The biotechnological potential of natural populations of *Burkholderiales* bacteria for antibiotic production

Thesis presented for the degree of Philosophiae Doctor

By

Amal Saeed Alswat

In candidature for the degree of Philosophiae Doctor

Organisms and the Environment Division

Cardiff School of Biosciences

Cardiff University

10 July 2020

Summary

Background. This study aimed to discover novel antibiotics produced by *Burkholderia* and *Paraburkholderia* bacteria as interesting sources of specialized metabolites. A novel collection of environmental *Burkholderia* and *Paraburkholderia* from the Bornean jungle, Sabah, Malaysia was assembled for the analysis.

Methods. After evaluation of different growth media, a total of 98 environmental samples were screened for the presence of these bacteria by enrichment on *Pseudomonas cepacia* azelaic acid tryptamine (PCAT) medium. Molecular identification using the *recA* and 16S rRNA gene was performed and a total of 57 isolates were genome sequenced ahead of phylogenomic analyses. Antibiotic production by the jungle strains was evaluated using an antimicrobial antagonism overlay assay, genome mining, and high performance liquid chromatography (HPLC) analysis.

Results. Over 50% of the 98 jungle samples were *Burkholderia/Paraburkholderia recA* PCR-positive. A total of 123 jungle isolates were purified and draft genome sequences were obtained for 57 of them. Genomic taxonomy identified that 45 isolates (85%) were within the newly described *Paraburkholderia*, and 13 isolates (15%) grouped within *Burkholderia*. Within the *Paraburkholderia*, 22 isolates were likely *Paraburkholderia tropica*, with a further 22 representing potentially novel taxa. All 12 *Burkholderia* isolates were members of the *Burkholderia cepacia* complex, including three that were likely a novel species group. The *Burkholderia* strains were all bioactive, but *Paraburkholderia* did not show antimicrobial activity under the conditions tested. Genome mining using antiSMASH showed that the capacity of *Burkholderia* to encode antibiotic biosynthesis gene clusters (BGCs) was greater than *Paraburkholderia*.

Conclusions. The Bornean jungle in Sabah, Malaysia, is a rich source of *Burkholderia* and *Paraburkholderia* bacteria. Classification of these bacteria using genomic taxonomy approaches accurately identified known and novel species within the collection. The

environmental *Burkholderia cepacia* complex (Bcc) jungle strains were promising antimicrobial producers, however, despite their taxonomic diversity, the *Paraburkholderia* did not show any antimicrobial activity

Acknowledgement

I would like to express my sincere gratitude to my PhD supervisor Prof. Eshwar Mahenthiralingam for the continuous support of my PhD study, and for his patience, and motivation during my study. His valuable guidance helped me all the time of my PhD and writing of this thesis.

I gratefully acknowledge the funding received towards my PhD research from Taif university, and I would like to thank the Royal Embassy of Saudi Arabia 'Cultural Bureau' in London for their encouragement and ultimate supports during my PhD journey.

I greatly appreciate my second supervisors Prof. Andrew Weightman for giving the encouragement and sharing insightful suggestions. I am grateful to Esh lab colleagues for providing me with assistance throughout my study, especially to Dr Gordon Webster for always being so supportive of my work.

I would also like to say a heartfelt thank you to my dad, mum and brothers for all the support you have shown me through this research. A very special thank you to my friends Mrs. Roaa Alnemari and Ms. Hanouf Bafhaid for helping in whatever way they could during this challenging time of my life.

Contents

List of figures

List of Tables

List of Abbreviation

16S rRNA	16S ribosomal RNA
AMR	Antimicrobial resistance
ANI	Average nucleotide identity
AntiSMASH	Antibiotics and secondary metabolite analysis shell
ATCC	American type Culture Collection
AutoMLST	Automated multi-locus species tree
Barrnap	Basic rapid ribosomal rna predictor
BCC	*Burkholderia* cardiff university collection
Bcc	*Burkholderia cepacia* complex
BCEM	*Burkholderia cepacia* complex enrichment
BCSA	*Burkholderia cepacia* selective agar
BGC	Biosynthesis gene clusters
BLAST	Basic local alignment search tool algorithm
BNF	Biological nitrogen fixation
Bp	Base pairs
BSM-G	Basal salts medium supplied with 0.4% (w/v) glycerol
BSM-GBP	Basal salts medium supplied with 0.4% (w/v) glycerol bromoacetic acid and polymixin
BSM-GC	Basal salts medium supplied with 0.4% (w/v) glycerol and 100 mg/µL cycloheximide
CDPS	tRNA-dependent cyclodipeptide synthases
CF	Cystic fibrosis
CLIMB	Cloud infrastructure for microbial bioinformatics
Cm	Centimeter
COGs	Clusters of orthologous groups

DDH	DNA-DNA hybridization
DMSO	Dimethylsulphoxide
DNA	Dioxyribonucleic acid
dNTPs	Deoxynucleoside triphosphates
DNA	Deoxyribonucleic acid
EU	European union
Fix	Nitrogen fixation genes
GC	Guanine-cytosine
H	Hour
hglE-KS	Heterocyst glycolipid synthase-like PKS
HPLC	High performance liquid chromatograph
Kb	Kilobases
KEGG	Kyoto encyclopedia of genes and genomes
Mb	Megabases
MDR-TB	Multidrug-resistant tuberculosis
MEGA	Molecular evolutionary genetics analysis
MIC	Minimum inhibitory concentration
Min	Minute
MLST	Multilocus sequence typing
ML	Maximum likelihood
Ml	Millimeter
MRSA	Methicillin-resistant *Staphylococcus aureus*
NBRC	NITE biological resource center
NCBI	National center for biotechnology information
Nif	Nitrogen fixation genes
NJ	Neighbor joining
NRPS	Non-ribosomal peptide synthetase cluster

NMR	Nuclear magnetic resonance
nod	Nodulation genes
OD	Optical density
OSMAC	One strain many compounds
RAPD-PCR	Random amplification of polymorphic DNA PCR
PCAT	*Pseudomonas cpacia* azelaic acid
PCR	Polymerase chain reaction
PKS	Polyketide synthase
PGPB	Plant growth-promoting bacteria
PGPR	Plant growth-promoting rhizobacteria
PpyS-KS	PPY-like pyrone cluster
recA	DNA recombination
RiPP	Ribosomally synthesized and post-translationally modified peptides
transAT-	Trans-AT PKS
TSA	Tryptone soya agar
TSB	Tryptone soya broth
TS	Type strain
WHO	World health organization

Chapter 1. General introduction and aim

1.1 General introduction

Burkholderia bacteria are relatively untapped as a source of antimicrobial and specialised metabolites. Recent studies have highlighted the ability of *Burkholderia* as a promising source of antibiotics that can overcome antimicrobial resistance in a range of pathogens (Mahenthiralingam et al. 2011; Song et al. 2017; Mullins et al. 2019). Bacteria within the genus *Burkholderia* have broad and diverse lifestyles. They are both pathogens in humans, animals, and plants, as well as symbiotic organisms that interact with multiple hosts for benefit (Depoorter et al. 2016). From a biotechnological perspective, they are producers of a wide range of antimicrobials and enzymes of interest, and have been used as bioremediation and biopesticial agents (Depoorter et al. 2016; Eberl and Vandamme 2016). Over the last 7 years, the taxonomy of the genus *Burkholderia* has been split into five further genera: *Paraburkholderia*, *Caballeronia*, *Robbsia*, *Mycetohabitans*, *Trinickia* (Table 1.1) (Dobritsa and Samadpour 2019). Thus, the expression of *Burkholderia* sensu lato will be used throughout this chapter to cover all of these genera (Kunakom and Eustáquio 2019). This study explored the diversity of *Burkholderia* sensu lato bacteria that occur in a tropical rain forest environment, the jungle in Sabah, Malaysia, seeking to understand their diversity and ability to produce novel antibiotics. Specific features of *Burkholderia* sensu lato and the need for antibiotic discovery are discussed below.

1.2 *Burkholderia* sensu lato bacteria (1992-2020)

Burkholderiaceae is a bacterial family that belongs to the *Burkholderiales* order under the *Betaproteobacteria* class within the *Proteobacteria* phylum, a Gram-negative bacterial phylum. Bacteria within the *Burkholderiaceae* family have an extremely diverse ecology, containing environmental organisms and opportunistic pathogens of humans, plants and animals.

Working with the *Burkholderiaceae* family over the past 80 years, microbiologists have shed considerable light on the biology of the *Burkholderia* genus. In the mid-1940s, Walter H. Burkholder described *Pseudomonas* (*Burkholderia*) *cepacia* as a plant pathogenic bacterium that caused onion bulb rot disease (Burkholder 1950). The genus was renamed in reference to Burkholder in 1992, with the new name *Burkholderia* proposed, and the following seven *Pseudomonas* species were transferred to the *Burkholderia: Pseudomonas* (*Burkholderia*) *gladioli*, *Pseudomonas* (*Burkholderia*) *pseudomallei*, *Pseudomonas* (*Burkholderia*) *caryophylli*, *Pseudomonas* (*Burkholderia*) *pickettii*, *Pseudomonas* (*Burkholderia*) *cepacia*, *Pseudomonas* (*Burkholderia*) *mallei* and *Pseudomonas* (*Burkholderia*) sloanacearum (Yabuuchi et al. 1992). Later, *Pseudomonas plantarii* and *Pseudomonas glumae* were transferred to the *Burkholderia* genus (Urakami et al. 1994), and *Pseudomonas glathei* and *Pseudomonas pyrrocinia* were reclassified as *Burkholderia* species (Vandamme et al. 1997). *Burkholderia pickettii* and *Burkholderia solanacearum* were transferred later to the new genus *Ralstonia* (Yabuuchi et al. 1995).

Over the last 28 years since initial designation, the *Burkholderia* genus and multiple related species have been discovered and studied (Yabuuchi et al. 1992; Vandamme and Peeters 2014; Depoorter et al. 2016; Ragupathi and Veeraraghavan 2019). The genus currently comprises at least 90 validly named species and will likely include many more as research progresses (Depoorter et al. 2016). Recently, between 2014 and 2020 the diversity within *Burkholderiaceae*, and specifically the genus *Burkholderia* was extended, splitting it and adding new groups to give an additional six genera (Dobritsa and Samadpour 2016). Multiple *Burkholderia* members have been transferred to the following new genera: *Paraburkholderia, Caballeronia, Robbsia, Mycetohabitans, Trinickia* and others retained within the remaining *Burkholderia* classification (Table 1.1) (Dobritsa and Samadpour 2019). In this thesis, *Burkholderia* sensu lato bacteria refers to all the species left within the originally described *Burkholderia* genus as follows.

Table 1.1 The history of *Burkholderia* sensu lato (1992-2020)

***Burkholderia* sensu lato genera**	**Proposed**	**Genera discribtion**	**References**
Burkholderia	1992	This genus contains pathogenic species divided into three subgroups (*Burkholderia cepacia* complex (Bcc), *Burkholderia pseudomallei*, and *Burkholderia gladioli* groups).	(Yabuuchi et al. 1992)
Paraburkholderia	2014	This genus contains non-pathogenic and environmental isolates.	(Sawana et al. 2014)
Caballeronia	2016	This genus contains environmental isolates and non-nitrogen-fixing bacteria	(Dobritsa and Samadpour 2016)
Robbsia	2017	This genus contains only one species (*Robbsia andropogonis*). This species has large genotypic differences from other species of *Burkholderia* sensu lato.	(Lopes-Santos et al. 2017)
Mycetohabitans	2018	This genus contains two fungal symbiont species (*Mycetohabitans rhizoxinica* and *Mycetohabitans endofungorum*).	(Estrada-de los Santos et al. 2018)
Trinickia	2018	This genus contains four species (*Trinickia caryophylli, Trinickia dabaoshanensis, Trinickia symbiotica and Trinickia soil*), whichhave large genotypic differences from other species of *Burkholderia* sensu lato.	(Estrada-de los Santos et al. 2018)

1.2.1 The *Burkholderia* genus and examples of well-described species groups within it

1.2.1.1 *Burkholderia cepacia* complex (Bcc)

The *Burkholderia cepacia* complex (Bcc) is considered the prototypic representative of the *Burkholderia* genus species because it contains the original Walter Burkholder characterised onion rot strain (Vandamme et al. 1997). As a group Bcc contains plant, human and animal pathogens, and a wide range of environmental strains and species (Depoorter et al. 2016; Ragupathi and Veeraraghavan 2019). The Bcc is a closely related group of *Burkholderia* species that reside within the *Burkholderia* clade, as demonstrated in multiple taxonomic studies (Sawana et al. 2014; Depoorter et al. 2016). Initially, novel species were described as genomovars and were hard to distinguish phenotypically (Vandamme et al. 1997). The term ‘genomovars’ in bacterial taxonomy includes several meanings, such as genomic species, genomic groups, genospecies or genomospecies, and it is used to define distinct groups of strains that are genotypically separate but phenotypically similar (Vandamme et al. 1997). Vandamme and colleagues proposed in 1997 that phenotypically variable *B. cepacia* isolates,

many from lung infections in people with cystic fibrosis (CF), were members of at least five closely related species within the *B. cepacia* complex (Vandamme et al. 1997). Further genomovar analysis by 2010 showed that the *B. cepacia* complex group contained at least seventeen species, with those found in CF being commonly members of: *Burkholderia cepacia* (genomovar I), *Burkholderia multivorans* (genomovar II), *Burkholderia cenocepacia* (genomovar III), *Burkholderia stabilis* (genomovar IV), *Burkholderia vietnamiensis* (genomovar V), *Burkholderia dolosa* (genomovars VI), *Burkholderia ambifaria* (genomovar VII), *Burkholderia anthina* (genomovar VIII) and *Burkholderia pyrrocinia* (genomovar IX).

The taxonomy of *Burkholderia* genus in particular and *Burkholderia* lato senso has continued to undergo considerable changes over the last decade based on multiple phenotypic and genotypic analysis methods, and in particular the wider application of whole genome sequencing. Further novel Bcc species have been defined, and this group now includes over 20 closely-related species with valid names that occupy diverse ecological niches: *Burkholderia cepacia, Burkholderia pyrrocinia, Burkholderia ubonensis, Burkholderia stabilis, Burkholderia vietnamiensis, Burkholderia multivorans, Burkholderia territorii, Burkholderia anthina, Burkholderia seminalis, Burkholderia metallica, Burkholderia arboris, Burkholderia contaminans, Burkholderia lata, Burkholderia latens, Burkholderia dolosa, Burkholderia diffusa, Burkholderia paludis* and *Burkholderia stagnalis* (De Smet et al. 2015; Ong et al. 2016; Jin et al. 2020). The term genomovar is not used to describe species within the Bcc group unless the isolates concerned are novel and less well characterised. With the application of DNA sequence-based classification methods, the *Burkholderia* genus was reclassified and separated from *Paraburkholderia*, a group of environmentally-associated and largely non-pathogenic species (Sawana et al. 2014). The species left within *Burkholderia* contained more of the well know opportunistic human and plant pathogenic species. This *Burkholderia* genus currently comprises at least three main groups: the *Burkholderia cepacia* complex (Bcc), *Burkholderia pseudomallei* group and *Burkholderia gladioli* group (described below) (Sawana et al. 2014; Depoorter et al. 2016).

Bcc thrives in a wide range of environments such as soil, water, plant, animal, hospitals and industrial environments. Their properties represent a double-edged sword that can be either beneficial or harmful. Bcc have many benefits for agriculture as a plant growth-promoting bacteria and biocontrol agents (Parke and Gurian-Sherman 2001). At the same time, they can be problematic infections causing lung infections in people with CF (LiPuma 2010). Bcc bacteria are the most common *Burkholderia* that cause chronic respiratory infections in CF patients, and *B. cenocepacia* and *B. multivorans* are the most frequently recovered species associated with CF (LiPuma 2010). CF is a chronic and progressive disorder that causes sticky mucus in the lungs and digestive system. CF patients require lifelong medical care that becomes increasingly costly and complex with age. This lethal condition leads to lung damage, which causes death for the majority of CF sufferers. CF patients are predisposed to bacterial infections caused by several pathogens including the *Burkholderia cepacia* complex*, Pseudomonas aeruginosa* and *Staphylococcus aureus* (LiPuma 2010; Martina et al. 2013; Coutinho et al. 2015; Medina-Pascual et al. 2015). Because Bcc species are resistant to a large range of antibiotics, such as polymyxin B, tetracycline, chloramphenicol and tobramycin, CF patients with *Burkholderia* infections cannot easily be treated with currently available antibiotics, which leads to their morbidity and mortality (Aaron et al. 2000; Nzula et al. 2002). Additionally, the Bcc group contains causative agents of plant disease such as prototypic *Burkholderia* species, *Burkholderia cepacia,* which was initially known as *Burkholderia 'Psedomonas' cepacia,* that cause the rot disease in onion bulbs (Burkholder 1950).

These CF associated species can also be found in the natural environment and a number of other locations including as contaminants in non-sterile industrial products (Rushton et al. 2013; Cunningham-Oakes et al. 2019), within terrestrial water bodies (Peeters et al. 2016), soil (Peeters et al. 2016) and associated with the rhizosphere of plants (Parke and Gurian-Sherman 2001). Moreover, Bcc bacteria have also been shown to have beneficial interactions with plants. For example, *B. vietnamiensis* is a nitrogen-fixing bacteria (Suárez-Moreno et al.

2012), while *B. cenocepacia* and *B. ambifaria* have been shown to promote biocontrol and plant growth-promoting activity (Parke and Gurian-Sherman 2001; Ho and Huang 2015).

1.2.1.2 *Burkholderia pseudomallei* group

The *B. pseudomallei* group or complex is a phylogenetically distinct clade within the *Burkholderia* spp. This group contains five identified members: *B. pseudomallei*, *B. mallei*, *B. thailandensis*, *B. humptydooensis* and *B. oklahomensis* (Depoorter et al. 2016; Sahl et al. 2016; Tuanyok et al. 2017). *B. pseudomallei* is a causative bacterium of melioidosis, which is a fatal disease in humans and animals (Dance 1990; Yabuuchi et al. 1992). In animals, *B. mallei* causes the infectious disease glanders, which can be transmitted to humans through direct contact with infected animals (Schadewaldt 1975; Yabuuchi et al. 1992; Neubauer et al. 2005). *B. thailandensis*, *B. humptydooensis* and *B. oklahomensis*, the other three species in the *B. pseudomallei* group, are isolated from the environment and considered non-pathogenic species (Tuanyok et al. 2017). *B. thailandensis* is a soil bacterium isolated in Thailand (Brett et al. 1997); *B. oklahomensis* is a clinical sample from a thigh wound contaminated with soil from a 27-year-old farmer in Oklahoma, USA (McCormick et al. 1977; Glass et al. 2006b). A few strains of *B. oklahomensis* and *B. thailandensis* have been reported to cause human infections (Glass et al. 2006a; Glass et al. 2006b; Gee et al. 2018). The last *B. pseudomallei* group member is *B. humptydooensis*, which was isolated from a bore water sample in the Northern Territory in Australia (Tuanyok et al. 2017).

1.2.1.3 *Burkholderia gladioli* group

Burkholderia gladioli group contains three identified members. *Burkholderia plantarii, Burkholderia gladioli* and *Burkholderia glumae* are plant pathogenic *Burkholdera* that infect rice (Maeda et al. 2006; Nandakumar et al. 2009) and other commercial crops such as mushrooms (Chowdhury and Heinemann 2006). *B. gladioli* is a common *Burkholderia* species associcated with lung infections in CF patients (LiPuma 2010), *B. gladioli* BCC0238 being isolated from the sputum of a CF patient (Song et al. 2017). In a survey of 2204 US CF patients infected with *Burkholderia*, *B. gladioli* was the third most common species (15% of case) from

1997 to 2007 (LiPuma 2010). A recent survey of the UK from 2013 to 2015 showed that *B. gladioli* was also the encountered in 16% of the *Burkholderia* infected patients (Kenna et al. 2017).

1.2.2 *Paraburkholderia* genus (2014-2020)

Based on molecular approaches that considered single gene sequence analysis as well as others that examined whole genome sequence data, a number of studies have determined that certain *Burkholderia* species isolated from the natural environment, and generally non-pathogenic, clustered phylogenetically in a separate clade of *Burkholderia* sensu lato (Bontemps et al. 2010; Gyaneshwar et al. 2011; Estrada-De Los Santos et al. 2013; Sawana et al. 2014; Zuleta et al. 2014). By 2014, alongside conserved protein gene and 16S rRNA gene phylogenies, sequence data was used to divide the *Burkholderia* genus into two genera, that retained several species as *Burkholderia* but also proposed the new group *Paraburkholderia* gen. nov. (Sawana et al. 2014). *Burkholderia* Clade 1, which contains environmental, animal and plant pathogenic species, has retained the genus name. Clade 2, which contained a predominance of environmental species, has been named *Paraburkholderia* (Sawana et al. 2014; Depoorter et al. 2016; Dobritsa et al. 2016).

As a result, the *Paraburkholderia* genus contains predominantly environmental species (i.e. the genus was proposed to accommodate the environmental *Burkholderia*). Examples of the first candidates transferred from *Burkholderia* to *Paraburkholderia* include *Paraburkholderia* '*Burkholderia*' *caribensis* (Achouak et al. 1999), *Paraburkholderia* '*Burkholderia*' *kururiensis* (Zhang et al. 2000), *Paraburkholderia 'Burkholderia' tropica* (Reis et al. 2004), *Paraburkholderia 'Burkholderia' xenovorans* (Goris et al. 2004), *Paraburkholderia* '*Burkholderia*' *ferrariae* (Valverde et al. 2006), *Paraburkholderia* '*Burkholderia*' *caballeronis* (Martínez-Aguilar et al. 2013), *Paraburkholderia* '*Burkholderia*' *aspalathi* (Mavengere et al. 2014) and others, as stated in Sawana's and his group work (Sawana et al. 2014). Dobritsa and Samadpour in 2016 and 2017 transferred 14 *Burkholderia* species to the *Paraburkholderia* genus: *Paraburkholderia dipogonis, Paraburkholderia ginsengiterrae, Paraburkholderia*

humisilvae, Paraburkholderia insulsa, Paraburkholderia kirstenboschensis, Paraburkholderia metalliresistens, Paraburkholderia monticola, Paraburkholderia panaciterrae, Paraburkholderia rhizosphaerae, Paraburkholderia solisilvae, Paraburkholderia susongensis, Paraburkholderia panaciterrae, *Paraburkholderia ginsengiterrae* and *Paraburkholderia jirisanensis* (Dobritsa et al. 2016; Dobritsa and Samadpour 2016; Dobritsa et al. 2017).

The number of studies that describe species within *Paraburkholderia* genus have increased over the last six years, and show that they are ecologically widespread. This genus is composed of plant growth-promoting bacteria (PGPB) and plant growth-promoting rhizobacteria (PGPR) species. *Paraburkholderia nodosa*, *Paraburkholderia dilworthii*, *Paraburkholderia caballeronis*, *Paraburkholderia tuberum*, *Paraburkholderia mimosarum* and others are species that promote nodulation and nitrogen fixation, which are prominent environmental features of *Paraburkholderia* (Suárez-Moreno et al. 2012). Furthermore, *Paraburkholderia* imparts other benefits with multiple environmental hosts including: to protect against plant pathogens, to improve nutrient uptake, to induce systemic resistance and to increase stress tolerance (Segura et al. 2009; Sawana et al. 2014; Vitorino and Bessa 2017). For example, *P. phytofirmans* (PsJN) promotes the growth of *Arabidopsis thaliana*, a model flowering plant (Timmermann et al. 2017), as well being a promising biological control agent for reducing disease caused by the plant pathogen *Pseudomonas syringae* (Esmaeel et al. 2019). *Paraburkholderia caballeronis* (TNe-841T) is a tomato plant-associated bacterium with the ability to fix nitrogen (Rojas-Rojas et al. 2017). Multiple studies describe the plant-bacteria interactions of *Paraburkholderia* species, indicating their importance and potential applications in plant pathogen biocontrol, biofertilization and bioremediation (Sawana et al. 2014; Eberl and Vandamme 2016). The use of plant-associated bacteria is a promising alternative biocontrol strategy in agriculture, as it could reduce the overuse of pesticides and the resulting negative environmental impacts (Esmaeel et al. 2019; Mullins et al. 2019).

New *Paraburkholderia* members with beneficial traits have been isolated and added to this genus in the last few years. For example, in China, *Paraburkholderia caffeinilytica* was isolated

from the soil of a tea plantation (Gao et al. 2016), while, in Korea, *Paraburkholderia azotifigens*, a nitrogen-fixing bacterium, was isolated from rice paddy soil (Choi and Im 2017). *Paraburkholderia panacihumi* was isolated from ginseng-cultivated soil and shown to be antagonistic against root rot fungal pathogens in Korea (Huo et al. 2018). *Paraburkholderia phosphatilytica*, a phosphate-solubilizing bacterium, and *Paraburkholderia caseinilytica* were both isolated from forest soil in China (Gao et al. 2018a; Gao et al. 2018b). *Paraburkholderia fynbosensis* is a bacterium symbiotic with *Lebeckia ambigua* root nodules that has shown the ability to nodulate and fix nitrogen in South Africa (De Meyer et al. 2018). An aromatic hydrocarbon-degrading *Paraburkholderia* species called *Paraburkholderia aromaticivorans* was isolated from gasoline-contaminated soil, and first described in 2018, and another strain of *Paraburkholderia aromaticivorans* (BTEX) was isolated from petroleum-contaminated soil associated with the biodegradation of naphthalene (Lee and Jeon 2018; Lee et al. 2019). The well characterised *Paraburkholderia xenovorans* was also isolated as polychlorinated hydrocarbon pollutant degrading species (Chain et al. 2006).

Currently, the genus *Paraburkholderia* contains at least 75 validly recognized species (http://www.bacterio.net/paraburkholderia.html). With the expanded understanding of the species within *Burkholderia* sensu lato and newly developed genomic-based taxonomic approaches, more *Paraburkholderia* species members have been added to this genus. This has led to the reclassification of certain *Burkholderia* species to *Paraburkholderia,* or from *Paraburkholderia* to other more recently proprosed genera. For example, the *Caballeronia* genus was proposed during the reclassification of a clade of *Burkholderia* species to *Paraburkholderia* (Dobritsa and Samadpour 2016). Current taxonomic studies have reclassified '*Burkholderia ultramafica'* as *Paraburkholderia ultramafica* (Dobritsa and Samadpour 2019). Several *Burkholderia* genus species have been transferred to the novel genera *Caballeronia, Robbsia, Mycetohabitans* and *Trinickia* (Dobritsa and Samadpour 2019). Most recently, various *Paraburkholderia* species have been reclassified to other genera, as

described below, because they are distinctly grouped from both *Burkholderia* and *Paraburkholderia* genera by phylogenetic analysis.

1.2.3 *Caballeronia* genus (2016 – 2020)

The proposal of the *Caballeronia* genus came through the classification of *Paraburkholderia* based on 16S rRNA gene phylogeny and sequence data (Sawana et al. 2014; Depoorter et al. 2016). *Caballeronia* was established in 2016, and by 2019 it comprised at least 29 validly established species (https://lpsn.dsmz.de/genus/caballeronia). One species (*Caballeronia glathei*) was reclassified from the *Pseudomonas* genus, and 26 species belonged to the *Burkholderia* genus (Dobritsa and Samadpour 2019; Quan et al. 2019; Uroz and Oger 2019). Just over the last year, more *Caballeronia* species have been added. *Caballeronia ginsengisoli* and *Caballeronia mineralivorans* are new *Caballeronia* species that were isolated in 2019 from ginseng cultivating soil and oak-Scleroderma citrinum mycorrhizosphere, respectively (Quan et al. 2019; Uroz and Oger 2019). Nearly all *Caballeronia* species are environmental isolates, except for *Caballeronia zhejiangensis,* as strains of this species were isolated from human clinical samples (Vandamme et al. 2013; Dobritsa and Samadpour 2016). The *Caballeronia* species are not nitrogen-fixing bacteria, as the *nif* genes have not been predicted on them (Estrada-de los Santos et al. 2018).

1.2.4 Other genera not accommodated by the *Burkholderia*, *Paraburkholderia* and *Caballeronia* genera: *Robbsia*, *Mycetohabitans* and *Trinickia*

The taxonomic analyses and studies of *Burkholderia* sensu lato based on 16S rRNA gene, housekeeping gene sequences and whole genome comparisons by several groups of researchers led them to propose the new genus of *Robbsia*. Moreover, it was clear from previous studies that had reclassified the historical *Burkholderia* genus to *Burkholderia*, *Paraburkholderia* and *Caballeronia,* that one species "*Robbsia andropogonis"* was a highly distinct clade (Viallard et al. 1998; Estrada-De Los Santos et al. 2013; Sawana et al. 2014; Dobritsa and Samadpour 2016). *Robbsia andropogonis* was originally isolated from stripe

disease in sorghum and known as *Bacterium andropogoni* (Smith 1911). In 1925, it was reclassified as *Pseudomonas andropogonis*, and in 1995 transferred to the *Bukholderia* genus (Yabuuchi et al. 1995). By 2014, *Burkholderia* sensu lato had been split into two main clusters, *Burkholderia* and *Paraburkholderia*, with *Burkholderia andropogonis* placed in *Paraburkholderia andropogonis* (Sawana et al. 2014). By 2017, the taxonomy *Robbsia andropogonis* was clearly showing that it is genotypic differences from other species of *Burkholderia* sensu lato, and it represented *Robbsia* Gen. Nov (Lopes-Santos et al. 2017).

The research efforts to update the taxonomy of *Burkholderia* sensu lato (*Burkholderia*, *Paraburkholderia*, *Caballeronia* and *Robbsia*) revealed that there were two further unique groups of species that could not be accommodated to other validly determined genera. These groups, *Mycetohabitans* and *Trinickia,* were later classified as new genera, a determination strongly supported by multiple analyses including: phylogenetic analysis based on conserved genes and protein amino acid sequence, together with comprehensive genomic approaches (average nucleotide identity, ANI; and amino acid identity, AAI), as well as other lifestyle experiments (plant growth promotion and nodulation features) (Estrada-de los Santos et al. 2018). The first group contains two species*, Mycetohabitans 'Paraburkholderia' rhizoxinica* and *Mycetohabitans 'Paraburkholderia' endofungorum*; these species are all fungal symbionts species (Partida-Martinez et al. 2007; Estrada-de los Santos et al. 2018). The second group contains four species: *Trinickia 'Paraburkholderia' caryophylli, Trinickia 'Burkholderia' dabaoshanensis, Trinickia 'Paraburkholderia' symbiotica* and *Trinickia 'Paraburkholderia' soil* (Yabuuchi et al. 1992; Yoo et al. 2007; Sheu et al. 2012; Zhu et al. 2012; Estrada-de los Santos et al. 2018).

As can be seen above, from 1992 to 2020, the taxa originally classified within "*Burkholderia*" have grown rapidly, and gradually divided to encompass six distinct genera: *Burkholderia* sensu lato*, Paraburkholderia, Caballeronia, Robbsia, Trinickia* and *Mycetohabitans*. To help our understanding of this group of organisms, it is vital to elucidate their taxonomy and characterize new species and their interactions. This PhD study has expanded the field by

adding a unique collection of environmental isolates to *Burkholderia* sensu lato, *Paraburkholderia,* and *Caballeronia,* and identifying potential novel taxa within both using genomic methods. A second component of the study was to understand if these environmental isolates were bioactive and produced novel antibiotics. The threat of antibiotic resistance and the need to discover new antibiotics is discussed next.

1.3 The threat of antibiotic-resistant superbugs

Antimicrobial resistance (AMR) threatens the world and is considered one of the biggest contemporary health problems. According to the World Health Organization (WHO), AMR increases the rate of morbidity of patients with microbial infections and mortality, especially for vulnerable patients undergoing chemotherapy, surgery and dialysis. AMR causes economic losses for patients, the pharmaceutical industry and hospitals. Bacteria become tolerant to antibiotics through naturally occurring mutations, or by acquiring resistance-encoding genes from other bacteria. The overuse of antibiotics plays a significant role in the selection of resistant strains (Tacconelli et al. 2018).

According to the data from the WHO, AMR superbugs pose serious global threats (Tacconelli et al. 2018). For example, the number of mortalities due to multidrug-resistant bacteria reaches 25,000 people per year in the European Union (EU). Moreover, the economic effect of AMR in the US approaches $34 billion per year. Hence, medical experts have raised the alarm regarding the emergence of new strains of microorganisms, such as methicillin-resistant *Staphylococcus aureus* (MRSA), which are highly resistant to existing treatments. More than 20 classes of antibiotics have been discovered from 1930 to 2000 (Figure 1.1). Since 2000, the discovery of new antibiotic classes has been very limited, with only a few examples such as daptomycin and linezolid (Coates et al. 2011). To this end, new antibiotics, especially those effective against resistant strains of bacteria, are needed more than ever.

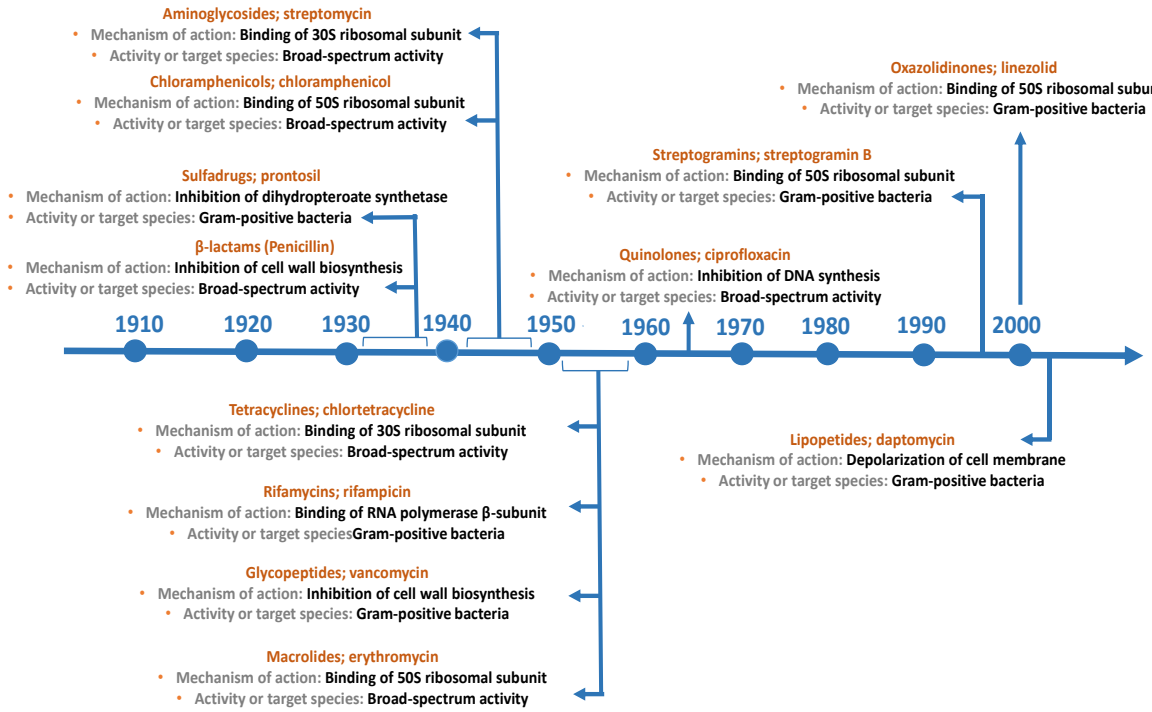


Figure 1.1 The timeline of antibiotics discovery and introduction of new drug classes. The major period of discovery of antibiotic classes was from 1940 to 1970, and few antibiotics have been discovered since 1970.

1.3.1 Urgent need for new antibiotics: *Burkholderia* as promising novel antibiotic producers

Actinomycetes, especially those within the genus *Streptomyces* are one of the largest and most well-known groups of antibiotic-producing bacteria. However, resistance to these classical antibiotics (Figure 1.1) has emerged and these agents have become less effective (de Lima Procópio et al. 2012). *Burkholderia* represents a relatively untapped source of novel antibiotics (Kunakom and Eustáquio 2019). The ability of *Burkholderia* to produce antifungal compounds is well known, with examples such as pyrrolnitrin, phenylpyrroles, cepaciamide, phenazines, quinolones and xylocandins (Parke and Gurian-Sherman 2001; Vial et al. 2007) (Table 1.2). But their ability to produce a wide range of specialized metabolites was not well characterised until more recently with the advent of genomic led methods. In 2011 the ability of *B. ambifaria* to antagonise *B. multivorans* was observed, and Mahenthiralingam et al. (2011) identified the biosynthetic pathway that *B. ambifaria* used to produce the anti-Gram

negative antibiotic, enacyloxin IIa (Mahenthiralingam et al. 2011). This study used conventional chemistry, genetics and genomics to fully characterize how *B. ambifaria* made enacyloxin.

The interest in antibiotics produced by *Burkholderia* has grown considerably in the last ten years. *Burkholderia* bacteria are a rich antibiotic resource, with unusual biosynthetic gene clusters encoding a range of bioactive natural products, many of which have been discovered by genome mining (Kunakom and Eustáquio 2019). The review re-emphasises that *Burkholderia* are an underexplored taxon with great potential for the discovery of new active compounds. At Cardiff University, the discovery of the potent activity of enacyloxins in 2011 by Mahenthiralingam and his group, led to the establishment of a *Burkholderia* antibiotic discovery program using genome-mining approaches (Mahenthiralingam et al. 2011). The Mahenthiralingam group has a large and unique collection of *Burkholderia* bacteria assembled from multiple studies, mainly characterizing them as CF pathogens (Mahenthiralingam et al. 2005; Mahenthiralingam et al. 2008); > 1900 isolates). Analysis to date has shown that individual *Burkholderia* strains may encode over 10 antibiotic biosynthetic pathways and produce multiple antibiotics (Depoorter et al. 2016). In 2017, the antibiotic discovery program between the Challis group (Warwick University) and Mahenthiralingam group, published the discovery and biosynthesis of gladiolin, which is produced by *Burkholderia gladioli*. Gladiolin shows promising antibacterial activity against various strains of *Mycobacterium tuberculosis* with MIC range (0.4 to > 2.3 *μ*g/mL for non-resistant strains) (Song et al. 2017). Building on a study that had isolated novel *Burkholderia*-like bacteria interacting with wood-decay fungi (Christofides et al. 2019), genomic analysis at Cardiff identified them as novel *Paraburkholderia* species and found that they had potential for antimicrobial biosynthesis (Webster et al. 2019). From a genome-mining approach, cepacin has been identified as a plant-protective metabolite and is produced by the biopesticidal bacterium *Burkholderia ambifaria* (Mullins et al. 2019).

Table 1.2 A selection of antibiotics produced by *Burkholderia*

Antibiotics	*Burkholderia* species or strain	References
Enacyloxin	*B. ambifaria*	(Mahenthiralingam et al. 2011)
Gladiolin	*B. gladioli* BCC0238	(Song et al. 2017)
Mupirocin-like polyketide	*B. thailandensis*	(Smith et al. 1997)
Quinolinone	*B. cepacia* strain PCII	(Moon et al. 1996)
AFC-BC11 (Lipo- peptides)	*B. cepacia* isolate BC11	(Kang et al. 1998)
Altericidins	*B. cepacia*	(Kirinuki et al. 1977)
Cepaciamids A and B	*B. cepacia*	(Jiao et al. 1996)
Cepacin A and Cepacin B	*B. cepacia*	(Parker et al. 1984)
CF66I	*B. cepacia* strain CF66	(Li et al. 2007; Li et al. 2008)
Phenazine	*B. cepacia* strain 5.5B	(Cartwright et al. 1995)
Pseudane	*B. cepacia* strain RB425	(Yoshihisa et al. 1989)
Pseudoanes	*B. cepacia* strain RB425	(Meyers et al. 1987)
Pyrrolnintrin	*B. cepacia* strain RB425 *B. cepacia* isolate B37w *B. cepacia* strain 5.5B *B. cepacia* NB-I *B. cepacia* 5.5B and RR 21-2 *B. cepacia* J2535, LMG 1222 and ATCC 51671 *B. cepacia* strain B23	(Arima et al. 1964) (Yoshihisa et al. 1989) (Burkhead et al. 1994) (Cartwright et al. 1995) (Roitman et al. 1990) (El-Banna and Winkelmann 1998) (Kadir et al. 2008; Sultan et al. 2008)
Toxoflavin, caryoynencin, macrolide lagriene and isothiocyanate sinapigladioside	*B. gladioli*	(Flórez et al. 2017)

*** This table is adapted from the thesis of Othman Boaisha (Cardiff, 2012).**

1.4 PhD aim and objectives

In light of the antibiotic resistance crisis and the lack of antibiotic discovery, there is an urgent need to find new sources of antibiotics. *Burkholderia* spp. are promising as novel antibiotic producers, as they are known to produce a range of antimicrobial metabolites and are rich in biosynthetic pathways encoding the production of these metabolites (Depoorter et al. 2016). Over the past few years, multiple environmental *Burkholderia* species have been characterised as producers of bioactive compounds (Mahenthiralingam et al. 2011; Kunakom and Eustáquio 2019; Mullins et al. 2019). However, the majority of these strains have been characterised in isolation for their antimicrobial properties. Only a few have come from systematic environmental screens which examine specific habitats, isolate *Burkholderia* or *Paraburkholderia,* and characterise them for the production of novel antibiotics.

Overall, this PhD project aimed to discover novel antibiotics produced by *Burkholderia* and *Paraburkholderia,* isolated specifically from environmental habitats. In addition, because it resulted in a unique collection of *Paraburkholderia*, which has not been widely associated with

antibiotic production (Webster et al. 2019), the PhD was able to characterise their potential for specialized metabolite production. The overarching hypothesis behind the study was "**Environmental *Burkholderia* and *Paraburkholderia* are a rich source of novel antibiotics**."

The following specific objectives were proposed to answer this hypothesis:

Objective 1. Develop cultivation-based and cultivation-independent methods to screen environmental samples for the presence of *Burkholderia* and *Paraburkholderia* bacteria (Chapter 3).

A former PhD student, George Payne, was successful in developing polymerase chain reaction (PCR) methods based on the *recA* gene to identify *Burkholderia* species (Payne et al. 2005), and these were successfully applied using cultivation-independent techniques to characterise the diversity of these bacteria in the maize rhizosphere (Payne et al. 2006). The first objective of the present study was to develop both growth-based and genetic tests capable of isolating and identifying *Burkholderia* and *Paraburkholderia.* This would expand on these initial single-gene based identification studies (Payne et al. 2005), and use the full complement of genomic resources now available for PCR product design to target these bacteria. Optimising and applying these approaches on well characterized *Burkholderia* and *Paraburkholderia* enabled the subsequent screening of the natural environment samples for these bacteria.

Objective 2. Apply cultivation and cultivation-independent methods to detect *Burkholderia* bacteria in the natural environment (Chapter 4).

This objective sought to assemble a collection of pure *Burkholderia* and *Paraburkholderia* isolates from the natural environment. Specifically, rhizosphere-associated soil samples and additional environmental samples from the jungle in Sabah, Borneo, Malaysia, were screened using PCR-based approaches. Subsequent growth-based analysis was used to purify *Burkholderia* and *Paraburkholderia* isolates from these samples.

Objective 3. Identify environmental *Burkholderia* and *Paraburkholderia* species using whole-genome sequencing (WGS) and bioinformatics tools (Chapter 5).

The collection of novel *Burkholderia* and *Paraburkholderia* isolates assembled was examined using whole genome sequence analysis and a range of molecular methods to achieve accurate identification. Phylogenetic trees based on single genes such as the 16S rRNA and *recA* gene were used to identify the isolates to the genus level and for selecting the closest taxonomic type strain (TS) neighbours. To increase taxonomic resolution and accurately define species, whole-genome comparisons were performed using methods such as genome sequence-based DNA-DNA hybridization (DDH), average nucleotide identity (ANI), and core gene-based phylogenetic trees.

Objective 4. Screen the *Burkholderia* and *Paraburkholderia* collections for the presence of novel antibiotics using bioactivity and genome-mining analysis (Chapter 6).

The assembled collection of environmental *Burkholderia* and *Paraburkholderia* were screened for the production of novel antimicrobials using conventional bioactivity screening approaches (Mahenthiralingam et al. 2011), as well as genome mining (Rutledge and Challis 2015). This combined genomic and biochemical approach was used to identify specific strains of interest, that produced novel antimicrobials or encoded novel antibiotic biosynthesis pathways worthy of further investigation**.**

Chapter 2. Materials and methods

2.1 Growth cultures and chemicals

Several different media were used in this study as follows: tryptone soya agar/broth (TSA/TSB), *Burkholderia cepacia* selective agar (BCSA) (Henry et al. 1997), basal salts medium (BSM) (Hareland et al. 1975; Mahenthiralingam et al. 2011) and *Pseudomonas cepacia* azelaic acid tryptamine (PCAT) (Burbage and Sasser 1982; Peeters et al. 2016). They were either purchased from suppliers or made up specifically as follows.

TSA/TSB are complex rich media and were used for general culture of *Burkholderia* and *Paraburkholderia* reference strains as well as other bacterial species used in this study. BCSA is a Bcc-selective medium containing the following antibiotics: Polymixin B (75,000 units/L), Gentamicin (2.5 mg/L), and Ticarcillin (50.0 mg/L) (Henry et al. 1997). All these growth media were supplied by Oxoid Ltd. (Hampshire, UK) and prepared according to the manufacturers' instructions.

BSM is a minimal growth medium developed by Hareland et al. (1975) and more recently used by Mahenthiralingam et al. (2011) to promote antibiotic production in *Burkholderia* when supplemented with glycerol (4 g/L) as a carbon source (BSM-G). The components of BSM-G were as follows (per liter): Glycerol 4 g; 50 ml phosphate Salts 20x stock ($K_2HPO_4.3H_2O$, 4.25g and $NaH_2PO_4.H_2O$, 1.0 g); 50 ml ammonium chloride 20x Stock (NH_4Cl, 2.0 g); 10 ml nitrilotriacetic acid 100x stock ($C_6H_9NO_6$, 0.1 g); 10 ml metal salts 100x stock ($MgSO_4.7H_2O$, 0.2g; $FeSO_4.7H_2O$, 0.012g; $MnSO_4.H_2O$, 0.003g; $ZnSO_4.7H_2O$, 0.003g; $CoSO_4.7H_2O$, 0.001g); 0.5 g CAS Amino Acids; and 0.5 g yeast extract.

A bromoacetic acid enrichment medium (BSM-GBP) was developed from BSM-G by adding a 1M Bromoacetic acid solution to achieve final test concentrations ranging from 50 mM to 0.1 mM; Polymixin B was added to this medium (final concentration 600 U/ml) as a known selective agent for *Burkholderia* and BSM-GBP adjusted to pH 7±2. PCAT is an enrichment medium used to isolate environmental *Burkholderia* and *Paraburkholderia* (Peeters et al.,

2016). The components of PCAT were as follows (per liter): 2 g azelaic acid; 10 ml magnesium sulfate heptahydrate 2% stock ($MgSO_4.7H_2O$, 2 g); 100 ml potassium hydrogen phosphate 0.4% stock (K_2HPO_4, 4 g); 100 ml potassium phosphate monobasic 0.4% stock (KH_2PO_4, 4 g); 0.5 g Yeast Extract; 200 mg Tryptamin. Cycloheximide was added to this medium (final concentration 200 mg/L) and the pH adjusted to 5.7.

All the solid media contained 15 g purified agar per L (1.5% w/v agar). All media were prepared with double-deionised water and sterilised by autoclaving at 121°C for 15 minutes before mixing with other supplements such as antibiotics. Bromoacetic acid, Polymixin B and Cycloheximide stock solutions were sterilised by filtration (0.2 µm filters; Sartorius Stedim Biotech, UK) and stored at -20°C; they were all added to media after autoclaving and cooling to 50 °C. The chemicals and commercial growth media used in this study were supplied by Sigma-Aldrich (Dorset, UK) and Fisher Scientific (Loughborough, UK).

2.2 Bacterial strains and routine growth and storage conditions

2.2.1 Strains panels and bacterial samples

The bacterial strains used on this study were a selection panel from BCC (*Burkholderia* Cardiff Collection), ESH (E. Mahenthiralingam's wider strain collection) and European Coordination Action for Research in Cystic Fibrosis (EURO-CARE CF) collection (EC FP6 project no. LSHM-CT-2005-018932) that were all held at Cardiff University. The environmental isolates for the present study were also obtained (see below) from the deep frozen stocks of environmental material collected by Eshwar Mahenthiralingam from Sabah, Malaysia, in 2008 (Figure 2.1). Sampling was carried out with permission from the Sabah Wildlife Department, Sabah, Malaysia, and followed the protocol as described below (see section 2.3).

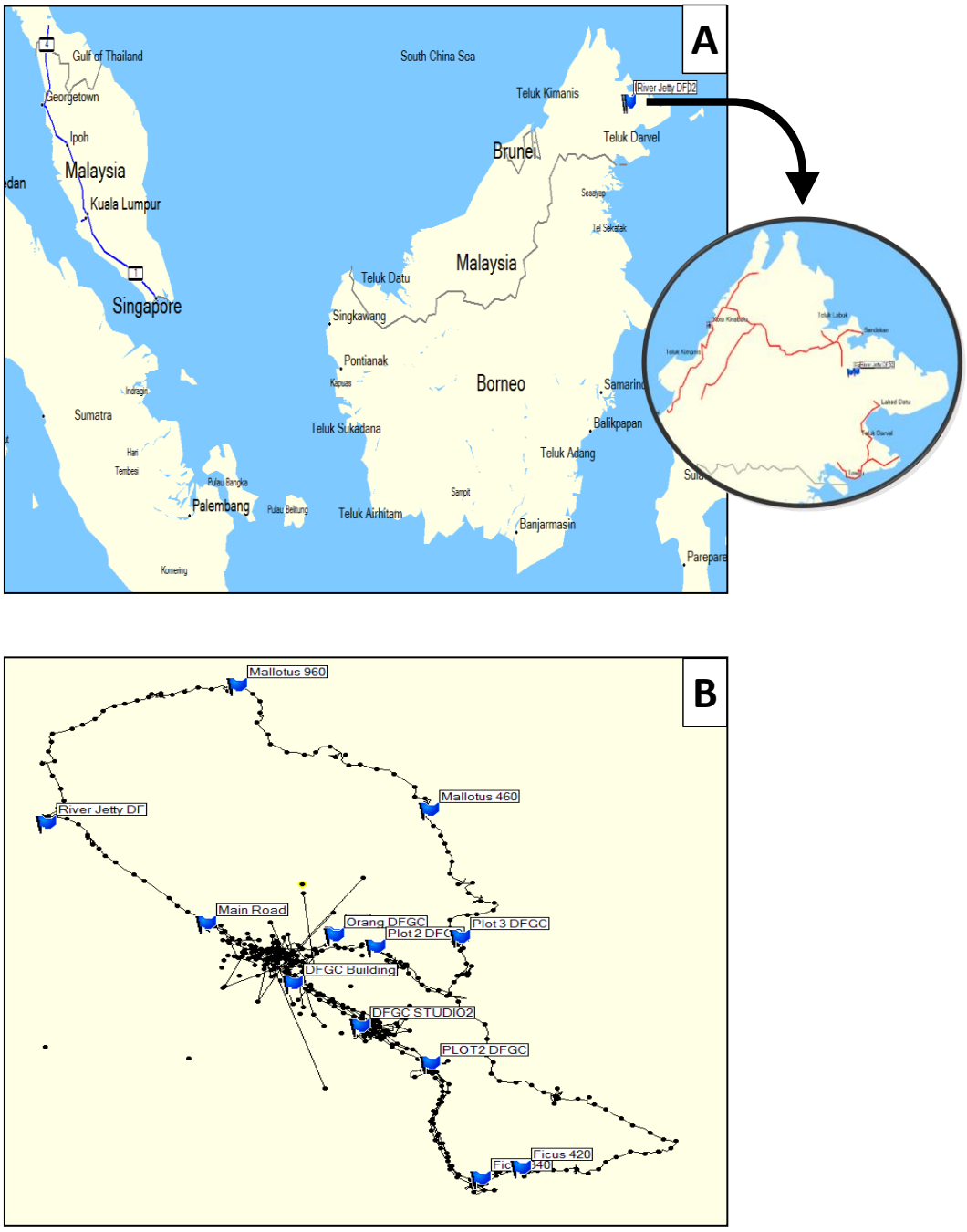


Figure 2.1 Location of rhizosphere and jungle sampling in Sabah, Malaysia. The location of the Danau Girang Field Centre, in Sabah Malaysia, in relation to the peninsula of Malaysia, is shown in panel A and circled in red. The field centre lies on the Kinabatangan River, just south of Sandakan, on the north-east coast of Sabah (panel B). The GPS recording of all the trails that were sampled around the field centre is shown in panel C, and the trail names and locations correspond to each of the sample points examined. Figure and information provided by E. Mahenthiralingam, Cardiff University.

2.2.2 Growth and storage of bacterial isolates

Bacterial isolates were grown on solid or liquid media (TSA/TSB for most general revival and growth controls) and BSM-G, BSM-GBP or PCAT dependent on the screen being performed. Incubation time ranged from overnight (18 to 24 h) to 1 week. Environmental strains and samples were incubated at 30°C, and other bacteria tolerant of higher temperatures grown at 30°C. Overnight cultures were prepared by inoculating 5 ml of TSB or other required medium within a 15 mL sterile tube and shaken (rocked gently on a platform) for between 48 h to 336 h. The storage of bacteria isolates and bacterial samples were at -80°C in TSB, BSM-G or PCAT containing 8% Dimethylsulfoxide (DMSO) Fisher Scientific (Loughborough, UK). The results were recorded for the bacterial growth up to one week.

2.2.3 Bacterial growth in 96-bioscreen plate and data analysis protocol(Chapter 3)

The Bioscreen C instrument (Labsystems, Finland) was used as an automated method to generate bacterial growth curves. It was used to test the impact of bromoacetic acid concentrations on the rate of bacterial growth. The general Bioscreen C set up and screening protocol is illustrated in Figure 2.2 and each strain was tested in quadruplicate for each run. The instrument settings were as follows: the bacterial growth monitored for 48 hours at 30°C; turbidity measurements were taken at 15 minute intervals using a wide band filter (450-580 nm), with the microplates shaken for 10 seconds at an intermediate intensity prior to each reading.

The data generated by the BioScreen C was exported into Microsoft Excel sheet and manipulated as follows. The optical density (OD) for each sample and the mean value of control wells (broth without cells) were pooled onto Microsoft Excel sheet for the growth curves analysis. The length of lag phase, the maximum growth rate and culture density reached were the 2 growth parameters estimated from the logarithmically transformed data by using the

gcFit function of the grofit statistical package, which uses R software for analysis (R-Core-Team 2013) (Kahm et al., 2010). The protocols and macros to convert BioScreen C data and to create the growth curves using R were described by Rebecca Weiser (PhD thesis, Cardiff University, 2014). Figure 2.3 presents the preparation of Bromoacetic acid, as enrichment agent for *Burkholderia* bacteria, with BSM-G media. This protocol conducted to study the impact of Bromoacetic acid on *Burkholderia* growth, and the growth parameters evaluated using BioScreen C.

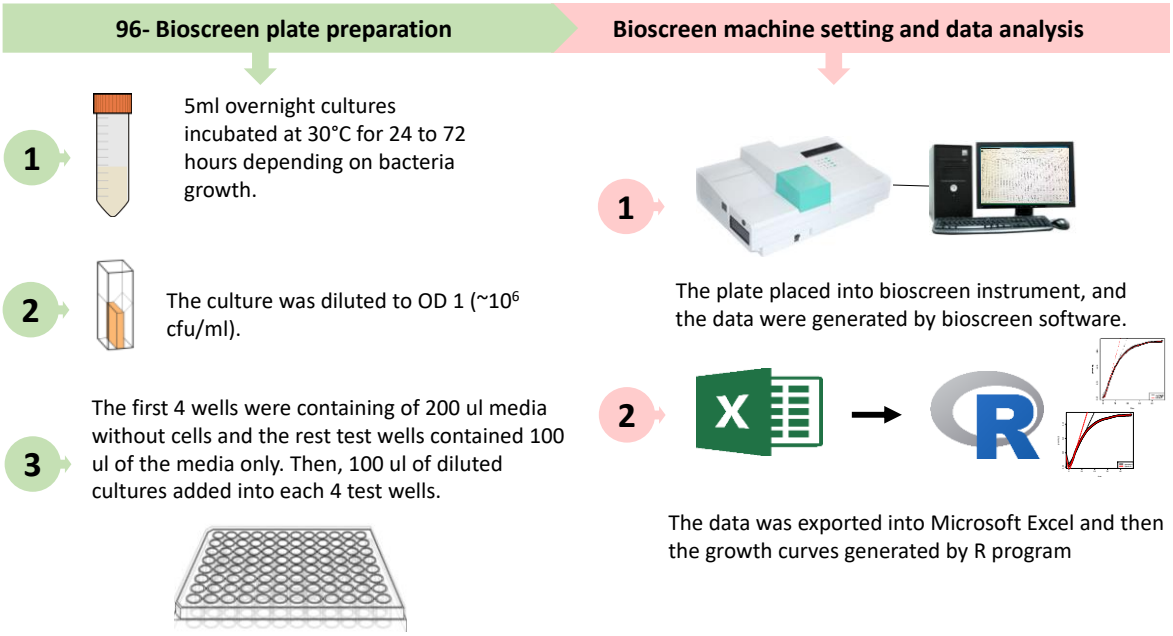


Figure 2.2 The Bioscreen C protocol used for bacterial growth rate determination. The protocol for growth analysis is shown with the first step preparing the 96-Bioscreen microplate and inoculating this with standardised dilutions of overnight bacterial cultures. The second step was to run the Bioscreen C growth instrument which automatically read the optical density generated throughout microbial growth, and after exporting the data to Excel, statistical software was used to generate growth curves.

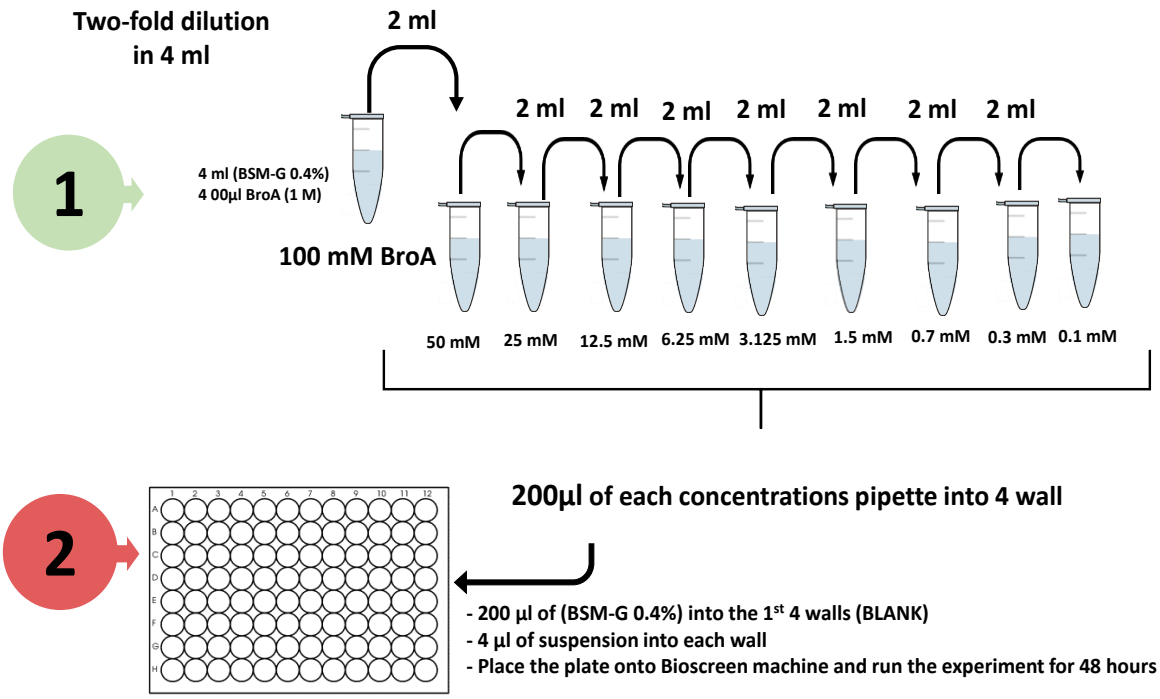


Figure 2.3 Evaluation of bromoacetic acid on bacteria growth using Bioscreen C screening. (1) 400 µl 1 M Bromoacetic acid added to 4 ml BSM-G (0.4%). Then the two-fold dilution had been done and the obtained bromoaceticacid concentrations were from 100 mM to 0.1 mM. (2) 200 µl of each diluted media loaded into bioscreen plate.

2.2.4 Rapid screening method for determining bacterial growth on enrichment media (Chapter 3)

Using 96-well growth plates and a replica plating tool, multiple bacterial strains were screened simultaneously for growth the different media tested in this study. This was a time-saving method used to transfer multiple liquid cultures simultaneously onto agar growth media, and since the master plate cultures could be stored deep frozen, it also enabled replicate testing. The following procedure was performed to create a 96-microplate containing bacterial cultures: 4 µl of a fresh overnight bacterial culture was inoculated into 200 µl of growth media in the 96-well plate (TSB or other test media such as BSM-G, BSM-GBP or PCAT) and the plate incubated at 30°C for 24 to 72 h depending on the bacterial growth rate (Figure 4). After that, the test cultures were transferred to a square agar plate using a microplate replicator device. The plate inoculated at 30°C for 24 h to 1 week. The bacterial visual growth was

observed daily. The result for each strain was recorded as (+) for a visual growth and (–) for no growth (See Tables in appendices).

2.3 Sampling in Sabah and storage

The following information about the environmental sampling procedure and processing in Sabah, Malaysia was provided by E. Mahenthiralingam, personal communication. A total of 98 jungle plant-associated and environmental samples were collected around the Danau Girang field centre, near Kota Kinabatangan, Sabah, Malaysia in 2008 (Figure 1). For rhizosphere sampling, a metal spatula was used to dig up small plants on the jungle floor, and at least 2 cm cut from a small section of their roots. The samples were placed into sealable plastic bags, and each labelled with a site number. The information about each sample location, and photographs for each plant or sample site were also taken. To maintain sterility as far as possible during sampling, the metal utensils used such as spatulas and scissors, were wiped down with ethanol in between each sample. Rhizosphere samples were processed as follows: small roots, 1 to 3 mm in diameter were cut to a length of 1 cm and each placed into a sterile microtube containing 1 ml of basal salts medium (Hareland et al. 1975); this was supplemented with 4 g/L of glycerol as a carbon source and 100 mg/µL cycloheximide to reduce fungal growth (the medium was designated BSM-GC). The sample was then homogenised using a disposable plastic plunger and incubated at the ambient temperature in Sabah (approximately 30°C) for up to 2 days. The growth enriched samples were then shipped to Cardiff University, Cardiff, Wales, UK (approximately 10 days), and 80 µl of DMSO was added to each sample as a cryopreservant, and the samples stored frozen at -80°C.

2.3.1 Reviving samples and initial screening to isolate single colonies using spiral platter technology (Chapter 4)

The deep frozen environmental samples were thawed for 30 mins and 100 µl of collected jungle sample suspension inoculated in 3 ml of PCAT broth. The cultures were incubated at 30°C for 72 hrs (with gentle shaking) to allow the microbial community to grow. A series of 10-fold dilutions were made to the cultures for six 10 fold dilutions (until 10^{-6}). 10^{-4} and 10^{-6} diluted cultures were spread on PCAT plates to enable growth of single colonies using a spiral platter instrument (Don Whitley Scientific). PCAT plates were incubated for up to 72 hrs. Depending on the differences observed in colony morphology, colony sizes and morphotypes were selected as indicative of a different bacteria strain and species type being present. The single colonies were inoculated into fresh tubes containing 1.5 ml of PCAT broth with 8% DMSO, and then stored at -80°C prior to further analysis. Subsequent purification of isolates was carried out by re-streaking to single colonies, and passaging them on TSB growth media at least twice before a final pure culture freezer stock was made.

2.4 Identification of environmental bacteria by molecular methods and bioinformatics analysis (Chapter 4 and 5)

2.4.1 DNA extraction from environmental enrichment and pure bacterial cultures

Overnight cultures of *Burkholderia*, *Paraburkholderia* and other bacteria used in this study were grown in 5 ml TSB and the cultures were incubated at 30°C with shaking for 24 hours. The overnight bacterial cultures were harvested by centrifugation at 2054 g for 10 min. The DNA was extracted from the cultures by 2 methods in this study: an automated Maxwell® 16 system (Promega, USA) and Chelex®100 resin (Biorad, UK) protocol. The same methods were also applied to the environmental enrichments cultures to test these for the presence of *Burkholderia*.

The Maxwell® instrument (Promega, USA) is an automated system to extract up to 16 DNA samples from multiple different sample types and cells in a time of 45 min. The extraction protocol was applied according to manufacturer's instructions with the following modifications. Bacterial pellets from harvested cultures were re-suspended in 500 ul TSB and then 500 µl of 4M guanidine isothiocyanate solution (Invitrogen, USA) was added. The suspension was then transferred to a 2 ml screw-cap micro-centrifuge tubes containing approximately 0.5 g of 0.1 mm zirconium beads (Stratech Scientific Limited, Ely). The tubes were placed onto a bead-beater machine to help to disrupt the bacterial cells. The tubes then centrifuged for 2 min and 300 µl of the lysate was transferred into cartridge, placed into the Maxwell robot and the extraction protocol set to extract the DNA > tissue culture. After extraction, the eluted DNA was placed into a sterile microtube and stored at -20°C.

Chelex®100 resin was used as a rapid DNA extraction method to yield PCR amplifiable DNA. Single bacterial colonies were picked using a sterilised tip and transferred into 50 µl of 5% Chelex® 100 resin solution (Bio-rad, Hertfordshire, UK; the solution was made with deionised water and autoclaved before use). The bacterial suspensions were heated to 95°C in Thermal Cycler (Bio-Rad) for 5 minutes, then placed at -80°C for 5 minutes. This process was repeated twice before centrifuging the sample for 1 minute to sediment the Chelex® 100 resin and cellular debris. The DNA from the supernatant was used directly for PCR and not stored for further analysis.

2.4.2 The quantitation and quality assessment of DNA

Gel electrophoresis and Qubit Fluorometer were the two methods used to measure DNA quantitation and quality. The gel electrophoresis was performed as follows: 1.5 % (w/v) agarose (molecular grade agarose; Severn Biotech Ltd.) was heated in 100 ml of 1X TAE (Tris-acetate-EDTA; Severn Biotech Ltd.) within a microwave to achieve dissolution, 10 µl SafeView stain (NBS Biologicals Ltd., Cambridgeshire, UK) added to the agarose after

melting, the gel was cooled to 50°C, poured into a gel tank mold with a comb for the sample wells, and allowed to set. Gels were immersed into TAE buffer within a gel tank and samples loaded with loading dye. The samples then run at 80-120 V for approximately 1.5 hours. The gel then was visualised with VersaDocTM Imaging Systems (Bio-Rad, UK) using quantity one software. The size, quantity and quality of the DNA was estimated from the gel images by eye.

Accurate DNA quantification was performed with the Quantifluor® dsDNA system (Promega, Wisconsin, USA) kit and the Qubit™ fluorometer (Invitrogen, USA) according to following the manufacturer's instructions. If the dilutions of DNA were needed, these were made in sterile nuclease free water (Severn Biotech Ltd.).

2.4.3 Primer design and synthesis

Amplification of 16S rRNA gene, *recA* gene as well as Random Amplified Polymorphic (RAPD-PCR) were performed using published primers shown in (Table 2.1). Oligonucleotides were synthesized by Eurofins Ltd (Wolverhampton, UK).

Table 2.1 16S rRNA and *recA* genes primers and RAPD-PCR primer

Target gene	Primer pair	Primer sequence	Target	Reference
16S rRNA	27F 1492R	**5'-**AGAGTTTGATCCTGGCTCAG-'**3** **5'**-GGTTACCTTGTTACGACTT-'**3**	Bacteria (Universal)	(Lane et al.,1985)
recA	BUR1 BUR2	**5'-**GATCGA(AG)AAGCAGTTCGGCAA-'**3** **5'-**TTGTCCTTGCCCTG(AG)CCGAT-'**3**	*Bukholderia*	(Payne et al. 2005)
RAPD	270	**5'-**TGCGCGCGGG-'**3**	*Burkholderia cepacia*	(Mahenthiralingam al., 1996)

2.4.4 Marker gene amplification using polymerase chain reaction (PCR)

Polymerase chain reaction (PCR) sample constituents (Table 2.2 and Table 2.3) and thermal cycle settings (Table 2.4) are shown. PCR reactions were set up with a total volume of 50 µl. Two PCR kits were used in this study. Initially, reagents from Qiagen (Limburg, Netherlands)

were used comprising: a *Taq* PCR core kit (1000 units) containing 1X Coralload buffer, 1X Q-solution, 200 µm deoxynucleoside triphosphates (dNTPs), 1000 U *Taq* polymerase. Reactions were set up as shown (Table 2.2) and contained 2 µl DNA template (approximately 20 ng was added for most PCRs). DreamTaq Green PCR master mix (2X) (Thermo Scientific, UK) was evaluated as the second PCR system, and this contained nuclease-free water and DreamTaq green PCR master mix (2X) (Dream*Taq* DNA Polymerase, 2X dreamTaq green buffer, dNTPs, and 4 mM $MgCl_2$). PCR was set up as shown (Table 2.3) and contained 5 µl DNA or bacterial culture (1 to 3 ul of an overnight culture). The PCR cycles were performed on a c1000™ Thermal Cycler (Bio-Rad), and PCR programmes are presented in Table 2.4 and 2.5. PCR products were analysed by gel electrophoresis as described above. PCR products were purified (see section 2.4.5 below) before submission to Eurofins Eurofins Ltd (Wolverhampton, UK) for Sanger sequencing. RAPD-PCR was used as a basic strain fingerprinting analysis to remove identical isolates recovered from the same individual jungle enrichment. The RAPD fingerprints were visualised by gel electerophoresis.

Table 2.2 The Qiagen PCR reaction components

Per 50 µl PCR reaction:	
1:1 primer solution (0.4 pmol/µl final concentration of each primer)	2 µl
dNTPs 10 mM each dNTP (final concentration 200 µM each)	1 µl
10X Buffer (contains Mg at 15 mM)	5 µl
MgCl2	2 µl
Taq (1 unit per reaction)	0.25 µl
Sterile polished water	36.25 µl
Pured DNA or cultures	5 µl

Table 2.3 The DreamTaq Green PCR master mix (2X) PCR components

Per 50 µl PCR reaction:	**All PCR reaction**	**RAPD-PCR**
1:1 primer solution (0.4 pmol/µl final concentration of each primer)	2 µl	8 µl
DreamTaq green PCR master mix (2X)	25µl	25µl
Nuclease-free water	18µl	13µl
DNA	5µl	4µl

Table 2.4 PCR programme for 16s rDNA gene and *recA* gene amplification

Steps		Temp (°C)		Time	No. of cycles
		16S rDNA	*recA*		
1.	**Initial denaturation**	95	95	5min	1
2.	**Denaturation**	94	94	1 min	35
3.	**Annealing**	52	58	1 min	
4.	**Extension**	72	72	1 min 30 sec	
5.	**Final Extension**	72	72	5 min	1

Table 2.5 PCR programme for RAPD-PCR

PCR steps	Tempreture	Duration	Cycles
1	94**°C**	5 min	-
2	36**°C**	5 min	4 cycles
3	72**°C**	5 min	
4	94**°C**	5 min	
5	94**°C**	1 min	30 cycles
6	36**°C**	1 min	
7	72**°C**	2 min	
8	72**°C**	10 min	-
9	10**°C**	Forever	-

2.4.5 PCR purification for 16S rRNA and *recA* gene phylogenies

PCR products were purified using Monarch® PCR & DNA Cleanup Kit (5 µg) (NEB, Ipswich Massachusetts) containing Monarch DNA blean up binding buffer, wash buffer, elution buffer and cleanup columns (5 µg). The purification process was performed according to the manufacturer's instructions, and the 20 µl pure DNA obtained submitted for Sanger sequencing (Eurofins, UK) (see chapter 4 and 5).

2.4.6 16S rRNA and *recA* genes sequencing and bacterial species identification using phylogenetic analysis

The purified 16S rRNA and *recA* amplified genes were forwarded to the MWG Eurofins DNA Sanger for sequencing following the sample submission guidelines 'MWG Eurofins Value Read Service in Tubes'. The partial 16S rRNA and *recA* genes sequences data was received via an email (see Chapter 4). Later, the complete length of 16S rRNA and *recA* genes sequences were extracted using command line Barrnap (https://github.com/tseemann/barrnap), and Artemis (Rutherford et al. 2000; Lagesen et al.

2007) (see Chapter 5). The 16S rRNA and *recA* gene sequences were analysed and phylogenetic trees constructed as follows.

The National Center for Biotechnology Information (NCBI) was used for database searching (Sayers et al. 2010) using the Basic Local Alignment Search Tool (BLAST) algorithm (Altschul et al. 1990) to rapidly infer species identities from the 16S rRNA and *recA* consensus gene sequences. As a more accurate method of sequence-based identification, sequence alignment to reference species gene sequences and phylogenetic tree construction was conducted by molecular evolutionary genetics analysis software virsion 7.0.26 (MEGA 7) (https://www.megasoftware.net/) (Tamura et al. 2013; Kumar et al. 2016). The gene sequences were imported into MEGA 7, trimmed, and aligned using ClustalW2 (Larkin et al. 2007). The *recA* tree was drawn to scale and evolutionary distances computed using the neighbour-joining method with 1000 bootstrap replicates within MEGA 7 software (see Chapter 4 and 5).

2.4.7 Whole genome sequencing and bioinformatics analysis (Chapter 5)

3 ml of the overnight culture used for the DNA extraction was prepared from the freezer stock pure isolates. Genomic DNA from 62 purified jungle isolates were extracted using the automated Maxwell® 16 system. The quantitation and quality assessment and the draft genomes (WGS) were carried out in Genomics Research Hub (Cardiff School of Biosciences, Cardiff university). The library preparation was carried out using NEB Ultra ll DNA kit with a Fragmentise step used on the genomic DNA before going in to the preparation. The sequencing was carried out on the Illumina Nextseq platform using a (2 x 150) Mid output cartridge with 130 Million paired-end reads. The output of each genome draft comes in FastQ files. All subsequent bioinformatics analysis was carried out using a virtual machine, hosted by the cloud infrastructure for microbial bioinformatics (CLIMB) consortium (Connor et al. 2016). FastQC command line had been run to provide a quality control check on the WGS

data. Unwanted sequences from sequencing process were removed Trim-Galore (https://github.com/FelixKrueger/TrimGalore), and the paired-reads were assembled using unicycler (Wick et al. 2017a). For bacterial genomes assembly, Unicycler pipeline was used. Prokka was used to annotate the whole genomes sequences (Wick et al. 2017b); this tool can identify and label the features of DNA sequences, such as genes locations and the coding region. (Seemann 2014). Finally, the draft genomes were subjected further genomics and bioinformatics identification and antiSMASH analysis.

The identification of novel isolates was based on using whole genome output as follows. The quality of draft genomes was evaluated quality assessment tool for genome assemblies (QUAST) (Gurevich et al. 2013). Kraken was used as a rapid taxonomic tool to identify the bacterial genomes (https://ccb.jhu.edu/software/kraken/MANUAL.html). Multi-Locus Species Tree (autoMLST) was used as an alternative and rapid MLST method from the manual MLSA to identify the unknown bacterial isolates. It was developed to build MLST phylogeny for uploaded genomes with its huge updated genomics database. It is useful tool for high-throughput species and genus classification (Alanjary et al. 2019). Average Nucleotide Identity (ANI) and in silico DNA-DNA hybridization (DDH) were used to delineate species (https://ggdc.dsmz.de/ggdc_background.php#) (Konstantinidis and Tiedje 2005; Meier-Kolthoff et al. 2013). Two different tools were used to assess the ANI value. First, Pyani (https://github.com/widdowquinn/pyani), was useful to assess the ANI value between this study isolates with the closest type strains references based on 16S rRNA and *recA* genes phylogenies. Second, FastANI (https://github.com/ParBLiSS/FastANI) was performed by Alex Mullins (Cardiff University), to compare the genomes from this study with those assembled from the genetic database (a collection of 4000 genomes is at the time of analysis). Roary is a rapid bioinformatic tool to identify the core and accessory genes within collections of genomes. Roary was used to identify the core genes conserved within genome of different *Burkholderia* and *Paraburkholderia* species, which were aligned and then built a phylogenetic tree to evaluate the relationship between the bacterial isolates (Page et al. 2015). The core

gene based phylogenies were constructed by fasttree based on maximum-likelihood phylogenetic trees (ML) method (Price et al. 2010).

2.5 Antimicrobial production analysis (Chapter 6)

2.5.1 Screening of antimicrobial activity using antagonism assays for Jungle isolates

To test the ability of antibiotic production from the environmental (jungle) *Burkholderia* and *Paraburkholderia* isolates, a conventional bioactivity screen was carried out (Mahenthiralingam et al. 2011). A 2 µl drop of each *Burkholderia* and *Paraburkholderia* isolates' frozen inoculum culture was spotted onto the centre of BSM-G supplied with glycerol as a carbon source. The pH was varied between 5 and 7, as well as growth at two different temperatures 22°C and 30°C tested. Incubation time was 72 h unless otherwise stated. After growth, chloroform vapour was used to kill the bacterial growth for 10 min. The antagonism assay was used as a rapid technique to detect which *Burkholderia and Paraburkholderia* isolates had produced antimicrobial compounds. The preparation of the overlay media was carried out as follows. The molten iso-sensitest agar was cooled to 45^{o}C after autoclaving. The agar was then aseptically mixed with the susceptibility testing organisms. *Candida albicans,* Gram-negative *Pectobacterium carotovorum* and Gram-positive Methicillin-resistant *Staphylococcus aureus* (MRSA) were selected as the organisms to be tested for susceptibility to the *Burkholderia* antimicrobial metabolites. These organisms were chosen both for their range of multi-drug resistant properties as well as being AMR priority species. Fifteen ml of the overlay culture was poured onto the surface of each isolated colony, and then the soft agar allowed to set at room temperature for 30-40 min. The plates were incubated at 30°C for 18 h. The antimicrobial activity was revealed by the presence of a zone clearing, where the susceptibility testing microorganism failed to grow. The diameter of the zone of inhibition was

measured in mm. All antagonisms assays were completed in triplicate to allow estimation of the mean and standard error for the zone of inhibition observed.

The collection of 58 jungle isolates (thirteen *Burkholderia*, forty-four *Paraburkholderia* and one *Caballeronia*) was tested alongside the following control *Burkholderia* antibiotic producers: (1) *B. ambifaria* AMMD^T (BCC0207), an enacyloxin producer (Mahenthiralingam et al. 2011) with activity against methicillin-resistant *Staphylococcus aureus* (MRSA), *Pectobacterium carotovorum* and *Candida alb icans*; (2) *B. gladioli* (BCC0238), a gladiolin producer (Song et al. 2017) with the same broad spectrum anti-Gram positive, anti-Gram negative and antifungal activity as *B. ambifaria* BCC0207; and (3) *B. ambifaria* (BCC0191), a cepacin producer with activity against Gram-positive bacteria (MRSA) and fungi (*C. albicans)* (Payne et al. 2006; Mullins et al. 2019). An additional jungle *B. vietnamiensis* isolate J17-3, identified from *recA* gene analysis (Chapter 4), but for which no genome sequence was available (Chapter 5), was included in the screening.

2.5.2 Metabolite analysis by high performance liquid chromatography (HPLC)

High performance liquid chromatography (HPLC) is a highly automated and extremely sensitive method. It is a chemical analysis method that has an ability to predict and quantify components dissolved in a liquid solvent. The HPLC instrument contains the following parts: a solvent reservoir, high-pressure pump, injector system, column, detector and waste. The principle of HPLC is that the sample mixture is moved by pressurized liquid (a mixture of solvents or water), and this allows the compounds to be separated and then analysed in two different main phases (mobile phase and stationary phase). Small amounts of samples can be introduced into the system where the solvent is holed by the reservoir (mobile phase). The high-pressure pump produces the pressure to move the solvent to the injector (to be ready for the stationary phase). The stationary phase where is the samples were injected to HPLC column and allowing the components in mixture to be separated and analysed. The time to

allow the separated compounds to travel through column to the detector is called retention time, and each compound has its own time. Finally, the detector observes each compound in peaks showed the quantities of the compounds and display the final analytical graphs on the computer. To start this automated HPLC flow, the bacterial metabolites preparation was conducted by the following procedure.

The thirteen bioactive environmental *Burkholderia* isolates were subjected to metabolite extraction from their spent growth media and the extracts examined using HPLC as described (Mullins et al. 2019). Characterising a number of known *Burkholderia* metabolite producers using HPLC (Song et al. 2017; Mullins et al. 2019) enabled the identification of known antimicrobial compounds, such as pyrrolnitrin and cepacin, as well as novel unknown compounds, which were worthy of further analysis.

To analyse metabolite extracts from the cultured bacteria, TSA and TSB were used as starter media to allow the bacteria to grow at 30°C. Overnight cultures of *Burkholderia Paraburkholderia,* and control antibiotic producing species were streaked in BSM-G (pH 5 and 7). Using sterilized cotton swabs the cultures were streaked with 10 parallel horizontal lines in BSM-G plates. The plates were incubated at 22°C and 30°C for 72 h to allow the bacteria to produce the metabolites. The bacterial growth was removed from the agar surface using a sterile microbiological cell scraper inside a class 2 safety cabinet. A 2 cm disc of agar was removed from the centre of each plate and placed into small dark glass bottle. 500 µl of dichloromethane was added to the agar disc, and then the bottles wereshaken for 2 h at room temprature (20 C) to allow metabolite extraction to take place. After clearing by centrifugation, this crude metabolite extract was fractionated on an acetonitrile gradient (5% to 95%) using a Waters® AutoPurification™ HPLC System fitted with a reverse phase analytical column (Waters® XSelect CSH C18, 4.6 x 100 mm, 5 µm). Then, metabolites eluted from the column were detected by a photodiode array and the peaks corresponding to cepacin, pyrrolnitrin, and other unknown compounds identified by their retention time (Mullins et al. 2019). Three replicate extract samples were prepared and run for each isolate examined.

2.5.3 Genome mining approach by means of antiSMASH 0.0.5 and graph pad data presentation

AntiSMASH stands for antibiotics and Secondary Metabolite Analysis Shell (Blin et al. 2013; Blin et al. 2019), and it is a rapid website tool that allows the whole genome sequence data to be run in FASTA format files, and predicate and analyse the secondary metabolite biosynthetic gene clusters within minutes. It was used in the present study to rapidly profile the BGC coding capacity and predict putative and known pathways for the antimicrobials within the *Burkholderia* and *Paraburkholderia* genomes. in Chapter 6. Multiple classes of secondary metabolite biosynthesis gene clusters can be predicted by antiSMASH, such as non-ribosomal peptides (NRPS), terpenes, polyketide synthese (PKS), butyrolactones, bacteriocins, nucleosides, beta-lactams, siderophores and others. It predicts the gene clusters and then aligns them to identify the cluster type from a database. The antiSMASH output provided information on clusters type and length (Kb), NRPS/PKS domain analysis, the chemical structure of predicted metabolites encoded by the cluster, and the cluster blast for the relative gene clusters. The total of 97 genomes were run through the web server (http://antismash.secondarymetabolites.org/) in this study.

A total collection of 97 *Burkholderia* and *Paraburkholderia* genomes were analysed: twelve environmental *Burkholderia* genomes from this study (the genome of isolate J17-3 was not determined), 20 reference genomes for well-classified *Burkholderia* spp., 44 jungle *Paraburkholderia* genomes, one *Caballeronia* genome, and 20 reference *Paraburkholderia* genomes. The reference genomes were downloaded from NCBI and were annotated in parallel with the de novo genome sequence data obtained in this study. The genome mining using antiSMASH identified the following features of the *Burkholderia* and *Paraburkholderia* genomes

2.6 Statistical analysis

Each experiment was performed as three biological replicates and frequently multiple technical replicates were also included where appropriate, as indicated inappropriate result chapters. For the statical analysis in Chapter 6, the length of BGC, the total of BGCs in *Burkholderia* and *Paraburkholderia* genomes, and other numerical metrics, such as genomes sizes, were collected within a Microsoft Excel spreadsheet. Basic statistical parameters such as the mean, median and ± standard error (SD) of the mean were collected and analysed by Microsoft Excel (see Tables 6.2 and 6.3). Analysis of the average and range of genome sizes, the range of predicted BGCs and the average predicted BGCs per Mb for the 96 de novo and reference genomes was performed using a nonparametric unpaired t-test (GraphPad Prisim version 8.2.1; in Chapter 6). The t-test was used to compare the means of two groups including: the total of BGCs in *Burkholderia* versus *Paraburkholderia* (see section 6.2.3.3), references *Burkholderia* versus jungle *Burkholderia* (see section 6.2.3.1), and references *Paraburkholderia* versus jungle *Paraburkholderia* (see section 6.2.3.2). The statistical significance differences between samples were expressed as the exact *P* value with $P < 0.05$ taken as significant.

Chapter 3. Enrichment media to identify environmental *Burkholderia* and *Paraburkholderia*

3.1 Introduction

Up to 99% of bacteria within soil and related environments cannot be grown easily in the laboratory (Amann et al. 1995). In relation to the primary aim of the current research, it was important to develop growth media for *Burkholderia* to help isolate and enrich them from the complexity of organisms present in natural environments. Surveillance using enrichment and culture-based approaches will increase knowledge of the habitats and diversity of *Burkholderia,* and also isolate novel *Burkholderia* strains that have the potential to produce natural bioactive compounds.

The majority of studies where *Burkholderia* bacteria have been isolated, have focused on the pathogenic species found in human and animal infections (Table 3.1). Ashdown's medium is a selective medium specific to the clinically important species *Burkholderia pseudomallei*, and it is derivative from tryptone agar TSA supplied by gentamicin (Ashdown 1979). *Burkholderia cepacia* selective agar (BCSA) and *Burkholderia cepacia* complex enrichment medium (BCEM) are selective media to enrich *Burkholderia cepacia* complex species from cystic fibrosis (Henry et al. 1997; Flanagan and Paull 1998). In the absence of patient-to-patient or nosocomial spread, these pathogenic *Burkholderia* infections are generally acquired from the natural environment, but there is a lack of research into the direct isolation of environmental *Burkholderia* (Peeters et al. 2016). A range of different growth media for the enrichment of *Burkholderia* from different sample types have been developed (Table 3.1). *Pseudomonas cepacia* azelaic acid tryptamine (PCAT) is one of the most effective media for isolation of *Burkholderia* from soil and water samples (Peeters et al. 2016).

Table 3.1 Growth media for *Burkholderia* bacteria

Media	Media type	Ref
Basal salt media (BSM)	It is a minimal defined medium and it supports the growth of a variety of bacteria.	(Hareland et al. 1975)
BSM-G	It is derived from BSM and supplied by Glycerol as antibiotics inducer and the bacteria could use it at a carbon and hydrogen sources.	(O'Sullivan et al. 2007)
BSM-GBP	It is derived from BSM-G and supplemented by Glycerol, bromoacetic acid and polymixin B as selective agents. The bromoacetic acid bacteria could be used as a carbon and hydrogen sources by bacteria.	This study
Asdown's medium	It is a selective medium to isolate the clinical species of *Burkholderia pseudomallei*. It is derivative from tryptone agar TSA supplied by gentamicin.	(Ashdown 1979)
***Pseudomonas cpacia* azelaic acid (PCAT)**	It is a selective medium to isolate the environmenta *Burkholderia* spp. It contains salts and Azelaic acid and Tryptamin as selective agents.	(Burbage and Sasser 1982; Peeters et al. 2016)
Trypan blue-tetracycline (TBT)	It is a selective medium for selectivity o *Pseudomonas cepacia* from soil.	(Hagedorn et al. 1987)
***Burkholderia cepacia* selective agar (BCSA)**	It is a selective medium contains crystal violet phenol red, lactose, sucrose and three selective agents: polymyxin B, gentamicin, and vancomycin	(Henry et al. 1997)
***Burkholderia Pseudomallei* Selective Agar (BPSA)**	It is a selective medium for *B.pseudomallei, B pseudomallei, B. cepacia,* and *P.aeruginosa.*	(Howard and Inglis 2003)
***Burkholderia cepacia complex* enrichment (BCEM)**	It is a selective medium for *B. cepacia* from cystic fibrosis from patients' sputa	(Flanagan and Paull 1998)
***Pseudomonas cepacia (*PC)**	It is a selective media for *B. mallei* and *B.pseudomallei* and contains crystal violet, bile salts ticarcillin, and polymyxin B.	(Gilligan and Schidlow 1984; Glass et al. 2009)
Stewart	It is a selective media for BCCs from CF patients	(Stewart 1971; Vanlaere et al. 2006)
Mast *B. cepacia* medium	It is selective media to isolate *B. cepacia* from CF patients	(Gilligan and Schidlow 1984)
LAB M *B. cepacia* médium	It is selective media to isolate *B. cepacia* from CF patients	LAB M,Ltd, Burry, UK and (Vermis et al. 2003)
Oxoid *B. cepacia* medium	It is selective media to isolate *B. cepacia* from CF patients	Oxoid ltd, Basingstoke, UK and (Vermis et al. 2003)

3.1.1 Aim and objective

To develop enrichment media for environmental *Burkholderia* and *Paraburkholderia*, bromoacetic acid was chosen as a selective agent to potentially enrich them and was used in combination with the Basal salt media (BSM-G) (Mahenthiralingam et al. 2011) which actively supports the growth of these species. The rationale for this was based on the fact that strains such as *Burkholderia cepacia* MBA4 can utilize this halogenated compound, bromo-acetic acid, as a sole carbon source (Yu et al. 2007). *Pseudomonas cepacia* azelaic acid tryptamine was recently shown to be an effective medium for the isolation of *Burkholderia* from soil and water samples (Peeters et al. 2016). Therefore, PCAT was also evaluated in this study, alongside the BSM-G with bromoacetic acid as a novel enrichment medium. The hypothesis

behind this chapter was that the longterm incubation of sediment samples in a minimal medium Basal salt medium (BSM-G) containing bromoacetic acid demonstrated that *Burkholderiales* bacteria were enriched by the presence of this halogenated hydrocarbon. The overall aim of this chapter was to develop culture-dependent methods to isolate *Burkholderia* and *Paraburkholderia* from the natural environment. This was achieved with the following objective:

- **The development of a broth-based growth approach to identify *Burkholderia* and *Paraburkholderia* from rhizospheres and soil samples.** PCAT was chosen as a control medium to be used in this chapter, alongside the evaluation of a bromoacetic acid based enrichment medium.

3.2 Results

3.2.1 Bromoacetic Acid is not 100% selective for *Burkholderia* and *Paraburkholderia*

While a number of growth media capable for selection and enrichment of *Burkholderia* have been developed (Table 3.1), with increasing diversity associated with newly defined species and genera such as *Paraburkholderia*, the development of novel media to improve isolation of these bacteria from the natural environment was the first experimental objective for this study. To begin media development, a growth analysis of selected reference *Burkholderia* and *Paraburkholderia* species was screened in TSB and BSM-G using a BioScreen C to examine planktonic growth rates. The collection of species screened was selected to represent the current diversity of *Burkholderiales*, and had example species from both *Burkholderia* and *Paraburkholderia*. After initial screening the results showed that *Burkholderia* could be divided into fast (μ_{max} 3 or greater), medium (μ_{max} 2 or greater), and slow-growers (μ_{max} 1 or less), in terms of *in vitro* growth (Table 3.2). At the beginning of this study all the strains selected were classified as either *Burkholderia* or *Paraburkholderia* species. However, by the time this thesis was written several of the species had been reclassified into additional new genera as shown (Table 3.2).

Table 3.2 The growth rate of *Burkholderia* and *Paraburkholderia* species and the effect of bromoacetic acid on this

Category	Species	category	μ_{max} (TSB)	µmax (BSM-G + BroA)	MICs (BroA) *
Fast growth	*B. vietnamiensis* (BCC0042)	*Burkholderia*	0.32	0.15	1.5 **m**M
	B. gladioli (BCC0238)	*Burkholderia*	0.34	0.24	1.5-0.7 mM
	P. xenovorans (BCC0657)	*Paraburkholderia*	0.34	0.30	1.5 mM
	Robbsia andropogonis (*BCC*0766)	*Robbsia*	0.32	0.05	0.1 mM
	Burkholderia sp (*BCC*1635)	*Burkholderia*	0.32	0.3	1.5-0.7 mM
	P. thailandensis (BCC0779)	*Burkholderia*	0.30	0.25	1.5 mM
Medium growth	*P. caryophylli* (BCC0769)	*Paraburkholderia*	0.23	0.26	3.1 mM
	P. phytofirmans (BCC1604)	*Paraburkholderia*	0.20	0.27	12.5 mM
	P. bryophila (BCC1876)	*Paraburkholderia*	0.26	0.26	6.25 mM
	P. phymatum (BCC1607)	*Paraburkholderia*	0.26	0.23	3.1 mM
Slow growth	*B. glumae* (BCC0773)	*Burkholderia*	0.13	0.25	0.7 mM
	B. graminis (BCC0774)	*Paraburkholderia*	0.086	0.26	25 mM
	P. tuberum (BCC1610)	*Paraburkholderia*	0.03	0.1	3.1 mM
	P. tropica (BCC1637)	*Paraburkholderia*	0.084	0.2	6.25 mM

*** BSM-G (0.4%) with with different concentrations of Bromoacatic Acid (0.1mM to 100mM)**

To evaluate a new selective medium, BSM-G was modified by adding different concentrations of bromoacetic acid, and this medium designated “BSM-G with bromoacetic acid” BSM-GB (Table 3.2). Since *Burkholderia* are known to be resistant to polymixin B, providing the selective basis for media such as BCSA (Henry et al. 1997) and BCEM (Flanagan and Paull 1998) (Table 3.2), it was also added to BSM-GB to bring additional selectivity (designated BSM-GBP). Thus, BSM-GBP was developed as an enrichment media for *Burkholderia* and *Paraburkholderia*, based on bromoacetic acid and polymixin B as selective agents (Table 3.2).

Initial liquid growth experiments showed that all the reference *Burkholderia* and *Paraburkholderia* species were capable of growth in the presence of bromoacetic acid, with certain strains such as *Robbsia andropogonis* being very susceptible (MIC = 0.1 mM) and other strains such as *Paraburkholderia phytofirmans* being highly tolerant (MIC = 12.5 mM) of the halogenated hydrocarbon (Table 3.2). The efficacy of BSM-GBP was further investigated by increasing the number of strains and species evaluated as follows: 70 *Burkholderia*, 26 *Paraburkholderia* and 32 non-*Burkholderia* (30 Gram-negative and 2 Gram-positive) control species were screened for growth on BSM-GBP media. Other well-characterised media PCAT (Peeters et al. 2016), TSA and BSM-G (Mahenthiralingam et al. 2011) were also compared to the growth on BSM-GBP (Table 3.3). To screen this large collection of *Burkholderia, Paraburkholderia* and control bacterial, high throughput replica plating of strains was carried out on the different growth media (Figure 3.1).

Table 3.3 *Burkholderia*, *Paraburkholderia* and Non-*Burkholderia* screening on a range of enrichment and control media

Media	*Burkholderia* (n= 70)	*Paraburkholderia* (n= 26)	*Non-Burkholderia* (n=32)
Control media			
TSA	100%	100%	100%
BSM-G	100%	100%	100%
Enrichment media			
PCAT	96%	77%	15%
BSM-GBP 0.5 mM	98%	80%	47%
BSM-GBP 1 mM	98%	80%	50%
BSM-GBP 1.5 mM	97%	73%	50%
BSM-GBP 2 mM	93%	65%	47%
BSM-GBP 2.5 mM	88%	50%	25%

The growth for *Burkholderia, Paraburkholderia* and non-*Burkholderia* in TSA and BSM-G (control media), PCAT (an enrichment control medium) was compared with the novel BSM-GBP medium containing increasing concentrations of bromoacetic-acid. All *Burkholderia, Paraburkholderia* and non-*Burkholderia* grew well on TSA and BSM-G, except one *Paraburkholderia* species isolate (BCC1885); this strain took more than one week to grow in BSM-G and but did grow within a week on PCAT (Table 3.3).

At low concentrations of bromoacetic acid (1 mM; and when 600 units/mL of polymixin B was present), 98 % *Burkholderia and* 80% *Paraburkholderia* isolates grew well in the BSM-GBP medium. In contrast, 88% *Burkholderia* and 50% *Paraburkholderia* isolates were able *to* tolerate concentrations of bromoacetic acid at 2.5 mM (Table 3.3, Figure 3.1 and 3.3).

For non-*Burkholderia* species, 47% grew at 0.5mM bromoacetic acid, with tolerant species including *Staphylococococcus aureus* (Gram-positive), *Serratia marcescens* and *Enterobacter cloacae* (Gram-negative). Increasing the bromoacetic acid to 2.5 mM was able to prevent growth of 25% non-*Burkholderia* organisms like *Candida albicans* (an ascomycete fungus) and *Ralstonia mannitolylitica* (Gram-negative) (Table 3.3, Figure 3.2). 8 out of 32 non-*Burkholderia* grew at 2.5 mM bromoacetic acid, such as *P. aeruginosa* and *R. mannitolylitica* (both Gram-negative) (Table 3.3, Figure 3.2). Thus, bromoacetic acid was not 100% selective for *Burkholderia* and *Paraburkholderia*, as certain non-*Burkholderia* species could grow during the test enrichments.

Comparing BSM-GBP media with PCAT, both media were not 100% selective for *Burkholderia* or *Paraburkholderia*. High concentrations of bromacetic acid that were 2.5 mM and above had a general toxic effect on the species screened. The result confirmed that *Burkholderia* and *Paraburkholderia* isolates grew well at low concentrations of bromoacetic acid up to 2.5 mM, when 600 units/mL of polymixin B was also present in the BSM-GBP medium (Figure 3.3). The high concentration of bromoacetic acid (2.5 mM) led to growth inhibition of 8 of 70 *Burkholderia* such as *B. vietnamiensis*, *B. pyrrocinia and B. multivorans,* and 8 out of 26 *Paraburkholderia*

such as *B. caryophylli* (BCC0769) and *B. glathei* (BCC0772). In comparison, certain *Burkholderia* species (*Burkholderia stagnalis*, *Burkholderia ambifaria*, and *Burkholderia cenocepacia*) and *Paraburkholderia* species (*Paraburkholderia 'Robbsia' andropogonis* and *Paraburkholderia hospital*) were selected by PCAT and prevented by 2.5 mM bromoacetic acid (Figure 3.3).

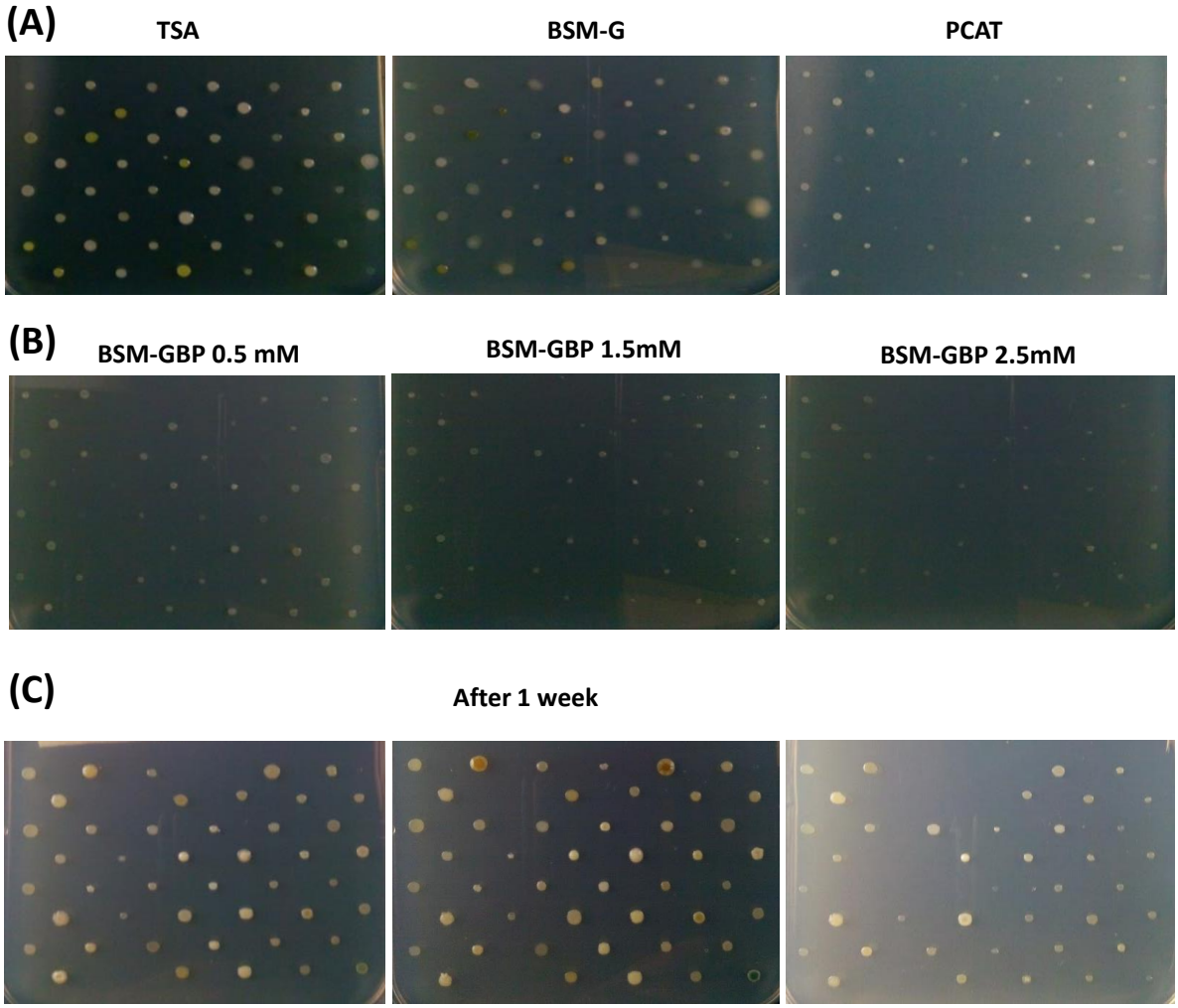


Figure 3.1 The growth of the *Burkholderia* and *Paraburkholderia* strain panel on different media. (A) *Burkholderia* grew well in the control medium TSA, BSM-G and PCAT. (B) Weak growth of all species was seen after 1 day of growth in bromacetic acid containing medium. (C) After 1 week of growth *Burkholderia and Paraburkholderia* strains grew better in BSM-GBP (0.5 and 1.5mM), but 2.5mM bromoacetic acid inhibited 8 of 48 strains.

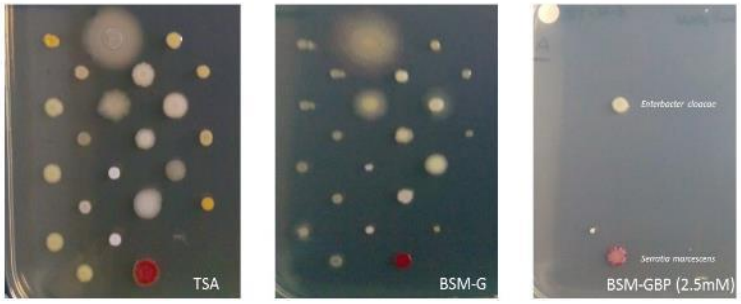


Figure 3.2 Inhibition of non-*Burkholderia* by the presence of bromoacetic acid. The Non-*Burkholderia* species evaluated grew well on TSA and BSM-G. Two Gram-negative non-*Burkholderia* grew BSM-GBP with the high, 2.5 mM concentration of bromoacetic acid.

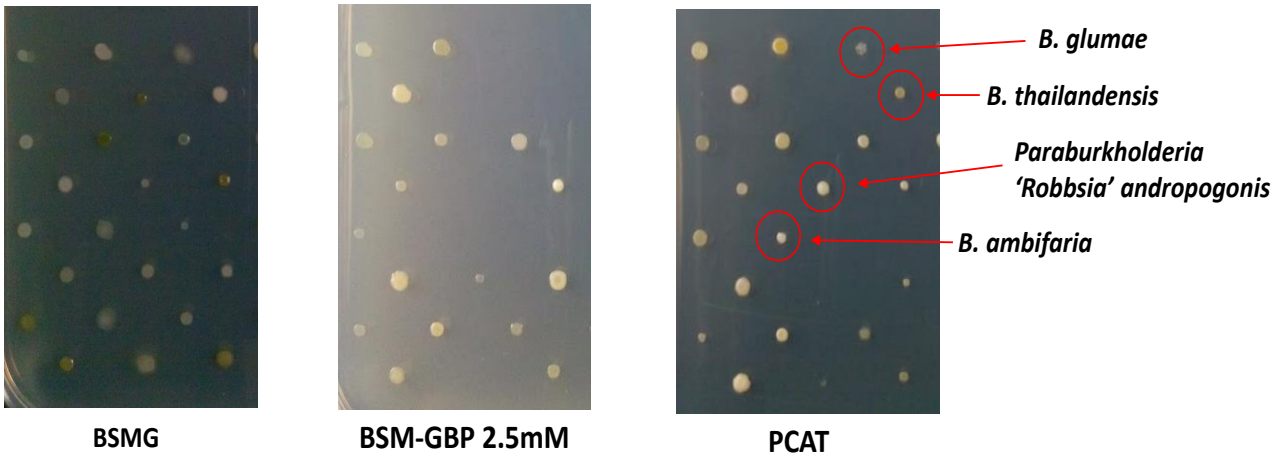


Figure 3.3 The comparison of BSM-G, PCAT and BSM-GBP 2.5 mM. PCAT was less inhibitory toward certain species and better able to select species, such as *Burkholderia ambifaria*, *Paraburkholderia 'Robbsia' andropogonis, Burkholderia glumae* and *Burkholderia thailandensis.* These species were prevented from growing by BSM-GBP (2.5 mM) (see red circle for absent growth).

3.3 Discussion

Bromoacetic acid was not 100% effective at enriching for *Burkholderia* and *Paraburkholderia*, even though it was capable of inhibiting the growth of non-*Burkholderia* bacteria. Thus, bromoacetic acid was not selected as an enrichment media for subsequent environmental enrichment of *Burkholderia* and *Paraburkholderia.* Instead, a screen of jungle environmental samples from Sabah, Malaysia, was initiated using PCAT as the selective medium to isolate *Burkholderia* and *Paraburkholderia.* In this initial testing, PCAT was less inhibitory towards

these bacteria than the bromoacetic acid-based medium, corroborating its recent successful use by others to isolate these species from natural sources (Peeters et al. 2016).

There are multiple growth media that can be used for *Burkholderia* species isolation (Table 3.1), and most of them were developed for pathogenic *Burkholderia* species. PCAT was not originally developed as a medium to isolate pathogenic *Burkholderia* (Burbage and Sasser 1982), and recently it has been shown to be effective for isolation of a range of *Burkholderiales* species from soil and water samples (Peeters et al. 2016). In contrast, the ability of *Burkholderia* to grow on and to degrade recalcitrant aromatic organic compounds, such as polychlorinated biphenyls, is a feature of environmental *Burkholderia* bacteria. It has been found in biodegradation studies that *Burkholderia cepacia* MBA4 can utilize the halogenated compound bromoacetic acid (Yu et al. 2007).

As described in the introductory chapter, the taxonomy of *Burkholderia* and *Paraburkholderia* genera changed considerably in the last 10 years, and it is still evolving with a need to fully define the diversity within these genera (Dobritsa and Samadpour 2019). The selected panel of *Burkholderia* and *Paraburkholderia* used to test BSM-GBP encompassed a well characterised collection of reference strains from the Cardiff University collection and other international strain repositories. However, even during its assembly and testing as a collection, it had included species that were subsequently reclassified. No clear evidence of BSM-GBP enriching any specific *Burkholderia* or *Paraburkholderia* species was observed in this study, and this was not unexpected because of the complex classification and diversity of these bacterial groups. Several members of the *Burkholderia* genus were reclassified to five genera (*Paraburkholderia*, *Trinickia*, *Mycetohabitans*, *Robbsia* and *Caballeronia*) over the last five years. For example, *Paraburkholderia andropogonis*, *Paraburkholderia caryophylli* and *Paraburkholderia glathei* were used on the BSM-GBP screening as *Paraburkholderia* species, and they were subsequently reclassified as *Robbsia andropogonis*, *Trinickia caryophylli* and *Caballeronia glathei* (Aaron et al. 2000; Dobritsa and Samadpour 2016; Lopes-Santos et al. 2017; Estrada-de los Santos et al. 2018). This may have hindered the development of the

bromoacetic acid-based selective agar (BSM-GBP). For examples, *Robbsia andropogonis* was found to be susceptible to 0.1 mM of bromoacetic acid as quite a low level of the hydrocarbon, and this would have been expected as it is not a *Burkholderia* or *Paraburkholderia* species. In future, it would be more appropriate to select type strains which have been accurately placed in relation to new classifications as references for each *Burkholderia* and *Paraburkholderia* with this type of screening and evaluation of growth media.

A recent study has evaluated a selective genome-guided method for isolation of environmental *Burkholderia* isolates, which also enabled growth media development in multiple ways (Haeckl et al. 2019). They evaluated 18 selective media, systematically examining what carbon and nitrogen sources and inhibitory compounds would be selective in enrichment of environmental *Burkholderia* species. The conclusion of this study prioritised five media (BIB, BIC, BID, BIE, BIF) as well as PCAT media for testing. PCAT was also used as a control medium to evaluate the developed basal growth media, and it was successfully applied to isolate environmental *Burkholderia* in this study (Haeckl et al. 2019). All the five developed media were variations of basal medium with additional additives to target environmental *Burkholderia* species, which included carbon and nitrogen sources, as well as selective antimicrobial agents to which *Burkholderia* species are resistant to. They prioritised fusaric acid, bacitracin, and acriflavine as selective antibiotics, and copper and nickel compounds as promising metal candidates to force enrichment of *Burkholderia*. They found *Burkholderia* were effectively resistant to these compounds compared with untargeted microorganisms. For identification of the optimum carbon and nitrogen sources metabolized by *Burkholderia*, there was no unique sources specifically identified for *Burkholderia*, correlating to the ability of these organisms to grow in a range of diverse environments. However, they found that L-sorbose, hydroxyproline, L-arginine, and D-glucosamine were utilized by *Burkholderia* and poorly metabolized by their non-*Burkholderia* panel. Finally, with one of their validated media they replaced the yeast extract on the basal media with an additive as a sole carbon source (Haeckl et al. 2019). For future study, it would be interesting to apply bromoacetic acid as a sole carbon source in this

basal medium, replacing the yeast extract. Haeckle and colleagues (2019) and other studies priotised BSM media for the development of *Burkholderia* media, as an optimal minimal media for the growth of *Burkholderia* bacteria. The same finding was observed in the present study, as BSM-G was shown to be the optimum growth medium for the *Burkholderia* and *Paraburkholdria* selection panels.

3.4 Conclusion

Overall, via comparative testing of reference strains, PCAT was shown to be effective and successfully able to support the growth of *Burkholderia* and *Paraburkholderia*. A concentration of 2.5 mM bromoacetic acid has a toxic effect on 12% *Burkholderia* and 50% *Paraburkholderia* isolates, and it also allowed the growth of certain non-*Burkholderia* species. Overall this initial attempt at development of a bromoacetic acid based enrichment medium for *Burkholderia* and *Paraburkholderia* was not successful and further analysis using this selective agent was not performed.

Chapter 4. Screening natural environments for the presence of *Burkholderiales*

4.1 Introduction

After gaining an understanding of how different *Burkholderia* and *Paraburkholderia* grow and can be enriched in different culture media (Chapter 3), assembling a novel environmental collection of these types of bacteria was the main aim of this chapter. *Burkholderia* and *Paraburkholderia* genera thrive in many natural environments (Sawana et al. 2014; Depoorter et al. 2016; Dobritsa and Samadpour 2016), yet the broad importance of these genera as antimicrobial producers has not been intensively studied (Depoorter et al. 2016). To date, environmental habitats screened for these bacteria include, soils (Hall et al. 2015), rhizosphere of crops such as maize (Ramette et al. 2005), and river samples (Peeters et al. 2016). An abundance of *B. cepacia* complex (Bcc) species were isolated from seven samples of maize rhizosphere (Ramette et al. 2005). The following Bcc species were recovered from the maize samples: *B. ambifaria*, *B. cepacia*, *B. multivorans*, *B. cenocepacia*, *B. stabilis*, *B. dolosa*, and *B. pyrrocinia* (Ramette et al. 2005). Notably, the maize rhizosphere samples were an environmental source of the species most frequently isolated in patients with cystic fibrosis (CF), which are *B. multivorans* and *B. cenocepacia* (LiPuma 2010). However, *B. multivorans* was not found at high density in this US maize study (Ramette et al. 2005), and was not found at all in a study of Italian maize rhizospheres (Dalmastri et al. 2007). Few studies have reported on *Burkholderia* species isolated from tropical environments, and especially members of the Bcc. However, *Burkholderia pseudomallei*, a highly pathogenic *Burkholderia*, has been routinely isolated from such environments (Radua et al. 2000; Sadiq et al. 2018). Furthermore, a recent study showed that *B. paludis,* is a unique Bcc strain isolated from Malaysian tropical peat swamp soil (Ong et al. 2016). Several new *Paraburkolderia* species have been isolated

from diverse environments. For example, *Paraburkholderia caffeinilytica*, *Paraburkholderia aromaticivorans*, and *Paraburkholderia caseinilytica* were isolated from tea plantations, gasoline-contaminated soil, and forest soil, respectively (Gao et al. 2016; Gao et al. 2018b; Lee and Jeon 2018).

Multiple studies have shown that the genus *Burkholderia* is a promising source of antimicrobial agents that encodes specialised metabolite biosynthetic gene clusters (Depoorter et al. 2016). The ability of *B. cepacia* complex bacteria to limit the growth of certain multidrug-resistant pathogens has also been observed, and additionally the ability to produce antifungal and biopesticidal agents in natural environments (Parke and Gurian-Sherman 2001; LiPuma 2010; Vandamme and Peeters 2014; Depoorter et al. 2016; Mullins et al. 2019). Recent studies on antibiotic production in these species have shown that *B. ambifaria* produces enacyloxin IIa (Mahenthiralingam et al. 2011) and that *B. gladioli* BCC0238 produces gladiolin (Song et al. 2017), both of which are potent antibiotics able to restrict the growth of AMR pathogens. Gladiolin is a novel polyketide belonging to the macrolide antibiotic class that shows activity against the various strains of *Mycobacterium tuberculosis* with MIC range (0.4 to >2.3 *μ*g/mL for sensitive strains). Gladiolin also has limited activity towards Gram-negative, such as *Pseudomonas aeruginosa* DSM29239 and *Burkholderia multivorans* ATCC17616 with the MIC= >64 *μ*g/mL for both; it does have good activity against Gram-positive bacteria, such as *Staphylococcus aureus* DSM21979 (MIC= 8 μg/mL), and the yeast *C.albicans* (MIC= 4 μg/mL) (Song et al. 2017). Enacyloxins are another polyketide antimicrobial produced by *B. ambifaria*, showing activity against *B. multivorans*, *C. albicans*, and *S. aureus* (Mahenthiralingam et al. 2011). Cepacin, quinolinones, cepaciamide, xylocandins, phenylpyrroles, and phenazines are well-known antifungal compounds produced by Bcc species (Parke and Gurian-Sherman 2001; Vial et al. 2007) (see Table 1.1 in chapter 1). The importance of the Bcc group as bioremediation or plant control agents has also been well established (Eberl and Vandamme 2016). Exploring new and diverse ecological niches in search of new antibiotic producing bacteria is essential to combat AMR. Malaysian jungle plants from Sabah have not yet been

investigated for the presence of antibiotic producers from the *Burkholderiaceae* family of bacteria. This study provided a unique opportunity to screen a tropical jungle environment for the presence of these bacteria.

4.1.1 Aim and objectives

Cardiff University has a field centre in Sabah, Malaysia. This provided an opportunity to obtain environmental samples from the Bornean jungle. Permission was obtained from the Sabah Wildlife board to sample the environment, and rhizosphere-associated samples from jungle plants and additional environmental samples were collected by E. Mahenthiralingam, Cardiff University in August 2008. Enriched microbial cultures of the soil and rhizosphere material were made and stored frozen at -80°C, but no further analysis was performed until the screening described within this PhD. The overall aim of this chapter was to assemble a systematic collection of *Burkholderia* and *Paraburkholderia* isolates from the Sabah jungle as an uncharacterised natural environment. This was achieved with the following objectives:

1. **The application of a developed *Burkholderia* and *Paraburkholderia* isolation strategy.** Established culture-dependent (PCAT enrichment medium) and culture-independent (*recA* gene PCR) methods were used to detect and isolate environmental *Burkholderia* and *Paraburkholderia* from the Bornean jungle samples.
2. **Identification of isolated bacteria using single-gene-based phylogenetic analysis (16S rRNA and *recA* genes).** The use of single-gene phylogenetic analysis was initially applied to investigate the diversity of *Burkholderiales* in environmental jungle samples.
3. **Application of random amplified polymorphic DNA fingerprinting (RAPD) PCR to reduce sequential strains.** De-replication of the number of identical strains in the collection was carried out using a basic RAPD-PCR- strain genotyping method.

4.2 Results

4.2.1 The strategy for screening the Malaysian jungle samples

To establish a large collection of environmental *Burkholderia* and *Paraburkholderia* isolates for subsequent bioactivity and genome mining analysis, a combination of culture-dependent and -independent methods was used. It was important in the beginning of this screening to test a screening strategy for a total of 98 rhizosphere-associated samples collected from Bornean jungle. This chapter evaluated a combination of PCR-based molecular identification and growth-based methods for the enrichment and isolation of *Burkholderia* and *Paraburkholderia*. The details of each method are described in detail in Chapter 2, however the overall screening strategy employed in this chapter references and links how each method was used to assemble a collection of bacteria. A preliminary screen of 57 jungle samples was initially performed to evaluate growth on different media and determine which PCR methods would be useful. Reference *Burkholderia* and *Paraburkholdria* strains (see Chapter 3) were included as positive controls. Since bromoacetic acid was shown to be inhibitory for several *Burkholderia* and *Paraburkholderia* species and was not sufficiently counter-selective against non-*Burkholderia* and -*Paraburkholderia* organisms (Chapter 3), screening was performed using PCAT as a media with proven ability to isolate both of these genera (Peeters et al. 2016) and BSM-G as a non-selective control medium which broadly supports growth of *Burkholderiaceae* (Mahenthiralingam et al. 2011).

PCR detection of the bacterial 16S rRNA gene was positive for all 57 BSM-G enriched samples, showing that bacteria had successfully grown in this non-selective minimal medium (Figure 4.1, panel A, Table 4.1). However, enrichment with PCAT showed that only 85% of these cultures had bacteria that could be detected by 16S rRNA gene PCR (Table 4.1). PCR screening using the *Burkholderiaceae*-specific *recA* gene (Payne et al. 2006) showed that 78% of PCAT enrichment cultures were positive for *Burkholderia/Paraburkholderia,* but only 63% were positive from BSM-G cultures (Table 4.1, Figure 4.1, panel B). The results of this

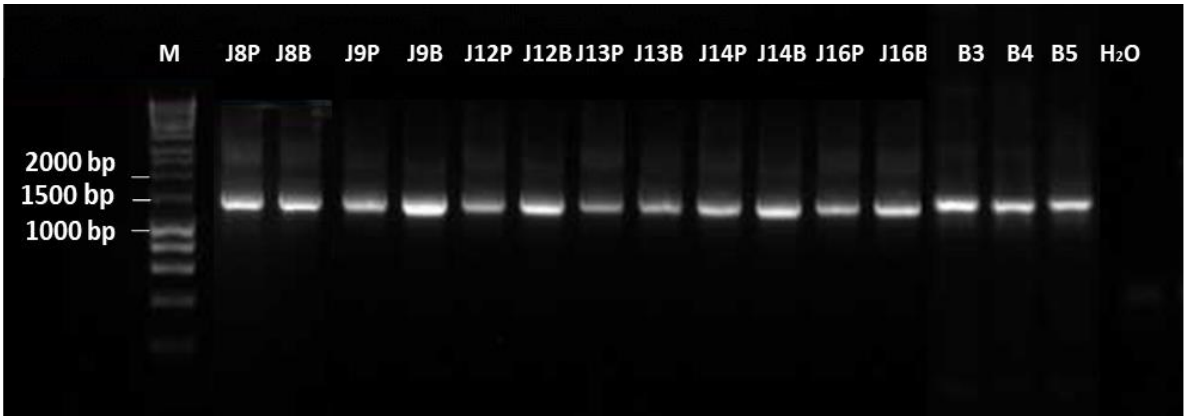
initial screen (Table 4.1) demonstrated that PCAT media performed better as an enrichment media for *Burkholderia/Paraburkholderia* and that *recA* gene amplification from these enrichments could identify positive samples worthy of follow-up. Using these results, a final strategy to isolate and characterise the jungle *Burkholderia* and *Paraburkholderia* isolates was developed (Figure 4.2). This was applied to all 98 samples to ensure that the time-consuming purification of individual *Burkholderia/Paraburkholderia* strains was performed only on those environmental samples most likely to be positive.

The final strategy (Figure 4.2) began with growing the samples in PCAT selective medium to enrich the environmental *Burkholderiaceae* species present (Peeters et al. 2016). After allowing the microbial community to grow in PCAT for a week at 30°C, DNA was extracted from the cultures for PCR screening (16S rRNA and *recA* gene PCR). The *recA* gene is a useful taxonomic discriminator for multiple *Burkholderia* species (Payne et al. 2005). Finally, *recA* gene-PCR positive samples were diluted and plated to single colonies on PCAT agar using a spiral plater. Single colonies were then re-purified on a nutrient-rich TSA growth medium (3 passages) prior to being stored at -80°C. Random amplified polymorphic DNA fingerprinting PCR (RAPD-PCR) was applied to these individual pure cultures to remove identical isolates. Once individual strains were isolated high-quality DNA was extracted from them using an automated system (Maxwell 16, Promega, UK). Within this chapter their preliminary identification was carried out using 16S rRNA and *recA* gene sequencing. Whole-genome sequencing (WGS) of the environmental *Burkholderia* and *Paraburkholderia* collection was performed to improve the resolution of their identification (Chapter 5) and subsequently antimicrobial bioactivity screening and genome mining was applied to identify antibiotic biosynthesis clusters of interest (Chapter 6).

Table 4.1 The results of screening different enrichment media by 16S rRNA and *recA* gene PCR

Growth media	Type of PCR	Number of +PCR cultures	% of +PCR cultures
BSM-G	16S rRNA PCR	57/57	100
	recA PCR	36/57	63
PCAT	16S rRNA PCR	49/57	85
	recA PCR	45/57	78

(A) 16S rRNA gene PCR



(B) *recA* gene PCR

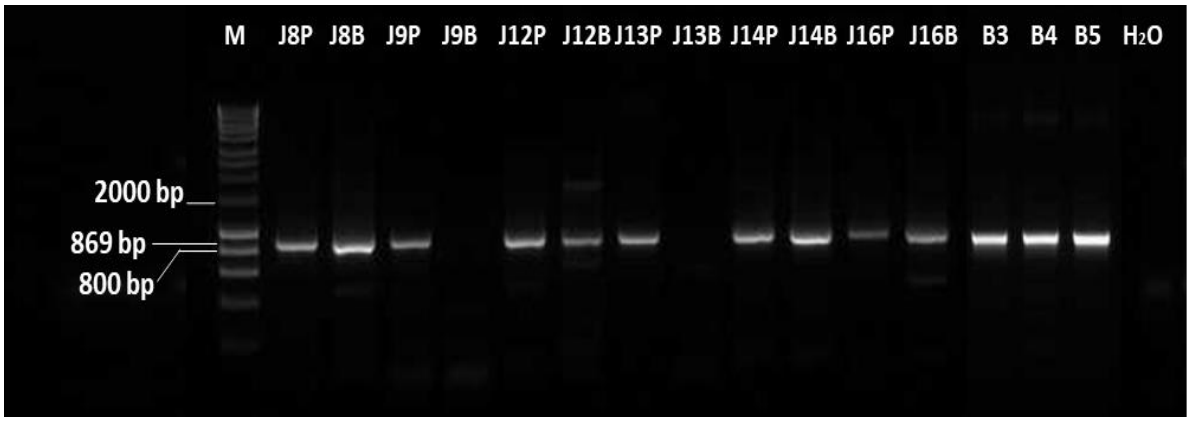


Figure 4.1 PCR amplification of 16S rRNA and *recA* genes from jungle sample enrichments. DNA was extracted from PCAT and BSM-G cultures for six jungle samples (J8, J9, J12, J13, J14, and J16). Each lane is labelled with the jungle sample identifier—J refers to the Jungle source—followed by the sample collection number and P referes to the isolate from PCAT enrichment and from a BSM-G culture. Panel A shows the 16S rRNA gene PCR and Panel B the *recA* gene PCR; molecular size markers are shown in lane M together with relevant DNA fragment sizes. The control DNA samples from reference strains were included as follows: B3: *Burkholderia vietnamiensis* (BCC0042), B4: *Paraburkholderia tropica* (BCC1608) and B5: *Paraburkholderia phenoliruptrix* (BCC1611). Jungle samples 9 and 13 were enriched using PCAT but did not grow bacteria in BSM-G, illustrating the improved ability of PCAT to enrich the *Burkholderia* and *Paraburkholderia* bacteria.

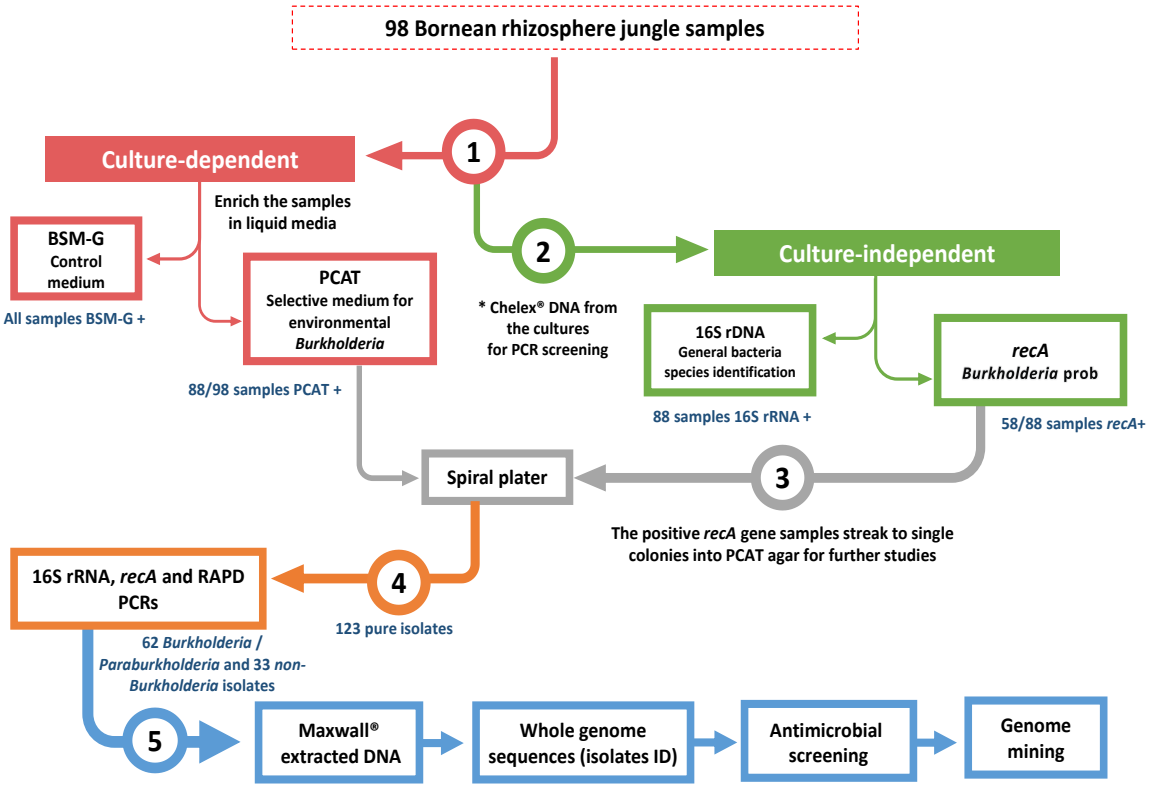


Figure 4.2 The strategy for isolating environmental *Burkholderia* and *Paraburkholderia* bacteria from jungle samples. First, enrichment of the samples in a liquid medium called *Pseudomonas cepacia* azelaic acid (PCAT) selective medium was used to isolate the environmental Burkholderiaceae species (Peeters et al. 2016). Secondly, DNA was extracted from the cultures for PCR screening using 16S rRNA and *recA* gene PCR. Note the *recA* gene is a useful taxonomic discriminator for multiple *Burkholderia* species (Payne et al. 2005). Thirdly, *recA*-positive samples were diluted and plated to single colonies on PCAT agar using a spiral plater. Single colonies were re-purified on Trytic Soya Agar (3 passages) prior to the storage of pure isolates. Fourthly, RAPD PCR was used to de-replicate duplicate isolates from the same jungle samples and 16S rRNA gene and recA gene PCR followed by sequencing used to identify the *Burkholderia* and *Paraburkholderia* isolates, and differentiate these from non- *Burkholderia* and *Paraburkholderia* bacteria. DNA was extracted from selected isolates using an automated system (Maxwell 16, Promega, UK). The whole-genome sequencing (WGS) of the environmental *Burkholderia* and *Paraburkholderia* collection was performed. Finally, bioactivity screening via an antimicrobial antagonism overlay assay and genome mining was applied to identify antibiotic biosynthesis clusters of interest and novel specialized metabolites.

4.2.2 Initial PCAT enrichment of Malaysian environmental samples

After mapping out a sample screening strategy (Figure 4.2), all 98 jungle samples were enriched in PCAT broth for 72 h to one week at 30°C with the aim of enriching the *Burkholderia*/*Paraburkholderia* community within the samples. The initial result of this screening showed that 10 of the 98 jungle samples could not be enriched in PCAT broth (samples 25, 58, 59, 64, 66, 77, 84, 87, 90, and 96) (Table 4.2). The remaining 88 jungle samples that grew on PCAT broth showed the following growth pattern: 46 of the jungle samples grew well within 72 h, and clear single colonies were obtained from these samples

when plated onto PCAT agar; sixteen samples grew slowly, taking between 72 h and one week to reach a culture density that could be followed up by plating onto PCAT medium; nine samples grew slowly but contained small amounts of fungal contamination; and seventeen samples developed significant fungal contamination that ultimately spread widely throughout the PCAT growth plates and precluded follow-up isolation of bacterial colonies. For example, the seventeen samples that were heavily contaminated by fungi, could not be purified by performing enrichment in the presence of 200 or 400 mg/L cycloheximide within the PCAT, as the fungal contaminants were highly resistant to cycloheximide (Table 4.2).

Table 4.2 Jungle sample growth observation in PCAT broth

Growth in PCAT	Samples no.	Samples IDs
Good	46	1, 6, 7,8, 10,11,12, 15, 16, 17, 18, 19, 23, 24, 26, 27, 34, 35, 36,41, 42, 45, 47, 48, 49, 50, 63, 67, 69, 70, 72, 74, 75, 76, 78, 79, 80, 85, 86, 88, 91, 92, 94, 95, 97, 98
Slow	16	2, 3, 5, 9, 13, 40, 51, 53, 54, 57, 60, 62, 71, 73, 89
Slow with fungus contamination	9	14, 20, 21, 30, 33, 38, 44, 55, 81
Overwhelming fungal contamination	17	4, 22, 28, 29, 31, 32, 37, 39, 43, 46, 52, 56, 61, 65, 68, 83, 93
No growth	10	25, 58, 59, 64, 66, 77, 84, 87, 90., 96

A cultivation-independent approach, *recA* gene PCR, was used to rapidly detect the potential presence of *Burkholderia* and *Paraburkholderia* species in the 88 PCAT culture enrichments. The analysis showed that 58 out of 88 (65%) of rhizosphere samples were *recA*-positive and potentially contained *Burkholderia* and *Paraburkholderia* (Table 4.3, Figure 4.3). Jungle samples 3, 5, and 9 were examples of slow-growing samples in PCAT (Table 4.3), but despite this the *recA* gene amplification was successful showing that they should be followed up by individual colony isolation. Slow-growing and fungal-contaminated samples J3 and J12 (Table 4.3, Figure 4.3) also tested positive for *recA* gene amplicons of the correct size predicted for *Burkholderia* and *Paraburkholderia*. Overall, this initial enrichment and molecular analysis demonstrated that the Malaysian rhizosphere samples from the Sabah jungle were a rich

source potential *Burkholderia* and *Paraburkholderia* bacteria worthy of full isolation by microbial culture.

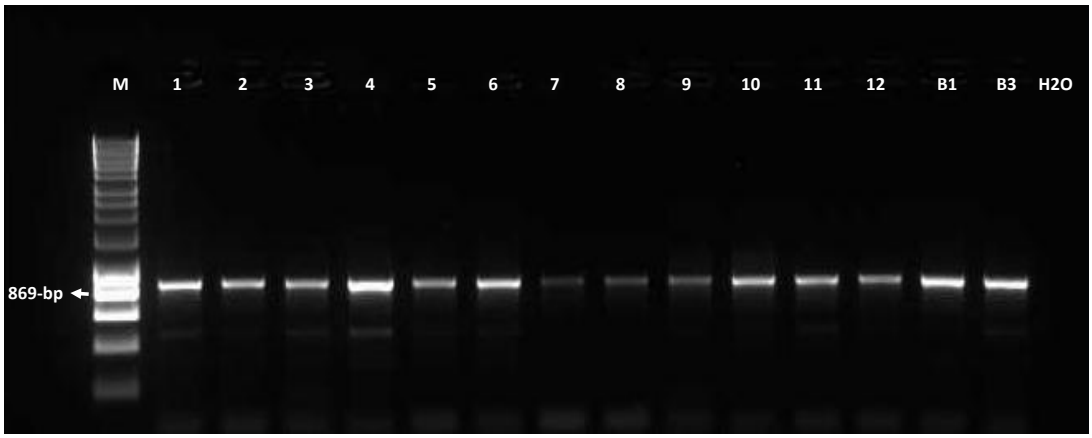


Figure 4.3 Detection of the *recA* gene for jungle isolates. Lane 1 is the 1 KB molecular marker, and lane 2 to lane 13 are the first 12 jungle samples (jungle sample number 1 to jungle sample number 12) grown in different plates as shown in Table 2 above. Lane 14 (B1) is *B. vietnamiensis* (BCC0042) and lane 15 (B3) is *P. tropica* (BCC1608). Lane 16 is the water negative control. The expected size of the *recA* gene amplicon is labelled as 869 bp. This screening confirmed that the first 12 jungle samples were contained *Burkholderia* and *Paraburkholderia* bacteria.

4.2.3 Isolation of pure cultures of putative *Burkholderia* and *Paraburkholderia*

The *recA*-positive enriched culture samples were diluted tenfold and plated to single colonies on PCAT agar using a spiral plater. The spiral plater instrument is an automated system for plating microorganisms onto agar Petri dishes that dilutes the sample from the inside to the outside of the plate in a spiral pattern. This technology helped to isolate single colonies from the microbial communities enriched within the selective media. Variable incubation times for the PCAT cultures were applied ranging from 72 h to one week at 30°C to allow the individual colonies to grow (Figure 4.4). Colonies representative of the different morphotypes present (Figure 4.4) were picked using a sterilised tip, and selected single colonies were subsequently re-plated and further purified on TSA. Figure 4.4 shows examples of the spiral plate dilutions achieved, and in panel B jungle sample 2 is specifically shown. The isolated colonies from sample 1 had two different sizes (small and large) and representatives of each were picked (Figure 4.4B) and isolated from these plates. Thus, two different isolates were selected from

jungle sample number 1 (J1-1 and J1-2) relying on these differences in colony morphotype. The same approach was used for other samples, such as jungle sample number 50 (four isolate morphologies) and 23 (three isolate morphologies). Finally, a total of three growth passages of isolates on TSA were applied to purify a single strain, prior to the storage of pure isolates. In total, 123 pure isolates were produced from the 88 jungle samples that were *recA* gene positive by PCR.

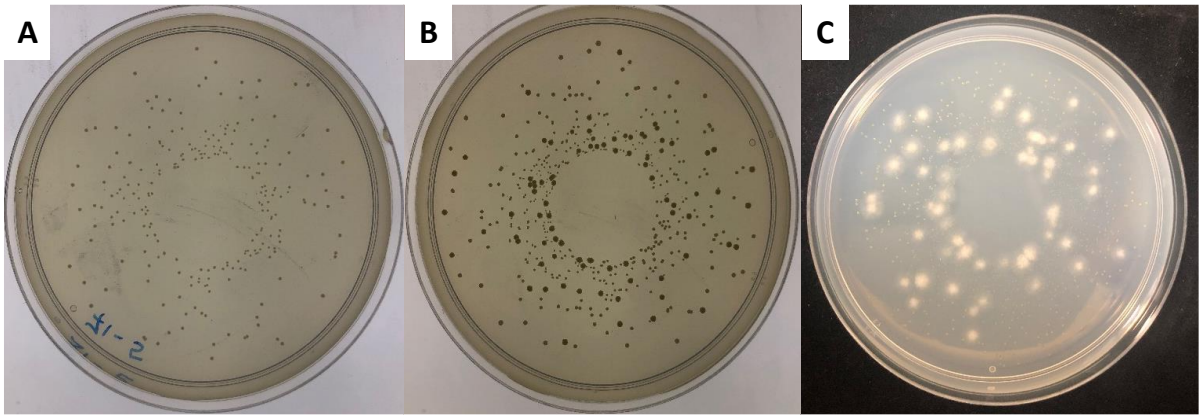


Figure 4.4 The morphotypes of bacterial colonies isolated on PCAT plates using the spiral plater. The plated jungle samples had the following types of morphology: (A) the purified single colonies all had the same morphotype; only a single isolate was selected from these plates. (B) The isolated colonies from jungle sample 1 had two different sizes (small and large) and representatives of each were picked and isolated from these plates. (C) This PCAT plate enriched fungal growth as well as small bacterial-like colonies; the purification of bacteria from these contaminated plates could not be achieved.

4.2.4 Molecular identification of purified isolates: 16S rRNA and *recA* gene-based identification

All 123 purified bacterial isolates were subjected to 16S rRNA and *recA* gene PCR followed by Sanger sequencing to provide preliminary species identification and potential *Burkholderia*/*Paraburkholderia* subclassification. The initial results for 16S rRNA gene sequencing from the bacterial specific 27F and 1492R PCR primers (Lane et al. 1985) demonstrated that the predicted 1500 bp amplicon could be amplified by PCR for all 123 isolates (Figure 4.5). Each successful 16S rRNA gene amplification was then subjected to DNA sequencing, and the resulting sequences were checked at NCBI using the nucleotide

basic local alignment search tool (BLAST) search tool. This preliminary molecular identification analysis showed that a total of 33 non-*Burkholderia* and 90 putative *Burkholderia*/*Paraburkholderia* isolates had been recovered (Table 4.3). Thus, 90 putative *Burkholderia*/*Paraburkholderia* isolates were identified and recovered from 54 jungle enrichment samples that had already been confirmed as *recA*-gene positive (Table 4.3).

This basic molecular identification using BLAST alignment searching at NCBI showed that certain non-*Burkholderia* species were also enriched on the PCAT media, including *Enterobacter sp.* (J87), *Klebsiella* sp. (J34), *Pantoea* sp. (J76), and *Leifsonia* sp. (J36). These results corroborated the results of Chapter 3 examining the performance of different growth media, and indicated that PCAT was a suitably selective media for *Burkholderia* and *Paraburkholderia* species. Overall, a few non-*Burkholderia* were enriched on PCAT media with the Gram-negative species including *Enterobacter* sp., *Klebsiella* sp., *Pantoae* sp., *Ochrobactrum* sp., and *Pandoraea* spp., and Gram-positive species including *Leifsonia* sp., *Bacillus* sp., *Curtobacterium* sp., and *Lactococcus* sp. (Table 4.4).

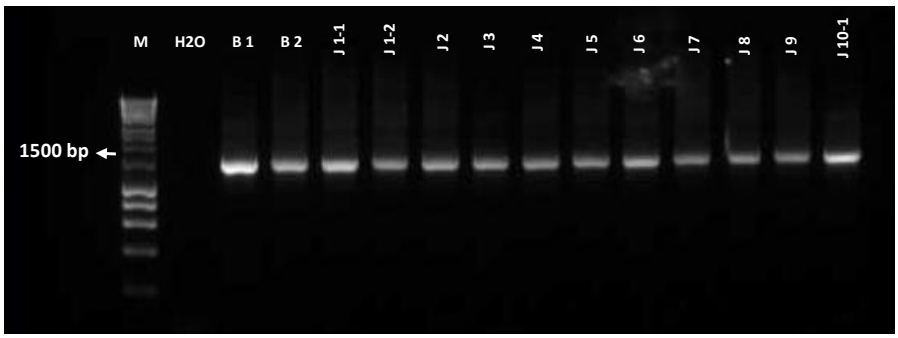


Figure 4.5 16S rRNA gene PCR of pure jungle isolates. The jungle isolate number is shown above each lane. Lane 1 is the 1 KB molecular marker and the amplicon size of 1500 bp for the 16S rRNA gene product is indicated. Lane 2 is the H_2O negative control. Lane 3 is B1 (*B. vietnamiensis* [BCC0042]) and Lane 4 is B2 (*B. graminis* [BCC0774]). Lanes 5 to 15 are purified jungle isolates produced by PCAT and purified in TSA—J1-1, J1-2, J2, J3, J4, J5, J6, J7, J8, J9, and J10-1.

To further resolve the identification of the purified isolates, the *recA* gene was amplified for the 90 putative *Burkholderia* and *Paraburkholderia* isolates. The *recA* gene (869-bp) was

successfully amplified from all 90 isolates and gel electrophoresis of the amplicon is shown for 13 jungle isolates in Figure 4.6 (J1-1, J1-2, J10-1, J10-2, J11-1, J11-2, J15-1, J15-2, J80-1, J80-2, J91-1, J91-2, and J91-3). However, not all sequences of *recA* genes in this section were evaluated, as the extraction of the full length *recA* gene sequences from selected isolates after whole genome sequencing was planned and performed in Chapter 5.

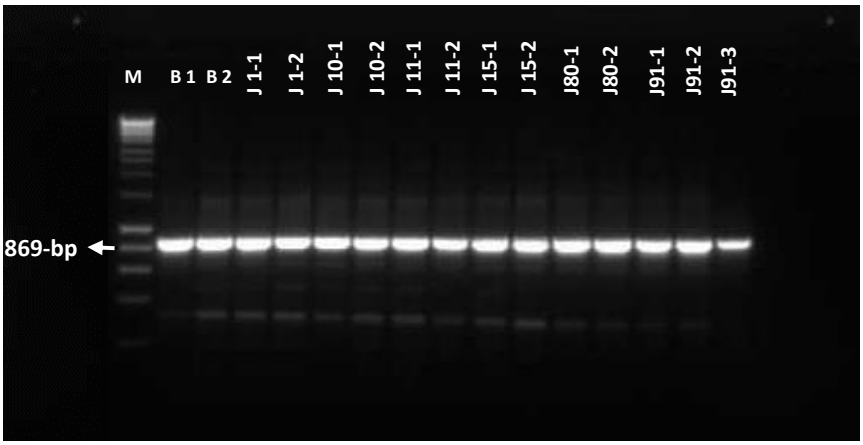


Figure 4.6 *recA* gene PCR confirmation for selected putative *Burkholderia* and *Paraburkholderia* isolates. Lane 1 is the 1 KB molecular marker. Lanes 2 and 3 are the B1 (*B. vietnamiensis* [BCC0042]) and B2 (*B. graminis* [BCC0774], which were positive controls. Lanes 4 to 16 are 13 jungle *Burkholderia/Paraburkholderia* isolates—J1-1, J1-2, J10-1, J10-2, J11-1, J11-2, J15-1, J15-2, J80-1, J80-2, J91-1, J91-2, and J91-3.

4.2.5 De-replicating duplicate strains from the jungle isolate collection

Since the initial screen to differentiate isolates was based on colony morphotypes (Figure 4.4), there was a high potential that duplicate isolates of the same strain type could be enriched and purified from a given jungle sample. Multiple colonies were isolated from individual jungle samples and to reduce the number of duplicate isolates from the set of 90 putative *Burkholderia*/*Paraburkholderia* strain genotyping by random amplified polymorphic DNA PCR (RAPD-PCR) fingerprinting was carried out as described (Mahenthiralingam et al. 1996). The genotyping analysis showed both strain diversity as well as strain conservation among the isolates (Figure 4.7). For example, 11 *recA*-positive isolates had been recovered from jungle

sample number 91. Evaluation of the RAPD-PCR fingerprints by eye filtered them down to 7 distinct strains. Furthermore, 6 isolates were purified from jungle sample number 8 and found to be *recA*-positive, but RAPD-PCR identified only 3 strains with visually distinct fingerprints within them (Figure 4.7). Overall, after combining the RAPD strain genotyping analysis with the species identification from 16S rRNA and *recA* genes, the 90 pure *Burkholderia/Paraburkholderia* isolates were reduced to 62 genetically unique strains in the jungle collection.

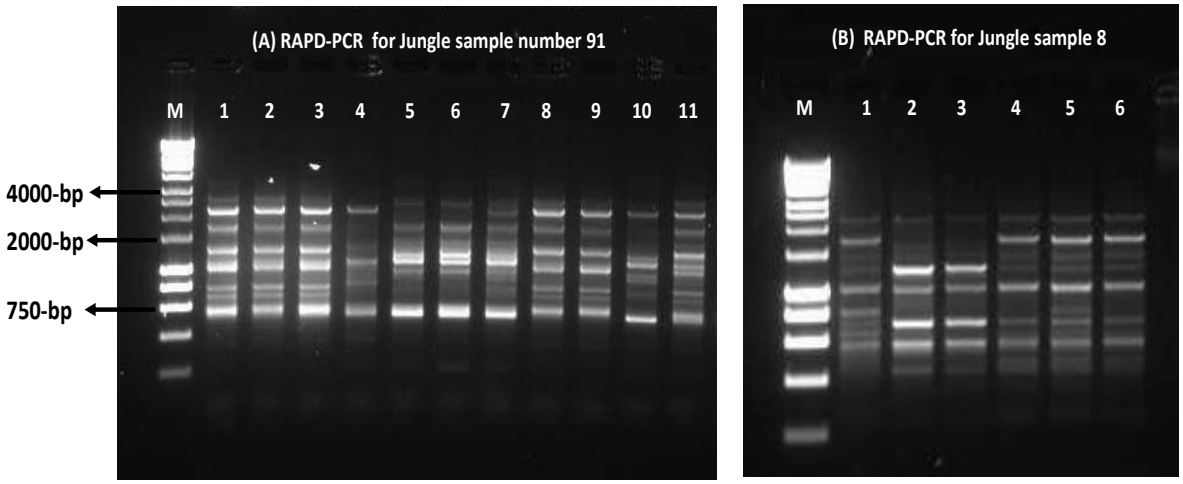


Figure 4.7 Random amplified polymorphic DNA (RAPD-PCR) fingerprinting of selected jungle isolates. (A) RAPD-PCR for jungle sample number 91 (J91). Lane 1 shows the molecular marker. Lanes 2 to 12 are 11 different isolates from sample J91. Conserved fingerprints suggesting the presence of single strains were observed for (see lanes): 1, 2, and 3; 5 and 6; and 8 and 9. Unique strains with distinct RAPD fingerprints from sample J91 are shown in lane 4, 7, 10 and 11. Overall, seven RAPD-distinct isolates were recovered from jungle sample number 91. (B) RAPD-PCR for jungle sample number 8 (J8). Lane 1 is a molecular marker and lanes 2 to 7 are six isolates from sample J8. There was one unique (lane 1), with isolates shown in lanes 2 and 3, and 4, 5, and 6, respectively representing a specific RAPD strain type. Thus, three diiferent strains were present in the isolates recovered from jungle sample 8.

Table 4.3 PCAT growth properties and *recA* gene screening for the collection of 98 jungle samples.

Jungle samples	Growth in PCAT broth	*recA*+/-	Jungle samples	Growth in PCAT broth	*recA*+/-	Jungle samples	Growth in PCAT broth	*recA*+/-
J1	+	+	**J34**	+	+	**J67**	+	+
J2	Slow	+	**J35**	+	+	**J68**	FC	-
J3	Slow	+	**J36**	+	+	**J69**	+	+
J4	FC	+	**J37**	FC	-	**J70**	+	+
J5	Slow	+	**J38**	Slow+FC	+	**J71**	SLOW	-
J6	+	+	**J39**	FC	-	**J72**	+	+
J7	+	+	**J40**	slow	-	**J73**	SLOW	-
J8	+	+	**J41**	+	+	**J74**	+	+
J9	Slow	+	**J42**	+	+	**J75**	+	+
J10	+	+	**J43**	FC	-	**J76**	+	+
J11	+	+	**J44**	Slow+FC	-	**J77**	-	-
J12	+	+	**J45**	+	+	**J78**	+	+
J13	Slow	+	**J46**	FC	-	**J79**	+	+
J14	Slow +FC	+	**J47**	+	+	**J80**	+	+
J15	+	+	**J48**	+	+	**J81**	SLOW+FC	-
J16	+	+	**J49**	+	+	**J82**	SLOW	-
J17	+	+	**J50**	+	+	**J83**	FC	-
J18	+	+	**J51**	SLOW	-	**J84**	-	-
J19	+	+	**J52**	FC	-	**J85**	+	+
J20	Slow +FC	-	**J53**	SLOW	-	**J86**	+	+
J21	Slow +FC	-	**J54**	SLOW	-	**J87**	-	-
J22	FC	-	**J55**	SLOW+FC	-	**J88**	+	+
J23	+	+	**J56**	FC	-	**J89**	SLOW	+
J24	+	+	**J57**	SLOW	-	**J90**	-	-
J25	-	-	**J58**	-	-	**J91**	+	+
J26	+	+	**J59**	-	-	**J92**	+	+
J27	+	+	**J60**	SLOW	-	**J93**	FC	-
J28	FC	-	**J61**	FC	-	**J94**	+	+
J29	FC	-	**J62**	SLOW	+	**J95**	+	+
J30	Slow +FC	+	**J63**	+	+	**J96**	-	-
J31	FC	-	**J64**	-	-	**J97**	+	+
J32	FC	-	**J65**	FC	-	**J98**	+	+
J33	Slow+FC	+	**J66**	-	-			

+: growth in PCAT media and amplification with *recA* PCR primer
-: no growth in PCAT broth and no amplification with recA PCR primer
FC: fungus conataminated samples

4.2.6 Phylogenetic analysis of 16S rRNA and *recA* genes of the jungle isolates

Phylogenetic analysis of partial 16S rRNA gene sequences was obtained for 62 jungle *Burkholderia*/*Paraburkholderia* isolates, and 33 additional purified isolates that turned out not to be members of these species (collectively designated non-*Burkholderia*). Table 4.5 presents the partial 16S rRNA gene sequence identification of these 95 isolates. The selection approach to the systemic collection assembly in this study was based on partial 16S rRNA gene sequences, and positive *recA* amplification of the 62 isolates deemed to be *Burkholderia*/*Paraburkholderia*. The initial identification was obtained using a BLASTn search from the NCBI and it showed that for the 62 isolates, 58% of isolates were *Paraburkholderia* and 42% were *Burkholderia*. The preliminary identification the analysis also showed that there

had been a false discovery rate of approximately 35%, with 33 of the 95 isolates in total falling into species groups outside of the *Burkholderia*/*Paraburkholderia* (Table 4.4).

Table 4.4 The partial 16S rRNA gene identification based on a BLASTn search for 62 putative jungle *Paraburkholderia*/*Burkholderia* and 33 non-*Burkholderia* species isolated

Putative *Paraburkholderia* isolates (*n*= 36)		Putative *Burkholderia* isolates (*n*=26)		Non-Burkholderia (*n*-=33)	
Original Name	**16S rRNA gene ID (BLASTn)**	**Original Name**	**16S rRNA gene ID (BLASTn)**	**Original Name**	**16S rRNA gene ID (BLASTn)**
J94	*P. tropica*	**J75-2**	*Burkholderia* sp.	**J3**	*Curtobacterium sp.*
J92	*P. tropica*	**J75-1**	*Burkholderia* sp.	**J5**	*Curtobacterium lsp.*
J88	*Paraburkholderia* sp.	**J69-1**	*Burkholderia* sp.	**J13**	*Curtobacterium sp.*
J7	*P. unamae*	**J67**	*Burkholderia* sp.	**J20**	*Klebsiella sp.*
J76-2	*Paraburkholderia* sp.	**J62**	*Burkholderia* sp.	**J21**	*Pantoea sp.*
J74	*P. metalliresistens*	**J50-4**	*Burkholderia* sp.	**J22**	*Curtobacterium sp.*
J72	*P. metalliresistens*	**J50-2**	*B. cepacia*	**J30**	*Curtobacterium sp.*
J6	*P. tropica*	**J45-1**	*B. cepacia*	**J33**	*Leifsonia sp.*
J69-2	*P. tropica*	**J41**	*Burkholderia* sp.	**J36**	*Curtobacterium sp.*
J63-2	*P. tropica*	**J11-2**	*Burkholderia* sp.	**J73**	*Curtobacterium sp.*
J50-3	*P. tropica*	**J23-1**	*Burkholderia* sp.	**J41-2**	*Curtobacterium sp.*
J50-1	*P. tropica*	**J17-4**	*B. vietnamiensis*	**J52**	*Curtobacterium sp.*
J45-2	*P. tropica*	**J17-1**	*Burkholderia cepacia*	**J53**	*Enterobacter sp.*
J42	*P. heleia*	**J48-2**	*Burkholderia* sp.	**J54**	*Enterobacter sp.*
J35	*P. tropica*	**J47-3**	*Burkholderia* sp.	**J56**	*Curtobacterium sp.*
J26	*P. tropica*	**J47-2**	*B. cepacia*	**J58**	*Curtobacterium sp.*
J24	*P. tropica*	**J70**	*Burkholderia* sp.	**J59**	*Klebsiella sp.*
J1-1	*P. tropica*	**J86-2**	*B. cenocepacia*	**J61**	Ochrobactrum sp.
J1-2	*P. tropica*	**J86-1**	*B. cepacia*	**J64**	*Bacillus* sp.
J16	*P. bannensis*	**J80-2**	*Burkholderia* sp.	**J65**	Enterobacter sp.
J15-2	*P. metalliresistens*	**J91-2**	*Burkholderia* sp.	**J66**	Lactococcus sp.
J15-1	*P. metalliresistens*	**J91-1**	*Burkholderia sp.*	**J89**	Lactococcus sp.
J11-1	*P. metalliresistens*	**J49**	*B. cepacia*	**J78-4**	*Pantoeasp.*
J10-2	*P. nodosa*	**J78-5**	*Burkholderia sp.*	**J78-2**	*Curtobacterium sp.*
J10-1	*P. unamae*	**J78-9**	*Burkholderia sp.*	**J85**	*Klebsiella sp.*
J8-2	*P. tropica*	**J78-10**	*Burkholderia sp.*	**J95**	Pantoeasp.
J27	*P. tropica*			**J38**	*Pandoraea sp.*
J23-3	*P. tropica*			**J87**	*Klebsiella* aerogenes
J23-2	*P. tropica*			**J8-8**	*Enterococcus* faecalis
J8-1	*P. tropica*			**J18**	*Curtobacterium* sp.
J19-1	*P. tropica*			**J36-1**	*Leifsonia* sp.
J19-2	*P. sacchari*			**J34-3**	*Klebsiella* sp.
J97	*P. heleia*			**J76-1**	*Pantoea* sp.
J12	*P. unamae*				
J98	*P. bannensis*				
J78-1	*P. tropica*				

A *recA*-gene-based phylogeny was obtained for nine isolates and demonstrated the presence of two major evolutionary branches (Figure 4.8). One branch was deep and comprised of nine jungle isolates (J72, J15-1, J15-2, J94, J26, J24, J97, and J12; Figure 4.8). These isolates were most closely related to *recA* gene sequences from reference isolates *P. aspalathi* (LMG 27731^T), *P. terricola* (LMG 20594^T), and *P. phenazinium* (GAS86). Thus, this preliminary analysis demonstrated the jungle isolates were likely *Paraburkholderia* species. The second

clade within the *recA* gene phylogeny suggested that jungle isolate J91-1 was a member of the Bcc since it grouped between reference Bcc species (*Burkholderia ubonensis* LMG 20358^T and *B. cepacia* ATCC17759; Figure 4.8).

Based on the 16S rRNA gene phylogenetic analysis, the tree was divided into three major evolutionary clades (Figure 4.9). The largest clade contained reference members of the *Paraburkholderia* genus and included 50 jungle isolates (Figure 4.9). Another clade included reference members of the *Burkholderia* genus with 12 jungle isolates also falling within this grouping (Figure 4.9). The final clade included the 33 non-*Burkholderia* species isolates that had been recovered as part of the enrichments. In the *Paraburkholderia* clade, the clustering of 52 jungle isolates showed that the jungle isolates were grouped into four sub-groups. The first group comprised 26 isolates placing adjacent to *Paraburkholderia tropica* (Figure 4.9). The other three groups did not cluster closely to *Paraburkholderia* reference species (Figure 4.9). Four isolates (J70, J76-2, J42-2, and J50-4) clustered together, and the closest neighbour to one of them (J42-2) was *Paraburkholderia heleia.* In the *Burkholderia* clade, the eleven jungle isolates were likely members of the Bcc closely related to *Burkholderia cepacia* (Figure 4.9).

The third clade confirmed that non-*Burkholderia* bacteria had been enriched by the PCAT growth media; these were clearly grouped separately from the *Burkholderia* and *Paraburkholderia* isolates and references. These isolates were removed from the final collection of isolates. It is important to mention that the 16S rRNA-based phylogeny shown at this point in the study was conducted based on 900 bp, as the full-length of 16S rRNA gene is 1529 bp. The full length of the 16S rRNA and *recA* gene phylogenies will be shown in the next chapter. In summary, the phylogenetic analysis of 16S rRNA and *recA* genes demonstrated the presence of two major clades: one that contained jungle isolates and reference species from *Paraburkholderia* and a second that contained jungle isolates grouped with *Burkholderia* species (Figures 4.8 and 4.9). Overall, a collection of 62 *Burkholderia* and *Paraburkholderia* was assembled and selected for genome sequencing. This high-resolution analysis would

enable further taxonomic analysis as well as genome mining analysis for antibiotic biosynthesis.

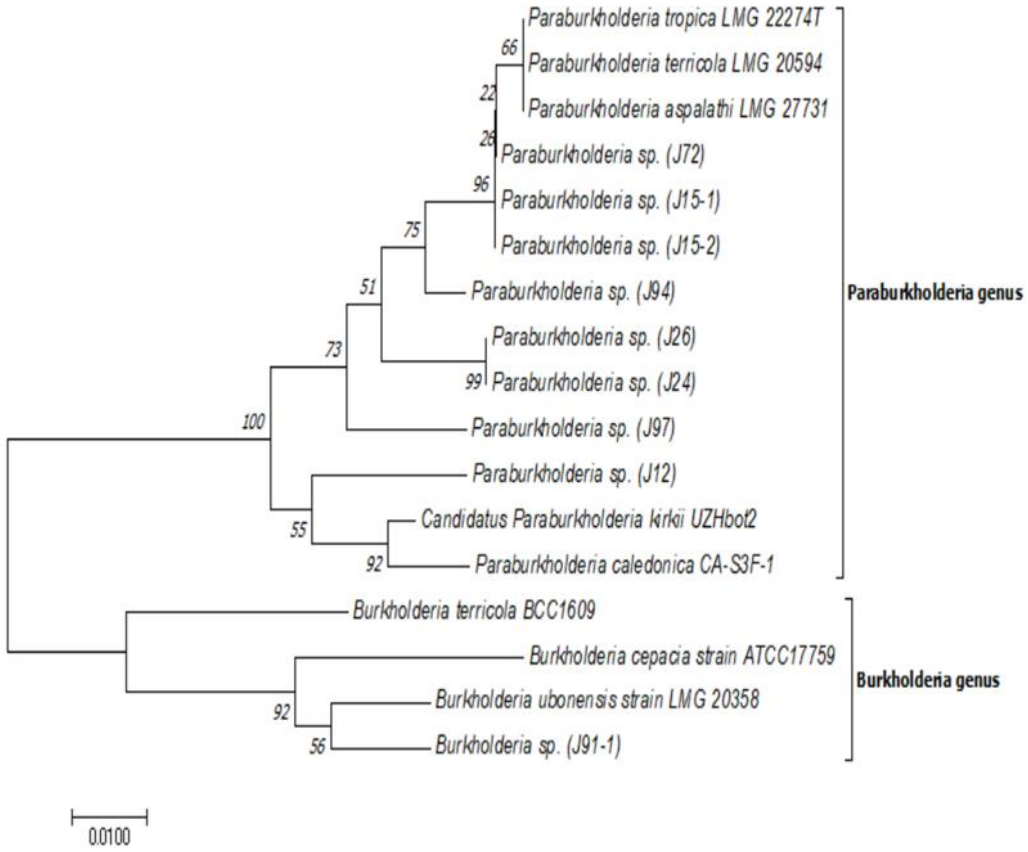


Figure 4.8 *recA* gene phylogenetic sequences analysis. Five *Paraburkholderia* and four *Burkholderia* references were aligned with nine jungle isolates. The alignment, phylogenetic, and molecular evolutionary analyses were conducted. The alignment and phylogenetic analyses were performed using MEGA.7 software. The alignment of *recA* genes was constructed using ClustalW2. Phylogenetic and molecular evolutionary analyses were conducted based on the neighbour-joining method, and the bootstrap values were 1000, as shown at the nodes. The *recA* gene analysis showed two main clades of jungle isolates. The first clade shows the *Paraburkholderia* isolates.

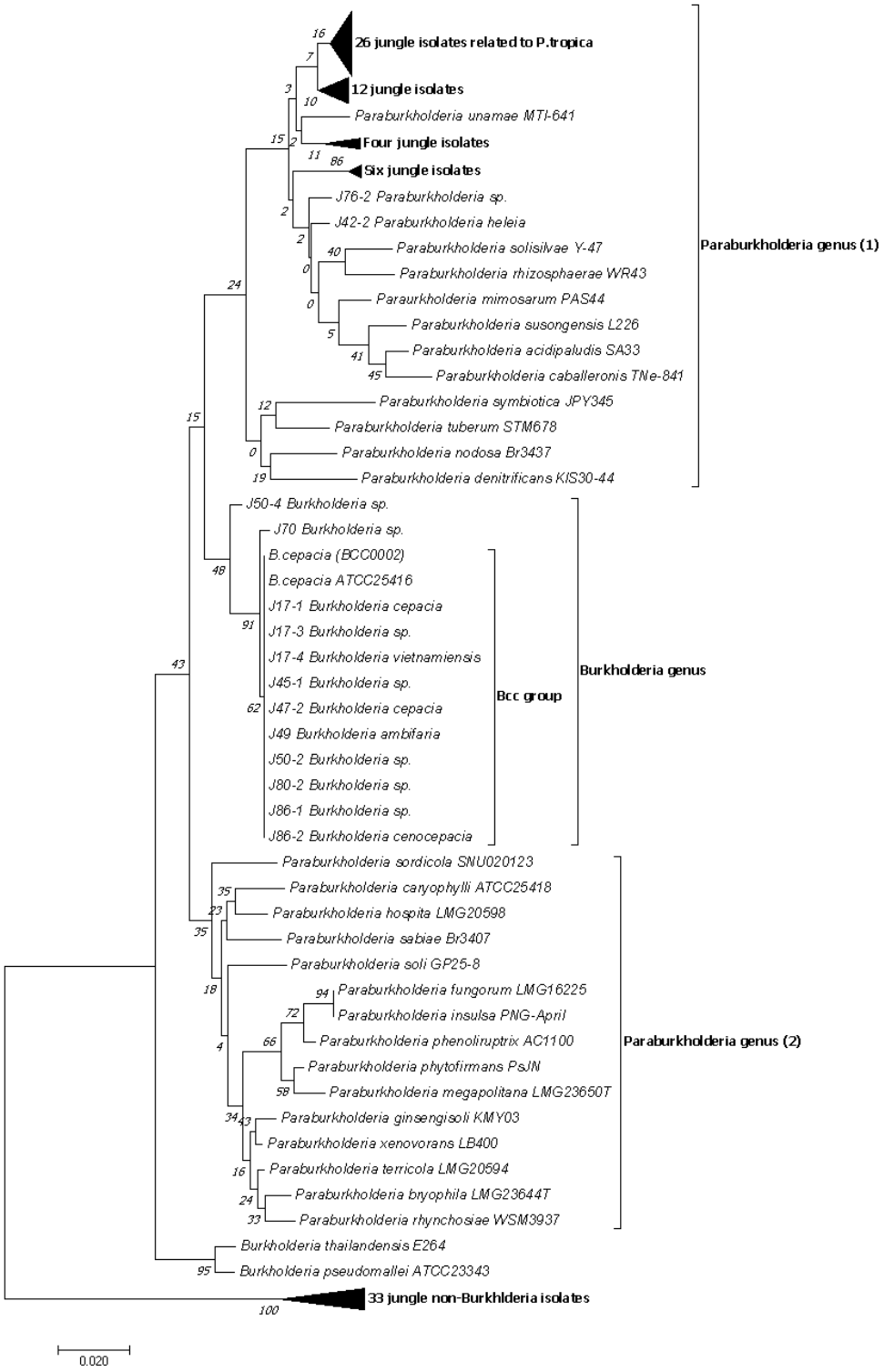


Figure 4.9 Phylogenetic analysis of 16S rRNA gene sequences from the entire collection of jungle *Burkholderia, Paraburkholderia*, and non-*Burkholderia* isolates. 32 16S rRNA sequences (references) were aligned with 95 jungle isolates (62 *Burkholderia* and *Paraburkholderia* and 33 non-*Burkholderia*). Clade 1 contained the greatest number of jungle isolates, all falling within the *Parabrkholderia* genus, with four distinct groups of isolates. Group 1 contains 26 isolates and having *P. tropica* as the closest related species, Group 2, 3 and 4 contain 12, 4 and 6 jungle isolates, subsequently. There were two isolates (J76-2 and J42-2) placed individually within *Paraburkholderia* genus. Clade 2 represent the *Burkholderia* genus. 12 jungle *Burkholderia* isolates fell within the *Burkholderia* genus and ten of them were showing to belong to Bcc group having *Burkholderia cepacia*. Clade 3 shows *Paraburkholderia* refrences. The final clade shows the 33 non-*Burkholderia* isolates. MEGA.7 software was used to construct the alignment and phylogenetic analyses. The alignments of the 16S rRBA genes were constructed using ClustalW2, and bootstrap values of 1000 were used (% similarity shown). The scale bar shows (0.020), which represents substitutions per site.

4.3 Discussion

Overall, the strategy to use a combination of culture-dependent and culture-independent analyses in parallel (Figure 4.2) was successful in allowing the assembly of a collection of *Burkholderia* and *Paraburkholderia* jungle isolates. The 16S rRNA and *recA* gene PCR applied to the culture enrichments and pure isolates showed that more than 50% of the Malaysian jungle samples were positive for *Burkholderia*/*Paraburkholderia*. To confirm their potential identity, 62 isolates were subjected to 16S rRNA and *recA* gene sequencing and phylogenetic analysis. Two key results were observed from this initial genetic screening as follows.

PCAT successfully enriched *Burkholderia* and *Pareburkholderia* in jungle samples. Previously, PCAT showed its selectivity in isolating *Burkholderia* from water and soil (Peeters et al. 2016). Peeters and colleagues found that PCAT was the optimal selective medium for Bcc species recovery from a range of river water and soil samples (Peeters et al. 2016). Their study compared three media to evaluate which was the best for isolating Bcc species from the environment. They found that PCAT yielded *Burkholderia* species more than other media, including BCEM (*B. cepacia* complex enrichment medium) and BCSA (*B. cepacia* selective agar) (Henry et al. 1997; Flanagan and Paull 1998). Non-*Burkholderia* species were observed during this study of jungle samples (e.g., Gram-negative species *Enterobacter* sp., *Klebsiella* sp., and *Pantoae* sp., and Gram-positive species *Leifsonia* sp., and *Curtobacterium* sp.,) (Peeters et al. 2016). This could be linked to the PCAT, as these species could ultilize tryptamine or the other components of the medium as a carbon or nitrogen sources, and hence grow along side *Paraburkholderia/Burkholderia*.

Furthermore, in this jungle study, the 16S rRNA and *recA* gene PCR successfully identified the recovered isolates and could easily distinguish *Paraburkholderia/Burkholderia* isolates from the non-*Burkholderia* isolates. The *recA* gene is a conserved housekeeping gene within all *Burkholderia*, and the PCR primers used in this study were developed to help identify *Burkholderia* sp. over fifteen years ago (Payne et al. 2005). By testing it on known control

strains (Figure 4.3), and subsequently applying it to the jungle isolate collection, it is clear that this historical *recA* gene PCR is also effective for the preliminary identification of *Paraburkholderia* species. Regardless of the strengths of using a single gene (16S rRNA and *recA* genes) to identify bacteria to the genus level, the resolution of using the single genes to differentiate the closely related species is limited. Within the current study I therefore proceeded to use the whole-genomes sequences (WGS) data to provid high-resolution species and strains ideintifications (see Chapter 5).

It is known that *Burkholderia* and *Paraburkholderia thrive* widely in multiple natural environments (Sawana et al. 2014; Depoorter et al. 2016). Several new species of Bcc, and a potentially diverse range of novel *Paraburkholderia* species, were isolated from Malaysian jungle samples, demonstrating they are prevalent at the rhizosphere of jungle plants and the soil. An abundance of Bcc species, such as *B. ambifaria*, *B. cepacia*, *B. stabilis*, *B. dolosa*, and *B. pyrrocinia,* were isolated from maize rhizosphere samples (Ramette et al. 2005). Few studies had previously examined *Burkholderia* isolates from tropical environments, especially Bcc. *Burkholderia pseudomallei* was isolated from Sarawak, Malaysia (Podin et al. 2014), and *Burkholderia paludis* is a Bcc species that was isolated from a tropical peat swamp forest soil in Malaysia (Ong et al. 2016). The Malaysian Borneo samples yielded a diversity of *Paraburkholderia* species, and this leads to questions of whether it was related to the environment, or the growth media and culture conditions used for their recovery. Further work will be needed to understand which was the most important parameter, but with 36 *Paraburkholderia* and 26 Bcc isolated from the jungle samples, it does appear the jungle environment in Sabah, Malaysia is a rich source for these bacteria.

4.4 Conclusion

A strategy was successfully developed and applied for screening the Malaysian jungle samples for the presence of *Burkholderia* and *Paraburkholderia* bacteria. This was successfully applied and produced a systemic collection of 62 isolates, of which 26 were

Burkholderia and 36 were *Paraburkholderia* when basic 16S rRNA gene and *recA* gene was applied. Further study was needed on the Borneo jungle sample isolates using genomic approacheds to identify known and potentially novel species (Chapter 5), and then a combination of genome mining and an antimicrobial activity analysis to look at their potential as antibiotic producers (Chapter 6). The use of single-gene analysis (16S rRNA gene or *recA* gene) has limitations in terms of accurate taxonomic analysis, although within this chapter it helped us identify the isolates to the genus level.

Chapter 5. Genomic identification of putative 57 Jungle *Burkholderia*/*Paraburkholderia* isolates

5.1 Introduction

Since the 1990s, scientists have established that *Burkholderia* sensu lato bacteria exhibit extensive diversity (Coenye and Vandamme 2003). Moreover, *Burkholderia* sensu lato bacteria have among the largest bacterial genomes, ranging from 6 to 9 MB. This group of bacteria has a unique genomes feature, consisting of at least 2 replicons, which is organized into two essential chromosomes and, for the *B. cepacia* complex contains one extra non-essential megaplasmid, as well as in some cases multiple other plasmids (Agnoli et al. 2012). This multireplicon genomic arrangement gives *Burkholderia* genomes considerable potential for high genomic plasticity and diversity. One of the consequences for this genomic diversity has been that the taxonomy and phylogeny of the *Burkholderia* sensu lato genera has had to be revised and reclassified several times over the past decades. This has been carried out through the analysis of phenotypic traits, and genotypic criteria such as 16S rRNA gene sequences, *recA* gene sequence, and multilocus sequence typing of housekeeping gene sequences, and recently, using whole genome sequences (Sawana et al. 2014; Dobritsa and Samadpour 2016; Lopes-Santos et al. 2017; Estrada-de los Santos et al. 2018). Early taxonomic studies of *Burkholderia* sensu lato were generally based on the guanine-cytosine (GC) content and the phylogenetic analysis of a single conserved gene, including the 16S rRNA, *recA, acdS*, *gyrB*, and *rpoB* gene sequences (Gyaneshwar et al. 2011).

There have been an increased number of genetically driven taxonomic studies of *Burkholderia* sensu lato within the past five years, especially in relation to the extensive genomic datasets that are now being generated. In 2014, the phylogenetic clustering of *Burkholderia* was further divided into two major genera: 1) the *Burkholderia* spp. was retained; and 2) *Paraburkholderia* gen. nov was proposed (Sawana et al. 2014). This result was based on both a genotypic

(phylogenetic tree of 21 conserved proteins and the 16S rRNA gene sequence) and phenotypic analysis (42 conserved sequence indels [CSIs]). It was found that six of these CSIs were specific for the first main clade and two for the second clade. *Burkholderia* in Clade 1 contained multiple animal and plant pathogenic species and retained the genus name because the original reference species, *Burkholderia cepacia,* is within this group. Clade 2 predominantly contained environmental species and was named '*Paraburkholderia'* based on the evolution of genomic approaches (Sawana et al. 2014; Dobritsa and Samadpour 2016). The *Caballeronia* genus was proposed in 2016 based on previous methods, DNA GC content, multiple phylogenetic analyses, CSI, and DNA-DNA hybridization values (Dobritsa and Samadpour 2016). In 2017, the genus *Robbsia* was also defined based on a combination of genotypic (16S rRNA gene phylogenetic tree, multilocus sequence analysis [MLSA], and average nucleotide identity [ANI]) and phenotypic (tetranucleotide signature frequency and percentage of conserved proteins [POCP]) analyses (Lopes-Santos et al. 2017). In 2018, two additional novel *Burkholderia* sensu lato genera were revealed: 1) *Mycetohabitans* gen. nov.; and 2) *Trinickia* gen. nov, which were based on a type of whole genome analysis - phylogenetic trees for the amino acid sequence of conserved genes and overall amino acid identity (Estrada-de los Santos et al. 2018).

To date, the original *Burkholderia* genus has been divided into at least six well-discriminated genera (*Paraburkholderia*, *Caballeronia*, *Robbsia*, *Mycetohabitans*, *Trinickia*, and the remaining *Burkholderia*), which are strongly supported by the taxonomic thresholds such as conserved gene phylogenies and average nucleotide identity. However, it is clear that there are remaining novel lineages and considerable genetic diversity within *Paraburkholderia* suggesting considerable potential for the definition of future novel species and genera (Parte 2013; Vandamme and Peeters 2014; Depoorter et al. 2016). The richness of the *Burkholderia* sensu lato genomic database indicates that the taxonomy of this bacteria has not yet been fully described and should be investigated using advanced genomic and bioinformatic tools.

5.1.1 Aim and objectives

In the present study, selective culture in combination with *recA* gene PCR enabled a highly diverse collection of environmental *Burkholderia* and *Paraburkholderia* to be assembled from the Bornean jungle. Moreover, there was substantial diversity within the *Burkholderia* and *Paraburkholderia* isolates recovered, which included potentially novel taxa, especially for the *Paraburkholderia*. Detailed phenotypic and genomic characterisation of this unique collection of environmental *Burkholderiaceae* strains was used to uncover their potential for antibiotic production (see next chapter 6). In the current chapter, the aim was to identify and fully characterise the taxonomic diversity within the environmental *Burkholderiaceae* collection based on whole genome sequencing (WGS) in combination withd multiple phylogenomic and bioinformatics approaches. The following objectives were addressed:

1. **16S rRNA and *recA* genes analysis**. Full length 16S rRNA and *recA* genes were extracted from the genomes and phylogenetic analysis used to classify them in relation to taxonomic type strain and reference genomes. 16S rRNA gene sequence-based phylogenies have been widely used as a rapid and accurate classification tool for *Burkholderia* sensu lato bacteria in several studies. However, since the 16S rRNA gene can only provide the resolution up to the genus level for several bacterial groups, it frequently does not have the accuracy to delineate species. Hence, comparative taxonomic analysis with the *recA* gene as a highly characterised *Burkholderia* sensu lato house-keeping gene was also performed. Collectively this combined approach of full length 16S rRNA and *recA* gene analysis was used to establish preliminary taxonomic relationships between the well-published species and the Borneo jungle isolates, identifying their closest neighbours and forming the basis for further genomic analyses.
2. **Application of web-based classification tools**. The automated multi-locus species tree (AutoMLST) tool (Alanjary et al. 2019), multilocus sequence typing (MLST) databases (Godoy et al. 2003) and National Centre for Biotechnology Information

(NCBI) was then used to extend the single gene phylogenetic analysis. This enabled further selection of the closest neighbours 'type strains' and genomic sequences that could be downloaded for subsequent analysis.

3. **Evaluation by genomic taxonomy methods.** The use of whole genome comparison methods, such as *in silico* DNA-DNA hybridization (DDH), and average nucleotide identity (ANI), were then applied for species delineation. ANI has been replaced with DDH in more recent classification studies, since it became used as a gold standard for defining bacterial species (Konstantinidis and Tiedje 2005; Meier-Kolthoff et al. 2013).
4. **Core gene phylogenies.** Once clear classification of isolates was established, core genes were extracted and alignments of these were performed to identify the isolates at both the species and strain level. This could only be carried out for more closely related groups within *Paraburkholderia* and *Burkholderia* sensu lato, because the genomic diversity across the entire family was extensive.

5.2 Results

5.2.1 Genome Sequences of novel *Burkholderia* and *Paraburkholderia* isolated from jungle Borneo

A set of *Paraburkholderia* (*n* = 44), *Burkholderia* (*n* = 12) and *Caballeronia* (*n* = 1) isolates were recovered from soil samples collected from a jungle in Borneo (see Chapter four) and successfully genome sequenced. Table 5.1 presents the assembly metrics for each draft genome as follows: the number of contigs, the largest contig (bp), total length (bp), GC (%), and N50. To assess the data quality of the 57 draft genomes, the statistical report for the assemblies was generated using the quality assessment tool for genome assemblies (QUAST) (Gurevich et al. 2013). The draft genome sequences assembled into the contigs ranged from 44 to 170. The N50 contig size ranged from 155,444 to 768,013 bp. The total genome length was estimated to range between ~6.84 Mb and ~9.43 Mb. The range of a G+C % content was between 62.66% and 67.08%. Kraken-based genomic identities for putative *Burkholderia* and *Paraburkholderia* isolates was also evaluated using this preliminary genomic analysis. Kraken is a fast and informative bioinformatics tool to identify the DNA sequences (Wood and Salzberg 2014). The Kraken data confirmed that all 57 environmental isolates belonged to the order *Burkholderiales*. This data confirmed the earlier partial 16S rRNA identification, which shows that they were a mix of *Burkholderia* and *Paraburkholderia* (Table 5.1). However, Kraken is not an accurate approach to identify the isolates at the genus or species level, and this collection was therefore identified by means of high resolution analyses such as ANI. Overall, these analyses demonstrated that the quality of the draft genomes was high and could be interrogated in subsequent analyses to provide detailed insights into this unique collection of *Burkholderiaceae* bacteria.

Table 5.1 The metrics, quality and partial 16S rRNA and Kraken identities of 57 draft genomes for jungle Borneo isolates

BCC#		Original Name	#contigs	Largest contig (bp)	Total length (Mbp)	GC (%)	N50	16S rRNA gene ID (BLASTn)	Kraken.1 ID (genus)
1.	**BCC1909**	J94	55	752,036	8,5	65	447,235	*P. tropica*	*Paraburkholderia/Burkholderia*
2.	**BCC1910**	J92	82	827,158	8,32	64.9	309,465	*P. tropica*	*Paraburkholderia /Burkholderia*
3.	**BCC1911**	J88	94	586,468	8,5	64.71	223,564	*Paraburkholderia* sp.	*Paraburkholderia*
4.	**BCC1912**	J7	97	696,542	9	63.93	279,902	*P. unamae*	*Paraburkholderia*
5.	**BCC1913**	J76-2	56	773,954	7,07	64.59	357,328	*Paraburkholderia* sp.	*Paraburkholderia*
6.	**BCC1914**	J75-2	44	1,487,362	8,74	63.71	768,013	*Burkholderia* sp.	*Paraburkholderia*
7.	**BCC1915**	J75-1	55	1,337,444	8,57	63.78	501,129	*Burkholderia* sp.	*Paraburkholderia*
8.	**BCC1916**	J74	77	1,110,779	8,83	63.53	374,821	*P. metalliresistens*	*Paraburkholderia*
9.	**BCC1917**	J72	101	628,205	8,72	63.71	166,145	*P. metalliresistens*	*Paraburkholderia*
10.	**BCC1918**	J6	73	941,985	7,96	65.05	265,916	*P. tropica*	*Paraburkholderia*
11.	**BCC1919**	J69-2	100	612,386	9,03	62.66	253,607	*P. tropica*	*Paraburkholderia*
12.	**BCC1920**	J69-1	95	693,893	9,03	62.66	367,675	*Burkholderia* sp.	*Paraburkholderia*
13.	**BCC1921**	J67	97	950,484	9,34	63.57	274,084	*Burkholderia* sp.	*Paraburkholderia*
14.	**BCC1922**	J63-2	170	513,543	8,74	65.07	159,965	*P. tropica*	*Paraburkholderia*
15.	**BCC1923**	J62	103	534,516	8,23	64.98	200,970	*Burkholderia* sp.	*Paraburkholderia*
16.	**BCC1924**	J50-4	103	707,577	8,31	64.93	280,271	*Burkholderia* sp.	*Paraburkholderia*
17.	**BCC1925**	J50-3	104	707,553	8,30	64.93	244,042	*P. tropica*	*Paraburkholderia*
18.	**BCC1926**	J50-2	122	486,077	8,28	64.94	187,754	*B. cepacia*	*Paraburkholderia*
19.	**BCC1927**	J50-1	123	497,694	8,29	64.94	223,578	*P. tropica*	*Paraburkholderia*
20.	**BCC1928**	J45-2	88	642,214	8,08	64.96	244,908	*P. tropica*	*Paraburkholderia*
21.	**BCC1929**	J45-1	60	676,080	7,93	65.13	347,688	*B. cepacia*	*Paraburkholderia*
22.	**BCC1930**	J42	105	554,426	7,77	64.58	255,736	*P. heleia*	*Paraburkholderia*
23.	**BCC1931**	J41	101	426,007	7,01	65.14	187,711	*Burkholderia* sp.	*Paraburkholderia*
24.	**BCC1932**	J35	62	741,103	8,61	63.75	343,593	*P. tropica*	*Paraburkholderia*
25.	**BCC1933**	J26	77	602,288	8,36	64.85	334,594	*P. tropica*	*Paraburkholderia*
26.	**BCC1934**	J24	78	602,252	8,36	64.85	287,287	*P. tropica*	*Paraburkholderia*
27.	**BCC1935**	J1-1	81	726,501	8,55	64.77	316,706	*P. tropica*	*Paraburkholderia*
28.	**BCC1936**	J1-2	73	726,457	8,56	64.77	316,900	*P. tropica*	*Paraburkholderia*
29.	**BCC1937**	J16	77	461,275	7,96	65.13	266,003	*P. bannensis*	*Paraburkholderia*
30.	**BCC1938**	J15-2	51	1,353,767	8,50	63.88	641,780	*P. metalliresistens*	*Paraburkholderia*
31.	**BCC1939**	15-1	78	779,750	8,89	63.6	435,228	*P. metalliresistens*	*Paraburkholderia*
32.	**BCC1940**	J11-2	109	505,552	9,16	63.86	236,207	*Burkholderia* sp.	*Paraburkholderia*
33.	**BCC1941**	J11-1	53	1,572,796	8,83	63.63	540,172	*P. metalliresistens*	*Paraburkholderia*
34.	**BCC1942**	J10-2	150	975,028	9,27	63.62	308,842	*P. nodosa*	*Paraburkholderia*
35.	**BCC1954**	J10-1	71	1,557,849	9,11	63.85	448,252	*P. unamae*	*Paraburkholderia*
36.	**BCC1943**	J8-2	118	1,214,788	9,43	63.73	272,312	*P. tropica*	*Burkholderia*
37.	**BCC1945**	J27	105	693,037	8,51	64.87	258,721	*P. tropica*	*Burkholderia*
38.	**BCC1946**	J23-3	107	693,037	8,51	64.87	258,721	*P. tropica*	*Burkholderia*
39.	**BCC1947**	J23-2	103	693,037	8,52	64.87	258,721	*P. tropica*	*Burkholderia*
40.	**BCC1948**	J23-1	115	526,046	8,53	64.85	258,721	*Burkholderia* sp.	*Burkholderia*
41.	**BCC1949**	J8-1	76	776,238	8,24	64.92	258,927	*P. tropica*	*Paraburkholderia*
42.	**BCC1950**	J19-1	92	668,051	8,46	64.87	310,799	*P. tropica*	*Burkholderia*

43.	**BCC1951**	J19-2	91	668,051	8,46	64.87	310,799	*P. sacchari*	*Burkholderia*
44.	**BCC1952**	J97	98	951,256	7,78	63.27	199,475	*P. heleia*	*Burkholderia*
45.	**BCC1953**	J12	50	1,052,774	6,92	64.95	344,856	*P. unamae*	*Burkholderia*
46.	**BCC1955**	J17-4	83	514,066	6,86	67.07	200,345	*B. vietnamiensis*	*Burkholderia*
47.	**BCC1956**	J17-1	114	284,604	6,84	67.08	155,444	*Burkholderia cepacia*	*Burkholderia*
48.	**BCC1957**	J48-2	74	1,119,754	7,34	66.33	223,626	*Burkholderia* sp.	*Burkholderia*
49.	**BCC1958**	J47-3	97	926,403	8,51	66.8	222,912	*Burkholderia* sp.	*Burkholderia*
50.	**BCC1959**	J47-2	99	461,209	8,26	66.9	201,276	*B. cepacia*	*Burkholderia*
51.	**BCC1960**	J70	204	550,763	8,67	66.57	181,965	*Burkholderia* sp.	*Burkholderia*
52.	**BCC1961**	J86-2	120	468,273	8,82	66.37	212,588	*B. cenocepacia*	*Burkholderia*
53.	**BCC1962**	J86-1	106	550,927	8,82	66.37	229,690	*B. cepacia*	*Burkholderia*
54.	**BCC1963**	J80-2	106	550,928	8,85	66.31	239,734	*Burkholderia* sp.	*Burkholderia*
55.	**BCC1964**	J91-2	94	963,054	7,17	66.5	217,336	*Burkholderia* sp.	*Burkholderia*
56.	**BCC1965**	J91-1	93	963,054	7,17	66.5	217,336	*Burkholderia sp.*	*Burkholderia*
57.	**BCC1967**	J49	89	1,010,429	8,51	66.33	223,626	*B. cepacia*	*Burkholderia*

5.2.2 Non-*Burkholderia* and -*Paraburkholderia* within Borneo collection genomes

The analysis also uncovered five *Enterobacter* spp. and four *Klebsiella* spp. that had enriched with the *Burkholderia* and *Paraburkholderia* strains during the culture screening. Initial 16S rRNA gene sequence analysis showed that they were non-*Burkholderia* and *Paraburkholderia* bacteria, and hence they were not followed up in detail. However, because the classification of *Enterobacter* and *Klebsiella* is also difficult, genomic analysis was carried out to confirm exactly which genera were able to come through on the PCAT selective media and show positivity for the *recA* gene PCR analysis. Table 5.2 shows that the metrics of draft genomes analysis for these non-*Burkholderia* isolates confirmed to be *Enterobacter* and *Klebsiella* species. These strains were all placed with the general "Esh" laboratory collection as examples of these genera that had been isolated from the natural environment.

Table 5.2 Metrics, quality and Kraken identity of nine draft genomes of *Enterobacter* and *Klebsiella* species isolates from jungle Borneo

Strain number	Name of original sample	#contigs	Largest contig (bp)	Total length (Mbp)	GC (%)	N50	Kraken ID
Esh 1097	N1	33	644,249	4,76	53.88	364,555	*Enterobacter*
Esh 1098	N5	32	644,249	4,76	53.88	364,555	*Enterobacter*
Esh 1099	N9	27	1,214,140	4,76	53.89	364,555	*Enterobacter*
Esh 1100	N10	28	1,214,140	4,76	53.88	364,555	*Enterobacter*
Esh 1101	N15	27	828,347	4,76	53.88	364,555	*Enterobacter*
Esh 1102	N24	47	818,488	5,66	55.21	298,110	*Klebsiella*
Esh 1103	N26	56	657,019	5,66	55.21	303,967	*Klebsiella*
Esh 1104	N27	47	818,539	5,66	55.21	298,059	*Klebsiella*
Esh 1105	N34	54	818,421	5,66	55.22	298,110	*Klebsiella*

5.2.3 The automated multi-locus species tree (autoMLST)

AutoMLST was recently launched as an online web-based bacterial genome analysis and classification tool (Alanjary et al. 2019). Initial characterization of 75 draft genomes of *Burkholderia* and *Paraburkholderia* was performed using this tool. The autoMLST phylogenies

were useful to select the closest type strain neighbors, for further 16S rRNA- and *recA*-gene based phylogenetic analysis. The analysis showed the taxonomic positions of the study isolates among other related *Burkholderia*, *Paraburkholderia* and *Caballeronia* species/ strains on the databases (data not shown).

5.2.4 *Burkholderia* classifications based on genomic data

5.2.4.1 16S rRNA- and *recA*-gene based phylogenetic analysis for twelve environmental *Burkholderia*

Phylogenetic analyses of complete 16S rRNA and *recA* gene sequences showed that twelve isolates clustered within the *Burkholderia*, forty-four isolates clustered within the genus *Paraburkholderia* and one isolate clustered with the *Caballeronia* genus. Thus, the presentation of classification results was divided into three sections, the first explaining the *Burkholderia* diversity, the second part covering the detailed analysis of *Paraburkholderia*, and the third part covering the analysis of *Caballeronia*.

A phylogenetic tree was constructed with the full length 16S rRNA gene sequences from the twelve draft *Burkholderia* genomes (Figure 5.1). Phylogenetic analysis demonstrated that these twelve environmental *Burkholderia* spp. cladded into four main groups based on the closest neighbouring type strains. The first group (Group 1; Figure 5.1) contained the following three isolates: BCC1963 (J80), BCC1962 (J86-1), and BCC1961 (J86-2), which were closely related to the type strain of *B. territorii* CCUG 65687^T and *B. cepacia* ATCC 25416^T. The second group was a new group of species closely related to *B. lata* 383^T and contained four jungle *Burkholderia* isolates: BCC1959 (J47-2), BCC1958 (J47-3), BCC1967 (J49) and BCC1960 (J70) (Group 2; Figure 5.2). Two jungle *Burkholderia* isolates: BCC1955 (J17-4) and BCC1956 (J17-1), were closely related to *B. vietnamiensis* LMG 10929^T, presenting as the third group (Group3; Figure 5.1). The fourth group contained three *Burkholderia* jungle

isolates: BCC1957 (J48), BCC1964 (J91-1), and BCC1964 (J91-2), which placed near to the type strain for *B. diffusa* CCUG 54557^{T} (Group 4; Figure 5.1).

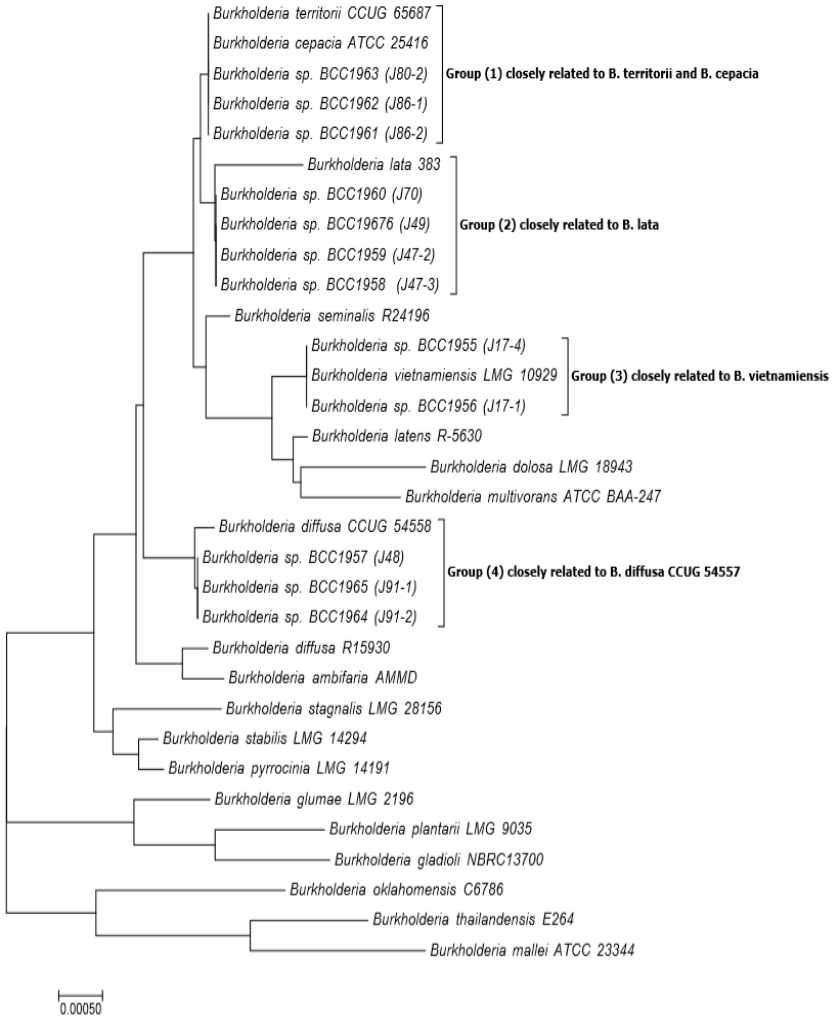


Figure 5.1 Neighbour-joining phylogenetic tree based on full length 16S rRNA gene sequence for environmental *Burkholderia*. The phylogeny based on an alignment of 1529 bp for twelve environmental *Burkholderia* and other type strain refrences of *Burkholderia*. The type strain reference sequences were downloaded from the NCBI. 16S rRNA gene sequences of twelve environmental *Burkholderia* isolates and other *Burkholderia* species were aligned in MEGA using ClustalW. The tree shows the phylogenetic positions of twelve jungle Burkholderia isolates within *Burkholderia*. Phylogenetic analysis was conducted using MEGA 7, and the bootstrap values based on 1000 replications (Kumar et al. 2016). The scale bar shows (0.0005), which represents substitutions per site.

The *recA* gene-based phylogeny demonstrated similar overall patterns to the 16S rRNA gene analysis for the closest known neighbouring type strains to the twelve *Burkholderia* isolates (Figure 5.2). The main difference in the *recA* gene tree was that the first and second groupings for the 16S rRNA gene analysis combined into one group of seven isolates. They formed a

well-supported group of potential *B. cepacia* strains, encompassing the type strain for *B. cepacia* (Figure 5.2). The second *recA* group clustered with *B. vietnamiensis* LMG 10929^{T}, and it again contained two jungle *Burkholderia* isolates: BCC1955 (J17-4) and BCC1956 (J17-1). Similarly, to the 16S rRNA gene analysis, BCC1957 (J48), BCC1964 (J91-1), and BCC1964 (J91-2), grouped together in a separate cluster from all other *Burkholderia* species, and via *recA* analysis they were also clearly distinct from the *B. diffusa* type strain. This suggested the three jungle isolates were a novel taxonomic clade within the Bcc group (see Group 3, novel Bcc; Figure 5.2).

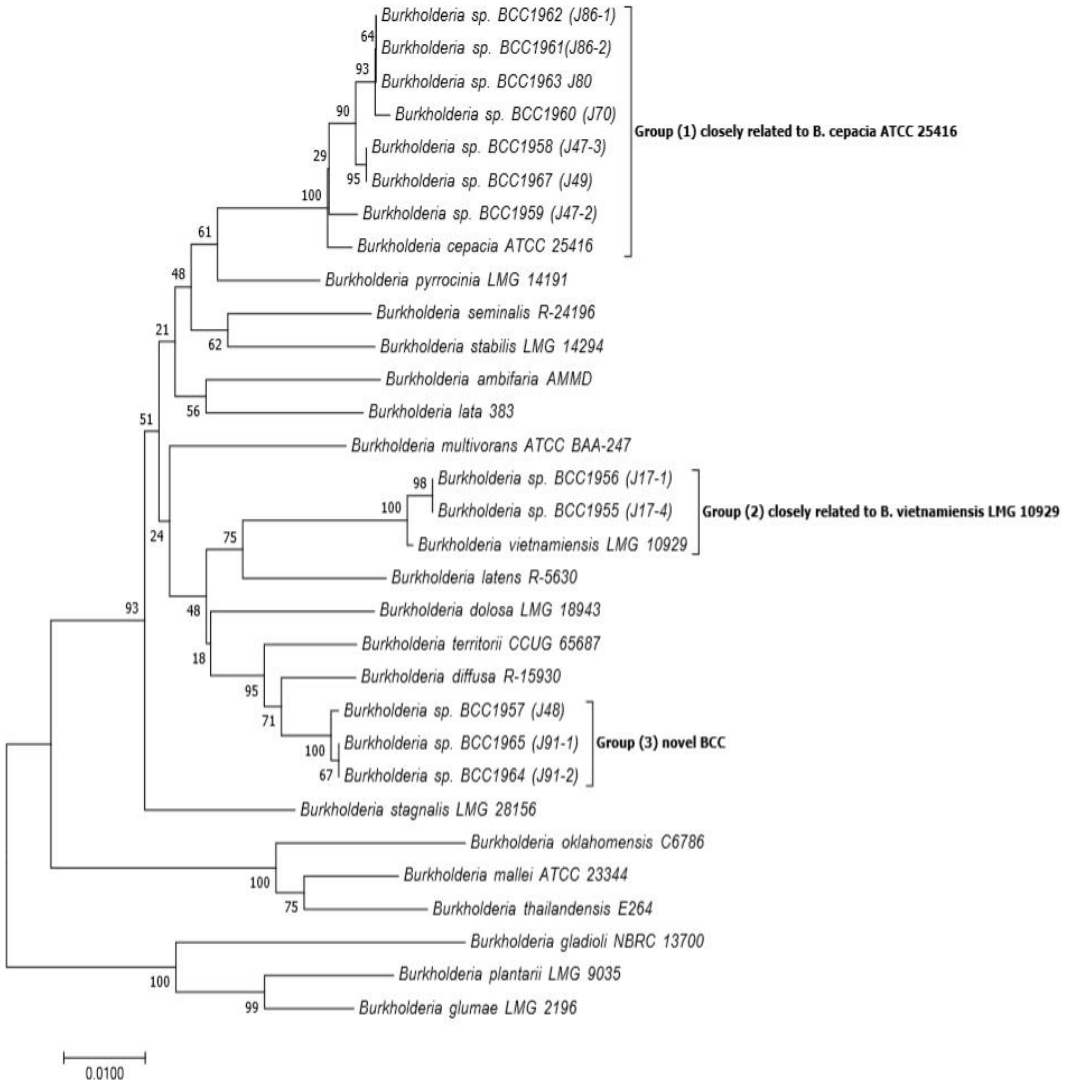


Figure 5.2 Neighbour-joining phylogenetic tree based on full length *recA* gene sequence for environmental *Burkholderia*. The phylogeny based on 1083 bp for twelve environmental *Burkholderia* and *Burkholderia cepacia* complex (Bcc) members. The type strains references were downloaded from the NCBI. *recA* gene sequences of twelve environmental *Burkholderia* isolates and other Bcc species were aligned in MEGA using ClustalW. The tree is showing the positions of twelve *Burkholderia* isolates and Bcc references. Phylogenetic analysis was conducted using MEGA 7, and the bootstrap values based on 1000 replications (Kumar et al. 2016). The scale bar shows (0.0100), which represents substitutions per site.

5.2.4.2 MLST analyses scheme

The same twelve draft genomes were also evaluated via the *Burkholderia cepacia* complex MLST database (https://pubmlst.org/bcc/). The seven MLST housekeeping genes alleles were automatically predicted from the genomes as follows: ATP synthase beta chain (*atpD*), glutamate synthase large subunit (*gltB*), DNA gyrase subunit B (*gyrB*), recombinase A (*recA*), GTP binding protein (*lepA*), acetoacetyl-CoA reductase (*phaC*) and tryptophan synthase subunit B (*trpB*). The predicted alleles from the PubMLST scheme is presented in Table 5.3. Interestingly, none of the twelve *Burkholderia* isolates matched a known sequence type (ST) within the database, showing that all were novel environmental Bcc strains with no linkages to strain type that had also been found in infection.

Table 5.3 MLST results of 12 environmental Bcc isolate show they are novel strain types

Isolates ID	*atpD*	*gltB*	*gyrB*	*recA*	*lepA*	*phaC*	*trpB*	ST
***Burkholderia* sp. BCC1959 (J47-2)**	6	~182	~613	~5	34	101	~418	-
***Burkholderia* sp. BCC1958 (J47-3)**	~2	~631	543	~183	~187	~101	3	-
***Burkholderia* sp. BCC1967 (J49)**	~2	~631	543	~183	~187	~101	3	-
***Burkholderia* sp. BCC1960 (J70)**	6	~635	993	373	~591	~325	3	-
***Burkholderia* sp. BCC1963 (80)**	102	~631	~543	114	~365	4	3	-
***Burkholderia* sp. BCC1962 (J86-1)**	102	~631	~543	114	~365	4	3	-
***Burkholderia* sp. BCC1961 (J86-2)**	102	~631	~543	114	~365	4	3	-
***Burkholderia* sp. BCC1956 (J17-1)**	27	20	~15	96	36	11	81	-
Burkholderia sp. BCC1955 (J17-4)	27	20	~15	96	36	11	81	-
***Burkholderia* sp. BCC1957 (J48)**	~459	~666	~687	~407	~465	~349	~603	-
Burkholderia sp. BCC1965 (J91-1)	~410	~459	~369	407	~465	~349	~603	-
***Burkholderia* sp. BCC1964 (J91-2)**	~410	~459	~369	407	~465	~349	~603	-

5.2.4.3 Whole-genome comparison of *Burkholderia*: *in silico* DDH and ANI

DDH was performed for the twelve environmental *Burkholderia* isolates by comparing them with the nearest neighbouring type strains based on 16S rRNA and *recA* gene phylogenetic analyses (Table 5.4). For the *recA* Group 1 strains (Figure 5.2), analysis by *in silico* DDH produced values ranging from 86.90% to 88.10% for all seven *B. cepacia*-like isolates when compared with type strain of *B. cepacia* (Table 4.5). The other strains when compared with *B. cepacia* ATCC 25416^{T} fell below the 70% species threshold, including the types strains for

B. vietnamiensis LMG 10929 and *B. diffusa* CCUG 54558^{T} as would be expected. The DNA–DNA relatedness values between two isolates BCC1956 (J17-1) and BCC1955 (J17-4) (Group 2) with the type strains of *B. vietnamiensis* LMG 10929^{T} were 90.90% and 90.80%, respectively. The DNA–DNA hybridization (DDH) values of the third *recA* group of related isolates *Burkholderia* sp. BCC1957 (J48), *Burkholderia sp.* BCC1964 (J91-1), and *Burkholderia sp.* BCC1964 (J91-2) ranged from 38.80% to 59.20% when compared with the selected type strains of *B. cepacia* ATCC 25416^{T}, *B. vietnamiensis* LMG 10929^{T}, and *B. diffusa* CCUG 54558^{T} (Table 5.4), as well as all other known Bcc genomes (data not shown). This DDH analysis supported that the *recA* gene group (Figure 5.2) isolates were most likely novel members of the *B. cepacia* complex.

Table 5.4 Pairwise DDH hybridization between the environmental *Burkholderia* novel species and strains.

Isolates genomes ID	*B. cepacia* ATCC 25416 ^{T}	*B. vietnamiensis* LMG 10929 ^{T}	*B. diffusa* CCUG 54558 ^{T}
***Burkholderia* sp. BCC1959 (J47-2)**	88.10*	35.50%	38.40%
***Burkholderia* sp. BCC1958 (J47-3)**	86.90%*	35.50%	38.30%
***Burkholderia* sp. BCC1967 (J49)**	86.90%*	35.50%	38.30%
***Burkholderia* sp. BCC1960 (J70)**	87.50%*	35.50%	38.40%
***Burkholderia* sp. BCC1962 (J86-1)**	88.10%*	35.50%	38.40%
***Burkholderia* sp. BCC1961 (J86-2)**	88.10%*	35.50%	38.40%
***Burkholderia* sp. BCC1963 (80)**	88.10*	35.50%	38.40%
***Burkholderia* sp. BCC1956 (J17-1)**	35.60%	90.90%*	38.10%
Burkholderia sp. BCC1955 (J17-4)	35.60%	90.80%*	38.10%
***Burkholderia* sp. BCC1957 (J48)**	39.50%	38.80%	59.10%
Burkholderia sp. BCC1965 (J91-1)	39.50%	38.80%	59.20%
***Burkholderia* sp. BCC1964 (J91-2)**	39.50%	38.80%	59.20%

*** DDH similarity above the 70% species threshold is shaded in green**

The whole genome comparisons based on the average nucleotide identity (ANI) analysis were carried out using PyANI (https://github.com/widdowquinn/pyani). Comparisons were evaluated for the twelve environmental *Burkholderia* spp. based on the nearest type strains selected using the *recA*-based phylogeny (Figure 5.2). Additionally, fastANI analysis was also carried out for these twelve environmental *Burkholderia* isolates by comparing them to a dataset of 7480 assembled *Burkholderia* genomes downloaded from the databases (genomic dataset assembled by Alex Mullins, Cardiff University). Figure 5.3 and Table 5.5 present the collective ANI analysis for the three *recA*-based groups of jungle Bcc. The ANI analysis clearly

demonstrated that the jungle isolates placed them within *B. cepacia*, *B. vietnamiensis*, and novel Bcc species (Figure 5.3).

The ANI values for the seven putative *B. cepacia* species isolates ranged from 98.63% to 99.99% when compared to one another (Table 5.5). The ANI values between the seven isolates and the type strain of *B. cepacia* was also well above the 95% species threshold, ranging from 98.40% to 98.57% (Table 5.5). Overall, collectively ANI analysis confirmed that the closest relative type strain to the seven jungle isolates was *B. cepacia* ATCC 25416^T, with ANI values collectively ≥ 98%, unequivocally defining them as this species of Bcc. For the two jungle isolates that were nearest to *B. vietnamiensis* LMG 10929^T in the prior analysis, they shared ANI values with each other (99.98%), and were also highly related to the type strain *B. vietnamiensis* LMG 10929^T (ANI 98.93% and 98.93%, respectively) (Table 5.5). The third *recA*-defined group (Figure 5.2) presented ANI results supportive of the fact they were likely new species within the Bcc. The collective ANI analysis for the three isolates were not higher than 95% to any Bcc type strains within the larger assembled *Burkholderia* genomes database (Table 5.5). Their closest *Burkholderia* relative was *B. diffusa* strains with 94% ANI value. Thus, these three novel jungle isolates BCC1957 (J48), BCC1964 (J91-1) and BCC1964 (J91-2) were closely related to each other (ANI values between them ranging from 98.79% to 99.99%), were not within any currently named *Burkholderia* species, and represented novel taxa (Figure 5.3, Table 5.5).

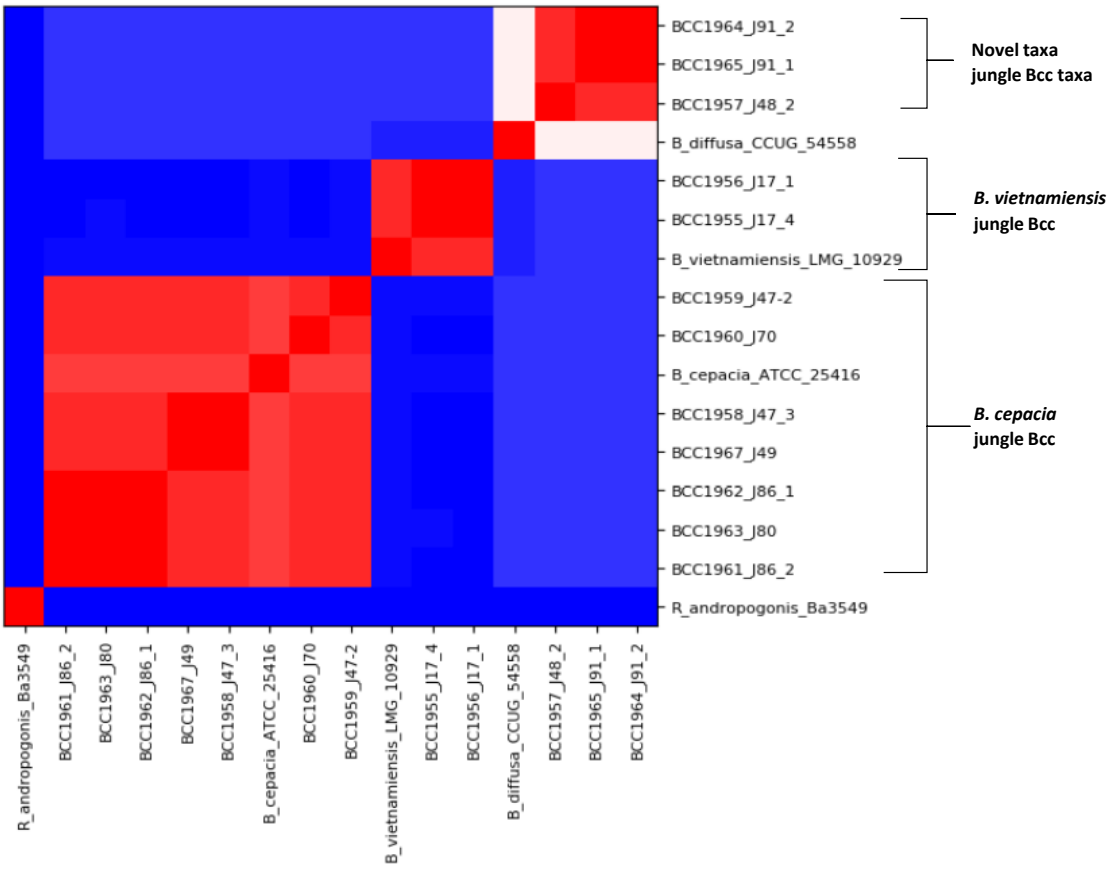


Figure 5.3 The whole-genome comparisons of 20 Bcc species/strains based on the ANI analysis. The twelve *Burkholderia* strains from this study were compared to three closest related type strains based on *recA* gene phylogenetic analysis. The ANI analysis was performed by means of Python module (Pyani). The analysis clustered the jungle isolates into three main species groups. The first cluster (lower left red box) shows the seven jungle isolates most closely related to the type strain of *B. cepacia*. The second cluster (central red box) grouped the closely related jungle strains (BCC1956 and BCC1955) with the type strain of *B. vietnamensis* and two other well characterised genomes of this species. The third ANI cluster consisted of three closely related jungle isolates comprising a novel species within Bcc group. They were not related to published type strains based on the ANI analysis. The colour key a scale for percentage ANI identity is shown (top left). The strains with ANI values >90 are shown within blue, and the strains with ANI and >95% are shown in red squares.

Table 5.5 FastANI values for twelve environmental Burkholderia isolates and closest neighbour type strains

Bcc genomes ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	*15	16
1. BCC1959 (J47-2)	-	98.81	98.79	98.77	98.71	98.70	98.71	88.99	89.01	90.71	90.77	90.74	98.57	89.09	~90	90.55
2. BCC1958 (J47-3)	98.81	-	99.99	98.65	98.74	98.72	98.72	88.89	88.86	90.71	90.76	90.74	98.43	89.02	~90	90.56
3. BCC1967 (J49)	98.80	99.99	-	98.61	98.72	98.69	98.71	88.89	88.89	90.75	90.74	90.71	98.40	88.99	90.12 - 90.68	90.45
4. BCC1960 (J70)	98.77	98.65	98.64	-	98.69	98.68	98.70	89.03	88.93	90.77	90.80	90.76	98.42	90.42	90.25 - 90.41	90.62
5. BCC1962 (J86-1)	98.72	98.73	98.72	98.65	-	99.99	99.99	89.05	89.07	90.79	90.83	90.81	98.53	89.14	90.22 - 90.44	90.52
6. BCC1961 (J86-2)	98.70	98.68	98.70	98.65	99.99	-	99.99	89.02	89	90.79	90.86	90.79	98.46	89.05	90.32 - 90.73	90.51
7. BCC1963 (80)	98.69	98.66	98.66	98.63	99.99	99.99	-	89.07	89.10	90.82	90.88	90.82	98.51	89.11	90.29 - 90.44	90.55
8. BCC1956 (J17-1)	89.11	89.03	88.99	89.07	89.02	89.12	89.12	-	99.99	90.44	90.42	90.43	89.11	98.99	90.22 - 90.33	90.39
9. BCC1956 (J17-4)	89.10	88.99	88.91	89.06	89.02	89.08	89.08	99.98	-	90.46	90.41	90.40	89.05	98.93	90.19 - 90.31	90.37
10. BCC1957 (J48)	90.89	90.93	90.89	90.86	90.92	90.98	90.95	90.40	90.43	-	98.78	98.81		90.42	*94.43 - 95.09	92.82
11. BCC1965 (J91-1)	90.91	90.92	91.01	90.85	90.97	90.98	90.96	90.47	90.48	98.80	-	99.99	90.88	90.47	* 94.73 - 95.10	92.86
12. BCC1964 (J91-2)	90.89	90.85	90.92	90.85	90.92	90.93	90.94	90.41	90.48	98.79	99.99	-	90.91	90.42	*94.73 - 95.08	92.82

***B. diffusa* like strains (MSMB 583 WGS, MSMB 377 WGS, INT4-BP16, MSMB 1060, MSMB 378 WGS, MSMB 375 WGS, MSMB 866)**
13 *B. cepacia* ATCC 25416^T; 14 *B. vietnamiensis* LMG 10929^T; 15 *B. diffusa* CCUG54558^T; 16 *B. ambifaria* AMMD^T
ANI values supportive of species identity are highlighted in green, and those supportive of a novel taxa closely related but distinct from *B. diffusa* are highlighted in blue

5.2.4.4 Pan-genome analysis of the twelve environmental *Burkholderia*

A core gene based phylogeny of the twelve environmental *Burkholderia* isolates and their closest type strain neighbours (limited to *B. vietnamiensis* and *B. cepacia*) was constructed using roary (Page et al. 2015). This analysis identified 1606 core genes shared by the *Burkholderia* genomes. In addition, 13558 shell genes and 5557 cloud genes were also present, and no soft-core genes were predicted. A phylogenetic tree of twelve environmental members of the Bcc was constructed after alignment of the 1606 core genes. The core gene phylogeny confirmed all the previous results, with seven jungle strains closely aligning with *B. cepacia*, two jungle strains clustering with *B. vietnamiensis* and the three-remaining jungle *Burkholderia* separating as a distinct clade (Figure 5.4). Within the *B. cepacia* group, jungle isolates J86-1 (BCC1962) and J86-2 (BCC1962) were highly clonal by core gene analysis (Figure 5.4), correlating to the fact they were isolated from the same sample site. Interestingly, isolate J80 (BCC1963) was also closely related to these two strains but was obtained from a different sample site. Similarly, two of the novel Bcc taxa, jungle isolates J91-1 (BCC1965) and J91-2 (BCC1964) appeared to be clonal; they were also recovered from the same sample (Figure 5.3). The two jungle *B. vietnamiensis* isolates were also essentially identical by core gene analysis and again derived from the same sample site. Overall, the core genome phylogeny demonstrated high resolution positioning of twelve environmental jungle *Burkholderia* isolates, showing they represented three taxonomic groups within the Bcc and comprised seven different strains (Table 5.6).

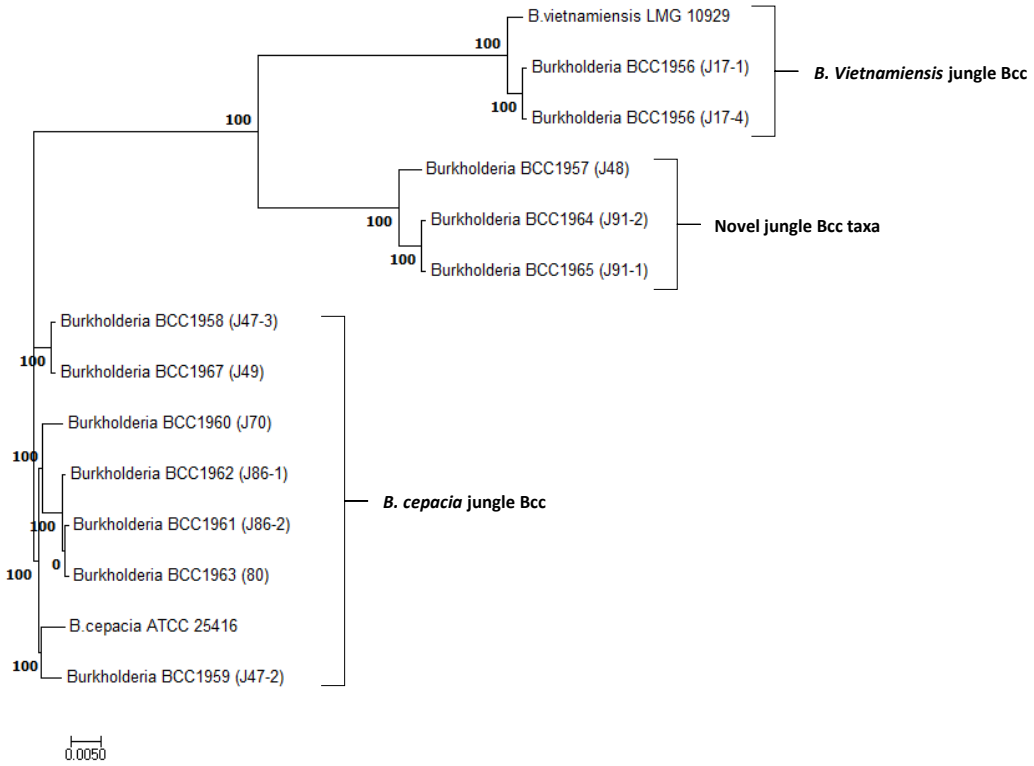


Figure 5.4 Core genes based phylogenetic tree for twelve environmental *Burkholderia* isolates. Core genes (1606) were extracted and aligned by Roary. Phylogenetic tree was constructed by fasttree based on maximum-likelihood (ML) method. The position of twelve Bcc species/strains and its closest neighbour type strain of *B. cepacia* and *B. vietnammensis* are shown into three different group of strains. The genetic distance scale shows that (0.0005), which represents substitutions per site.

Table 5.6 Final identification of *B. cepacia* complex jungle isolates characterised in this study

Group/Genus and Strain ID	Final ID	16S rRNA gene	*recA* gene	DDH	ANI	*Strain prediction
J17-1 (BCC1956)	*B. vietnamiensis*	*B. vietnamiensis*	*B. vietnamiensis*	*B. vietnamiensis*	*B. vietnamiensis*	1
J17-4 (BCC1955)	*B. vietnamiensis*	*B. vietnamiensis*	*B. vietnamiensis*	*B. vietnamiensis*	*B. vietnamiensis*	1
J47-2 (BCC1959)	*B. cepacia*	*B. lata*	*B. cepacia*	*B. cepacia*	*B. cepacia*	2
J47-3 (BCC1958)	*B. cepacia*	*B. lata*	*B. cepacia*	*B. cepacia*	*B. cepacia*	3
J49 (BCC1967)	*B. cepacia*	*B. lata*	*B. cepacia*	*B. cepacia*	*B. cepacia*	3
J70 (BCC1960)	*B. cepacia*	*B. lata*	*B. cepacia*	*B. cepacia*	*B. cepacia*	4
J80-2 (BCC1963)	*B. cepacia*	*B. cepacia*	*B. cepacia*	*B. cepacia*	*B. cepacia*	5
J86-1 (BCC1962)	*B. cepacia*	*B. cepacia*	*B. cepacia*	*B. cepacia*	*B. cepacia*	5
J86-2 (BCC1961)	*B. cepacia*	*B. cepacia*	*B. cepacia*	*B. cepacia*	*B. cepacia*	5
J48-2 (BCC1957)	Bcc	*B. diffusa*	Bcc (novel Group 3)	Bcc	Bcc (novel Group 3)	6
J91-1 (BCC1965)	Bcc	*B. diffusa*	Bcc (novel Group 3)	Bcc	Bcc (novel Group 3)	7
J91-2 (BCC1964)	Bcc	*B. diffusa*	Bcc (novel Group 3)	Bcc	Bcc (novel Group 3)	7

***shared MLST sequence type or and core gene analysis**

5.2.5 *Paraburkholderia* classifications based on genomic data

5.2.5.1 Full length 16S rRNA and *recA* gene phylogenies

The position and relatedness of the 44 *Paraburkholderia* isolates to other members of this genus were evaluated via the full length of 16S rRNA gene and *recA* gene sequence phylogenies. It is important here to mention that two isolates, BCC1912 (J7) and BCC1932 (J35), were excluded from the 16S rRNA-based phylogenies, because only short 16S rRNA gene sequences (approximately 500 bp) could be extracted from their genomic data. The genus identification of *Paraburkholderia* for these two isolates was predicted using Kraken and a BLASTn search using the available ~500 bp 16S rRNA gene sequences (Table 5.1). Despite the issues with extracting single 16S rRNA gene sequences, the draft genomes of BCC1912 (J7) and BCC1932 (J35) were judged as high quality using QUAST (Table 5.1), and were suitable for whole genome comparison using ANI.

Full length 16S rRNA gene phylogenetic trees were generated for 42 jungle *Paraburkholderia* isolates and 25 reference species (Figures 5.5 and 5.6). The reference sequences were from sixteen validly named *Paraburkholderia* species and included: *Paraburkholderia tropica, Paraburkholderia caledonica, Paraburkholderia ferrariae* and *Paraburkholderia aspalathi*; seven *Caballeronia* species references were also included. The same dataset of reference species was used to generate a *recA* gene phylogenetic tree, that also included jungle isolates BCC1912 (J7) and BCC1932 (J35), as no problems were encountered with extracting their *recA* gene sequences. The same overall clustering of jungle isolates was observed in both 16S rRNA gene and *recA* gene phylogenies as follows.

For the 16S rRNA gene analysis, a large cluster of 25 *Paraburkholderia* jungle isolates formed a well-supported and distinct group (Group 1; Figure 5.5). The Group 1 strains included the type strain of *P. tropica* LMG 22274^{T} as well as other characterised *P. tropica* strains, including Ppe8, KACC 13422, SIr-6563, and SIr-6529. A second novel set of 16S rRNA genes, Group 2, was also identified and comprised eight jungle *Paraburkholderia* isolates [BCC1917 (J72), BCC1914 (J75-2), BCC1921 (J67), BCC1915 (J75-1), BCC1941 (J11-1), BCC1938 (J15-2),

BCC1916 (J74) and BCC1939 (J15-1)] (Figure 5.6). This group clustered closely to the Group 1 *Paraburkholderia tropica* related isolates but was clearly distinct from other known *Paraburkholderia* species. A third group (Group 3) of isolates comprised four jungle *Paraburkholderia* [J8-2 (BCC1943), J10-1 (BCC1954), J10-2 (BCC1942), and J11-2 (BCC1940)] (Figure 5.6). This group is placed separately from other groups of jungle isolates and reference sequences. Five additional jungle *Paraburkholderia* isolates [(BCC1931 (J41-1), BCC1913 (J76-2), BCC1930 (J42-2), BCC1953 (J12) and BCC1922 (J63-2)] were also distinct in terms of 16S rRNA gene inferred lineages (Group 4; Figure 5.6). However, *P. ferrariae* NBRC 106233^T was the closest *Paraburkholderia* type strain to the jungle *Paraburkholderia* species BCC1953 (J12), and *P. oxyphila* NBRC 105797^T was clustered nearby jungle isolate BCC1922 (J63-2) (Figure 5.6).

Figures 5.7 and 5.8 present *recA* gene phylogenies for all 44 *Paraburkholderia* isolates. The same overall clustering as seen in 16S rRNA gene phylogenies was observed within the *recA* gene tree. Group 1 contained the 22 isolates which represented a *P. tropica* group. Group 2 and 3 contained eight and four isolates, respectively, and clustered distinctly via *recA* gene analysis (Figure 5.7). Furthermore, ten individual *Paraburkholderia* isolates did not place within the groups of related isolates and *Paraburkholderia* references.

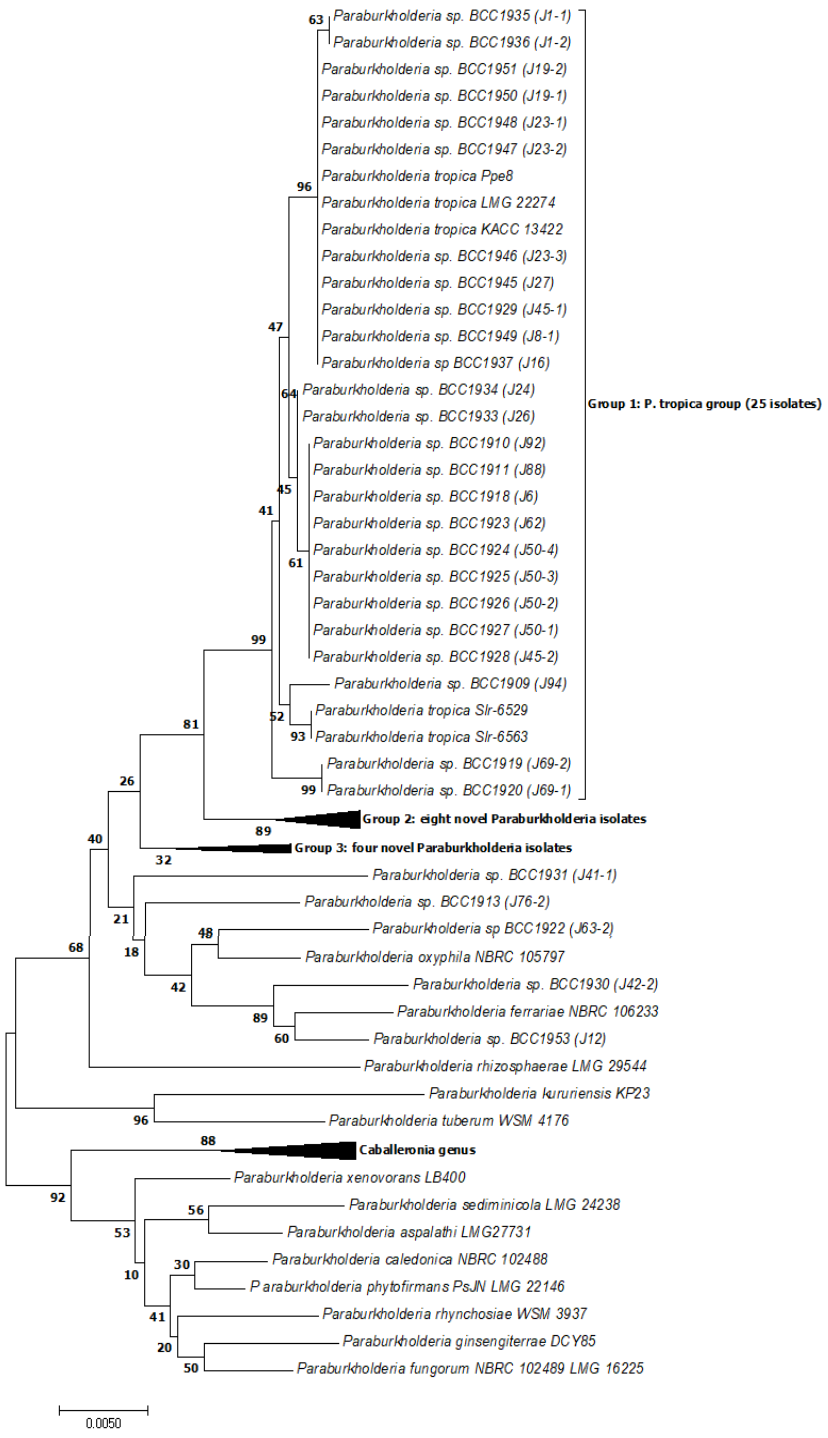


Figure 5.5 Neighbour-joining phylogenetic tree based on 16S rRNA sequences for *Paraburkholderia tropica* group. The phylogeny based on the full length of the 16S rRNA gene (1529 bp) for 25 *Paraburkholderia* isolates (Group 1: *P. tropica* group), seventeen *Paraburkholderia* isolates, *Caballeronia* group, and 25 references (eighteen *Paraburkholderia* and seven *Caballeronia* spp.). All extracted 16S rRNA sequences were aligned in MEGA using ClustalW. The positions of 25 *Paraburkholderia tropica* isolates were shown comprised with the type strains of *P. tropica*. All *Paraburkholderia* 16R rRNA sequences were aligned in MEGA using ClustalW. The tree is showing the positions of 25 *Paraburkholderia* isolates clustered within *Paraburkholderia tropica* type strain and other strains references. Phylogenetic analysis was conducted using MEGA 7, and the bootstrap values based on 1000 replications (Kumar et al. 2016). The genetic distance scale shows (0.0050) that represents substitutions per site.

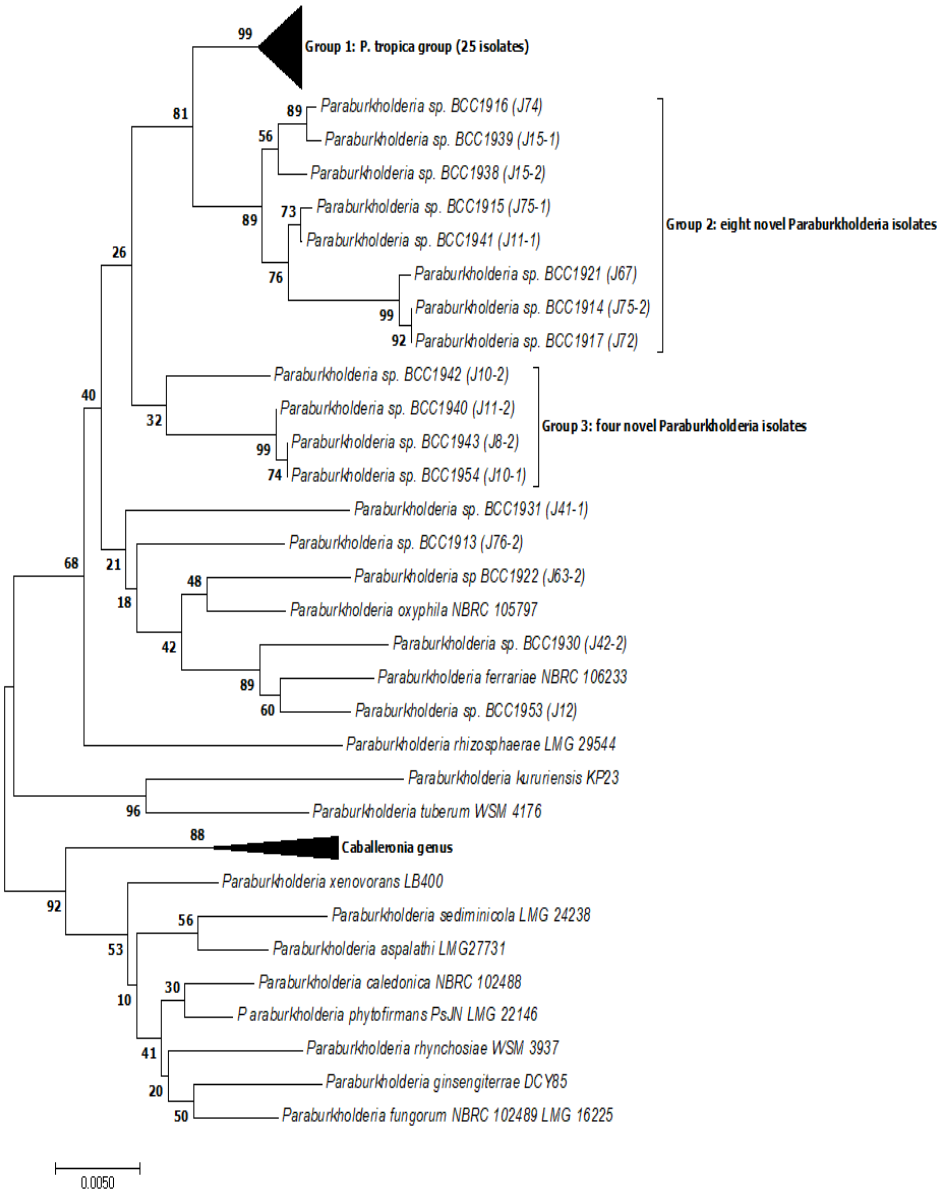


Figure 5.6 Neighbour-joining phylogenetic tree based on 16S rRNA sequences for novel *Paraburkholderia* isolates. The phylogeny based on the full length of 16S rRNA (1529 bp) for Group 1: *P. tropica* group containing 25 *P. tropica* strains, Group 2 contains eight novel *Paraburkholderia* isolates, Group 3 contains four novel *Paraburkholderia* isolates, Group 4 contains five different novel *Paraburkholderia* species, *Caballeronia* group, and 25 references (eighteen *Paraburkholderia* and seven *Caballeronia* spp.). All extracted 16S rRNA gene sequences were aligned in MEGA using ClustalW. The positions of seventeen *Paraburkholderia* were shown as follow: Group 2 (eight closely related *Paraburkholderia* isolates), Group 3 (four closely related *Paraburkholderia* isolates) and Group 4 (five individual *Paraburkholderia* species). Group 2 and 3 were shown clearly on the tree as they are formed two clusters of isolates within *Paraburkholderia* genus that are separated from type strain and other strains references. The four individual *Paraburkholderia* isolates. Each group contains related isolates clearly separated from the other groups of isolates and references. For Group 4, it is clearly that five individual *Paraburkholderia* isolates represent a group of novel species within *Paraburkholderia* genus. *Paraburkholderia* sp. BCC1922 (J63-2) was placed closely to the type strain of *Paraburkholderia oxyphila*, and *Paraburkholderia* sp. BCC1953 (J12) was placed closely to *Paraburkholderia ferrariae* NBRC 106233^{T}. Phylogenetic analysis was conducted using MEGA 7, and the bootstrap values based on 1000 replications (Kumar et al. 2016). The genetic distance scale shows (0.0050) that represents substitutions per site.

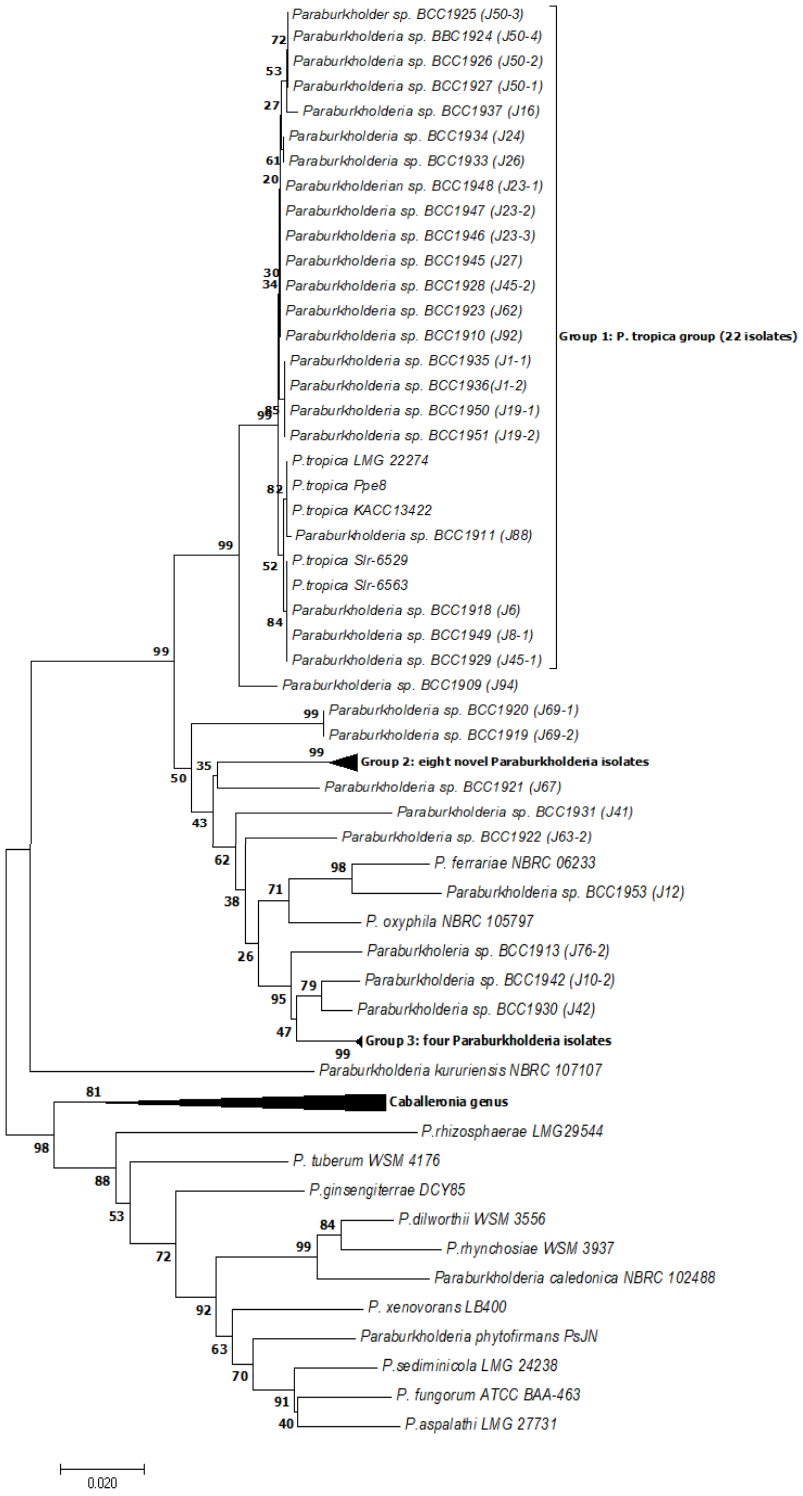


Figure 5.7 Neighbour-joining phylogenetic tree based on *recA* gene sequence for Group 1 of *Paraburkholderia* isolates (*P. tropica* group). The phylogeny based on 1086 bp for 45 isolates (44 putative *Paraburkholderia* and one *Caballeronia*) as well as references (nineteen *Paraburkholderia* and seven *Caballeronia*). The complete length of *recA* gene sequences were aligned in MEGA using ClustalW. The tree is showing the positions of 22 *Paraburkholderia* isolates grouped within *P. tropica* type strain and references. Phylogenetic analysis was conducted using MEGA 7, and the bootstrap values based on 1000 replications (Kumar et al. 2016). The genetic distance scale shows that (0.020), which represents substitutions per site.

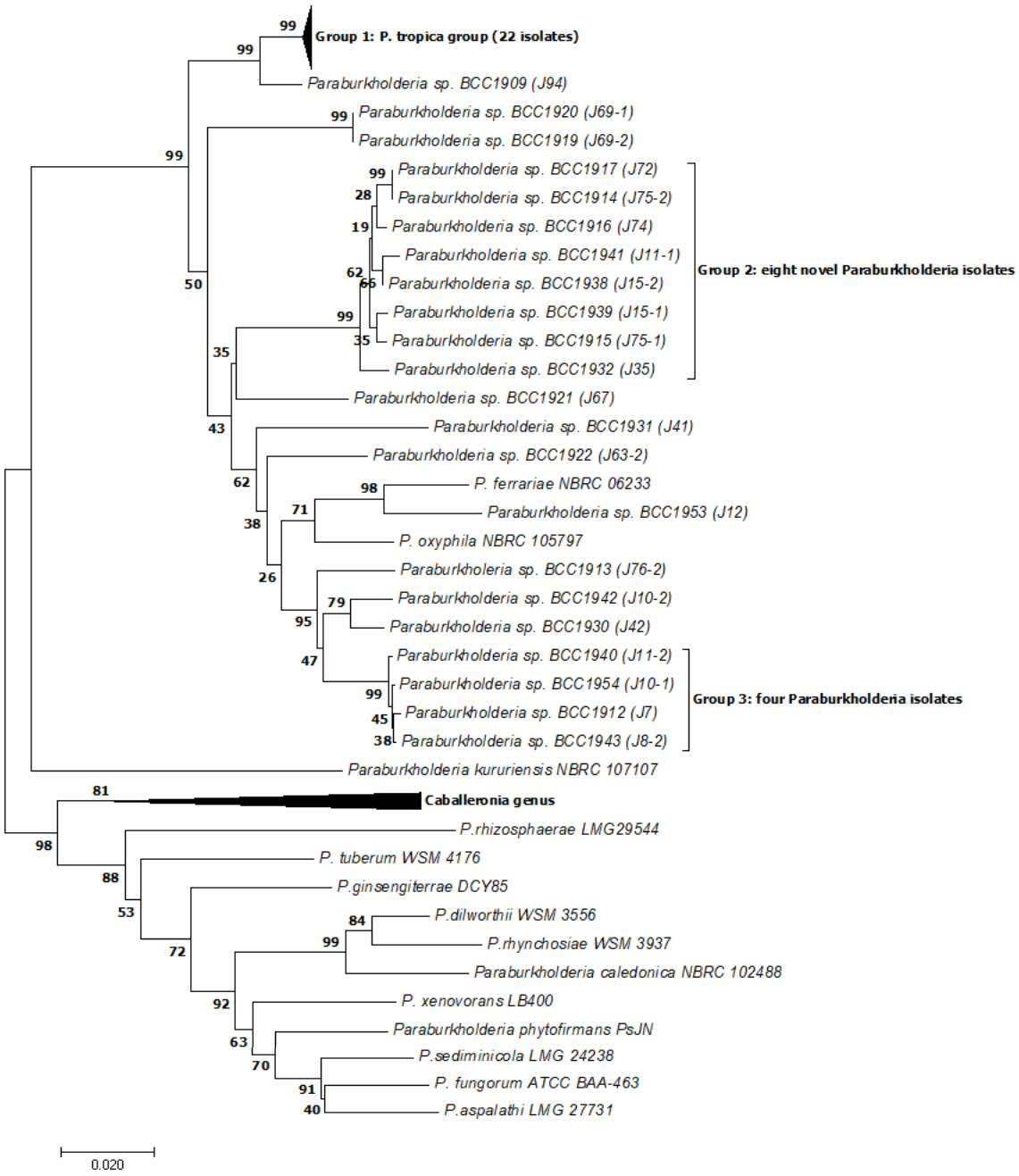


Figure 5.8 Neighbour-joining phylogenetic tree based on *recA* gene sequence for novel *Paraburkholderia* isolates.The phylogenetic analysis was based on 1086 bp for 45 isolates (44 putative *Paraburkholderia* and one *Caballeronia*) as well as references (nineteen *Paraburkholderia* and seven *Caballeronia*). The complete length of *recA* gene sequences were aligned in MEGA using ClustalW. The tree is showing the positions of two separate groups of isolates. Group 2 composes eight related isolates, and Group 3 that contains four related isolates. Ten different isolates are placed individually separated from others groups of isolates and references. Phylogenetic analysis was conducted using MEGA 7, and the bootstrap values based on 1000 replications (Kumar et al. 2016). The genetic distance scale shows that (0.020), which represents substitutions per site.

5.2.5.2 The collective ANI analysis of *P. tropica* group and novel *Paraburkholderia* spp./strains groups of jungle *Paraburkholderia* isolates

ANI analysis in this study was based on two methods. Initially, a FastANI method was used to identify closest-related *Paraburkholderia* type strains (data not shown) with the 44 jungle *Paraburkholderia* and one jungle *Caballeronia* isolate compared against a dataset of 205 assembled *Paraburkholderia* genomes downloaded from the databases (genomic dataset assembled by Alex Mullins, Cardiff University). Secondly, PyANI analysis was performed on selected subsets of reference *Paraburkholderia* and jungle isolates as presented below. Collectively, the ANI analysis confirmed that the jungle *Paraburkholderia* isolates could be divided into three major clusters as well as a total of ten *Paraburkholderia* species including multiple novel taxonomic groups (Figure 5.9). As expected from the full length 16S rRNA gene (Figure 5.5) and *recA* gene (Figure 5.7) analysis, the average nucleotide identity of the 22 most closely related jungle *Paraburkholderia* isolates represented a single taxonomic group, Group 1. This, *P. tropica* group, had ANI values that were greater than 99% between themselves and *P. tropica* LMG 22274^{T} and other reference strains of this species, *P. tropica* P-31, *P. tropica* BE15, *P. tropica* Ppe8, *P. tropica* SIr-6563 and *P. tropica* SIr-6529. This ANI based result unequivocally classified them as *P. tropica* (Figure 5.9).

The remaining 22 jungle *Paraburkholderia* isolates shared less than 95% of their nucleotide identity (ANI) with the other known *Paraburkholderia* species (Figure 5.9). Group 2 contained the following novel eight *Paraburkholderia* spp. strains, BCC1941 (J11-1), BCC1939 (15-1), BCC1938 (J15-2), BCC1932 (J35), BCC1917 (J72), BCC1916 (J74), BCC1915 (J75-1) and BCC1914 (J75-2), represented a district genotaxonomic group within the *Paraburkholderia* species (Figure 5.9). When compared to themselves, their ANI values ranged from 96.04 to 97.70, demonstrating they were likely a novel single *Paraburkholderia* taxa. ANI analysis of these isolates with their closest phylogenetic neighbours, *Paraburkholderia bannensis* NBRC 103871 and *Paraburkholderia eburnea* JCM 18070, gave similarities of less than 95% and 90%, respectively (data not shown). The third ANI taxonomic group contained four isolates of

jungle *Paraburkholderia*: BCC1912 (J7), BCC1943 (J8-2), BCC1954 (J10-1) and BCC1940 (J11-2), which shared ANI values ranging from 97.58% to 97.86% when compared to themselves (Figure 5.9). The ANI values between *Paraburkholderia heleia* NBRC 101817^T, as the nearest type strain of known *Paraburkholderia* species, and the four jungle *Paraburkholderia* ranged from 91.06% to 91.14% (Figure 5.9).

ANI analysis showed that the following jungle *Paraburkholderia* comprised ten novel taxa *Paraburkholderia* isolates: BCC1942 (J10-2), BCC1953 (J12), BCC1931 (J41), BCC1930 (J42), BCC1922 (J63), BCC1919 (J69-1), BCC1920 (69-2), BCC1921 (J67), BCC1913 (J76-2) and BCC1909 (J94) (Figure 5.9). Thus, the ANI data analysis showed that these ten different species are relatively distinct from other reference *Paraburkholderia* species and the previous three jungle *Paraburkholderia* groups, because they share less than 90% of the ANI values between each other (Figure 5.9). Table 5.7 shows the finalised identification for the total of 44 *Paraburkholderia* isolates characterised in this study.

5.2.5.3 Pan-genome analysis

The close taxonomic relatedness *Paraburkholderia* Group 1 (22 isolates), Group 2 (eight isolates), and Group 3 (four isolates) enabled core gene phylogenies to be generated using roary (Figure 5.10) (Page et al. 2015). The core genome phylogenies for these different groups demonstrated the position of isolates within each group. For *P. tropica* group 1, 6120 core genes of 22 jungle isolates and *Paraburkholderia tropica* strains (including the type strains) were aligned by roary. The phylogeny for this group indicated that 22 jungle isolates and the reference strains were all genetically distinct strains of *P. tropica* (Figure 5.10, panel A). For Group 2, the phylogenetic tree of the eight novel species jungle isolates was based on the alignment of 5808 core genes and showed that six of them were genetically different strains (Figure 5.10, panel B). Isolates BCC1917 (J72) and BCC 1914 (J75-2) were very closely related from the core gene phylogeny (Figure 5.10, panel B), but were recovered from different sampling sites. Based on 6421 core genes, the phylogenetic alignments of novel

Paraburkholderia Group 3 isolates also demonstrated they were all genetically distinct. In summary, each jungle *Paraburkholderia* group isolates contained genetically diverse strains.

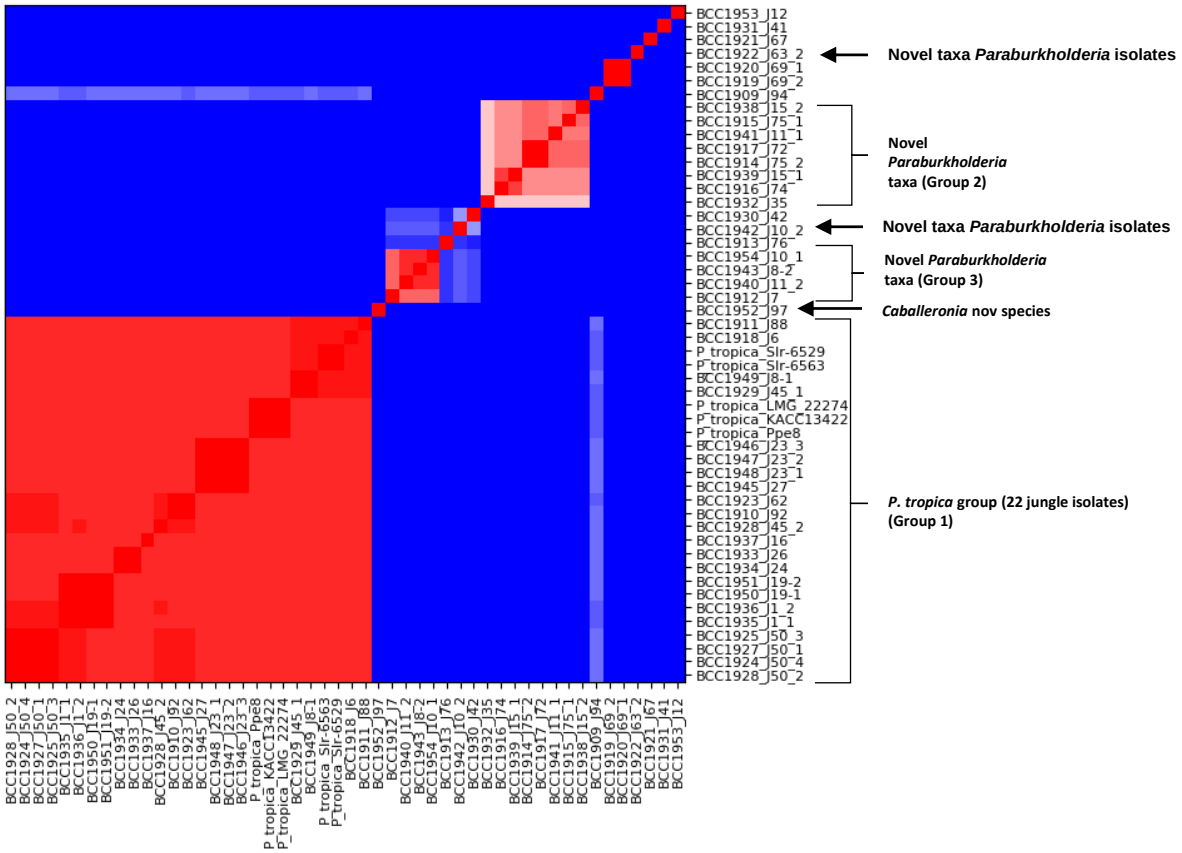


Figure 5.9 The whole-genome comparisons of 44 *Paraburkholderia* based on the ANI analysis. A total of 44 jungle *Paraburkholderia*, 1 *Caballeronia* species (BCC1952; J97), and five *P. tropica* refrences are shown in the analysis. The ANI analysis was performed by means of Python module (Pyani). The jungle isolates and references clustered in three main species groups as follows. Group 1 is *P. tropica* group (red box in the bottom). This group contains *P. tropica* references, including the type strain, and the 22 closely related jungle isolates. Group 2 jungle *Paraburkholderia* represented four closely related isolates (red box in the middle). Group 3 comprised eight closely related jungle strains (red box in the top). The small red boxes represent ten novel *Paraburkholderia* species. The strains with ANI values >90 are shown within blue, and the strains with ANI >95% are shown in red squares.

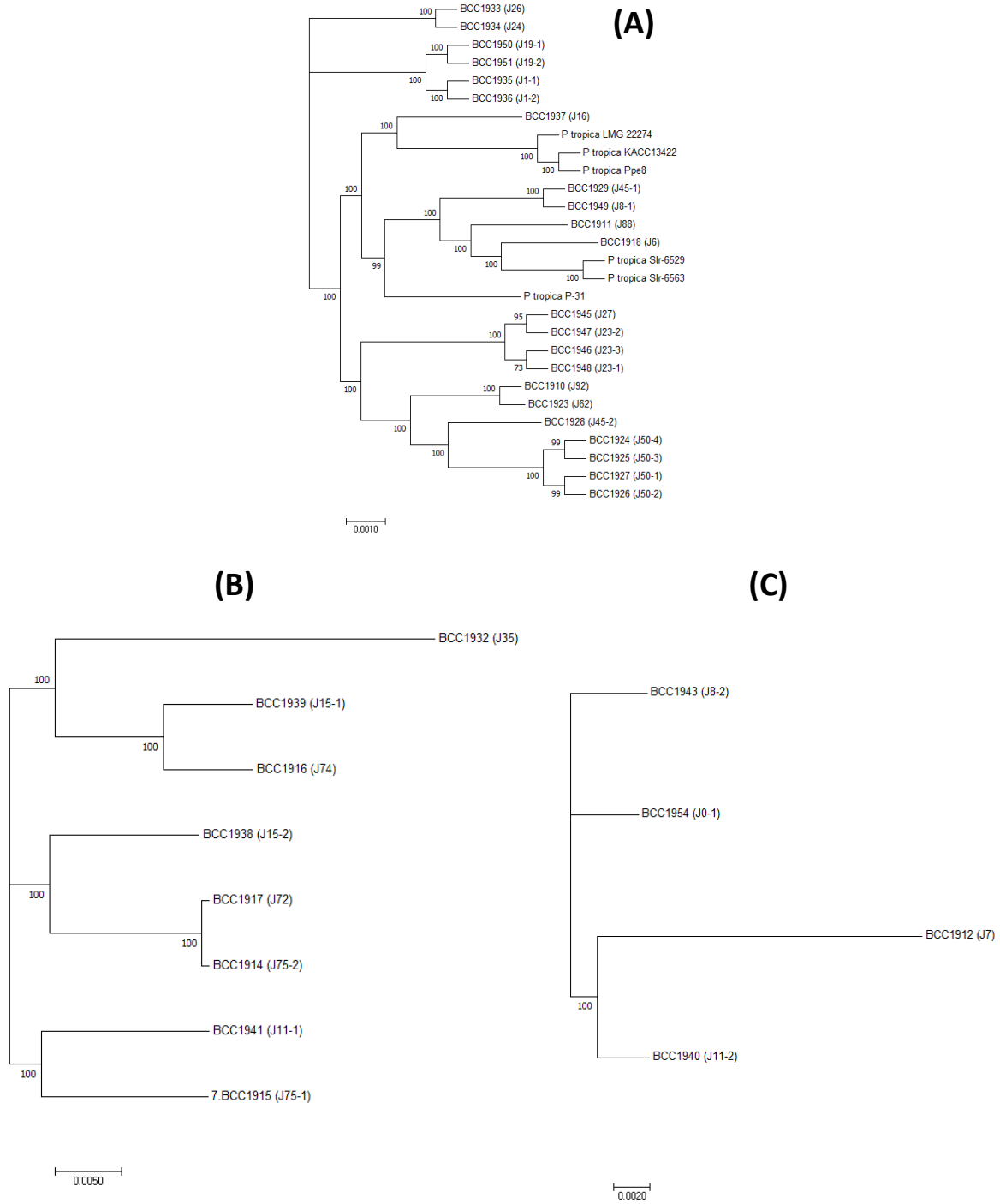


Figure 5.10 Core genes based phylogenetic tree for three main groups of *Paraburkholderia* jungle isolates. The phylogeny shows the relationship between each of the *P. tropica* jungle isolates. Core genes of each group were extracted and aligned by Roary, and the phylogenetic tree was constructed by fasttree based on maximum-likelihood (ML) method. Panel A shows the diversity within the *P. tropica* Group 1 isolates (22 jungle strains and five reference isolates) obtained by the analysis of 6120 core genes. Panel B shows the diversity of novel jungle *Paraburkholderia* Group 2 isolates (*n* =8) based on the alignment of 5808 core genes. Panel C shows the phylogeny of the four novel *Paraburkholderia* groups three isolates based on the alignment of 6421 core genes. The genetic distance scale shows (0.0010) for *P. tropica* group (Panel A), (0.0050) for Group 2 (panel B) and (0.0020) for Group 3 (panel C).

Table 5.7 Final identification of *Paraburkholderia* jungle isolates characterised in this study.

Group/Genus and Strain ID	Final ID	16S rRNA gene	*recA* gene	ANI
***Paraburkholderia tropica* Group 1 (*n* = 22)**				
J1-1 (BCC1935)	*P. tropica*	*P. tropica*	*P. tropica*	*P. tropica*
J1-2 (BCC1936)	*P. tropica*	*P. tropica*	*P. tropica*	*P. tropica*
J6 (BCC1918)	*P. tropica*	*P. tropica*	*P. tropica*	*P. tropica*
J8-1 (BCC1949)	*P. tropica*	*P. tropica*	*P. tropica*	*P. tropica*
J16 (BCC1937)	*P. tropica*	*P. tropica*	*P. tropica*	*P. tropica*
J19-1 (BCC1950)	*P. tropica*	*P. tropica*	*P. tropica*	*P. tropica*
J19-2 (BCC1951)	*P. tropica*	*P. tropica*	*P. tropica*	*P. tropica*
J23-1 (BCC1948)	*P. tropica*	*P. tropica*	*P. tropica*	*P. tropica*
J23-2 (BCC1947)	*P. tropica*	*P. tropica*	*P. tropica*	*P. tropica*
J23-3 (BCC1946)	*P. tropica*	*P. tropica*	*P. tropica*	*P. tropica*
J24 (BCC1934)	*P. tropica*	*P. tropica*	*P. tropica*	*P. tropica*
J26 (BCC1933)	*P. tropica*	*P. tropica*	*P. tropica*	*P. tropica*
J27 (BCC1945)	*P. tropica*	*P. tropica*	*P. tropica*	*P. tropica*
J45-1 (BCC1929)	*P. tropica*	*P. tropica*	*P. tropica*	*P. tropica*
J45-2 (BCC1928)	*P. tropica*	*P. tropica*	*P. tropica*	*P. tropica*
J50-1 (BCC1927)	*P. tropica*	*P. tropica*	*P. tropica*	*P. tropica*
J50-2 (BCC1926)	*P. tropica*	*P. tropica*	*P. tropica*	*P. tropica*
J50-3 (BCC1925)	*P. tropica*	*P. tropica*	*P. tropica*	*P. tropica*
J50-4 (BCC1924)	*P. tropica*	*P. tropica*	*P. tropica*	*P. tropica*
J62 (BCC1923)	*P. tropica*	*P. tropica*	*P. tropica*	*P. tropica*
J88 (BCC1911)	*P. tropica*	*P. tropica*	*P. tropica*	*P. tropica*
J92 (BCC1910)	*P. tropica*	*P. tropica*	*P. tropica*	*P. tropica*
***Paraburkholderia* sp. nov. Group 2 (*n*= 8)**				
J11-1 (BCC1941)	*Paraburkholderia* sp. nov. Group 2	*Paraburkholderia* sp.	*Paraburkholderia* sp.	*Paraburkholderia* sp.
J15-1 (BCC1939)	*Paraburkholderia* sp. nov. Group 2	*Paraburkholderia* sp.	*Paraburkholderia* sp.	*Paraburkholderia* sp.
J15-2 (BCC1938)	*Paraburkholderia* sp. nov. Group 2	*Paraburkholderia* sp.	*Paraburkholderia* sp.	*Paraburkholderia* sp.
J35 (BCC1932)	*Paraburkholderia* sp. nov. Group 2	*Paraburkholderia* sp.	*Paraburkholderia* sp.	*Paraburkholderia* sp.
J72 (BCC1917)	*Paraburkholderia* sp. nov. Group 2	*Paraburkholderia* sp.	*Paraburkholderia* sp.	*Paraburkholderia* sp.
J74 (BCC1916)	*Paraburkholderia* sp. nov. Group 2	*Paraburkholderia* sp.	*Paraburkholderia* sp.	*Paraburkholderia* sp.
J75-1 (BCC1915)	*Paraburkholderia* sp. nov. Group 2	*Paraburkholderia* sp.	*Paraburkholderia* sp.	*Paraburkholderia* sp.
J75-2 (BCC1914)	*Paraburkholderia* sp. nov. Group 2	*Paraburkholderia* sp.	*Paraburkholderia* sp.	*Paraburkholderia* sp.
***Paraburkholderia* sp. nov. Group 3 (*n*=4)**				
J7 (BCC1912)	*Paraburkholderia* sp. nov. Group 3	*Paraburkholderia* sp.	*Paraburkholderia* sp.	*Paraburkholderia* sp.
J8-2 (BCC1943)	*Paraburkholderia* sp. nov. Group 3	*Paraburkholderia* sp.	*Paraburkholderia* sp.	*Paraburkholderia* sp.
J10-1 (BCC1954)	*Paraburkholderia* sp. nov. Group 3	*Paraburkholderia* sp.	*Paraburkholderia* sp.	*Paraburkholderia* sp.
J11-2 (BCC1940)	*Paraburkholderia* sp. nov. Group 3	*Paraburkholderia* sp.	*Paraburkholderia* sp.	*Paraburkholderia* sp.
***Paraburkholderia* sp. nov. (single strain novel species; *n*=10)**				
J10-2 (BCC1942)	*Paraburkholderia* sp. nov. J10-2	*Paraburkholderia* sp.	*Paraburkholderia* sp.	*Paraburkholderia* sp.
J12 (BCC1953)	*Paraburkholderia* sp. nov. J12	*P. ferrariae*	*P. ferrari*	*Paraburkholderia* sp.
J41 (BCC1931)	*Paraburkholderia* sp. nov. J41	*Paraburkholderia* sp.	*Paraburkholderia* sp.	*Paraburkholderia* sp.
J42 (BCC1930)	*Paraburkholderia* sp. nov. J42	*Paraburkholderia* sp.	*Paraburkholderia* sp.	*Paraburkholderia* sp.
J63 (BCC1922)	*Paraburkholderia* sp. nov. J63	*P. oxyphila*	*Paraburkholderia* sp.	*Paraburkholderia* sp.
J69-1 (BCC1919)	*Paraburkholderia* sp. nov. J69-1	*Paraburkholderia* sp.	*Paraburkholderia* sp.	*Paraburkholderia* sp.
J69-2 (BCC1920)	*Paraburkholderia* sp. nov. J92-2	*Paraburkholderia* sp.	*Paraburkholderia* sp.	*Paraburkholderia* sp.
J67 (BCC1921)	*Paraburkholderia* sp. nov. J67	*Paraburkholderia* sp.	*Paraburkholderia* sp.	*Paraburkholderia* sp.
J76 (BCC1913)	*Paraburkholderia* sp. nov. J76	*Paraburkholderia* sp.	*Paraburkholderia* sp.	*Paraburkholderia* sp.
J94 (BCC1909)	*Paraburkholderia* sp. nov. J94	*Paraburkholderia* sp.	*Paraburkholderia* sp.	*Paraburkholderia* sp.

5.2.6 Nitrogen-fixing (*nif*) genes predicted on four closely related *P. tropica* isolates and two *B. vietnamiensis*

The nitrogen fixation (*nif*) genes encode a nitrogenase enzyme, which reduces atmospheric nitrogen to ammonia, which can then be metabolised or transferred to plants (De Meyer et al.

2016). At least 20 *nif* gene types have been reported that are linked to nitrogen fixation and organized in eight operons (Patel et al. 2019). The organization and complexity of *nif* genes varies in different groups of microorganisms. The enzyme complex of nitrogenase consists of two subunits that are encoded by *nifDK* and *nifH*. The regulation of *nif* genes is controlled by *nifA* gene (a positive activator) and *nifL* gene (a negative regulator) (Patel et al. 2019). Both genes are effected by the concentration of oxygen and nitrogen, and this controls *nif* gene regulation in the natural environment (Merrick and Edwards 1995).

Within the present study, the prokka annotated draft genomes of the jungle isolates were reviewed to identify predicted *nif* genes. This method is not accurate for prediction of the complete *nif* genes operon, as it was only identifying individual *nif* genes. A set of orthologous *nif* genes were found in four *Paraburkholderia* and two *Burkholderia* isolates. BCC1927 (J50-1), BCC1926 (J50-2), BCC1925 (J50-3) and BCC1924 (J50-4) were all *P. tropica* strains, correlating with the discovery of nitrogen fixation within this *Paraburkholderia* species (Reis et al. 2004). Interestingly all the latter jungle *P. tropica* were recovered from a single sample site, and were closely related forming a single clade, but were distinct strains from core gene analysis (Figure 5.10, panel A). The jungle strains, *Burkholderia vietnamiensis* BCC1955 (J17-4) and *Burkholderia vietnamiensis* BCC1956 (J17-1), were also found to encode nitrogen-fixation genes, correlating with the known property of this species (Gillis et al. 1995). The nif genes predicted in this study for the jungle *B. vietnamiensis* isolates were *nifH*, *nifD* and *nifK* (nitrogenise synthesis genes), *nifA* (activator gene for *nif* genes operons), and other *nif* genes (*nifB, nifH, nifW* and *nifS*). *Caballeronia* sp. nov BCC1952 (J97) isolated from the Bornean jungle.

One jungle isolate BCC1952 (J97) was most similar to species of the recently proposed genus *Caballeronia* based on local BLASTn alignment tool for full length of 16S rRNA and *recA* gene sequence analysis. 16S rRNA gene (Figure 5.11) and *recA* gene (Figure 5.12) phylogenies were generated, which included validly named *Paraburkholderia* and seven *Caballeronia* species (*Caballeronia arvi* LMG 29317^T, *Caballeronia cordobensis* LMG 27620^T, *Caballeronia*

glebae LMG 29325^T, *Caballeronia choica* LMG 22940^T, *Caballeronia fortuita* LMG 29320^T, *Caballeronia concitans* LMG 29315^T and *Caballeronia humi* LMG 22934^T) (Figure 5.11 and 5.12). The phylogenies showed that the jungle isolate BCC1952 (J97) grouped separately from *Paraburkholderia* species, but was clearly within *Caballeronia*, and most likely constituted a novel species. Comparison of the genome for *Caballeronia sp.* BCC1952 (J97) with those of closely related type strains generated ANI values below ≤ 90% (Figure 5.13). The ANI analysis was also corroborated with DDH experiments for *Caballeronia* sp. BCC1952 (J97) and other *Caballeronia* spp. references (Table 5.5). All the DDH values were below the accepted 70% threshold for species ranging between (24 to 39.40%) (Achtman and Wagner 2008) (Table 5.6). Overall, this genotaxonomic analysis confirmed that *Caballeronia sp.* BCC1952 (J97) was a novel species of this newly proposed genus.

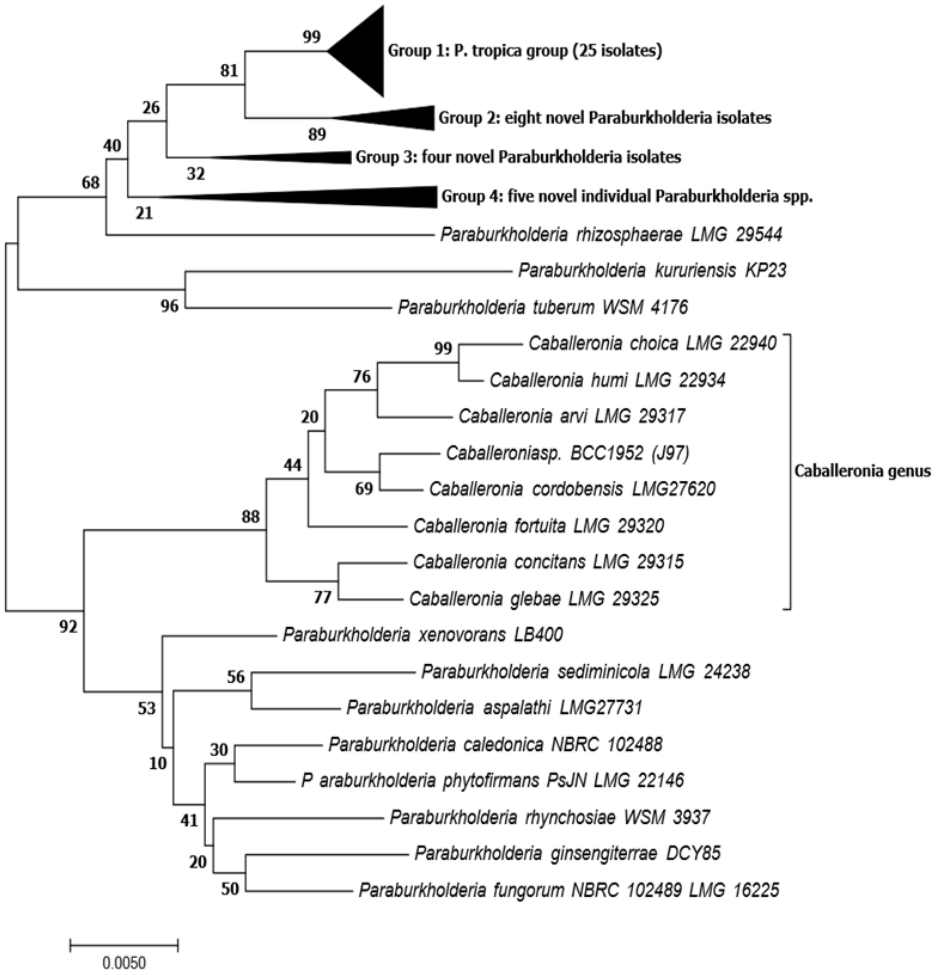


Figure 5.11 Neighbour-joining phylogenetic tree based on 16S rRNA gene sequence for novel *Caballeronia* isolate BCC1952 (J97). The phylogeny full length of 16S rRNA sequence (1529 bp) for jungle *Paraburkholderia* isolates and *Caballeronia* and 23 references (sixteen *Paraburkholderia* and seven *Caballeronia* spp.). All extracted 16S rRNA sequences were aligned in MEGA using ClustalW. The position of BCC1952 (J97) isolate is shown within *Caballeronia* references. Phylogenetic analysis was conducted using MEGA 7, and the bootstrap values based on 1000 replications (Kumar et al. 2016). The genetic distance scale shows that (0.0050), which represents substitutions per site.

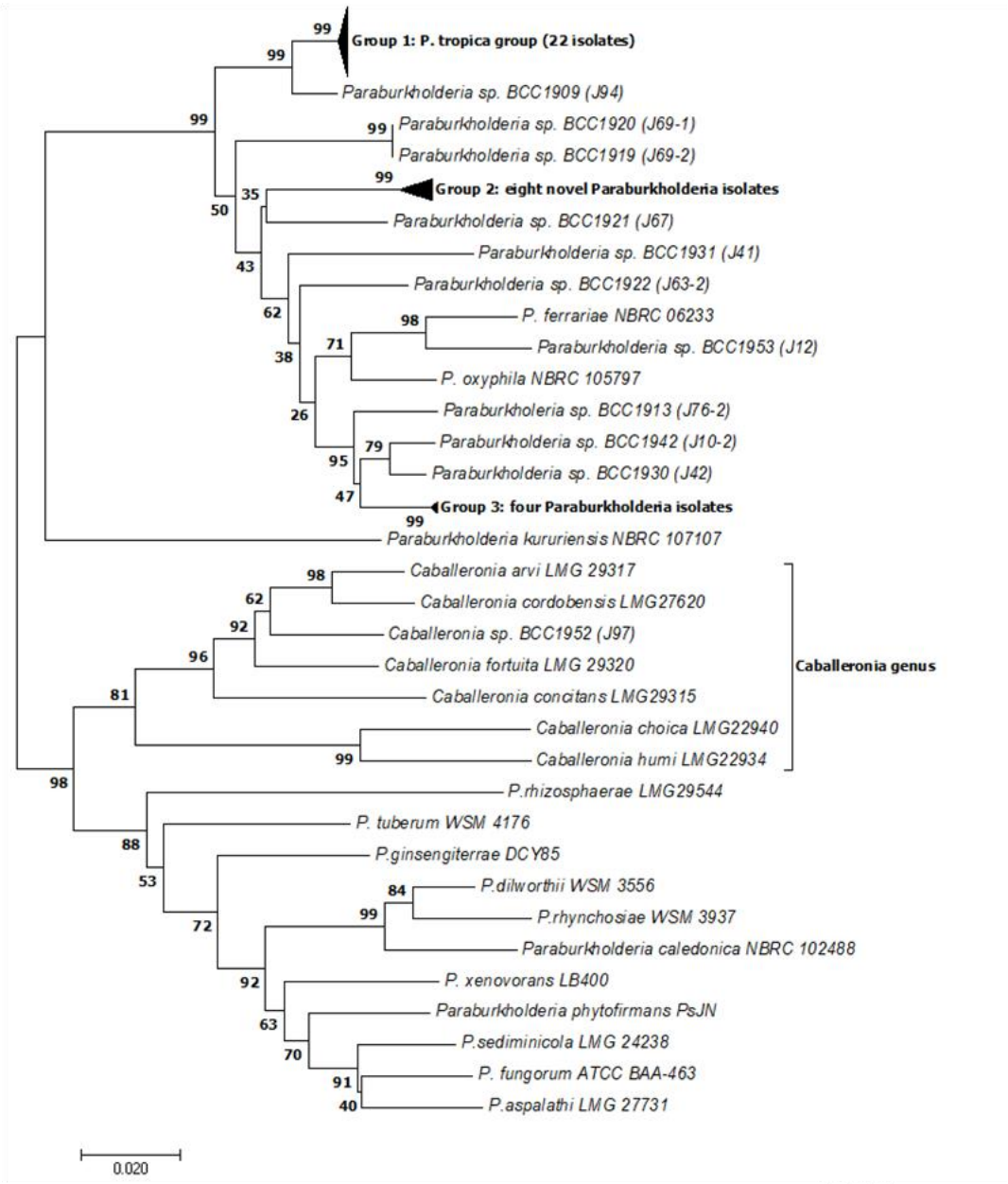


Figure 5.12 Neighbour-joining phylogenetic tree based on *recA* gene sequence for *Caballeronia*. The phylogenetic analysis was based on 1086 bp for jungle *Paraburkholderia* and *Caballeronia* and references (sixsteen *Paraburkholderia* and seven *Caballeronia*). The complete length of *recA* gene sequences were aligned in MEGA using ClustalW. The tree is showing the positions of BCC1952 (J97) isolate within *Caballeronia* refrences. Phylogenetic analysis was conducted using MEGA 7, and the bootstrap values based on 1000 replications (Kumar et al. 2016). The genetic distance scale shows that (0.020), which represents substitutions per site.

Table 5.8 Pairwise DDH hybridization between *Caballeroniea* sp. nov BCC1952 (J97) and type strains of *Caballeronia* genus.

DDH for ***Caballeroniea* sp. BCC1952 (J97)**	
***Caballeronia arvi* LMG 29317**	33.30%
***Caballeronia cordobensis* LMG 27620**	32.60%
***Caballeronia glebae* LMG 29325**	39.40%
***Caballeronia choica* LMG 22940**	24.20%
***Caballeronia fortuita* LMG 29320**	36.60%
***Caballeronia concitans* LMG 29315**	27.40%
***Caballeronia humi* LMG 22934**	24.20%

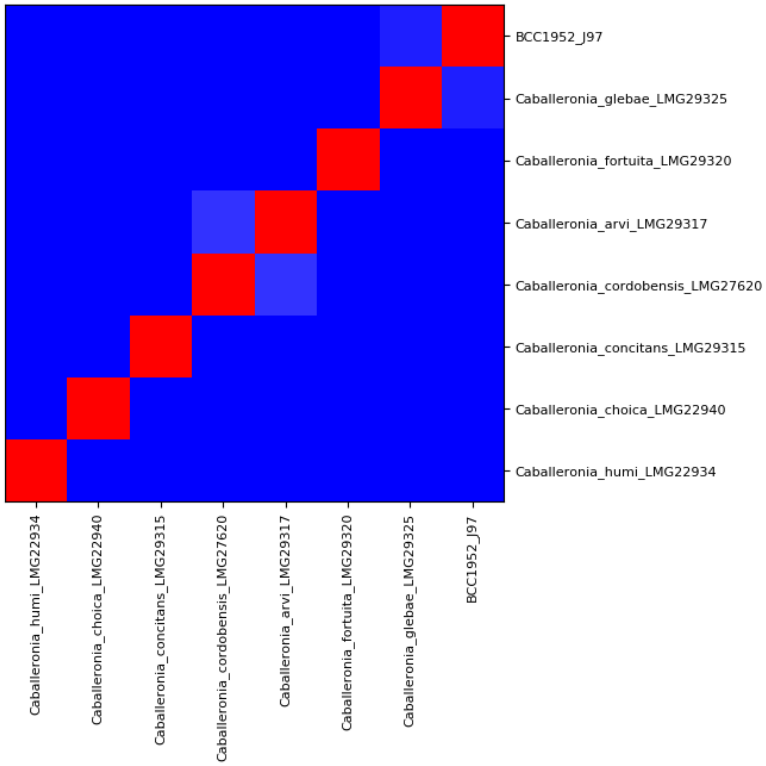


Figure 5.13 The whole-genome comparisons of *Caballeronia* sp. nov BCC1952 (J97) and type strains of *Caballeronia* genus. The ANI analysis was performed by means of Python module (Pyani). *Caballeroniea* sp. nov BCC1952 (J97) and type strains of *Caballeronia* were not related to each other's. No closer ANI was found with any of the published type strains. The strains with ANI values >90 are shown within blue, and the strains with ANI and >95% are shown in red squares.

5.3 Discussion

From single gene-based phylogeny to whole-genome comparison methods, the unique environmental collection from the Bornean jungle represented both novel taxa and known members within the *Burkholderiaceae* family. The study by Vandamme et al. (2017) and other taxonomic studies of *Burkholderia* sensu lato bacteria have reported the DNA G+C content

ranges of *Burkholderia*, *Paraburkholderia* and *Caballeronia* genera. The GC content of *Burkholderia* genomes is comparatively high, ranging between 65.7–68.5 mol%. The GC content of *Paraburkholderia* and *Caballeronia* is lower than *Burkholderia.* The range of the DNA G+C contents of *Paraburkholderia* members is from 61.4 to 65 mol%, and *Caballeronia* overlaps with *Paraburkholderia* by a range of between 58.9 and 65.0 mol% (Gyaneshwar et al. 2011; Sawana et al. 2014; Vandamme et al. 2017; Dobritsa and Samadpour 2019). In addition to these figures, and according to 16S rRNA and *recA* gene-based phylogenetic trees, the isolates from this study are categorised into three genera. Twelve isolates belong to the *Burkholderia* genus, specifically within the Bcc group, which have a G+C content of 66–67%. Forty-four isolates cluster within the *Paraburkholderia* genus with G+C content 63% to 65%, and one isolate belongs to the *Caballeronia* genus with a G+C content of 63.27%.

DDH and ANI are powerful tools for discriminating species. For the pairwise comparison between two genomes, a DDH value of ≤ 70% indicates that the tested genome belongs to a different species, a DDH value of > 70% indicates that it belongs to the same species and a DDH value of > 79% indicates that it is part of the same subspecies (Meier-Kolthoff et al. 2013). Additionally, an ANI value of greater than 95% was proposed for species delineation (Konstantinidis and Tiedje 2005). For the environmental *Burkholderia* isolates, the ANI and DDH values have revealed that the three main species groups differ from each other, as the ANI values between the species groups were 90% and below. The phylogenetic analysis based on the 16S rRNA and *recA* gene sequences using the NJ algorithm indicated that the twelve environmental *Burkholderia* isolates formed three distinct phylogenetic lineages within the genus *Burkholderia* (Bcc group). The first group of seven isolates clustered with the type strain of *Burkholderia cepacia*, and the second group of two isolates clustered with the type strain of *Burkholderia vietnamiensis* (Figures 5.1 and 5.2). The third group of *Burkholderia*, which formed a novel group of isolates, was not related to the published type strains of Bcc within the 16S rRNA and *recA* analysis, suggesting that they represent novel species within Bcc (Figure 5.2). The ANI and *in silico* DDH values confirm this finding; the values were clearly

lower than the thresholds of 95 and 70%, respectively, which are generally accepted for species delineation. The novel group of isolates needs further analysis in order to develop a type strain definition and characterisation.

To date, the diversity within the genus *Paraburkholderia* is extensive (Sawana et al. 2014; Dobritsa et al. 2016; Dobritsa and Samadpour 2016; Dobritsa et al. 2017) and this study supported these findings. The phylogenetic analysis indicated that 44 jungle *Paraburkholderia* isolates formed three distinct taxonomic groups, and a further ten isolates clearly formed four distinct phylogenetic lineages within the genus *Paraburkholderia*. Only one group clustered within the type strains of *P. tropica*. The ANI and *in silico* DDH values were calculated between the group of *P. tropica* isolates and closely related type strains (*P. tropica*). The whole-genome sequence of *P. tropica* and 22 *P. tropica* strains showed ≥ 98% ANI and >78.9% *in silico* DDH, clearly above the recommended species thresholds (95% and 70%). The further groups of isolates (Group 2 and 3) and the additional isolates represented novel *Paraburkholderia* taxa and species. This suggests that they are novel taxa within *Burkholderia* sensu lato bacteria following the evolution of this bacteria over the last five years. Thus, this collection has a diversity of isolates with promising novel taxa that need further phenotypical, morphological and chemical analysis to define the type strains.

Symbiotic biological nitrogen fixation (BNF) process requires a specific communication process between the plant and its bacterial symbionts, and this complicated prosses includes the expression of nodulation (*nod*) and nitrogen fixation (*nif*, *fix*, and *fdx*) genes (De Meyer et al. 2016). Several N_2-fixing *Paraburkholderia* species, reclassified from *Burkholderia* to *Paraburkholderia*, have been reported, and include *Paraburkholderia tropica* (Reis et al. 2004), *Paraburkholderia unamae* (Caballero-Mellado et al. 2004), *Paraburkholderia xenovorans* (Goris et al. 2004), *Paraburkholderia mimosarum* (Chen et al. 2006), *Paraburkholderia silvatlantica* (Perin et al. 2006), *Paraburkholderia nodosa* (Chen et al. 2007) and *Paraburkholderia heleia* (Aizawa et al. 2010a). They were all isolated from various plant rhizosphere environments (e.g., maize, mimosas, sugar cane, and teosinte [a Mexican grass

type]) and soils. This phenotypic feature has been reported in multiple members of the *Paraburkholderia* as well as environmental *Burkholderia* species. For example, *Burkholderia vietnamiensis* was the first described N_2-fixing species from the Bcc (Gillis et al. 1995). Six isolates in this present study (two *Burkholderia vietnamiensis* and four *Paraburkholderia tropica*) isolated from plant roots showed the genomic potential to fix nitrogen. Despite the *nif* genes prediction, there was a limitation to clearly determine other N_2-fixing genes (*nif*, *fix*, and *fdx* genes). It would be worth using further bioinformatics tools to predict the N_2-fixing genes, and/or to determine active nitrogen fixation in these isolates using methods such as the acetylene reduction assay (Gillis et al. 1995).

5.4 Conclusion

In summary, comprehensive genotyping techniques, including 16S rRNA, *recA*, MLST and whole-genome-comparisons, were used to support the existence of novel taxa, species and strains of *Burkholderia*, *Paraburkholderia* and *Caballeronia* genera. The new proposed taxa and species are distinct but genetically related to other validly published *Burkholderia* (Bcc), *Paraburkholderia* and *Caballeronia* members. A limited number of jungle *Burkholderia* (two *B. vietnamiensis*) and *Paraburkholderia* (four *P. tropica*) were found to have nitrogen fixation genes and hence may offer agricultural benefits if characterized further. The addition of this collection within the *Burkholderiaceae* bacteria family will benefit evolutionary studies of *Burkholderia* sensu lato bacteria, environmental Bcc, *Paraburkholderia* and *Caballeronia* species, which all share a similar ecological niche with these new taxa and species.

Chapter 6. Antimicrobial screening of environmental *Burkholderia* and *Paraburkholderia* isolates

6.1 Introduction

Burkholderia bacteria are a valuable source of natural bioactive molecules (Liu and Cheng 2014; Haeckl et al. 2019; Kunakom and Eustáquio 2019). These natural products have biotechnological potential in agriculture and pharmaceuticals, which can impact human health (Singh et al. 2017). The ability of microorganisms to produce active compounds was first recognised in 1928 by Sir Alexander Fleming, leading to the discovery of penicillin (Ligon 2004). Since that time, the pharmaceutical industry and researchers have been deeply engaged in developing new strategies for the purification and production of antibiotics (Zerikly and Challis 2009). During this antimicrobial discovery period, it also became clear that certain antibiotic-producing microorganisms could produce multiple antibiotics, leading to the one strain many compounds (OSMAC) approach to discovery (Bode et al. 2002). OSMAC is a cultivation-based method in which a single strain is induced to produce a variety of bioactive metabolites by varying its growth conditions, such as nutrient content and temperature.

Such conventional methods for discovering novel antimicrobials, along with synthetic chemical approaches for modifying natural products, take considerable time and effort to develop. With the advent of the genomic era and whole genome sequencing, genome mining strategies were developed to improve the process of discovering natural products in bacteria. Nowadays, the combination of conventional bioactivity-guided isolation and genome-guided approaches offers a very powerful way to explore potential methods to encode and produce specialized metabolites (Katz and Baltz 2016).

The availability of genome sequence data and the power of genome-guided approaches has revealed the huge diversity of natural products that can be found within *Burkholderia* species making this genus an exciting new source of bioactive natural compounds. For example, *B.*

thailandensis E264 is a rich source of diverse compounds, such as betulinan, terferol, malleilactone, burkholderic acid, thailandepsins and burkholdacs (Liu and Cheng 2014). It also produces capistruin as peptide antibiotic. Meanwhile, several antifungal antibiotics are produced by the *Burkholderia cepacia* complex (Bcc) species, including cepaciamide, xylocandins, quinolinones, phenylpyrroles, pyrrolnitrin, phenazines and burkholdines (1097 and 1229) (Parke and Gurian-Sherman 2001; Vial et al. 2007; Sultan et al. 2008). Furthermore, various promising antibacterial compounds are produced from several *Burkholderia* spp., including gladiolin (Song et al. 2017), enacyloxin (Mahenthiralingam et al. 2011) and cepacin (Mullins et al. 2019).

Moreover, several reports have demonstrated the capacity of *Burkholderia* sensu lato genomes to encode multiple biosynthetic gene clusters (BGCs) of natural products. *Burkholderia* sensu lato have large and complex genomes ranging from 6 to 9 Mb in size and encoding multiple genes, giving them ecologically and metabolically diverse phenotypes (Kunakom and Eustáquio 2019). Ultimately, the importance of the *Burkholderia* genus as a rich and relatively untapped source of diverse natural products has considerable therapeutic and biotechnological relevance. This chapter investigated the potential of *Burkholderia* and *Paraburkholderia* to produce bioactive molecules and encode BGCs that may have medical and biotechnological benefits.

6.1.1 Aim and objectives

Chapter 5 assembled a novel collection of *Burkholderia* and *Paraburkholderia* isolates from the jungle environment in Borneo, and this chapter examined the ability of these bacteria to produce antimicrobials and encode specialized metabolites. A key research question behind the study was whether *Burkholderia* and *Paraburkholderia* have equivalent potential to produce bioactive molecules or encode specialized metabolite BGCs. This aim was investigated through the following objectives:

1. **Application of antimicrobial antagonism assays.** The environmental *Burkholderia* and *Paraburkholderia* genomes were screened for novel antimicrobial production using microbial antagonism assays, as a fast and inexpensive bioactivity screening approach.
2. **High-performance liquid chromatography (HPLC) analysis.** Secreted metabolites were extracted and analysed by HPLC to estimate the number of molecules and the identity of the compounds produced by the bioactive isolates.
3. **A genome mining approach using AntiSMASH.** This is a rapid and published software for specialized metabolite gene cluster identification, annotation and analysis in bacteria (Blin et al. 2013; Blin et al. 2019). AntiSMASH was used to identify and compare the BGCs encoded within the environmental *Burkholderia* and *Paraburkholderia* genomes to explore the potential of specialized metabolism within the jungle isolates.

6.2 Results

6.2.1 Antimicrobial production using antagonism assay

6.2.1.1 Jungle *Burkholderia* isolates produce anti-Gram-positive and antifungal compounds

The most striking result of this screening was that the 44 jungle *Paraburkholderia* and one *Caballeronia* isolates did not show antagonistic activity against the Gram-positive, Gram-negative or fungal microorganisms tested. Alterations in the environmental growth conditions, standard BSM-G media (Mahenthiralingam et al. 2011) in terms of pH (pH 5 or 7) or the incubation temperature (22°C or 30°C), did not result in antimicrobial activity being observed. In contrast, all the jungle *Burkholderia* isolates produced antimicrobial activity when grown on the BSM-G medium at pH 7 at 22°C and 30°C against MRSA (Gram-positive) and *C. albicans* (fungus), as follows.

Figures 6.1 and 6.2 show the antimicrobial activity of thirteen environmental *Burkholderia* isolates on BSM-G medium at pH 7 at 22°C and 30°C. Five of these isolates showed novel antimicrobial activity against MRSA and *C. albicans* under these initial screening conditions: *B. vietnamiensis* J17-1 (BCC1956), *B. vietnamiensis* J17-3 (no genome was available for this strain), *B. vietnamiensis* J17-4 (BCC1955), *Burkholderia* sp. J91-1 (BCC1965) and *Burkholderia* sp. J91-2 (BCC1964). Furthermore, greater antagonistic activity against the Gram-positive bacteria (MRSA) was observed at the lower temperature (22°C), with the zones of inhibition being much larger than those observed at 30°C (Figure 6.2 and 6.3).

Antifungal activity was also evaluated for these five bioactive isolates. The inhibition zones of the novel Bcc isolates (*Burkholderia* sp. J91-1 [BCC1965] and *Burkholderia* sp. J91-2 [BCC1964]) increased at 22°C compared to the antifungal activity at 30°C (Figure 6.1). However, *B. vietnamiensis* J17-1 (BCC1956), *B. vietnamiensis* J17-3 and *B. vietnamiensis* J17-4 (BCC1955) each had a similar zone of clearing against *C. albicans* at both temperatures

(Figure 6.1). All three *B. vietnamiensis* strains showed greater antifungal activity than the two novel *Burkholderia spp.* isolates, at both temperatures (Figure 6.1).

The antimicrobial activity of the thirteen jungle *Burkholderia* isolates was subsequently investigated at pH 5, and this change in growth condition showed interesting results in combination with growth temperature. Growth at pH 5 had a strong influence on the antibacterial activity at 22°C and 30°C (Figure 6.3; panel A). All thirteen environmental *Burkholderia* isolates displayed antibacterial activity against MRSA at pH 5. The seven previously negative jungle *B. cepacia* isolates all showed a limited amount of anti-Gram positive activity. This was consistent at both growth temperatures (22°C and 30°C), except for novel *Burkholderia* sp. J48 (BCC1957), which did not show antibacterial activity at 30°C. Figure 6.3 (A) shows the effect of low pH at 22°C in all thirteen jungle *Burkholderia* isolates, including the varied zone clearings among the active isolates. Low temperature also influenced the pigmentation of three isolates (*Burkholderia sp.* J47-3, *Burkholderia sp*. J86-1 and *Burkholderia sp.* J86-2) (Figure 6.4). Regarding antifungal activity, Figure 6.3 (panel B) and Figure 6.5 show that three of the thirteen isolates were active against *C. albicans*: *B. vietnamiensis* J17-1 (BCC1956), *B. vietnamiensis* J17-3 and *B. vietnamiensis* J17-4 (BCC1955).

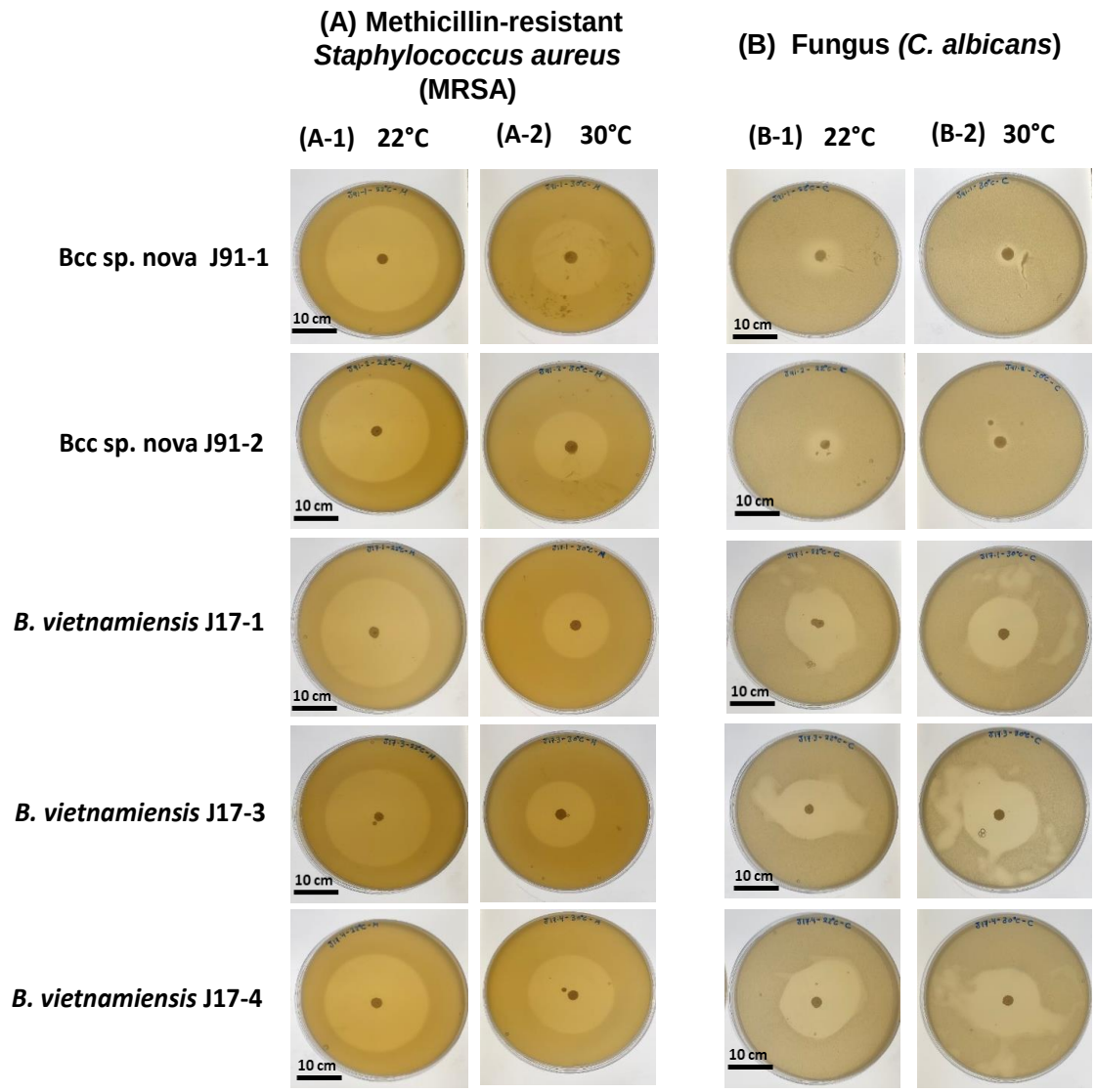


Figure 6.1 The effect of temperature on the antimicrobial activity of environmental *Burkholderia* isolates on BSM-G at pH 7. Panel (A) illustrates the activity of five environmental *Burkholderia* isolates against MRSA. A-1 and A-2 show the antibacterial activity at 22°C and 30°C, respectively. The lower temperature had a greater antibacterial effect than the higher temperature in all five *Burkholderia* isolates. Panel (B) shows the antifungal activity at 22°C (B-1) and 30°C (B-2). The lower temperature had a greater effect on the antifungal activity than the higher temperature in two isolates (*Burkholderia* sp. J91-1 and *Burkholderia* sp. J91-2). The antagonism assay shows the same clearing zones in all three *B. vietnamiensis* strains, indicating that antifungal activity was not affected by temperature. Antibacterial and antifungal activity defined by measuring the the zones of inhibition (mm), and *n*= 3 antagonism overlays of *Burkholderia* isolates against MRSA and *C. albicans*.

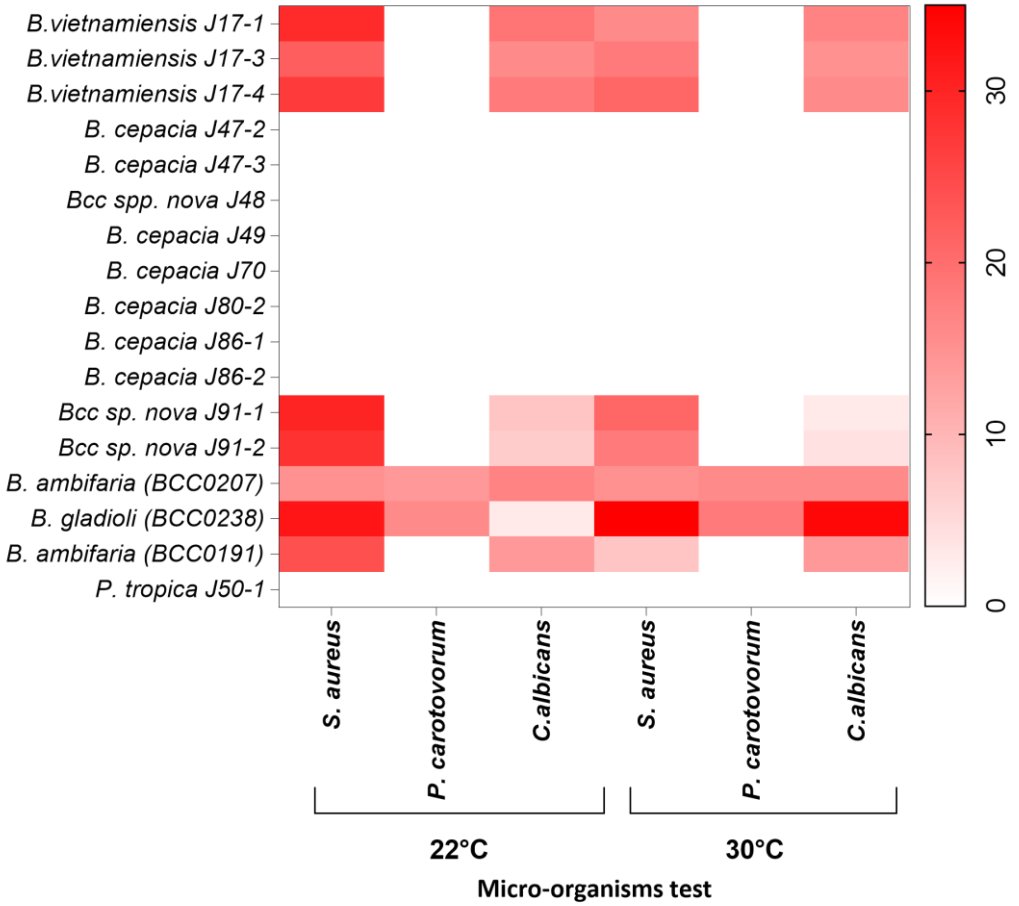


Figure 6.2 Antagonism assay for novel jungle *Burkholderia* isolates on BSM-G media at pH 7. The inhibition zones of thirteen environmental *Burkholderia* isolates were tested for antimicrobial activity at 22°C and 30°C. Three well-characterized antimicrobial *Burkholderia* producers were used as positive controls, and one environmental *Paraburkholderia* sp. was used as a negative control for non-antimicrobial producer bacteria. The antimicrobial activity was variable against the micro-organisms tested and in the various temperatures used. The antimicrobial activity was defined by the measurement of zone of inhibition in (mm) for each *Burkholderia* against methicillin-resistant *Staphylococcus aureus* (MRSA), *Pectobacterium carotovorum* and *Candida albicans*. The heatmap shows the mean of inhibition zone (Y axes on the right hand side), and n = 3 antagonism overlays was performed for each *Burkholderia*. Five environmental jungle *Burkholderia* strains showed antimicrobial activity. All five were active at both temperatures, and the lower temperature increased the activity of all antimicrobial producers against *S. aureus* and *C. albicans*. Among the antimicrobial producers, *B. ambifaria* (BCC0207) and *B. gladioli* (BCC0238) displayed variable antimicrobial activity against all tested micro-organisms, while *B. ambifaria* (BCC0191) showed antibacterial activity with gram-positive bacteria (*S. aureus*) and antifungal activity.

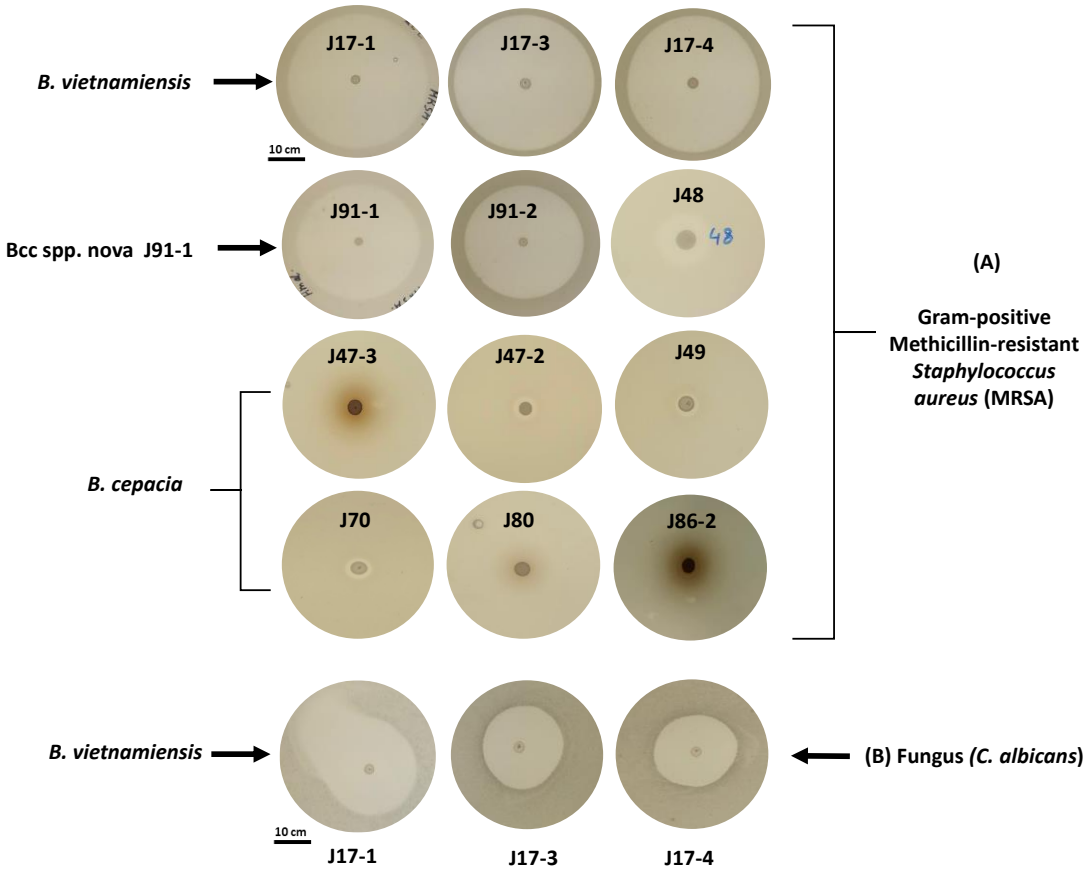


Figure 6.3 Antagonism assay of environmental *Burkholderia* isolates on BSM-G at 22°C and pH 5. (A) All 13 *Burkholderia* isolates displayed activity against MRSA (12 isolates shown in this figure, as well as *Burkholderia* sp. J86-1, which is shown in Figure 5). The zone of clearing (representing bacterial inhibition) varied among the 13 isolates. Five isolates had large zones, while the other eight were smaller. (B) The three strains of *B. vietnamiensis* (J17-1, J17-3 and J17-4) displayed greater antifungal activity at this pH. Antibacterial and antifungal activity defined by measuring the the zones of inhibition (mm), and *n*= 3 antagonism overlays of *Burkholderia* isolates against MRSA and *C. albicans*.

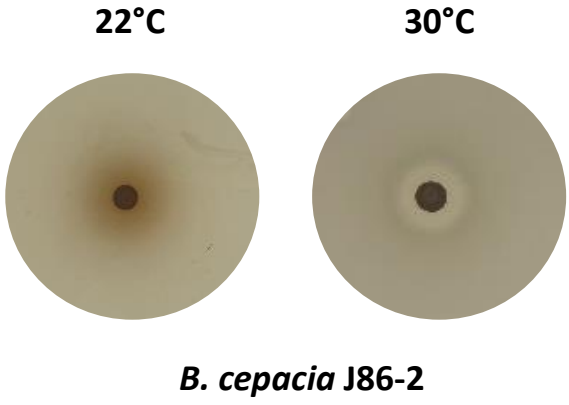


Figure 6.4 The effect of low temperature (22°C) and pH 5 on *B. cepacia* J86-2 (BCC1962). Dark brown pigmentation was observed under the lower pH and temperature conditions. In addition, the bacterial inhibition zone at 30°C was bigger than at 22°C. Antibacterial activity defined by measuring the the zones of inhibition (mm), and *n*= 3 antagonism overlays of *B. cepaia* J86-2 against MRSA.

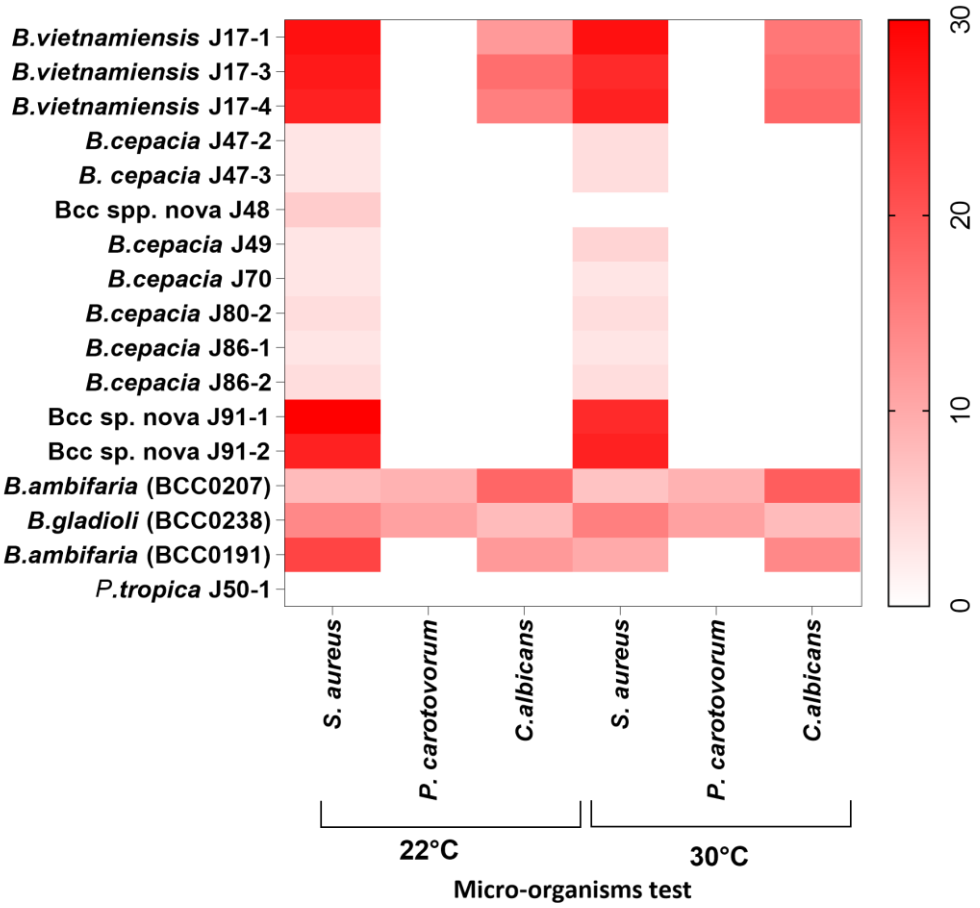


Figure 6.5 Antagonism assay for novel jungle *Burkholderia* isolates on BSM-G media (pH 5) at 22°C and 30°C. In total, seventeen bacteria were tested for antimicrobial activity, as depicted in Figure 1. Thirteen were environmental *Burkholderia* spp./strains (seven *B. cepacia*, three Bcc sp. nova and three *B. vietnamiensis*). Further three bacteria were well-characterized antimicrobial *Burkholderia* producers, which were used as positive controls; and one was an environmental *Paraburkholderia* sp., which was used as a negative control for non-antimicrobial producer bacteria. The antimicrobial activity was defined by the measurement of zone of inhibition in (mm) for each *Burkholderia* against methicillin-resistant *Staphylococcus aureus* (MRSA), *Pectobacterium carotovorum* and *Candida albicans*. The heatmap shows the mean of inhibition zone (Y axes on the right hand), and $n = 3$ antagonism overlays was performed for each *Burkholderia*. Variable antimicrobial activity was found against the micro-organism tests and in the various temperatures. The biggest inhibition zones were observed in five isolates (Bcc sp. nova J91-1, Bcc sp. nova J91-2, *B. vietnamiensis* J17-1, *B. vietnamiensis* J17-3 and *B. vietnamiensis* J17-4). The three environmental *B. vietnamiensis* strains showed antimicrobial activity against MRSA and fungus. The other two isolates showed strong antibacterial activity against MRSA. All five isolates were active at both temperatures and showed the same results. The other eight *Burkholderia* isolates were active at both temperatures except *Burkholderia* sp. J48, which was not active at 30°C. Among the antimicrobial producers, *B. ambifaria* (BCC0207) and *B. gladioli* (BCC0238) both showed variable antimicrobial activity against all micro-organism tests, while *B. ambifaria* (BCC0191) showed antibacterial activity with gram-positive bacteria (*S. aureus*) and antifungal activity.

6.2.2 Metabolites produced under different growth conditions in bioactive Bcc species

The antimicrobial antagonism assays above were performed for for growth conditions pH 5 and 7, and temperatures 22 and 30°C. Based on this, combinations of the two different growth conditions were selected for HPLC analysis to screen the ability of *Burkholderia* isolates to produce metabolites which could account for the antimicrobial activity. In general, metabolite peaks were seen at HPLC retention times ranging from 5.53 to 7.88 minutes. The selected growth conditions for HPLC analyses were: (i) pH 5 BSM-G at 30°C and (ii) pH 7 BSM-G at 22°C (Table 6.1) (both were grown for 72 h). Under these conditions, single and multiple peaks were observed in all bioactive Bcc species, except novel *Burkholderia* sp. J48 (BCC1957). The peak associated with the polyyne compound cepacin was observed in most of the environmental Bcc strains, such as *B. vietnamiensis* J17-1 (BCC1956), *B. cepacia* J70 (BCC1960) and Bcc sp. nova J91-1 (BCC1965) (Figure 6.6, Table 6.1). The cepacin peak, with a retention time of 5.89 minutes, was associated with the most Bcc isolates, especially at pH 7 and 22°C (Figure 6.6, Table 6.1). The peak for the compound pyrrolnitrin was found in most of the *B. cepacia* strains, such as J47-2 (BCC1959), J70 (BCC1960 and J86-1 (BCC1962), especially at pH 7 and 22°C (Figure 6.6, Table 6.1). The retention time of the known antifungal compound pyrrolnitrin, produced by *B. cepacia* strains, was approximately 7.63 under the HPLC conditions used (Figure 6.6). Multiple Bcc isolates produced unknown multiple compounds at both conditions, especially in *B. cepacia* strains [(BCC1967), J70 (BCC1960) and J80 (BCC1963)] (Figure 6.6 and Table 6.1).

The comparison of growth conditions from pH 5 and 30°C, to pH 7 and 22°C, confirmed that growing the bacteria in BSM-G at pH 7 and 22°C was more effectivein eliciting the production of metabolites observable by HPLC (Table 6.1). For the first tested condition at pH5 and 30°C, of the 13 isolates, 6 were cepacin positive, 4 were pyrrolnitrin positive, and 7 produced novel unknown compounds. Changing the growth conditions to pH 7 and 22°C altered the results, with 9 becoming cepacin positive, 7 being pyrrolnitrin positive, and 11 of the 13 showing the

production of multiple unknown compounds. Figure 6.6 shows the predicted known and unknown compounds observed by HPLC analysis of metabolite extracts from three different environmental Bcc isolates. The HPLC results for the two different growth condtions are presented in the Table 6.1 below.

Table 6.1 The number of predicted peaks for environmental Bcc species (*n*=13) on BSM-G at two growth conditions.

The HPLC predicted peaks at pH 5 and 30°C			
ID	**Cepacin RT***	**Pyrrolnitrin RT***	**Novel metabolites (RT)**
***B. vietnamiensis* J17-1 (BCC1956)**	5.88	-	-
***B. vietnamiensis* J17-3**	5.88	-	-
***B. vietnamiensis* J17-4 (BCC1955)**	5.89	-	-
***B. cepacia* J47-2 (BCC1959)**	-	7.62	6 peaks (5.50, 5.80, 6.30, 6.52, 7.20 and 7.61)
***B. cepacia* J47-3 (BCC1958)**	-	-	2 peaks (6.30, 7.21 and 7.60)
Bcc sp. nova J48-2 (BCC1957)	-	-	3 peaks (6.29, 7.20 and 7.60)
***B. cepacia* J49 (BCC1967)**	-	7.62	5 peaks (5.50, 6.30, 6.52, 7 and 7.20)
***B. cepacia* J70 (BCC1960)**	-	7.63	2 peaks (6.30 and 7.20)
***B. cepacia* J80 (BCC1963)**	-	7.62	5 peaks (5.50, 6.30, 6.52. 6.90 and 7.20)
***B. cepacia* J86-1 (BCC1962)**	-	7.63	5 peaks (5.50, 6.30, 6.55, 6.90 and 7.24)
***B. cepacia* J86-2 (BCC1961)**	5.85	7.63	5 peaks (5.50, 6.30, 6.55, 6.90 and 7.24)
Bcc sp. nova J91-1 (BCC1965)	5.89	-	-
Bcc sp. nova J91-2 (BCC1964)	-	-	-
The HPLC predicted peaks at pH 7 and 22°C			
ID	**Cepacin RT***	**Pyrrolnitrin RT***	**Novel metabolites (RT)**
***B. vietnamiensis* J17-1 (BCC1956)**	5.88	-	1 peak (~6.58)
***B. vietnamiensis* J17-3**	5.88	-	-
***B. vietnamiensis* J17-4 (BCC1955)**	5.89	-	1 peak (~6.58)
***B. cepacia* J47-2 (BCC1959)**	5.90	7.63	5 peaks (5.66, 6.30, 6.53, 6.88 and 7.22)
***B. cepacia* J47-3 (BCC1958)**	5.90	7.63	5 peaks (5.66, 6.30, 6.53, 6.88 and 7.22)
Bcc sp. nova J48-2 (BCC1957)	-	-	-
***B. cepacia* J49 (BCC1967)**	-	7.62	5 peaks (5.65, 6.30, 6.53, 6.90 and 7.20)
***B. cepacia* J70 (BCC1960)**	5.89	7.63	6 peaks (5.40, 5.65, 6.26, 6.53, 6.87 and 7.20)
***B. cepacia* J80 (BCC1963)**	-	7.62	4 peaks (5.65, 6.52. 6.90 and 7.20)
***B. cepacia* J86-1 (BCC1962)**	-	7.62	2 peaks (5.65, 6,52)
***B. cepacia* J86-2 (BCC1961)**	5.89	7.63	4 peaks (5.66, 6.55, 6.89 and 7.20)
Bcc sp. nova J91-1 (BCC1965)	5.89	-	2 peaks (6.95 and 7.40)
Bcc sp. nova J91-2 (BCC1964)	5.89	-	2 peaks (6.95 and 7.40)

***RT=** retention time (mins)

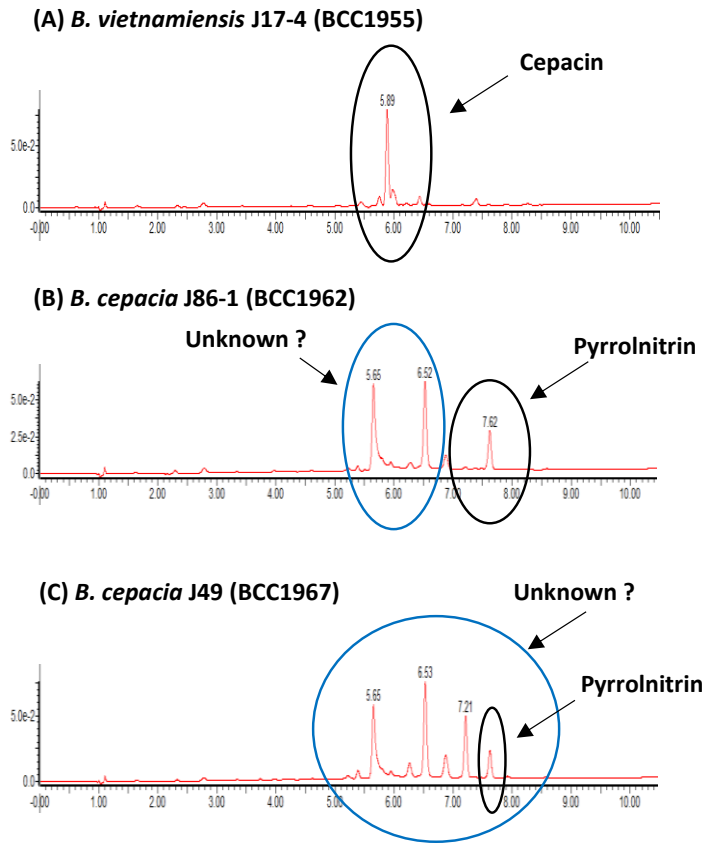


Figure 6.6 Examples of HPLC profiles of three different environmental Bcc isolate extracts from bacteria grown at pH 7 and 22°C. (A) Cepacin peak identified in *B. vietnamiensis* J17-4 (BCC1955). (B) Three different peaks were observed in *B. cepacia* J86-1 (BCC1962) extracts; two peaks are novel unknown peaks and the third peak is for pyrrolnitrin. (C) Five different novel unknown peaks were identified in *B. cepacia* J49 (BCC1967) and pyrrolnitrin.

6.2.3 Genome mining approach using antiSMASH software (5.0.0)

6.2.3.1 The diversity and capacity for antibiotic production of novel environmental *Burkholderia*

Table 6.2 presents the average genome size and predicted number of BGCs per Mb of genome in the *Burkholderia* strains examined. The stoical comparisons were performed using unpaired t tests by GraphPad prism. The average genome size of the twelve environmental *Burkholderia* and 20 *Burkholderia* references were 7.98 ± 0.82 Mb and 7.4 ± 0.8 Mb, respectively. No significant differences between the genomes size between environmental and references *Burkholderia* was observed, as the *P* value was greater than < 0.05 based on unpaired t test performed by GraphPad prism. The predicted number of BGCs in twelve

environmental *Burkholderia* were 15.5 ± 2.6), and in 20 *Burkholderia* references were 18.8 ± 8.7). The predicted BGCs in references were significantly much higher than the BGCs in environmental *Burkholderia* genomes (*P* value was 0.0002). Furthermore, despite similar genome lengths, the mean BGC difference per Mb in the reference *Burkholderia* species was much higher than the BGC difference per Mb in environmental *Burkholderia* ($P < 0.05$). The measured average of BGCs per Mb of the genome were 1.9 ± 0.2 per Mb in environmental genomes and 2.5 ± 1 per Mb in references *Burkholderia* genomes.

In addition, 21 and 11 predicted BGC types were observed in the *Burkholderia* references and jungle strains, respectively. The 21 predicted BGC types in the *Burkholderia* references were: terpene, arylpolyne, phosphonate, homoserine lactone, non-ribosomal peptide synthetase (NRPS), polyketide synthase (PKS), bacteriocin/RiPP, hybrid PKS/NRPS, heterocyst glycolipid synthase-like PKS (hglE-KS), other/unknown (a specialized metabolite that did not fit into any other BGC types), tRNA-dependent cyclodipeptide synthases (CDPS), polyketide synthases lacking acyltransferase (transAT-PKS), phenazine, betalactam, butyrolactone, lanthipeptide, thiopeptide, betalactone, siderophore, ladderane and ectoine. Only the first eleven types (terpene, arylpolyne, phosphonate, homoserine lactone, NRPS, PKS, bacteriocin/RiPP, hybrid PKS/NRPS, hglE-KS, other/unknown and CDPS) were also predicted in the environmental *Burkholderia* strain genomes. Figure 6.7 shows the BGC types and lengths predicted for all of the *Burkholderia* strains examined. The types and lengths of the BGCs in the collection of *Burkholderia* strains varied considerably, with reference genomes for species such as *B. gladioli, B. plantarii, B. oklahomensis* and *B. ambifaria,* showing more diversity and overall BGC content than the jungle strains, none of which belonged to these species (Figure 6.7). For all *Burkholderia* examined NRPS and terpenes were the first and second most abundant BGC classes predicted.

Table 6.2 Genome sizes, number of predicted BGCs and average predicted BGCs per 1 Mb in the environmental and reference *Burkholderia*.

Environmental *Burkholderia* (*n*=13)			
Strain ID	**Genome size (Mb)**	**# BGCs**	**BGCs per 1 Mb**
***B. vietnamiensis* J17-1 (BCC1956)**	6.84	12	1.75
***B. vietnamiensis* J17-4 (BCC1955)**	6.86	11	1.60
***B. cepacia* J47-2 (BCC1959)**	8.26	16	1.93
***B. cepacia* J47-3 (BCC1958)**	8.51	17	1.99
***B. cepacia* J49 (BCC1967)**	8.51	17	1.99
***B. cepacia* J70 (BCC1960)**	8.67	18	2.07
***B. cepacia* J80-2 (BCC1963)**	8.85	18	2.29
***B. cepacia* J86-1 (BCC1962)**	8.82	18	2.04
***B. cepacia* J86-2 (BCC1961)**	8.82	18	2.04
Bcc sp. nova J48-2 (BCC1957)	7.34	13	1.77
Bcc sp. nova J91-1 (BCC1965)	7.17	14	1.95
Bcc sp. nova J91-2 (BCC1964)	7.17	14	1.95
Mean and standard error	**7.98 ± 0.8**	**15.5 ± 2.6**	**1.9 ± 0.2**
Reference *Burkholderia* (*n*=20)			
Strain ID	**Genome size (Mb)**	**# BGCs**	**BGCs per 1 Mb**
***B. territorii* LMG 28158 (LK023503)**	6.89	12	1.74
***B. cepacia* ATCC 25416 (U96927)**	8.54	17	1.99
***B. seminalis* LMG 24067**	7.64	18	2.35
***B. lata* 383 (CP000151)**	8.61	14	1.62
***B. vietnamiensis* LMG 10929 (AF097534)**	6.88	13	1.88
***B. latens* R-5630 (AM747628)**	7.07	13	1.83
***B. dolosa* LMG 18943 (JX986970)**	6.36	9	1.41
***B. multivorans* ATCC BAA-247**	6.95	14	2.01
***B. diffusa* R-15930 (AM747629)**	7.09	12	1.69
***B. ambifaria* AMMD (AF043302)**	7.47	21	2.81
***B. stabilis* LMG 14294 (AF097533)**	8.46	14	1.65
***B. stagnalis* LMG 28156 (LK023502)**	8.09	31	3.83
***B. pyrrocinia* LMG 14191 (U96930**	7.90	17	2.15
***B. glumae* LMG 2196 (U96931)**	6.82	25	3.66
***B. gladioli* NBRC 13700 LMG 2216**	8.71	48	5.51
***B. plantarii* LMG 9035 (U96933)**	8.02	25	3.11
***B. thailandensis* E264 (U91838)**	6.67	20	2.99
***B. mallei* ATCC 23344 (AF 110188)**	5.83	17	2.91
***B. oklahomensis* C6786 (DQ108388)**	6.94	22	3.17
***B. cenocepacia* H111**	7.71	15	1.94
Mean and standard error	**7.4 ± 0.8**	**18.85 ± 8.7**	**2.5 ± 1**

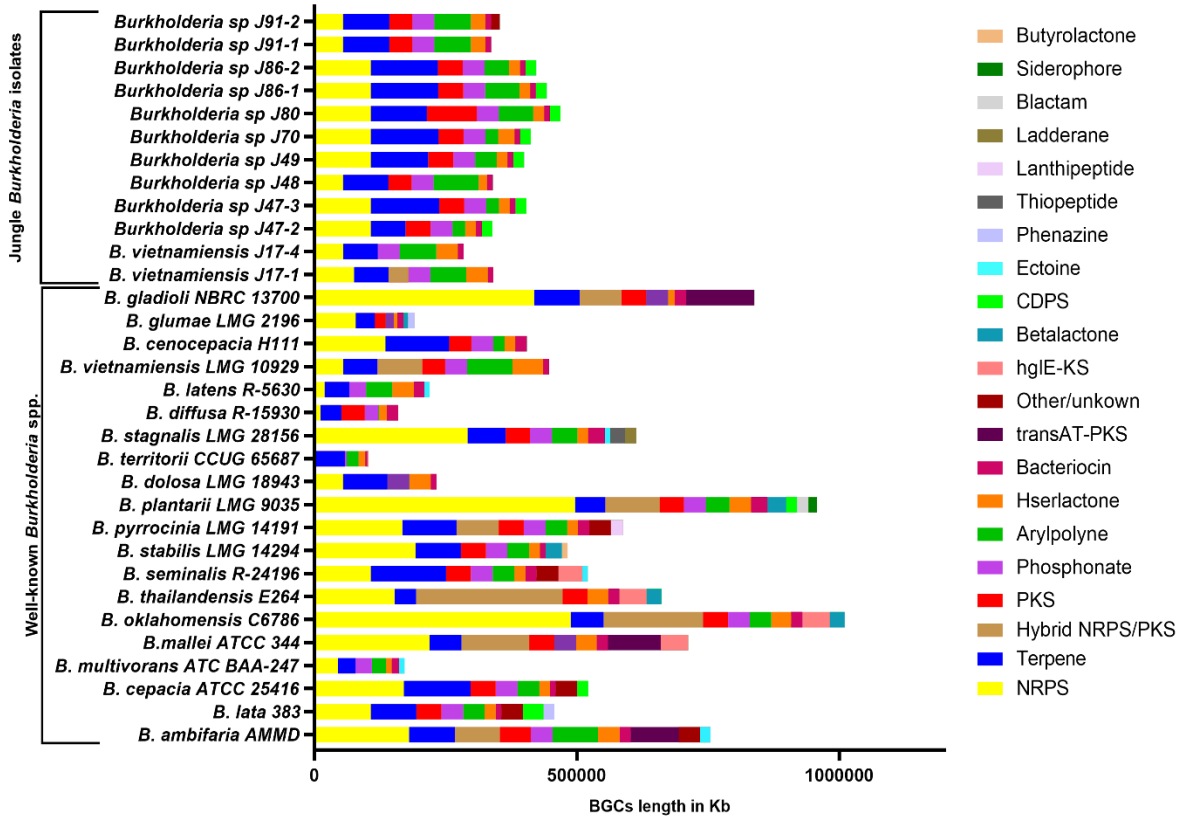


Figure 6.7 BGC types and lengths in Kb of 32 *Burkholderia* species/strains, as predicted by antiSMASH. The y-axis presents the *Burkholderia* IDs, and the x-axis presents the BGC lengths in kb. The 21 clusters found in all 32 *Burkholderia* species/strains are indicated on the right-hand key using different colours, and they are ordered from the least predicted metabolite BGC class (top) to the most predicted (bottom).

6.2.3.2 Understanding the diversity and capacity for antibiotic production of novel jungle *Paraburkholderia*

Table 6.3 shows the average genome and BGC sizes in the *Paraburkholderia* species/strains. The genome lengths are significantly different ($P < 0.05$), P value = <0.0001, as the jungle *Paraburkholderia* genomes are larger than the references *Paraburkholderia*. The average of jungle *Paraburkholderia* genomes is 8.4 ± 0.5 compared with 8.5 ± 1.02 for the 20 reference *Paraburkholderia* genomes. The BGC length averages were 14.1 ± 1.9 and 12.4 ± 3.4 for the 44 jungle *Paraburkholderia/ Caballeronia* genomes, and the 20 *Paraburkholderia* references, respectively. This showed that the length of BGCs encoded into jungle *Paraburkholderia* were significantly higher than the references (P value = 0.0017; $P < 0.05$). Moreover, the average

of encoded BGCs per Mb are higher in jungle genomes (1.7 ± 0.2 per Mb), as the $P < 0.05$. The average of BGCs per Mb in references are 1.4 ± 0.4.

Table 6.3 Genome sizes, number of predicted BGCs and average predicted BGCs per 1 Mb in the jungle and reference *Paraburkholderia*

Jungle *Paraburkholderia/Caballeronia* spp./strains (*n*=45)			
Strain ID	**Genome size (Mb)**	**# BGCs**	**BGCs per 1 Mb**
***P. tropica* J16 (BCC1937)**	7.96	13	1.6
***P. tropica.* J23-1(BCC1948)**	8.5	14	1.6
***P. tropica* J19-2 (BCC1951)**	8.46	14	1.6
***P. tropica* J19-1 (BCC1950)**	8.46	17	2
***P. tropica* J8-1 (BCC1949)**	8.24	13	1.57
***Paraburkholderia* sp. nova J8-2 (BCC1943)**	9.43	14	1.48
***P. tropica* J26 (BCC1933)**	8.36	14	1.67
***P. tropica* J27 (BCC1945)**	8.51	14	1.64
***P. tropica* J6 (BCC1918)**	7.96	13	1.63
***P. tropica* J1-2 (BCC1936)**	8.56	14	1.63
***Paraburkholderia* sp. nova J35-1 (BCC1932)**	8.61	13	1.5
***P. tropica* J1-1 (BCC1935)**	8.55	14	1.63
***P. tropica* J23-3 (BCC1946)**	8.51	14	1.64
***P. tropica* J23-2 (BCC1947)**	8.52	14	1.64
***P. tropica* J24 (BCC1934)**	8.36	13	1.55
***Paraburkholderia* sp. nova J7 (BCC1912)**	9	14	1.55
***Paraburkholderia* sp. nova J10-1 (BCC1954)**	9.11	13	1.42
***Paraburkholderia* sp. nova 10-2 (BCC1942)**	9.27	16	1.72
***Paraburkholderia* sp. nova J15-1 (BCC1939)**	8.89	14	1.57
***Paraburkholderia* sp. nova J11-1 (BCC1941)**	8.83	14	1.58
***Paraburkholderia* sp. nova J11-2 (BCC1940)**	9.16	15	1.63
***Paraburkholderia* sp. nova J12 (BCC1953)**	6.92	19	2.74
***P. tropica* J45-2 (BCC1928)**	8.08	13	1.60
***Paraburkholderia* sp. nova J42-2 (BCC1930)**	7.77	13	1.67
***Paraburkholderia* sp. nova J15-2 (BCC1938)**	8.50	15	1.76
***P. tropica.* 50-1 (BCC1927)**	8.29	13	1.56
***Paraburkholderia* sp. nova J41 (BCC1931)**	7.01	11	1.56
***P. tropica* J50-4 (BCC1924)**	8.31	13	1.56
***P. tropica* J50-2 (BCC1926)**	8.28	13	1.57
***P. tropica* J50-3 (BCC1925)**	8.30	13	1.56
***P. tropica* J62 (BCC1923)**	8.23	13	1.57
***Paraburkholderia* sp. nova J63 (BCC1922)**	8.74	20	2.28
***Paraburkholderia* sp. nova J74 (BCC1916)**	8.83	14	1.58
***Paraburkholderia* sp. nova J67 (BCC1921)**	9.34	14	1.49
***P. tropica.* J88 (BCC1911)**	8.53	14	1.64
***Paraburkholderia* sp. nova J69-1 (BCC1919)**	9.03	17	1.8
***Paraburkholderia* sp. nova J72 (BCC1917)**	8.72	16	1.83
***Paraburkholderia* sp. nova J69-2 (BCC1920)**	9.03	18	1.99
***Paraburkholderia* sp. nova J75-1 (BCC1915)**	8.57	15	1.75
***Paraburkholderia* sp. nova J76 (BCC1913)**	7.07	13	1.83
***Paraburkholderia* sp. nova J75-2 (BCC1914)**	8.74	15	1.71
***P. tropica* J92 (BCC1910)**	8.32	13	1.6
***Paraburkholderia* sp. nova J94 (BCC1909)**	8.51	13	1.5
***Caballeronia sp*. nova J97**	7.78	9	1.2
Mean and standard error	8.4 ± 0.5	14.1 ± 1.9	1.7 ± 0.2
Reference *Paraburkholderia* spp. (*n*=20)			
Strain ID	**Genome size (Mb)**	**#BGCs**	**BGCs per Mb**
***P. ginsengiterrae* DCY85 ^{T}**	8.5	8	0.9
***P. insulsa* LMG 28183 ^{T}**	9.2	13	1.41
***P. sediminicola* LMG 24238 ^{T}**	10.7	13	1.21
***P. tropica* LMG 22274 ^{T}**	8.58	12	1.39
***P. heleia* NBRC 101817 ^{T}**	8	16	2
***P. rhynchosiae* LMG 27174 ^{T}**	8.02	11	1.36
***P. silvatlantica* SRCL-318 ^{T}**	8.8	13	1.47

***P. bannensis* NBRC 103871^T**	7.22	11	1.52
***P. tuberum* WSM 4176^T**	9.06	9	0.99
***P. ferrariae* NBRC 106233^T**	7.93	15	1.89
***P. rhizosphaerae* LMG 29544^T**	7.50	9	1.2
***P. caledonica* NBRC 102488^T**	7.28	10	1.37
***P. graminis* LMG 18924^T**	7.47	9	1.20
***P. nodosa* DSM 21604^T**	9.62	16	1.66
***P. kururiensis* KP23^T**	7.52	10	1.32
***P. phytofirmans* PsJN LMG 22146^T**	8.21	11	1.33
***P. fungorum* ATCC BAA-463^T**	8.99	11	1.22
***P. xenovorans* LB400^T**	9.66	11	1.13
***P. oxyphila* NBRC 105797^T**	10.59	18	1.69
***P. mimosarum* LMG 23256^T**	8.46	21	2.4
Mean and standard error	8.5 ± 1.02	12.4 ± 3.4	1.4 ± 0.4

Thus, there is a significant difference in the BGC length between the jungle and reference *Paraburkholderia spp.*, indicating that the BGC content was higher in the jungle genomes than the reference genomes. The following 13 BGC types were predicted in the jungle *Paraburkholderia*: terpene, arylpolyne, phosphonate, hserlactone, non-ribosomal peptide synthetase (NRPS), polyketide synthase (PKS), bacteriocin/RiPP, betalactone, siderophore, ladderane, microviridin, PPY-like pyrone cluster (PpyS-KS) and hybrid PKS/NRPS. The same types were found in the reference *Paraburkholderia*, as well as an additional one called ectoine found in *P. mimosarum* (Table 6.3, Figure 6.8). Terpene biosynthesis was the most abundant BGC class, and its length in the *Paraburkholderia* species/strains is roughly consistent (Figure 6.8). NRPS and arylpolyne were the second and third most common classes, while ection is produced by only *P. oxyphila* NBRC 105797^T. Interestingly, the BGCs of PpyS-KS was encoded only in *P. tropica* strains (Figure 6.8).

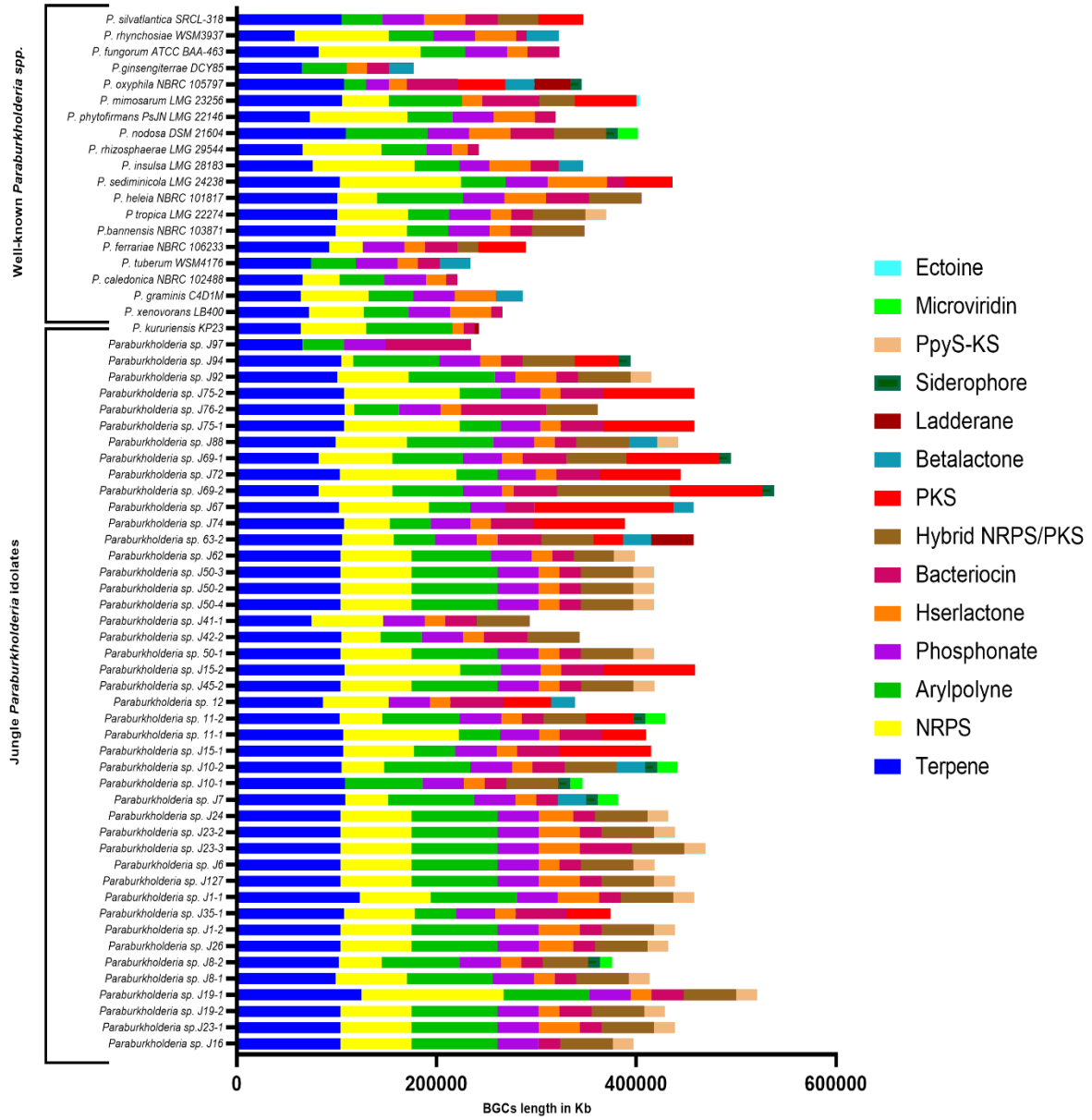


Figure 6.8 BGC types and lengths in Kb of 65 *Paraburkholderia* species/strains, as predicted by antiSMASH. In total, 44 environmental *Paraburkholderia* species/strains, one *Caballeronia* nov sp., and 20 reference *Paraburkholderia* species were run through antiSMASH to predict BGC types and lengths in kb. The y-axis presents the *Paraburkholderia* IDs, and the x-axis presents the BGC lengths in Kb. The 14 cluster types found on the 65 *Paraburkholderia* species/strains, including one *Caballeronia* nov sp, are indicated on the right-hand key using different colours, and they are ordered from the most produced class (bottom) to the least (top).

6.2.3.3 *Burkholderia* species encode more BGCs than *Paraburkholderia* species

Paraburkholderia genomes were larger than *Burkholderia* genomes based on (p < 0.0001). The average genome sizes of the *Burkholderia* and *Paraburkholderia* isolates were 7.6 ± 0.8, 8.5 ± 0.7, P <0.0001. The average number of predicted BGCs in *Burkholderia* genomes was

higher than *Paraburkholderia* genomes (17.6 ± 7.5, 13.55 ± 2.5, *P* <0.0001). When the overall genome size was considered, the average predicted BGCs per Mb in the *Burkholderia* genomes was also higher than in the *Paraburkholderia* genomes (2.3 ± 0.8, 1.6 ±0.3 per Mb, *P* <0.0001).

The overall number of BGCs predicted in the 96 genomes indicated that twelve BGCs types were shared between *Burkholderia* and *Paraburkholderia* (terpene, arylpolyne, phosphonate, hserlactone, NRPS, PKS, bacteriocin/RiPP, Hybrid PKS/NRPS, betalactone, siderophore, ladderane and ectoine). Nine BGCs types were predicted in *Burkholderia* genomes (hglE-KS, CDPS, transAT-PKS, phenazine, betalactam, butyrolactone, lanthipeptide, thiopeptide and other/unknown). Microviridin and PpyS-KS gene clusters were unique for *Paraburkholderia* genomes. *Burkholderia* genomes encoded a total of 21 cluster types while forteen cluster types were encoded in *Paraburkholderia* genomes. This indicated that the overall diversity of predicted cluster types encoded in *Burkholderia* genomes was greater than that for *Parburkholderia* (Table 6.4).

Table 6.4 Number and types of clusters in the 97 *Burkholderia* and *Paraburkholderia* genomes, as predicted by antiSMASH.

Cluster types predicted by antiSMASH	*Burkholderia* (*n*= 32)	*Paraburkholderia* (*n*=65)
1. Terpene	32	65
2. Arylpolyne	27	62
3. Phosphonate	31	62
4. Hserlactone	32	61
5. NRPS	32	58
6. PKS	26	20
7. Bacteriocin/RiPP	32	64
8. Hybrid PKS/NRPS	9	40
9. hglE-KS	15	0
10. Other/unknown	13	0
11. CDPS	10	0
12. transAT-PKS	3	0
13. Phenazine	2	0
14. Betalactam	1	0
15. Butyrolactone	1	0
16. Lanthipeptide	1	0
17. Thiopeptide	1	0
18. Betalactone	5	12
19. Siderophore	1	11
20. Ladderane	1	3
21. Ectoine	5	1
22. Microviridin	0	7
23. PpyS-KS	0	22

***Shared BGCs types highlighted in green**

6.3 Discussion

The key findings of this study examining strains from the jungle confirm that the *Burkholderia* genus is a promising source of bioactive compounds. Significantly, no evidence was found indicating that *Paraburkholderia* has the ability to produce active antimicrobiol compounds under the *in vitro* growth condition screened. Overall, the bioactivity-guided, HPLC analysis and genomic-guided approaches revealed that all thirteen environmental Bcc species and strains were antimicrobial producers. These findings are consistent with much of the current literature, which shows *Burkholderia* produce an array of natural products (Kunakom and Eustáquio 2019).

Kunakom and Eustáquio (2019) also analysed a set of publicly available *Burkholderia* and *Paraburkholderia* genomes using antiSMASH and also used other published results about BGCs in *Burkholderia*. Their findings indicate that the *Burkholderia* genus has a large number of BGCs compared to *Paraburkholderia*. In total, 55% of the compounds were identified using the bioactivity-guided method and 32% using genome mining methods. For the bioactivity of the analysed compounds, 39% of the identified compounds were also antimicrobial, 25% were toxic compounds, 23% were unknown or other compounds and 13% were functional compounds (7% iron chelating, 6% swarming and biofilm formation) (Kunakom and Eustáquio 2019). This PhD provides further evidence supporting Kunakom and Eustáquio's (2019) review, specifically adding a unique collection of novel environmental *Burkholderia* and *Paraburkholderia* bacteria to those available in public genome databases.

The antiSMASH analysis of the *Burkholderia* and *Paraburkholderia* clearly demonstrated that *Burkholderia* spp. genomes yield more predicted BGCs than *Paraburkholderia* species. In contrast, it showed that *Paraburkholderia* genomes are larger than *Burkholderia* genomes by nearly 1 Mb, showing that while they likely encode more genes, these are not specifically specialized metabolite classes predicted by antiSMASH. This was partly supported by the bioactivity data which indicated that *Paraburkholderia* bacteria did not produce any bioactive

compounds under the conditions screened in the present study. This was also evident in literature analysis which characterized very few *Paraburkholderia* associated natural products, such as *P. xenovorans* LB400 (Kunakom and Eustáquio 2019). However, three *Paraburkholderia* spp. were active against Gram-negative plant pathogens (*Dickeya solani, Pectobacterium carotovorum, Pseudomonas savastanoi, Pseudomonas syringae,* and *Rhizobium radiobacter*) (Webster et al. 2019). Beyond antiSMASH further genomic function predictions using software, such as Kyoto encyclopedia of genes and genomes (KEGG) or clusters of orthologous groups (COGs), could be used to investigate if the *Paraburkholderia* genomes encode any promising specialized metabolites (Tatusov et al. 1997; Kanehisa and Goto 2000).

Regardless of the genome length, it is important to note that the number of putative BGCs varies and is not necessarily correlated with genome size in *Burkholderia* and *Paraburkholderia*, with multiple factors such as the specific species and strain analysed making a difference (Kunakom and Eustáquio 2019). For example, the same number of putative BGC clusters (*n* = 17) were predicted in two *Burkholderia* species with different genome sizes (*B. mallei* ATCC 23344 with 5.83 Mb and *B. cepacia* ATCC 25416^T with 8.54 Mb). This indicates that the small genome of the former encodes much more than the larger genome of the latter. Specifically, *B. mallei* ATCC 23344 analysis yielded 2.91 BGCs per Mb, while *B. cepacia* ATCC 25416 showed the presence of 1.99 BGCs per Mb. Thus, in some cases small genomes can encode more specialized metabolite pathways than large genomes, indicating that genome size might not be linked to the number of BGCs.

The ability of *Burkholderia cepacia* complex members to produce cepacin and pyrrolnitrin has been reported previously (Sultan et al. 2008; Mullins et al. 2019). It has been found that the environmental *Burkholderia* in the present study produced cepacin, pyrrolnitrin and multiple novel metabolites**.** Since some of these are from novel members of the Bcc, it would be worth characterizing the cepacin and pyrrolnitrin pathways and compare them with the published pathways. Identifying and characterising the pathways of novel metabolites found in this study

would also lead to exploration of new antimicrobial compounds that could help combat the current antimicrobial resistance issue.

6.4 Conclusion

To understand the diversity and capacity for antibiotic production by *Burkholderia* and novel *Paraburkholderia* species isolated from Bornean jungle samples, genome- and bioactivity based approaches were conducted. The *Burkholderia* strains examined were all bioactive, but no evidence for the ability of *Paraburkholderia* to produce antimicrobial compounds was observed under the conditions tested. Genome mining showed that the capacity of *Burkholderia* to encode antibiotic biosynthetic gene clusters (BGCs) was also greater than *Paraburkholderia*. Non-ribosomal peptide synthases (NRPS) was the largest BGC class found in *Burkholderia* genomes, while terpene biosynthesis was the most represented BGC class within *Paraburkholderia*.

Chapter 7. Conclusion and future directions

7.1 Conclusion

As discussed in the review of *Burkholderiales* bacteria in Chapter 1, the initial knowledge of *Burkholderia* came from studying it as a plant pathogen in 1952 (Yabuuchi et al. 1992; Burkhead et al. 1994). Since then, most research has focused on understanding the role of *Burkholderia* species in pathogenicity towards a range of animal, plant and human hosts. The key known and widely studied *Burkholderia* groups were associated with infections: **(1)** *Burkholderia cepacia* complex (opportunistic pathogens caused lung infections in people with cystic fibrosis [CF]), **(2)** *Burkholderia pseudomallei* (a human and animal pathogen) and **(3)** *Burkholderia gladioli* (plant pathogens) (Sawana et al. 2014; Depoorter et al. 2016). *Burkholderia* bacteria are not only associated with infection; they live freely in multiple natural environments and in contrast occur as symbionts in a range of host organisms, most notably within fungi and insects (Depoorter et al. 2016). The growing interest in *Burkholderia*, from negative infectious traits to beneficial features, has led to further exploration of *Burkholderia* as a valuable source of natural products (Haeckl et al. 2019; Kunakom and Eustáquio 2019). Over the last five years, there have been multiple evaluations of the evolutionary history of *Burkholderia,* and this has expanded our understanding of the diversity of the genus and in particular expanded knowledge on *Paraburkholderia (Sawana et al. 2014)*. The majority of *Paraburkholderia* species have been isolated from the natural environment and are extensively associated with the plant rhizosphere and soil (Dobritsa and Samadpour 2019). The data from the present study demonstrates that the Borneo jungle (Sabah, Malaysia) is an interesting tropical environment, that was particularly rich in *Paraburkholderia.* These *Paraburkholderia* were not found to be antibiotic producers, however, their density within the jungle rhizosphere samples suggests they play a key role in maintaining beneficial plant and soil interactions (**results Chapter 5**).

The interest in exploiting members of *Burkholderia* genus as antimicrobial sources has been increased, and it is clear they produce multiple specialised metabolites (Haeckl et al. 2019;

Kunakom and Eustáquio 2019). Genomic analysis has showed that *Burkholderia* genomes encode multiple biosynthetic gene clusters, and in combination with analytical chemistry this has shown that single strains may produce several novel specialised metabolites (Kunakom and Eustáquio 2019). Multiple natural products have been identified in *Burkholderia* using these genome mining approaches. A set of natural compounds – Bactobolin A, Thailanstatin A, Malleilactone (Burkholderic acid), Thailandamide A, and Thailandepsin A (Burkholdac B) – was isolated from *Burkholderia thailandensis* (Liu and Cheng 2014). Enacyloxin IIa and cepacin A are potential antimicrobial compounds that can be isolated from *Burkholderia ambifaria* and genome mining played an extensive role in their discovery (Mahenthiralingam et al. 2011; Mullins et al. 2019). *Burkholderia gladioli* produces lagriamide, gladiolin and sinapigladioside (Flórez et al. 2017; Song et al. 2017; Flórez et al. 2018). In contrast, relatively few specialised metabolites have been characterised in *Paraburkholderia* strains with the siderophore, gramibactin recently discovered in *Paraburkholderia graminis*, being one of the few examples (Hermenau et al. 2018).

The overarching hypothesis behind the study was "**Environmental *Burkholderia* and *Paraburkholderia* are rich source of novel antibiotics".** The results of this study performed using parallel culture and genome-guided strategies, has confirmed:

1. The strategies developed to isolate *Burkholderia* and *Paraburkholderia* from a tropical environment were successful **(Results Chapter 4)**.
2. Systemic *Burkholderia* and *Paraburkholderia* environmental collections were successfully assembled **(Results Chapter 4).**
3. The diversity of this collection encompassed known and novel species. It identified multiple of novel taxa within the *Paraburkholderia* genus (44 *Paraburkholderia* isolates, 22 of them did not belong to any described *Paraburkholderia* species). In addition of the thirteen environmental *Burkholderia* strains isolated, three of them did not belong to any described *Burkholderia* species. The study also resulted in the isolation of a novel *Caballeronia* species **(Results Chapter 5)**.

4. All isolated environmental *Burkholderia*, including the novel Bcc taxa, demonstrated antimicrobial activity and produced one or more specialised metabolites. In contrast, none of the *Paraburkholderia* isolates showed antimicrobial bioactivity under the same growth conditions **(Results Chapter 6)**.
5. Overall, the study is the first to isolate such a large diversity of *Paraburkholderia* and *Burkholderia* from a single tropical natural environment, and apply genomic taxomy to characterise them in detail.

7.1.1 The development a strategy to identify and characterise bacteria in this study

The importance of developing new methods for isolation specific genera is required by natural products research to ensure a diversity of species can be examined. The discovery of natural products and diverse microbial sources are increasingly supported by the power of genome sequence data. Genome data provides a rapid insight into both the specialised metabolic potential and the taxonomic identity of microorganisms (Liu and Cheng 2014; Haeckl et al. 2019; Kunakom and Eustáquio 2019). In this study, the molecular and genome guided classification methods ranged from single gene and whole genome sequencing, and provided a powerful and accurate means of bacterial strain and species identification **(Results Chapter 4 and 5).** Multiple recent studies have also employed a genome mining approaches to understand what metabolic and biosynthetic capabilities *Burkholderia* have. For example by understanding the diversity of their metabolism and the genes that encode this, different growth and enrichment media have been recently designed (Haeckl et al. 2019). Combining both genome guided taxonomy and metabolic genome mining provides a powerful strategy for future exploration of environmental *Burkholderia* and *Paraburkholderia* species as new sources of natural products **(Results Chapter 5 and 6)**. This study has shown that the isolated *Burkholderia* bacteria are most likely to produce antimicrobials in the laboratory **(Results Chapter 6)**, but further analysis of genome encoded *Paraburkholderia* metabolic

pathways can now be exploited to design improved antibiotic production media (Haeckl et al. 2019).

This study has laid the foundations of rapid culture and molecular approaches to identify *Burkholderiales* bacteria from the natural environment. The application of a spiral plater to dilute cultures prior to growth on agar media greatly accelerated the screening phase in comparison to using the conventional manual single plates spreading techniques **(Results Chapter 4)**. This growth based isolation of pure bacteria was combined with a molecular approach, specifically using a *recA*-based PCR identification for *Burkholderia* species that was developed over 15 years ago (Payne et al. 2006). Despite the increases in diversity of the *Burkholderia* genus encompassing new taxa such as *Paraburkholderia* and *Caballeronia*, these *recA* PCR primers (Payne et al. 2006) performed very well in identifying relevant *Burkholderia* and *Paraburkholderia* from the organisms that grew. The combined growth on enrichment and molecular approaches applied to the jungle samples enabled 98 samples to be screened for the presence of *Burkholderiales* bacteria within less than three months **(Results Chapter 4)**. Expanding the analysis, and optimising it further for screening large numbers of environmental samples can now be carried out.

7.1.2 Novel taxa and strains added within *Burkholderia*, *Paraburkholderia*, and *Caballeronia*

At the start of this PhD, all of the constituent genera within *Burkholderia* were classified as *Burkholderia* (Table 1.1). Keeping pace with this rate of change in terms of taxonomic classification is difficult, but with genome guided taxonomy identification of known species and the definition of novel taxa has become much clearer. However, the accurate classification of different *Burkholderiales* genera, especially *Burkholderia* genus, is important, because it includes the human pathogenic group *Burkholderia cepacia* complex (Bcc) species. The taxonomic status of *Burkholderia* genus and other relative genera within *Burkholderiales*

bacteria remains a challenge, as new members from the present study and other studies were added to Bcc group and to *Paraburkholderia* genus as well (Jin et al. 2020). Hence, the accurate classification of these bacteria is needed.

The genome guided taxonomy analysis was ultimately used for accurate species identification, confirmed that the present study had resulted in the isolation of a unique novel set of *Paraburkholderia* and environmental *Burkholderia* species, as well as a novel *Caballeronia* species. Interestingly, three isolates were confirmed as novel taxa within the Bcc group **(Results Chapter 5)**. This was confirmed on several reports about the existence of Bcc in environments and added new members of this important group. The last reported new species within the Bcc group was *Burkholderia paludis* isolated from Malaysian tropical peat swamp soil and an antibiotic-siderophore producing (Ong et al. 2016). Further taxonomic groups will have to be added to the 22 species within the Bcc group to account for novel taxa from this study (jungle *Burkholderia*) alongside the new members published in 2020 (Jin et al. 2020). Another notable result of the present study was the diversity of novel *Paraburkholderia* species that were recovered **(Results Chapter 5)**. What led to this would be an interesting scientific question that could be answered by further studies, such as there was thought that PCAT media may enrich *Paraburkholderia* species more favorably than *Burkholderia* species. This question could be answered by rescreening well-classified *Paraburkholderia* species alondside *Burkholderia* species into PCAT media. Since this is one of the first studies to investigate the presence of environmental *Burkholderia* and *Paraburkholderia* in a tropical natural environment. Further analysis of other jungle environments using different enrichment media, but including PCAT, will be needed to answer these questions.

For the genomic taxonomy based approach to bacterial species identification, average nucleotide identity (ANI) and DNA–DNA hybridisation were used to propose novel species boundaries of 95–96% for ANI and 70% for DDH (Konstantinidis and Tiedje 2005; Meier-Kolthoff et al. 2013). Both methods were fast and highly effective for delineating the new *Burkholderia* and *Paraburkholderia* species in the present study **(Results Chapter 5)**,

corroborating what others have seen in recent reclassification studies (Dobritsa and Samadpour 2016; Lopes-Santos et al. 2017; Estrada-de los Santos et al. 2018). Other researchers have used multilocus sequence analysis (MLSA) of four housekeeping genes (*recA*, *gyrB*, *trpB* and *gltB*) to confirm the position of novel strains in relation to references for *Paraburkholderia* (Sawana et al. 2014; Dobritsa et al. 2016). Using genomic taxonomy based on whole genome ANI and DDH it is ultimately more accurate than these MLSA methods limited to four genes (Sawana et al. 2014; Dobritsa et al. 2016), because these methods are based on whole genomes comparisions and analysis and they are both the most recent powerfull methods for species delineation. However, a comparison of the minimal number of genes required to provide accurate identification of *Paraburkholderia* and *Burkholderia* species in future would be worthwhile.

Identification of new species based solely on genetics analysis and complete genome data is not sufficient to classify and name new species; combining this genetic information with phenotypic information is required for this (Depoorter et al. 2016). Looking to the most recent classification studies of *Burkholderia*, *Paraburkholderia* and *Caballeronia* species, the phenotypic features of these bacteria needs to be combined with the genomic analysis to describe novel type strains for novel taxa (De Smet et al. 2015; Quan et al. 2019; Paulitsch et al. 2020). The description of the phenotypic features for the novel environmental *Burkholderia*, *Paraburkholderia* and *Caballeronia* should be future direction of this PhD data to complete the data available for species identification. The phenotypic tests should include morphological, physiological and biochemical characteristics, and take advantage of high resolution phenotypic methods such as matrix-assisted laser desorption ionisation time of flight mass spectroscopy (Van Belkum et al. 2013). The differential characteristics for the new species should be compared with the most closely related species type strains of *Burkholderia*, *Paraburkholderia* and *Caballeroniea* (Li et al. 2016). Finally, applying the combination of phylogenomic and phenotypic characterisation would enable naming of the novel taxa identified in the present study.

7.1.3 Development of new growth and PCR-based approaches for isolation specific genus *Paraburkholderia* or environmental *Burkholderia*

The majority of methods and growth media developed for *Burkholderia* have been focused on pathogenic species (Table 3.1). Given the diversity of this genus and the fact that it now encompasses five genera; it would be interesting to develop enrichment methods to separate the constituent species. However, the development of new media for specific genera is challenging **(Results Chapter 3)**. Utilisation of different aromatic compounds is a feature of the *Burkholderia* and *Paraburkholderia* genera (as seen below). Based on the results in Chapter 3, bromoacetic acid successfully enriched >50% of a selected panel of *Burkholderia* and *Paraburkholderia* chosen from the reference species available in 2016–2017. Since that time several selected members of this strain panel were reclassified. In addition, this study has produced multiple novel *Paraburkholderia* species that could be included to test growth media. It would be worthwhile to extend this screening and focus on one specific group of environmental *Burkholderia* or *Paraburkholderia*. Altering the pH is worth to examining. The ability of *Burkholderia* strains to tolerate moderate pH values has been with the greatest selectivity of strains seen between pH values ranging from 5.0 to 6.0 (Stopnisek et al. 2014; Haeckl et al. 2019). The prediction of the selective additives such as carbon, nitrogen, or metal sources as well as antibiotics, could be selected for each species/genus based genome data analysis, following the strategies developed by Haeckl et al (2019). This was a highly effective strategy and that developed five selective media applied to enrich *Burkholderiales* bacteria from a set of 49 environmental samples with a greater than 95% success rate (Haeckl et al. 2019).

To improve PCR-based methods for identification, it would be useful to develop *Paraburkholderia*-specific primers and environmental *Burkholderia*-specific primers. The *recA* gene approaches applied in this study were developed for the *Burkholderia* genus some time ago (Payne et al. 2005). As seen in the Chapter 1, the taxonomy of *Burkholderia* is expanding

considerably and the*recA*-gene PCR used in this study was not 100% specific for *Burkholderia* and *Paraburkholderia*, as multiple non-*Burkholderia* species were *recA* PCR positive (Chapter 3 and 4). A *repA*-based multiplex PCR, based on this gene from the third replicon, was developed as a rapid identification tool for clinical Bcc species (Drevinek et al. 2008). Based on the fact that a unifying feature of both *Paraburkholderia* and *Burkholderia* is that they have a multi replicon genome, developing new primers that target genes, such as *repA*, *parA* and *parB* genes, on the 2nd replicon could be useful as new specific identification method for these bacteria. Only a limited number of other Gram-negative bacteria within environmental samples have such multi replicon genomes and the second replicon of *Burkholderia* and *Paraburkholderia* is highly specific to these organisms. These genes are essential for bacterial replicating and partitioning during bacterial growth and cell division. Other options include testing *Burkholderia*-specific primers developed in a recent study characterising and enriching environmental *Burkholderia* (Haeckl et al. 2019).

7.1.4 Characterisation of promising specialised metabolites and antimicrobial compounds

Complete genome sequences should be obtained for all novel specialised metabolite *Burkholderia* producers. The characterisation of biosynthetic pathways in interesting *Burkholderia* antibiotic producer strains can be upscaled and accelerated based on genomics and biochemistry methods. Analysis of genomes and cryptic biosynthetic gene clusters (BGCs) identification maybe characterised using bioinformatics tools (Blin et al. 2013; Blin et al. 2019). Molecular genetic methods, such as transcriptomic and gene mutagenesis, may be used to systematically understand antibiotic production, and analytical chemistry is used to identify the novel specialised metabolites. Structure characterisation of novel metabolites would initially involve scaling up their production and purification using probative HPLC. After this a combination of high resolution analytical chemistry methods would be used to determine the constituent metabolites and their structural relationship to one another. Characterisation

of the novel metabolites would done by using a combination of high-resolution mass spectrometry and nuclear magnetic resonance (NMR) spectroscopy, as has been carried out for enacyloxin, gladiolin and cepacin (Mahenthiralingam et al. 2011; Song et al. 2017; Mullins et al. 2019). Silent antibiotic biosynthesis gene clusters which are not expressed in the laboratory can also be targeted for activation using genetic and synthetic biology approaches (Rutledge and Challis 2015).

Since we have assembled a novel and highly diverse set of *Paraburkholderia* members from this study, they could have potential biotechnological uses, and this would be future investigation studies. Example biotechnological use of *Paraburkholderia* include the capacity of *Paraburkholderia susongensis* to sequester the minerals from weathered rock surfaces (Gu et al., 2015) and *Paraburkholderia metalliresistens* to solubilise phosphate from polluted soil (Guo et al. 2015). A limited number of *Paraburkholderia* species have been described as acid-tolerant (*Paraurkholderia acidipaludis, Paraburkholderia heleia* and *Paraburkholderia bannensis*; (Aizawa et al. 2010a; Aizawa et al. 2010b; Aizawa et al. 2011) and metal-tolerant (*Paraburkholderia insulsa* and *Paraburkholderia metalliresistens*; (Guo et al. 2015; Rusch et al. 2015). *Paraburkholderia phytofirmans* PsJN has been shown to act as a biocontrol agent against plant disease caused by *Pseudomonas syringae* in *A. thaliana* plants (Timmermann et al. 2017). *Paraburkholderia phytofirmans*, *Paraburkholderia bryophila* and *Paraburkholderia megapolitana* have also been described as plant-growth-promoting bacteria in several studies (Sessitsch et al. 2005; Vandamme et al. 2007). *Paraburkholderia caffeinilytica* CF1 can utilise caffeine as its sole carbon and nitrogen source, which could overcome the environmental pollution caused by the overuse of caffeine in food and drug industries (Sun et al. 2020). Members of *Paraburkholderia* have also been described for their capacity to degrade aromatic and phenolic acid and were reported as principle contributors to the soil priming effect (Otsuka et al. 2011; Morya et al. 2020; Zwetsloot et al. 2020).

The ability of several species of *Paraburkholderia* and environmental *Burkholderia* to nodulate that effectively fixes atmospheric N_2 has been reported as key plant beneficial feature of these

bacteria (De Meyer et al. 2016). The results of the present study extend these observations for *Paraburkholderia* species isolated specifically from jungle soils. Six isolates from this study (4 *Paraburkholderia tropica* strains and 2 *Burkholderia vietnamensis* strain) were putative nitrogen-fixing species since they encoded component nif genes (as seen in Chapter 5). In several studies of the N_2-fixing *Burkholderia* and *Paraburkholderia* species, phylogenetic analysis based on nodulation capacity has been used to identify the N_2-fixing *Burkholderia* and *Paraburkholderia* strains/species and also showed the differences between these strains/species (Gillis et al. 1995; Reis et al. 2004). Further characterisation of the nitrogen fixing ability of the jungle *Paraburkholderia* should be carried out since this may be a key feature of their contribution in maintaining the fertility of jungle environments.

In conclusion, the outputs from this PhD can initiate multiple further interesting avenues of research encompassing both evolution and taxonomic studies or novel specialised metabolite characterisation studies. Key future study directions are as follows:

1. To fully describe the new *Burkholderia* and *Paraburkholderia* strains based on extensive phenotypic characterisation.
2. To characterise the novel specialised metabolites, they encoded as potential antimicrobial compounds or biotechnological agents, and examine the utility of strains for biodegradation, biocontrol, or plant promotion.
3. Further optimise the cultivation-dependent and -independent methods for the isolation of environmental *Burkholderia* and *Paraburkholderia* from other natural environments of interest.

Chapter 8. References

Aaron, s. d. et al. 2000. Multiple combination bactericidal antibiotic testing for patients with cystic fibrosis infected with *Burkholderia cepacia*. *American Journal of Respiratory and Critical Care Medicine* 161(4), pp. 1206-1212.

Achouak, W. et al. 1999. *Burkholderia caribensis* sp. nov., an exopolysaccharide-producing bacterium isolated from vertisol microaggregates in Martinique. *International Journal of Systematic and Evolutionary Microbiology* 49(2), pp. 787-794.

Achtman, M. and Wagner, M. 2008. Microbial diversity and the genetic nature of microbial species. *Nature Reviews Microbiology* 6(6), pp. 431-440.

Agnoli, K. et al. 2012. Exposing the third chromosome of *Burkholderia cepacia* complex strains as a virulence plasmid. *Molecular Microbiology* 83(2), pp. 362-378.

Aizawa, T. et al. 2010a. *Burkholderia heleia* sp. nov., a nitrogen-fixing bacterium isolated from an aquatic plant, *Eleocharis dulcis*, that grows in highly acidic swamps in actual acid sulfate soil areas of Vietnam. *International Journal of Systematic and Evolutionary Microbiology* 60(5), pp. 1152-1157.

Aizawa, T. et al. 2010b. *Burkholderia acidipaludis* sp. nov., aluminium-tolerant bacteria isolated from Chinese water chestnut (Eleocharis dulcis) growing in highly acidic swamps in South-East Asia. *International Journal of Systematic and Evolutionary Microbiology* 60(9), pp. 2036-2041.

Aizawa, T. et al. 2011. *Burkholderia bannensis* sp. nov., an acid-neutralizing bacterium isolated from torpedo grass (*Panicum repens*) growing in highly acidic swamps. *International Journal of Systematic and Evolutionary Microbiology* 61(7), pp. 1645-1650.

Alanjary, M. et al. 2019. AutoMLST: an automated web server for generating multi-locus species trees highlighting natural product potential. *Nucleic Acids Research* 47(W1), pp. W276-W282.

Altschul, S. F. et al. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215(3), pp. 403-410.

Amann, R. I. et al. 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Mol. Biol. Rev.* 59(1), pp. 143-169.

Arima, K. et al. 1964. Pyrrolnitrin, a new antibiotic substance, produced by *Pseudomonas*. *Agricultural and Biological Chemistry* 28(8), pp. 575-576.

Ashdown, L. 1979. An improved screening technique for isolation of *Pseudomonas pseudomallei* from clinical specimens. *Pathology* 11(2), pp. 293-297.

Blin, K. et al. 2013. antiSMASH 2.0—a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Research* 41(W1), pp. W204-W212.

Blin, K. et al. 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Research* 47(W1), pp. W81-W87.

Bode, H. B. et al. 2002. Big effects from small changes: possible ways to explore nature's chemical diversity. *ChemBioChem* 3(7), pp. 619-627.

Bontemps, C. et al. 2010. *Burkholderia* species are ancient symbionts of legumes. *Molecular Ecology* 19(1), pp. 44-52.

Brett, P. et al. 1997. Characterization of *Burkholderia pseudomallei* and *Burkholderia pseudomallei*-like strains. *Epidemiology & Infection* 118(2), pp. 137-148.

Burbage, D. and Sasser, M. eds. 1982. *A medium selective for Pseudomonas-cepacia*. *Phytopathology.* AMER PHYTOPATHOLOGICAL SOC 3340 PILOT KNOB ROAD, ST PAUL, MN 55121.

Burkhead, K. D. et al. 1994. Pyrrolnitrin production by biological control agent *Pseudomonas cepacia* B37w in culture and in colonized wounds of potatoes. *Appl. Environ. Microbiol.* 60(6), pp. 2031-2039.

Burkholder, W. H. 1950. Sour skin, a bacterial rot of onion bulbs. *Phytopathology* 40(1),

Caballero-Mellado, J. et al. 2004. *Burkholderia unamae* sp. nov., an N2-fixing rhizospheric and endophytic species. *International Journal of Systematic and Evolutionary Microbiology* 54(4), Pp. 1165-1172.

Cartwright, D. K. et al. 1995. Pyrrolnitrin and phenazine production by *Pseudomonas cepacia*, strain 5.5 B, a biocontrol agent of Rhizoctonia solani. *Applied Microbiology and Biotechnology* 43(2), pp. 211-216.

Chain, P. S. et al. 2006. *Burkholderia xenovorans* LB400 harbors a multi-replicon, 9.73-Mbp genome shaped for versatility. *Proceedings of the National Academy of Sciences* 103(42), pp. 15280-15287.

Chen, W.-M. et al. 2007. *Burkholderia nodosa* sp. nov., isolated from root nodules of the woody Brazilian legumes Mimosa bimucronata and Mimosa scabrella. *International Journal of Systematic and Evolutionary Microbiology* 57(5), pp. 1055-1059.

Chen, W.-M. et al. 2006. *Burkholderia mimosarum* sp. nov., isolated from root nodules of Mimosa spp. from Taiwan and South America. *International Journal of Systematic and Evolutionary Microbiology* 56(8), pp. 1847-1851.

Choi, G.-M. and Im, W.-T. 2017. *Paraburkholderia azotifigens* sp. nov., a nitrogen-fixing bacterium isolated from paddy soil. *International Journal of Systematic and evolutionary Microbiology* 68(1), pp. 310-316.

Chowdhury, P. R. and Heinemann, J. A. 2006. The general secretory pathway of *Burkholderia gladioli* pv. agaricicola BG164R is necessary for cavity disease in white button mushrooms. *Appl. Environ. Microbiol.* 72(5), pp. 3558-3565.

Christofides, S. R. et al. 2019. Fungal control of early-stage bacterial community development in decomposing wood. *Fungal Ecology* 42, p. 100868.

Coates, A. R. et al. 2011. Novel classes of antibiotics or more of the same? *British journal of pharmacology* 163(1), pp. 184-194.

Coenye, T. and Vandamme, P. 2003. Diversity and significance of *Burkholderia* species occupying diverse ecological niches. *Environmental Microbiology* 5(9), pp. 719-729.

Connor, T. R. et al. 2016. CLIMB (the Cloud Infrastructure for Microbial Bioinformatics): an online resource for the medical microbiology community. *Microbial Genomics* 2(9),

Coutinho, C. P. et al. 2015. Incidence of *Burkholderia contaminans* at a cystic fibrosis centre with an unusually high representation of *Burkholderia cepacia* during 15 years of epidemiological surveillance. *Journal of Medical Microbiology* 64(8), pp. 927-935.

Cunningham-Oakes, E. et al. 2019. Understanding the challenges of non-food industrial product contamination. *FEMS microbiology letters*,

Dalmastri, C. et al. 2007. Investigating *Burkholderia cepacia* complex populations recovered from Italian maize rhizosphere by multilocus sequence typing. *Environmental Microbiology* 9(7), pp. 1632-1639.

de Lima Procópio, R. E. et al. 2012. Antibiotics produced by Streptomyces. *The Brazilian Journal of Infectious Diseases* 16(5), pp. 466-471.

De Meyer, S. E. et al. 2016. Symbiotic *Burkholderia* species show diverse arrangements of nif/fix and nod genes and lack typical high-affinity cytochrome cbb3 oxidase genes. *Molecular Plant-Microbe Interactions* 29(8), pp. 609-619.

De Meyer, S. E. et al. 2018. Symbiotic and non-symbiotic *Paraburkholderia* isolated from South African Lebeckia ambigua root nodules and the description of *Paraburkholderia fynbosensis* sp. nov. *International Journal of Systematic and Evolutionary Microbiology* 68(8), pp. 2607-2614.

De Smet, B. et al. 2015. *Burkholderia stagnalis* sp. nov. and *Burkholderia territorii* sp. nov., two novel *Burkholderia cepacia* complex species from environmental and human sources. *International Journal of Systematic and Evolutionary Microbiology* 65(7), pp. 2265-2271.

Depoorter, E. et al. 2016. *Burkholderia*: an update on taxonomy and biotechnological potential as antibiotic producers. *Applied Microbiology and Biotechnology* 100(12), pp. 5215-5229. doi: 10.1007/s00253-016-7520-x

Dobritsa, A. P. et al. 2016. Reclassification of *Paraburkholderia panaciterrae* (Farh et al. 2015) Dobritsa & Samadpour 2016 as a later synonym of *Paraburkholderia ginsengiterrae* (Farh et al. 2015) Dobritsa & Samadpour 2016. *International Journal of Systematic and Evolutionary Microbiology* 66(10), pp. 4085-4087.

Dobritsa, A. P. et al. 2017. Transfer of 13 species of the genus *Burkholderia* to the genus *Caballeronia* and reclassification of *Burkholderia jirisanensis* as *Paraburkholderia jirisanensis* comb. nov. *International Journal of Systematic and Evolutionary Microbiology* 67(10), pp. 3846-3853.

Dobritsa, A. P. and Samadpour, M. 2016. Transfer of eleven species of the genus *Burkholderia* to the genus *Paraburkholderia* and proposal of *Caballeronia* gen. nov. to accommodate twelve species of the genera Burkholderia and Paraburkholderia. *International Journal of Systematic and Evolutionary Microbiology* 66(8), pp. 2836-2846.

Dobritsa, A. P. and Samadpour, M. 2019. Reclassification of *Burkholderia insecticola* as *Caballeronia insecticola* comb. nov. and reliability of conserved signature indels as molecular synapomorphies. *International Journal of Systematic and Evolutionary Microbiology* 69(7), pp. 2057-2063.

Drevinek, P. et al. 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(1), p. 44.

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

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(1), pp. 69-78.

Esmaeel, Q. et al. 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(3), p. 40.

Estrada-de los Santos, P. 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(8), p. 389.

Estrada-De Los Santos, P. et al. 2013. Phylogenetic analysis of *Burkholderia* species by multilocus sequence analysis. *Current Microbiology* 67(1), pp. 51-60.

Flanagan, P. and Paull, A. 1998. Isolation of *Burkholderia cepacia* by enrichment. *Journal of clinical pathology* 51(7), pp. 557-558.

Flórez, L. V. et al. 2017. Antibiotic-producing symbionts dynamically transition between plant pathogenicity and insect-defensive mutualism. *Nature communications* 8(1), pp. 1-9.

Flórez, L. V. et al. 2018. An antifungal polyketide associated with horizontally acquired genes supports symbiont-mediated defense in Lagria villosa beetles. *Nature communications* 9(1), pp. 1-10.

Gao, Z.-h. et al. 2018a. *Paraburkholderia phosphatilytica* sp. nov., a phosphate-solubilizing bacterium isolated from forest soil. *International Journal of Systematic and Evolutionary Microbiology* 69(1), pp. 196-202.

Gao, Z.-h. et al. 2018b. *Paraburkholderia caseinilytica* sp. nov., isolated from the pine and broad-leaf mixed forest soil. *International Journal of Systematic and Evolutionary Microbiology* 68(6), pp. 1963-1968.

Gao, Z. et al. 2016. *Paraburkholderia caffeinilytica* sp. nov., isolated from the soil of a tea plantation. *International Journal of Systematic and Evolutionary Microbiology* 66(10), pp. 4185-4190.

Gee, J. E. et al. 2018. *Burkholderia thailandensis* isolated from infected wound, Arkansas, USA. *Emerging Infectious Diseases* 24(11), p. 2091.

Gilligan, P. H. and Schidlow, D. V. 1984. The role of *Pseudomonas cepacia* in pulmonary disease of cystic fibrosis patients. *Clinical Microbiology Newsletter* 6(6), pp. 42-44.

Gillis, M. et al. 1995. Polyphasic taxonomy in the genus *Burkholderia* leading to an emended description of the genus and proposition of *Burkholderia vietnamiensis* sp. nov. for N2-fixing isolates from rice in Vietnam. *International Journal of Systematic and Evolutionary Microbiology* 45(2), pp. 274-289.

Glass, M. B. et al. 2009. Comparison of four selective media for the isolation of *Burkholderia mallei* and *Burkholderia pseudomallei*. *The American Journal of Tropical Medicine and Hygiene* 80(6), pp. 1023-1028.

Glass, M. B. et al. 2006a. Pneumonia and septicemia caused by *Burkholderia thailandensis* in the United States. *Journal of Clinical Microbiology* 44(12), pp. 4601-4604.

Glass, M. B. et al. 2006b. *Burkholderia oklahomensis* sp. nov., a *Burkholderia pseudomallei*-like species formerly known as the Oklahoma strain of *Pseudomonas pseudomallei. International Journal of Systematic and Evolutionary Microbiology* 56(9), pp. 2171-2176.

Godoy, D. et al. 2003. Multilocus sequence typing and evolutionary relationships among the causative agents of melioidosis and glanders, *Burkholderia pseudomallei* and *Burkholderia mallei. Journal of Clinical Microbiology* 41(5), pp. 2068-2079.

Goris, J. et al. 2004. Classification of the biphenyl-and polychlorinated biphenyl-degrading strain LB400T and relatives as *Burkholderia xenovorans* sp. nov. *International Journal of Systematic and Evolutionary Microbiology* 54(5), pp. 1677-1681.

Guo, J. K. et al. 2015. *Burkholderia metalliresistens* sp. nov., a multiple metal-resistant and phosphate-solubilising species isolated from heavy metal-polluted soil in Southeast China. *Antonie van Leeuwenhoek* 107(6), pp. 1591-1598.

Gurevich, A. et al. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29(8), pp. 1072-1075.

Gyaneshwar, P. et al. 2011. Legume-nodulating betaproteobacteria: diversity, host range, and future prospects. *Molecular Plant-Microbe Interactions* 24(11), pp. 1276-1288.

Haeckl, F. J. et al. 2019. A selective genome-guided method for environmental *Burkholderia* isolation. *Journal of Industrial Microbiology & Biotechnology* 46(3-4), pp. 345-362.

Hagedorn, C. et al. 1987. A selective medium for enumeration and recovery of *Pseudomonas cepacia* biotypes from soil. *Appl. Environ. Microbiol.* 53(9), pp. 2265-2268.

Hall, C. M. et al. 2015. Diverse *Burkholderia* species isolated from soils in the southern United States with no evidence of *B. pseudomallei*. *PloS one* 10(11),

Hareland, W. A. et al. 1975. Metabolic function and properties of 4-hydroxyphenylacetic acid 1-hydroxylase from *Pseudomonas acidovorans*. *Journal of Bacteriology* 121(1), pp. 272-285.

Henry, D. et al. 1997. Identification of *Burkholderia cepacia* isolates from patients with cystic fibrosis and use of a simple new selective medium. *Journal of Clinical Microbiology* 35(3), pp. 614-619.

Hermenau, R. et al. 2018. Gramibactin is a bacterial siderophore with a diazeniumdiolate ligand system. *Nature Chemical Biology* 14(9), pp. 841-843.

Ho, Y.-N. and Huang, C.-C. 2015. Draft genome sequence of *Burkholderia cenocepacia* strain 869T2, a plant-beneficial endophytic bacterium. *Genome Announc.* 3(6), pp. e01327-01315.

Howard, K. and Inglis, T. 2003. Novel selective medium for isolation of *Burkholderia pseudomallei*. *Journal of clinical microbiology* 41(7), pp. 3312-3316.

Huo, Y. et al. 2018. *Paraburkholderia panacihumi* sp. nov., an isolate from ginseng-cultivated soil, is antagonistic against root rot fungal pathogen. *Archives of Microbiology* 200(8), pp. 1151-1158.

Jiao, Y. et al. 1996. Structural identification of cepaciamide A, a novel fungitoxic compound from *Pseudomonas cepacia* D-202. *Tetrahedron letters* 37(7), pp. 1039-1042.

Jin, Y. et al. 2020. Genome-based classification of *Burkholderia cepacia* complex provides new insight into its taxonomic status. *Biology Direct* 15(1), pp. 1-14.

Kadir, J. et al. 2008. Extraction of antifungal substances from *Burkholderia cepacia* with antibiotic activity against *Colletotrichum gloeosporioides* on papaya (Carica papaya). *Int. J. Agric. Biol* 10, pp. 15-20.

Kanehisa, M. and Goto, S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* 28(1), pp. 27-30.

Kang, Y. et al. 1998. Characterization of genes involved in biosynthesis of a novel antibiotic from *Burkholderia cepacia* BC11 and their role in biological control of *Rhizoctonia solan*i. *Appl. Environ. Microbiol.* 64(10), pp. 3939-3947.

Katz, L. and Baltz, R. H. 2016. Natural product discovery: past, present, and future. *Journal of Industrial Microbiology & Biotechnology* 43(2-3), pp. 155-176.

Kenna, D. T. et al. 2017. Prevalence of *Burkholderia* species, including members of Burkholderia cepacia complex, among UK cystic and non-cystic fibrosis patients. *Journal of Medical Microbiology* 66(4), pp. 490-501.

Kirinuki, T. et al. 1977. Altericidins, a complex of polypeptide antibiotics, produced by *Pseudomonas* sp. and their effect for the control of black spot of pear caused by Alternaria kikuchiana Tanaka. *Science Reports of Faculty of Agriculture-Kobe University (Japan)*,

Konstantinidis, K. T. and Tiedje, J. M. 2005. Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences* 102(7), pp. 2567-2572.

Kumar, S. et al. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* 33(7), pp. 1870-1874.

Kunakom, S. and Eustáquio, A. S. 2019. *Burkholderia* as a source of natural products. *Journal of Natural Products* 82(7), pp. 2018-2037.

Lagesen, K. et al. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research* 35(9), pp. 3100-3108.

Lane, D. J. et al. 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proceedings of the National Academy of Sciences* 82(20), pp. 6955-6959.

Larkin, M. A. et al. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23(21), pp. 2947-2948.

Lee, Y. and Jeon, C. O. 2018. *Paraburkholderia aromaticivorans* sp. nov., an aromatic hydrocarbon-degrading bacterium, isolated from gasoline-contaminated soil. *International Journal of Systematic and Evolutionary Microbiology* 68(4), pp. 1251-1257.

Lee, Y. et al. 2019. Biodegradation of naphthalene, BTEX, and aliphatic hydrocarbons by *Paraburkholderia aromaticivoran*s BN5 isolated from petroleum-contaminated soil. *Scientific reports* 9,

Li, Q. et al. 2016. Cultural, physiological, and biochemical identification of actinobacteria. *Actinobacteria-Basics and Biotechnological Applications*, pp. 87-111.

Li, X. et al. 2008. Multiple effects of a novel compound from *Burkholderia cepacia* against Candida albicans. *FEMS microbiology letters* 285(2), pp. 250-256.

Li, X. et al. 2007. Antifungal activity of a novel compound from *Burkholderia cepacia* against plant pathogenic fungi. *Letters in applied microbiology* 45(5), pp. 508-514.

LiPuma, J. J. 2010. The changing microbial epidemiology in cystic fibrosis. *Clinical Microbiology Reviews* 23(2), pp. 299-323.

Liu, X. and Cheng, Y.-Q. 2014. Genome-guided discovery of diverse natural products from *Burkholderia* sp. *Journal of Industrial Microbiology & Biotechnology* 41(2), pp. 275-284.

Lopes-Santos, L. et al. 2017. Reassessment of the taxonomic position of *Burkholderia andropogonis* and description of Robbsia andropogonis gen. nov., comb. nov. *Antonie van Leeuwenhoek* 110(6), pp. 727-736.

Maeda, Y. et al. 2006. Phylogenetic study and multiplex PCR-based detection of *Burkholderia plantarii*, *Burkholderia glumae* and *Burkholderia gladioli* using *gyrB* and *rpoD* sequences. *International Journal of Systematic and Evolutionary Microbiology* 56(5), pp. 1031-1038.

Mahenthiralingam, E. et al. 2008. *Burkholderia cepacia* complex bacteria: opportunistic pathogens with important natural biology. *Journal of Applied Microbiology* 104(6), pp. 1539-1551.

Mahenthiralingam, E. et al. 1996. Epidemiology of *Burkholderia cepacia* infection in patients with cystic fibrosis: analysis by randomly amplified polymorphic DNA fingerprinting. *Journal of Clinical Microbiology* 34(12), pp. 2914-2920.

Mahenthiralingam, E. et al. 2011. Enacyloxins Are Products of an Unusual Hybrid Modular Polyketide Synthase Encoded by a Cryptic *Burkholderia ambifaria* Genomic Island. *Chemistry & Biology* 18(5), pp. 665-677. doi: http://dx.doi.org/10.1016/j.chembiol.2011.01.020

Mahenthiralingam, E. et al. 2005. The multifarious, multireplicon *Burkholderia cepacia* complex. *Nature Reviews Microbiology* 3(2), pp. 144-156.

Martina, P. et al. 2013. Genetic diversity of *Burkholderia contaminans* isolates from cystic fibrosis patients in Argentina. *Journal of Clinical Microbiology* 51(1), pp. 339-344.

Martínez-Aguilar, L. et al. 2013. *Burkholderia caballeronis* sp. nov., a nitrogen fixing species isolated from tomato (Lycopersicon esculentum) with the ability to effectively nodulate *Phaseolus vulgaris*. *Antonie van Leeuwenhoek* 104(6), pp. 1063-1071.

Mavengere, N. R. et al. 2014. *Burkholderia aspalathi* sp. nov., isolated from root nodules of the South African legume *Aspalathus abietina* Thunb. *International Journal of Systematic and Evolutionary Microbiology* 64(6), pp. 1906-1912.

McCormick, J. B. et al. 1977. Wound infection by an indigenous *Pseudomonas pseudomallei*-like organism isolated from the soil: case report and epidemiologic study. *Journal of Infectious Diseases* 135(1), pp. 103-107.

Medina-Pascual, M. J. et al. 2015. Increase in isolation of *Burkholderia contaminans* from Spanish patients with cystic fibrosis. *Clinical Microbiology and Infection* 21(2), pp. 150-156.

Meier-Kolthoff, J. P. et al. 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC bioinformatics* 14(1), p. 60.

Merrick, M. and Edwards, R. 1995. Nitrogen control in bacteria. *Microbiol. Mol. Biol. Rev.* 59(4), pp. 604-622.

Meyers, E. et al. 1987. Xylocandin: a new complex of antifungal peptides. *The Journal of Antibiotics* 40(11), pp. 1515-1519.

Moon, S.-S. et al. 1996. Plant growth promoting and fungicidal 4-quinolinones from *Pseudomonas cepacia*. *Phytochemistry* 42(2), pp. 365-368.

Morya, R. et al. 2020. *Burkholderia*: An Untapped but Promising Bacterial Genus for the Conversion of Aromatic Compounds. *Trends in Biotechnology*,

Mullins, A. J. et al. 2019. Genome mining identifies cepacin as a plant-protective metabolite of the biopesticidal bacterium *Burkholderia ambifaria*. *Nature Microbiology* 4(6), pp. 996-1005.

Nandakumar, R. et al. 2009. *Burkholderia glumae* and *B. gladioli* cause bacterial panicle blight in rice in the southern United States. *Plant Disease* 93(9), pp. 896-905.

Neubauer, H. et al. 2005. Serodiagnosis of *Burkholderia mallei* Infections in Horses: State-of-the-art and Perspectives. *Journal of Veterinary Medicine, Series B* 52(5), pp. 201-205.

Nzula, S. et al. 2002. Influence of taxonomic status on the in vitro antimicrobial susceptibility of the *Burkholderia cepacia* complex. *Journal of Antimicrobial Chemotherapy* 50(2), pp. 265-269.

O'Sullivan, L. A. et al. 2007. Identifying the genetic basis of ecologically and biotechnologically useful functions of the bacterium *Burkholderia vietnamiensis*. *Environmental Microbiology* 9(4), pp. 1017-1034.

Ong, K. S. et al. 2016. *Burkholderia paludis* sp. nov., an antibiotic-siderophore producing novel Burkholderia cepacia complex species, isolated from Malaysian tropical peat swamp soil. *Frontiers in Microbiology* 7, p. 2046.

Otsuka, Y. et al. 2011. *Burkholderia oxyphila* sp. nov., a bacterium isolated from acidic forest soil that catabolizes (+)-catechin and its putative aromatic derivatives. *International Journal of Systematic and Evolutionary Microbiology* 61(2), pp. 249-254.

Page, A. J. et al. 2015. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31(22), pp. 3691-3693.

Parke, J. L. 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(1), pp. 225-258.

Parker, W. L. et al. 1984. Cepacin A and cepacin B, two new antibiotics produced by *Pseudomonas cepacia*. *The Journal of Antibiotics* 37(5), pp. 431-440.

Parte, A. C. 2013. LPSN—list of prokaryotic names with standing in nomenclature. *Nucleic Acids Research* 42(D1), pp. D613-D616.

Partida-Martinez, L. P. et al. 2007. *Burkholderia rhizoxinica* sp. nov. and *Burkholderia endofungorum* sp. nov., bacterial endosymbionts of the plant-pathogenic fungus Rhizopus microsporus. *International Journal of Systematic and Evolutionary Microbiology* 57(11), pp. 2583-2590.

Patel, C. B. et al. 2019. Microbial genes involved in interaction with plants.*New and Future Developments in Microbial Biotechnology and Bioengineering*. Elsevier, pp. 171-180.

Paulitsch, F. et al. 2020. *Paraburkholderia atlantica* sp. nov. and *Paraburkholderia franconis* sp. nov., two new nitrogen-fixing nodulating species isolated from Atlantic forest soils in Brazil. *Archives of Microbiology*, pp. 1-12.

Payne, G. W. et al. 2006. Application of a *recA* gene-based identification approach to the maize rhizosphere reveals novel diversity in *Burkholderia* species. *FEMS microbiology letters* 259(1), pp. 126-132.

Payne, G. W. et al. 2005. Development of a recA gene-based identification approach for the entire Burkholderia genus. *Appl. Environ. Microbiol.* 71(7), pp. 3917-3927.

Peeters, C. et al. 2016. Extensive cultivation of soil and water samples yields various pathogens in patients with cystic fibrosis but not *Burkholderia multivorans*. *Journal of Cystic Fibrosis* 15(6), pp. 769-775.

Perin, L. et al. 2006. *Burkholderia silvatlantica* sp. nov., a diazotrophic bacterium associated with sugar cane and maize. *International Journal of Systematic and Evolutionary Microbiology* 56(8), Pp. 1931-1937.

Podin, Y. et al. 2014. *Burkholderia pseudomallei* isolates from Sarawak, Malaysian Borneo, are predominantly susceptible to aminoglycosides and macrolides. *Antimicrobial Agents and Chemotherapy* 58(1), pp. 162-166.

Price, M. N. et al. 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. *PloS one* 5(3),

Quan, X.-T. et al. 2019. *Caballeronia ginsengisol*i sp. nov., isolated from ginseng cultivating soil. *Archives of Microbiology* 201(4), pp. 443-449.

Radua, S. et al. 2000. Characterization of *Burkholderia pseudomallei* isolated in Thailand and Malaysia. *Diagnostic Microbiology and Infectious Disease* 38(3), pp. 141-145.

Ragupathi, N. K. D. and Veeraraghavan, B. 2019. Accurate identification and epidemiological characterization of *Burkholderia cepacia* complex: an update. *Annals of Clinical Microbiology and Antimicrobials* 18(1), p. 7.

Ramette, A. et al. 2005. Species abundance and diversity of *Burkholderia cepacia* complex in the environment. *Appl. Environ. Microbiol.* 71(3), pp. 1193-1201.

Reis, V. et al. 2004. *Burkholderia tropica* sp. nov., a novel nitrogen-fixing, plant-associated bacterium. *International Journal of Systematic and Evolutionary Microbiology* 54(6), pp. 2155-2162.

Roitman, J. N. et al. 1990. A new chlorinated phenylpyrrole antibiotic produced by the antifungal bacterium *Pseudomonas cepacia*. *Journal of Agricultural and Food Chemistry* 38(2), pp. 538-541.

Rojas-Rojas, F. U. et al. 2017. Draft genome of *Paraburkholderia caballeronis* TNe-841 T, a free-living, nitrogen-fixing, tomato plant-associated bacterium. *Standards in Genomic Sciences* 12(1), p. 80.

Rusch, A. et al. 2015. *Burkholderia insulsa* sp. nov., a facultatively chemolithotrophic bacterium isolated from an arsenic-rich shallow marine hydrothermal system. *International Journal of Systematic and Evolutionary Microbiology* 65(1), pp. 189-194.

Rushton, L. et al. 2013. Key role for efflux in the preservative susceptibility and adaptive resistance of Burkholderia cepacia complex bacteria. *Antimicrobial Agents and Chemotherapy* 57(7), pp. 2972-2980.

Rutherford, K. et al. 2000. Artemis: sequence visualization and annotation. *Bioinformatics* 16(10), pp. 944-945.

Rutledge, P. J. and Challis, G. L. 2015. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nature Reviews Microbiology* 13(8), pp. 509-523.

Sadiq, M. A. et al. 2018. Phylogenetic Diversity of *Burkholderia pseudomallei* isolated from veterinary cases and the environments in Peninsular Malaysia. *Veterinary and Animal Science* 6, pp. 21-28.

Sahl, J. W. et al. 2016. The effects of signal erosion and core genome reduction on the identification of diagnostic markers. *MBio* 7(5), pp. e00846-00816.

Sawana, A. et al. 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, p. 429.

Sayers, E. W. et al. 2010. Database resources of the national center for biotechnology information. *Nucleic Acids Research* 39(suppl_1), pp. D38-D51.

Schadewaldt, H. 1975. Discovery of glanders bacillus. *Deutsche medizinische Wochenschrift (1946)* 100(44), pp. 2292-2295.

Seemann, T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30(14), pp. 2068-2069.

Segura, A. et al. 2009. Life of microbes that interact with plants. *Microbial Biotechnology* 2(4), p. 412.

Sessitsch, A. et al. 2005. *Burkholderia phytofirmans* sp. nov., a novel plant-associated bacterium with plant-beneficial properties. *International Journal of Systematic and Evolutionary Microbiology* 55(3), pp. 1187-1192.

Sheu, S.-Y. et al. 2012. *Burkholderia symbiotica* sp. nov., isolated from root nodules of Mimosa spp. native to north-east Brazil. *International Journal of Systematic and Evolutionary Microbiology* 62(9), pp. 2272-2278.

Singh, R. et al. 2017. Microbial metabolites in nutrition, healthcare and agriculture. *3 Biotech* 7(1), p. 15.

Smith, E. F. 1911. *Bacteria in relation to plant diseases: history, general considerations, vascular diseases*. Carnegie institution of Washington.

Smith, M. et al. 1997. Arabinose assimilation defines a nonvirulent biotype of *Burkholderia pseudomallei*. *Infection and Immunity* 65(10), pp. 4319-4321.

Song, L. 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(23), pp. 7974-7981.

Stewart, D. 1971. A composite arginine glucose medium for the characterization of *Pseudomonas aeruginosa* and other gram negative bacilli. *Journal of Applied Bacteriology* 34(4), pp. 779-786.

Stopnisek, N. et al. 2014. Genus-wide acid tolerance accounts for the biogeographical distribution of soil *Burkholderia* populations. *Environmental Microbiology* 16(6), pp. 1503-1512.

Suárez-Moreno, Z. R. et al. 2012. Common features of environmental and potentially beneficial plant-associated *Burkholderia*. *Microbial Ecology* 63(2), pp. 249-266.

Sultan, M. Z. et al. 2008. Novel oxidized derivatives of antifungal pyrrolnitrin from the bacterium *Burkholderia cepacia* K87. *The Journal of Antibiotics* 61(7), p. 420.

Sun, D. et al. 2020. Novel caffeine degradation gene cluster is mega-plasmid encoded in *Paraburkholderia caffeinilytica* CF1. *Applied Microbiology and Biotechnology*, pp. 1-12.

Tacconelli, E. et al. 2018. Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *The Lancet Infectious Diseases* 18(3), pp. 318-327.

Tamura, K. et al. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* 30(12), pp. 2725-2729.

Tatusov, R. L. et al. 1997. A genomic perspective on protein families. *Science* 278(5338), pp. 631-637.

Timmermann, T. et al. 2017. *Paraburkholderia phytofirmans* PsJN protects *Arabidopsis thaliana* against a virulent strain of *Pseudomonas syringae* through the activation of induced resistance. *Molecular Plant-Microbe Interactions* 30(3), pp. 215-230.

Tuanyok, A. et al. 2017. *Burkholderia humptydooensis* sp. nov., a new species related to *Burkholderia thailandensis* and the fifth member of the *Burkholderia pseudomalle*i complex. *Appl. Environ. Microbiol.* 83(5), pp. e02802-02816.

Urakami, T. et al. 1994. Transfer of *Pseudomonas plantarii* and *Pseudomonas glumae* to *Burkholderia* as *Burkholderia* spp. and Description of Burkholderia vandii sp. nov. *International Journal of Systematic and Evolutionary Microbiology* 44(2), pp. 235-245.

Uroz, S. and Oger, P. 2019. *Corrigendum* to" *Caballeronia mineralivorans* sp. nov., isolated from oak-Scleroderma citrinum mycorrhizosphere"[Syst. Appl. Microbiol. 40 (2017) 345-351]. *Systematic and Applied Microbiology* 42(3), pp. 423-424.

Valverde, A. et al. 2006. *Burkholderia ferrariae* sp. nov., isolated from an iron ore in Brazil. *International Journal of Systematic and Evolutionary Microbiology* 56(10), pp. 2421-2425.

Van Belkum, A. et al. 2013. Rapid clinical bacteriology and its future impact. *Annals of Laboratory Medicine* 33(1), pp. 14-27.

Vandamme, P. et al. 2013. *Burkholderia humi* sp. nov., *Burkholderia choica* sp. nov., *Burkholderia telluris* sp. nov., *Burkholderia terrestris* sp. nov. and *Burkholderia udeis* sp. nov.: Burkholderia glathei-like bacteria from soil and rhizosphere soil. *International Journal of Systematic and Evolutionary Microbiology* 63(12), pp. 4707-4718.

Vandamme, P. 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 and Evolutionary Microbiology* 47(4), pp. 1188-1200.

Vandamme, P. et al. 2007. *Burkholderia bryophila* sp. nov. and *Burkholderia megapolitana* sp. nov., moss-associated species with antifungal and plant-growth-promoting properties. *International Journal of Systematic and Evolutionary Microbiology* 57(10), pp. 2228-2235.

Vandamme, P. and Peeters, C. 2014. Time to revisit polyphasic taxonomy. *Antonie van Leeuwenhoek* 106(1), pp. 57-65.

Vandamme, P. et al. 2017. Comparative genomics of *Burkholderia singularis* sp. nov., a low G+ C content, free-living bacterium that defies taxonomic dissection of the genus Burkholderia. *Frontiers in Microbiology* 8, p. 1679.

Vanlaere, E. et al. 2006. Growth in Stewart's medium is a simple, rapid and inexpensive screening tool for the identification of *Burkholderia cepacia* complex. *Journal of Cystic Fibrosis* 5(2), pp. 137-139.

Vermis, K. et al. 2003. Isolation of *Burkholderia cepacia* Complex Genomovars from Waters. *Systematic and Applied Microbiology* 26(4), pp. 595-600. doi: http://dx.doi.org/10.1078/072320203770865909

Vial, L. et al. 2007. *Burkholderia* diversity and versatility: an inventory of the extracellular products. *Journal of Microbiology and Biotechnology* 17(9), pp. 1407-1429.

Viallard, V. et al. 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 and evolutionary microbiology* 48(2), pp. 549-563.

Vitorino, L. C. and Bessa, L. A. 2017. Technological microbiology: development and applications. *Frontiers in Microbiology* 8, p. 827.

Webster, G. et al. 2019. The Genome Sequences of Three *Paraburkholderia* sp. Strains Isolated from Wood-Decay Fungi Reveal Them as Novel Species with Antimicrobial Biosynthetic Potential. *Microbiology Resource Announcements* 8(34), pp. e00778-00719.

Wick, R. R. et al. 2017a. Completing bacterial genome assemblies with multiplex MinION sequencing. *Microbial Genomics* 3(10),

Wick, R. R. et al. 2017b. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Computational Biology* 13(6), p. e1005595.

Yabuuchi, E. et al. 1992. Proposal of *Burkholderia* gen. nov. and transfer of seven species of the genus *Pseudomonas* homology group II to the new genus, with the type species *Burkholderia cepacia* (Palleroni and Holmes 1981) comb. nov. *Microbiology and Immunology* 36(12), pp. 1251-1275.

Yabuuchi, E. et al. 1995. Transfer of Two *Burkholderia* and An Alcaligenes Species to *Ralstonia* Gen. Nov. Proposal of *Ralstonia pickettii* (Ralston, Palleroni and Doudoroff 1973) Comb. Nov., *Ralstonia solanacearum* (Smith 1896) Comb. Nov. and *Ralstonia eutropha* (Davis 1969) Comb. Nov. *Microbiology and Immunology* 39(11), pp. 897-904.

Yoo, S.-H. et al. 2007. *Burkholderia soli* sp. nov., isolated from soil cultivated with Korean ginseng. *International Journal of Systematic and Evolutionary Microbiology* 57(1), pp. 122-125.

Yoshihisa, H. et al. 1989. Production of antibiotics by *Pseudomonas cepacia* as an agent for biological control of soilborne plant pathogens. *Soil Biology and Biochemistry* 21(5), pp. 723-728.

Yu, M. et al. 2007. Isolation and characterization of a novel haloacid permease from *Burkholderia cepacia* MBA4. *Appl. Environ. Microbiol.* 73(15), pp. 4874-4880.

Zerikly, M. and Challis, G. L. 2009. Strategies for the discovery of new natural products by genome mining. *ChemBioChem* 10(4), pp. 625-633.

Zhang, H. et al. 2000. *Burkholderia kururiensis* sp. nov., a trichloroethylene (TCE)-degrading bacterium isolated from an aquifer polluted with TCE. *International Journal of Systematic and Evolutionary Microbiology* 50(2), pp. 743-749.

Zhu, H. et al. 2012. *Burkholderia dabaoshanensis* sp. nov., a heavy-metal-tolerant bacteria isolated from Dabaoshan mining area soil in China. *PloS one* 7(12), p. e50225.

Zuleta, L. F. G. et al. 2014. The complete genome of *Burkholderia phenoliruptrix* strain BR3459a, a symbiont of Mimosa flocculosa: highlighting the coexistence of symbiotic and pathogenic genes. *BMC genomics* 15(1), p. 535.

Zwetsloot, M. J. et al. 2020. Prevalent root-derived phenolics drive shifts in microbial community composition and prime decomposition in forest soil. *Soil Biology and Biochemistry*, p. 107797.

Chapter 9. Appendices

A-1 The result of *Burkholderia* Cardiff Collection (BCC) screening into different medium TSA, BSM-G, PCAT and BSM-GBP (0.5, 1, 11.5, 2, 2.5mM) of Bromoacetic acid and 600 units/mL of polymixin

	BCC	DIG	TSA	BSM-G	BSM-GBP 0.5mM		BSM-GBP 1mM		BSM-GBP 1.5mM		BSM-GBP 2mM		BSM-GBP 2.5mM		PCAT	
					24hrs	1week	24hrs	1 week	24hrs	1week	24hrs	1week	24hrs	1week	24hrs	1week
1.	*B. cepacia gv I* (BCC0002)	ENV	+	+	+	+	+	+	+	+	+	+	±	+	+	+
2.	*B. multivorans* (BCC0005)	CF	+	+	+	+	+	+	+	+	±	+	±	+	+	+
3.	*B. multivorans (*BCC0008)	CF	+	+	+	+	+	+	+	+	+	+	+	+	+	+
4.	*B. multivorans* (BCC0011)	CF	+	+	+	+	+	+	+	+	+	+	±	+	+	+
5.	*B. vietnamiensis* (BCC0042)	ENV	+	+	+	+	±	+	±	+	±	+	±	+	±	+
6.	BCC6 (BCC0044)	CF	+	+	+	+	+	+	+	+	+	+	+	+	+	+
7.	*B. arboris* (BCC0049)	NON	+	+	+	+	+	+	+	+	+	+	+	+	+	+
8.	*B. metallic* (BCC0095)	CF	+	+	+	+	+	+	+	+	+	+	+	+	+	+
9.	*B. cenocepacia* (BCC0097)	CF	+	+	+	+	+	+	+	+	+	+	+	+	+	+
10.	*B. diffusa* (BCC2) (BCC0106)	CF	+	+	+	+	+	+	±	+	+	+	+	+	+	+
11.	*B. diffusa* (BCC0109)	CF	+	+	+	+	+	+	+	+	+	+	+	+	+	+
12.	*B. contaminans* (BCC0123)	CF	+	+	+	+	+	+	+	+	+	+	+	+	+	+
13.	*B. cenocepacia* (BCC0127)	CF	+	+	+	+	+	+	+	+	+	+	+	+	+	+
14.	*B. pyrrocinia (gv IX)* (BCC0171)	ENV	+	+	+	+	+	+	+	+	+	+	+	+	+	+
15.	*B. lata* (BCC0147)	ENV	+	+	+	+	+	+	+	+	+	+	+	+	+	+
16.	*B. vietnamiensis* (BCC0190)	CF	+	+	+	+	+	+	±	+	±	+	±	+	±	+
17.	*B. vietnamiensis* (BCC0194)	ENV	+	+	+	+	+	+	+	+	+	+	-	-	+	+
18.	*B. cepacia gv I* (BCC0196)	ENV	+	+	+	+	+	+	+	+	+	+	+	+	+	+
19.	*B. cenocepacia* (BCC0202)	ENV	+	+	+	+	+	+	+	+	+	+	+	+	+	+
20.	*B. ambifaria* (BCC0203)	ENV	+	+	+	+	+	+	+	+	+	+	+	+	+	+
21.	*B. ambifaria* (BCC0207)	ENV	+	+	+	+	+	+	+	+	+	+	+	+	+	+
22.	*B. lata* (BCC0217)	CF	+	+	+	+	+	+	+	+	+	+	+	+	+	+
23.	*B. stabilis* (BCC0237)	ENV	+	+	+	+	+	+	+	+	+	+	+	+	+	+
24.	*B. gladioli* (BCC0238)	CF	+	+	+	+	+	+	+	+	+	+	±	+	+	+
25.	*B. stabilis* (BCC285)	NON	+	+	+	+	+	+	-	+	-	+	-	+	-	+
26.	*B. dolosa* (BCC0305)	CF	+	+	+	+	+	+	+	+	+	+	+	+	±	±
27.	*B. dolosa* (BCC0306)	CF	+	+	+	+	+	+	+	+	+	+	+	+	±	+
28.	*B. cepacia gv I* (BCC0310)	NON	+	+	+	+	+	+	±	+	+	+	±	+	+	+
29.	BCC4 (BCC0322)	CF	+	+	+	+	+	+	+	+	+	+	+	+	+	+
30.	*B. ambifaria* (BCC0372)	CF	+	+	±	+	-	+	-	+	-	-	-	-	+	+

31.	**BCC5 (BCC0397)**	NON	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
32.	**BCC6 (BCC0398)**	ENV	+	+	+	+	+	+	+	+	+	+	±	+	+	+	
33.	***B. anthina* (BCC0403)**	ENV	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
34.	***B. cenocepacia IIID* (BCC0506)**	CF	+	+	+	+	±	+	±	+	-	-	-	-	+	+	
35.	***B. contaminans* (BCC0633)**	NON	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
36.	***B. cenocepacia* (BCC0665)**	CF	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
37.	***B. glumae* (BCC0773)**	ENV	+	+	-	+	+	+	-	-	-	-	-	-	-	-	
38.	***B. plantarii* (BCC0777)**	ENV	+	+	+	+	+	+	+	+	+	+	±	+	+	+	
39.	***B. thailandensis* (BCC0779)**	ENV	+	+	+	+	+	+	+	+	-	+	-	+	-	+	
40.	***B. lata* (BCC0803)**	ENV	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
41.	***B. contaminans* (BCC0823)**	CF	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
42.	***B. cenocepacia* (BCC0999)**	ENV	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
43.	***B. pseudmultivorans* (BCC1191)**	CF	+	+	+	+	+	+	±	+	±	+	+	+	±	+	
44.	***B. cenocepacia* (BCC1202)**	CF	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
45.	***B. cenocepacia* (BCC1203)**	ENV	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
46.	***B. pyrrocinia* (BCC1346)**	ENV	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
47.	***B. ubonensis* (BCC1603)**	ENV	+	+	+	+	+	+	+	+	+	+	±	+	+	+	
48.	***B. oklahomensis* (BCC1605)**	ENV	+	+	+	+	+	+	+	+	±	+	-	+	-	+	
49.	***B. latens* (BCC1625)**	CF	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
50.	***B. seminalis (*BCC1627)**	CF	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
51.	***B. gladioli* (BCC1650)**	ENV	+	+	+	+	+	+	+	+	+	+	+	+	±	+	
52.	***B. stagnalis* (BCC1887)**	ENV	+	+	+	+	±	+	±	+	-	-	-	-	±	+	
53.	***B. territorii* (BCC1888)**	ENV	+	+	+	+	+	+	+	+	+	+	-	+	-	+	
54.	***B. ubonesis* (BCC1890)**		+	+	+	+	+	+	-	+	-	+	-	+	±	+	
55.	***B. lata* (BCC1736)**	ENV	+	+	+	+	+	+	+	+	±	+	±	+	±	+	
56.	***B. lata* (BCC1737)**	ENV	+	+	+	+	+	+	+	+	±	+	±	+	±	+	
57.	***B. cepacia* (CF001)**		+	+	+	+	+	+	+	+	+	+	+	+	+	+	
58.	***B. multivorans* (CF002)**		+	+	+	+	+	+	+	+	+	+	±	+	+	+	
59.	***B. multivorans* (CF003)**		+	+	+	+	+	+	+	+	±	+	±	+	+	+	
60.	***B. multivorans* (CF004)**		+	+	+	+	+	+	+	+	+	+	-	-	+	+	
61.	***B. cenocepacia* (CF005)**		+	+	+	+	+	+	+	+	+	+	+	+	±	+	
62.	***B. cenocepacia* (CF006)**		+	+	-	-	-	-	-	-	-	-	-	-	±	+	
63.	***B. cenocepacia* (CF007)**		+	+	+	+	+	+	+	+	+	+	+	+	+	+	
64.	***B. cenocepacia* (CF008)**		+	+	±	+	-	+	-	+	-	+	-	+	±	+	
65.	***B. cenocepacia* (CF009)**		+	+	+	+	+	+	+	+	+	+	±	+	+	+	
66.	***B. stabilis* (CF010)**		+	+	+	+	+	+	+	+	±	+	±	+	+	+	
67.	***B. vietnamiensis* (CF011)**		+	+	+	+	+	+	+	+	±	+	±	+	+	+	
68.	***B. dolosa* (CF012)**		+	+	+	+	+	+	+	+	±	+	±	+	-	-	
69.	***B. pyrrocinia* (CF013)**		+	+	+	+	+	+	+	+	+	+	-	-	+	+	
70.	***B.gladioli* (CF014)**		+	+	+	+	+	+	+	+	+	+	+	+	+	+	

A-2 The result of *Paraburkholderia* Cardiff Collection (BCC) screening into different medium TSA, BSM-G, PCAT and BSM-GBP (0.5, 1, 11.5, 2, 2.5mM) of Bromoacetic acid and 600 units/mL of polymixin

	BCC	DIG	TSA	BSM-G	BSM-GBP 0.5mM		BSM-GBP 1mM		BSM-GBP 1.5mM		BSM-GBP 2mM		BSM-GBP 2.5mM		PCAT	
					24hrs	1week	24hrs	1week	24hrs	1week	24hrs	1week	24hrs	1week	24hrs	1week
1.	***P. xenovorans* (BCC0657)**	ENV	+	+	+	+	+	+	+	+	±	+	-	+	±	+
2.	***P. andropogonis* (BCC0766)**	ENV	+	+	±	+	±	+	-	-	-	-	-	-	+	+
3.	***P. caledonica* (BCC0767)**	ENV	+	+	+	+	+	+	+	+	+	+	+	+	-	+
4.	***P. caledonica (BCC0768)***	ENV	+	+	-	-	-	-	-	-	-	-	-	-	+	+
5.	***P. caryophylli* (BCC0769)**	ENV	+	+	±	+	±	+	-	+	-	-	-	±	-	-
6.	***P. fungorum* (BCC0770)**	CF	+	+	+	+	+	+	+	+	+	+	+	+	+	+
7.	***P. glathei* (BCC0772)**	ENV	+	+	±	+	-	+	-	+	-	-	-	-	+	+
8.	***P. graminis* (BCC0774)**	ENV	+	+	+	+	+	+	+	+	+	+	±	+	+	+
9.	***P. kururiensis* (BCC0775)**	ENV	+	+	+	+	±	+	-	-	-	-	-	-	-	-
10.	***P. phenzinium* (BCC0776)**	ENV	+	+	-	-	-	-	-	-	-	+	-	+	-	+
11.	***P. sacchari* (BCC0778)**	ENV	+	+	+	+	+	+	+	+	+	+	+	+	+	+
12.	***P. phytofirmans* (BCC1209)**	ENV	+	+	+	+	+	+	+	+	+	+	+	+	+	+
13.	***P. phytofirmans* (BCC1604)**	ENV	+	+	+	+	+	+	+	+	+	+	±	+	-	+
14.	***P. mimosarum* (BCC1606)**	ENV	+	+	-	-	-	-	-	-	-	-	-	-	-	+
15.	***P. phymatum* (BCC1607)**	ENV	+	+	-	-	-	-	-	+	-	-	-	-	±	±
16.	***P. tropica* (BCC1608)**	ENV	+	+	+	+	+	+	+	+	+	+	-	+	+	+
17.	***P. terricola* (BCC1609)**	ENV	+	+	+	+	+	+	+	+	+	+	+	+	+	+
18.	***P. tuberum* (BCC1610)**	ENV	+	+	+	+	+	+	+	+	-	+	-	-	-	-
19.	***P. phenoliruptrix* (BCC1611)**	ENV	+	+	+	+	+	+	+	+	+	+	+	+	+	+
20.	***P. hospita* (BCC1612)**	ENV	+	+	+	+	+	+	-	-	-	-	-	-	+	+
21.	***Paraburkholderia* sp.(BCC1637)**	ENV	+	+	+	+	+	+	+	+	+	+	±	+	+	+
22.	***P. phenazinium* (BCC1873)**	ENV	+	+	+	+	+	+	+	+	+	+	±	+	+	+
23.	***P. bryophila* (BCC1876)**	ENV	+	+	+	+	+	+	+	+	±	+	±	+	±	+
24.	***Paraburkholderia* sp. (BCC1884)**	ENV	+	+	+	+	±	+	±	+	±	+	±	+	±	+
25.	***Paraburkholderia* sp. (BCC1885)**	ENV	-	-	-	-	-	-	-	-	-	-	-	-	±	±
26.	***Paraburkholderia* sp. (BCC1886)**	ENV	+	+	+	+	±	+	±	+	±	+	±	+	-	-

A-3 The result of *Non-Burkholderia* Cardiff Collection (BCC) screening into different medium TSA, BSM-G, PCAT and BSM-GBP (0.5, 1, 11.5, 2, 2.5mM) of Bromoacetic acid and 600 units/mL of polymixin

	BCC	DIG	TSA	BSM-G	BSM-GBP 0.5mM		BSM-GBP 1mM		BSM-GBP 1.5mM		BSM-GBP 2mM		BSM-GBP 2.5mM		PCAT	
					24hrs	1week	24hrs	1week	24hrs	1week	24hrs	1week	24hrs	1week	24hrs	1week
1.	***A.xylosoxidans* (BCC1386)**	Gram-negative	+	+	-	+	-	+	-	+	-	-	-	-	-	-
2.	***P. fluorescens* (BCC1388)**	Gram-negative	+	+	+	+	+	+	+	+	+	+	+	+	-	-
3.	***P. putida* (BCC1389)**	Gram-negative	+	+	-	-	-	-	-	-	-	-	-	-	-	-
4.	***P. stutzeri* (BCC1390)**	Gram-negative	+	+	+	+	+	+	-	+	-	+	-	+	-	+
5.	***R. mannitolytica* (BCC1391)**	Gram-negative	+	+	+	+	+	+	+	+	+	+	±	+	±	+
6.	***R. pickettii* (BCC1392)**	Gram-negative	+	+	+	+	+	+	+	+	+	+	+	+	+	+
7.	***S. aureus* (BCC1393)**	Gram-positive	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8.	***S. maltophilia* (BCC1394)**	Gram-negative	+	+	-	-	-	-	-	-	-	-	-	-	-	-
9.	***P. acidophila* (BCC1891)**	Gram-negative	+	+	+	+	+	+	+	+	+	+	+	+	+	+
10.	***P. aeruginosa* (Esh487)**	Gram-negative	+	+	-	-	-	-	-	-	-	-	-	-	-	±
11.	***E. coli* (Esh488)**	Gram-negative	+	+	-	-	-	-	-	-	-	-	-	-	-	-
12.	***S. aureus* (Esh489)**	Gram-positive	+	+	-	-	-	-	-	-	-	-	-	-	-	-
13.	***C. albicans* (Esh498)**	Gram-positive	+	+	+	-	+	+	+	+	+	+	-	-	±	±
14.	***K. pneumonia* (Esh510)**	Gram-negative	+	+	+	-	+	+	+	+	+	+	+	+	±	±
15.	***E. cloacae* (Esh511)**	Gram-negative	+	+	-	-	-	-	-	-	-	-	-	-	-	±
16.	***B. subtilis* (Esh566)**	Gram-positive	+	+	-	-	-	-	-	-	-	-	-	-	-	-
17.	***S. marcescens* (Esh781)**	Gram-negative	+	+	+	-	+	+	+	+	+	+	+	+	-	±
18.	***R. radiobacter* (Esh836)**	Gram-negative	+	+	-	-	-	-	-	-	-	-	-	-	-	-
19.	***P. syringae pv. syringae* (Esh837)**	Gram-negative	+	+	-	-	-	-	-	-	-	-	-	-	-	-

20.	***D. solani* (Esh841)**	Gram-negative	**+**	**+**	**-**	**±**	**-**	**±**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**
21.	***C. fungivorans* (Esh842)**	Gram-negative	**+**	**+**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**
22.	***R. mannitolylitica* (CF015)**	Gram-negative	**+**	**+**	**+**	**-**	**+**	**+**	**+**	**+**	**+**	**+**	**-**	**-**	**-**	**-**
23.	***R. respiraculi* (CF016)**	Gram-negative	**+**	**+**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**
24.	***R. pickettii* (CF017)**	Gram-negative	**+**	**+**	**+**	**-**	**+**	**+**	**+**	**+**	**+**	**+**	**-**	**-**	**-**	**-**
25.	***Ralstonia insidiosa* (CF018)**	Gram-negative	**+**	**+**	**+**	**-**	**+**	**+**	**+**	**+**	**+**	**+**	**-**	**-**	**-**	**-**
26.	***P. sputorum* (CF020)**	Gram-negative	**+**	**+**	**+**	**-**	**+**	**+**	**+**	**+**	**+**	**+**	**-**	**-**	**-**	**-**
27.	***P. pulmonicola* (CF021)**	Gram-negative	**+**	**+**	**+**	**-**	**+**	**+**	**+**	**+**	**+**	**+**	**-**	**-**	**-**	**-**
28.	***P. pnomenusa* (CF022)**	Gram-negative	**+**	**+**	**+**	**-**	**+**	**+**	**+**	**+**	**+**	**+**	**-**	**-**	**-**	**-**
29.	***S. maltophilia* (CF023)**	Gram-negative	**+**	**+**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**
30.	***P. apista* (CF019)**	Gram-negative	**+**	**+**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**
31.	***A. xylosoxidans* (CF024)**	Gram-negative	**+**	**+**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**
32.	***I. limonsus* (CF025)**	Gram-negative	**+**	**+**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**
33.	***A. baumannii* (BCC0807)**	Gram-negative	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**	**+**

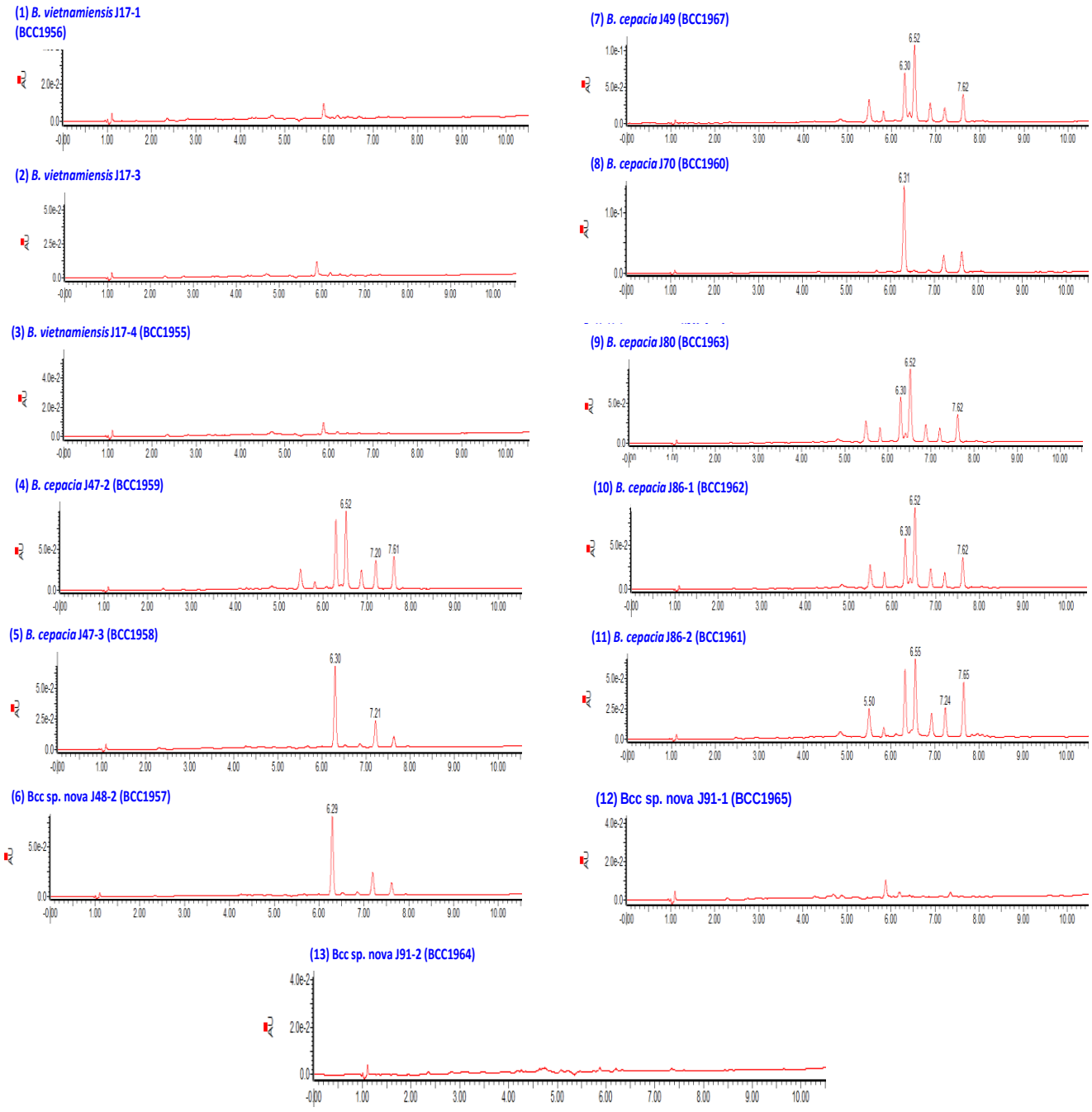


Figure A-1 HPLC analysis for 13 Bcc isolates in BSM-G pH 5 at 30°C

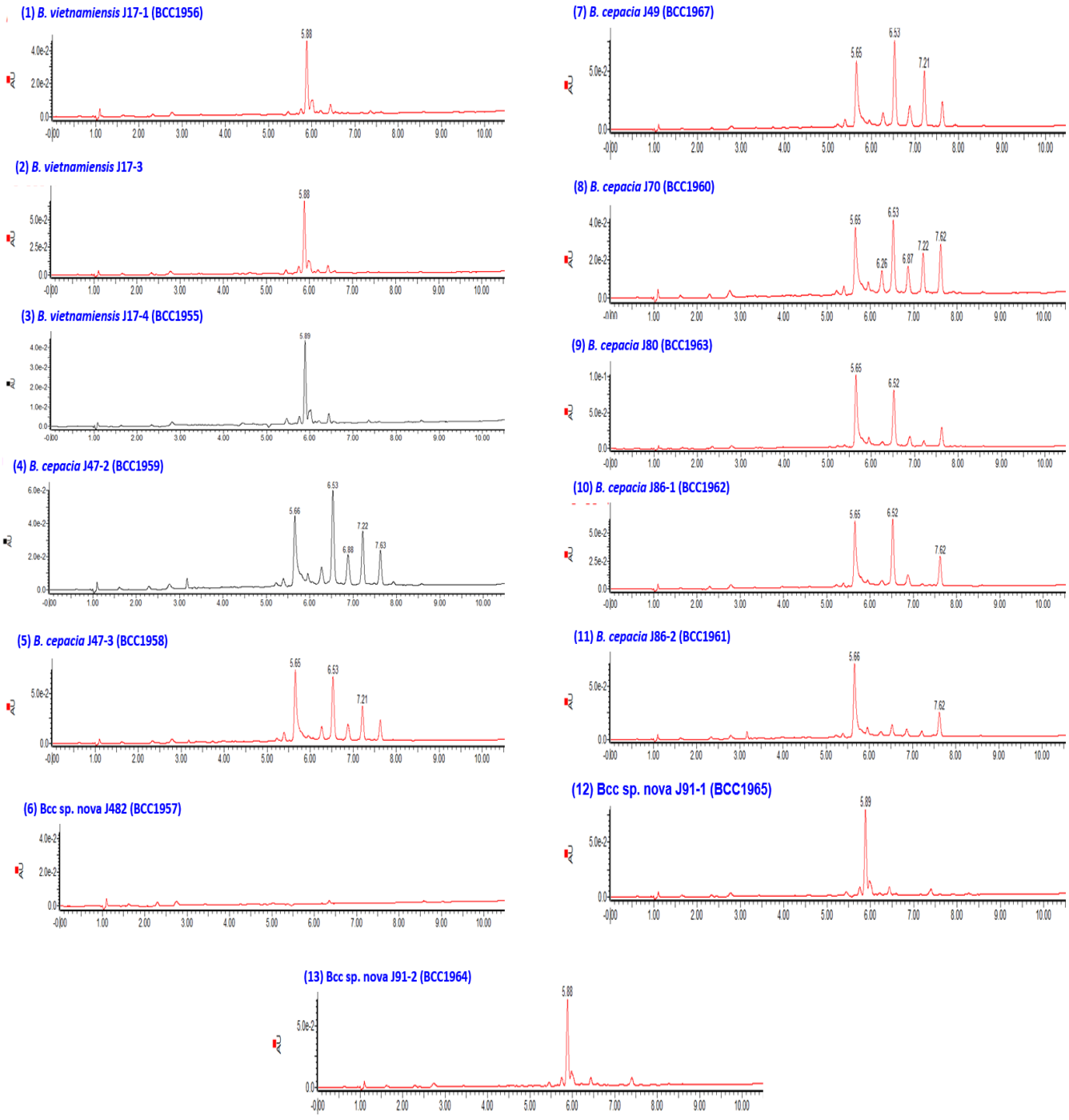


Figure A.2 HPLC analysis for 13 Bcc isolates in BSM-G pH 7 at 22°C