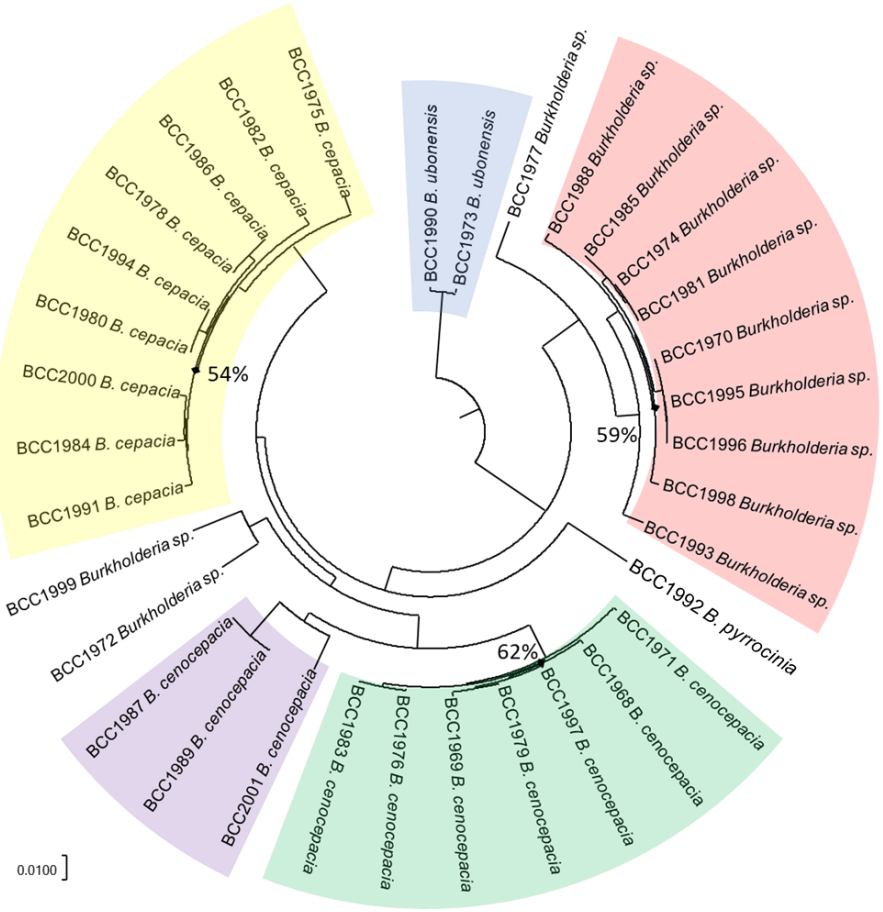



**Figure 4.4. Core-gene phylogeny of 34 *Burkholderia* genomes from the Mysore collection.**
A core-gene alignment of 985 genes was constructed with Roary, and the phylogeny created using RaxML (100 bootstraps). Nodes (3) with bootstrap values <70% are indicated with black diamonds and their relative bootstrap values. The *B. ubonensis, B. cenocepacia* and *B. cepacia* isolates form distinct clades. Scale bar represents the number of substitutions per base position.

### 4.2.7. Genome statistics of the *B. cenocepacia* strains

Of the 42 *Burkholderia cenocepacia* genomes sequenced, 10 were from the Mysore *Burkholderia* collection. The remaining 32 strains were taken from the Cardiff University in-house collection as well characterised representatives of environmental strains and epidemic and non-epidemic CF strains. All 42 genomes were sequenced on the Illumina platform and contigs were assembled from the 150 nucleotide paired-end reads.

The total number of contigs for the 42 *B. cenocepacia* strains ranged between 41-237 total contigs, with a mean average of 130 contigs (Table 4.3). The total number of contigs for the environmental isolates ranged between 41-128 total contigs with a mean of 89 contigs (Table 4.3; bold text). The N50 values ranged between 81 - 608 kbp with a mean of 185 kbp for the entire *B. cenocepacia* collection (Table 4.3). The N-50 value of the environmental isolates (Table 4.3; bold text) genomes ranged 156-608 kbp with a mean of 259 kbp. The highest N50 value (608 kbp) from strain BCC1989 varied considerably when compared to all other sequenced *B. cenocepacia* genomes. The second highest N50 value was from strain BCC1197 with only 285 kbp. All *B. cenocepacia* genomes were deemed suitable quality draft sequences to take forward for analysis.

### 4.2.8. Genome size and GC content of *B. cenocepacia* genomes

The total sequence length of the 42 *B. cenocepacia* strains ranged between 6.46-8.84 Mbp with a mean sequence length of 7.81 Mbp (Table 4.3). The total sequence length for the 10 Mysore collection *B. cenocepacia* ranged between 7.5-8 Mbp with a mean sequence length of 7.8 Mbp. The GC content of the entire *B. cenocepacia* collection ranged between 66.53-67.45% with mean GC content of 67.07%. The GC content of the 10 Mysore collection *B. cenocepacia* ranged between 66.67-67.42% with a mean GC content of 67.1% (Table 4.3).

**Table 4.3. *B. cenocepacia* genomic statistics.**

| | Strain | ST type[a,b] | GC% | Total contigs | No. contigs > 1000 bp | N50 (bp) | Total length (Mbp) |
|---|---|---|---|---|---|---|---|
| Mysore *B. cenocepacia* | BCC1968 | ST259[a] | 67.18 | 115 | 105 | 195436 | 7.91 |
| | BCC1969 | ST32[b] | 67.17 | 87 | 81 | 277448 | 7.74 |
| | BCC1971 | ST964[b] | 67.26 | 94 | 85 | 197210 | 7.67 |
| | BCC1976 | ST31[a] | 67.06 | 128 | 113 | 156475 | 8.00 |
| | BCC1979 | ST32[a] | 67.23 | 103 | 94 | 185537 | 7.76 |
| | BCC1983 | ST31[a] | 67.12 | 103 | 92 | 162830 | 7.90 |
| | BCC1987 | ST1654[b] | 66.98 | 59 | 53 | 276828 | 7.74 |
| | BCC1989 | ST1654[b] | 66.94 | 41 | 38 | 608302 | 7.99 |
| | BCC1997 | ST32[a] | 67.42 | 74 | 70 | 271063 | 7.50 |
| | BCC2001 | ST330[b] | 66.67 | 81 | 73 | 259956 | 7.78 |
| Cardiff University *B. cenocepacia* in-house collection | BCC0051 | ST242 | 67.24 | 171 | 142 | 105896 | 7.52 |
| | BCC0052 | ST234 | 67.30 | 103 | 93 | 186191 | 7.84 |
| | BCC0055 | ST358 | 67.31 | 52 | 49 | 277247 | 6.65 |
| | BCC0090 | ST214[a] | 66.88 | 206 | 162 | 139200 | 8.22 |
| | BCC0112 | ST209 | 67.07 | 155 | 136 | 155278 | 8.03 |
| | BCC0146 | ST241 | 66.86 | 168 | 132 | 178356 | 7.98 |
| | BCC0173 | ST219[a] | 67.15 | 155 | 112 | 176332 | 8.09 |
| | BCC0211 | ST32 | 67.27 | 237 | 208 | 97158 | 7.86 |
| | BCC0275 | ST216 | 67.32 | 99 | 92 | 182165 | 7.78 |
| | BCC0291 | ST807 | 67.21 | 93 | 87 | 152578 | 7.97 |
| | BCC0315 | ST38[b] | 66.57 | 143 | 118 | 151888 | 7.98 |
| | BCC0366 | ST208 | 67.19 | 154 | 136 | 127708 | 7.45 |
| | BCC0369 | ST40 | 66.95 | 93 | 85 | 175116 | 6.95 |
| | BCC0434 | ST309 | 67.30 | 117 | 81 | 240606 | 7.70 |
| | BCC0460 | ST201 | 67.45 | 107 | 93 | 185133 | 7.52 |
| | BCC0523 | ST230 | 67.12 | 122 | 96 | 186556 | 7.64 |
| | BCC0524 | ST234 | 66.54 | 142 | 127 | 156147 | 8.84 |
| | BCC0535 | ST33 | 67.08 | 152 | 134 | 139359 | 8.12 |
| | BCC0536 | ST210 | 67.02 | 135 | 125 | 140224 | 7.67 |
| | BCC0575 | ST234 | 67.09 | 173 | 158 | 115355 | 8.02 |
| | BCC0644 | ST234 | 66.53 | 138 | 123 | 160685 | 8.80 |
| | BCC0687 | ST218 | 67.45 | 68 | 59 | 239330 | 6.48 |
| | BCC0749 | ST280[b] | 67.34 | 97 | 89 | 171560 | 6.74 |
| | BCC0961 | ST32 | 67.16 | 235 | 210 | 80541 | 7.90 |
| | BCC1122 | ST32 | 67.14 | 206 | 169 | 112345 | 8.01 |
| | BCC1197 | ST306 | 67.22 | 53 | 51 | 284896 | 7.40 |
| | BCC1283 | ST250 | 67.08 | 88 | 86 | 182302 | 7.84 |
| | BCC1295 | ST241 | 66.57 | 168 | 134 | 164306 | 8.37 |
| | BCC1552 | ST28 | 67.02 | 139 | 123 | 168449 | 7.87 |
| | BCC1558 | ST241 | 66.58 | 225 | 192 | 121824 | 8.69 |
| | BCC1561 | ST28[a] | 67.08 | 171 | 149 | 141330 | 8.11 |
| | BCC1562 | ST258 | 66.94 | 223 | 190 | 93134 | 8.09 |

[a] Novel MLST profile

[b] Unknown MLST profile

### 4.2.9. ANI analysis of the *B. cenocepacia* genomes

Previous identifications made using *recA* gene sequencing and MLST of the *B. cenocepacia* strains from the Cardiff University in-house collection confirmed all strains as *B. cenocepacia*. Utilising the fastANI tool (v1.3) (Jain, 2017) a k-mer based pairwise comparison of the assembled genomes of the 10 Mysore collection *B. cenocepacia* to all the *Burkholderia* genomes in the European Nucleotide Archive (http://www.ebi.ac.uk/ena) was conducted. The 3 closest identity matches of all 10 isolates were *B. cenocepacia* genomes. The *B. cenocepacia* strains, except for BCC1987 and BCC1989, shared ANI values over 98% with their 3 closest identity matches. Strain BCC1987 shared ANI values with its 3 closest identity matches ranging between 95.4% to 96.6% and strain BCC1989 the same with 95.4% to 96.6%. Utilising the fastANI (v1.3) (Jain, 2017) tool the entire collection of 42 assembled *B. cenocepacia* genomes was compared against the *B. cenocepacia* type strain J2315. The ANI of all *recA* IIIA lineage strains when compared to the *B. cenocepacia* type strain J2315 ranged between 98.7% to 100%. Strains BCC1987 and BCC1989 had an ANI of 94.5% when compared against the *B. cenocepacia* type strain J2315, whereas the remaining *recA* IIIB lineage strains had an ANI of 95.1%.

To investigate the strain to strain ANI relatedness of the assembled genomes, all 42 *B. cenocepacia* genomes were compared. Using the PyANI (v0.2.9) (Pritchard *et al*., 2016) tool, alignment-based pairwise ANI values were generated (Figure 4.5). The 42 strains possessed an average nucleotide identity ranging between 94.5-100% to one another. When the strains (BCC1987, BCC1989, BCC2001, BCC0315 and BCC0369) previously identified as belonging to the IIIB *recA* lineage (Table 4.4) were removed from the analysis, the ANI of the remaining 37 strains ranged between 98.9% to 100%. The 5 IIIB lineage strains had ANI values ranging between 94.9% to 100% when compared to one another. Strains BCC1987 and BCC1989 had the lowest ANI value 94.9% when compared to the remaining 3 *B. cenocepacia* IIIB strains. However, strains BCC1987 and BCC1989 were most closely related to the strains BCC2001, BCC0315 and BCC0369 (ranging from 94.9% to 95%) when compared against the entire *B. cenocepacia* collection. Strain BCC2001 has its closest identity match (97.8%) with strains BCC0315 and BCC0369 when compared against the entire *B. cenocepacia* collection.

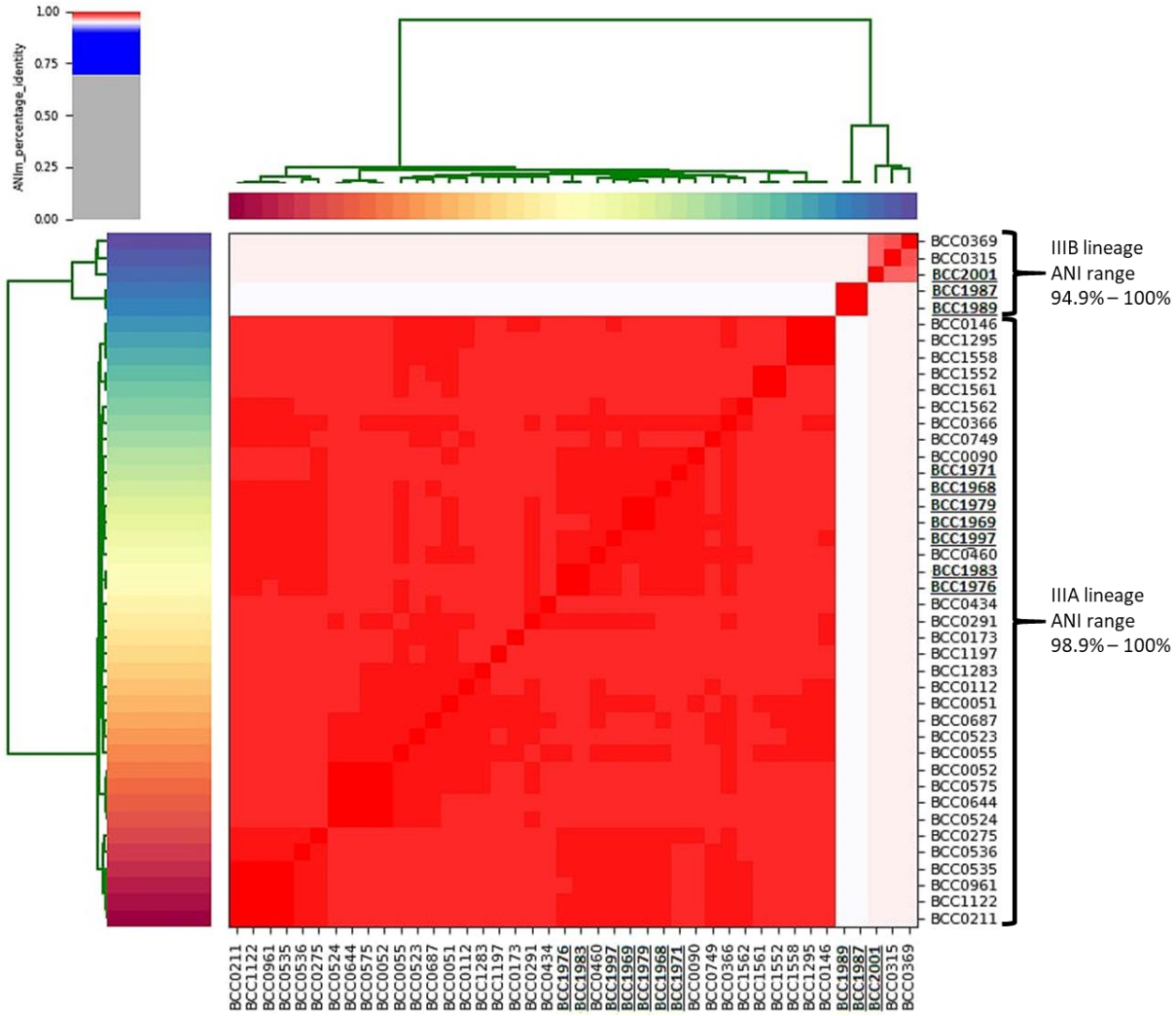



**Figure 4.5. Graphical representation of the *B. cenocepacia* genomes separating into 2 major ANI groups.**

Visual representation of ANI analysis generated by PyANI script for the *B. cenocepacia* strains. The heatmap indicates the degree of nucleotide similarity between the 42 *Burkholderia cenocepacia* strains. Each cell represents the pairwise comparison between the genomes of the strains on the rows and columns. The red highlights strains that possess >95% nucleotide similarity; the darker the red, the greater the similarity. The dendrograms are single-linkage clustering trees generated from the matrix of pairwise coverage results. Strains from the Mysore collection are underlined. Strains comprising the *recA* IIIA and IIIB lineages are shown by the brackets on the right, together with their numerical ANI relatedness.

### 4.2.10. Multi-locus sequence typing matches the Mysore *B. cenocepacia* collection to known epidemic *B. cenocepacia* strains

To gain an initial understanding of the strain-level diversity of the 42 *B. cenocepacia* isolates their MLST profiles were compared to all the Bcc MLST profiles on PubMLST (https://pubmlst.org/bcc/). A total of 32 unique MLST profiles were identified amongst the 42 *B. cenocepacia* strains. The MLST profiles of 37 strains were associated with the STs of *B. cenocepacia recA* IIIA lineage strains (Table 4.4) and 26 strains were identical with those of characterised STs, including the STs from epidemic and/or globally distributed lineages (ST28, ST234, ST241, ST32). Of the remaining 5 strains, the MLST profile of 3 strains were associated with IIIB lineage strains (BCC0315, BCC0369, BCC2001) and 2 strains (BCC1987, BCC1989) had MLST profiles associated with an uncharacterized ST (Table 4.4).

Among the collection several unknown and novel ST types were identified (Table 4.4). For strains with novel or unknown sequence type (ST) the nearest ST match found in the PubMLST database was reported (Table 4.4). For the Mysore *B. cenocepacia* collection the ST of 5 strains (BCC1969, BCC1971, BCC1987, BCC1989, BCC2001) were considered unknown, while the ST of 5 strains (BCC1979, BCC1997, BCC1983, BCC1968, BCC1976) were considered novel (Table 4.4). Strains BCC1976 (ST-31), BCC1983 (ST-31), BCC1997 (ST-32), BCC1971 (ST-964) and BCC1968 (ST-259) only differed in the *gyrB* allele. Strains BCC1979 and BCC1969, associated with ST-32, varied in the *gyrB* and *recA* alleles. Strain BCC2001 had polymorphisms in 5 out of the 7 alleles and strains BCC1987 and BCC1989 had unknown polymorphisms in every single allele, compared to their associated STs.

**Table 4.4. Details of isolation, strain type and genomic characteristics of the 42 *B. cenocepacia* strains.**

| | Strain | Predicted or closest ST type[a,] | Source of isolation[c, d] | Location | *recA* lineage[e] | BCESM[f] |
|---|---|---|---|---|---|---|
| Mysore *B. cenocepacia* | BCC1968 | ST259[a] | Rhizosphere | Pilikula, India | IIIA | + |
| | BCC1969 | ST32[b] | Rhizosphere | Pilikula, India | IIIA | - |
| | BCC1971 | ST964[b] | Rhizosphere | Pilikula, India | IIIA | - |
| | BCC1976 | ST31[a] | Rhizosphere | Pilikula, India | IIIA | + |
| | BCC1979 | ST32[a] | Rhizosphere | Pilikula, India | IIIA | + |
| | BCC1983 | ST31[a] | Rhizosphere | Pilikula, India | IIIA | + |
| | BCC1987 | ST1654[b] | Soil | Mani, India | Other Bcc | - |
| | BCC1989 | ST1654[b] | Soil | Mani, India | Other Bcc | - |
| | BCC1997 | ST32[a] | Soil | Mercara, India | IIIA | + |
| | BCC2001 | ST330[b] | Soil | Kukke Subramanya, | IIIB | - |
| Cardiff University *B. cenocepacia* in-house collection | BCC0051 | ST242 | NON-CF[c] | Edinburgh, UK | IIIA | |
| | BCC0052 | ST234 | NON-CF[c] | USA | IIIA | + |
| | BCC0055 | ST358 | NON-CF[c] | Strasbourg, France | IIIA | + |
| | BCC0090 | ST214[a] | Cystic fibrosis | BC, Canada | IIIA | + |
| | BCC0112 | ST209 | Cystic fibrosis | BC, Canada | IIIA | + |
| | BCC0146 | ST241 | CGD[d] | USA | IIIA | - |
| | BCC0173 | ST219[a] | NON-CF[c] | USA | IIIA | - |
| | BCC0211 | ST32 | Cystic fibrosis | Montreal, Canada | IIIA | + |
| | BCC0275 | ST216 | NON-CF[c] | Newark, USA | IIIA | + |
| | BCC0291 | ST807 | Cystic fibrosis | Sheffield, UK | IIIA | + |
| | BCC0315 | ST38[b] | Soil | USA | IIIB | + |
| | BCC0366 | ST208 | Cystic fibrosis | USA | IIIA | - |
| | BCC0369 | ST40 | Cystic fibrosis | USA | IIIB | + |
| | BCC0434 | ST309 | Cystic fibrosis | No data | IIIA | - |
| | BCC0460 | ST201 | Cystic fibrosis | No data | IIIA | - |
| | BCC0523 | ST230 | NON-CF[c] | Toronto, Canada | IIIA | + |
| | BCC0524 | ST234 | Cystic fibrosis | France | IIIA | + |
| | BCC0535 | ST33 | Cystic fibrosis | BC, Canada | IIIA | + |
| | BCC0536 | ST210 | Cystic fibrosis | BC, Canada | IIIA | + |
| | BCC0575 | ST234 | Cystic fibrosis | BC, Canada | IIIA | + |
| | BCC0644 | ST234 | Cystic fibrosis | No data | IIIA | + |
| | BCC0687 | ST218 | Cystic fibrosis | No data | IIIA | + |
| | BCC0749 | ST280[b] | Cystic fibrosis | Portugal | IIIA | + |
| | BCC0961 | ST32 | Cystic fibrosis | Prague, Czechia | IIIA | + |
| | BCC1122 | ST32 | Cystic fibrosis | France | IIIA | + |
| | BCC1197 | ST306 | No data | No data | IIIA | - |
| | BCC1283 | ST250 | Environmental | USA | IIIA | + |
| | BCC1295 | ST241 | Environmental | No data | IIIA | - |
| | BCC1552 | ST28 | Cystic fibrosis | No data | IIIA | + |
| | BCC1558 | ST241 | Environmental | No data | IIIA | - |
| | BCC1561 | ST28[a] | Cystic fibrosis | Canada | IIIA | + |
| | BCC1562 | ST258 | NON-CF[c] | USA | IIIA | + |

[a] *Novel MLST profile*
[b] *Unknown MLST prolife*
[c] *NON-CF: Non-cystic fibrosis patients*
[d] *CGD: Chronic granulomatous disease*
[e] *The recA lineage associated with the ST in the PubMLST database*
[f] *The BCESM presence or absence was determined using Abricate.*

### 4.2.11. Pan, core and accessory genome analysis

The pan, core and accessory genome of the 42 *B. cenocepacia* strains was determined by Roary-based analysis (v3.13.0) (Page *et al*., 2015). Following coding sequence prediction, a BLASTP sequence identity threshold of 95% was imposed to identify homologous proteins across species. Based on this threshold, the pan genome of the entire collection constituted 25,336 genes, 3094 genes were identified across all 42 genomes and represented the core genome. The accessory genome consisted of 22,242 genes and most genes (17,881) were encoded by less than 15% of the 42 genomes. That is, for a given accessory gene out of the 17,881 genes, the gene was encoded by less than 15% of the genomes. The number of encoded genes per genome ranged between 5837 to 8140 genes per genome, with *B. cenocepacia* strain BCC0687 encoding the least genes and *B. cenocepacia* strain BCC0524 encoding the most.

### 4.2.12. Core genome phylogenomics

To determine strain-to-strain relatedness of the 42 *B. cenocepacia* strains, the 3094 core genes as identified by Roary were aligned to construct a maximum likelihood predicted phylogeny (Figure 4.6). Strains BCC1968, BCC1969, BCC1971, BCC1976, BCC1983 and BCC1997 form a clade with all the strains belonging to the *B. cenocepacia recA* IIIA lineage (Figure 4.6; blue highlight). Environmental strain BCC2001 forms a clade together with the two strains identified as *recA* IIIB lineage (Figure 4.6; red highlight). Environmental strains BCC1987 and BCC1989 also form their own clade and were the most distantly related to the entire collection, forming a natural outgroup of the phylogeny (data not shown) which was used to root the tree (Figure 4.6; highlighted in yellow). The environmental strains BCC1969, BCC1979 and BCC1997 have a sequence type closely related to ST-32 (Table 4.4), a known epidemic lineage of *B. cenocepacia*, and are positioned within the same clade as the ST-32 strains BCC0211, BCC0961, BCC1122 from the in-house collection.

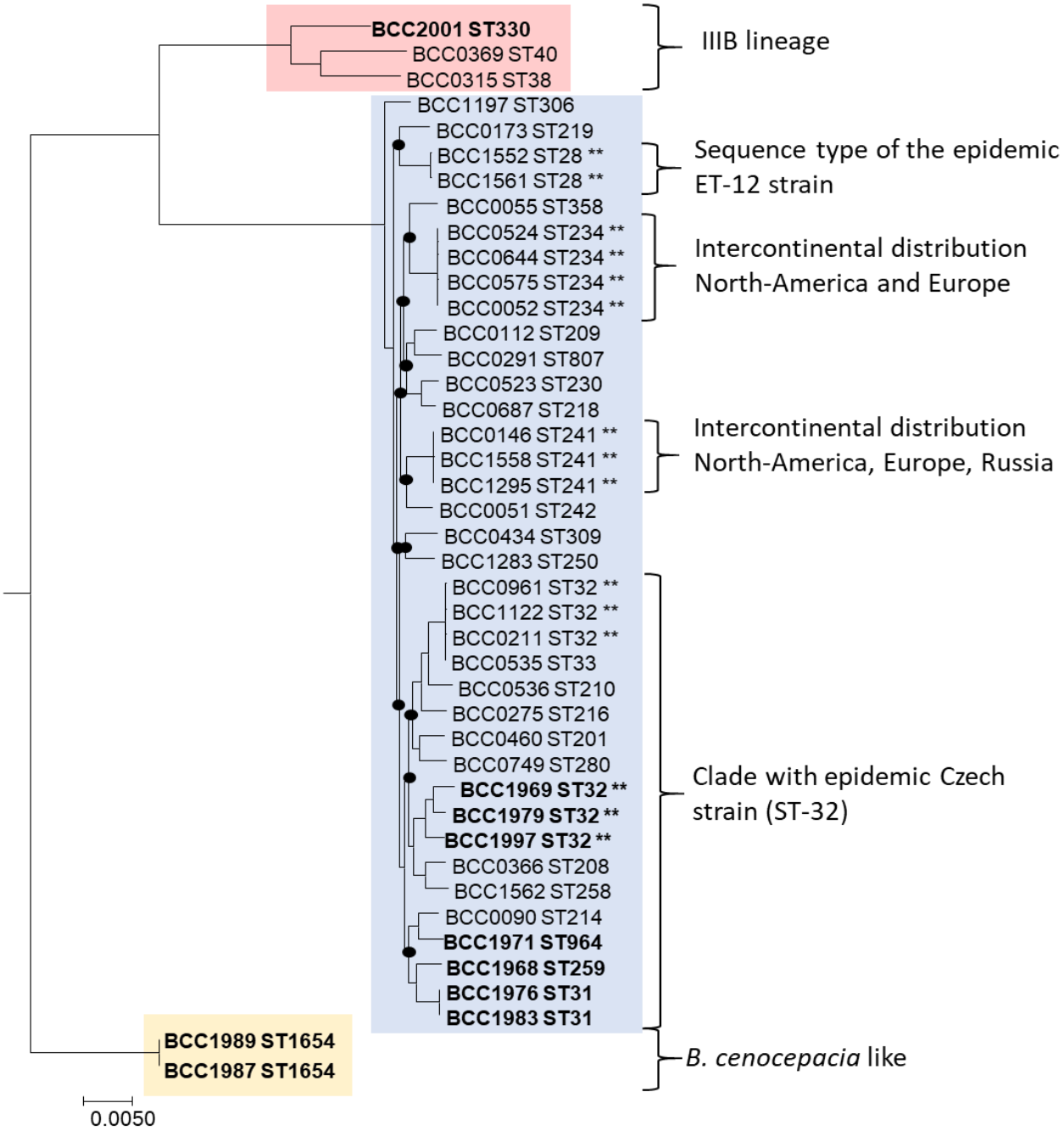



**Figure 4.6. Core-gene phylogeny of the Mysore *B. cenocepacia* collection and 32 characterised *B. cenocepacia.***

A core-gene alignment of 3094 genes from the 42 *B. cenocepacia* genomes was constructed with Roary, and the maximum likelihood phylogeny predicted using RaxML (100 bootstraps). Nodes with bootstrap values of <70% are indicated with black circles. Strains related to an epidemic and/or globally distributed ST lineage are indicated with two asterix. The Mysore collection *B. cenocepacia* are indicated in bold. Scale bar represents the number of substitutions per base position.

### 4.2.13. Screening for *B. cenocepacia* island homologs

Initial characterisation of the Mysore *B. cenocepacia* collection (Figure 3.3) demonstrated that three isolates were positive for the presence of the BCESM clinical risk marker (Mahenthiralingam *et al.*, 1997). The 42 *B. cenocepacia* strain genomes were therefore investigated for the presence of the BCESM sequence (Mahenthiralingam *et al.*, 1997) and the wider *B. cenocepacia* island (*cci*) as described by Baldwin *et al.* (2004). Baldwin *et al.* (2004) originally described a 43-kb region including 47 predicted coding regions within the genome of *B. cenocepacia* strain J2315. This 43-kb region included a putative genomic island, the *cci,* a 31.7-kb region with 35 predicted coding regions (BCAM0239a – BCAM0275) flanked by arsenic tolerance, antimicrobial resistance and stress protein-like genes. A total of 46 gene sequences, from the 47 coding regions predicted from the genome of strain J2315, were assembled in a database to screen against the 42 *B. cenocepacia* genomes using Abricate (Figure 4.7 and 4.8). Coding region BCAM0269 was not submitted to the GenBank database and was omitted from the analysis. Sequences which had an ≥85% sequence coverage, with a ≥75% nucleotide identity match, were considered as a potential homolog. The *esmR* gene*,* which is part of the BCESM PCR marker (Mahenthiralingam *et al*., 1997), was identified in 31 *B. cenocepacia* strains, of which 3 (BCC2001, BCC0315 and BCC0369) belonged to the *B. cenocepacia* IIIB lineage (Figure 4.7). For strains BCC0315 and BCC0369, 24 and 25 homologous sequences, respectively, of the *cci* were identified. Strains BCC2001, BCC1558 and BCC1295 carry a homolog of the *esmR* (Figure 4.7)*,* however most of the remaining *cci* was missing (Figure 4.7 and 4.8). For strain BCC2001 only 2 amino acid metabolism gene homologs of the *cci* were identified and for strains BCC1295 and BCC1558 only a putative amino acid transport gene (BCAM0260). Strains BCC1552 and BCC1561 with the same sequence type as strain J2315 (ST-28) carried the full complement of the *cci* except for gene BCAM0245 a putative decarboxylase enzyme (Figure 4.7 and 4.8).

In 6 of the *B. cenocepacia* from the Mysore collection (Figure 4.7; bold text) (BCC1968, BCC1976, BCC1979, BCC1997, BCC1983 and BCC2001) a homolog for *esmR* was identified. The 6 *B. cenocepacia* strains BCC0055, BCC1295, BCC1558, BCC1976, BCC1983, BCC2001 had an *esmR* homolog, but carried none of the other transcriptional regulators (BCAM0257, BCAM0258, BCAM0259) (Figure 4.7 and 4.8). In strain BCC0055 a total of 25 (71%) homologs from the *cci* were identified, and strains BCC1976 and BCC1983 each had 23 (66%) homologs. Strains BCC1979 and BCC1997 are closely related to ST-32 with 27 and 26 homologs from the *cci* identified, respectively. Interestingly, for strain BCC1969, also closely related to ST-32, the *cci* was missing almost entirely with only two homologs identified from the *cci* (Figure 4.7 and 4.8). Strains (BCC0146, BCC1295, BCC1558) with ST-241, the same sequence type of strains reported to have spread through Canada and the US, carried less than 4 homologs of the *cci* (Figure 4.7 and 4.8). However, strains BCC1295, BCC1558 both carried *esmR* homologous sequences.


**Figure 4.7. The distribution of downstream *cci* gene homologs (BCAM0239a to BCAM0257) across the 42 *B. cenocepacia* strains.**
All 42 *B. cenocepacia* genome sequences were screened using Abricate for 46 coding sequences, including the *cci* (BCAM0239a to BCAM0257), metal tolerance and antimicrobial resistance genes (BCAM0233 to BCAM0239) and stress protein-like genes (BCAM0276 to BCAM0280a) that flank the island. The associated ST of each strain is indicated after the strain ID. The dashed line indicates the start of the *cci* as identified by Baldwin *et al.* (2004). The *esmR* gene (BCAM0256) is indicated with a black box. The tree in Figure 4.6 was used to make a topology only tree to align with the table. Mysore collection strains are indicated in bold. The pink cells indicate where more than one homolog has been identified but at least one homolog has ≥85% gene coverage. If only one gene homolog has been identified, with ≥85% sequence coverage and a ≥75% nucleotide identity match, the cell is in red. AHL stands for *N*-acyl homoserine lactone.


**Figure 4.8. The distribution of the upstream *cci* gene homologs (BCAM0258 to BCAM0275a) across the 42 *B. cenocepacia* strains.**
All 42 *B. cenocepacia* genome sequences were screened using Abricate for 46 coding sequences, including the *cci* (BCAM0239a to BCAM0257), metal tolerance and antimicrobial resistance genes (BCAM0233 to BCAM0239) and stress protein-like genes (BCAM0276 to BCAM0280a) that flank the island. The tree in Figure 4.6 was used to make a topology only tree to align with the table. The associated ST of each strain is indicated after the strain ID. The dashed line indicates the end of the *cci* as identified by Baldwin *et al.* (2004). Mysore collection strains are indicated in bold. The pink cells indicate where more than one homolog has been identified but at least one homolog has ≥85% gene coverage. If only one gene homolog has been identified, with ≥85% sequence coverage and a ≥75% nucleotide identity match, the cell is in red.

## 4.3. Discussion

Following *RecA* and 16S rRNA gene analysis of the Mysore *Burkholderia* collection as described in Chapter 3, whole genome sequences were obtained for MLSA, genomics ANI and phylogenomic analysis to increase resolution at the species level and determine the taxonomic relationship of the environmental *Burkholderia* to characterized Bcc strains. Genomic ANI, MLSA and phylogenomic analysis revealed 22 strains to belong to known Bcc species lineages and 12 strains which represent putative novel Bcc species lineages. Additionally, 10 environmental strains were identified as *B. cenocepacia* including 7 strains which belong to the IIIA lineage, a lineage which is rarely isolated outside of clinical infections. To determine the relationship of the Mysore *B. cenocepacia* collection to problematic clinical isolates, MLST and phylogenomic analysis of the 10 environmental and 32 previously characterized *B. cenocepacia* strains was conducted. In addition, all *B. cenocepacia* strains were screened for a pathogenicity island (*cci*) associated with enhanced virulence and transmissibility in *B. cenocepacia*. The isolation of 34 *Burkholderia* strains including several putative novel species suggests the Western Ghats, India is a source rich in novel *Burkholderia* species. Furthermore, the isolation of 7 *B. cenocepacia* IIIA strains in a single environmental sampling study is the first time a major natural source of *B. cenocepacia* IIIA strains has been identified.

### 4.3.1. Identification of the environmental *Burkholderia* isolates using phylogenomics

Whole genome comparison statistics such as ANI has been widely used for calculating the relatedness between genomes to identify and delineate species (Chun and Rainey, 2014). Richter and Rossello-Mora (2009) reported ANI values of ≈95-96% to be a suitable threshold value for species delineation. Additionally, Richter and Rossello-Mora (2009) suggest adopting an upper boundary threshold value of 96% would be suitable to strengthen obscure species boundaries. Interestingly, the 22 strains which were assigned to characterized species lineages (Table 4.2.), except for *B. cenocepacia* strains BCC1987, BCC1989, BCC2001 and *B. pyrrocinia* strain BCC1992, shared ANI values ≥97.3% with the species type strains providing strong evidence for their species designations. *B. pyrrocinia* strain BCC1992 shared an ANI value of 95% with the *B. pyrrocinia* type strain, which is at the lower ANI value boundary for species delineation as suggested by Richter and Rossello-Mora (2009). Nevertheless, strain BCC1992 clustered with the two *B. pyrrocinia* reference strains, including the type strain LMG 14191, in the MLSA phylogeny (Figure 4.1.) and its position was well supported with strong bootstrap values.

The *recA* gene phylogeny (Figure 3.1; green highlight) revealed strains BCC1987, BCC1989 and BCC2001 to belong to the *B. cenocepacia recA* IIIB lineage. *B. cenocepacia* strain BCC2001 formed a clade with the two *recA* IIIB lineage reference strains (PC184 and ATCC17765) in the MLSA phylogeny

(Figure 4.1.). Additionally, BCC2001 shared an ANI of 95.1% with strain LMG16656$^T$ and its associated ST (ST-330) was characterized as *B. cenocepacia,* providing strong support for its species designation. In contrast, strains BCC1987 and BCC1989 were positioned outside the IIIB *B. cenocepacia* clade in the MLSA phylogeny (Figure 4.1.; green highlight), had an ANI of 94.5% when compared against LMG16656$^T$ and their associated ST (ST1654) has not yet been characterized. Recently, Wallner *et al.* (2019) proposed a division of the *B. cenocepacia* species into two separate species lineages, based on genomic ANI, phylogenomics and the distribution of key virulence factors and plant associated genes. They proposed the name *Burkholderia servocepacia* for a new species lineage*,* although not formally accepted, which consists mainly of *recA* IIIB lineage strains (Vandamme *et al.,* 2003) with *B. cenocepacia* strain Tatl-371 as the type strain. The genomes of BCC1987 and BCC1989 were compared to Tatl-371 to determine if these strains could be characterized as *B. servocepacia* to support this new species definition. However, the strains shared an ANI of 94.9% with Tatl-371 falling just below the species cutoff value suggested for species delineation by Richter and Rossello-Mora (2009). Instead, strains BCC1987 and BCC1989 were considered *B. cenocepacia*-like as they shared an ANI of 95% with BCC2001 (Figure 4.2.) and clustered together in the core-genome phylogeny (Figure 4.4.; purple highlight).

### 4.3.2. Environmental *Burkholderia* and their association with medicinal plants

*Burkholderia cepacia* complex bacteria are ubiquitous in the natural environment and are common soil inhabitants (Stopnisek *et al.,* 2013). Various soil characteristics such as pH, organic carbon concentration, salinity, texture and available nitrogen have been proposed to influence microbial soil communities (Fierer, 2017). However, several studies have reported soil pH as a primary factor driving microbial diversity, community composition and relative abundance (Fierer and Jackson, 2006; Wang *et al.*, 2019). Interestingly, acid tolerance is a general phenotypic trade of the genus *Burkholderia* which provides a competitive advantage in acidic soils (Stopnisek *et al.*, 2013). Although soil pH was not recorded for this study, acid tolerance has been linked to *Burkholderia* prevalence in soil communities and could play a role in their distribution in the soil of the Western Ghats, India. Overall, the recovery of 35 *Burkholderia* strains from 31 of the 66 samples collected, indicates the soil and rhizosphere environment in the Western Ghats, India is a rich source of naturally occurring *Burkholderia.*

*Burkholderia* are widely reported as plant-associated bacteria either free-living in the rhizosphere or as epiphytic or endophytic bacteria (Suarez-Moreno *et al.,* 2012). However, the association of *Burkholderia* species with medicinal plants has not been well investigated. The rhizospheres sampled in this study were likely colonised through contact with soil dwelling *Burkholderia,* as microbiota of

the rhizosphere are mostly acquired through horizontal transfer i.e. roots are colonised by microbes from the surrounding soil environment (Fierer, 2017). However, contact and colonisation with *Burkholderia* could have occurred through other mechanisms as bacteria can also be passed down to successive generations via seeds or transmitted via pollen (Koberl *et al.,* 2013). *Burkholderia* have been isolated as endophytic bacteria and are able to translocate to other parts of the plant which indicates they are capable of colonising seeds and pollen (Mendes *et al.,* 2007; Compant *et al.,* 2008b). Indeed, Oteros *et al.* (2018) reports the isolation of *Bukholderia* from airborne pollen of *Artemisia vulgari,* which suggests pollen transfer can be a possible route through which plants are colonised by *Burkholderia*. Furthermore, seed-borne *B. glumae* have been identified as a source from which rice seedlings are infected by *B. glumae* (Pedraza *et al.*, 2018). Through which route the plants sampled in this study acquired *Burkholderia* remains to be investigated. Nevertheless, the recovery of 21 Bcc strains including 3 novel species lineages from the rhizosphere of medicinal plants reveals a rich diversity of Bcc bacteria within this ecological niche.

Soil microorganisms are attracted to the nutrient rich environment of the plant rhizosphere (Koberl *et al.*, 2013). Root exudates feed the microbial community and influence their composition and activities, it therefore stands to reason plants produce exudates to regulate the rhizosphere environment to their benefit. For example, Rudrappa *et al.* (2008) showed *Arabidopsis thaliana* recruits the beneficial rhizobacterium *Bacillus subtilis* using malic-acid-mediated stimulation. Medicinal plants produce unique and structurally diverse specialised metabolites which likely influence microbial community composition. For example, Zhao *et al.* (2012) assessed the *Actinobacteria* diversity of 7 medicinal plants and found each plant was colonized by a unique rhizosphere and endophyte community. Even plants grown in the same soil hosted a remarkedly different collection of strains, which suggests plant species, and thus their root exudates, can affect the strain combinations found in the rhizosphere. Indeed, several studies have shown plants have specificity towards certain groups of bacteria (Berg and Smalla, 2009; Li *et al.,* 2014). However, none of the plant species sampled in this study were observed to have specificity towards a species of *Burkholderia* (Table 3.1). Nevertheless, it is reasonable to consider a relationship with *Burkholderia* is supported by the plant if such a relationship was beneficial. Several enzymes and metabolic pathways have been identified in *Burkholderia* allowing them to catabolize plant hydrocarbon derivatives, which highlights their ability to form a close association with plants (Wallner *et al.*, 2019). Ultimately, the effect of medicinal plants and their root secretions, specifically their specialized metabolites, on *Burkholderia* prevalence in the rhizosphere would have to be investigated for each strain individually.

Several species of *Burkholderia* have been reported with plant beneficial characteristics such as pathogen protection, promoting plant growth and plant nutrient acquisition (Suarez-Moreno *et al.,*

2012). Furthermore, multiple strains were isolated in this study from species lineages with evidence of members forming plant beneficial relationships. This study identified strains from 4 known species lineages including: *B. pyrrocinia* (n=1), *B. cepacia* (n=9), *B. ubonensis* (n=2) and *B. cenocepacia* (n=8) (Table 4.2). These 4 known species lineages have been isolated in previous soil and rhizosphere sampling studies, and members from those species' lineages have been reported with plant beneficial properties (Madhaiyan *et al.,* 2008; Lee *et al.,* 2011; Price *et al.,* 2017; Jung *et al.,* 2018; Wallner *et al.,* 2019). Lee *et al.* (2011) reports the isolation of a *B. pyrrocinia* strain (CH-67) from forest soil with antifungal activity against several phytopathogenic fungi. Additionally, genome analyses of CH-67 by Song *et al.* (2012) revealed several genetic factors beneficial to plants. Similarly, Jung *et al.* (2018) reports the isolation of a *B. cepacia* strain (JBK9) from the rhizosphere of cultivated crops with antifungal activity against the phytopathogens *P. capsici*, *F. oxysporum*, and *R. solani.* Additionally, Jung *et al.* (2018) demonstrated JBK9 was able to protect pepper plants from infection by the fungal pathogen *P. capsica. B. ubonensis* and *B. cenocepacia* have also frequently been isolated from soil and plant rhizospheres, and plant beneficial properties for members of those species' lineages have been reported (Madhaiyan *et al.* 2008*,* Ho *et al.,* 2015; Price *et al.,* 2017; Wallner *et al.,* 2019). Moreover, antifungal activity is a widely reported feature of *Burkholderia* bacteria. *Burkholderia,* and most fungi, prefer soils with an acidic pH and previous work has shown that members of the *Burkholderia* genus often co-occur with fungi in the soil environment (Stopnisek *et al.,* 2013; Stopnisek *et al.,* 2015). The antagonism of fungi by *Burkholderia* has been well investigated and many antifungal compounds from *Burkholderia* have been described (Vial *et al.,* 2007; Kunakom and Eustáquio, 2019). The association between medicinal plants and *Burkholderia* in the Western Ghats could provide the plants with protection against fungal pathogens. Additionally, *Burkholderia* can also protect plants by stimulating plant defences through the production of reactive oxygen species and phytohormones (Esmaeel *et al.,* 2019). In fact, the interaction between medicinal plants and their microbiome has been reported to have a direct or indirect influence on the production of bioactive phytochemicals such as antimicrobials (Koberl *et al.*, 2013). The interaction between rhizosphere *Burkholderia* and medicinal plants could play a role in their specialized metabolite production, however this would have to be investigated further.

### 4.3.4. Investigating the relationship and pathogenic potential of environmental *B. cenocepacia* to clinically characterized strains

A collection 42 *B. cenocepacia* strains including 32 previously characterized *B. cenocepacia* strains from the Cardiff university in-house collection, and 10 *B. cenocepacia* environmental strains from the Western Ghats, India, was assembled for MLST and phylogenomic analysis. None of the Mysore environmental strains grouped with characterized strains of the IIIC and IIID lineages (Vandamme *et al.,* 2003) in the *recA* gene phylogeny (Figure 3.2.) and therefore no representative strains were included in the final analysis. The 37 strains which were previously identified as *B. cenocepacia recA* IIIA lineage (Table 4.4.) (Figure 3.1.; green highlight), shared ANI values ranging 98.7% to 100% with *B. cenocepacia* LMG16656$^{T}$ (a IIIA strain) confirming their identification. The remaining 5 *B. cenocepacia* strains (BCC1987, BCC1989, BCC2001, BCC0315 and BCC0369) were classified as *recA* IIIB lineage strains (Table 4.4.) (Figure 3.1.; green highlight). Previous studies supporting the division of *B. cenocepacia* into IIIA, IIIB, IIIC and IIID subgroups have relied almost exclusively on *recA* gene analysis and MLSA (Mahenthiralingam *et al.,* 2000b; Vandamme *et al.,* 2003; Bladwin *et al.*, 2005; Jin *et al.,* 2020). Whole genome analysis offers greater resolution at the species level and several phylogenomic studies have proposed the *recA* IIIB lineage represents a separate species lineage (Wallner *et al.* 2019; Jin *et al.* 2020). Richter and Rossello-Mora (2009) suggest adopting an ANI value of 96% to be suitable for demarcating obscure species boundaries. Strains BCC2001, BCC0315 and BCC0369 share ANI values ranging 97.8% to 100% amongst each other but an ANI of 95.1% with type strain LMG16656$^{T}$, which suggests they belong to a single novel species. Strains BCC1987 and BCC1989 which share ANI values ranging 94.9% to 95% with the other IIIB strains, and an ANI of only 94.5% with strain LMG16656$^{T}$, would represent an additional novel species. Furthermore, strains BCC1987 and BCC1989 form an outgroup in the core-genome phylogeny, therefore their current classification should be re-examined for future analyses (Figure 4.6.; yellow highlight). For the purposes of this study all five of the *recA* IIIB lineage strains were designated as *B. cenocepacia.* However, the classification of *B. cenocepacia* subgroups into separate species lineages should be evaluated in future phylogenomic analyses, especially in light of Wallner et al. (2019) proposing the novel species name *B. servocepacia* as a substitute for the IIIB lineage.

Most clinically relevant *B. cenocepacia* isolates have been reported to reside in subgroups IIIA and IIIB (Drevinek *et al.,* 2008). Furthermore, *B. cenocepacia* subgroup IIIA has been almost exclusively isolated from clinical sources whereas subgroup IIIB has been isolated from both clinical and environmental sources (Drevinek *et al.,* 2008). A shift from a free-living/plant-associated to a host-associated

pathogen lifestyle has been accredited to the isolation of very few IIIA lineage strains outside a clinical setting (Holden *et al.*, 2008; Wallner *et al.*, 2019). Out of the 10 *B. cenocepacia* strains isolated from the soil or rhizosphere in the Western Ghats, India, 7 strains were classified as IIIA lineage strains (Figure 3.1.; green highlight) (Table 3.1.), which suggests IIIA lineage strains could be more prevalent in the natural environment than initially suspected (Drevinek and Mahenthiralingam, 2010). Whether environmental IIIA lineage strains have the same pathogenic potential as clinical isolates has not been well investigated. However, few systematic studies looking at environmental samples and attempting to isolate Bcc bacteria have identified *B. cenocepacia* IIIA isolates. Pirone *et al.* (2008) reported environmental IIIA lineage strains Mex-1 (maize, rhizosphere) and POPR8 (raddish, rhizosphere) to have a similar capacity as clinical isolates to maintain a chronic lung infection in a mouse infection model. Additionally, Wallner *et al.* (2019) proposed a division of *B. cenocepacia* into two separate species lineages based on the distribution of key virulence factors and environmental and plant associated traits. Only 1 environmental *B. cenocepacia* (soil) strain (F01) of the IIIA lineage was used in their analysis, however all virulence factors investigated were identified in this strain (Wallner *et al.,* 2019). The abovementioned studies suggest virulence associated traits are naturally occurring in environmental IIIA lineage strains; however, this warrants further investigation with a greater number of representatives.

MLST of the *B. cenocepacia* strains was conducted to determine if strains belonged to notable ST lineages or clonal complexes. Sequence types can be associated with clinical epidemiology and allows the identification of strains with the propensity to cause infection and disease (Baldwin *et al.,* 2005; Drevinek and Mahenthiralingam, 2010). In addition, the relationship of environmental strains to strains of characterized ST lineages was further evaluated using core-genome phylogeny. The MLST profiles of the 42 *B. cenocepacia* strains were associated with 32 unique ST's, representing a broad diversity of STs (Table 4.4.). The MLST profiles of 27 strains were identical with those of characterized STs, including the STs from epidemic and/or globally distributed lineages (ST-28, ST-234, ST-241, ST-32) (Table 4.4.). The MLST profiles of the environmental *B. cenocepacia* revealed 7 ST's associated with *B. cenocepacia* IIIA lineage strains, one ST with *B. cenocepacia* IIIB lineage strain and two uncharacterized STs (Table 4.4.). The Bcc MLST scheme utilizes 7 genes to assign a sequence type however the use of only 6 genes has been reported accurate enough for species and strain identification (Mahenthiralingam *et al.,* 2008).

The MLST profile of IIIB lineage strain BCC2001 had unknown polymorphisms in 5 out of the 7 alleles and BCC1987 and BCC1989 had unknown polymorphisms in every single allele. However, IIIA lineage strains BCC1968 (ST-259), BCC1971 (ST-964), BCC1976 (ST-31), BCC1983 (ST-31), BCC1997 (ST-32) only differed in a single allele, *gyrB,* from their associated STs. Strains BCC1979 and BCC1969 varied only in

two alleles, *gyrB* and *recA,* to their associated ST (ST-32). These minor differences in MLST type indicate they all belong to the same clonal complex (Baldwin *et al.,* 2005; Baldwin *et al.,* 2007). ST-32 has been detected in CF populations in Italy, France, UK and Canada and has spread epidemically in Canada (strain type RAPD01) and the Czech Republic (strain CZ1) (Drevinek and Mahenthiralingam, 2010). The same *B. cenocepacia* ST-32 strain also caused multiple problematic infections in non-CF individuals within intensive care units in France (Graindorge *et al.,* 2010). Core-genome phylogeny positioned all 7 environmental strains in the same clade with clinical reference strains of ST-32 demonstrating a close genetic relationship (Figure 4.6.; blue highlight). Furthermore, the ST's of IIIA strains BCC1968 (ST-259), BCC1976 (ST-31), BCC1983 (ST-31), BCC1997 (ST-32), BCC1979 (ST-32) and BCC1969 (ST-32) are all linked to clonal complex 31, which also includes ST-28, an epidemic CF lineage which spread intercontinentally between Canada and the UK during the 1990s (Drevinek and Mahenthiralingam, 2010). Interestingly, the STs of the environmental IIIA strains, except for ST-259, have clinical isolates with the same STs submitted to the PubMLST database (https://pubmlst.org/bcc/). The close relationship of the environmental isolates in this study to epidemic clinical strains suggests the Western Ghats, India is a rich environmental source of potentially virulent *B. cenocepacia* strains.

### 4.3.5. Characterizing known *B. cenocepacia* virulence markers: the *B. cenocepacia* pathogenicity island

To further evaluate the pathogenic potential of the environmental *B. cenocepacia* strains they were investigated for the presence of the BCESM (Figure 3.3.) (Mahenthiralingam *et al.*, 1997; Baldwin *et al.,* 2004). The 1.4-kb BCESM DNA encodes a putative negative transcriptional regulator (BCAM0256 - *esmR*) which resides on a pathogenicity island (*cci*) encoding genes linked to virulence and metabolism (Baldwin *et al.*, 2004; Malott *et al.,* 2005). Analysis of CF patients within British Columbia, Canada demonstrated that *B. cenocepacia* IIIA strains that were BCESM positive resulted in the greatest mortality of CF individuals over a 30-year study period (Mahenthiralingam *et al.*, 2001; Zlosnik *et al.,* 2015). Furthermore, previous studies have demonstrated attenuated virulence and inflammatory potential in knockout mutants of the following 4 *cci* genes: *cciI* (BCAM0239a - AHL synthase gene), *cciR* (BCAM0240 - AHL synthesis response regulator gene), *amiI* (BCAM0265 - amidase synthase) and *opcI* (BCAM0267 - porin gene) (Baldwin *et al.,* 2004; Malott *et al.,* 2005). The *cci* is not essential for *B. cenocepacia* pathogenicity, as *cci* negative strains have been isolated from infections, but does contribute to inflammation and virulence in the CF lung (Baldwin *et al.,* 2004; Pirone *et al.*, 2008).

To establish if the 42 *B. cenocepacia* strains encoded the *esmR* and the wider *B. cenocepacia* island (*cci*), their genomes were screened using Abricate (v.0.9.8) (https://github.com/tseemann/abricate)

against a database containing the *cci* genes as described by Baldwin *et al.* (2004). The *esmR* was identified in 31 out of the 42 *B. cenocepacia* strains, including 3 strains (BCC2001, BCC0315 and BCC0369) which belonged to the *B. cenocepacia* IIIB lineage (Figure 4.7). Interestingly, the Mysore collection strain BCC2001 (IIIB lineage), and the in-house collection strains BCC1295 (IIIA lineage) and BCC1558 (IIIA lineage) carried an *esmR* homolog but less than 4 additional homologs from the *cci,* and none of the *cciI*, *amiI*, *cciR* and *opcI* genes (Figure 4.7.; Figure 4.8.). The presence of the *esmR,* but absence of the 4 characterized genes, suggests a single gene such as *esmR* is not a good marker to determine if the whole pathogenicity island (*cci*) is present.

Southern hybridization mapping of selected *cci* genes against 88 *B. cenocepacia* strains by Baldwin *et al.* (2004) revealed that the presence of the *esmR, cciI*, *amiI*, and *opcI* genes in a strain was always associated with an intact *cci*. Based on the presence of *esmR, cciI*, *amiI*, and *opcI* homologs in a strain, the *cci* was identified in 26 out of the 37 IIIA lineage strains (Figure 4.7.; Figure 4.8.), including 5 out of the 7 environmental IIIA-lineage strains (BCC1968, BCC1976, BCC1979, BCC1997 and BCC1983). The 35 genes from the *cci* as described by Baldwin *et al.* (2004) were well conserved (66% to 88% of genes within the island) amongst the environmental IIIA lineage strains (Figure 4.7.; Figure 4.8.), which suggests the *cci* may also have a functional role in the rhizosphere and soil environment. In addition, several other studies have reported the *cci* in environmental strains which demonstrates that the *cci* cannot be used to distinguish clinical and environmental isolates (Baldwin *et al.,* 2004; Pirone *et al.*, 2008). Nevertheless, the *cci* promotes survival and pathogenesis in the CF lung and its prevalence amongst environmental strains highlights their intrinsic ability for virulent opportunistic human infection (Baldwin *et al.,* 2004). Future analyses should evaluate the role and prevalence of key virulence factors, such as the *cci,* amongst environmental *B. cenocepacia* IIIA lineage strains to determine their potential for human infection.

# Chapter 5 - Antimicrobial discovery of the Mysore *Burkholderia* collection

## 5.1. Introduction

The Mysore *Burkholderia* collection was investigated for antimicrobial production utilising strategies to enhance or elicit novel antimicrobial production. The background to the methods and techniques used for antimicrobial discovery in this chapter are introduced as follows:

### 5.1.1. Antimicrobial discovery in the genomics era

With most antibiotic classes having been discovered from soil dwelling *Actinobacteria,* the need for new antibiotics has encouraged the exploration of new sources for clinical drugs such as unique ecological niches and unusual bacteria genera (Aminov, 2010). Among these genera is the *Burkholderia* from which various compounds have already been identified with bioactivity as antifungal, antibacterial, herbicidal, insecticidal, cytotoxic and anticancer agents (Depoorter *et al.*, 2016; Kunakom and Eustáquio, 2019). A recent review by Kunakom and Eustáquio (2019) reports 66 structural classes of compounds across the *Burkholderia* genus and closely related genera (*Paraburkholderia, Caballeronia and Robbsia*), highlighting the biosynthetic potential of this genus.

With the advent of cost-effective whole genome sequencing, and a growing body of knowledge of the genetics and enzymology behind natural product biosynthesis, new strategies to antimicrobial discovery have started to emerge (Zelrikly and Challis, 2009). Bioinformatic tools such as antiSMASH (Blin *et al.*, 2017) has enabled researchers to readily genome mine the vast quantities of DNA sequence data in publicly available databases for novel natural products (Zelrikly and Challis, 2009). Genome mining methods have already identified several useful natural products from *Burkholderia* such as the antifungal compound fragin from *B. cenocepacia* (Jenul *et al.,* 2018), the antibacterial lassopeptide capistriun from *B. thailandensis* (Knappe *et al.,* 2008) and the antibacterial icosalide from *B. gladioli* (Jenner *et al.,* 2019). Although genome mining allows the identification of silent or cryptic BGCs, the structure of the encoded compound is often difficult to predict. Conventional chemical analyses such as HPLC, LC-MS, NMR and X-ray crystallography are still needed to predict the chemical structure of a BGC product.

The expression of seemingly silent BGCs can be activated by modifying lab growth conditions or through genetic engineering to facilitate purification and identification of metabolites that are otherwise expressed at undetectable levels or at low concentrations. Koch *et al.* (2002) reported a significant increase in amphisin production by the *Pseudomonas* sp. strain DSS73 by addition of sugar beet seed exudate to minimal growth media, compared against media lacking the seed exudate. Thus,

modifications to growth conditions present a relatively simple strategy to activate or enhance BGC expression. In addition, various genetic engineering strategies such as promoter exchange or heterologous expression can be used to facilitate the expression of silent or cryptic BGCs so the derived BGC product can be purified and analysed (Kunakom and Eustáquio, 2019).

Advances in whole genome sequencing and the development of bioinformatic tools for phylogenomic analysis and specialised metabolite discovery has contributed significantly to characterising the taxonomic diversity and metabolic capabilities of the *Burkholderia* genus. Furthermore, extensive research concerning the pathogenesis of *Burkholderia,* and the development of tools such as MLST, have contributed to a greater understanding of the epidemiology and virulence mechanisms of pathogenic groups such as the *Burkholderia cepacia* complex (Drevinek and Mahenthiralingam, 2010). Nevertheless, gaps in our knowledge remain as an environmental source of *B. cenocepacia* IIIA strains has not been defined, despite their widespread occurrence as virulent and transmissible CF pathogens (Drevinek and Mahenthiralingam, 2010). Additionally, the *Burkholderia* with their large genomes, diverse lifestyles and metabolic capabilities present a relatively underexploited resource with possible applications in multiple biotechnology industries (Depoorter *et al.*, 2016).

In the following chapter a select group *Burkholderia* are investigated for their specialised metabolite biosynthetic potential by adopting strategies to elicit novel or enhanced antimicrobial production. The following objectives will be addressed in this chapter:

1. Elicit novel or enhanced antimicrobial production in a panel of characterised *Burkholderia* and selected strains from the Mysore collection through modifying lab growth conditions.
2. Characterise the antimicrobial producing potential and biosynthetic pathway distribution of selected *Burkholderia* strains using reductive genome engineering, bioactivity guided isolation and genome mining.

# 5.2. Results

### 5.2.1. Antimicrobial activity screening of the Mysore *Burkholderia* collection

To evaluate the antimicrobial activity of the Mysore *Burkholderia* collection, all strains were tested against a panel of Gram-negative, Gram-positive and fungal pathogens (Figure 5.1; panel a). Following screening a panel of 4 strains (*B. ubonensis* strains BCC1973 and BCC1990, *Burkholderia* sp. strains BCC1970 and BCC1977) with interesting bioactivity were carried forward for further investigation (Figure 5.1; panel b; bold text). Strains were considered bioactive if the zone of inhibition was >7mm in diameter. Of the 34 strains which were screened 25 strains exhibited inhibitory activity against one of the susceptibility organisms (Figure 5.1; panel b). *Burkholderia* sp. strains BCC1988, BCC1998, BCC1972 and BCC1999, *B. pyrrocinia* strain BCC1992 and *B. cepacia* strains BCC1991, BCC1984, BCC1978 and BCC1986 did not exhibit any inhibitory activity against the susceptibility organisms. Of the 34 strains, 11 strains exhibited antifungal activity, 3 strains anti-Gram-negative activity and 19 strains exhibited anti-Gram-positive activity. *B. ubonensis* strains BCC1973 and BCC1990 exhibited noticeably strong activity against *P. carotovorum* and *C. albicans* and were therefore chosen for further investigation. Five novel *Burkholderia* sp. strains BCC1970, BCC1974, BCC1981, BCC1985, BCC1995 and BCC1996 all exhibited antifungal activity and were also closely related in terms of phylogenetic clustering (Figure 5.1; panel b). Strain BCC1970 exhibited the strongest antifungal activity of all the novel *Burkholderia* sp. strains and was therefore selected for further analysis. Strain BCC1977 exhibited weak Gram-positive activity and strong antifungal activity, and along with its evolutionary position in relation to the rest of the *Burkholderia* sp. strains, also made it an interesting candidate for further investigation (Figure 5.1; panel b). The *B. cenocepacia* IIIA strain cluster (Figure 5.1; panel b; strains BCC1971 to BCC1983) and strain BCC2001 all exhibit inhibitory activity against *S. aureus*. In contrast, the putative *B. cenocepacia* IIIB strains BCC1987 and BCC1989 lack inhibitory activity against *S. aureus* but showed strong antagonism of *C. albicans* (Figure 5.1; panel b). The *B. cepacia* isolates exhibited either no inhibitory activity or weak activity (<10mm diameter zone of inhibition) against only *S. aureus* (Figure 5.1; panel b).

To investigate the distribution of BGCs and the role of the third replicon in antimicrobial activity, third replicon mutants were created as described by Agnoli *et al.* (2012), for the 4 strains BCC1970, BCC1973, BCC1977 and BCC1990. Loss of the third replicon was confirmed by PCR (data no shown).

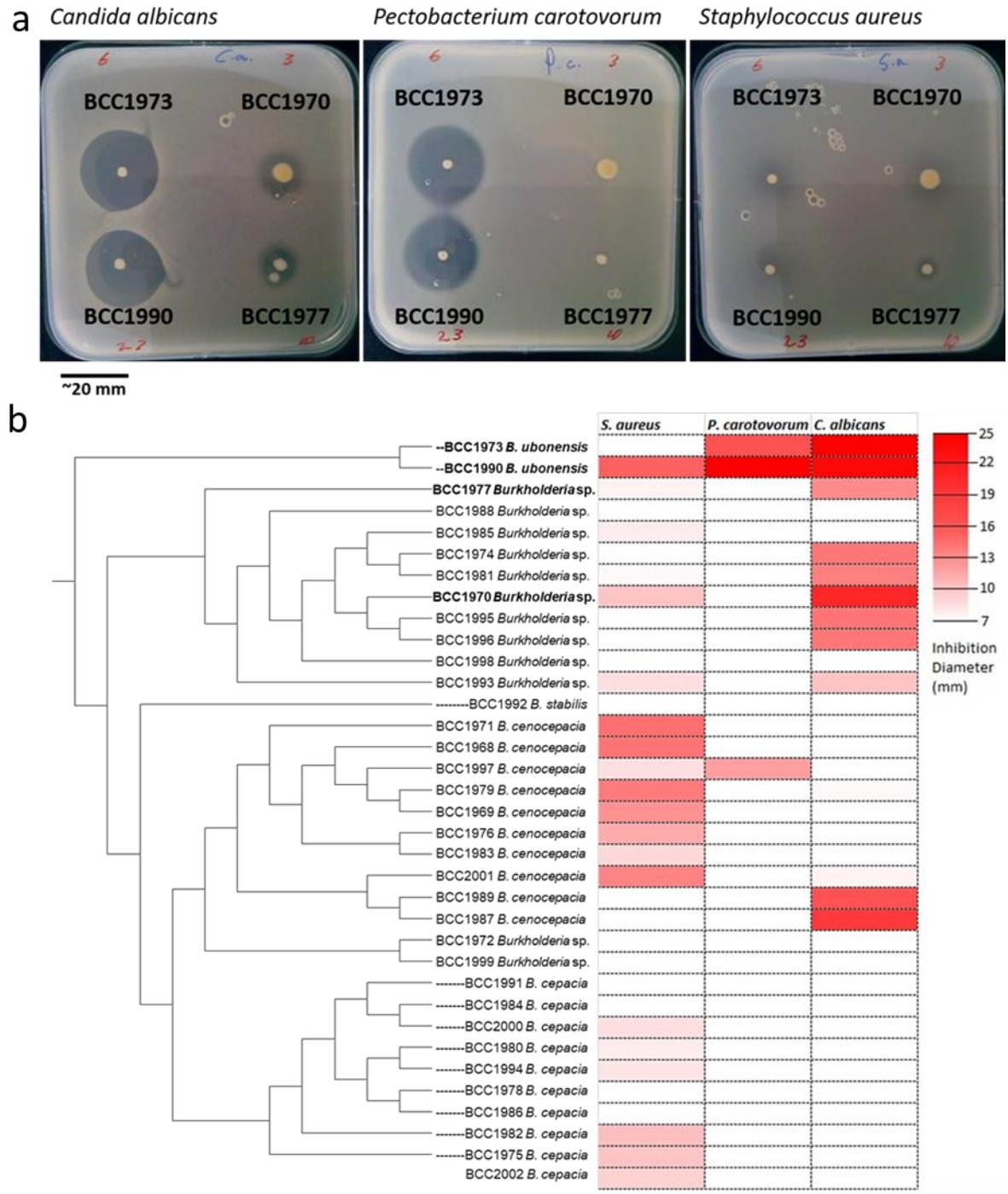



**Figure 5.1. Antimicrobial activity of the Mysore *Burkholderia* collection in relation to their core genome inferred evolutionary relationship.**

**(a)** Zones of inhibition against *Staphylococcus aureus* NCTC 12981, *Pectobacterium carotovorum* LMG 2464 and *Candida albicans* SC 5314 by environmental isolates *B. ubonensis* (BCC1973 and BCC1990) and *Burkholderia* sp. (BCC1970 and BCC1977). Scale bar represents 20 ± 1 mm. **(b)** A core-gene alignment of 985 genes from the 34 environmental isolate genomes was constructed with Roary, and the cladogram created using RaxML (100 bootstraps). Strains with good antimicrobial activity that were chosen for further analysis are highlighted in bold. Antimicrobial activity was defined by measuring the diameter of the zone of inhibition (n=3). Heatmap shows mean zone of inhibition. Initial screening was carried out on BMSG (pH 7) with a 3-day incubation of the environmental strains at 30°C prior to being overlaid with the susceptibility organisms.

### 5.2.2. Investigating the effect of seed exudate media and biomimetic carbon sources within growth media on antimicrobial activity

A panel of 13 *Burkholderia* strains, including the 4 Mysore collection strains BCC1970, BCC1973, BCC1977, BCC1990 and their pC3 replicon mutants, were assembled to investigate if specific root exudate components within growth media could activate silent biosynthetic gene clusters or enhance antimicrobial production. Soil microorganisms are attracted to the nutrient rich environment of the plant rhizosphere and the community structure of root associated microbes is influenced by the composition of plant root secretions and other rhizodeposits (Koberl *et al.,* 2013). The sugars glucose, adonitol, citrate, and the amino acid glutamate were reported to be upregulated by maize in nutrient limited soil and were utilized as the sole carbon source in minimal media (D-glucose, adonitol, tri-Sodium citrate, and monosodium glutamate) (Carvalhais *et al.,* 2010). *Burkholderia cepacia* complex bacteria have been frequently isolated from the maize rhizosphere and the suppression of soilborne fungal and oomycete pathogens by maize root associated *Burkholderia* has been reported, which provided the rational for utilizing maize associated root exudates (Hebbar *et al.,* 1992; Hebbar *et al.,* 1998; Bevivino *et al.,* 2011). Antimicrobial activity was determined by measuring the diameter of the zones of inhibition and was considered novel or altered in relation to the antimicrobial activity observed after growth on BSMG as a control. Wheat and pea seed exudate media were ineffective at inducing novel or enhancing antimicrobial activity for the panel of strains and were not investigated further. The third replicon mutants of *B. ubonensis* strains BCC1973 and BCC1990 lost all activity when tested against the panel of susceptibility organisms. In contrast, the pC3 deletion mutants of *Burkholderia* sp. strains BCC1970 and BCC1977 retained the same antimicrobial activity as observed for the wild type strains (Figure 5.2).

Two carbon sources induced antimicrobial activity not observed for glycerol in 3 strains (BCC1973, BCC1977, BCC1990) (Figure 5.2). Glucose induced Gram-positive antagonism for *B. ubonensis* strains BCC1973 and BCC1990. Additionally, adonitol and glucose induced weak activity against *S. aureus* (≤ 9 mm diameter) for *Burkholderia* sp. strain BCC1977 (Figure 5.2). No antimicrobial activity was observed for the *B. anthina* (BCC0639) and *B. ubonensis* (BCC1783) species type strains. Glutamate was effective at enhancing antimicrobial activity against *C. albicans* for strains BCC1970 and BCC0238. (Figure 5.2). An independent t-test was conducted to compare the zones of inhibition, when *B. gladioli* strain BCC0238 was tested against *C. albicans,* following growth on media with glutamate or glycerol as the sole carbon source. There was a clear significant difference in the scores for glutamate (Mean = 68.67, SD = 2.31) and glycerol (Mean = 48.33, SD = 2.52) as carbon sources; t (4) = 10.3109, p = 0.0005. Additionally, an independent t-test was conducted to compare the zones of inhibition, when *Burkholderia* sp. strain BCC1970 was tested against *C. albicans,* following growth on media with

glutamate or glycerol as the sole carbon source. There was a clear significant difference in the scores for glutamate (Mean = 35, SD = 2.00) and glycerol (Mean = 17.33, SD = 2.52) as carbon sources; t (4) = 9.5191, p = 0.0007.

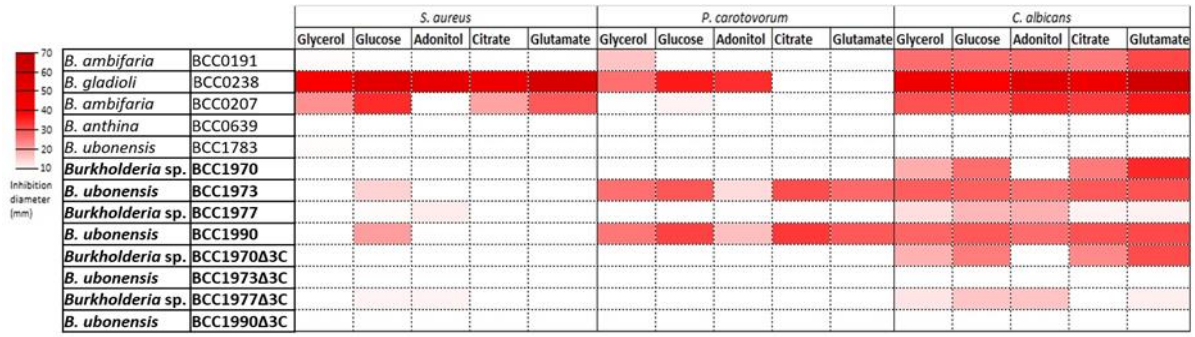
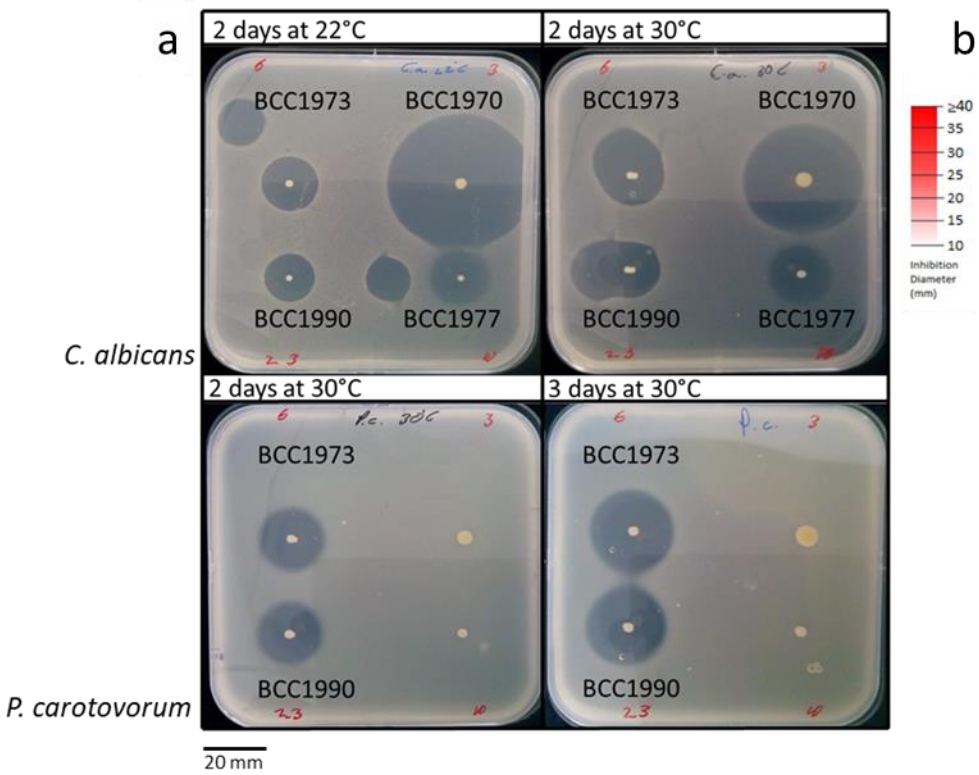



**Figure 5.2. Antimicrobial activity of 13 selected strains when grown on minimal growth media supplemented with carbon sources associated with plant root exudates.**
The antimicrobial activity of 13 *Burkholderia* strains was defined by measuring the diameter of the zones of inhibition (mm); n = 3 overlays of each *Burkholderia* strain for each carbon source against each susceptibility organisms. Heatmap shows mean zone of inhibition. Strains from the Mysore collection and their 3rd replicon mutants (Δ3C) are indicated in bold.

### 5.2.3. Investigating the effect of changes in incubation period and temperature on antimicrobial activity

Incubation temperature and time have been demonstrated to influence antimicrobial activity in *Burkholderia* (El-Banna and Winkelmann, 1997; Funston *et al.,* 2016). The panel of 4 Mysore collection strains were therefore investigated for the optimal incubation period and temperature that elicits the strongest antimicrobial activity. Strains were incubated at either 2 or 3 days at either 22°C or 30°C. Antimicrobial activity against *C. albicans* and *P. carotovorum* was strongest at 3 days for *B. ubonensis* strains BCC1973 and BCC1990. Antimicrobial activity was weakest when the strains were incubated at 22°C (Figure 5.3). In contrast, antimicrobial activity against *C. albicans* for *Burkholderia* sp. strains BCC1970 and BCC1977 was the strongest when incubated for 2 days at 22°C (Figure 5.3).




b

| Strain | Susceptbility organism | Incubation period | | Temperature (°C) |
|---|---|---|---|---|
| | | 3 Days (n=3) | 2 Days (n=3) | |
| Burkholderia spp. BCC1970 | *C. albicans* | | | 30 °C |
| | *C. albicans* | | | 22 °C |
| | *S. aureus* | | | 30 °C |
| | *S. aureus* | | | 22 °C |
| | *P. carotovorum* | | | 30 °C |
| | *P. carotovorum* | | | 22 °C |
| Burkholderia ubonensis BCC1973 | *C. albicans* | | | 30 °C |
| | *C. albicans* | | | 22 °C |
| | *S. aureus* | | | 30 °C |
| | *S. aureus* | | | 22 °C |
| | *P. carotovorum* | | | 30 °C |
| | *P. carotovorum* | | | 22 °C |
| Burkholderia spp. BCC1977 | *C. albicans* | | | 30 °C |
| | *C. albicans* | | | 22 °C |
| | *S. aureus* | | | 30 °C |
| | *S. aureus* | | | 22 °C |
| | *P. carotovorum* | | | 30 °C |
| | *P. carotovorum* | | | 22 °C |
| Burkholderia ubonensis BCC1990 | *C. albicans* | | | 30 °C |
| | *C. albicans* | | | 22 °C |
| | *S. aureus* | | | 30 °C |
| | *S. aureus* | | | 22 °C |
| | *P. carotovorum* | | | 30 °C |
| | *P. carotovorum* | | | 22 °C |

**Figure 5.3. Antimicrobial screening of environmental *B. ubonensis* (BCC1973 and BCC1990) and *Burkholderia* sp. (BCC1970 and BCC1977) at different temperatures and incubation periods.**
**(a)** Zones of inhibition when strains BCC1970, BCC1973, BCC1977 and BCC1990 are tested against *P. carotovorum* LMG 2464 and *C. albicans* SC 5314. Scale bar represents 20 mm. **(b)** Heatmap showing the difference in antimicrobial activity for strains BCC1970, BCC1973, BCC1977 and BCC1990 when incubated at 2 days or 3 days and at different temperatures. Antimicrobial activity was defined by measuring the diameter of the zones of inhibition (mm); n = 3 overlays of each *Burkholderia* strain for each growth condition against each susceptibility organism. The diameter of the zone of inhibition in mm is indicated by the colour coded key (top left).

### 5.2.4. Detection of specialised metabolite production by HPLC analysis

The Mysore strains BCC1970, BCC1973, BCC1977 and BCC1990 were cultured on BSMG and the extracts derived from the spent BSMG media analysed using HPLC as previously described (Chapter 2; section 2.10.). HPLC analysis of *Burkholderia* sp. strains BCC1970 and BCC1977 (Figure 5.4) identified compounds with novel absorbance spectra with maximum absorbance wavelength $\lambda_{max}$=300 nm that could not be attributed to known metabolites previously observed by HPLC analysis (Mullins *et al.,* 2019). Peak height and integrated HPLC-peak area have a proportional relationship to metabolite concentration. The peak height of metabolites detected with a maximum absorbance wavelength $\lambda_{max}$=300 nm for *Burkholderia* sp. strains BCC1970 and BCC1977 demonstrated a decrease in metabolite concentration as incubation period was increased (Figure 5.4).

*B. ubonensis* strains BCC1973 and BCC1990 exhibited metabolite peaks with max absorbance wavelength $\lambda_{max}$=360 nm (Figure 5.5), with retention times similar to those observed for enacyloxin IIa from *B. ambifaria* strain AMMD, the prototypic enacyloxin producer strain (Mahenthiralingam *et al.,* 2011). The enacyloxin-like compounds identified in strains BCC1973 and BCC1990 did not diminish in concentration but stayed the same when incubated for 3 or 7 days. Interestingly, both *B. ubonensis* strains produce a metabolite with retention time 6.73 min (Figure 5.5), which was suspected to be a different isomer of enacyloxin based on their UV spectra. No novel metabolites could be detected using HPLC analysis of spent growth media supplemented with carbon sources other than glycerol (Figure 5.2).

HPLC analysis was carried out to investigate if the concentration of the metabolites identified in *Burkholderia* sp. strains BCC1970 and BCC1977 (Figure 5.4) responded to changes in temperature and incubation period (Figure 5.5). The peak area for the metabolite in strain BCC1977 seen at retention time 7.32 min (Figure 5.5; panel a), decreased over time from 9162 AU (Day 2) to 6070 AU (Day 3) when incubated at 22°C. The peak area for the metabolite from strain BCC1970 seen at retention time 7.88 min (Figure 5.5; panel a) also decreased over time from 13094 AU (Day 2) to 315 AU (Day 3) when incubated at 22°C. Analysis of the UV spectra for the compound at retention time 7.88 min revealed 3 peaks with spectral characteristics associated with polyyne compounds (Figure 5.5; panel b) such as cepacin (Mullins *et al.,* 2019), collimonins (Kai *et al.,* 2018) and caryoynencin (Ross *et al.*, 2014). LC-MS analysis of total metabolite extracts from strain BCC1970 showed that N-decanoyl-L-homoserine lactone and the polyynes collimonins C and D were present in spent growth media when it was incubated at 22°C for 2 days (Figure 5.6). The observed mass ions, and mass spectrum for the putative compounds correspond to the proposed structures based on the observed and theoretical mass(es) for these compounds (Figure 5.6).

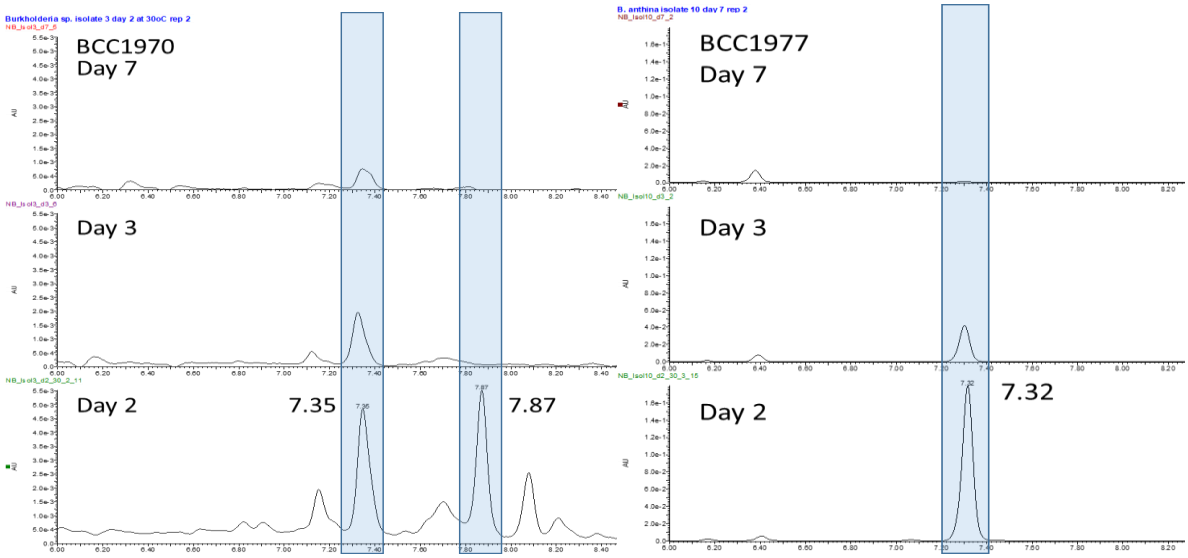



**Figure 5.4. HPLC metabolite traces at 300 nm for the *Burkholderia* sp. strains BCC1970 and BCC1977 when incubated for 2, 3 or 7 days.**
Metabolites with $\lambda_{max}$=300 nm were detected for strains BCC1970 and BCC1977 when grown on BSMG at 30 °C. Retention times in minutes are indicated next to the largest peaks. The ordinate axes have the same range for each day to show the difference in concentration. n=3 independent HPLC analyses of BCC1970 and BCC1977 cultures. Dr Gordon Webster performed the HPLC analyses.

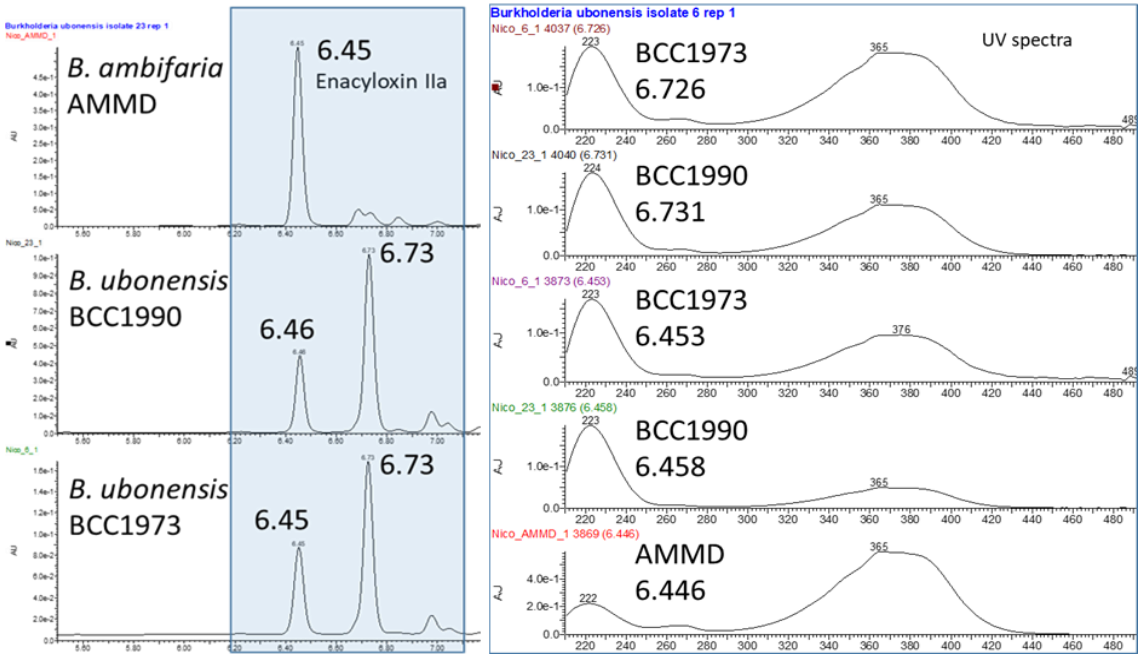



**Figure 5.5. HPLC metabolite traces at 360 nm for the *Burkholderia ubonensis* environmental strains compared to *B. ambifaria* AMMD.**
Scanning of metabolite extracts at 360 nm indicated chromatogram patterning similar to *B. ambifaria* enacyloxin IIa for environmental strains BCC1973 and BCC1990 when grown on BSMG at 30°C for 3 days. Absorbance spectra for each peak is shown on the right, with the retention times in minutes relating to each peak below the strain identification; n=3 independent HPLC analyses of BCC1973 and BCC1990 cultures. Enacyloxin IIa from *B. ambifaria* AMMD has retention time of 6.45 min that was closely mirrored by the environmental strains. A metabolite peak at 6.73 min was also seen in the Mysore collection strains and suspected to be an isomer of enacyloxin IIa. Dr Gordon Webster performed the HPLC analyses.

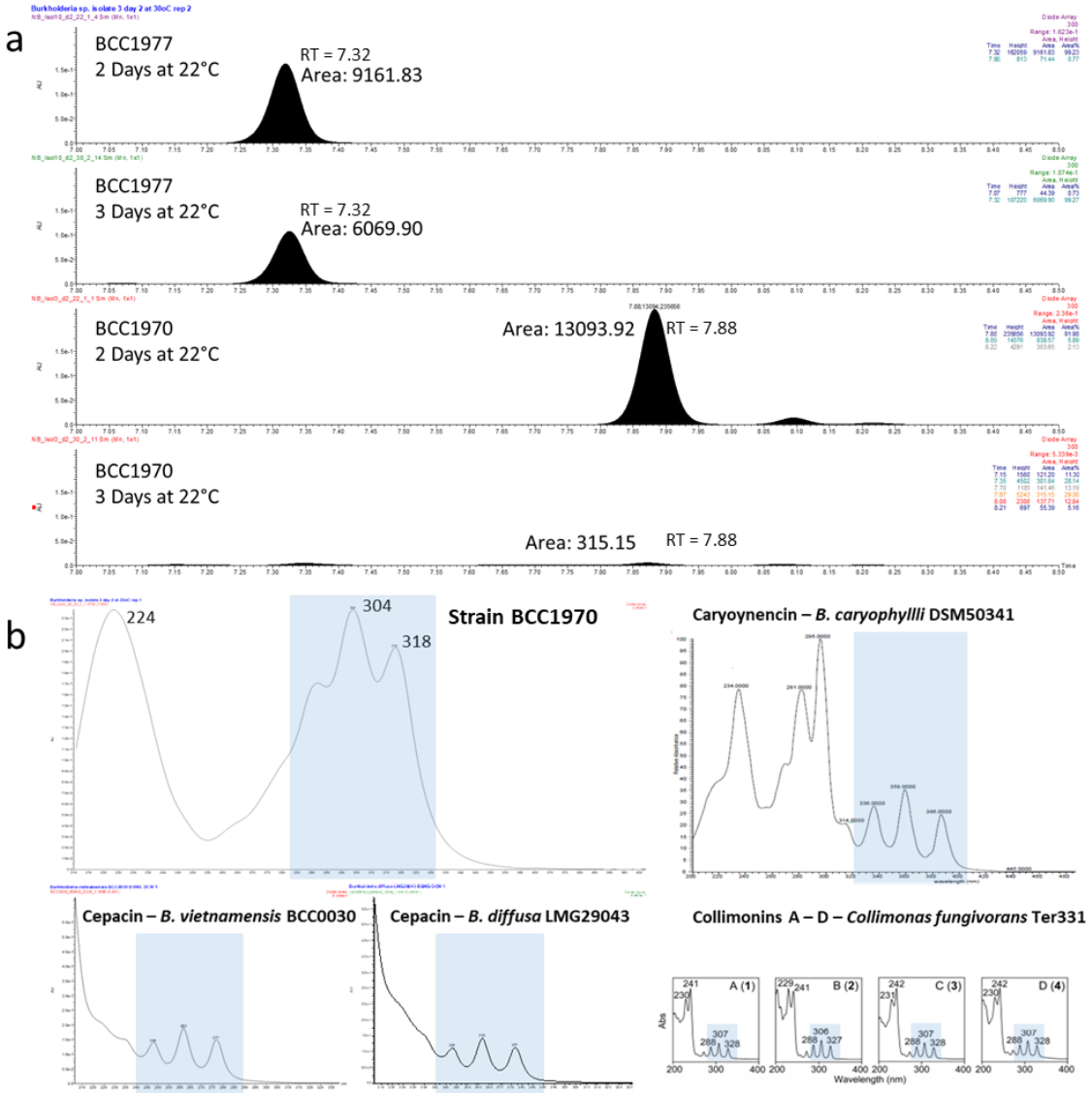



**Figure 5.6. HPLC metabolite traces (at 300 nm) and peak UV spectra for *Burkholderia* sp. BCC1970.** **(a)** Metabolites detected for strains BCC1970 and BCC1977 when grown on BSMG measured at 300 nm. The retention time (RT) in minutes of each peak is written next to the curve for which the area underneath was calculated. The ordinate axes have the same range for each sample. **(b)** The UV spectra of the compound with retention time 7.88 min for strain BCC1970 exhibits 3 small peaks, highlighted in blue, characteristic for polyyne compounds. The UV spectra of cepacin in *B. vietnamensis* and *B. diffusa* are shown with the 3 peaks characteristic for polyynes highlighted in blue (Dr Gordon Webster - unpublished data). The UV spectra for caryoynencin and collimonins A – D are shown with the 3 characteristic peaks for the polyynes highlighted in blue. Dr Gordon Webster performed the HPLC analyses. UV spectrum of caryoynencin was adapted from: Ross *et al.* (2014). UV spectra of collimonins A – D was adapted from: Kai *et al.,* (2018).

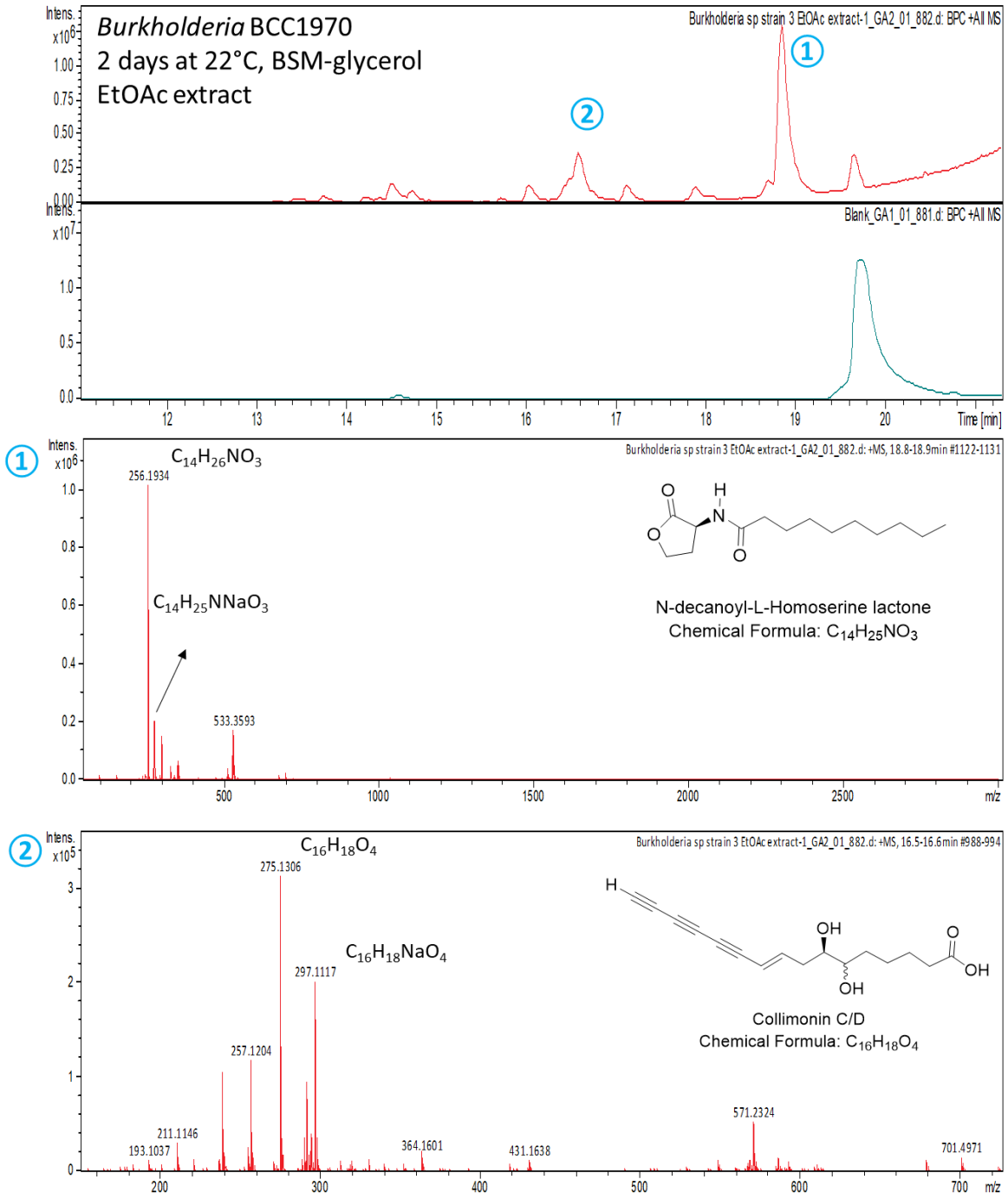



**Figure 5.7. Detection of homoserine lactone and collimonins C/D using LC-MS analysis of *Burkholderia* sp. strain BCC1970.**

Extracted ion chromatograms with the structure based on observed mass ions, and mass spectrum for the homoserine lactone and collimon C/D compounds in strain BCC1970. The green trace represents a blank sample used to identify and subtract background noise from the culture sample trace. UHPLC-ESI-Q-TOF-MS analyses were performed by Dr Jinlian Zhao at the University of Warwick, who also provided the putative compound structures associated with the mass ion peaks detected.

### 5.2.5. Isolating enacyloxin isomers from *B. ubonensis* to investigate their antimicrobial activity

HPLC analysis of *B. ubonensis* strains BCC1973 and BCC1990 identified compounds with maximum absorbance wavelengths similar to known enacyloxin compounds (Figure 5.5). LC-MS analysis confirmed strains BCC1973 and BCC1990 to have identical metabolite profiles (Figure 5.7; panel a). The mass ions, and mass spectrum of enacyloxin IIa and IVa compounds were detected and corresponded to the proposed structures based on the observed and theoretical mass(es) for these compounds (Figure 5.7; panel b). HPLC analysis of strains BCC1973 and BCC1990 indicated the enacyloxin-like compound with retention time 6.73 min (Figure 5.5) to be produced in higher concentrations to the compound with similar retention time as enacyloxin IIa (Figure 5.5; 6.45 min).

Enacyloxin IVa is a known intermediate in the biosynthesis of enacyloxin IIa in *Frateuria* sp. W-135 (Oyama *et al.*, 1994). The difference in chemical structure between enacyloxin IIa and IVa is the oxidative state of C-15' (Figure 5.7; panel b), with enacyloxin IIa forming following the dehydrogenation of enacyloxin IVa (Oyama *et al.,* 1994). To investigate the antimicrobial activity of enacyloxin IVa, *B. ubonensis* strain BCC1990 was grown in bulk and the suspected enacyloxin IVa compound purified using preparative HPLC (Figure 5.8; panel a). The purified compound peaks were analysed using LC-MS to confirm their identity prior to the antimicrobial MIC assay (Figure 5.8). The observed mass ions, and mass spectrum of the compounds in fractions 2b and 3 (Figure 5.8; panel b) both corresponded to the proposed structures based on the observed and theoretical mass(es) of enacyloxin IVa (Figure 5.7). Fraction 2b with the highest purity of enacyloxin IVa was subsequently used for the MIC assay.

### 5.2.6. Minimum inhibitory concentration of purified enacyloxin IVa

To investigate the efficacy of enacyloxin IVa as an antimicrobial, the purified compound from fraction 2b (Figure 5.8) was investigated against a panel of plant pathogens (Mullins *et al.*, 2019) and medically relevant strains (Mahenthiralingam *et al.*, 2011) that had been tested previously against enacyloxin IIa. The putative enacyloxin IVa inhibited the growth of *Acinetobacter baumannii* OXA23 clone 2 and *B. multivorans* strain ATCC17616 at lower MICs, >3-fold and >4-fold decrease in metabolite concentration respectively, than reported for enacyloxin IIa (Table 5.1). *C. albicans* SC5314 and *S. aureus* NCTC 12981 did not respond to the concentrations of the putative enacyloxin IVa investigated. The putative enacyloxin IVa, at a concentration ranging between 6.25 to 12.5 µg/ml, caused an 80% knock down in optical density in *P. carotovorum.* Interestingly, the concentration reported by Mullins *et al.* (2019) at which enacyloxin IIa causes an 80% knock down in optical density in *P. carotovorum* falls within this range.

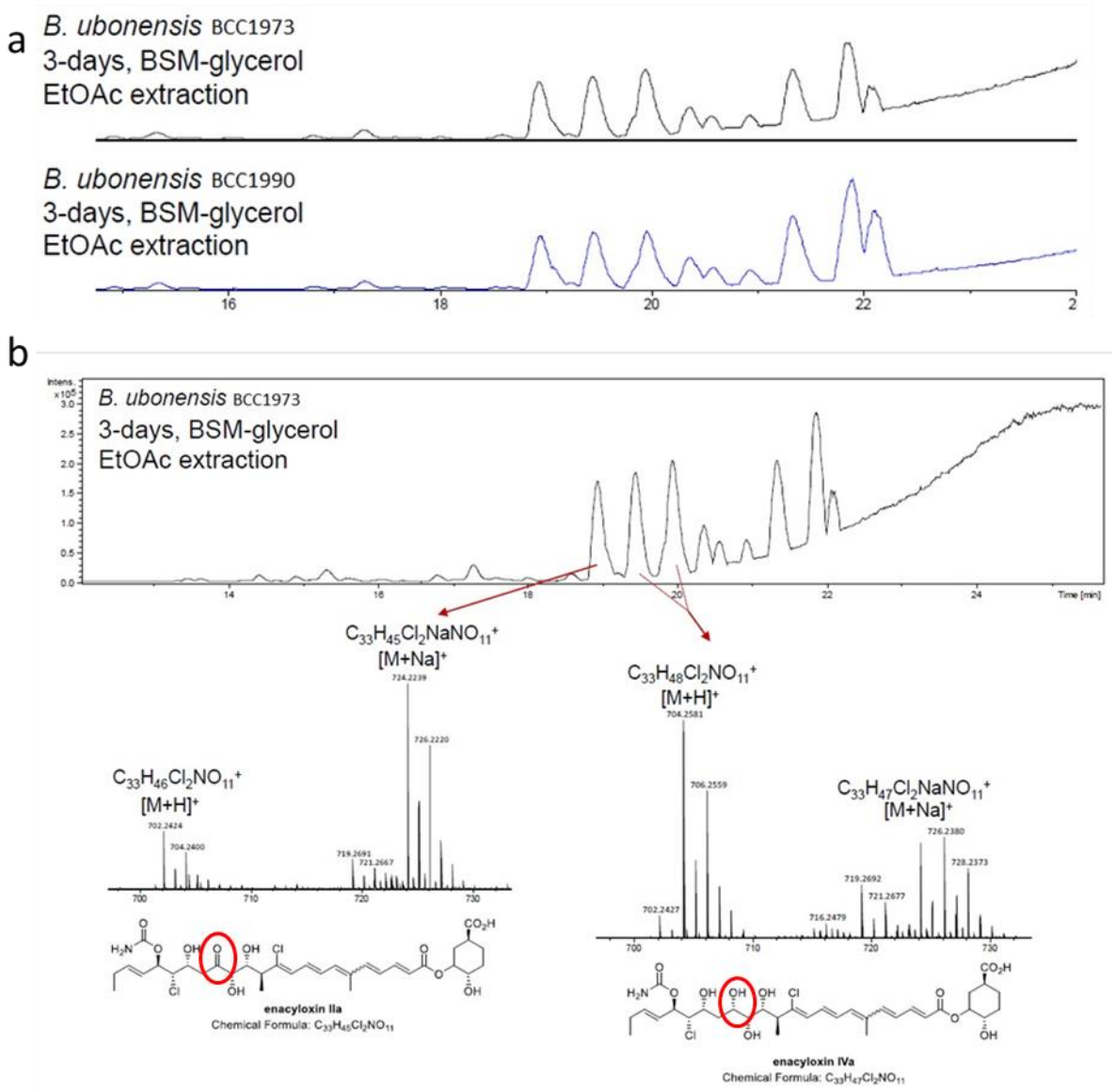



**Figure 5.8. Detection of enacyloxin compounds in *B. ubonensis* strains BCC1973 and BCC1990 using LC-MS analysis.**

**(a)** Chromatograms for BCC1973 and BCC1990 following UHPLCESI-Q-TOF-MS analysis showing identical metabolic profiles for strain BCC1973 and BCC1990. **(b)** Extracted ion chromatograms with the structure based on observed mass ions, and mass spectrum for the enacyloxin compounds in strain BCC1973. The C-15' has been circled in red to illustrate the difference in its oxidative state between the two compounds. UHPLC-ESI-Q-TOF-MS analyses were performed by Dr Matthew Jenner at the University of Warwick, who also provided the putative compound structures associated with the mass ion peaks detected.

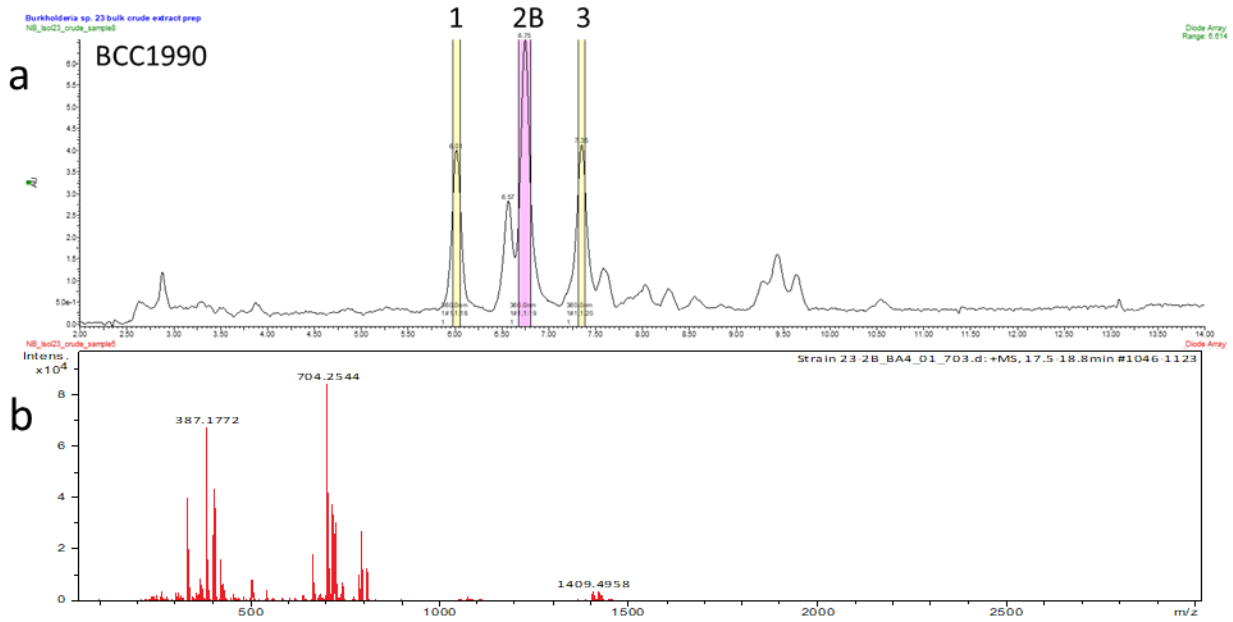



**Figure 5.9. Preparative HPLC chromatograms and LC-MS spectra showing the purification of suspected enacyloxin isomers from *B. ubonensis* strain BCC1990.**
**(a)** Enacyloxin compounds produced by strain BCC1990 were extracted and purified from BSMG using preparative HPLC. Fraction 2b was considered to have the highest purity. (**b)** Extracted ion chromatograms of fraction 2b at m/z 704.2544±0.01 correspond to [M+H]+=704.2599 for enacyloxin IV. Dr Gordon Webster performed the HPLC analyses and assisted in purifying the enacyloxin metabolite.

**Table 5.1. MIC assay results with previously reported MICs for enacyloxin IIa and the MICs for enacyloxin IVa purified from *B. ubonensis* strain BCC1990.**

| | | Minimal Inhibitory Concentration (µg/ml) | |
|---|---|---|---|
| **Organism** | **Source[a]** | **Enacyloxin IIa** | **Enacyloxin IVa** |
| *Staphylococcus aureus* 209 P | Prezioso *et al.* 2017 | 50 | |
| *Staphylococcus aureus* NCTC 12981 | | | >100 |
| *Pectobacterium carotovorum* LMG 2464 | Mullins *et al.* 2019 | 6.3 | 6.3 – 12.5 |
| *Candida albicans* SC 5314 | Mahenthiralingam *et al.* 2011 | >100 | >100 |
| *Acinetobacter baumannii* OXA23 clone 2 | Mahenthiralingam *et al.* 2011 | 3.0 | 0.4 – 0.8 |
| *Burkholderia multivorans* ATCC 17616 | Mahenthiralingam *et al.* 2011 | 6.5 | 0.8 – 1.5 |

[a] The MIC of enacyloxin IIa for the assayed organisms as reported in these papers.

### 5.2.7. Genome mining for *in vitro* expressed antimicrobial biosynthetic pathways

To correlate the detected metabolite production with biosynthetic pathways in the bioactive *Burkholderia* strains, AntiSMASH (v5.1.1) (Blin *et al.,* 2019) analysis on their draft genome sequences was performed (Figure 5.9). In default mode, antiSMASH v5.1.1. (Blin *et al.*, 2019) predicted 12, 18, 12 and 18 secondary metabolite gene clusters for strains BCC1970, BCC1973, BCC1977 and BCC1990, respectively. A hybrid NRP and polyketide pathway was predicted in *B. ubonensis* strains BCC1973 and BCC1990 with close homology to the *B. ambifaria* AMMD enacyloxin IIa pathway. All the genes characterised for the enacyloxin IIa gene cluster in *B. ambifaria* AMMD shared homologs with strains BCC1973 and BCC1990 (Figure 5.9). BLASTP analysis of the translated amino acid sequence for each gene queried ranged between 78 – 100% when compared against the enacyloxin IIa pathway gene products from *B. ambifaria* AMMD. The coverage of each AA sequence ranged between 48% – 100%, however when excluding the sequence with 48% coverage the total ranged between 89% – 100% coverage.

AntiSMASH analysis of the genomes of *B. ubonensis* strains BCC1973 and BCC1990 also predicted a hybrid NRP and polyketide pathway with close homology to the occidiofungin A biosynthetic gene cluster associated with *B. pyrrocinia* strain Lyc2 (Wang *et al*., 2016) (Figure 5.9). The genes from the query sequence for strain BCC1990 shared homologs with 76% of the genes in the cluster (Figure 5.9; panel b). For strain BCC1990 the AA sequence identity for each gene product ranged between 79% - 96% with the sequence coverage ranging between 99 – 100% when compared against the occidiofungin A pathway gene product AA sequences. For strain *B. ubonensis* strain BCC1990 only 4 genes found on the original occidiofungin A cluster (Wang *et al.*, 2016) were not identified, these included a *luxR* family transcriptional regulator, a hypothetical protein, *luxR* family transcriptional regulator and an oleoyl-ACP hydrolase.

For *B. ubonensis* strain BCC1973, 52% of the genes from the occidiofungin A BGC showed similarity, however, the cluster appeared to be split over two contigs (Figure 5.9; panel a). Nevertheless, at least 9 out of the 17 genes shared homology with those found in *B. pyrrocinia* strain Lyc2 (Wang *et al.*, 2016). For strain BCC1973 the AA sequence identity for each gene product ranged between 87% - 96% with the sequence coverage ranging between 99 – 100% when compared against the occidiofungin A gene product AA sequences. No homologs for the following genes could be identified oleoyl-ACP hydrolase, amino acid adenylation protein, glycosyl transferase, hypothetical protein, ATP-binding protein, *luxR* family transcriptional regulator, hypothetical protein and *luxR* family transcriptional regulator.

AntiSMASH analysis of *Burkholderia* sp. strain BCC1970 (Figure 5.10) predicted a BGC with homologs to 68% of the genes from the polyyne cepacin A cluster (Mullins *et al.*, 2019) present in *Burkholderia ambifaria* IOP40-10. Additionally, the BGC shared gene homologs with 38% of the genes from the polyyne caryoynencin BGC of *Burkholderia gladioli* BSR3. The following genes found on the cepacin A cluster were not identified in strain BCC1970: flavin-dependent monooxygenase, MFS transporter, 2 fatty acid desaturases and rubredoxin (Figure 5.10). Nevertheless, 11 gene homologs to the cepacin A cluster have been identified. The AA sequence identity for each gene product ranged between 88 - 96% with the sequence coverage ranging between 93 – 100% when compared against the AA sequence of the cepacin A cluster gene products. The query sequence sits on a contig edge possibly excluding several genes which are involved in the BGC predicted in strain BCC1970.

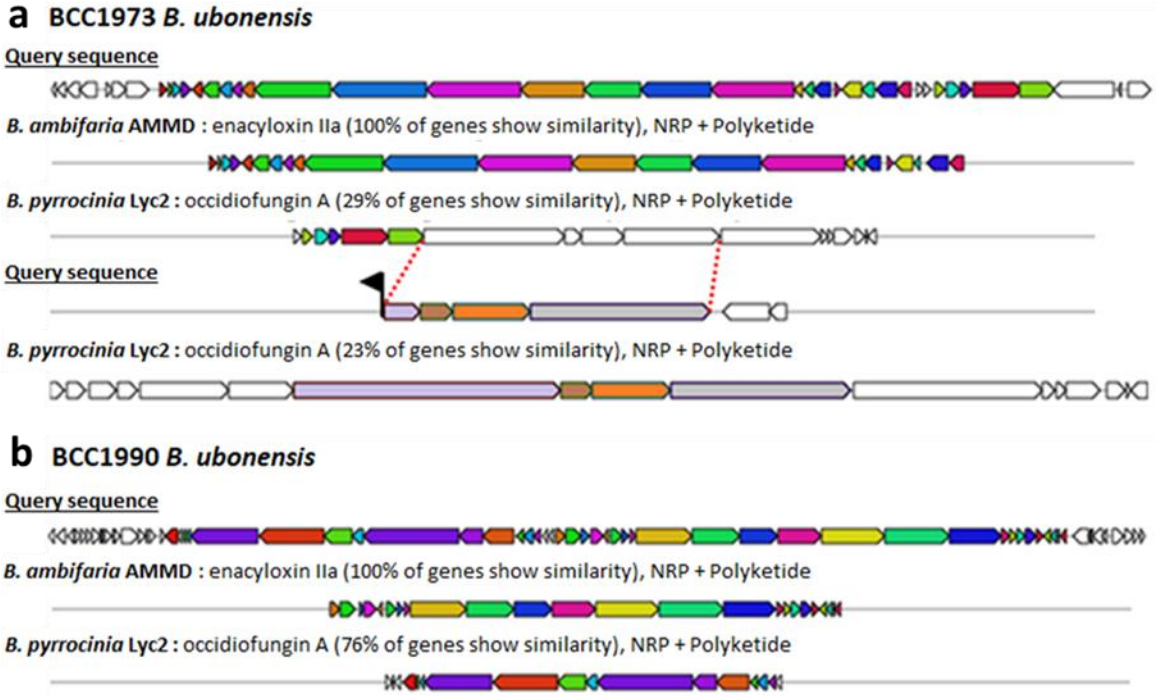



**Figure 5.10. Comparison of the predicted BGCs in *B. ubonensis* strains BCC1973 and BCC1990 to the enacyloxin IIa BGC identified in *B. ambifaria* AMMD and the occidiofungin A BGC identified in *B. pyrrocinia* Lyc2.**
Graphic was generated using antiSMASH v5.1.1. using the KnownClusterBlast function and edited using Paint v.1083 (Microsoft). The query sequence includes all the genes located together on a contig which were used to search the Minimum Information on Biosynthetic Gene cluster (MiBig) database. The genes are colour coded according to the gene homolog identified in *B. ambifaria* AMMD and *B. pyrrocinia* Lyc2. The black flag indicates the contig edge.

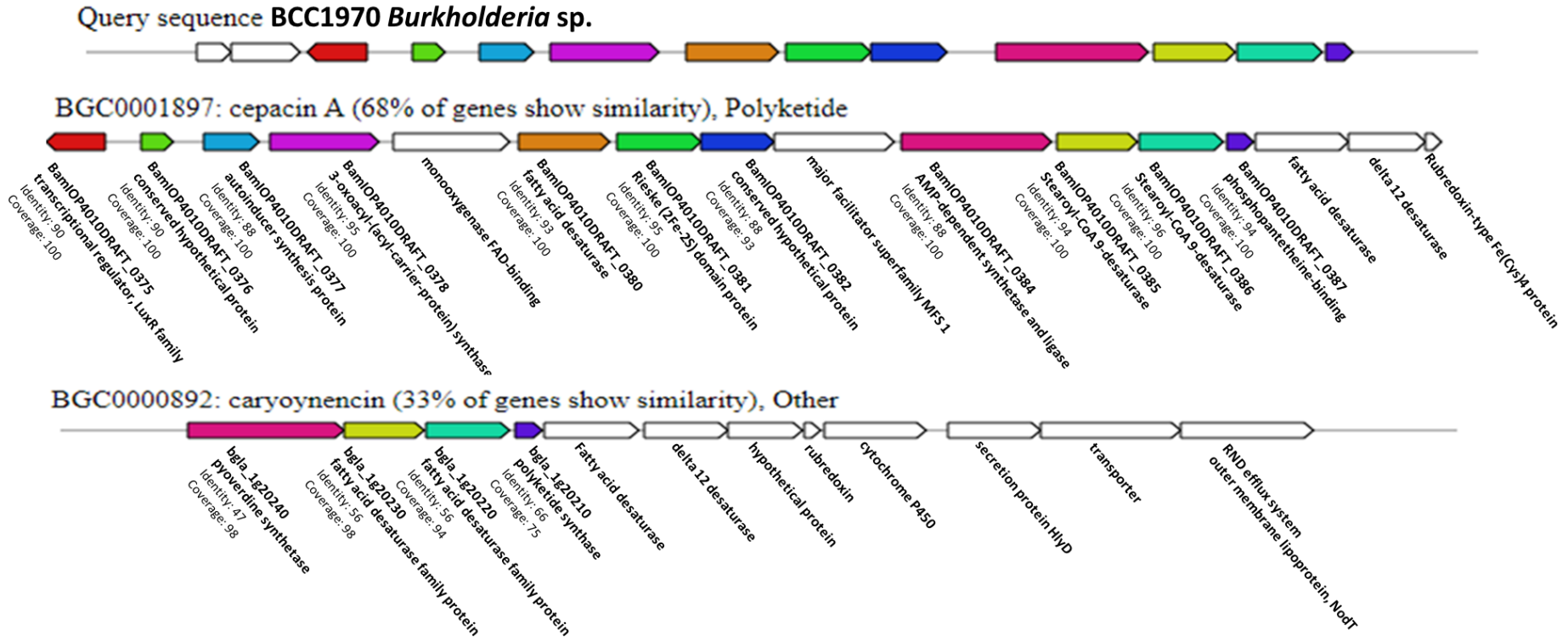



**Figure 5.11. Comparison of the BGC predicted in *Burkholderia* sp. strain BCC1970 to the cepacin A BGC identified in *B. ambifaria* IOP40-10 and the caryoynencin BGC in *Burkholderia gladioli* BSR3.**

Graphic was generated using antiSMASH v5.1.1. using the KnownClusterBlast function and edited using Powerpoint v.2002 (Microsoft). The query sequence includes all the genes located together on a contig which were used to search the Minimum Information on Biosynthetic Gene cluster (MiBig) database. The genes are colour coded according to the predicted gene homologs in *B. ambifaria* IOP40-10 and *B. gladioli* BSR3. The translated amino acid sequence of each gene in the query sequence was aligned using BLASTP to the translated amino acid sequence of the genes from *B. ambifaria* IOP40-10 and *B. gladioli* BSR3. The putative protein product, the gene identifier and the BLASTP similarity scores are given for each gene.

## 5.3. Discussion

The following section discusses the antimicrobial screening assays, strategies to optimize antimicrobial production, chemical analysis of culture extracts and genome mining analysis utilized to reveal the biosynthetic diversity and antimicrobial activity of the Mysore *Burkholderia* collection. In order to identify strains with antimicrobial activity, the strains from the Mysore *Burkholderia* collection were screened against a panel of Gram-positive, Gram-negative and fungal pathogens (Figure 5.1.). *Burkholderia* strains with interesting antimicrobial activity were selected for further investigation and screened for activity under a variety of different growth conditions (Figure 5.2 and 5.3). To identify antimicrobial metabolites, spend culture extracts were analysed by HPLC (Figure 5.4.; 5.5; 5.6; 5.9) and LC-MS (Figure 5.6.; 5.8; 5.9). Putative antimicrobial compounds were found to be structurally related to compounds previously identified in *Burkholderia* and subsequent genome mining predicted BGCs associated with known metabolites detected by HPLC and LC-MS analysis (Figure 5.9. and 5.10.).

### 5.3.1. Screening the entire Mysore *Burkholderia* collection

The entire Mysore *Burkholderia* collection (n=35) was investigated for antimicrobial activity against previously characterized human and plant pathogens (Table 2.7.). Antimicrobial activity was observed for 26 strains (Figure 5.1) and included 13 strains with antifungal activity (50%), 3 strains with anti-Gram-negative activity (12%) and 20 strains with anti-Gram-positive activity (77%). The 10 *B. cepacia* strains recovered from the Mysore environment were the least antimicrobial bioactive strains with only weak anti-Gram-positive activity (Figure 5.1) correlating to the lack of activity seen in the species in previous studies (Mahenthiralingam *et al.,* 2011). The most antimicrobially bioactive strains were *B. ubonensis* and the cluster of novel *Burkholderia* complex species (Figure 5.1). *Pseudomonas mesoacidophila* ATCC 31433 was recently reclassified as *B. ubonensis* and shown to encode multiple specialised metabolites (Loveridge *et al.,* 2017). The antimicrobial activity of the environmental *B. ubonensis,* and the identification of multiple specialized metabolite BGCs in strain ATCC 31433, suggests the *B. ubonensis* species is a potential source for the discovery of novel bioactive compounds.

The largest zones of inhibition were observed when the environmental *Burkholderia* were tested against the fungal pathogen *C. albicans* (SC 5314) (Figure 5.1). The antagonism of fungi by *Burkholderia* has been well investigated and many antifungal compounds from *Burkholderia* have been described (Vial *et al.,* 2007; Kunakom and Eustáquio, 2019). *Burkholderia,* and most fungi, prefer soils with an acidic pH and previous work has shown that members of the *Burkholderia* genus often co-occur with fungi in the soil environment (Stopnisek *et al.,* 2013; Stopnisek *et al.,* 2015). Competition with fungi in the soil environment provides a positive selective pressure for antifungal activity which would explain

the high prevalence of antifungal activity observed across the Mysore *Burkholderia* collection (Figure 5.1).

In addition to their close relationships with fungi, close association of *Burkholderia* with plants is also a widely reported feature of these bacteria (Compant *et al.,* 2008b; Suarez-Moreno *et al.,* 2012; Wallner *et al.,* 2019). *Burkholderia* bacteria can utilise a wide variety of carbon sources including plant exudates, which have been shown to influence bacterial community structure (Doornbos *et al.,* 2011; Kost *et al.,* 2014). From the Mysore *Burkholderia* collection 21 strains were recovered from the rhizosphere of medicinal plants demonstrating a close relationship with plants. To investigate if plant exudates could elicit novel or enhance antimicrobial production in the Mysore *Burkholderia* collection, wheat or pea seed exudates were extracted and incorporated into an agar-based medium, with no further nutrient supplements. However, wheat and pea seed exudate media did not elicit novel or enhance antimicrobial production compared to BSMG and was therefore not investigated further (data not shown). Koch *et al.* (2002) demonstrated that the addition of sugar beet seed exudate to minimal media enhanced antimicrobial production in the sugar beet rhizosphere isolate *Pseudomonas* sp. DSS73, compared to media lacking the seed exudate. The bacteria and plant host relationship should be considered in future growth media design as perhaps incorporating exudates from associated medicinal plants could have elicited or enhanced antimicrobial production in the Mysore *Burkholderia* collection.

### 5.3.2. Investigating growth conditions to elicit novel or enhance antimicrobial production

Following screening of the Mysore *Burkholderia* collection, a panel of 13 *Burkholderia* strains (Table 2.7), including 4 selected environmental strains and their third replicon deletion mutants (BCC1970, BCC1973, BCC1977, BCC1990), was assembled to investigate if alterations to culture conditions could enhance or induce novel antimicrobial production. Antimicrobial activity was determined by measuring the diameter of the zones of inhibition and was considered novel or altered in relation to the antimicrobial activity observed after growth on BSMG as a control. To investigate if specific root exudate components could enhance antimicrobial production, like sugars and amino acids, which were reported to be upregulated by maize in nutrient limited soil, were utilized as the sole carbon source in minimal media (Carvalhais *et al.,* 2010). *Burkholderia cepacia* complex bacteria have been frequently isolated from the maize rhizosphere and the suppression of soilborne fungal and oomycete pathogens by maize root associated *Burkholderia* has been reported, which provided the rational for utilizing maize associated root exudates (Hebbar *et al.,* 1992; Hebbar *et al.,* 1998; Bevivino *et al.,* 2011).

For the 4 Mysore *Burkholderia* strains investigated, glucose induced novel Gram-positive antagonism for *B. ubonensis* strains BCC1973 and BCC1990. Additionally, adonitol and glucose induced novel but weak activity against *S. aureus* (≤ 9 mm diameter) for *Burkholderia* sp. strain BCC1977 (Figure 5.2). However, HPLC analysis of culture extracts did not detect any novel bacterial metabolites when compared to the HPLC data of culture extracts from BSMG (data not shown). Glutamate induced a statistically significant increase in antimicrobial activity relative to glycerol when *Burkholderia* sp. BCC1970 ($p = 0.0007$) and *B. gladioli* BCC0238 ($p = 0.0005$) were tested against *C. albicans* (Figure 5.2.). Glutamate is a key precursor in the synthesis of specialized metabolites, such as pyrrolnitrin and the addition of L-glutamic acid to culture media has been shown to increase pyrrolnitrin production in *Burkholderia* (Pawar *et al.,* 2019). Whether glutamate is a key precursor for the antimicrobial compounds produced by BCC0238 and BCC1970 remains to be investigated. Nevertheless, the increased antimicrobial activity observed for strains BCC0238 and BCC9170 (Figure 5.2.) suggests glutamate could enhance antimicrobial production in some *Burkholderia* strains.

In addition to media components, incubation temperature and time has also been demonstrated to influence antimicrobial activity in *Burkholderia*. For example, El-Banna and Winkelmann (1997) reported *B. cepacia* strain NB-1 to reach maximal antimicrobial activity after 120 hours. Funston *et al.* (2016) demonstrated increased rhamnolipid production in *Burkholderia thailandensis* E264 when grown at 25˚C compared to 30˚C. Thus, three different incubation temperatures (22°C, 30°C and 37°C) and incubation periods of 2, 3 and 7 days were investigated for their effect on antimicrobial production. Antimicrobial activity of the 4 environmental strains (BCC1970, BCC1973, BCC1977, BCC1990) following incubation at 37°C for 3 days or 7 days at 30°C was not presented as neither growth conditions elicited novel antimicrobial activity. Antimicrobial activity against *C. albicans* and *P. carotovorum* by *B. ubonensis* strains BCC1973 and BCC1990, in terms of the diameter of the zone of inhibition, was the greatest after 3 days of incubation at 30°C (Figure 5.3.). In contrast, antimicrobial activity of *Burkholderia* sp. strains BCC1970 and BCC1977 against *C. albicans* was the greatest following a 2-day incubation period at 22°C (Figure 5.3.). These data indicate that a variety of culture conditions should be considered to identify antimicrobial products from *Burkholderia*. Nevertheless, the culture conditions which generated the largest zones of inhibition were adopted to obtain maximum metabolite yields for HPLC and LC-MS analysis.

### 5.3.3. Chemical analysis and genome mining for antimicrobial discovery

HPLC analysis of culture extracts from the novel *Burkholderia* sp. strains BCC1970 and BCC1977 identified metabolites with unique maximum absorbance wavelength $\lambda_{max}$=300 nm (Figure 5.4.; Figure 5.5.). The concentrations of the detected metabolites inferred from the HPLC peak sizes, after

incubation at either 22˚C (Figure 5.5.) or 30˚C (Figure 5.4.), was the greatest following a 2-day incubation period. This correlated to when the zones of clearing were the largest (Figure 5.3.) and were therefore suspected as the metabolites responsible for the observed antimicrobial activity. LC-MS analysis of culture extracts from *Burkholderia* sp. strain BCC1977 identified compounds with similar mass ion spectra as 2,4-diacetylphloroglucinol and pyoluteorin. However, no associated BGCs were identified for these metabolites using antiSMASH (Blin *et al.*, 2019) and because only limited chemical analysis had been performed to support these findings it was not presented. LC-MS analysis of culture extracts from strains *Burkholderia* sp. strain BCC1970 (Figure 5.6.) identified compounds with similar mass ion spectra as N-decanoyl-L-homoserine, a member of N-acyl-homoserine lactone (AHL) family (Vial *et al.*, 2007), and the polyynes collimonins C/D (Fritsche *et al.,* 2014; Kai *et al*., 2018). In addition, HPLC analysis of the metabolite with retention time 7.88 min (Figure 5.5.) revealed three UV absorbance maxima peaks, a pattern characteristic for polyynes such as collimonins (Kai *et al.*, 2018) and caryoynencins (Ross *et al.*, 2014). No antimicrobial activity for N-decanoyl-L-homoserine lactone has been reported however these quorum sensing (QS) molecules and related QS systems have been shown to regulate the biosynthesis of several antimicrobials in *Burkholderia* such as enacyloxin (Mahenthiralingam *et al.,* 2011), cepacin (Mullins *et al.,* 2019) and fragin (Jenul *et al.,* 2018).

The collimonins are characterized as antifungal compounds and explains the antagonism of *C. albicans* by BCC1970 (Fritsche *et al.,* 2014; Kai *et al*., 2018). Collimonins, and its associated BGC originally designated cluster K, were first identified in Gram-negative bacterium *Collimonas fungivorans* Ter331 (Fritsche *et al.,* 2014; Kai *et al.,* 2018). Currently no collimonin BGC has been submitted to the MiBig database (https://mibig.secondarymetabolites.org/) which antiSMASH uses to predict BGCs (Blin *et al.*, 2019). Nevertheless, Fritsche *et al.* (2014) reports a near-complete orthologous version of cluster K in *Burkholderia ambifaria* IOP40-10, which was later predicted to encode cepacins (Mullins *et al.,* 2019). Genome mining of *Burkholderia* sp. strain BCC1970 identified a BGC (Figure 5.10) with homology to 68% of the genes from the *B. ambifaria* strain IOP40-10 cepacin pathway (Mullins *et al.*, 2019). The predicted BGC in BCC1970 was located on a contig edge which could have excluded several genes from the query sequence. Nevertheless, the translated amino acid sequences of 10 genes from the predicted BGC in BCC1970 shared 88% to 96% identity match over 93% to 100% of their length with the cepacin pathway (16 genes) from *B. ambifaria* IOP40-10 (Figure 5.10). Cepacin A and B are reported to have strong anti-staphylococcal activity (Parker *et al.,* 1984) however strain BCC1970 did not exhibit antimicrobial activity against *Staphylococcus aureus* to correlate to this finding (Figure 5.1; Figure 5.2). Overall, the close association of the predicted *Burkholderia* sp. strain BCC1970 BGC to the

collimonin BGC (Figure 5.10), the polyyne metabolites detected by HPLC (Figure 5.5) and LC-MS (Figure 5.6) analysis suggest BCC1970 is a collimonin producer.

Various polyynes have been found in nature however few polyynes have been described from bacteria (Kai *et al.,* 2018). So far, only the following biosynthetic pathways for bacterial polyynes have been elucidated: cepacins (Mullins *et al.,* 2019), caryonencins (Ross *et al.,* 2014) and collimonins (Fritsche *et al.*, 2014; Kai *et al.,* 2018) which have been reported to share a high degree of similarity (Fritsche *et al.*, 2014; Ross *et al.* 2014; Mullins *et al.*, 2019). Currently, no collimonins have been described from *Burkholderia,* although closely related pathways to cluster K have been identified in other *Burkholderia* and *Pseudomonas* species (Fritsche *et al.,* 2014). Nevertheless, the identification of collimonins from *Burkholderia* sp. strain BCC1970 provides further insight into the biosynthetic diversity of *Burkholderia* species and distribution of bacterial polyyne BGCs. To fully characterise the polyyne metabolite and BGC, the complete genome sequence for *Burkholderia* sp. strain BCC1970 should be determined to understand the full extent of its biosynthetic pathway, and whether this can be correlated to its production of collimonin.

HPLC analysis of the culture extracts from *B. ubonensis* strains BCC1973 and BCC1990 (Figure 5.5.) identified compounds with retention times and maximum absorbance wavelengths characteristic for the enacyloxin antibiotic produced by *B. ambifaria* (Mahenthiralingam *et al.,* 2011). LC-MS analysis confirmed *B. ubonensis* strains BCC1973 and BCC1990 to have identical metabolite profiles (Figure 5.7.; panel a), and to produce compounds with mass ions spectra predictive of enacyloxin IIa and enacyloxin IVa (Figure 5.7.; panel b). Enacyloxin IIa and IVa have potent anti-Gram-negative activity (Oyama *et al.,* 1994; Mahenthiralingam *et al.*, 2011) which would explain the strong antagonism of *P. carotovorum* by *B. ubonensis* strains BCC1973 and BCC1990 (Figure 5.1; panel b). Interestingly, HPLC analysis demonstrated enacyloxin IVa to be produced in higher amounts than enacyloxin IIa, which was inferred from the differences in HPLC peak heights (Figure 5.5.; Figure 5.8.; panel a). Oyama *et al.* (1994) reports a pyrroloquinoline quinone (PQQ)-dependent extracellular enzyme to catalyse the conversion of enacyloxin IVa to enacyloxin IIa in *Frateuria* sp. W-315. Mahenthiralingam *et al.* (2011) propose a similar PQQ-dependent enzyme encoded by Bamb_5932 to convert enacyloxin IVa to enacyloxin IIa in *B. ambifaria* AMMD, a strain which produces mainly enacyloxin IIa. Genome mining of *B. ubonensis* strains BCC1973 and BCC1990 (Figure 5.9.) identified a complete enacyloxin IIa BGC orthologous to the prototypic pathway identified in *B. ambifaria* AMMD, including homologs for strains BCC1973 and BCC1990 which share 92% amino acid sequence identity with Bamb_5932 (100% coverage) (Mahenthiralingam *et al.,* 2011). Perhaps strains BCC1973 and BCC1990 produce more enacyloxin IVa than IIa because their putative PQQ-dependent enzymes are less efficient than the enzyme encoded by Bamb_5932, however, this remains to be investigated. Nevertheless, the

identification of the enacyloxin IIa BGC and enacyloxin compounds in the culture extracts confirms strains BCC1973 and BCC1990 as enacyloxin producers, uniquely linking this metabolite and BGC to *B. ubonensis* for the first time.

*B. ubonensis* strains BCC1973 and BCC1990 also exhibited strong antifungal activity when tested against *C. albicans* (Figure 5.3.) however neither enacyloxin IIa nor IVa inhibited growth of *C. albicans* when tested as pure compounds purified from producing strains (Table 5.1.). Genome mining of BCC1973 and BCC1990 (Figure 5.9.) predicted BGCs with close homology to the occidiofungin A BGC from *B. pyrrocinia* Lyc2 as characterized by Wang *et al.* (2016). Although production of occidiofungin, a known antifungal compound, would explain the antifungal activity observed for strains BCC1973 and BCC1990, LC-MS analyses of culture extracts could not detect mass ions spectra predictive of occidiofungin. Nevertheless, antiSMASH (Blin *et al.,* 2019) analyses identified a near complete orthologous version of the original occcidiofungin BGC in strain BCC1990 (Figure 5.9.; panel a). Furthermore, antiSMASH (Blin *et al.,* 2019) analysis of *B. ubonensis* strain BCC1973 identified 9 genes (53%) with close homology, 87% to 96% AA sequence identity match over 99% to 100% of their length, to the occidiofungin BGC from Lyc2 (Figure 5.9.; panel b). However, the cluster appeared to be split over two contigs and several genes could have gone undetected by the antiSMASH algorithm (4.18.; panel a). The predicted BGC from strain BCC1990 shared close homology, 79% to 96% AA sequence identity match over 99% to 100% of their length, with 76% of the genes from the occidiofungin A BGC (4.18.; panel b). No homologs were identified for only 4 genes including 2 *LuxR* family transcriptional regulators, a hypothetical protein and an oleoyl-ACP hydrolase (Wang *et al.*, 2016). The absence of *LuxR* homologs suggests BCC1990 encodes different transcriptional regulators for the predicted BGC. Additionally, proteins with functional similarity to oleoyl-ACP hydrolase and the hypothetical protein, but different amino acid sequences, would have gone undetected by the BLASTP algorithm antiSMASH uses (Blin *et al.,* 2017). Interestingly, the predicted occidiofungin BGC in BCC1990, and part of the predicted occidiofungin BGC in BCC1973, were located directly adjacent to the enacyloxin BGC (Figure 5.9.). Enacyloxins are responsible for the observed anti-Gram-negative activity from strains BCC1973 and BCC1990 (Figure 5.2 and 4.11.). The complete loss of antimicrobial activity for the third replicon deletion mutants of *B. ubonensis* strains BCC1973 and BCC1990 (Figure 5.2) suggests the enacyloxin BGC, and therefore by its proximity the occidiofungin BGC, are located on the third replicon.

The loss of antifungal activity in conjunction with the loss of the occidiofungin BGC in third replicon deletion mutants further suggests occidiofungin or occidiofungin-like compounds are responsible for the observed antifungal activity by strains BCC1973 and BCC1990. Overall, additional chemical analyses of culture extracts will be needed to confirm the production of occidiofungin and complete genome sequence analysis to confirm the location of the BGCs

# Chapter 6 – Conclusion

## 6.1. Conclusion

This project sought to expand our understanding of the diversity of *Burkholderia* bacteria in the natural environment. A total of 66 environmental samples were collected from the rhizosphere and surrounding soil of 12 different medicinal plants located across 6 different sites in the Western Ghats, Karnataka region, India in order to answer the following hypothesis:

**"The rhizosphere and surrounding soil of medicinal plants in the Western Ghats, India is a rich source of antimicrobial producing *Burkholderia* bacteria."**

### 6.1.1. The natural environment in Western Ghats, India is a rich source of *Burkholderia cepacia* complex bacteria.

Isolation, enrichment and purification of bacteria from the 66 environmental samples resulted in the recovery of 73 pure bacteria isolates (Table 3.1 and 3.2) from 54 of the samples. PCR amplification of the 16S rRNA and *recA* genes and subsequent sequencing of DNA products, permitted sequence alignment with the NCBI database for species identification which revealed 35 *Burkholderia* isolates within the collection (Table 3.1). Strikingly, based on the 16S rRNA gene sequence, the closest relatives of the 35 *Burkholderia* isolates in the NCBI database were all Bcc species; no additional *Burkholderia* species outside this closely related group were identified. Alignment of the *recA* gene sequence (Table 3.1) identified the 35 *Burkholderia* isolates as *Burkholderia* sp. (n=3), *Burkholderia cenocepacia* (n=10), *Burkholderia cepacia* (n=9), *Burkholderia arboris* (n=9), *Burkholderia ubonensis* (n=2), *Burkholderia anthina* (n=1), *Burkholderia pyrrocinia* (n=1) (Table 3.1.). Although *recA* and 16S rRNA gene sequence analysis grouped all *Burkholderia* isolates in the Bcc (Figure 3.1), several strains remained unclassified at the species level, demonstrating the limitations of single gene sequence analysis for bacterial species identification.

In order to enhance resolution and obtain accurate species identifications, whole genome sequences were obtained for 34 of the 35 *Burkholderia* isolates, and this data was used to carry out MLST, MLSA and phylogenomic analysis (Table 4.1). Richter and Rossello-Mora (2009) have reported that an ANI value of 95% between genomes is a suitable threshold value for bacterial species delineation. The following species identifications were made based on genomic ANI values ≥95% after comparison with species type strains: *B. pyrrocinia* (n=1), *B. cepacia* (n=9), *B. ubonensis* (n=2) and *B. cenocepacia* (n=8) (Table 4.2). Strains BCC1987 and BCC1989 shared an ANI value of 94.5% with the *B. cenocepacia* type strain, falling just below the threshold value of 95% as suggested by Richter and Rosello-Mora (2009). Furthermore, BCC1987 and BCC1989 clustered separately from the *B. cenocepacia* reference strains

used in the MLSA (Figure 4.1) and formed an outgroup in a core-phylogeny constructed from 42 *B. cenocepacia* genomes (Figure 4.6). Collectively this data suggests BCC1987 and BCC1989 may belong to a novel species taxon within the Bcc. Interestingly, strains BCC1987 and BCC1989 were identified as *B. cenocepacia* IIIB lineage strains based on *recA* phylogeny (Figure 3.1) and shared an ANI value of 95% with the only other *B. cenocepacia* IIIB lineage strain (BCC2001) in the Mysore collection (Figure 4.2). Phylogenomic studies of the *B. cenocepacia* species (Wallner *et al.* 2019) and the wider Bcc (Jin *et al.* 2020) have proposed the *B. cenocepacia recA* IIIB lineage represents a separate species lineage. With the enhanced resolution that whole genome sequence analysis now provides over single gene or multi-locus sequence analysis, the classification of *B. cenocepacia* subgroups into separate species lineages should be evaluated in future phylogenomic analyses. Nevertheless, BCC1987 and BCC1989 were considered *B. cenocepacia*-like species isolates for the purposes of this study.

The remaining 12 environmental strains were designated as novel *Burkholderia* sp. (Table 4.2) based on genomic taxonomy. However, at the time of analysis the collection of available species type strains genomes for the Bcc was incomplete with no type strain genomes available for *B. diffusa*, *B. anthina* and *B. arboris*. Several of the *Burkholderia* sp. strains exhibited a close relationship to those type strains based on MLSA phylogeny (Figure 4.1.) and obtaining type strain genomes should be a priority for future analyses to establish accurate species identifications (Chun and Rainey, 2014). In summary, excluding BCC1987 and BCC1989, genomic ANI, MLSA and phylogenomic analysis identified 20 strains of 4 defined Bcc species lineages and 12 strains which represent 4 putative novel species lineages (Table 4.2).

### 6.1.2. The natural environment in the Western Ghats, India was identified as a source of *B. cenocepacia* IIIA lineage strains – a virulent lineage with a previously undefined environmental source.

Initial analysis of the Mysore *Burkholderia* collection using the *recA* gene sequence (Figure 3.1 and 3.2) identified 10 isolates as *B. cenocepacia*, including 7 isolates which belonged to the *recA* IIIA lineage, a strain lineage which has rarely been isolated outside of clinical infection (Baldwin *et al.*, 2007). Interestingly, 6 Mysore *B. cenocepacia* IIIA lineage strains were isolated from Pilikula Biological Park, Mangalore, India and 1 strain about 140 km south-east of Mangalore in Mercara, India indicating a potentially localized environmental source of *B. cenocepacia* IIIA lineage strains (Table 3.1). MLST of the 10 Mysore *B. cenocepacia* isolates revealed 7 ST's associated with *B. cenocepacia* IIIA lineage strains, 1 ST with *B. cenocepacia* IIIB lineage strain and 2 uncharacterized STs, corroborating the identification of 7 *B. cenocepacia* IIIA lineage strains following *recA* gene sequence analysis (Table 4.4). The isolation of 7 *B. cenocepacia* IIIA lineage strains in a single environmental sampling study is the

first time a major natural source of strains from this lineage has been identified (Drevinek and Mahenthiralingam, 2010).

Sequence types and clonal relationships can be associated with clinical epidemiology to permit the identification of strains with the propensity to cause infection and disease (Baldwin *et al.,* 2005; Drevinek and Mahenthiralingam, 2010). Although there were minor differences in the MLST profiles of 6 Mysore IIIA lineage strains the analysis demonstrated they all belonged to clonal complex 31, a clonal complex which includes several ST's associated with epidemic *B. cenocepacia* strains (Baldwin *et al.,* 2005; Baldwin *et al.,* 2007). Strikingly, the MLST profiles of strains BCC1997, BCC1979 and BCC1969 were associated with ST-32, an ST which has been detected in CF populations in Italy, France, UK and Canada and has spread epidemically in Canada (strain type RAPD01) and the Czech Republic (strain CZ1) (Drevinek and Mahenthiralingam, 2010). Comparative genomic analysis of the 10 Mysore *B. cenocepacia* strains and 32 characterized *B. cenocepacia* strains (Table 4.4), which included strains with STs from epidemic and/or globally distributed lineages (ST-28, ST-234, ST-241, ST-32), revealed a close relationship of the 7 Mysore IIIA lineage strains to clinical reference strains with ST-32 (Figure 4.5 and 4.6). The close relationship of the environmental *B. cenocepacia* strains to clinically problematic and globally distributed strains suggests that acquisition of virulent *B. cenocepacia* IIIA strains can occur directly from the natural environment. However, the majority of *B. cenocepacia* IIIA acquisitions within CF were down to transmissible strains which likely spread from patient to patient (Drevinek and Mahenthiralingam, 2010). Further environmental sampling in the Western Ghats, India, is required to understand the distribution and prevalence of *B. cenocepacia* IIIA lineage strains in this particular natural environment, and to investigate the capacity of these environmental strains to cause infections through appropriate infection model systems.

To evaluate the pathogenic potential of the Mysore *B. cenocepacia* strains at a genomic level, the strains were investigated for the presence of the BCESM, a clinical risk marker (Mahenthiralingam *et al.* 1997) linked to a pathogenicity island (*cci*) which promotes survival and pathogenesis in the CF lung (Baldwin *et al.,* 2004). PCR amplification of the 1.4 kb BCESM fragment (Figure 3.3) demonstrated 3 Mysore *B. cenocepacia* IIIA lineage strains carried the BCESM. In order to determine if BCESM positive strains encoded the wider *cci,* the genomes of the 42 *B. cenocepacia* strains used for comparative analysis were screened against a database containing the *cci* genes as described by Baldwin *et al.* (2004) (Figure 4.7 and 4.8). From the 7 Mysore *B. cenocepacia* IIIA lineage strains 5 encoded an *esmR* homolog, which is a putative negative transcriptional regulator encoded by the BCESM. Genes from the *cci* were conserved amongst the 5 Mysore *B. cenocepacia* strains (ranging from 66% to 88% of the genes within the island) which were *esmR* positive, which suggests the *cci* may also have a functional role for survival within the rhizosphere and soil environment. Previous studies have demonstrated

attenuated virulence and inflammatory potential in knockout mutants of the *cci* genes *cciR, cciI*, *amiI*, and *opcI* (Baldwin *et al.,* 2004; Malott *et al.,* 2005). Homologs for each gene were identified in all 5 *esmR* positive Mysore IIIA lineage strains (Figure 4.7 and Figure 4.8). In conclusion, the *cci* promotes survival and pathogenesis in CF lung infections and its prevalence amongst environmental strains indicates an intrinsic ability for virulent opportunistic human disease (Baldwin *et al.,* 2004). The capacity for infection and disease of environmental *B. cenocepacia* IIIA lineage strains should be investigated further by evaluating their pathogenesis in infection models and screening for key virulence genes.

### 6.1.3.Environmental *Burkholderia cepacia* complex strains from the Western Ghats, India can produce multiple antimicrobials.

Following the isolation and identification of 35 *Burkholderia* strains, the Mysore *Burkholderia* collection was investigated for antimicrobial activity. Antimicrobial activity screening of the Mysore *Burkholderia* collection (n=35) revealed 26 strains with antimicrobial activity. Strategies for eliciting novel or enhancing antimicrobial production by supplementing minimal growth media with alternative carbon sources, or by using seed exudates to produce a biomimetic growth source, had limited success. In contrast, altering basic growth conditions such as incubation temperature and period revealed varying growth conditions for optimal antimicrobial production in selected environmental *Burkholderia* strains. HPLC and LC-MS analyses of culture extracts identified several metabolites closely related to known antimicrobial compounds. Subsequent genome mining predicted several BGCs that could be associated with the detected compounds. *Burkholderia* sp. BCC1970 was determined to produce collimonins C/D and is the first characterized collimonins producer in *Burkholderia*, further revealing the biosynthetic potential of the members of this genus. Although collimonin is structurally closely related to cepacin (Mullins *et al.*, 2019), the *Burkholderia* collimonin pathway provides further insight into the biosynthesis of polyynes in *Burkholderia* and this knowledge could be exploited for engineering biosynthetic pathways to generate novel derivatives.

*B. ubonensis* strains BCC1973 and BCC1990 were determined to produce potent antifungal and antibacterial compounds including enacyloxin IIa and IVa, which is the first time enacyloxin metabolites have been linked to *B. ubonensis* species. Reports of human *B. ubonensis* infections are almost entirely absent (LiPuma, 2010; Price *et al.,* 2017), and this epidemiological data, combined with the antimicrobial activity identified in this study, suggests *B. ubonensis* could be a useful species for developing *Burkholderia* based biopesticides and biotechnological heterologous expression strains.

In conclusion, the isolation of 35 *Burkholderia cepacia* complex strains with 74% of these exhibiting antimicrobial activity, reveals the rhizosphere and surrounding soil of medicinal plants in the Western

Ghats, India as a source rich in antimicrobial producing Bcc bacteria. Genome mining and chemical analysis of culture extracts identified the *B. ubonensis* strains as enacyloxin and occidiofungin producers, and *Burkholderia* sp. BCC1970 as a collimonin producer. Several other Bcc species have been identified as producers of enacyloxin (Mahenthiralingam *et al.,* 2011), occidiofungin (Lu *et al.,* 2009; Wang *et al.*, 2016) and compounds which are structurally closely related to collimonin (Kai *et al.*, 2018; Mullins *et al.,* 2019). These findings emphasize how the frequent rediscovery of compounds continues to be a major problem when attempting to discover new antibiotics from bacteria. The *Burkholderia* genus is a relatively underexplored genus for antimicrobial discovery and whether the findings in this study are a result of the technical limitations to antimicrobial discovery or due to a lack of biosynthetic diversity in the Bcc remains to be determined. This project has considerably expanded the collection of environmental Bcc isolates which can be studied in the context of specialized metabolite production, antibiotic discovery, and to understand how these organisms can cause opportunistic human infection.

# References

Agnoli, K., Schwager, S., Uehlinger, S., Vergunst, A., Viteri, D., Nguyen, D. and Sokol, P. *et al.* (2012). Exposing the third chromosome of *Burkholderia cepacia* complex strains as a virulence plasmid. *Molecular Microbiology* **83**:362-378.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J. (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology,* **215**:403-410.

Andrews, S. (2009). FastQC: A quality control tool for high throughput sequence data. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ [Accessed: 9 March 2020].

Aminov, R. (2010). A Brief History of the Antibiotic Era: Lessons Learned and Challenges for the Future. *Frontiers in Microbiology*, **1**:134.

Amies, C. R. (1967). A modified formula for the preparation of Stuart's Transport Medium. *Canadian Journal of Public Health,* **58**:296-300

Arima, K., Imanaka, H., Kousaka, M., Fukuta, A. and Tamura, G. (1964). Pyrrolnitrin, a New Antibiotic Substance, Produced by *Pseudomonas*. *Agricultural and Biological Chemistry* **28**:575-576.

Atlas, R. (2010). Handbook of Microbiological Media. 4th ed. Washington, D.C.: ASM Press.

Baldwin, A., Sokol, P., 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*. *Infection and Immunity,* **72**:1537-1547.

Baldwin, A., Mahenthiralingam, E., Thickett, K., Honeybourne, D., Maiden, M., Govan, J. and Speert, D. *et al.* (2005). Multilocus Sequence Typing Scheme That Provides Both Species and Strain Differentiation for the *Burkholderia cepacia* Complex. *Journal of Clinical Microbiology,* **43**:4665-4673.

Baldwin, A., Mahenthiralingam, E., Drevinek, P., Vandamme, P., Govan, J., Waine, D. and Chiarini, L. *et al.* (2007). Environmental *Burkholderia cepacia* Complex Isolates from Human Infections. *Emerging Infectious Diseases* **13**:458-461.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A., Dvorkin, M., Kulikov, A. and Lesin, V. *et al.* (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology* **19**:455-477.

Berg, G. and Smalla, K. (2009). Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere. *FEMS Microbiology Ecology* **68**:1-13.

Bevivino, A., Costa, B., Cantale, C., Cesarini, S., Chiarini, L., Tabacchioni, S. and Caballero-Mellado, J. *et al.* (2011). Genetic relationships among Italian and Mexican maize-rhizosphere *Burkholderia cepacia* complex (BCC) populations belonging to *Burkholderia cenocepacia* IIIB and BCC6 group. *BMC Microbiology* **11**:228.

Biggins, J., Liu, X., Feng, Z. and Brady, S. (2011). Metabolites from the Induced Expression of Cryptic Single Operons Found in the Genome of *Burkholderia pseudomallei*. *Journal of the American Chemical Society* **133**:1638-1641.

Blin, K., Kim, H., Medema, M. and Weber, T. (2017). Recent development of antiSMASH and other computational approaches to mine secondary metabolite biosynthetic gene clusters. *Briefings in Bioinformatics* **20**:1103-1113.

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. and Medema, M. *et al.* (2019). antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Research* **47**:W81-W87.

Brenner, D., Fanning, G., Rake, A. and Johnson, K. (1969). Batch procedure for thermal elution of DNA from hydroxyapatite. *Analytical Biochemistry* **28**:447-459.

Burbage, D. A., Sasser, M. (1982) A medium selective for *Pseudomonas cepacia. Phytopathology,* **76:**706

Burkholder, W. (1950) Sour skin, a bacterial rot of onion bulbs. *Phytopathology,* **64**:468–475.

Caputo, A., Fournier, P. and Raoult, D. (2019). Genome and pan-genome analysis to classify emerging bacteria. *Biology Direct* **14**:5

Cartwright, D., Chilton, W. and Benson, D. (1995). Pyrrolnitrin and phenazine production by *Pseudomonas cepacia*, strain 5.5B, a biocontrol agent of *Rhizoctonia solani*. *Applied Microbiology and Biotechnology* **43**:211-216.

Carvalhais, L., Dennis, P., Fedoseyenko, D., Hajirezaei, M., Borriss, R. and von Wirén, N. (2010). Root exudation of sugars, amino acids, and organic acids by maize as affected by nitrogen, phosphorus, potassium, and iron deficiency. *Journal of Plant Nutrition and Soil Science,* **174**:3-11.

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. *Letters in Applied Microbiology* **49**:580-588.

Chandler, D., Bailey, A., Tatchell, G., Davidson, G., Greaves, J. and Grant, W. (2011). The development, regulation and use of biopesticides for integrated pest management. *Philosophical Transactions of the Royal Society B: Biological Sciences* **366**:1987-1998.

Chun, J. and Rainey, F. (2014). Integrating genomics into the taxonomy and systematics of the Bacteria and Archaea. *International Journal of Systematic and Evolutionary Microbiology* **64**:316-324.

Coenye, T., Vandamme, P., Govan, J. and LiPuma, J. (2001). Taxonomy and Identification of the *Burkholderia cepacia* Complex. *Journal of Clinical Microbiology* **39**:3427-3436.

Coenye, T. (2010). Social interactions in the *Burkholderia cepacia* complex: biofilms and quorum sensing. *Future Microbiology* **5**:1087-1099.

Compant, S., Nowak, J., Coenye, T., Clément, C. and Ait Barka, E. (2008a). Diversity and occurrence of *Burkholderia spp.* in the natural environment. *FEMS Microbiology Reviews* **32**:607-626.

Compant, S., Kaplan, H., Sessitsch, A., Nowak, J., Ait Barka, E. and Clement, C. (2008b). Endophytic colonization of *Vitis vinifera L.* by *Burkholderia phytofirmans* strain PsJN: from the rhizosphere to inflorescence tissues. *FEMS Microbiology Ecology* **63**:84-93.

Connor, T., Loman, N., Thompson, S., Smith, A., Southgate, J., Poplawski, R. and Bull, M. *et al.* (2016). CLIMB (the Cloud Infrastructure for Microbial Bioinformatics): an online resource for the medical microbiology community. *Microbial Genomics* **2**:e000086.

Cunningham-Oakes, E., Weiser, R., Pointon, T. and Mahenthiralingam, E. (2019). Understanding the challenges of non-food industrial product contamination. *FEMS Microbiology Letters* **366**:1-5

Davies, J. (2013). Specialized microbial metabolites: functions and origins. *The Journal of Antibiotics* **66**:361-364.

De Soyza, A., Morris, K., McDowell, A., Doherty, C., Archer, L., Perry, J., Govan, J. R. W., Corris P. A., Gould, K. (2004). Prevalence and clonality of *Burkholderia cepacia* complex genomovars in UK patients with cystic fibrosis referred for lung transplantation. *Thorax,* **59**:526-528.

Depoorter, E., Bull, M., Peeters, C., Coenye, T., Vandamme, P. and Mahenthiralingam, E. (2016). *Burkholderia*: an update on taxonomy and biotechnological potential as antibiotic producers. *Applied Microbiology and Biotechnology* **100**:5215-5229.

Dobritsa, A. 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*. *International Journal of Systematic and Evolutionary Microbiology*, **66**:2836-2846.

Doornbos, R., van Loon, L. and Bakker, P. (2011). Impact of root exudates and plant defence signalling on bacterial communities in the rhizosphere. A review. *Agronomy for Sustainable Development* **32**:227-243.

Draghi, W., Degrossi, J., Bialer, M., Brelles-Mariño, G., Abdian, P., Soler-Bistué, A. and Wall, L. *et al.* (2018). Biodiversity of cultivable *Burkholderia* species in Argentinean soils under no-till agricultural practices. *PLOS One* **13**:e0200651.

Drancourt, M., Bollet, C., Carlioz, A., Martelin, R., Gayral, J. and Raoult, D. (2000). 16S Ribosomal DNA Sequence Analysis of a Large Collection of Environmental and Clinical Unidentifiable Bacterial Isolates. *Journal of Clinical Microbiology* **38**:3623-3630.

Drevinek, P., Baldwin, A., Dowson, C. and Mahenthiralingam, E. (2008). Diversity of the *parB* and *repA* genes of the *Burkholderia cepacia* complex and their utility for rapid identification of *Burkholderia cenocepacia*. *BMC Microbiology,* **8**:44.

Drevinek, P. and Mahenthiralingam, E. (2010). Burkholderia cenocepacia in cystic fibrosis: epidemiology and molecular mechanisms of virulence. *Clinical Microbiology and Infection,* **16**:821-830.

Eberl, L. and Vandamme, P. (2016). Members of the genus *Burkholderia*: good and bad guys. *F1000Research* **5**:1007.

El-Banna, N. and Winkelmann, G. (1998). Pyrrolnitrin from *Burkholderia cepacia*: antibiotic activity against fungi and novel activities against *Streptomycetes*. *Journal of Applied Microbiology* **85**:69-78.

Esmaeel, Q., Jacquard, C., Clément, C., Sanchez, L. and Ait Barka, E. (2019). Genome sequencing and traits analysis of *Burkholderia* strains reveal a promising biocontrol effect against grey mould disease in grapevine (*Vitis vinifera L.*). *World Journal of Microbiology and Biotechnology* **35**:40

Estrada-de los Santos, P., Palmer, M., Chávez-Ramírez, B., Beukes, C., Steenkamp, E., Briscoe, L. and Khan, N. *et al.* (2018). Whole Genome Analyses Suggests that *Burkholderia* sensu lato Contains Two Additional Novel Genera (*Mycetohabitans* gen. nov., and *Trinickia* gen. nov.): Implications for the Evolution of Diazotrophy and Nodulation in the *Burkholderiaceae*. *Genes* **9**:389.

Fierer, N. and Jackson, R. B. (2006) The diversity and biogeography of soil bacterial communities *Proceedings of the National Academy of Sciences* **103:**626-631

Fierer, N. (2017) Embracing the unknown: disentangling the complexities of the soil microbiome. *Nature Reviews Microbiology*, **15:**579–590.

Fila, L. and Dřevínek, P. (2017). *Burkholderia cepacia* complex in cystic fibrosis in the post-epidemic period: multilocus sequence typing-based approach. *Folia Microbiologica* **62**:509-514.

Funston, S., Tsaousi, K., Rudden, M., Smyth, T., Stevenson, P., Marchant, R. and Banat, I. (2016). Characterising rhamnolipid production in *Burkholderia thailandensis* E264, a non-pathogenic producer. *Applied Microbiology and Biotechnology* **100**:7945-7956.

Fritsche, K., van den Berg, M., de Boer, W., van Beek, T., Raaijmakers, J., van Veen, J. and Leveau, J. (2014). Biosynthetic genes and activity spectrum of antifungal polyynes from *Collimonas fungivorans* Ter331. *Environmental Microbiology* **16**:1334-1345.

Genilloud, O. (2018). Mining *Actinomycetes* for Novel Antibiotics in the Omics Era: Are We Ready to Exploit This New Paradigm? *Antibiotics* **7**:85.

Gerrits, G., Klaassen, C., Coenye, T., Vandamme, P. and Meis, J. (2005). *Burkholderia fungorum* Septicemia. *Emerging Infectious Diseases* **11**:1115-1117.

Glass, M., Wilkins, P., Hoffmaster, A. and Beesley, C. (2009). Comparison of Four Selective Media for the Isolation of *Burkholderia mallei* and *Burkholderia pseudomallei*. *The American Journal of Tropical Medicine and Hygiene,* **80**:1023-1028.

Goris, J., Klappenbach, J., Vandamme, P., Coenye, T., Konstantinidis, K. and Tiedje, J. (2007). DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology* **57**:81-91.

Graindorge, A., Menard, A., Neto, M., Bouvet, C., Miollan, R., Gaillard, S. and de Montclos, H. *et al.* (2010). Epidemiology and molecular characterization of a clone of *Burkholderia cenocepacia* responsible for nosocomial pulmonary tract infections in a French intensive care unit. *Diagnostic Microbiology and Infectious Disease* **66**:29-40.

Gurevich, A., Saveliev, V., Vyahhi, N. and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**:1072-1075.

Haeckl, F., Baldim, J., Iskakova, D., Kurita, K., Soares, M. and Linington, R. (2019). A selective genome-guided method for environmental *Burkholderia* isolation. *Journal of Industrial Microbiology & Biotechnology* **46**:345-362.

Hareland, W. A., Crawford, R. L., Chapman, P. J., and Dagley, S. (1975) Metabolic Function and Properties of 4-Hydroxyphenylacetic Acid 1-Hydroxylase from *Pseudomonas acidovorans*. *J. Bacteriol.,* **121**:272–285.

Hebbar, K., Atkinson, D., Tucker, W. and Dart, P. (1992). Suppression of *Fusarium moniliforme* by maize root-associated *Pseudomonas cepacia*. *Soil Biology and Biochemistry* **24**:1009-1020.

Hebbar, K., Martel, M. and Heulin, T. (1998). Suppression of pre- and postemergence damping-off in corn by *Burkholderia cepacia*. *European Journal of Plant Pathology* **104**:29-36.

Henry, D., Campbell, M., McGimpsey, C., Clarke, A., Louden, L., Burns, J. and Roe, M. *et al.* (1999). Comparison of Isolation Media for Recovery of Burkholderia cepacia Complex from Respiratory Secretions of Patients with Cystic Fibrosis. *Journal of Clinical Microbiology,* **37**:1004-1007.

Ho, Y. and Huang, C. (2015). Draft Genome Sequence of *Burkholderia cenocepacia* Strain 869T2, a Plant-Beneficial Endophytic Bacterium. *Genome Announcements* **3**:e01327-15.

Holden, M., Seth-Smith, H., Crossman, L., Sebaihia, M., Bentley, S., Cerdeno-Tarraga, A. and Thomson, N. *et al.* (2008). The Genome of *Burkholderia cenocepacia* J2315, an Epidemic Pathogen of Cystic Fibrosis Patients. *Journal of Bacteriology,* **191**:261-277.

Jain, C., Rodriguez-R, L., Phillippy, A., Konstantinidis, K. and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications* **9**: 1-8

Jenner, M., Jian, X., Dashti, Y., Masschelein, J., Hobson, C., Roberts, D. and Jones, C. *et al.* (2019). An unusual *Burkholderia gladioli* double chain-initiating nonribosomal peptide synthetase assembles 'fungal' icosalide antibiotics. *Chemical Science* **10**:5489-5494.

Jin, Y., Zhou, J., Zhou, J., Hu, M., Zhang, Q., Kong, N. and Ren, H. et al. (2020). Genome-based classification of *Burkholderia cepacia* complex provides new insight into its taxonomic status. *Biology Direct* **15**: 1972–1973

Jolley, K. and Maiden, M. (2010). BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC *Bioinformatics,* **11**:595

Jolley, K., Bliss, C., Bennett, J., Bratcher, H., Brehony, C., Colles, F. and Wimalarathna, H. *et al.* (2012). Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. *Microbiology* **158**:1005-1015.

Jones, A., Dodd, M., Govan, J., Barcus, V., Doherty, C. and Webb, A. (2004). *Burkholderia cenocepacia* and *Burkholderia multivorans*: influence on survival in cystic fibrosis. *Thorax* **59**:948-951.

Jung, B., Hong, S., Park, G., Kim, M. and Shin, J. (2018). Isolation of *Burkholderia cepacia* JBK9 with plant growth-promoting activity while producing pyrrolnitrin antagonistic to plant fungal diseases. *Applied Biological Chemistry* **61**:173-180.

Kai, K., Sogame, M., Sakurai, F., Nasu, N. and Fujita, M. (2018). Collimonins A–D, Unstable Polyynes with Antifungal or Pigmentation Activities from the Fungus-Feeding Bacterium *Collimonas fungivorans* Ter331. *Organic Letters* **20**:3536-3540.

Katoh, K. and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* **30**:772–80.

Kawanishi, T., Shiraishi, T., Okano, Y., Sugawara, K., Hashimoto, M., Maejima, K. and Komatsu, K. *et al.* (2011). New Detection Systems of Bacteria Using Highly Selective Media Designed by SMART: Selective Medium-Design Algorithm Restricted by Two Constraints. *PLoS ONE* **6**:e16512.

Kenna, D., Lilley, D., Coward, A., Martin, K., Perry, C., Pike, R., Hill, R. and Turton, J. (2017). Prevalence of *Burkholderia* species, including members of *Burkholderia cepacia* complex, among UK cystic and non-cystic fibrosis patients. *Journal of Medical Microbiology*, **66**:490-501.

Keum, Y. (2009). Effects of Nutrients on Quorum Signals and Secondary Metabolite Productions of Burkholderia sp. O33. *Journal of Microbiology and Biotechnology*, **19**:1142-1149

Khaldi, N., Seifuddin, F., Turner, G., Haft, D., Nierman, W., Wolfe, K. and Fedorova, N. (2010). SMURF: Genomic mapping of fungal secondary metabolite clusters. *Fungal Genetics and Biology* **47**:736-741.

Kikuchi, Y., Hosokawa, T. and Fukatsu, T. (2007). Insect-Microbe Mutualism without Vertical Transmission: a Stinkbug Acquires a Beneficial Gut Symbiont from the Environment Every Generation. *Applied and Environmental Microbiology* **73**:4308-4316.

Kim, M., Morrison, M. and Yu, Z. (2011). Evaluation of different partial 16S rRNA gene sequence regions for phylogenetic analysis of microbiomes. *Journal of Microbiological Methods* **84**: 81–87

King E. O., Ward, M. K. and Raney, D. E. (1954) Two Simple Media for the Demonstration of Pyocyanin and Fluorescein. *Journal of Laboratory and Clinical Medicine,* **44**:301- 307.

Köberl, M., Schmidt, R., Ramadan, E. M., Bauer, R. and Berg, G. (2013). The microbiome of medicinal plants: diversity and importance for plant growth, quality and health. *Frontiers in Microbiology*, **4:**400.

Koch, B., Nielsen, T., Sørensen, D., Andersen, J., Christophersen, C., Molin, S. and Givskov, M. *et al.* (2002). Lipopeptide Production in *Pseudomonas* sp. Strain DSS73 Is Regulated by Components of Sugar Beet Seed Exudate via the *Gac* Two-Component Regulatory System. *Applied and Environmental Microbiology* **68**:4509-4516.

Kost, T., Stopnisek, N., Agnoli, K., Eberl, L. and Weisskopf, L. (2014). Oxalotrophy, a widespread trait of plant-associated *Burkholderia* species, is involved in successful root colonization of lupin and maize by *Burkholderia phytofirmans*. *Frontiers in Microbiology* **4**:421

Krueger, F. 2016. Trim Galore! Available at: https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ [Accessed: 9 March 2020]

Kunakom, S. and Eustáquio, A. (2019). *Burkholderia* as a Source of Natural Products. *Journal of Natural Products,* **82**:2018-2037.

Kumar, S., Stecher, G. and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, **33**:1870-1874.

Lane, D. J. (1991). 16S/23S rRNA Sequencing. In: Stackebrandt, E., and Goodfellow, M. (eds.) *Nucleic acid techniques in bacterial systematics.* New York: John Wiley and Sons, pp. 115-175.

Lee, K., Kong, H., Choi, K., Lee, S. and Moon, B. (2011). Isolation and Identification of *Burkholderia pyrrocinia* CH-67 to Control Tomato Leaf Mold and Damping-off on Crisphead Lettuce and Tomato. *The Plant Pathology Journal* **27**:59-67.

Li, X., Rui, J., Mao, Y., Yannarell, A. and Mackie, R. (2014). Dynamics of the bacterial community structure in the rhizosphere of a maize cultivar. *Soil Biology and Biochemistry* **68**:392-401.

Lincke, T., Behnken, S., Ishida, K., Roth, M. and Hertweck, C. (2010). Closthioamide: An Unprecedented Polythioamide Antibiotic from the Strictly Anaerobic Bacterium *Clostridium cellulolyticum*. *Angewandte Chemie International Edition* **49**:2011-2013.

LiPuma, J., Spilker, T., Gill, L., Campbell, P., Liu, L. and Mahenthiralingam, E. (2001). Disproportionate Distribution of *Burkholderia cepacia* Complex Species and Transmissibility Markers in Cystic Fibrosis. *American Journal of Respiratory and Critical Care Medicine* **164**:92-96.

LiPuma, J., Spilker, T., Coenye, T. and Gonzalez, C. (2002). An epidemic *Burkholderia cepacia* complex strain identified in soil. *The Lancet* **359**:2002-2003.

LiPuma, J. (2003). Molecular epidemiology of *Burkholderia* species. *Frontiers in Bioscience* **8**:e55-67.

LiPuma, J. (2010). The Changing Microbial Epidemiology in Cystic Fibrosis. *Clinical Microbiology Reviews* **23**:299-323.

Lopes-Santos, L., Castro, D., Ferreira-Tonin, M., Corrêa, D., Weir, B., Park, D., Ottoboni, L., Neto, J. and Destéfano, S. (2017). Reassessment of the taxonomic position of *Burkholderia andropogonis* and description of *Robbsia andropogonis* gen. nov., comb. nov. *Antonie van Leeuwenhoek*, **110**:727-736.

Loutet, S. A. and Valvano, M. A. (2011). Extreme Antimicrobial Peptide and Polymyxin B Resistance in the Genus *Burkholderia*. *Frontiers in Microbiology* **2**: 159

Loveridge, E., Jones, C., Bull, M., Moody, S., Kahl, M., Khan, Z. and Neilson, L. *et al.* (2017). Reclassification of the Specialized Metabolite Producer *Pseudomonas mesoacidophila* ATCC 31433 as a Member of the *Burkholderia cepacia* Complex. *Journal of Bacteriology*, **13**: e00125-17**.**

Lowbury, E. and Collins, A. (1955). The Use of a New Cetrimide Product in a Selective Medium for *Pseudomonas pyocyanea*. *Journal of Clinical Pathology,* **8**:47-48.

Lu, S., Novak, J., Austin, F., Gu, G., Ellis, D., Kirk, M. and Wilson-Stanford, S. *et al.* (2009). Occidiofungin, a Unique Antifungal Glycopeptide Produced by a Strain of *Burkholderia contaminans*. *Biochemistry* **48**:8312-8321.

Madhaiyan, M., Poonguzhali, S., Kwon, S. W., Song, M. H., and Sa, T. (2008). Molecular characterization of *Burkholderia* strains isolated from rice cultivars (*Oryza sativa L.*) for species identification and phylogenetic grouping. *Journal of Microbiology and Biotechnology,* **18**:1005–1010.

Mahenthiralingam, E., Simpson, D. A., Speert, D. P. (1997) Identification and Characterization of a Novel DNA Marker Associated with Epidemic Burkholderia cepacia Strains Recovered from Patients with Cystic Fibrosis. *American Society for Microbiology,* **35**:808-816

Mahenthiralingam, E., Coenye, T., Chung, J., W., Speert, D. P., Govan, J. R. W., Taylor, P., Vandamme, P. (2000a). Diagnostically and Experimentally Useful Panel of Strains from the *Burkholderia cepacia* Complex. *Journal of Clinical Microbiology*, **38**:910-913

Mahenthiralingam, E., Bischof, J., Byrne, S., Radomski, C., Davies, J., Av-Gay, Y. and Vandamme, P. (2000b). DNA-Based Diagnostic Approaches for Identification of *Burkholderia cepacia* Complex, *Burkholderia vietnamiensis, Burkholderia multivorans, Burkholderia stabilis,* and *Burkholderia cepacia* Genomovars I and III. *Journal of Clinical Microbiology* **38**:3165-3173.

Mahenthiralingam, E., Vandamme, P., Campbell, M., Henry, D., Gravelle, A., Wong, L. and Davidson, A. *et al.* (2001). Infection with *Burkholderia cepacia* Complex Genomovars in Patients with Cystic Fibrosis: Virulent

Transmissible Strains of Genomovar III Can Replace *Burkholderia multivorans*. *Clinical Infectious Diseases,* **33**:1469-1475.

Mahenthiralingam, E., Baldwin, A. and Dowson, C. (2008). *Burkholderia cepacia* complex bacteria: opportunistic pathogens with important natural biology. *Journal of Applied Microbiology* **104**:1539-1551.

Mahenthiralingam, E., Song, L., Sass, A., White, J., Wilmot, C., Marchbank, A., *et al.* (2011). Enacyloxins Are Products of an Unusual Hybrid Modular Polyketide Synthase Encoded by a Cryptic *Burkholderia ambifaria* Genomic Island. *Chemistry & Biology*, **18**:665-677.

Maiden, M. (2006). Multilocus Sequence Typing of Bacteria. *Annual Review of Microbiology* **60**:561-588.

Maiden, M., van Rensburg, M., Bray, J., Earle, S., Ford, S., Jolley, K. and McCarthy, N. (2013). MLST revisited: the gene-by-gene approach to bacterial genomics. *Nature Reviews Microbiology* **11**:728-736.

Malott, R., Baldwin, A., Mahenthiralingam, E. and Sokol, P. (2005). Characterization of the *cciIR* Quorum-Sensing System in *Burkholderia cenocepacia*. *Infection and Immunity* **73**:4982-4992.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal* **17**:10

Medema, M., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J., Blin, K. and de Bruijn, I. *et al.* (2015). Minimum Information about a Biosynthetic Gene cluster. *Nature Chemical Biology,* **11**:625-631.

Mendes, R., Pizzirani-Kleiner, A., Araujo, W. and Raaijmakers, J. (2007). Diversity of Cultivated Endophytic Bacteria from Sugarcane: Genetic and Biochemical Characterization of *Burkholderia cepacia* Complex Isolates. *Applied and Environmental Microbiology* **73**:7259-7267.

Moore, J., Mcllhatton, B., Buchanan, J., Gilpin, D., Shaw, A., Hall, V. and Murphy, P. *et al.* (2002). Occurrence of *Burkholderia cepacia* in the hospital environment. *Irish Journal of Medical Science* **171**:131-133.

Mullins, A., Murray, J., Bull, M., Jenner, M., Jones, C., Webster, G. and Green, A. *et al.* (2019). Genome mining identifies cepacin as a plant-protective metabolite of the biopesticidal bacterium *Burkholderia ambifaria*. *Nature Microbiology,* **4**:996-1005.

Nampoothiri, K. M., Ramkumar, B., Pandey, A. (2013) Western Ghats of India: Rich source of microbial biodiversity. *Journal of Scientific and Industrial Research,* **72**:617-623

Ondov, B., Bergman, N. and Phillippy, A. (2011). Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics* **12**:385

O'Sullivan, L., Weightman, A., Jones, T., Marchbank, A., Tiedje, J. and Mahenthiralingam, E. (2007). Identifying the genetic basis of ecologically and biotechnologically useful functions of the bacterium *Burkholderia vietnamiensis*. *Environmental Microbiology*, **9**:1017-1034.

Oteros, J., Bartusel, E., Alessandrini, F., Núñez, A., Moreno, D., Behrendt, H. and Schmidt-Weber, C. *et al.* (2019). *Artemisia* pollen is the main vector for airborne endotoxin. *Journal of Allergy and Clinical Immunology* **143**:369-377

Oyama, R., Watanabe, T., Hanzawa, H., Sano, T., Sugiyama, T. and Izaki, K. (1994). An Extracellular Quinoprotein Oxidase That Catalyzes Conversion of Enacyloxin IVa to Enacyloxin IIa. *Bioscience, Biotechnology, and Biochemistry* **58**:1914-1917.

Page, A., Cummins, C., Hunt, M., Wong, V., Reuter, S., Holden, M. and Fookes, M. *et al.* (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics,* **31**:3691-3693.

Pallud, C., Viallard, V., Balandreau, J., Normand, P. and Grundmann, G. (2001). Combined use of a specific probe and PCAT medium to study Burkholderia in soil. *Journal of Microbiological Methods*, **47**:25-34.

Parke, J. and Gurian-Sherman, D. (2001). Diversity of the *Burkholderia Cepacia* Complex and Implications for Risk Assessment of Biological Control Strains. *Annual Review of Phytopathology* **39**:225-258.

Parker, W., Rathnum, M., Seiner, V., Trejo, W., Principe, P. and Sykes, R. (1984). Cepacin A and cepacin B, two new antibiotics produced by *Pseudomonas cepacia*. *The Journal of Antibiotics* **37**:431-440.

Pawar, S., Chaudhari, A., Prabha, R., Shukla, R. and Singh, D. (2019). Microbial Pyrrolnitrin: Natural Metabolite with Immense Practical Utility. *Biomolecules* **9**:443.

Payne, G., Vandamme, P., Morgan, S., LiPuma, J., Coenye, T., Weightman, A., Jones, T. and Mahenthiralingam, E. (2005). Development of a *recA* Gene-Based Identification Approach for the Entire *Burkholderia* Genus. *Applied and Environmental Microbiology*, **71:** 3917-3927.

Peacock, S., Chieng, G., Cheng, A., Dance, D., Amornchai, P., Wongsuvan, G. and Teerawattanasook, N. *et al.* (2005). Comparison of Ashdown's Medium, *Burkholderia cepacia* Medium, and *Burkholderia pseudomallei* Selective Agar for Clinical Isolation of *Burkholderia pseudomallei*. *Journal of Clinical Microbiology* **43**:5359-5361.

Pedraza, L. A., Bautista, J. and Uribe-Vélez, D. (2018). Seed-born *Burkholderia glumae* Infects Rice Seedling and Maintains Bacterial Population during Vegetative and Reproductive Growth Stage. *The Plant Pathology Journal*, **34:**393–402

Pirone, L., Bragonzi, A., Farcomeni, A., Paroni, M., Auriche, C., Conese, M. and Chiarini, L. *et al.* (2008). *Burkholderia cenocepacia* strains isolated from cystic fibrosis patients are apparently more invasive and more virulent than rhizosphere strains. *Environmental Microbiology* **10**:2773-2784.

Poirel, L., Jayol, A. and Nordmann, P. (2017). Polymyxins: Antibacterial Activity, Susceptibility Testing, and Resistance Mechanisms Encoded by Plasmids or Chromosomes. *Clinical Microbiology Reviews,* **30**:557-596.

Price, E., Sarovich, D., Webb, J., Hall, C., Jaramillo, S., Sahl, J. and Kaestli, M. *et al.* (2017). Phylogeographic, genomic, and meropenem susceptibility analysis of *Burkholderia ubonensis*. *PLOS Neglected Tropical Diseases* **11**:e0005928.

Pritchard, L., Glover, R., Humphris, S., Elphinstone, J. and Toth, I. (2016). Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Analytical Methods* **8**:12-24.

Ramette, A., LiPuma, J. and Tiedje, J. (2005). Species Abundance and Diversity of *Burkholderia cepacia* Complex in the Environment. *Applied and Environmental Microbiology* **71**:1193-1201.

Richardson, J., Stead, D., Elphinstone, J. and Coutts, R. (2002). Diversity of *Burkholderia* isolates from woodland rhizosphere environments. *Journal of Applied Microbiology* **93**:616-630.

Richter, M. and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences* **106**:19126-19131.

Rokas, A., Williams, B., King, N. and Carroll, S. (2003). Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* **425**:798-804.

Ross, C., Scherlach, K., Kloss, F. and Hertweck, C. (2014). The Molecular Basis of Conjugated Polyyne Biosynthesis in Phytopathogenic Bacteria. *Angewandte Chemie International Edition* **53**:7794-7798.

Rossolini, G., Arena, F., Pecile, P. and Pollini, S. (2014). Update on the antibiotic resistance crisis. *Current Opinion in Pharmacology*, **18**:56-60.

Rudrappa, T., Czymmek, K., Paré, P. and Bais, H. (2008). Root-Secreted Malic Acid Recruits Beneficial Soil Bacteria. *Plant Physiology* **148**:1547-1556.

Rushton, L., Sass, A., Baldwin, A., Dowson, C. G., Donoghue, D., Mahenthiralingam, E. (2013). Key Role for Efflux in the Preservative Susceptibility and Adaptive Resistance of *Burkholderia cepacia* Complex Bacteria. *Antimicrobial Agents and Chemotherapy,* **57:**2972–2980.

Salles, J. F., Samyn, E., Vandamme, P., van Veen, J. A., van Elsas, J. D. (2006) Changes in agricultural management drive the diversity of *Burkholderia* species isolated from soil on PCAT medium. *Soil Biology and Biochemistry,* **38**:661–673.

Sawana, A., Adeolu, M. and Gupta, R. (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. *Frontiers in Genetics*, **5**:429

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**:2068-2069.

Segerman, B. (2012). The genetic integrity of bacterial species: the core genome and the accessory genome, two different stories. *Frontiers in Cellular and Infection Microbiology* **2**:116

Seng, R., Saiprom, N., Phunpang, R., Baltazar, C., Boontawee, S., Thodthasri, T. and Silakun, W. *et al.,* (2019). Prevalence and genetic diversity of *Burkholderia pseudomallei* isolates in the environment near a patient's residence in Northeast Thailand. *PLOS Neglected Tropical Diseases* **13**:e0007348.

Seo, Y., Lim, J., Park, J., Kim, S., Lee, H., Cheong, H. and Kim, S. *et al*. (2015). Comparative genome analysis of rice-pathogenic *Burkholderia* provides insight into capacity to adapt to different environments and hosts. *BMC Genomics* **16**:349

Silva, I., Santos, P., Santos, M., Zlosnik, J., Speert, D., Buskirk, S. and Bruger, E. *et al.* (2016). Long-Term Evolution of *Burkholderia multivorans* during a Chronic Cystic Fibrosis Infection Reveals Shifting Forces of Selection. *mSystems,* **1:**e00029-16

Skinnider, M., Merwin, N., Johnston, C. and Magarvey, N. (2017). PRISM 3: expanded prediction of natural product chemical structures from microbial genomes. *Nucleic Acids Research* **45**:W49-W54.

Song, J., Kwak, M., Lee, K., Kong, H., Kim, B., Kwon, S. and Lee, S. et al. (2012). Draft Genome Sequence of the Antifungal-Producing Plant-Benefiting Bacterium *Burkholderia pyrrocinia* CH-67. *Journal of Bacteriology* **194**:6649-6650.

Song, L., Jenner, M., Masschelein, J., Jones, C., Bull, M., Harris, S. and Hartkoorn, R. *et al.* (2017). Discovery and Biosynthesis of Gladiolin: A Burkholderia gladioli Antibiotic with Promising Activity against Mycobacterium tuberculosis. *Journal of the American Chemical Society,* **139**:7974-7981.

Spilker, T., Baldwin, A., Bumford, A., Dowson, C., Mahenthiralingam, E. and LiPuma, J. (2009). Expanded Multilocus Sequence Typing for *Burkholderia* Species. *Journal of Clinical Microbiology* **47**:2607-2610.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics,* **30**:1312–3.

Stopnisek, N., Bodenhausen, N., Frey, B., Fierer, N., Eberl, L. and Weisskopf, L. (2013). Genus-wide acid tolerance accounts for the biogeographical distribution of soil *Burkholderia* populations. *Environmental Microbiology* **16**:1503-1512.

Stopnisek, N., Zühlke, D., Carlier, A., Barberán, A., Fierer, N., Becher, D. and Riedel, K. *et al.* (2015). Molecular mechanisms underlying the close association between soil *Burkholderia* and fungi. *The ISME Journal* **10**:253-264.

Storms, V., Van Den Vreken, N., Coenye, T., Mahenthiralingam, E., LiPuma, J., Gillis, M. and Vandamme, P. (2004). Polyphasic Characterisation of *Burkholderia cepacia*-Like Isolates Leading to the Emended Description of *Burkholderia pyrrocinia*. *Systematic and Applied Microbiology* **27**:517-526.

Suárez-Moreno, Z., Caballero-Mellado, J., Coutinho, B., Mendonça-Previato, L., James, E. and Venturi, V. (2012). Common Features of Environmental and Potentially Beneficial Plant-Associated *Burkholderia*. *Microbial Ecology,* **63**:249-266.

Tange, O. 2018. GNU Parallel 2018. Available at: https://zenodo.org/record/1146014#.WzyseVVKiUk [Accessed: 9 February 2020]

Tettelin, H., Masignani, V., Cieslewicz, M., Donati, C., Medini, D., Ward, N. and Angiuoli, S. *et al.* (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial "pan-genome." *Proceedings of the National Academy of Sciences* **102**:13950-13955.

Trân Van, V., Berge, O., Ngô Kê, S., Balandreau, J. and Heulin, T. (2000). Repeated beneficial effects of rice inoculation with a strain of *Burkholderia vietnamiensis* on early and late yield components in low fertility sulphate acid soils of Vietnam. *Plant and Soil* **2**:273-284.

Turton, J., Arif, N., Hennessy, D., Kaufmann, M. and Pitt, T. (2007). Revised Approach for Identification of Isolates within the *Burkholderia cepacia* Complex and Description of Clinical Isolates Not Assigned to Any of the Known Genomovars. *Journal of Clinical Microbiology,* **45**:3105-3108.

Uehlinger, S., Schwager, S., Bernier, S., Riedel, K., Nguyen, D., Sokol, P. and Eberl, L. (2009). Identification of Specific and Universal Virulence Factors in *Burkholderia cenocepacia* Strains by Using Multiple Infection Hosts. *Infection and Immunity,* **77**:4102-4110.

Van Tran, V., Mavingui, P., Berge, O., Balandreau, J., Heulin, T. (1996) *Burkholderia vietnamiensis*, a new nitrogen-fixing species associated with rice roots, isolated from an acid sulphate soil in Vietnam: plant-growth-promoting effects on rice. In: *Biological Nitrogen Fixation Associated with Rice Production.* Dordrecht: Springer Netherlands, pp. 181–190.

Vandamme, P., Holmes, B., Vancanneyt, M., Coenye, T., Hoste, B., Coopman, R. and Revets, H. *et al.* (1997). Occurrence of Multiple Genomovars of *Burkholderia cepacia* in Cystic Fibrosis Patients and Proposal of *Burkholderia multivorans sp. nov. International Journal of Systematic Bacteriology* **47**:1188-1200.

Vandamme, P., Henry, D., Coenye, T., Nzula, S., Vancanneyt, M., LiPuma, J. and Speert, D. *et al.* (2002). *Burkholderia anthina* sp. nov. and *Burkholderia pyrrocinia*, two additional *Burkholderia cepacia* complex bacteria, may confound results of new molecular diagnostic tools. *FEMS Immunology & Medical Microbiology,* **33**:143-149.

Vandamme, P., Holmes, B., Coenye, T., Goris, J., Mahenthiralingam, E., LiPuma, J. and Govan, J. (2003). *Burkholderia cenocepacia sp.* nov.—a new twist to an old story. *Research in Microbiology*, **154**:91-96.

Vanlaere, E., LiPuma, J., Baldwin, A., Henry, D., De Brandt, E., Mahenthiralingam, E. and Speert, D. et al. (2008). *Burkholderia latens sp. nov., Burkholderia diffusa sp. nov., Burkholderia arboris sp. nov., Burkholderia seminalis sp. nov.* and *Burkholderia metallica sp. nov.*, novel species within the *Burkholderia cepacia* complex. *International Journal of Systematic and Evolutionary Microbiology* **58**:1580-1590.

Vanlaere, E., Baldwin, A., Gevers, D., Henry, D., De Brandt, E., LiPuma, J. and Mahenthiralingam, E. *et al.* (2009). Taxon K, a complex within the *Burkholderia cepacia* complex, comprises at least two novel species, *Burkholderia contaminans* sp. nov. and *Burkholderia lata* sp. nov. *International Journal of Systematic and Evolutionary Microbiology* **59**:102-111.

Vial, L., Groleau, M. C., Dekimpe, V. and Déziel, E. (2007) *Burkholderia* diversity and versatility: an inventory of the extracellular products. *J Microbiol Biotechnol,* **17**:1407–1429

Viallard, V., Poirier, I., Cournoyer, B., Haurat, J., Wiebkin, S., Ophel-Keller, K. and Balandreau, J. (1998). *Burkholderia graminis* sp. nov., a rhizospheric *Burkholderia* species, and reassessment of [Pseudomonas] *phenazinium*, [*Pseudomonas*] *pyrrocinia* and [*Pseudomonas*] *glathei* as *Burkholderia*. *International Journal of Systematic Bacteriology*, **48**:549-563.

Vermis, K., Coenye, T., Vandamme, P., Nelis, H. and Mahenthiralingam, E. (2002). Evaluation of species-specific recA-based PCR tests for genomovar level identification within the *Burkholderia cepacia* complex. *Journal of Medical Microbiology* **51**:937-940.

Walker, B., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S. and Cuomo, C. *et al.* (2014). Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS ONE* **9**:e112963.

Wallner, A., King, E., Ngonkeu, E., 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:**803

Walsh, P., Metzger, D. and Higuchi, R. (2013). Chelex 100 as a Medium for Simple Extraction of DNA for PCR-Based Typing from Forensic Material. *BioTechniques* **54**:506-513

Wang, X., Liu, A., Guerrero, A., Liu, J., Yu, X., Deng, P. and Ma, L. *et al.* (2016). Occidiofungin is an important component responsible for the antifungal activity of *Burkholderia pyrrocinia* strain Lyc2. *Journal of Applied Microbiology* **120**:607-618.

Wang, C., Zhou, X., Guo, D., Zhao, J., Yan, L., Feng, G. and Gao, Q. *et al.* (2019) Soil pH is the primary factor driving the distribution and function of microorganisms in farmland soils in North-eastern China. *Annals of Microbiology* **69:**1461–1473

Webster, G., Parkes, R. J., Cragg, B. A., Newberry, C. J., Weightman, A. J., and Fry, J. C. (2006). Prokaryotic community composition and biogeochemical processes in deep subseafloor sediments from the Peru Margin. *FEMS microbiology ecology*, **58**: 65–85.

Wick, R., Judd, L., Gorrie, C. and Holt, K. (2017). Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLOS Computational Biology* **13**:e1005595.

Wood, D. and Salzberg, S. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology* **15**:R46.

Yabuuchi, E., Kosako, Y., Oyaizu, H., Yano, I., Hotta, H., Hashimoto, Y., Ezaki, T. and Arakawa, M. (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. *Microbiology and Immunology*, **36**:1251-1275.

Yabuuchi, E., Kawamura, Y., Ezaki, T., Ikedo, M., Dejsirilert, S., Fujiwara, N., *et al.* (2000) *Burkholderia uboniae* sp. nov., L-arabinose-assimilating but different from *Burkholderia thailandensis* and *Burkholderia vietnamiensis*. *Microbiology and Immunology,* **44**:307–317.

Yang, B., Wang, Y. and Qian, P. Y. (2016). Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC bioinformatics*, **17**:135.

Zerikly, M. and Challis, G. (2009). Strategies for the Discovery of New Natural Products by Genome Mining. *ChemBioChem* **10**:625-633.

Zhang, H., Hanada, S., Shigematsu, T., Shibuya, K., Kamagata, Y., Kanagawa, T. and Kurane, R. (2000). *Burkholderia kururiensis* sp. nov., a trichloroethylene (TCE)-degrading bacterium isolated from an aquifer polluted with TCE. I*nternational Journal of Systematic and Evolutionary Microbiology*, **50**:743-749.

Zhao, K., Penttinen, P., Chen, Q., Guan, T., Lindström, K., Ao, X. and Zhang, L. *et al.* (2012). The rhizospheres of traditional medicinal plants in Panxi, China, host a diverse selection of *Actinobacteria* with antimicrobial properties. *Applied Microbiology and Biotechnology* **94**:1321-1335.

Zlosnik, J., Zhou, G., Brant, R., Henry, D., Hird, T., Mahenthiralingam, E. and Chilvers, M. *et al.* (2015). *Burkholderia* species Infections In Patients With Cystic Fibrosis In British Columbia, Canada. 30 Years' Experience. *Annals of the American Thoracic Society,* **12:**70-78.