Understanding the genomics and specialised metabolites of the biopesticidal bacterium *Burkholderia ambifaria*

Thesis presented by

Alex James Mullins, BSc (Hons) Microbiology

In candidature for the degree of

Doctor of Philosophy

Microbiomes, Microbes, and Informatics Group

School of Biosciences

Cardiff University

August 2019

Acknowledgements

I express my sincere thanks to the following people:

My primary supervisor Professor Eshwar Mahenthiralingam for his continuous support and guidance throughout my PhD studies, and limitless expertise and knowledge of the field. It has been a pleasure to work as part of the Mahenthiralingam group, and has provided opportunities to progress my career in multiple directions.

My secondary supervisor Professor James Murray for his insights into plant biology.

My internship supervisors Dr Brian Anton and Sir Richard Roberts for the invaluable opportunity to work at New England Biolabs, Massachusetts, USA, and their support and guidance over the three month placement.

The Biotechnology and Biological Sciences Research Council (BBSRC) funding of the PhD through the South West Biosciences Doctoral Training Partnership (BB/M009122/1).

Dr Matthew Bull for his exhaustive knowledge of bioinformatics, valued advice, and overall encouragement throughout the four-year PhD.

Dr Simon Scofield and Joanne Kilby for their assistance in providing access to suitable facilities and technical support for plant biology lab-based experiments.

Thanks to all the members of the microbiology group, in particular, Dr Matthew Bull, Dr Laura Ruston, Dr Gordon Webster, Dr Rebecca Weiser, Dr Angharad Green, Dr Cerith Jones, Edward Cunningham-Oakes, and Hannah Salvage for their support, friendship, and for overall making the PhD a memorable experience.

Finally, I would to thank my mum, Rhian, my dad, Ray, and sister, Laura, for their encouragement and support throughout my studies.

Scientific Conferences and Awards

Exploiting and engineering the plant-protective and growth-promoting abilities of *Burkholderia* (2015). PhD Poster Session, Cardiff University, Cardiff, UK. Poster presentation.

Genomic and antimicrobial activity analysis of *Burkholderia ambifaria* as a prominent biological control species (2016). SWBio DTP First Year Event, University of Bristol, Bristol, UK. Poster presentation.

Understanding the biological control abilities of *Burkholderia ambifaria* using activity- and genomic-driven secondary metabolite analyses (2016). The Chemistry and Biology of Natural Products Symposium X (CBNP X), University of Warwick, Coventry, UK. Poster presentation.

Secondary metabolite bioactivities of *Burkholderia ambifaria* as a biological control agent (2016). Genomics Driven Approaches to Bioactive Natural Product Discovery, Syngenta, Jealott's Hill, UK. Poster presentation.

President's Fund (2017), £1200, Society for Applied Microbiology. Travel award

Genomic mining and plant pathogen inhibition analyses of *Burkholderia ambifaria* as a biological control agent (2017). American Society for Microbiology (ASM) Microbe, New Orleans, USA. Poster and oral presentation.

Genomic analysis updates the biotechnological potential of biopesticide *Burkholderia ambifaria* (2018). International *Burkholderia cepacia* Working Group (IBCWG) Meeting, Dublin, Ireland. Oral presentation.

Natural products of the biopesticide *Burkholderia ambifaria*: The functional basis of biological control (2018). The Chemistry and Biology of Natural Products Symposium XII (CBNP XII), University of Warwick, Coventry, UK. Oral presentation.

Exploring the antimicrobial properties of biopesticidal *Burkholderia*. (2018). The Chemistry and Biology of Natural Products Symposium XII (CBNP XII), University of Warwick, Coventry, UK. Poster presentation.

Genome mining specialised metabolites in *Burkholderia ambifaria* for novel antimicrobials (2019). Cardiff University Antimicrobial Resistance and Infection Network (CURE-infection), Early Career Researchers Event, Cardiff University, Cardiff, UK. Oral presentation.

The induction and suppression of biosynthetic gene clusters from *Burkholderia* species. (2018). Cardiff University Antimicrobial Resistance and Infection Network (CURE-infection), Early Career Researchers Event, Cardiff University, Cardiff, UK. Poster presentation.

Society Conference Grant (2019), £238, Microbiology Society. Travel award.

Genomic analysis of *Burkholderia ambifaria* identifies key specialised metabolites for biopesticidal applications (2019). Microbiology Society Annual Conference, Belfast, UK. Oral presentation. DOI: 10.1099/acmi.ac2019.po0224

Publications

Webster G, **Mullins AJ**, Bettridge AS, Jones C, Cunningham-Oakes E, Connor TR *et al.* Genome sequences of three *Paraburkholderia* spp. strains isolated from Wood-decay fungi reveals them as novel taxa with antimicrobial biosynthetic potential. *Microbiology Resource Announcements* 2019; 8: DOI: 10.1128/MRA.00778-19

Webster G, **Mullins AJ**, Watkins AJ, Cunningham-Oakes E, Weightman AJ, Mahenthiralingam E *et al.* Genome Sequences of Two Choline-Utilizing Methanogenic Archaea, *Methanococcoides* spp., Isolated from Marine Sediments. *Microbiology Resource Announcements* 2019; 8: e00342-19. DOI: 10.1128/MRA.00342-19

Mullins AJ, Murray JAH, Bull MJ, Jenner M, Jones C, Webster G *et al.* Genome mining identifies cepacin as a plant-protective metabolite of the biopesticidal bacterium *Burkholderia ambifaria*. *Nature Microbiology* 2019; 4: 996–1005. DOI: 10.1038/s41564-019-0383-z

Beaton A, Lood C, Cunningham-Oakes E, MacFadyen A, **Mullins AJ**, Bestawy W El *et al.* Community-led comparative genomic and phenotypic analysis of the aquaculture pathogen *Pseudomonas baetica* a390T sequenced by Ion semiconductor and Nanopore technologies. *FEMS Microbiology Letters* 2018; 365. DOI: 10.1093/femsle/fny069

Summary

Burkholderia ambifaria is a versatile bacterium frequently isolated from the environment in association with the rhizosphere of important crops species, and occasionally found as an opportunistic pathogen of cystic fibrosis patients. *B. ambifaria* strains were exploited successfully as biological pesticides during the 1990s, but declined in popularity following concerns over the pathogenicity of associated species in the *Burkhlderia cepacia* complex.

A collection of environmentally and clinically sourced *B. ambifaria* strains were sequenced with the purpose of developing a deeper understanding of the biopesticide. Comparative genomics were combined with *in vitro* metabolite analyses, antagonism assays, and agriculturally relevant biological control assays to determine the contribution of antimicrobial metabolites to biocontrol.

Genome mining for biosynthetic gene clusters (BGCs) revealed a considerable specialised metabolite potential, and multiple BGCs associated with characterised antimicrobials. Regulatory gene mining of quorum sensing associated *luxR* genes revealed an uncharacterised LuxRI system linked to an unknown BGC. Insertional mutagenesis and mass spectrometry confirmed the BGC as the biosynthetic origin of the historical *Burkholderia* polyyne metabolite cepacin. Comparison of the *B. ambifaria* BCC0191 wild-type with the cepacin-deficient mutant highlighted the importance of cepacin in the biological control of *Pythium ultimum* with a *Pisum sativum* crop model. The biocontrol phenotype was maintained following the deletion of the third replicon, and subsequent virulence testing in a murine respiratory inhalation model demonstrated a reduced persistence compared to the wild-type.

This study systematically defined the specialised metabolite biosynthetic potential of *B. ambifaria*, and demonstrated the importance of the polyyne cepacin in biological control. Maintenance of biocontrol and loss of virulence following third-replicon deletion presents an opportunity to attenuate *B. ambifaria* and address the pathogenicity concerns that led to the decline of *B. ambifaria* as a biopesticide.

Contents

List of Tables

List of Figures

Abbreviations

AHL	Acyl-homoserine-lactone
AMR	Antimicrobial resistance
AntiSMASH	Antibiotics and Secondary Metabolite Analysis Shell
ARG-ANNOT	Antibiotic Resistance Gene-ANNOTatio
ATCC	American Type Culture Collection
BCSA	*Burkholderia cepacia* selective agar
BIGSdb	Bacterial Isolate Genome Sequence Database
BGC	Biosynthetic gene cluster
BLAST	Basic local Alignment Search Tool
BLASTn	Nucleotide BLAST
BLASTp	Protein BLAST
bp	Base pairs
BSM-G	Basal salts media supplemented with glycerol
CARD	Comprehensive Antibiotic Resistance Database
CDS	Coding sequences
CF	Cystic Fibrosis
cfu	colony forming units
cfu/g	Colony forming units per gram
cfu/ml	Colony forming units per millilitre
CLIMB	Cloud Infrastructure for Microbial Bioinformatics
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DDH	DNA-DNA hybridisation
DI	De-ionised
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonucleotide triphosphates
ENA	European Nucleotide Archive
EPS	Extracellular polymeric substance
EU	European Union
GTR	General Time Reversible
HGT	Horizontal gene transfer
HMAQ	4-hydroxy-3-methyl-2-alkylquinoline
HMM	Hidden Markov model
HPLC	High Performance Liquid Chromatography
Kbp	Kilo base pairs
LB	Luria-Bertani
LMG	Belgian Co-ordinated Collections of Micro-organisms, Gent.
mbp	Mega base pairs

MIC	Minimum inhibitory concentration
MLST	Multi-locus sequence typing
MS	Mass spectrometry
m/z	Mass-to-charge (ratio)
NCBI	National Centre of Biotechnology and Information
NCTC	National Collection of Type Cultures
NRPS	Non-ribosomal peptide synthetase
OD	Optical Density
ONT	Oxford Nanopore Technology
PacBio	Pacific Biosciences
PBS	Phosphate buffer solution
PCR	Polymerase chain reaction
PDA	Potato dextrose agar
PKS	Polyketide synthase
QAC	Quaternary ammonium compound
QS	Quorum sensing
QUAST	Quality Assessment Tool for Genome Assemblies
R	R statistical software
RAPD	Random amplified polymorphic DNA
RiPP	Ribosomally synthesised and post-translationally modified peptide
RF	Radio frequency
RH	Relative humidity
rMLST	Ribosomal Multi Locus Sequence Typing
rRNA	Ribosomal ribonucleic acid
RND	Resistance Nodulating Division
RNA	Ribonucleic acid
rpm	Revolutions per minute
SMRT	Single molecule, real-time
SNPs	Single Nucleotide Polymorphisms
Taq	*Thermus aquaticus*
TOF-MS	Time of Flight-Mass spectrometry
TSA	Tryptone soya agar
TSB	Tryptone soya broth
TVC	Total viable count
UHPLC	Ultra high performance liquid chromatography
ESI-Q-TOF-MS	Electrospray ionization-quadrupole-time of flight-mass spectrometry
V	Volt
Vpp	Volts peak-to-peak
WT	Wild-type

1. Introduction

1.1. The origins and diversity of *Burkholderia* and associated genera

The evolutionary history of *Burkholderia* and closely related genera continues to be unravelled with the ongoing discovery of novel species and regular updates to the taxonomic distinction between genera. Members of the genus *Burkholderia* include clinically and agriculturally important pathogens, beneficial and commensal environmental bacteria, and species of biotechnological relevance. Originally isolated as a pathogen of onions and named *Pseudomonas cepacia* (Burkholder 1950), the bacterium lost its nomenclature validity following omission from the Applied Lists of Bacterial Names. However, the name was re-instated in accordance with Rule 28a of the International Code of Nomenclature of Bacteria (Palleroni and Holmes 1981; Lapage *et al.* 1992). The species was later re-classified as *Burkholderia cepacia* and accompanied by the migration of multiple species previously described as *Pseudomonas*, marking the origins of the genus (Yabuuchi *et al.* 1992). The genus *Burkholderia* was named after William Burkholder who originally isolated the organism (Burkholder 1950).

Interestingly, the designated type strain *Burkholderia cepacia* was later proven to represent a complex at least 5 genetically distinct species (genomovars), and subsequently referred to as the *Burkholderia cepacia* complex (Vandamme *et al.* 1997). Ten genomovars were eventually characterised and designated as distinct species as follows (genomovar I to X): *Burkholderia cepacia* (Vandamme *et al.* 1997), *Burkholderia multivorans* (Vandamme *et al.* 1997), *Burkholderia cenocepacia* (Vandamme *et al.* 2003), *Burkholderia stabilis* (Vandamme *et al.* 2000), *Burkholderia vietnamiensis* (Gillis *et al.* 1995), *Burkholderia dolosa* (Vermis *et al.* 2004), *Burkholderia ambifaria* (Coenye *et al.* 2001a), *Burkholderia anthina* (Vandamme *et al.* 2002), *Burkholderia pyrrocinia* (Vandamme *et al.* 2002), and *Burkholderia ubonensis* (Vermis *et al.* 2002). Members of the *B. cepacia* complex possess multireplicon genomes consisting of three primary replicons: c1, c2 and c3, and various plasmids (Mahenthiralingam *et al.* 2005). The smallest of the primary replicons, c3, was subsequently confirmed to be a large virulence and stress-tolerance plasmid (Agnoli *et al.* 2014). Despite the conservation of the c3 megaplasmid within the complex, the plasmid possesses considerable sequence plasticity, demonstrating an impressive ability to capture virulence factors and specialised metabolite biosynthetic gene clusters.

Differentiating these closely related species was necessary to improve understanding of the role of *Burkholderia* in CF epidemiology. Initial identification based on sequencing hypervariable regions of the 16S rRNA gene proved insufficient in differentiating the *B. cepacia* complex species with their considerable genomic heterogeneity. The housekeeping gene *recA* encoding a protein involved in homologous recombination was adopted as an alternative gene to distinguish members of the *B. cepacia* complex (Mahenthiralingam *et al.* 2000). Unlike the 16S rRNA gene sequence that provided an insufficient resolution of the complex at 98-99% sequence similarity, the *recA* gene sequence delivered a higher resolution 94-95% sequence similarity between species (Coenye *et al.* 2001c). Further developments in phylogenetic analysis of *Burkholderia* incorporated an expanded gene portfolio of multi-locus sequence typing (MLST) initially exploiting seven genes: *atpD*, *gltB*, *gyrB*, *recA*, *lepA*, *phaC*, and *trpB* (Baldwin *et al.* 2005). This was later expanded to ribosomal MLST (rMLST) with the increased uptake in genomic sequencing in microbiology (Jolley *et al.* 2012) which provided a distinction between *B. cepacia* complex species and the wider *Burkholderia* genus and sister clades (Depoorter *et al.* 2016). Novel *B. cepacia* complex species continue to be isolated and characterised, with the most recent addition *Burkholderia puraquae* validly named in 2018, bringing the total number of *B. cepacia* complex members to at least 21 species (Martina *et al.* 2018).

Beyond the *B. cepacia* complex two major groups are recognised, the Pseudomallei group and the phytopathogen group (Depoorter *et al.* 2016). The Pseudomallei group includes the Category B bioterrorism agents *Burkholderia pseudomallei* and *Burkholderia mallei* responsible for melioidosis and glanders, respectively (Biggins *et al.* 2012). The remaining species in the group include *Burkholderia thailandensis*, *Burkholderia oklahomensis* and *B. humptydooensis* (Ginther *et al.* 2015); the former species *B. thailandensis* functions as a useful model for the human pathogenic species. The phytopathogenic group is composed of three species: *Burkholderia gladioli*, *Burkholderia glumae* and *Burkholderia plantarii*, originally described as the etiologic agents of sheath rot, grain rot, and seedling blight in rice, respectively (Seo *et al.* 2015).

The expansion of the *Burkholderia* genus combined with high-throughput whole genome sequencing over the past decade led researchers to identify a major split in the genus phylogeny (Sawana *et al.* 2014). One clade retained the *Burkholderia* designation and consisted of the aforementioned opportunistic pathogens, including the *B. cepacia* complex, Pseudomallei group and phytopathogen group. While the second clade encompassed mainly commensal and beneficial environmental bacteria, and was subsequently designated

Paraburkholderia (Sawana *et al.* 2014). The oversimplification of defining the former clade as pathogenic and the latter as lacking human pathogenicity has been highlighted (Vandamme and Peeters 2014). Specifically, it was emphasised that most of the research into the beneficial properties of *Burkholderia* and *Paraburkholderia* has been focussed on members of the supposedly pathogenic clade; and several *Paraburkholderia* species have been associated with both plant, animal and human pathogenicity (Vandamme and Peeters 2014).

In 2016 a further sub-division was proposed based on the formation of a phylogenetic clade distinct from *Paraburkholderia*, referred to as *Caballeronia* (Dobritsa and Samadpour 2016). This was followed by the re-classification of *Burkholderia andropogonis* as *Robbsia andropogonis* (Lopes-Santos *et al.* 2017) which remains the only member of the genus. The most recent proposals in '*Burkholderia*' taxonomy was the re-classification of several outliers in the *Paraburkholderia* clade as either the plant and soil-associated *Trinickia* or the fungal-associated *Mycetohabitans* (Estrada-de los Santos *et al.* 2018).

1.2. Specialised metabolites and biosynthetic potential of *Burkholderia*

Bacterial specialised (secondary) metabolites are natural products not involved in primary metabolism but serve several broad functions centralised around survival and/or lifestyle, such as virulence, antagonism, stress tolerance and cell-to-cell signalling (Demain and Fang 2000). Anthropogenic uses of these bacterial natural products predominantly involve medicinal applications, such as antimicrobial agents, statins, anti-cancer agents and immunosuppressants. The first antimicrobial specialised metabolite to be extracted from the producing organism was arguably the antibiotic penicillin, discovered in 1928 and synthesised by the fungus *Penicillium chrysogenum* (formerly *P. notatum*) (Gould 2016). However, the antibiotic streptomycin derived from *Streptomyces griseus* (formerly *Actimomyces griseus*) was the first of many bacterial-derived antimicrobials to be exploited in medicine, agriculture or research (Gould 2016). Antimicrobials are arguably the most characterised class of natural products due to the relative simplicity in screening assays; that is, identifying bacteria or culture extracts capable of inhibiting the growth of fungi, nematodes or other bacteria. The high-throughput screening combined with a demand for medically and agriculturally important antimicrobials has inspired research into multiple bacterial genera other than *Streptomyces*. Gram-negative bacteria have been identified as an unexploited source of natural products, specifically antimicrobials, with a plethora of compounds discovered in genera *Pseudomonas* and *Burkholderia,* and broader taxonomic groups cyanobacteria and myxobacteria (Masschelein *et al.* 2017).

Specialised metabolites are the products of biosynthetic gene clusters (BGCs) which possess considerable diversity as exemplified by the structures of the resulting compounds. However, many of these BGCs remain silent under laboratory culture conditions, and therefore represent a considerable unexploited wealth of natural products for human use (Rutledge and Challis 2015). *Burkholderia* species dedicate a significant proportion of their large multireplicon genomes to specialised metabolism that is higher than both *Pseudomonas* and *Bacillus*, and similar in capacity to *Streptomyces* (Chevrette and Currie 2019). A large proportion of *Burkholderia* biosynthetic sequence capacity consists of large polyketide synthases (PKSs), non-ribosomal peptide synthetases (NRPSs) or hybrid pathways (NRPS-PKS), which are known to encode antimicrobials (Mahenthiralingam *et al.* 2011; Masschelein *et al.* 2017; Song *et al.* 2017; Mullins *et al.* 2019). This multiplicity of BGCs also presents an opportunity to discover rare or unusual chemistry such as the enacyloxin BGC that possesses both *cis*- and *trans*-acyltransferase domains (Mahenthiralingam *et al.* 2011), and unusual biosynthetic architecture of the *B. gladioli* derived icosalide BGC (Jenner *et al.* 2019).

The antifungal and antibacterial compound pyrrolnitrin was the first instance of a specialised metabolite to be discovered in *Burkholderia* (Arima *et al.* 1964). The compound was later found to be widespread in the *B. cepacia* complex with the gene cluster detected in at least five species (Schmidt *et al.* 2009a), and later detected in *Pseudomonas*, *Serratia* and *Pandorea* and *Myxococcus* (Masschelein *et al.* 2017). Further antimicrobials and cytotoxic compounds were characterised in the following decades in all three major clades of the *Burkholderia* genus (Masschelein *et al.* 2017). Notable examples of antimicrobial specialised metabolites discovered in *Burkholderia* species that attest to the diversity of compounds are enacyloxin IIa, gladiolin and capistruin. The polyketide antibiotic enacyloxin IIa found in *B. ambifaria* with activity against pan-resistant Gram-negative human pathogen *Acinetobacter baumannii* (Mahenthiralingam *et al.* 2011). *B. gladioli* in the phytopathogenic group synthesises the macrolide antibiotic gladiolin that exhibited strong anti-mycobacterial activity and low human cytotoxicity (Song *et al.* 2017); while the Pseudomallei group bacterium *B. thailandensis* encodes a 4.5 kbp gene cluster responsible for the synthesis of a ribosomally synthesised and post-translationally modified peptide (RiPP) capistruin with narrow antibiotic spectrum (Knappe *et al.* 2009).

In addition to antimicrobial compounds, *Burkholderia* produce specialised metabolites that contribute to pathogenicity in animals and humans. The virulence factor malleilactone is present in *B. thailandensis* and other members of the Pseudomallei group (Vial *et al.* 2008). A mutant knockout of this pathway in *B. thailandensis* attenuated its capacity to kill the nematode *Caenorhabditis elegans*, and enabled the amoeba *Dictyostelium discoideum* to complete differentiation into multicellular fruiting bodies (Biggins *et al.* 2012). Multiple *Burkholderia* species also synthesis siderophores for iron acquisition from the host during an infection. The siderophore ornibactin is widespread in the *B. cepacia* complex (Butt and Thomas 2017) and proven to contribute to persistence of *Burkholderia* in a murine lung infection model (Sokol *et al.* 1999), while members of the *Pseudomallei* group possess its counterpart, malleobactin (Butt and Thomas 2017). In addition to the primary siderophore, many *B. cepacia* complex species also possess secondary siderophores, such as cepaciachelin and pyochelin (Butt and Thomas 2017).

Signalling molecules are a type of specialised metabolite that facilitate and co-ordinate the gene expression in bacteria, and can influence phenotypes such as virulence, metabolite biosynthesis, biofilm formation and motility. Two well characterised signalling systems in *Burkholderia* involve homoserine lactones and hydroxyquinolines. Homoserine lactones are

associated with quorum sensing, the cell-density dependent co-ordinated gene expression in a bacterial population, and are known to regulate the biosynthesis of a plethora of metabolites: enacyloxin (Mahenthiralingam *et al.* 2011), pyrrolnitrin (Schmidt *et al.* 2009b), fragin (Jenul *et al.* 2018), bactobolins (Seyedsayamdost *et al.* 2010), burkholdines (Chapalain *et al.* 2013) and cepacin (Mullins *et al.* 2019). The hydroxyquinolines, specifically 4-hydroxy-3-methyl-2-alkylquinolines (HMAQs), are the methylated derivatives of the signalling molecules discovered in *Pseudomonas* (Vial *et al.* 2008). HMAQs were shown to repress the quorum sensing regulon in *Burkholderia* (Vial *et al.* 2008; Chapalain *et al.* 2017), and function as antimicrobials themselves (Kilani-Feki *et al.* 2011; Mahenthiralingam *et al.* 2011).

While there exists a multitude of publications attesting to the biosynthetic diversity and specialised metabolism of *Burkholderia* (Masschelein *et al.* 2017), a paucity of compounds have been discovered in *Paraburkholderia* and the recently classified related genera. Few examples exist of bioactivity in these environmental-termed clades in comparison to *Burkholderia*. The fungal endosymbiont *Trinickia rhizoxinica* (formerly *Paraburkholderia rhizoxinica*) was proven as the producer of the anti-mitotic compound rhizoxin (Partida-Martinez and Hertweck 2007). While antifungal activity was observed in the moss-associated *Paraburkholderia bryophila* and *Paraburkholderia megapolitana* (Vandamme *et al.* 2007), the compound or biosynthetic gene cluster responsible for this activity has yet to be identified. A member of the recently classified *Trinickia* genus, *Trinickia caryophilli* (formerly *Paraburkholderia caryophilli*), possessed potent anti-Gram-positive polyynes caryoynencins (Kusumi *et al.* 1987), for which the gene cluster was later identified (Ross *et al.* 2014).

1.3. Plant-microbe interactions and biological control

Microorganisms are known to interact with humans, animals, plants, as well as other microbes; and the type of interaction is classified by the impact on the host species, which can be beneficial, commensal or detrimental. Beneficial plant-microbe interactions have been exploited in agriculture for centuries in the form of nitrogen-fixing nodule-forming bacteria with species in the family Fabaceae, commonly referred to as legumes, such as clover, peas, and alfalfa. Specific microbes are also capable of stimulating plant growth and increasing crop yield; such the fungus *Trichoderma hamatum* through the secretion of plant growth-promoting compounds (Studholme *et al.* 2013).

Burkholderia and closely related genera have been isolated from the rhizospheres of many plants species, including major crop species globally, including rice in Vietnam (Van Tran *et al.* 1996), maize in the US (Ramette *et al.* 2005), sugarcane (Paungfoo-Lonhienne *et al.* 2014) and wheat (Balandreau *et al.* 2001) in Australia, onion in the UK (Richardson *et al.* 2002), and coffee in Mexico (Estrada-De Los Santos *et al.* 2001). These bacteria can exist in the rhizosphere in a beneficial capacity to the host plant, and multiple species have been described under the umbrella term of plant growth promoting rhizobacteria (PGPR). A variety of plant growth promotion mechanisms have been characterised, from atmospheric nitrogen fixation and phosphate sequestration, to phytohormone synthesis (Compant *et al.* 2008). These functions all support the growth of the plant and increase above and below ground biomass, resulting in greater agricultural yields.

The model organism *Arabidopsis thaliana* was subject to RNA-sequencing to understand the impact of *Paraburkholderia phytofirmans* on plant growth promotion (Poupin *et al.* 2013). The environmental bacterium was found to modulate plant gene expression to elicit growth hormones production resulting in significantly higher biomass, including greater chlorophyll content (Poupin *et al.* 2013). A dual RNA-sequencing approach was applied to the study of *Burkholderia* sp. Q208 and stimulation of sugarcane growth, to gain an insight into the gene expression changes in both organisms (Paungfoo-Lonhienne *et al.* 2016). Genes involved in biofilm formation and several metabolic pathways for energy production were upregulated in the *Burkholderia* (Paungfoo-Lonhienne *et al.* 2016). While genes associated with ethylene production were upregulated in the sugarcane plant resulting in the enhanced growth of the root aerenchyma facilitating oxygen transfer (Paungfoo-Lonhienne *et al.* 2016). *B. ambifaria* is also capable of stimulating plant growth in the absence of direct contact with the plant host through the release of volatile compounds (Groenhagen *et al.* 2013). The growth promotion

phenotype was shown to be independent of the strain origin, environmental or clinical, and the result of a plethora of volatile compounds (Groenhagen *et al.* 2013).

However, in contrast to PGPR, plant pathogenic microbes have a detrimental impact on the host and represent a considerable burden to agriculture. Notable examples of pathogens that inflict major economic consequences on the agricultural industry include the bacterial genus *Xanthomonas*, capable of infecting over 400 plant species including crop species of rice, banana and pepper (Ryan *et al.* 2011). Due to the diversity of the genus and range of host species, the pathogen causes a multitude of diseases from wilting of banana trees to leaf streak in cereal crops (Ryan *et al.* 2011). Damping-off disease causes the death of germinating seeds (pre-emergence) and emerging seedlings (post-emergence), and is caused by multiple fungal and oomycetal species (Lamichhane *et al.* 2017). The fungal pathogens *Fusarium* and *Rhizoctonia*, and oomycetal pathogens *Pythium* and *Phytophthora*, represent the most frequently isolated organisms during disease outbreaks, capable of reducing crop yield up to 80% (Lamichhane *et al.* 2017). Significant socio-economic costs are associated with crop diseases worldwide. Following an assessment of fungal disease prevalence in the five high priority agricultural crop species, it was estimated that mitigation of low level disease persistence would be sufficient to feed approximately 600 million people (Fisher *et al.* 2012). Substantial funds are invested into mitigating the impact of crop diseases, with the global pesticide market estimated at over $25 billion USD in 2010 (Bailey *et al.* 2010).

Despite the urgent requirement of synthetic pesticides in agriculture, there is a growing demand by consumers for non-synthetic alternatives due to concerns over detrimental effects to the environment and risk to human health (Glare *et al.* 2012). For example, the European Commission has not approved the renewal of synthetic fungicide thiram due to concerns over consumer and worker exposure and the risk to birds and mammals (European Food Safety Authority 2017; European Commission 2018). Public opinion and reductions to acceptable pesticide residue on crops has accelerated research into biological alternatives (Glare *et al.* 2012). Biopesticides are living organisms or their natural products that are capable of suppressing the growth or spread of crop pathogens or pests (Glare *et al.* 2012), and represent a small but expanding fraction of the global pesticide market (Bailey *et al.* 2010). Products based on *Bacillus thuringiensis* dominate the biopesticides market as the most popular form of biological pest control globally (Glare *et al.* 2012). *B. thuringiensis* as a species has the capacity to synthesise a myriad of toxins, mycolytic enzymes, and antimicrobials endowing the biopesticide with the ability to target a myriad of invertebrate classes (Mnif and Ghribi 2015). The diversity of pests targeted by specific toxins, combined with the lack of *B.*

thuringiensis human pathogenicity, has contributed to their rise in the biological control sector. *Pseudomonas* has also been exploited as a biopesticide due to its considerable biosynthetic capacity responsible for the synthesis of a myriad of antimicrobials (Mark *et al.* 2006). Several *Pseudomonas* species have been incorporated into commercial biocontrol products targeting fungal and oomycetal pathogens, such as *Pseudomonas chlororaphis* in the product Cedemon® manufactured by BioAgri designed for cereal crops (Mark *et al.* 2006).

Burkholderia represent a valuable biopesticidal resource for agriculture, and have been exploited previously in several products, however, concerns over pathogenicity triggered the decline in *Burkholderia*-based products (Parke and Gurian-Sherman 2001). Recently there has been a revival in the concept of *Burkholderia* biopesticides following advances in genomics and our understanding of *Burkholderia* in cystic fibrosis infections. Several *Burkholderia* heat-killed products possessing nematocidal and mitocidal properties have been marketed in the US by Marrone Bio Innovations®, using the potential Pseudomallei group member *Burkholderia rinojensis* A396 (Cordova-Kreylos *et al.* 2013).

1.4. Biotechnological applications of *Burkholderia*

In addition to the potential of *Burkholderia* and related genera as biological control agents, the multitalented bacteria are of interest to the wider field of biotechnology due to their large and diverse genomic and metabolic capacity. The fraction of the genome dedicated to specialised metabolism varies considerably between *Burkholderia* species, even within phylogenetic groups such as the *B. cepacia* complex (Depoorter *et al.* 2016). Aside from the exploitation of these compounds as biopesticides, they represent an untapped source of novel pharmaceuticals (Depoorter *et al.* 2016). The macrolide antibiotic gladiolin synthesised by *B. gladioli* represents a promising candidate as a novel *Mycobacterium tuberculosis* therapeutic, with activity against multi-drug resistant *M. tuberculosis* clinical isolates (Song *et al.* 2017). The polyketide antibiotic enacyloxin IIa exhibited anti-Gram-negative specificity towards CF pathogens *B. dolosa* and *B. multivorans*, and a multi-resistant clinical isolate of *A. baumannii* (Mahenthiralingam *et al.* 2011). The compound rhizoxin initially thought to be fungal-derived was synthesised by a *Burkholderia* endosymbiont, *Burkholderia rhizoxina* (Scherlach *et al.* 2006). This compound displayed anti-tumour properties in a murine-tumour model, and progressed to stage II clinical trials for breast cancer and melanoma, but displayed limited activity (Hanauske *et al.* 1996).

The diverse metabolic capacity of *Burkholderia* confers the ability to colonise a plethora of environments of variable nutrient availability due to their ability to exploit multiple compounds as a primary carbon source (Mahenthiralingam *et al.* 2005). This talent extends to metabolising pollutants in the environment and even antibiotics as carbon sources (Mahenthiralingam *et al.* 2005). Multiple *Burkholderia* and *Paraburkholderia* species possess the catabolic enzymes required to degrade phenol and benzene derivatives, as well as halogenated compounds illustrating potential as bioremediation agents (O'Sullivan and Mahenthiralingam 2005). The degradation, and thereby removal, of toxic substances from the environment via biological means is described as bioremediation. *B. vietnamiensis* G4 is an extensively characterised strain capable of degrading trichloroethylene, an abundant ground water aquifer contaminant in the US (O'Sullivan and Mahenthiralingam 2005). This ability derives from a catabolic-plasmid carrying a toluene *o*-monoxygenase gene, and has been exploited in model aquifer systems to remove simulated pollutant, resulting in two US patents on the G4 strain (O'Sullivan and Mahenthiralingam 2005).

While the organisms themselves have restricted use for environmental applications due to concerns over pathogenicity to immunocompromised individuals, the enzymes they synthesise have been exploited in industry as biological catalysts. A strain of *B. cepacia* (formerly *Pseudomonas cepacia*) was known to produce an extracellular lipase with a broad substrate specificity, heat resistance and solvent tolerance (Sánchez *et al.* 2018). These properties have led to the enzymes expansive usage in industrial biotechnology. The *B. cepacia* lipase (BCL) has been explored as a means of catalysing regioselective and stereoselective reactions to generate pure compounds for pharmaceutical applications (Sánchez *et al.* 2018). The structural and stereochemistry of drugs is important due to the lack of activity or detrimental effects of one isomer versus the beneficial biological function of the opposite isomer (Sánchez *et al.* 2018). Immobilised BCL was examined as a possible method of generating biodiesel through the transesterification of soybean oils with methanol, outperforming existing commercial immobilised lipase systems (Li *et al.* 2017). Unlike other genera such as *Pseudomonas, Burkholderia* have seen limited use as a source of enzymes in the detergent industry (Boran and Ugur 2016). A screen of wastewater samples for lipase activity identified a *B. multivorans* strain that exhibited promising activity in the presence of commercial detergents (Boran and Ugur 2016). Further characterisation revealed the purified lipase to tolerate several surfactants and oxidising agents, such as Tween 40 and sodium hypochloride, respectively; in addition to commercial detergent formulas (Boran and Ugur 2016).

With the exception of several extensively characterised strains and purified enzymes, *Burkholderia* remains a valuable but underexploited resource in multiple biotechnology industries from bioremediation to biofuels and pharmaceuticals. Advances in high throughput screening combined with cost-effective genome sequencing and refined genetic manipulation tools provides an opportunity to capitalise on the hidden diversity of *Burkholderia* natural products and extracellular enzymes.

1.5. Clinical history and pathogenicity of *Burkholderia*

Members of the genus *Burkholderia* have been well characterised regarding their pathogenic tendencies in humans, animals and plants. *B. mallei* and *B. pseudomallei* represent primary pathogens of humans and animals, and are the causative agents of mellioidosis and glanders, respectively (Biggins *et al.* 2012). The moderate fatality rate associated with these pathogens in the absence of appropriate antibiotic treatment, and the possibility of *B. pseudomallei* infection via inhalation, ingestion or cutaneous exposure (Limmathurotsakul *et al.* 2016) are reasons why these bacteria are classified as Category B bioterrorism agent (Biggins *et al.* 2012). Originally characterised as a plant pathogen, *B. gladioli* also represents a pathogen outside the *B. cepacia* complex in cystic fibrosis patients. *B. gladioli* is more frequently isolated from sputum samples of cystic fibrosis patients than most members of the *B. cepacia* complex, and represents a significant fraction of the *Burkholderia* burden in CF clinics (LiPuma 2010). In the US between 1997 and 2007 isolates were obtained from over 2000 *Burkholderia*-infected CF patients which revealed *B. gladioli* as the third most prevalent *Burkholderia* species responsible for 15% of cases (LiPuma 2010).

The *B. cepacia* complex is known for opportunistic pathogenicity in immunocompromised patients. Of the 10 *B. cepacia* complex members described up to 2010, 9 have been isolated from CF patients, while *B. ubonensis* is the only *B. cepacia* species yet to be associated with CF (LiPuma 2010). The prevalence of '*Pseudomonas cepacia*' was observed to increase between 1972 and 1981 among CF patients in Canada, and was therefore recognised as an emerging pathogen in the CF community (Isles *et al.* 1984). Several patients experienced a rapid decline in lung function following *P. cepacia* infection, resulting in greater than 60% fatality (Isles *et al.* 1984). A similar study in the US also reported an increase in patient colonisation with *P. cepacia* and unexpected, often fatal, declines in CF patient health (Thomassen *et al.* 1985). This clinical manifestation of chronic *P. cepacia* infection was later quantified as occurring in approximately 10% of patients, and referred to as cepacia syndrome (Mahenthiralingam *et al.* 2008). The rapid onset and high fatality rate associated with cepacia syndrome was a clinical outcome not observed in other chronic CF pathogen infections such as *Pseudomonas aeruginosa*, and therefore represented an emerging challenge for the CF research and clinical community (Mahenthiralingam *et al.* 2008). Exacerbating this virulence as an emerging pathogen, evidence of patient-patient transmission was observed within CF clinics (Smith *et al.* 1993), between regional CF clinics (Govan *et al.* 1993), and international transmission between Canadian and UK clinics (Sun *et al.* 1995). Subsequent analyses via genotyping revealed the epidemic was the result of the clonal *B. cenocepacia* strain ET12 that likely originated in Canada and spread to UK CF clinics (Mahenthiralingam *et al.* 2008).

Improvements in isolate identification and infection control through patient segregation based on the infecting *Burkholderia* reduced the spread of *B. cenocepacia* between patients, and ultimately ended the epidemic in CF clinics (Mahenthiralingam *et al.* 2001). Continued surveillance in the US of *Burkholderia* in CF patients has uncovered a shift in the dominant infecting *Burkholderia* species from *B. cenocepacia* to *B. multivorans* between 1997 and 2007 (LiPuma 2010). A recent study confirmed the dominance of *B. multivorans* in the UK representing 56% of CF patient isolates, while *B. cenocepacia* prevalence was only 15% (Kenna *et al.* 2017). *B. ambifaria* was not present in the 1,047-isolate dataset collected over a two-year study period (Kenna *et al.* 2017); and combined with several other *B. cepacia* complex species represented less than 3% of isolates in the 10-year US study (LiPuma 2010). A recent publication described a successful lung re-transplant in a patient with CF who developed cepacia syndrome from a *B. ambifaria* infection five years post-initial-transplant; the first example of this organism causing the clinical manifestation (Goodlet *et al.* 2019). *B. cepacia* complex infections also occur in non-CF patients and are isolated from a variety of clinical settings including cardiac units, intensive therapy units, and neonatal intensive care units (Kenna *et al.* 2017).

Despite the classification of *Paraburkholderia* and subsequent sub-genera *Trinickia* and *Mycetohabitans* as non-pathogenic environmental sister clades of *Burkholderia*, there are several exceptions to this distinction. *Trinickia caryophylli* (formerly *Paraburkholderia caryophylli*) represents well characterised plant pathogen known to cause wilt of carnation (Kusumi *et al.* 1987). *Paraburkholderia xenovorans* has been previously isolated from a human blood culture, although not necessarily in association with disease (Goris *et al.* 2004). A case study of a 4 month-old boy who developed septicaemia with *Paraburkholderia tropica* following surgery was documented in Malaysia, however, the child was heavily immunocompromised and had previously experienced several blood infections (Deris *et al.* 2010). *Paraburkholderia fungorum* has been isolated from several human and veterinary sources, including human cerebrospinal fluid, vaginal secretions, sputum of CF patients, murine noses, porcine brain, and deer brain stem (Coenye *et al.* 2001b; Coenye *et al.* 2002; Vandamme and Peeters 2014). In addition, a case study of septic arthritis in a 9 year-old girl was caused by community-acquired *P. fungorum* (Gerrits *et al.* 2005). Overall, these *Paraburkholderia* infection cases illustrate that host vulnerability is also a key factor in the ability of *Burkholderia* and *Paraburkholderia* bacteria to cause infections.

1.6. Developments in genomics and sequence-based technologies

Following the sequencing of the first bacterial genome in 1995 there has been a rapid progression in sequencing technology throughput and reduction in cost per base pair. The completion of the low-GC 1.8 Mbp *Haemophilus influenza* genome represented the beginning of whole genome microbial genomics, and was permitted by chain termination Sanger sequencing technology (Fleischmann *et al.* 1995). Several alternative approaches were subsequently developed with higher throughput that represented the second generation of sequencing technologies, including Illumina, 454, Ion Torrent, and SOLiD (Morganti *et al.* 2019). Illumina ultimately emerged as the dominant sequencing technology championing multiplexed paired-end short read sequencing generating between 7.5 Gb to 6000 Gb depending on the Illumina instrument (Morganti *et al.* 2019). Single molecule, real-time (SMRT) sequencing is considered the third generation of sequencing technologies (Schadt *et al.* 2010), and generates long and ultra-long reads in comparison to Illumina. Two of the major third generation platforms, Pacific Biosciences and Oxford Nanopore Technology vary not only in their underpinning technology, but also cost, base-pair calling accuracy, and a centralised vs de-centralised approach to sequencing.

The worldwide uptake in sequencing technologies by the microbiology research community has resulted in a rapid increase in available genomic data, with the European Nucleotide Archive (ENA) possessing 317,013 bacterial genomic assemblies as of 30th June 2019. To conceptualise the scale of genomic data deposition in the public domain, the number of bacterial genomes archived at ENA in the first-half of 2019 alone was greater than the number of genomes deposited over the 7-year period between 2010 and 2017 (Figure 1). This wealth of genomic data has supported a multitude of projects such as epidemiology, genome mining, population biology and taxonomy, antimicrobial and virulence profiling, and microbial evolution. Whole-genome sequencing has been used to track the transmission of multi-drug resistant pathogens in hospital outbreaks, and ultimately identify the source in infections (Quick *et al.* 2014). Long-read sequencing technologies have also permitted the real-time epidemiological monitoring of the largest known Ebola virus outbreak (Quick *et al.* 2016). A plethora of genome mining strategies have been developed to exploit the considerable biosynthetic diversity that exists in bacteria (Ziemert *et al.* 2016). Genome mining software, such as AntiSMASH (Blin *et al.* 2017) coupled to cluster networking methodologies has enabled researchers to condense the abundance of biosynthetic gene cluster data into a series of targets for metabolite identification (Ziemert *et al.* 2016).

Genomics and related fields of metagenomics and transcriptomics have influenced all of the aforementioned concepts and experiments in relation to *Burkholderia* biology. The progression and refinement of *Burkholderia* taxonomy, genome mining for specialised secondary metabolite biosynthetic gene clusters, understanding the intricacies of microbe-plant interactions, exploiting biotechnological properties, and exposing the underlying epidemiology of *Burkholderia* in CF was made possible through exploiting the genomic data available and the ensuing bioinformatics tools.

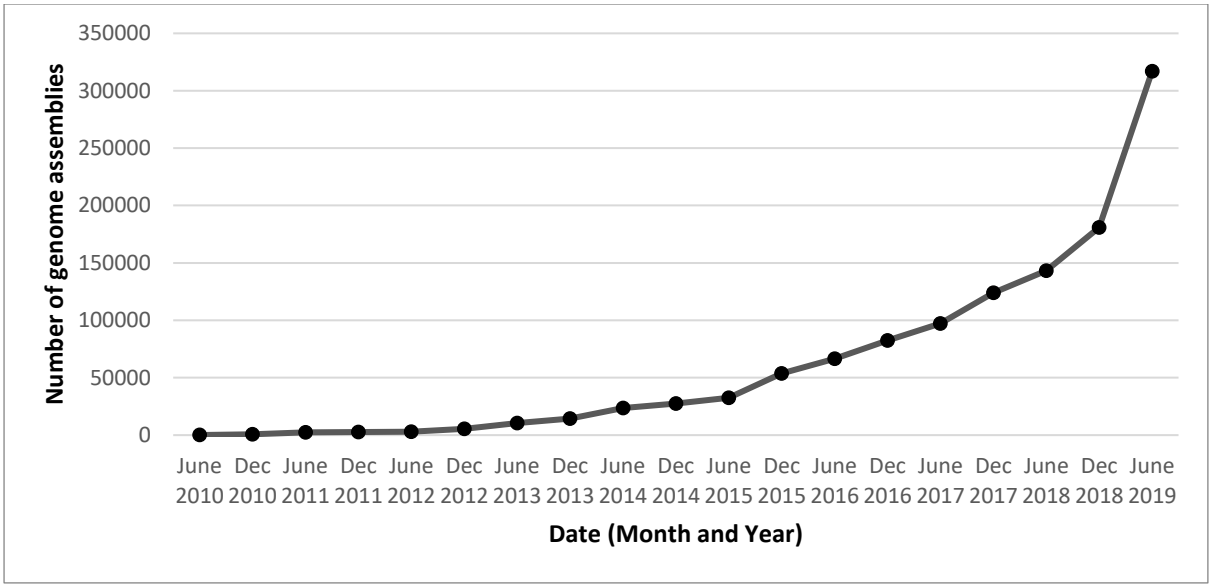


Figure 1. Bacterial genome assemblies available in the European Nucleotide Archive.

Only full genome representatives were included in the plot. Genome assembly levels include contig, scaffold, chromosome and complete genomes.

1.7. An introduction to *Burkholderia ambifaria*

B. ambifaria is a member of the *B. cepacia* complex, and while several other members of the complex are implicated in the majority of *Burkholderia*-associated CF infections, *B. ambifaria* is rarely encountered in these individuals (LiPuma 2010). The Type strain AMMD was isolated from the rhizosphere of a healthy pea plant (*Pisum sativum*) in Wisconsin, USA in 1985 (Coenye *et al.* 2001a), and possessed strong anti-oomycete activity (Parke *et al.* 1991; King and Parke 1993; Heungens and Parke 2000). *B. ambifaria* is unique among *Burkholderia* in its ability to synthesise multiple antimicrobial compounds, its low occurrence in CF and non-CF infections, and the species historical precedence as a successful biological control agent in the US. Compared to other *B. cepacia* complex species, *B. ambifaria* possesses significant potential as a biological control agent, with multiple publications attesting to its antimicrobial properties and biocontrol abilities (Parke and Gurian-Sherman 2001), and was the focus of this research as follows:

1.8. Project aims and objectives

The Biotechnology and Biological Research Council (BBSRC) funded this PhD through the South West Biosciences doctoral training partnership (SWBio). The aim of this research was to understand the genomics and specialised metabolite of the biopesticidal bacterium *B. ambifaria*. This overall aim was achieved by completing the following objectives:

1) **Comparative genomics (Chapter 3).** *B. ambifaria* strains were whole-genome sequenced by Illumina and Pacific Biosciences platforms. Comparative genomic analyses were applied to the assembled whole genomes of *B. ambifaria* to understand the genetic diversity of the species. The core, pan, and accessory genome was determined for the complete genome, in addition to specific replicons. The virulence and antimicrobial resistance profiles were characterised, and the presence of plasmids confirmed and investigated through Oxford Nanopore Technology.

2) **Specialised metabolite biosynthetic gene clusters (Chapter 4).** Biosynthetic gene clusters (BGCs) were predicted for all 64 *B. ambifaria* genomes, and de-replicated to determine the true biosynthetic diversity of the species. The distribution of core, accessory and antimicrobial BGCs was resolved, and the BGC local regulatory systems, specifically LuxRI-based systems, investigated.

3) ***In vitro* antimicrobial activity (Chapter 4).** The antimicrobial properties of *B. ambifaria* strains were assessed against a collection of plant and animal pathogens representing Gram-positive bacteria, Gram-negative bacteria, and fungi.

4) **Polyyne metabolite and biosynthetic gene clusters (Chapter 5).** The antimicrobial contribution of the polyyne cepacin to *B. ambifaria* was assessed, in addition to the role of temperature on the production of the polyyne. The distribution and diversity of polyyne BGCs was analysed by exploiting local and publically available genomes, and the gene architecture compared between different polyyne BGCs.

5) **Biological control and *In vivo* virulence (Chapter 6).** A biological control model was developed to assess the importance of cepacin in the biological control of the damping-off pathogen *P. ultimum* with the crop model *P. sativum*. The loss of the third replicon was assessed for biological control efficacy and rhizocompetence. *In vivo* virulence of wild-type and third-replicon deficient *B. ambifaria* was examined through the invertebrate *G. mellonella* wax-moth larvae model and murine respiratory inhalation model.

2. Materials and methods

2.1. Sources of *B. ambifaria* genome assemblies

Most *B. ambifaria* genomes were sequenced on the Illumina platform HiSeq 2000 and HiSeq X Ten instruments by the Wellcome Sanger Institute in collaboration with Prof. Julian Parkhill. The HiSeq 2000 and HiSeq X Ten instruments generated 125-nucleotide and 150-nucleotide paired-end reads, respectively. The Illumina platform was used to sequence 59 strains that were later *de novo* assembled into contigs (Table 1 and Table 2). The complete genome of *B. ambifaria* strain, BCF (BCC0203) was generated using Pacific Biosciences single molecule real time sequencing technology with the PacBio RS instrument at the Wellcome Sanger Institute. The remaining four strains: MC40-6 (BCC1212), AMMD (BCC0207), IOP40-10 and MEX-5 were available via public databases and sequenced by Illumina or 454 sequencing.

2.2. Background to bioinformatics analyses

All bioinformatics analyses were performed using the MRC-funded Cloud Infrastructure for Microbial Bioinformatics (CLIMB) (Connor *et al.* 2016). Bash scripts were parallelised using the shell tool GNU parallel (Tange 2018). Bioinformatics analyses that required scripts downloaded from GitHub are indicated as necessary.

Table 1. *B. ambifaria* strains and genomes used during this study.

Study Strain Name	Alternative Name	Source details (CF = cystic fibrosis; ENV = environmental)	Accession No. or Bioproject	Reference
BCC0118	CEP0617; LMG P-24636	CF (Sputum); USA	ERS784989	/
BCC0191	HI 2345 (J82); ATCC 51993; ARS BcB	ENV (Soil); USA	ERS784799	(Mao *et al.* 1997)
BCC0192	Ral-3; R-8863; HI2347; FC627	ENV (Corn rhizosphere, biocontrol strain); USA	ERS785047	(Coenye *et al.* 2001a)
BCC0197	ATCC 51671; LMG 19465; FC661; R-9945; B37w	ENV (Leaves of *Sesbania exaltata*, biocontrol strain)	ERS785076	(Schisler *et al.* 1991)
BCC0200	Formally *B.cepacia* gv I (BC-B)	ENV; USA	ERS785045	(Mao *et al.* 1997)
BCC0203	BCF/HG1-A; LMG-P 24640	ENV; USA	ERS782625	(Mao *et al.* 1998)
BCC0207	AMMD (LMG 19182^T)	ENV (Pea rhizosphere); USA	PRJNA13490	(Coenye *et al.* 2001a)
BCC0250	CEP0958;LMG P-24637; R-9927	CF (Sputum); Australia	ERS784819	/
BCC0267	LMG 19467; CEP0996; R-9935	CF (Sputum); Australia	ERS784835	(Coenye *et al.* 2001a)
BCC0284	ATCC 53267; LMG 17829; CEP0102; C2965	ENV (Corn roots); USA	ERS1328916	(Coenye *et al.* 2001a)
BCC0316	M54, HI 2347, R-5142	ENV (Soil); USA	ERS784850	(Coenye *et al.* 2001a)
BCC0338	ATCC 53266 LMG 17828; FC662	ENV (Corn roots); USA	ERS1371637	(Coenye *et al.* 2001a)
BCC0399	CEP1054	CF; USA	ERS784860	/
BCC0410	MVP/C1 64	ENV (Maize); Italy	ERS784866	(Dalmastri *et al.* 1999)
BCC0423	MCI 4	ENV (Maize); Italy	ERS1336067	(Bevivino *et al.* 1998)
BCC0477	AU0216	CF (Sputum); USA	ERS784882	/
BCC0478	AU1366	CF (Sputum); USA	ERS784897	/
BCC0480	HI-2427	ENV (Soil); USA	ERS784913	/
BCC1041	MVP-C2-51	ENV (Maize); Italy	ERS784930	(Dalmastri *et al.* 1999)
BCC1048	MVP-C2-69	ENV (Maize); Italy	ERS784886	(Dalmastri *et al.* 1999)
BCC1052	MCII-68	ENV (Maize); Italy	ERS1371632	(Di Cello *et al.* 1997)
BCC1062	MDII-130riz	ENV (Maize); Italy	ERS1328829	(Pirone *et al.* 2005)
BCC1065	MDIII-B-388	ENV (Maize); Italy	ERS784959	(Pirone *et al.* 2005)
BCC1066	MDIII-B-399	ENV (Maize); Italy	ERS784800	(Pirone *et al.* 2005)
BCC1072	MDIII-P-170	ENV (Maize); Italy	ERS1328913	(Pirone *et al.* 2005)
BCC1080	MDIII-T-2	ENV (Maize); Italy	ERS1371635	(Pirone *et al.* 2005)
BCC1083	MDIII-T-50	ENV (Maize); Italy	ERS1328835	(Pirone *et al.* 2005)
BCC1086	MDIII-T-401(s)	ENV (Maize); Italy	ERS1328827	(Pirone *et al.* 2005)
BCC1088	MDIII-T-474(s)	ENV (Maize); Italy	ERS1371633	(Pirone *et al.* 2005)
BCC1090	MVP-C1-40	ENV (Maize); Italy	ERS1328833	(Pirone *et al.* 2005)
BCC1092	MVP-C1-53	ENV (Maize); Italy	ERS784808	(Dalmastri *et al.* 1999)
BCC1093	MVP-C1-55	ENV (Maize); Italy	ERS784821	(Dalmastri *et al.* 1999)
BCC1095	MVP-C1-80	ENV (Maize); Italy	ERS784837	(Dalmastri *et al.* 1999)
BCC1098	MVP-C1-95	ENV (Maize); Italy	ERS784852	(Dalmastri *et al.* 1999)
BCC1100	MVP-C2-25	ENV (Maize); Italy	ERS784868	(Dalmastri *et al.* 1999)
BCC1103	MVP-C2-44	ENV (Maize); Italy	ERS784884	(Dalmastri *et al.* 1999)
BCC1105	MVP-C2-73	ENV (Maize); Italy	ERS1371636	(Dalmastri *et al.* 1999)
BCC1107	MVP-C2-79	ENV (Maize); Italy	ERS784899	(Dalmastri *et al.* 1999)
BCC1212	MC40-6	ENV (Rhizosphere); USA	PRJNA17411	/
BCC1213	MC80-27	ENV; USA	ERS1328957	/
BCC1214	MA80-5	ENV; USA	ERS784915	/
BCC1216	MW20-13	ENV; USA	ERS1328918	/
BCC1218	MW80-16	ENV; USA	ERS784932	/

BCC1220	MS5-3	ENV; USA	ERS1328837	/
BCC1223	MS80-4	ENV; USA	ERS784947	/
BCC1224	KS0-1	ENV (Maize); USA	ERS1328836	(Ramette and Tiedje 2007)
BCC1228	KA20-1	ENV (Maize); USA	ERS1328832	(Ramette and Tiedje 2007)
BCC1229	KA5-1	ENV (Maize); USA	ERS1371639	(Ramette and Tiedje 2007)
BCC1233	KC0-24	ENV (Maize); USA	ERS1328917	(Ramette and Tiedje 2007)
BCC1236	KC5-54	ENV (Maize); USA	ERS1328839	(Ramette and Tiedje 2007)
BCC1237	KC10-16	ENV (Maize); USA	ERS1371640	(Ramette and Tiedje 2007)
BCC1240	KC311-11	ENV (Maize); USA	ERS1328914	(Ramette and Tiedje 2007)
BCC1241	KC311-6	ENV (Maize); USA	ERS784961	(Ramette and Tiedje 2007)
BCC1246	KC20-40	ENV (Maize); USA	ERS1328834	(Ramette and Tiedje 2007)
BCC1248	KW0-1; LMG-P 24641	ENV (Maize); USA	ERS784801	(Ramette and Tiedje 2007)
BCC1249	KW0-5	ENV (Maize); USA	ERS1371634	(Ramette and Tiedje 2007)
BCC1252	KW10-1	ENV (Maize); USA	ERS784809	(Ramette and Tiedje 2007)
BCC1256	KW420-19	ENV (Maize); USA	ERS784823	(Ramette and Tiedje 2007)
BCC1258	KW318-1	ENV (Maize); USA	ERS1371641	(Ramette and Tiedje 2007)
BCC1259	KW20-2	ENV (Maize); USA	ERS784838	(Ramette and Tiedje 2007)
BCC1265	MC40-7	ENV; USA	ERS784854	/
BCC1270	KC20-17	ENV (Maize); USA	ERS784870	(Ramette and Tiedje 2007)
IOP40-10	/	ENV (Prairie grass rhizosphere)	PRJNA20669	/
MEX-5	/	ENV (Teosinte plants, *Zea perennis*)	PRJNA20667	/
BCC0191 ::*ccnJ*	/	/	/	This study
BCC1252 ::*ccnJ*	/	/	/	This study
BCC1241 ::*ccnJ*	/	/	/	This study
BCC0477 ::*ccnJ*	/	/	/	This study
BCC1259 ::*ccnJ*	/	/	/	This study
BCC1218 ::*ccnJ*	/	/	/	This study
BCC0191Δc3	/	/	/	This study
BCC0191 ::*ccnJ*Δc3	/	/	/	This study

2.3. Genome assembly

Two approaches were taken to assemble *B. ambifaria* genomes: short-read *de novo* based assembly, and hybrid assembly employing both short-reads and longer Oxford Nanopore Technology (ONT)-sourced reads as scaffolds. The former provided the foundation for the majority of analyses described in this thesis, while the latter contributed to our understanding of previously uncharacterised plasmids and broader genomic context of *B. ambifaria*.

2.3.1. Short read-(Illumina)-based assembly

Illumina adaptors were trimmed from 125-nucleotide or 150-nucleotide paired-end reads using Cutadapt v1.12 (Martin 2011), and the read quality assessed with FastQC v0.10.1 (Andrews 2009); both scripts were executed via the wrapper script Trim Galore v0.4.2 (Krueger 2016). Overlapping paired-end reads were identified and merged to form extended fragments with FLaSH v1.2.11 (Magoč and Salzberg 2011); this was performed to improve the contiguity of the assembled contigs. The remaining paired-end reads, along with the extended fragments, were input into the sequence assembler SPAdes v3.9.1 (Bankevich *et al.* 2012). Contigs were polished with Pilon v1.21 (Walker *et al.* 2014) to identify and amend any erroneously assembled contigs. The assembled contigs were screened for contamination by applying Kraken v0.10.5-beta (Wood and Salzberg 2014) with the Minikraken database. Kraken converts the query sequences into k-mers and compares the k-mers to a database to determine the likely taxonomic origin of the query sequence. Contigs that were classified outside the family Burkholderiaceae were discarded from the sequence data. Genome sequence quality was assessed using Quast v4.4 (Gurevich *et al.* 2013), which also provided N50 values, and further statistics attaining to the assemblies (Table 2).

Table 2. *B. ambifaria* genomic statistics.

Strain	Total Contigs (Mbp)	Mapped Contigs (Mbp)	Total Contig No.	Contigs >1000 bp	N50 (bp)
BCC0118	7.50	7.43	103	45	273879
BCC0191	7.58	7.55	97	57	302202
BCC0192	7.40	7.32	100	48	277504
BCC0197	7.38	7.36	82	33	386867
BCC0200	7.63	7.55	124	64	292321
BCC0203	7.93	7.53	4	4	2669373
BCC0207	7.53	7.48	4	4	2646969
BCC0250	7.36	7.33	88	32	396147
BCC0267	7.36	7.33	74	27	784052
BCC0284	7.47	7.38	88	34	423851
BCC0316	7.64	7.56	131	62	292321
BCC0338	7.45	7.42	61	28	514723
BCC0399	7.40	7.38	86	38	407605
BCC0410	7.38	7.36	44	18	849830
BCC0423	7.47	7.30	100	36	382759
BCC0477	7.81	7.43	118	54	255598
BCC0478	7.24	7.21	89	25	601161
BCC0480	7.84	7.52	87	24	570779
BCC1041	7.51	7.42	88	38	469567
BCC1048	7.51	7.43	76	35	469249
BCC1052	7.33	7.30	77	34	397259
BCC1062	7.45	7.36	90	37	382455
BCC1065	7.31	7.28	81	32	605534
BCC1066	7.32	7.29	112	37	381533
BCC1072	7.44	7.36	93	40	397405
BCC1080	7.29	7.26	85	35	383339
BCC1083	7.30	7.26	95	35	373328
BCC1086	7.60	7.40	302	49	395516
BCC1088	7.48	7.40	70	38	397405
BCC1090	7.33	7.29	114	33	340920
BCC1092	7.38	7.36	57	20	849686
BCC1093	7.40	7.38	91	37	427669
BCC1095	7.40	7.38	79	34	407577
BCC1098	7.38	7.37	48	15	849506
BCC1100	7.40	7.37	91	37	407581
BCC1103	7.40	7.37	77	38	427387
BCC1105	6.30	6.13	50	15	1592784
BCC1107	7.40	7.37	83	39	374059
BCC1213	7.47	7.42	107	40	425516
BCC1214	7.61	7.14	169	79	270842
BCC1216	7.46	7.33	166	73	234327
BCC1218	7.46	7.42	109	51	348377
BCC1220	7.64	7.14	188	74	393412
BCC1223	7.61	7.14	157	80	267383
BCC1224	7.27	7.19	152	53	360040

BCC1228	7.38	7.35	82	33	558864
BCC1229	7.51	7.48	100	51	298531
BCC1233	7.96	7.42	126	75	390415
BCC1236	7.63	7.58	123	47	358767
BCC1237	7.34	7.31	91	32	420767
BCC1240	7.40	7.37	63	29	503900
BCC1241	7.68	7.44	138	61	288780
BCC1246	7.48	7.44	72	32	450462
BCC1248	8.03	7.50	273	132	241997
BCC1249	7.58	7.55	71	27	447443
BCC1252	7.42	7.38	107	60	254705
BCC1256	7.54	7.47	104	49	281063
BCC1258	7.60	7.52	110	57	272850
BCC1259	7.63	7.30	162	75	431727
BCC1265	7.62	7.29	136	50	335885
BCC1270	7.60	7.53	115	51	315858
IOP40-10	7.69	6.88	629	571	24355
MC40-6	7.64	7.34	4	4	2769414
MEX-5	7.86	6.58	706	634	20649

2.3.1.1. Replicon scaffolding and re-circularisation

To provide genomic context to the genome assemblies, the contigs were re-arranged into the replicon structure of *B. ambifaria* with the Python script CONTIGuator v2.7.4 (Galardini *et al.* 2011). Three reference *B. ambifaria* genomes were used during contig re-arranging and scaffolding: BCF (BCC0203), MC40-6 (BCC1212) and AMMD (BCC0207). The BCF genome was generated from PacBio sequencing, MC40-6 and AMMD were sourced from the European Nucleotide Archive. All genome assemblies were scaffolded with each reference, and the reference providing greatest synteny and number of scaffolded contigs was chosen for the replicon structure. The command line option to fill gaps in scaffolded replicons with strings of "N" bases was disabled. Manual curation of the scaffolded contigs was performed to remove errors in the reconstituted replicons. Replicons c1, c2 and c3 were re-circulated with genes *dnaA*, *parA* and *parB*, respectively, using the software Circlator v1.2.1 (Li 2013). Scaffolded replicons and non-scaffolded contigs were annotated using Prokka v1.12-beta (Seemann 2014) with the 'Bacteria' annotation mode (default).

2.3.2. Hybrid-(Illumina and MinION)-based genome assembly

The hybrid assembly combined the Illumina reads with long-reads generated by ONT MinION (MIN-101B), using the MinION fluidics module R9 version flow cells (FLO-MINSP6). Genomic DNA was extracted twice per sample using the Maxwell Instrument 16 as described previously (section 2.12.1), and the DNA pooled per sample. Fragment size and concentration were assessed using the Tapestation and Qubit 3 fluorometer, respectively. Approximately 15 µg of genomic DNA per sample was sheared to ~20 kbp fragments using the g-TUBE Covaris instrument (5,800 rpm, 60 seconds), and followed by a quality check of fragment size using the Tapestation. Size exclusion was performed with 0.4x volume of SPRI beads to reduce the concentration of fragments <1500 bp; the resulting DNA was eluted in 20 µl of molecular grade water. Further size selection was achieved through 0.4x volume of AMPure beads to exclude fragments <1000 bp, and eluted into 20 µl of molecular grade water. Both SPRI beads and AMPure beads were used as per manufacturer's instructions (Beckman Coulter, Indianapolis, USA). DNA concentrations of the size-selected samples were determined with Qubit fluorimetry, and 400 ng per sample was diluted to a final volume of 7.5 µl in molecular grade water as required by the Nanopore rapid barcoding kit (SQK-RBK004) protocol. The library prep was performed as per the manufacturer's instructions (Oxford Nanopore Technologies, Oxford, UK). Barcoded samples were pooled at a ratio based on the Illumina-assembled genome sizes (e.g. 7.6 Mbp genome = 7.6 µl) prior to the final AMPure clean up and elution into 15 µl molecular grade water. A final Qubit fluorimetry of DNA concentration confirmed the total DNA quantity loaded onto the MinION flow cell in 10 µl of pooled barcoded sample.

Approximately 650 ng of pooled library prep representing twelve genomes was loaded into the MinION flow cell. The flow cell started with 375 active pores, and the run was allowed to continue until 100% of bases were called.

MinION reads exported from the Nanopore MinKNOW basecalling software, and reads that passed the average base score quality threshold were trimmed and de-multiplexed using Porechop available on GitHub (https://github.com/rrwick/Porechop). The default settings of 70% barcode sequence identity and a minimum of 5% difference between best and second-best matching barcode to MinION read barcode. Reads with middle adaptors were discarded rather than split. De-multiplexed raw reads were corrected to increase sequence accuracy with Canu v1.8. Statistics of the ONT MinION sequencing run were calculated and graphically displayed using the script NanoPlot, available at GitHub (https://github.com/wdecoster/NanoPlot). Hybrid genome assemblies were constructed using Unicycler v0.4.7 to assemble the trimmed Illumina reads (section 2.3.1) and scaffold the resulting contigs with corrected MinION reads with default settings. The resulting contigs were visualised using Bandage v0.8.1 to assess the level of completeness in unresolved replicons.

In an attempt to resolve incomplete replicons with single path bridges linking the nodes (contigs), corrected MinION reads were mapped to the contigs constituting the replicon. A short script written by Ryan Wick that implements the pairwise aligner minimap2 is available on GitHub (https://github.com/rrwick/Unicycler/wiki/Read-extraction). This script was modified to output a list of MinION reads that possessed sequence similarity to the contigs of interest. A BLAST database was created of the hybrid genome assembly in Bandage, and the extracted MinION reads with homology were searched against the database. Evidence of one of more reads spanning the single path bridges were used to support the merger and completion of the replicon.

2.4. Average nucleotide identity

To confirm the species designation of the sequenced isolates, all 67 available *B. ambifaria* genomes: 61 assembled Illumina genomes, one assembled PacBio genome (BCC0203), and five publically available genomes (AMMD, MC40-6, IOP40-6, MEX-5 and RZ2MS16), were subject to average nucleotide identity analysis (ANI). The Python script PyANI v0.2.7 (Pritchard *et al.* 2016) was used to calculate alignment-based pairwise ANI values for the 67 genomes, using MUMmer 3 (Kurtz *et al.* 2004) as the preferred method to generate pairwise sequence alignments. A threshold ANI of 95% was applied to delineate *B. ambifaria* genomes as suggested by Richter and Rossello-Mora (2009). All nucleotide similarity values were reported to three significant figures.

2.5. Gene prediction, sequence annotation and pan-genomics

Reconstituted replicons and non-scaffolded contigs were annotated using Prokka v1.12-beta (Seemann 2014) with the 'Bacteria' annotation mode (default). Prokka is a prokaryotic genome annotation pipeline that combines multiple scripts to generate an annotation file. Optional and recommended scripts included in the Prokka annotation pipeline included the RNA-encoding gene detection tools ARAGORN (Laslett and Canback 2004) and Barrnap (https://github.com/tseemann/barrnap). The core, accessory and pan-genomes of all *B. ambifaria* genomes were predicted using Roary v3.7.0 (Page *et al.* 2015), and applied to the non-scaffolded contigs to capture the complete genetic diversity of each strain. A minimum percentage identity of 95% for blastp comparisons, and a 99% core gene threshold (all strains) was applied to the *B. ambifaria* dataset. Antimicrobial resistance genes were predicted in the draft *B. ambifaria* genomes using publically available AMR databases: CARD, ARG-ANNOT, NCBI, and ResFinder; while virulence-conferring genes were predicted using the publically available virulence factor database (vfdb). All five databases were employed through the BLAST-based tool Abricate, available via GitHub (https://github.com/tseemann/abricate). The vfdb database was manually curated with *B. ambifaria* homologues in instances of fragmented operon detection to increase the accuracy of detection in the *B. ambifaria* genomes.

2.6. Sequenced-based phylogenetic trees

Sequence alignments for generating phylogenies were produced using MAFFT v7.305b (Katoh and Standley 2013). Core-gene alignments were generated using Roary (Page *et al.* 2015) which incorporates MAFFT as the alignment algorithm. Two scripts were used to construct phylogenies: RAxML v8.2.11 (Stamatakis 2014) and FastTree v2.1.9 (Price *et al.* 2010). FastTree was used to produce phylogenies, when speed was priority, by using an approximate maximum-likelihood model with double-precision and multi-thread compilation.

RAxML was employed to generate high-accuracy phylogenies using a maximum-likelihood with General Time Reversible (GTR) substitution and a GAMMA model of rate heterogeneity supported by 100 bootstraps. The RAxML executable was compiled with multi-thread and SSE3 functionality: raxmlHPC-PTHREADS-SSE3. Phylogenies constructed with FastTree and RAxML were visualised with phylogeny graphic viewer FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/) and edited using the graphic editing software Inkscape v0.91 (http://www.inkscape.org/). Ribosomal MLST (rMLST) profiles were downloaded from the Bacterial Isolate Genome Sequence Database (BIGSdb).

2.7. Plasmid identification and characterisation

To determine the prevalence of plasmids in the *B. ambifaria* strains collection the following pipeline was employed to assemble and detect contigs that were likely to originate from plasmids. To determine which contigs in a genome assembly were plasmid derived the contig assembly process was repeated using the plasmidSPAdes (Antipov *et al.* 2016) mode of the assembler SPAdes v3.8.0 (Bankevich *et al.* 2012). PlasmidSPAdes detects contigs of plasmid origin by read coverage of contigs compared to the median of long contigs (greater than 10,000 bp), and removal of chromosomal contigs based on the plasmidSPAdes algorithm during transformation of the assembly graph into the plasmid graph.

Following plasmidSPAdes contig assembly and identification, the contigs were annotated using Prokka, and the predicted protein and gene sequences extracted. Hidden Markov models were built for protein domains indicative of ParA and ParB proteins using the interproscan domains IPR002586 - CobQ/CobB/MinD/ParA nucleotide binding domain, and IPR004437 - ParB/RepB/Spo0J partition protein family, respectively. A random 10% subset of the available domain examples for both IPR002586 (2,949 sequences) and IPR004437 (3,308 sequences) were extracted for the HMMs to minimise time-consuming sequence alignments. The protein domains were aligned with MAFFT, and the HMMs built using Hmmer (Finn *et al.* 2011). The predicted protein sequences extracted from the plasmid contigs were scanned using the HMMs, with a 1e-5 e-value, to identify potential ParA and ParB homologues. Candidate homologues were pooled and examined for an adjacent *parA* or *parB* gene; candidate protein sequences without an encoded gene partner were excluded from further analyses. Separate ParA and ParB protein alignments were generated using MAFFT (Katoh and Standley 2013), and a phylogeny constructed with FastTreeMP (Price *et al.* 2010). The ParA and ParB protein sequences of replicons c1, c2 and c3 were extracted from a subset of

B. ambifaria genomes and included in the phylogenies to distinguish chromosomal and plasmid derived ParAB proteins.

To authenticate the plasmid-derived origin of the MEX-5 and IOP40-10 candidate proteins, the corresponding nucleotide sequences were searched against the nucleotide collection database on NCBI. Most returned hits were either dominated by chromosomal or plasmid derived homologues, and thus provided an unambiguous method of identifying and removing non-plasmid derived homologues. Complete plasmid sequences with gene homology to either *parA* or *parB* genes in MEX-5 or IOP40-10 were downloaded, and their ParA and ParB homologues extracted to determine if the plasmid derived ParA and ParB proteins in MEX-5 and IOP40-10 possessed amino acid sequence similarity with paired ParAB proteins or erroneously identified proteins (those lacking paired proteins). Protein sequences that lacked any *Burkholderia* homologues were discounted from further analyses. Plasmids with *parA* or *parB* homology in ambiguous blastn results with mixed plasmid and chromosomal homologues were downloaded to compare protein sequences with MEX-5 and IOP40-10 homologues via phylogeny analyses.

2.8. Identifying potential plasmids through pairwise k-mer matching

MinHash sketches were produced of the 4,479 locally assembled *Burkholderia* genomes and *B. ambifaria* BCC1248 plasmid p1 using Mash (Ondov *et al.* 2016) with default parameters. MinHash sketches are composed of k-mer sets generated by a sliding window of k-length along the sequence of interest. Genomic hits with an e-value < 1 were selected for further analysis. Non-*B. pseudomallei c*ontigs greater than 5 kbp were extracted and screened for the presence of *parB* genes, while *B. pseudomallei* contigs greater than 50 kb were extracted due to the large number of genomes. A pairwise comparison of contig mash sketches against the BCC1248 p1 plasmid identified which contigs shared k-mers with the plasmid of interest. Contig hits with a Jaccard index greater than 1/1000 were subject to *parB* screening as previously described in Section 2.7.

2.9. Mapping core and accessory genes

The core and accessory genome distribution across the three replicons was visualised for the *B. ambifaria* type strain AMMD (BCC0207). The scaffolded replicons c1, c2 and c3 were processed by Roary (Page *et al.* 2015) to generate replicon specific core, accessory and pan-genomic statistics. The Roary-derived core-gene codes for each replicon were used to extract the nucleotide sequences for the replicon-specific core genes, from the pan-genome output file. A similar approach was used to extract the accessory genes, following manipulation of the pan-genome file to display only gene codes for strains that shared an accessory gene with the strain of interest, AMMD. This was necessary as Roary outputs a single example of each gene to the pan-genome file, not necessarily the gene sequence from the strain of interest. Therefore, all strain gene codes that correspond to the accessory genes encoded by the strain of interest must be extracted and searched against the pan-genome output file. The lack of gene sequences from the strain of interest prevents direct mapping of the genes to the genome, necessitating the need for a blast search for the homologous genes in strain AMMD. Following identification of the homologous genes, two annotation files must be generated to indicate the location of each gene on the given replicon, and the core or accessory nature of the genes. To accomplish this a shell script, blast_to_gff_wrapper.sh, written by Alvar Almstedt was used to identify AMMD gene homologues and generate a general feature format (gff) annotation file specific to the replicon blast database. The shell and python scripts are available on GitHub: https://github.com/alvaralmstedt. The following edits were made to the shell script: One high scoring pair (HSP) was reported for each query sequence during the blastn search. The blast-short task, optimised for shorter query sequences, was implemented when a query gene was too short to detect the AMMD homologue. The general feature format files were visualised on the replicon using the circular vector visualising software Gview v1.7.

2.10. Mapping core and accessory BGCs

The identification of distinct BGCs, and the core and accessory nature of BGCs in *B. ambifaria* was defined in section 2.9. The nucleotide sequence for each BGC was used as a query against their respective replicon, and the coordinates formatted as a general feature format file using the blast_to_gff_wrapper shell script. The BGCs were visualised on the replicon using the circular vector visualising software Gview v1.7.

2.11. Culture-based enumeration of viable bacterial cells

Viable bacterial cells were enumerated using the Miles and Misra drop method (Miles *et al.* 1938). A 100 µl aliquot of the bacterial cell suspension requiring viable cell enumeration was serially diluted in 900 µl aliquots of 1x PBS, up to 10^{-7} dilution. Each dilution was vortexed to ensure a homogenous suspension. Three 10 µl drops of each dilution were pipetted onto TSA agar (supplemented with antibiotic if selection was required), and the drops allowed to dry on the agar surface. Once dry, the agar plates were incubated overnight at 30°C. Following incubation, colonies of the appropriate dilution were counted, and the viable count expressed as colony forming units per ml (cfu/ml), per seed (cfu/seed) or root section (cfu/root section).

2.12. DNA extraction from bacterial cultures

2.12.1. High quality DNA extraction using automated Maxwell 16 instrument

High quality genomic DNA was extracted from liquid bacterial cultures using the automated Maxwell® 16 instrument (Promega UK) with the Tissue DNA Purification Kit. *B. ambifaria* strains were grown overnight at 37°C in 3 ml TSB broth (supplemented with antbiotic if selection was required). Bacterial cells were pelleted by centrifugation for 10 minutes at 4000 rpm, and the re-suspended in 300 µl molecular biology grade guanidinium thiocyanate. The cell suspension was transferred into a cartridge of the Tissue DNA Purification Kit (Promega UK), and placed in the Maxwell® 16 instrument for automated DNA extraction. The DNA was eluted into 300 µl of molecular grade water. DNA extracted using the Maxwell® 16 instrument was stored at 4°C for short-term storage or -20°C for long-term storage.

2.12.2. Rapid DNA extraction using Chelex 100 resin

Chelex 100 resin (Bio-Rad)-based DNA extraction was performed on solid-surface bacterial growth to screen bacterial colonies via PCR. A 5% (w/v) chelex 100 suspension was autoclaved to sterilise the resin. Freshly grown (less than 48 hours following plate inoculation) bacterial growth was removed from the agar surface using a 200 µl pipette tip, and re-suspended in 50 µl sterilised 5% (w/v) chelex 100 resin within a 200 µl PCR tube. The PCR tube was subsequently cycled twice between 95°C and -80°C, and held at each temperature for 5 minutes. Following the freeze-thaw cycling, the resin was allowed to settle, or pelleted by brief centrifugation (3 seconds of pulse centrifugation), and 2 µl of the clear supernatant used as the template for PCR. DNA extracted using chelex 100 resin was used the same day as extraction.

2.13. Identifying biosynthetic gene clusters in *B. ambifaria*

2.13.1. Genome mining of secondary metabolite BGCs

Bioinformatics tool antiSMASH v.3.0.5 (Weber *et al.* 2015) detected secondary metabolite biosynthesis gene clusters in both scaffolded replicons and non-scaffolded contig sequences. AntiSMASH was run locally using the CLIMB infrastructure. Additional analyses were included: clusterblast, compared predicted BGCs against a database of other predicted BGCs; and knownclusterblast, compared predicted BGCs against the MIBiG database of characterised BGCs. Characterised gene clusters that were not detected by antiSMASH were identified with nucleotide-nucleotide BLAST v2.6.0+ (Morgulis *et al.* 2008). A local nucleotide BLAST database of *B. ambifaria* genomes was generated using the makeblastdb command with parse-seqids option. The nucleotide BLAST search used a manually extracted BGC representative from a *B. ambifaria* genome as a query sequence, with the following options: evalue 1e-5, max_hsps 1 and max_target_seqs 1. BLAST defined co-ordinates were used to extract the BGCs from multiple strains with the script extractSequence.pl written by Lee Katz available at GitHub (https://github.com/lskatz). Comparisons of the secondary metabolic potential of each replicon were visualised using the web tool BoxPlotR (Spitzer *et al.* 2014).

2.13.2. De-replication of BGCs and cluster network analysis

BGCs predicted by antiSMASH and manually extracted were de-replicated by clustering nucleotide sequences using a pairwise K-mer-matching approach with Mash v1.1.1 (Ondov *et al.* 2016), reporting a maximum p-value and maximum distance of 1 and 0.05, respectively. The ability to adjust the maximum-distance allowed continuous range of distance thresholds to be applied, and optimisation of the final parameters, reported above. Each BGC sequence was fragmented into short 21 bp strings (K-mers). Each unique K-mer was assigned a number, and the list of K-mer numbers of one BGC was compared to those of other BGCs to determine the number and fraction of shared K-mers (Jaccard index) in a pairwise manner. An additional metric calculated was the Mash-distance (estimated mutation rate between sequences). The resulting distance matrix generated by Mash (Ondov *et al.* 2016) was visualized with Cytoscape v3.4.0 (Shannon *et al.* 2003), applying the Jaccard index, p-value and Mash-distance as edge attributes.

2.13.3. Identifying known BGCs in cluster network

The gene architectures of characterised BGCs were compared to representatives of the de-replicated BGC dataset to identify potentially known BGCs in the *B. ambifaria* genomes. Following this initial comparison, representative de-replicated BGCs that possessed gene synteny to published BGCs were annotated and compared to the published annotation of the characterised BGC. Similar gene architecture combined with congruent gene annotations provided sufficient evidence to identify the *B. ambifaria* BGCs.

2.14. *In vitro* antimicrobial metabolite detection

All *B. ambifaria* strains were grown at 30°C on BSM-G for 72 hours. Single plates were extracted by addition of 4 mL of acetonitrile for 2 hours, followed by centrifugation to remove debris. Crude extracts were directly analysed by UHPLC-ESI-Q-TOF-MS. UHPLC-ESI-Q-TOF-MS analyses were performed using a Dionex UltiMate 3000 UHPLC connected to a Zorbax Eclipse Plus C-18 column (100 × 2.1 mm, 1.8 µm) coupled to a Bruker MaXis II mass spectrometer. Mobile phases consisted of water (A) and acetonitrile (B), each supplemented with 0.1% formic acid. A gradient of 5% B to 100% B over 30 minutes was employed at a flow rate of 0.2 mL/min. The mass spectrometer was operated in positive ion mode with a scan range of 50-3000 m/z. Source conditions were: end plate offset at −500 V; capillary at −4500 V; nebulizer gas (N_2) at 1.6 bar; dry gas (N_2) at 8 L min^{-1}; dry temperature at 180 °C. Ion transfer conditions were: ion funnel radio frequency (RF) at 200 Vpp; multiple RF at 200 Vpp; quadrupole low mass at 55 m/z; collision energy at 5.0 eV; collision RF at 600 Vpp; ion cooler RF at 50–350 Vpp; transfer time at 121 µs; pre-pulse storage time at 1 µs. Calibration was performed with 1 mM sodium formate through a loop injection of 20 µl at the start of each run.

2.15. Rapid screening for antimicrobial metabolites

B. ambifaria strains were grown overnight in TSB and used to inoculate agar plates with 10 streaks across the surface using a swab. The agar plates were incubated for 72 hours at either 22°C or 30°C. Microbial growth was removed from the surface, and a 22 mm diameter circular plug (approximately 1 g) removed from the centre of the plate, and added to 500 µl of ethyl acetate. The agar plug was incubated in the solvent, with agitation, for 2 hours at 22°C. Extracts of 20 µl were analysed on a Waters® AutoPurification™ HPLC system fitted with a reverse-phase analytical column (Waters® XSelect CSH C18, 4.6 x 100 mm, 5 µm) and a C18 SecurityGuard™ cartridge (Phenomenex) in series. A photo-diode array detector (PDA) scanning between 210-400 nm was used to detect metabolites. Mobile phases consisted of (A) water with 0.1% formic acid and (B) acetonitrile with 0.1% formic acid with a flow rate of 1.5 ml min^{-1}. Elution conditions were as follows: 0 to 1 minute, 95% phase A & 5% phase B; 1

to 9 minutes, gradient of phase A from 95 to 5% and gradient of phase B from 5% to 95%; 10 to 11 minutes, 5% phase A & 95% phase B; 11 to 15 minutes, 95% phase A & 5% phase B.

2.16. Detecting regulatory systems in *B. ambifaria*

2.16.1. Regulatory protein LuxR prediction and de-replication

LuxR-encoding gene homologues were identified by replicating the systematic *in silico* approach previously described (Gan *et al.* 2014). A hidden Markov model (HMM) was built with 7014 protein sequences possessing a specific domain architecture encoding the LuxR autoinducer-binding domain and LuxR C-terminal using HMMER v3.1b2 (http://hmmer.org). Protein sequences were predicted from the *B. ambifaria* genomes using Prodigal v2.6.2 (Hyatt *et al.* 2010) and subsequently scanned with the HMM, applying a threshold e-value of 1e-5. Candidate proteins identified were extracted from the protein multi-fasta using the script subset_fasta.pl (author: John Nash, National Research Council of Canada). Candidate protein sequences were further annotated for protein domains using InterProScan v5.22-61.0 (Jones *et al.* 2014). Protein sequences encoding the four protein domains characteristic of LuxR proteins: winged helix-turn-helix DNA-binding domain (formerly IPR011991, currently IPR036388), autoinducer-binding domain (IPR005143), transcriptional regulator LuxR C-terminal (IPR000792) and signal transduction response regulator (IPR016032), were extracted. Extracted sequences were aligned with MAFFT v7.305b (Katoh and Standley 2013) and a phylogeny constructed with FastTree v2.1.7 (Price *et al.* 2010).

2.16.2. Characterising alternative regulatory genes of de-replicated BGCs

Gene regulation of BGCs that lacked *luxR* genes was analysed by two methods. The first method exploited the web-based tool Predicted Prokaryotic Regulatory Proteins (P2RP) (Barakat *et al.* 2013) which predicted protein sequences in submitted BGC nucleotide sequences, and searched for the presence of DNA-binding domains or two-component system domains associated with regulatory proteins. The second approach relied on Prokka (Seemann 2014) annotation of the BGC sequences, coupled to manual extraction of regulatory genes and InterProScan annotation to confirm the presence of regulatory protein domains. These approaches were conducted in parallel, and the manual-InterProScan approach was used to support or disprove the predictions made by P2RP.

2.17. *In vitro* antimicrobial screening against pathogens

B. ambifaria strains were screened for antimicrobial activity using the an antagonism assay modified from Mahenthiralingam *et al.* (2011). All *B. ambifaria* strains were grown in tryptic

soy broth (TSB) at 37°C and aerated overnight. Six strains were distributed across each 96-well plate (wells A2, A10, C6, E2, E10 and G6); using 200 µl of overnight culture and 17.4 µl DMSO (8% final concentration). A 96-pin replicator was used to transfer the cultures on to 120x120 mm Petri plates containing basal salts medium (Hareland *et al.* 1975) supplemented with 4 g/l glycerol (BSM-G) (Mahenthiralingam *et al.* 2011), and grown for three days at 30°C. Following incubation, *B. ambifaria* growth was killed by chloroform exposure for 2 minutes. *B. ambifaria* growth was overlaid with seeded half-strength iso-sensitest agar, cooled to 42°C, containing 200 µl of susceptibility organism liquid culture in 50 ml half-strength iso-sensitest agar. Susceptibility organisms were grown overnight in TSB at their optimum growth temperatures (Table 3). Overlaid plates were incubated overnight at the optimum growth temperature of the susceptibility organism in the seeded iso-sensitest. Zones of inhibition diameter were measured for each *B. ambifaria* strain against each susceptibility organism. The assay was repeated to obtain duplicate measurements. A heatmap of antagonistic activity based on zones of inhibition was generated in statistics software R v3.2.3 via RStudio v0.99.484.

A modified bacterial antagonism assay was used to screen mycelial fungal plant pathogens *Alternaria alternata* and *Fusarium solani*. A 5 mm-diameter agar core was taken from the leading edge of mycelial growth and transferred onto a PDA agar plate. Inoculated plates were allowed to grow for 9 days at 22°C before re-suspending mycelial growth in 1 ml PBS taken from the fungal surface with a swab. Half-strength iso-sensitest agar was seeded with the mycelial suspension at a ratio of 320 µl suspension to 100 ml agar, then used as an overlay as described previously. Fungal overlay plates were incubated for two days at 22°C, and the zone of inhibition measured.

Table 3. Antimicrobial susceptibility organisms: host and disease phenotypes (Howden *et al.* 2009; Mansfield *et al.* 2012; Tsuge *et al.* 2013; Coleman 2016) **and incubation temperatures.**

Organism	Source/ID Number	Incubation Temperature (°C)	Host Range	Disease
Pathogenic bacteria, fungi and oomycetes				
Rhizobium radiobacter	LMG 187	30	Broad	Crown gall disease
Pseudomonas savastanoi pv. *phaseolicola*	LMG 2245	30	Common bean	Bean halo blight
Dickeya solani	LMG 25993	30	European potato	Potato tuber rot
Pectobacterium carotovorum	LMG 2464	30	Several crop species	Soft rot disease
Pseudomonas syringae pv. *tomato*	LMG 5093	30	Tomato	Bacterial speck
Xanthomonas campestris pv. *campestris*	8004	30	Cultivated Brassicaceae	Black rot
Pseudomonas syringae pv. *syringae*	LMG 1247	30	Common bean	Bacterial brown spot
Burkholderia multivorans	ATCC 17616	30	Human (immunocompromised)	Opportunistic (e.g. CF lung infection)
Staphylococcus aureus	NCTC 12981	37	Human	Multiple (cutaneous, systemic etc.)
Enterococcus faecalis	ATCC 51299	37	Human	Multiple (uninary, systemic etc.)
Candida albicans	SC 5314	37	Human	Candidiasis
Fusarium solani var. redolens (Wollenweber)	MUCL 14241	22	Multiple crop species	Root/foot rot, wilt
Alternaria alternata (Fries:Fries) von Keissler	MUCL 36	22	Several (pathovar specific)	Black-spot, stem canker
Pythium ultimum Trow var. *ultimum*	MUCL 16164	22	Broad	Damping-off
Other reference organisms				
Bacillus subtilis	ATCC 23857	30	N/A	N/A

2.18. Minimum inhibitory concentration (MIC) assay

MICs were performed in a microbroth dilution assay using iso-sensitest broth or TSB broth, as previously described by Rushton *et al.* (2013), with doubling-dilutions of enacyloxin IIa between 100 and 0.098 µg/ml. Bacteria were grown for 18-24 hrs at 30°C and optical density measurements taken at 600 nm. The MIC was determined by calculating the enacyloxin IIa concentration required to cause an 80% knockdown in optical density compared to the control grown in the absence of the antibiotic. Each microbroth dilution was performed in triplicate.

2.19. Temperature-dependent regulation of *B. ambifaria* cepacin production

The impact of temperature on the production of cepacin in *B. ambifaria* was assessed using the traditional overlay assay outlined in section 2.17. Briefly, each cepacin-encoding *B. ambifaria* was grown overnight in TSB at 37°C, and 2 µl spotted onto BSM-G. The agar plates were incubated for three days at either 22°C or 30°C. The *B. ambifaria* growth was killed by chloroform exposure, overlaid with *S. aureus*-seeded half-strength iso-sensitest, and incubated overnight at 37°C. The zone of inhibition diameter for each strain was measured, and the experiment repeated in triplicate.

2.20. Insertional mutagenesis of uncharacterised *Burkholderia* BGC

Insertional mutagenesis was used to disrupt the expression of an uncharacterised LuxRI-associated *Burkholderia* BGC that was subsequently shown to be the biosynthetic origin of cepacin A. Primers were designed to amplify a 649 bp region of the fatty AMP ligase-encoding gene (Table 4), yielding a final product of 707 bp. The product was amplified following a standard PCR thermal cycling protocol (Table 5) using the NEB Q5 high-fidelity DNA polymerase (M0491) (Table 6). The PCR product was purified using the QIAquick PCR Purification kit according to the manufacturer's protocol. Following purification, the product was ligated into the suicide vector pGp-omega-Tp (Flannagan *et al.* 2007). DNA concentrations of the purified vector and insert were quantified using the Qubit 3.0 Fluorometer according to the manufacturer's protocol. Double restriction digests of up to 1 µg of pGp-omega-Tp and insert with XbaI and EcoRI were performed for 1 hour at 37°C as described in the NEB optimal restriction endonuclease reaction protocol. The restricted vector and insert were quantified as described before using the Qubit Fluorometer. Ligation of the insert into the vector was performed at a 3:1 ratio of insert to vector molar mass using the Bioline Quick-Stick Ligase according to the Bioline product protocol. Ligation of the purified restricted vector in the absence of the insert was performed as a control. The ligation reaction was directly transformed into chemically-competent *E. coli* SY327 cells via heat-shock. A tri-parental

conjugation of the *E. coli* SY327 pGp-omega-Tp-insert (donor), *E. coli* HB101 pRK2013 (helper) and *B. ambifaria* (recipient) was performed to construct *B. ambifaria* strains with the pGp-omega-Tp-insert integrated into the genome at specific sites via homologous recombination. Insertional mutants in each *B. ambifaria* strain background were confirmed by PCR (Table 4) using *Taq* polymerase (Table 7).

2.21. Constructing insertional mutants in *B. ambifaria*

2.21.1. Preparation of chemically-competent cells

Chemically competent cells were produced to introduce the pGp-omega-Tp vector possessing the cloned *ccnJ* gene fragment into *E. coli* SY327. A 3 ml overnight TSB culture of *E. coli* SY327 at 37°C was used to inoculate 50 ml fresh TSB. The inoculated TSB was incubated at 37°C with 150 rpm shaking until an OD_{600nm} equal to 0.6-0.8 was achieved, and then pelleted by centrifugation for 10 minutes at 4000 rpm. The pellet was carefully re-suspended in 25 ml chilled 50 mM calcium chloride, and kept on ice for 20 minutes. The suspension was pelleted by repeat centrifugation, and carefully re-suspended in 5 ml chilled 50 mM calcium chloride. The cell suspensions were divided into 300 µl aliquots, and subsequently stored at -80°C until required.

2.21.2. Miniprep of plasmid vectors

The pGp-omega-Tp vector was prepared by growing 50 ml TSB *E. coli* SY327 pGp-omega-Tp cultures supplemented with 50 µl/ml trimethoprim overnight at 37°C. The pGp-omega-Tp vector was purified using the QIAprep Miniprep kit, following the manufacturer's protocol, with the exception of the input volume. The 50 ml culture was divided into two equal aliquots and cell pelleted by centrifugation for 10 minutes at 4000 rpm. Each pellet was treated separately, until the final elution step where both samples were eluted through the same column.

2.21.3. Heat-shock transformation of chemically-competent cells

Chemically competent *E. coli* SY327 cells was thawed on ice, and kept on ice unless stated otherwise. The heat-shock transformation protocol was adapted from the NEB High Efficiency Transformation Protocol (C2987H/C2987I). A 2 µl aliquot of ligase reaction (either ligated vector-insert reaction or vector-only reaction) or unrestricted vector were added to 100 µl of thawed chemically-competent *E. coli* SY327 cells. The vector-only ligation reaction was transformed to control for successful double digestion of the vector. The unrestricted vector was transformed to confirm cell competency. An aliquot of chemically-competent *E. coli* SY327 cells without the addition of ligase reaction or restricted plasmid was included as a negative

growth control on selection media. The vials were kept on ice for 30 minutes, and subsequently heat-shocked for 30 seconds at 42°C before returning to ice for 10 minutes. Heat-shocked cells were recovered by thoroughly mixing 50 µl with 950 µl of room temperature SOC medium, and incubating for 60 minutes at 37°C with 150 rpm shaking. A serial dilution of each sample to 10^{-2} was created, and 100 µl of the neat suspension along with dilutions were plated onto both TSA and TSA supplemented with 50 µg/ml trimethoprim. Plates were incubated overnight at 37°C. Potential transformant colonies were screened for the presence of the insert with the original PCR primers used to amplify the fragment.

2.21.4. Tri-parental conjugation

Overnight 5 ml TSB cultures of *E. coli* HB101 pRK2013 (kanamycin selection, 25 µg/ml), *E. coli* SY327 pGp-omega-Tp-insert (trimethoprim selection, 50 µg/ml) and *B. ambifaria* were grown at 37°C. Overnight cultures were pelleted by centrifugation for 10 minutes at 4,000 rpm, then washed in equal volume and re-suspended in LB broth supplemented with 50 mM $MgCl_2$ (4 ml per *E. coli* strain; and 3 ml per *B. ambifaria* strain). These cell suspensions were combined in equal volume, and 50 µl of each cell suspension along with the combined suspension were spotted onto LB agar supplemented with 50 mM $MgCl_2$. The spotted agar plate was incubated for 5-6 hours at 37°C. Each growth patch was re-suspended in 1 ml TSB, and serially diluted. The neat suspension and serial dilutions were spread plated (100 µl per plate) onto both TSA and TSA supplemented with 150 µg/ml and polymyxin 600 U/ml. Plates were incubated for 48 hours at 37°C.

2.22. Comparing gene architecture of polyyne BGCs

Cepacin, collimomycin and caryoynencin BGC were manually extracted from representative encoding strains as previously described in previously (section 2.13.1) The representative BGCs were annotated using Prokka v1.12-beta (Seemann 2014) in default mode. The GenBank annotation files were entered into the Python application Easyfig (Sullivan *et al.* 2011), which generated BLAST comparison files (crunch files) of the input sequences, and produced a graphical representation overlaid with annotation information from the Genbank files.

2.23. Distribution of bacterial polyyne BGCs

A local nucleotide BLAST database was constructed of all available *Burkholderia*, *Paraburkholderia* and related genera with BLAST v2.6.0+ (Morgulis *et al.* 2008). A representative sequence of each polyyne BGC was used as a query against the databases, and the best match from each genome returned. The following nucleotide-nucleotide BLAST settings were applied: evalue 1e-5, max_hsps 1 and max_target_seqs 1. BLAST defined co-ordinates were used to extract the BGCs from multiple strains with the script extractSequence.pl written by Lee Katz available at GitHub (https://github.com/lskatz). The extracted cepacin BGCs were aligned using MAFFT v7.305b (Katoh *et al.* 2002) and a phylogenetic tree generated using FastTree v2.1.9 (Price *et al.* 2010).

Table 4. Primers used during this study.

Primers[a]	Final Product Size (bp)	Primer Use	Annealing Temperature (Polymerase)	Source
Fwd: 5'- GCG TCT AGA GAC GTG ATC ATT GCC GGA AA -3' Rev: 5'- GCG GAA TTC TTG CCC GAT ACA TAG AGC GT -3'	707	Amplify product from fatty AMP ligase-encoding gene, *ccnJ*	72°C (Q5)	This study
Fwd: 5'- TTA YTT TTG YGC CGC TAC MG -3' Rev: 5'- CCM GAG CAG CTY TAT ACG AT -3'	582	Screen for presence of c3 replicon in *B. ambifaria* BCC0191Δc3 (degenerate for *B. ambifaria*)	51°C (*Taq*)	This study
Fwd: 5'- AAG AAA TCT GCT GCC GCT TG -3' Rev: 5'- CAC TTC GCT GTA CCT CAA GC -3'	608	Screen for presence of pMinic3 in *B. ambifaria* BCC0191Δc3	53°C (*Taq*)	This study
Fwd: 5'-CAA TCA CCG GAT CCC CG-3' Rev: 5'-GAA GGA TGA TGT CGT GAG C-3'	Approx. 500	Confirming integration site of suicide vector pGpΩTp in *ccnJ*	51°C (*Taq*)	Fwd: ??? Rev: This study
RAPD272: 5'- AGC GGG CCA A -3'	Variable	Random amplified polymorphic DNA genotyping	36°C (*Taq*)	(Mahenthiralingam *et al.* 1996)
1300 sequencing primer: 5'-TAACGGTTGTGGACAACAAGCCAGGG-3'	Variable	Confirm insert cloned into suicide vector pGpΩTp	/	/

[a] Restriction sites in primer sequences are underlined.

Table 5. Standard thermal cycling protocol for PCR.

PCR Step	Temperature (°C)	Time (s)	Number of Cycles
Initial Denaturation	95	180	1
Denaturation	95	30	30
Primer Annealing	Variable[a]	30	
Primer Extension	72	50	
Final Extension	72	300	1

[a] See Table 4 for specific primer annealing temperatures.

Table 6. Components of a 50 µl PCR reaction using Q5 polymerase.

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

Table 7. Components of a 25 µl PCR reaction using *Taq* polymerase.

Component	Volume (µl)	Final Concentration
5X Q5 Reaction Buffer	10	1X
10 mM dNTPs	1	200 µM
10 µM Forward Primer	1.25	0.25 µM
10 µM Reverse Primer	1.25	0.25 µM
Template DNA	1	Variable
Q5 Polymerase	0.5	0.02 U/µl
5X Q5 High GC Enhancer	10	1X
Nuclease-Free Water	25	/

Table 8. Random amplified polymorphic DNA thermal cycling protocol.

PCR Step	Temperature (°C)	Time (s)	Number of Cycles
Initial Denaturation	94	300	1
Primer Annealing	36	300	4
Primer Extension	72	300	
Denaturation	94	300	
Denaturation	94	60	30
Primer Annealing	36	60	
Primer Extension	72	60	
Final Extension	72	600	1

2.24. Biological control modelling of *B. ambifaria*

To ascertain the involvement of antimicrobial specialised metabolites synthesised by *B. ambifaria* in realistic plant-pathogen interaction scenarios we developed a lab-based biological control model.

2.24.1. Bacterial plant pathogen biocontrol assay

To assess the ability of bacterial plant pathogens to inhibit the germination and emergence of *Pisum sativum*, the pea seeds were planted in soil containing a selection of plant pathogens (Table 3). Each bacterial plant pathogen was grown in TSB for approximately 18 hours at 30°C. These broth cultures were used to swab inoculate the surface of three TSA plates per pathogen, which were incubated for 3 days at 30°C. Following incubation, the bacterial growth was scraped from the surface, weighed, and an equal volume of sterile de-ionised (DI) water added (1 ml per 1 g) to create a high cell density suspension. Each potting well contained one seed planted in a 5:1 ratio compost-sand mixture and watered with a diluted cell suspension of 200 µl neat suspension in 15 ml DI water. Each pathogen was tested for pathogenicity against three seeds. The read-out for the assay was successful vs unsuccessful seedling emergence and signs of disease in seedlings.

2.24.2. Preparing Oomycete-infested soil

Multiple crop seeds were assessed for vulnerability to the damping-off disease oomycete *Pythium ultimum* Trow var. *ultimum* (MUCL 16164), these included pea, maize, broad bean, alfalfa, wheat, and soy bean. The degree of vulnerability was determined by exposure of the germinating crop seeds to *Pythum* infested soil and scoring seedling emergence. An initial infested soil stock was prepared fresh for each experiment. A 0.5 mm diameter plug of leading edge growth was used to inoculate a standard 90 mm potato dextrose agar (PDA) plate, and incubated for three days at 22°C. Following incubation the *Pythium*-covered PDA was sliced into squares approximately 0.5 mm^2 and mixed into 120 g of a compost-sand mixture. The compost-sand mixture (potting soil) was composed of a 5:1 ratio of compost and sand, using standard potting compost. The potting soil was incubated for three days at 22°C to produce the infested potting soil inoculum. By combining the infested potting soil at different ratios with un-infested potting soil, a plant-pathogenic soil was produced to challenge crop seed germination and seedling emergence. Both compost and sand were used un-sterilised for all biological control experiments.

2.24.3. Determining the relationship between cell density and seed-coat carrying capacity

B. ambifaria BCC0191 was grown in TSB for approximately 18 hours at 37°C, washed and re-suspended in PBS; then adjusted to 1 $OD_{600\ nm}$ which approximated mid-10^8 cfu/ml.. The cell suspension was serially diluted in PBS and pea seeds the three pea seeds submerged in each dilution for 1 hour with agitation. Following submersion, each seed was vortexed in 1 ml PBS, and the resulting cell suspension quantified by miles misra drop count. The drop counts were grown on TSA and incubated for approximately 18 hours at 37°C.

2.24.4. Preparing *B. ambifaria* seed-coat

The *B. ambifaria* biocontrol inoculum was generated by growing the strain of interest in TSB overnight at 37°C. The cells were pelleted by centrifugation at 4,000 rpm for 10 minutes, washed in equal volume of sterile PBS, re-pelleted, and then re-suspended in 0.5x volume of sterile PBS. The optical density was measured and adjusted to 5 $OD_{600\ nm}$, equivalent to approximately $1x10^9$ cfu/ml, and diluted to produce inoculums of $1x10^8$ cfu/ml and $1x10^7$ cfu/ml. Crop seeds were coated by submerging the seeds in suspensions of *B. ambifaria*, followed by immediate planting in the infested or control potting soil.

2.24.5. Comparison of biocontrol capabilities between *B. ambifaria* BCC0191 and mutants

To compare the biological control potential of multiple strains or mutants, the default condition of 1% *P. ultimum* infested soil was used, prepared as described in section 2.24.2. Groups of 10 pea seeds per condition were used per experimental replicate and repeated in triplicate. Each seed well was watered with 10 ml de-ionised water following planting and watered as required through the 14-day experiment time period. Biological control experiments were grown in dedicated growth cabinets at 22°C with a 16-hour light/8-hour dark photoperiod and 70% relative humidity (RH) for 14 days. After 14 days survival was assessed by the height of the seedling shoot: seedlings with shoots greater than 30 mm in length were considered to have germinated successfully. Seedlings that failed to emerge or with shoots less than 30 mm were considered severely stunted in growth and therefore unsuccessfully survived the pathogen encounter. The statistical significance of survival between different inoculum conditions and strain backgrounds was determined by two-sample t-test or Welch's two sample t-test. Assumptions for the two-sample t-test were normally distributed data (Shapiro-Wilk test) and equal variances (Bartlett test); Welch's two-sample t-test did not assume equal variances.

2.25. Rhizocompetence of *B. ambifaria* in a soil environment

We assessed the root colonisation ability of *B. ambifaria* following seed inoculation in the biological control model to understand the impact of the third replicon deletion on fitness in a soil environment. Pea seeds were coated with approximately $1x10^9$ cfu/ml (equivalent to approximately $1x10^7$ cfu/seed) of either *B. ambifaria* BCC0191 wild-type, cepacin-deficient BCC0191::*ccnJ,* third-replicon mutant BCC0191Δc3, or PBS, in triplicate. These seeds were planted in non-sterile un-infested potting mix, and allowed to grow under the conditions previously described (section 2.24.5). Following 14-days growth the pea plants were removed from the potting mix. The 1^{st} to 2^{nd} cm of primary root excised, weighed, briefly submerged in sterile PBS to remove loosely attached soil particles, and macerated in 1 ml sterile PBS. The resulting root suspension was serially diluted onto *Burkholderia cepacia* selective agar (BCSA; Oxoid UK), and incubated for 24 hours at 37°C to determine the total viable count. The resulting TVC was related back to 1 g of fresh root. Statistical significance was determined by two-sample (unpaired) t-test with the assumption of equal variances (Barlett test) and normally distributed data (Shapiro-Wilk test). The non-parametric Wilcoxon rank sum test was employed failing these assumptions.

2.26. Preparing pea-exudate biomimetic media

To understand the plant-mediated induction of specialised metabolite biosynthetic gene clusters we prepared a solid agar medium based on the exudates that are passively released during seed imbibition and germination. Approximately 100 g of Early Onward variety pea seeds were washed in de-ionised water, and later suspended in 400 ml sterile de-ionised water for two days at room temperature (~21°C) with agitation to aid the diffusion of exudate into the surrounding water. Following incubation the pea-exudate was recovered and filter sterilised. The seed-exudate was first passed through a coarse fibreglass filter to remove larger debris, followed by a 0.2 µm filter. The sterile exudate was combined with autoclaved 3% purified agar cooled to approximately 55°C to generate a final agar composed of 50% seed-exudate and 1.5% purified agar base. *B. ambifaria* was grown on the pea-exudate agar for metabolite production in an analogous approach to that described previously for BSM-G (section 2.14).

2.27. Comparing antimicrobial production between different media types and mutants

B. ambifaria BCC0191 and BCC0203 were grown on pea exudate medium, BSM-G and TSA to compare the antimicrobial induction properties of the different media types. BCC0191 and BCC0203 were grown in TSB for approximately 18 hours at 37°C, and 2 µl of the liquid cultures spotted onto the agar. The BCC0191 and BCC0203 inoculated plates were incubated at 22°C and 30°C, respectively, for three days. Following incubation the bacterial growth was chloroform killed, and overlaid with half-strength isosensitest agar seeded with susceptibility organisms, as described in section 2.17. Plates for HPLC analysis were produced as described (section 2.15), with the same the incubation temperature used for the antimicrobial assay.

2.28. *In vivo G. mellonella* wax-moth larvae infection model

G. mellonella larvae were purchased from BioSystems Technology Ltd TruLarv to ensure uniform age and genetic background. The method described was adapted from Agnoli *et al.* (2012). *B. ambifaria* BCC0191 wild-type (WT) and c3-deficient mutant (Δc3) were grown for approximately 18 hours in TSB at 37°C. Liquid cultures were washed, re-suspended in PBS, and adjusted to approximately 1 x10^6 cfu/ml. Larvae were injected with 10 µl of bacterial suspension in the hindmost right proleg using a 25 G × 1” needle (BD Microlance 3) attached to a 1705TLL, 50 µl syringe (Hamilton). PBS was injected as a control, and 10 larvae were injected per condition. Larvae were monitored six times over 72 hours (time points: 18, 24, 42, 48, 66 and 72 h post-inoculation), and survival was determined by physical movement

following agitation. The experiment was conducted three times, and the inoculum total viable count confirmed for each replicate via Miles Misra drop-counts.

2.29. Murine lung inhalation infection model

The chronic murine respiratory infection model was conducted as previously described for *P. aeruginosa* (Fothergill *et al.* 2014; Bricio-Moreno *et al.* 2018). The experiments were conducted by Dr Angharad Green and supervised by Dr Daniel Neill at the University of Liverpool with approval from the UK Home Office and University ethics committee. Mice were randomised to cages (experimental groups) by animal unit staff with no involvement in study design. Researchers were not blinded to the experimental groupings. Good tolerance was observed in mice with a high infectious dose of 10^6 cfu/mouse during an initial dosing study. Sample size was calculated to give 95% power at $\alpha = 0.05$ to detect a >4-fold difference in lung cfu between BCC0191 WT and BCC0191Δc3, assuming a mean viable count of 100 in the BCC0191 WT lung and a standard deviation of 50% of the mean as determined in preliminary experiments. Groups of 6 mice (female BALB/c, 6–8-weeks old; Charles River) were infected intranasally with approximately 2 $x10^6$ cfu/mouse under anaesthesia (O2/isolfuorane). Mice were culled at 24 h, 48 h, and 5 d post-infection, and the nasopharyngeal tissue, lungs and spleens were removed. Tissues were homogenized in 3 ml PBS, serially diluted and plated onto *B. cepacia* selective agar (BCSA) to determine the total viable count. No animals were excluded from the analysis, and animals were monitored for symptoms of disease. Microbial growth was removed following growth on BCSA and subject to PCR genotyping by random amplified polymorphic DNA fingerprinting to confirm identify (Table 4 and Table 8).

3. Comparative genomic analysis of *B. ambifaria*

3.1. Introduction

Early endeavours to identify and distinguish the multi-species *B. cepacia* complex exploited a PCR-based identification method focused on the *recA* gene (Coenye *et al.* 2001c). The single copy *recA* gene possessed greater nucleotide variability than the more widely used 16S rRNA gene. Two *B. cepacia* complex species would possess 98-99% 16S rRNA gene nucleotide similarity; this was reduced to 94-95% with *recA*, providing a means of distinguishing members of the closely-related species complex (Coenye *et al.* 2001c). However, to better distinguish strains within a species, more discriminatory multi-locus typing schemes such as the seven-gene multi-locus sequence typing (MLST) (Baldwin *et al.* 2005) and 53-gene ribosomal MLST (rMLST) (Jolley *et al.* 2012) were developed. MLST exploits seven single copy housekeeping genes, in contrast to rMLST that uses conserved ribosomal protein-encoding genes. The enhanced resolution of these multi-locus typing schemes provided a robust means of assessing the diversity of *Burkholderia* species and their global distribution (Depoorter *et al.* 2016). Multi-locus typing schemes provided valuable information on the relatedness of strains, and distinguishing species, however, DNA-DNA hybridisation (DDH) remained the gold standard for delineating and defining novel bacterial species (Richter and Rossello-Mora 2009). DDH was developed in 1969 and relies on the nucleotide homology between two different genomes; the annealing of complementary DNA strands between genomes forms hybrid molecules, and the degree of hybridisation corresponds to the relatedness of the genomes (Brenner *et al.* 1969).

In the era of cost-effective genomics inundating public databases with high quality sequence data, the use of whole genome sequences to delineate strains and species for large genomic analyses is a reality for the scientific community. The availability of genome sequences presented an opportunity to exploit the entire genomic content of an isolate to accurately identify a clone, strain and species for a plethora of research and clinical applications. High quality genome sequences have enabled the detection of single nucleotide polymorphisms (SNPs) facilitating research studies from hospital-scale epidemiology (Lewis *et al.* 2010) to global geographical distribution and surveillance studies (Wong *et al.* 2016). *In silico* genome comparisons termed average nucleotide identity (ANI) analyses have been proposed as an alternative to traditional DDH assays to determine bacterial species boundaries. A DDH species threshold of 70% corresponded closely to an ANI of 95% (Goris *et al.* 2007). Initial ANI analyses relied on the nucleotide BLAST algorithm to generate pairwise comparisons between sequence fragments (Goris *et al.* 2007). Developments in bioinformatics tools and approaches have streamlined ANI analyses, providing high-throughput tools capable of

substituting BLAST-based sequence alignments in favour of rapid alignment algorithms (Kurtz *et al.* 2004; Pritchard *et al.* 2016), and even removing the need for alignments altogether (Jain *et al.* 2017).

The genus *Burkholderia* has benefited significantly from the rapid development of DNA sequencing technology. In December 2010, only 23 *Burkholderia* genomes and 21 Illumina paired-end read sets were publically available; this expanded to 591 genomes by December 2015, and as of December 2019, 3097 *Burkholderia* genomes and 4,828 Illumina paired-end read sets were publically available. The number of *B. ambifaria* genomes has expanded alongside the *Burkholderia* genus; starting with the first *B. ambifaria* genome sequence of the type strain AMMD (LMG 19182^T) in 2007, originally isolated from the rhizosphere of a pea plant (Coenye *et al.* 2001a). As of June 2019, 10 *B. ambifaria* genomes were publically available, and 70 Illumina paired-end read sets.

The substantial increase in *Burkholderia* genomes has enabled genomic-level analyses of *Burkholderia* species, clades, and the genus as a whole; identifying their core, accessory and pan-genomes, secondary metabolite potential, risk factors for virulence, and species classification. Several *Burkholderia* comparative genomics studies have focussed on understanding the pathogenic members of the genus in the *Burkholderia cepacia* complex (Peeters *et al.* 2017), *B. pseudomallei* complex (Ho *et al.* 2011; Spring-Pearson *et al.* 2015) and plant-pathogenic complex (Seo *et al.* 2015). Multi-genome analyses have proven useful for identifying targets capable of distinguishing species commonly misidentified by biochemical assays (Ho *et al.* 2011). The *in silico* extraction and analysis of proteins encoded by multiple genomes highlighted targets for multiplex PCR to distinguish between *B. cepacia* complex and *B. pseudomallei* clade species (Ho *et al.* 2011). Genome mining has emerged as a concept for assessing the specialised metabolite potential of a genome or collection of genomes by predicting biosynthetic pathways *in silico*. Genome mining combined with comparative genomics can determine the distribution of specialised metabolites, and therefore the distribution of predicted functional attributes of a given species or genus. This methodology has led to the discovery of the biosynthetic origin for a historical metabolite, cepacin (Mullins *et al.* 2019), as well as detecting gene clusters outside of their originally characterised species background. The simplest application of comparative genomics is species identification by comparing a query genome to a collection of known genomes. This approach is also capable of identifying previously un-typed genomes as representatives of characterised species (Vandamme *et al.* 2017), thereby curating publically available sequence data while also expanding genomic representatives of novel species.

3.1.1. Aims and objectives

This chapter summarises research that expanded the genomic understanding and population biology of *B. ambifaria* by addressing the following objectives:

1) To sequence, assemble and verify the identity of *B. ambifaria* genomes.

2) To define the core, accessory and pan-genome of *B. ambifaria*.

3) To investigate the genetic diversity and population biology of *B. ambifaria* using phylogenomics.

4) To analyse the genome architecture of *B. ambifaria* by exploiting long-read Oxford Nanopore Technology to generate hybrid complete genome assemblies.

5) To exploit the genomic data and predict functional phenotypes of individual strains and the collective species.

3.2. Results

3.2.1. Multi-locus sequence typing

To gain an initial understanding of the diversity of the bacterium *B. ambifaria*, a multi-locus sequence phylogeny was constructed as a precursor to whole genome sequencing. The multi-locus sequence typing (MLST)-based phylogeny highlighted 15 distinct clades that accommodated 106 of the 119 strains (87%) (Figure 2). Based on this low-resolution species phylogeny, 64 were chosen for whole genome sequencing that represented the species diversity, and for which the strains were locally available at Cardiff University.

3.2.2. ANI analysis of *B. ambifaria* genomes

At the time of analysis 67 draft or complete genomes identified as *B. ambifaria* were available: 62 genomes sequenced internally and identified as *B. ambifaria* in the Cardiff University BCC collection; and five available in public databases (MEX-5, IOP40-10, MC40-6, AMMD, and RZ2MS16). The majority (64 out of 67) of the genomes possessed pairwise ANI values greater than the 95% species delineation threshold applied to the analysis, with no strain-strain pairwise comparison below 96% (Figure 3). This ANI analysis confirmed the *B. ambifaria* species designation of the 64 genomes, and established the genomic dataset upon which the remainder of the bioinformatics and phylogenomic components of this research were underpinned.

Based on the ANI analysis the following three genomes were excluded from the dataset: BCC1630, BCC1638, and RZ2MS16. The genome assemblies for strains BCC1630 and BCC1638 possessed 99.9% nucleotide identity to one another, but 92.9-93.2% identity to the ANI-verified *B. ambifaria* genomes. The publically available genome RZ2MS16 possessed a maximum of 93.7% ANI to the ANI-verified *B. ambifaria* genomes. The ANI analysis highlighted erroneous identities of strains stored in the Cardiff University BCC collection as well as publically deposited genomes, and indicated BCC1630, BCC1638 and RZ2MS16 represented two *B. ambifaria*-like species. Closer examination of ANI strain grouping revealed three major clusters: ANI-1, ANI-2 and ANI-3 (Figure 3). All 38 *B. ambifaria* strains within the ANI-1 cluster possessed 97.6-100% sequence identity to one another, but 96.0-97.2% identity to the remaining *B. ambifaria* strains analysed. The ANI signal was stronger in clusters ANI-2 and ANI-3, as evidenced by the longer tree branches in the core-gene phylogeny. The nine cluster ANI-2 strains possessed pairwise sequence identity of 98.2-100%, but only 96.0-97.1

to the remaining strains; while the 15 strains in the ANI-3 cluster had 98.6-100% pairwise identity, but 96.2-97.0% outside of the cluster.

3.2.3. Illumina-sequenced genome statistics

Of the 64 ANI-confirmed *B. ambifaria* genomes included in the final analysis, 60 were sequenced internally as part of a study at the Wellcome Sanger Institute, and four were downloaded from public databases. Of the internal genomes, 59 were sequenced via the Illumina platform and one via the PacBio platform. Of the public database genomes, Illumina platform was used for AMMD and MC40-6, which represent complete genomes; the remaining two publically available genomes, MEX-5 and IOP40-10 were sequenced using the 454 pyrosequencing platform. The sequencing platform and consequent assembly pipeline influenced the final sequence assembly quality: final contig assembly number, related N50 values, and the degree of success scaffolding contigs into replicons as follows.

3.2.3.1. Contig numbers and N50 values

The complete genomes AMMD and MC40-6, in addition to the PacBio sequence BCC0203, assembled into the three replicons c1, c2 and c3 along with a plasmid encoded by each strain, totalling four contigs. Both 454 platform genomes MEX-5 and IOP40-10 possessed 706 and 629 contigs, respectively. The 59 internal Illumina genomes ranged between 44-302 total contigs, with a mean average of 106 contigs (Table 2). The N50 value varied considerably in the 59 Illumina genomes, with a range of 234-1,593 kbp, and mean of 432 kbp. The N50 value of BCC1105 (1,593 kbp) was almost twice that of the nearest genome BCC0410 (850 kbp), this was due to the natural absence of the third replicon c3 in BCC1105, affecting the total sequence length of the genome, and subsequently the N50 value (Table 2). The complete genomes for BCC0203, AMMD and MC40-6 possessed N50 values averaging 2.66 Mbp; while the earlier 454 platform-sequenced IOP40-10 and MEX-5 genomes possessed extremely low N50 values averaging 22.5 kbp.

3.2.3.2. Total genome sequences and scaffolding contigs

The total sequence length of the 64 *B. ambifaria* genomes varied considerably, with a minimum length of 6.30 Mbp in BCC1105 and a maximum of 8.03 Mbp in BCC1248 (Table 2). The minimum *B. ambifaria* total sequence increased to 7.24 Mbp if BCC1105 was excluded as this strain naturally lacks the third replicon c3. The mean average total sequence length of the 64 genomes was 7.49 Mbp. The existence of three complete genomes (BCC0203, AMMD and MC40-6) permitted the scaffolding of contigs into replicons, providing an opportunity to compare replicon sequence length and synteny among the 64 genomes.

Most genomes (50 out of 64) had >97% of the contig sequence data (% total basepairs) scaffolded into replicons; and 62 out of 64 had >93% scaffolded. Only the 454 sequenced genomes MEX-5 and IOP40-10 had <90% of their sequence scaffolded into replicons. Excluding BCC1105, MEX-5 and IOP40-10 due to lack of replicons or poor sequence quality, the remaining 61 scaffolded genomes ranged in size by 0.44 Mbp, with minimum and maximum lengths of 7.14 Mbp and 7.58 Mbp, respectively (Table 2). However, a comparison of all genomes with three replicons (only excluding BCC1105) more than doubles this genome size range to 0.99 Mbp. Comparing the mean total genome size of core phylogenomic clades 1-3 (see section 3.2.5) revealed minor differences. Clades 1-3 possessed average total sequence lengths of 7.43 Mbp, 7.60 Mbp and 7.57 Mbp, respectively. This observation was mirrored in the clade 1 sub-clades a, b, c and d (see section 3.2.5), with averages of 7.43 Mbp, 7.31 Mbp, 7.65 Mbp and 7.39 Mbp, respectively. However, comparing the range of each clade did reveal appreciable variation; clade 1 possessed the largest range at 1.73 Mbp (0.79 Mbp if BCC1105 was excluded), while clade 3 possessed the lowest range at 0.47 Mbp, the range of clade 2 was 0.68 Mbp. At the replicon level, there was considerable variation in length across c1, c2 and c3; with mean lengths of 3.46 Mbp, 2.74 Mbp and 1.15 Mbp, respectively. Relative to the average length of the replicons, c3 fluctuated most in length at 41.6% (size range of 0.48 Mbp). Replicon c1 fluctuated the least at 8.9% (size range of 0.30 Mbp), and c2 varied in size by 25.9% around the average c2 replicon length (size range of 0.71 Mbp).

3.2.3.3. Gene variation at replicon and genome level

The gene content of the total sequence for each *B, ambifaria* strain was assessed by extracting the gene presence-absence data outputted from the Roary analysis. The mean gene number encoded by *B. ambifaria* was 6,576 genes, with a range of 1,615 genes; this range reduced to 931 genes when BCC1105 was excluded due to the absence of a third replicon. Excluding BCC1105, the minimum gene content was encoded by BCC0748 (6,308 genes), and the maximum by BCC1233 (7,239 genes). Individual replicons were also subject to Roary analysis, enabling the gene content of each replicon to be defined. Replicons c1, c2 and c3 encoded 3,116, 2,398 and 954 genes, respectively. Replicon c2 boasted the greatest variation in gene content, with a range of 551 genes; MEX-5 encoded 2,205 genes and BCC1236 encoded 2,765 genes, representing the minimum and maximum gene content for replicon c2. The largest *B. ambifaria* replicon c1 encoded the least gene content variation, with a range of 288 genes; BCC0478 encoded the lowest number of c1 genes (3,004 genes), while BCC0480 encoded the highest c1 gene count at 3,292 genes. A maximum of 1,054 genes (BCC1041) was observed for the smallest replicon, c3, and a minimum of 721 genes in MEX-5; with a medium gene content range of all three replicons at 333 genes.

The difference between replicon gene content and total gene content was assessed to determine the approximate number of genes encoded by unscaffolded contigs. These contigs represent sequence data that is encoded as plasmids or genomic DNA that lacked homology with the three references used to construct the replicons. Only 17 of the 64 strains encoded more than 100 genes outside of the re-constructed replicons, the largest discrepancies were in strains MEX-5 (1,067 genes) and IOP40-10 (665 genes), likely the result of the lower quality sequence data in comparison to the Illumina and PacBio derived sequence data. Other strains, such as BCC0203 (470 genes) and MC40-6 (293 genes) are known to encode plasmids that are potentially responsible to the unscaffolded contigs.

3.2.3.4. Genomic and replicon GC content

The GC content of total genomic DNA was relatively consistent across the 64 strains, with a mean of 66.65% and range of 0.87%; in contrast, the GC content was not stable between the replicons. Replicons c1 and c2 possess a similar range in GC content, 0.52% and 0.57%, respectively, while c3 has almost double the range at 1.02%. The mean GC content for c1, c2 and c3 was 66.90%, 66.74% and 66.27%, respectively.

3.2.4. Pan, core and accessory genome analysis

The pan, core and accessory genome of the species was determined by Roary-based analysis of the 64 *B. ambifaria* genomes. Following coding sequence prediction, a blastp sequence identity threshold of 95% was imposed to identify homologous proteins across species. Based on this threshold, a pan-genome of 22,376 genes was defined for the 64 genomes, of which, 3,784 genes were encoded by all 64 *B. ambifaria* and represented the core genome. The remaining 18,592 genes comprised the accessory genome, approximately 4.9x larger than the core genome. Most of the accessory genes, 78.4% (14,582 genes) were encoded by a minority (<15%) of the *B. ambifaria* genomes. That is, for a given accessory gene out of the 14,582 genes, the gene was encoded by less than 15% of the *B. ambifaria* genomes. The pan-genome was approximately 3.4x larger than the mean gene content of a *B. ambifaria* strain, 6,576 genes. Strain BCC1233 encoded the most genes (7,239 genes), while BCC0478 encoded the least (6,308 genes) when excluding BCC1105 (5,624 genes). Exclusion of *B. ambifaria* strain BCC1105 (naturally lacking c3) from the core genome analysis resulted in a 382 gene increase in the core genome (total 4,166 genes). The majority of the additional genes were attributable to the third replicon, c3, which were excluded from the initial analysis due to the absence of these genes from one of the analysed strains (BCC1105).

The pan, core and accessory gene content of each re-constituted replicon was assessed to understand the distribution of accessory and core genes across the genome, as well as locally per replicon. Replicon c1 encoded the highest number of both core (2,289) and accessory (6,106) genes, as expected, considering c1 is the largest of the three replicons. Replicon c3 encoded the fewest core (299) and accessory (2,694) genes compared to c1 and c2; while c2 encoded 1,407 core genes and 5,646 accessory genes. On average, c1, c2 and c3 encoded 3,116, 2,398 and 954 genes, respectively. The size of the pan-genome relative to the mean gene content per replicon was relatively consistent between the replicons. Despite c1 physically encoding the largest replicon pan-genome, replicons c1, c2 and c3 possessed pan-genomes 2.7x, 2.9x and 3.1x larger than their mean replicon gene content, respectively. However, an appreciable variation was observed in the distribution of the core and accessory genes across the replicons relative to their size. The accessory genome of c3 was approximately 9x larger than the c3 core genome; while the accessory genomes of c1 and c2 were only 2.7x and 4.0x greater than the c1 and c2 core genomes, respectively. A comparison of the percentage of core genes relative to the average gene content per replicon unequivocally demonstrated the replicons that possess the greatest variability in the *B. ambifaria* genome. More than half of the average gene content of replicons c1 and c2 consisted of core genes, 73% and 59%, respectively; whereas only 31% of the average gene content of c3 encoded core genes.

The distribution of core and accessory genes across each replicon was determined by extracting the respective gene datasets predicted by Roary and mapping them via BLAST nucleotide homology to a representative *B. ambifaria* genome, type strain AMMD (BCC0207) (Figure 4). The core genes of c1 and c2 were uniformly distributed across both replicons, with the exception of several regions predicted to encode genomic islands or specialised metabolite BGCs. In contrast, the core genome of replicon c3 was considerably fragmented and interspersed with large regions, up to 210 kbp, of accessory genes (Figure 4). These accessory gene regions included large BGCs and predicted genomic islands, in addition to long stretches of genes without a defined collective function based on specialised metabolite and genomic island predictive software.

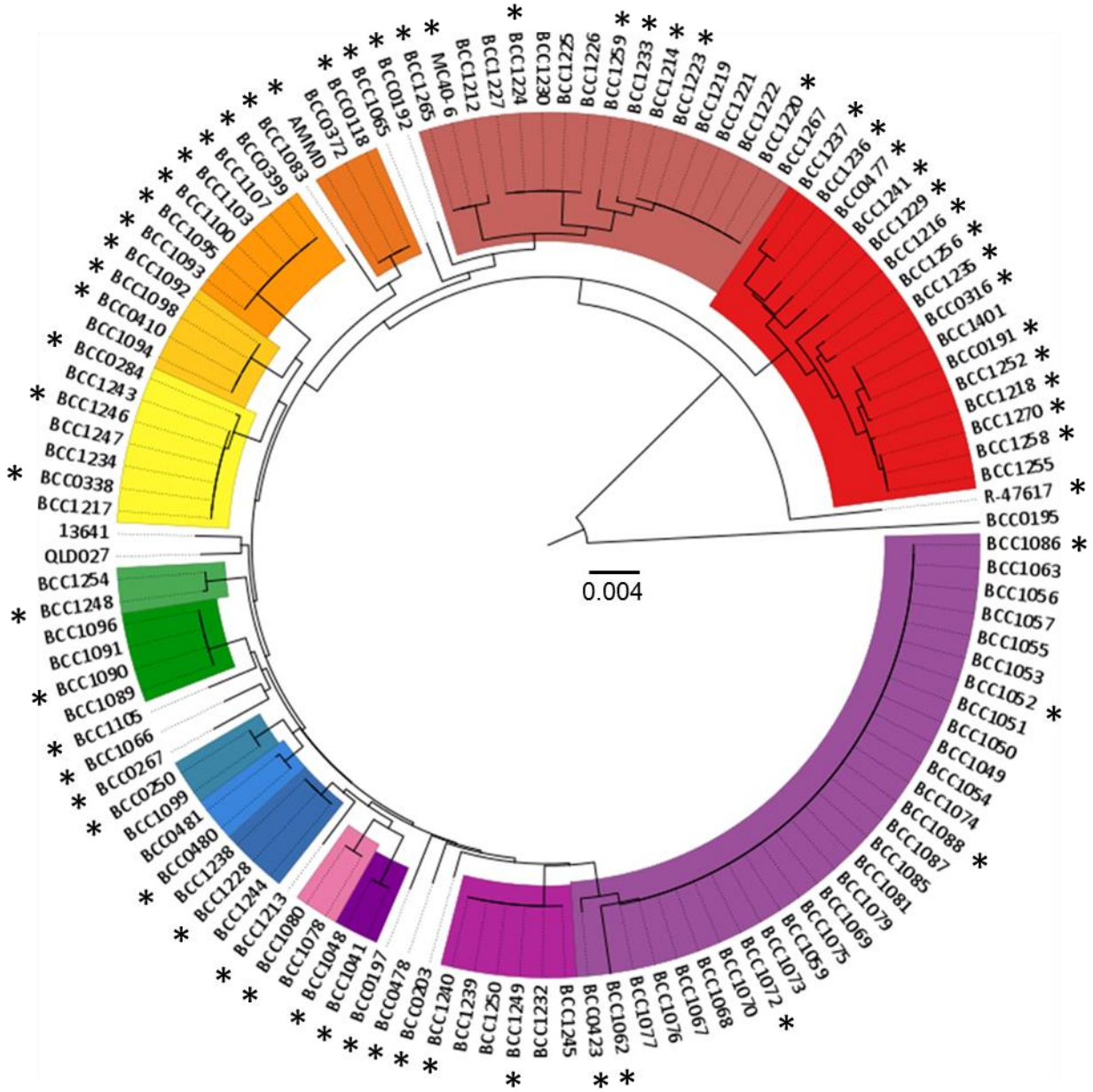


Figure 2. Multi-locus sequence typing phylogeny of *B. ambifaria* strains.
Concatenated alignment of seven essential genes: *atpD, gltB, gyrB, lepA, phaC, recA,* and *trpB*. Applying an 80% main support bootstrap threshold and maximum genetic distance of 4.5% revealed 15 major clades (colour coded). Whole genome sequenced strains are indicated with asterisks. Scale bar represents substitutions per base position.

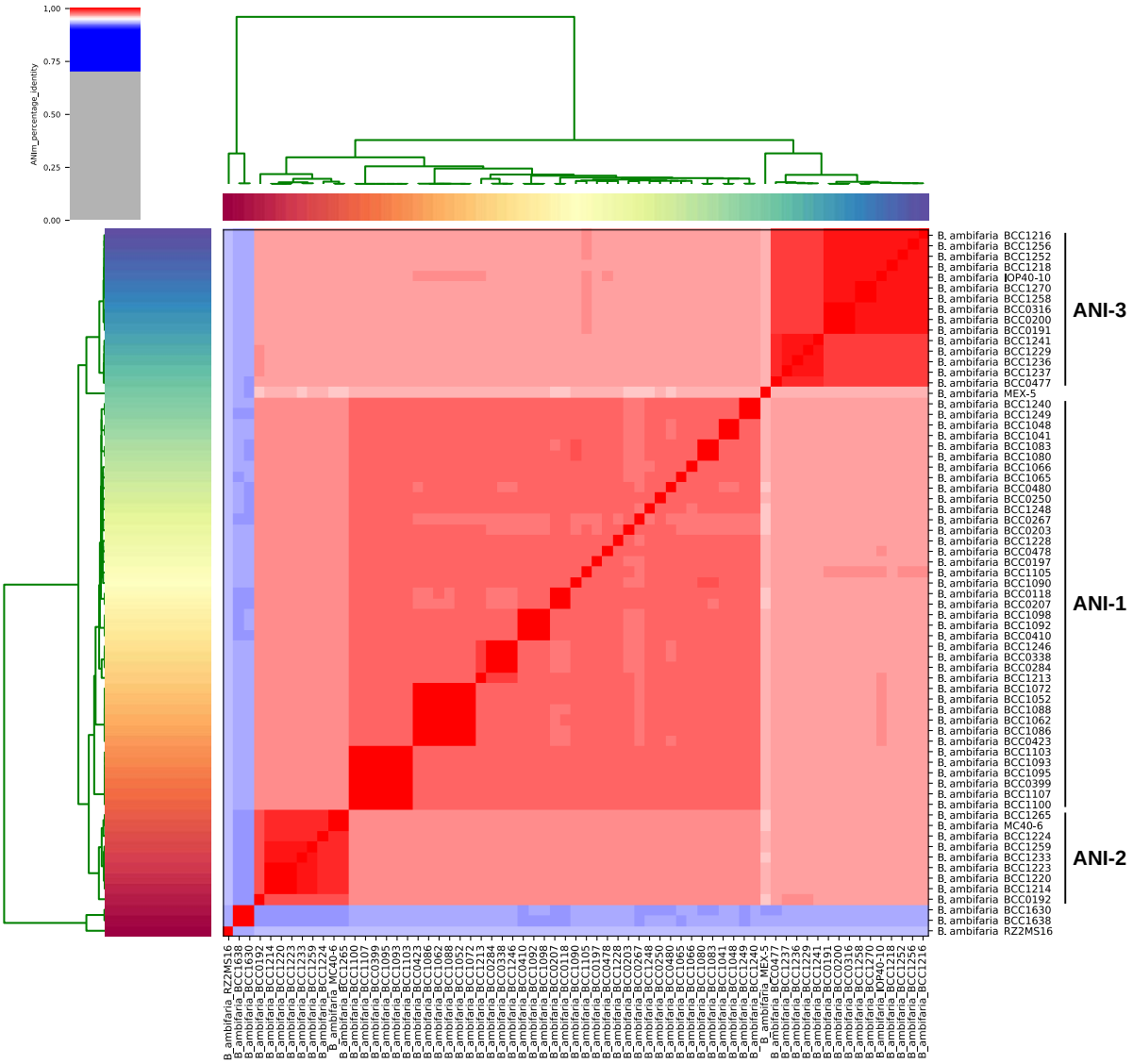


Figure 3. Visual representation of ANI analysis generated by PyANI script.
The heatmap indicates the degree of nucleotide similarity between the 67 putative *B. ambifaria* genomes. The red highlights strains that possess >95% nucleotide similarity; the darker the red, the greater the similarity. The blue highlights strains with <95% similarity. The heatmap was ordered by row and column dendrograms constructed based on the nucleotide similarity value. The three ANI clades are labelled to the right of the heatmap.

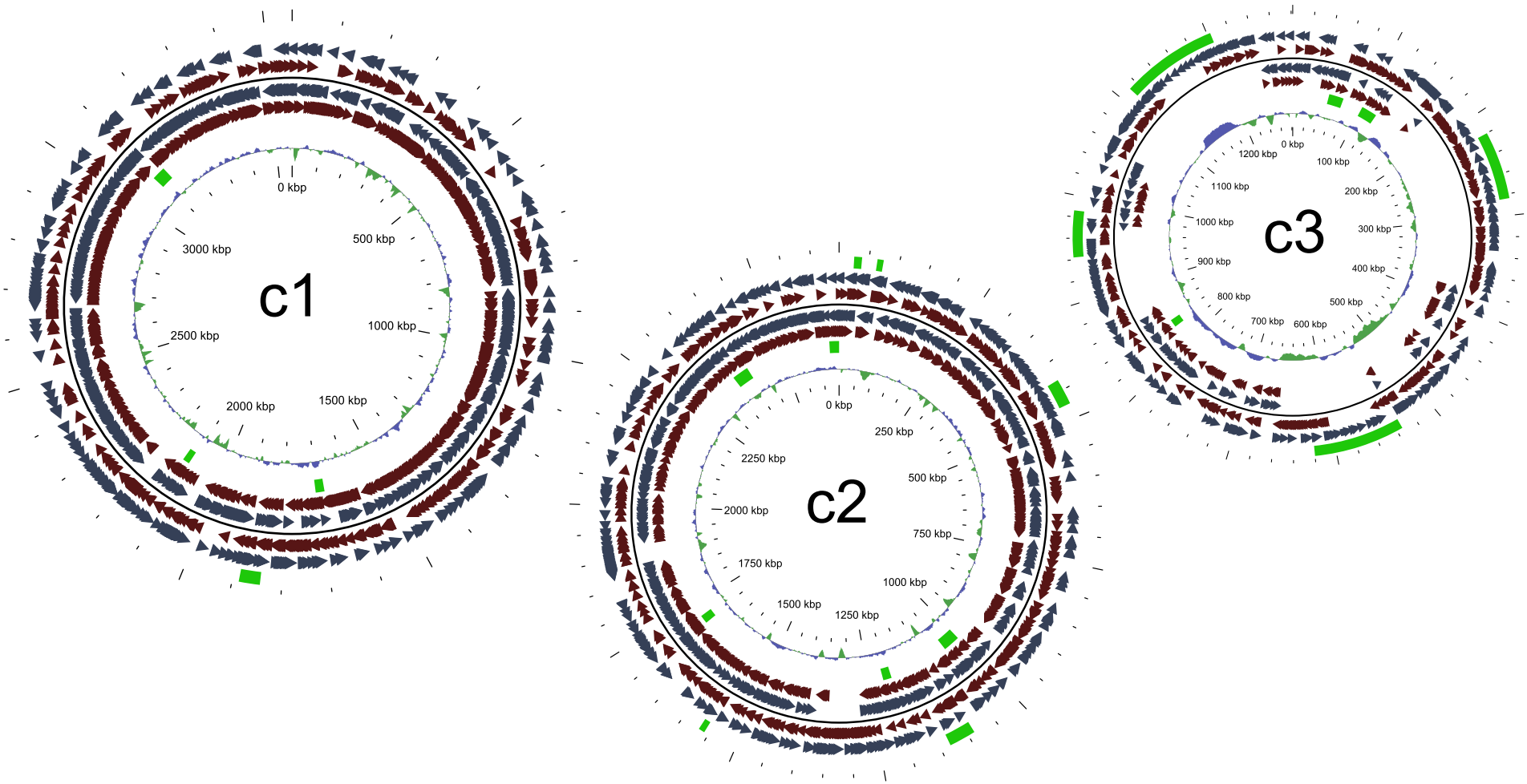


Figure 4. Distribution of core and accessory genome, and BGCs across the three *B. ambifaria* replicons in strain BCC0207.

The following features are represented by tracts in each replicon figure (innermost to outermost): GC content (1); core BGCs (2); core genes, forward (3) and reverse (4) reading frames; accessory genes, forward (5) and reverse (6) reading frames; accessory BGCs (7).

3.2.5. Core-gene phylogenomics

Accurate strain-to-strain relatedness was determined by aligning the 3,784 core genes identified by Roary and constructing a maximum likelihood phylogeny (Figure 5). The phylogeny revealed that *B. ambifaria* was divided into three major clades (clades 1-3), with several sub-clades present in clade 1 (sub-clades a-d). The largest clade, clade 1, harboured 59% (38 of 64) of the *B. ambifaria* strains. Clades 2 and 3 harboured 14% (9 of 64) and 23% (15 of 64) of the strains, respectively; while two strains, BCC1066 and MEX-5 did not reside within the defined three clade phylogeny structure. *B. ambifaria* MEX-5 was the most distantly related of all 64 strains, forming a natural outgroup of the phylogeny (Figure 5). To support the use of MEX-5 as the outgroup the Roary core gene analysis was repeated with the inclusion of *B. vietnamiensis* G4 genome, generating a 1,594 core gene alignment. A core gene phylogeny based on this alignment highlighted MEX-5 as the *B. ambifaria* strain most closely related to the outgroup *B. vietnamiensis* G4 (Figure 6). The majority of branch nodes in the core gene phylogeny possessed bootstrap values >70%, of which, most were greater than 90%. Nodes with weaker bootstrap values (≤69%) were infrequent (12 nodes) and occurred mainly in clade 1 (11 of 12 nodes).

3.2.6. Comparative replicon phylogenomics

The distribution of the *B. ambifaria* genome across three replicons, and the discovery that the smallest replicon (c3) is a large virulence plasmid (Agnoli *et al.* 2012) led to the question of whether natural mobilisation of c3 to other strain backgrounds could occur. To address this question high-resolution core-gene phylogenies were constructed for each replicon, and their topology compared to that of the core-gene phylogeny constructed from the entire genome. The core-gene phylogenies for replicons c1, c2 and c3 were based on alignments of 2,289, 1,407 and 299 genes, respectively (section 3.2.4). High congruence was observed between the whole-genome and replicon c2 core-gene phylogenies, with minimal variation in topology and strain positioning. Clade 1b diverged before clade 1a in the c2 core-gene phylogeny, strain BCC0480 migrated from clade 1b to clade 1a, and the formally clade-less BCC1066 was positioned in clade 1a. High congruence of the main branches of clade 2 and clade 3 were observed across the four phylogenies. Minor differences were observed in the positioning of strains in closely related clades, and no major changes were observed in phylogeny topology across the c1, c2 and c3-based trees (Figure 7). Variations in strain positioning was explained by the high sequence identity between species within clades (Figure 3). Highly congruent tree topologies supported vertical acquisition of all three replicons; and that no mobilisation and transfer of *B. ambifaria* replicons, especially c3, have occurred within the species or with members of the *B. cepacia* complex.

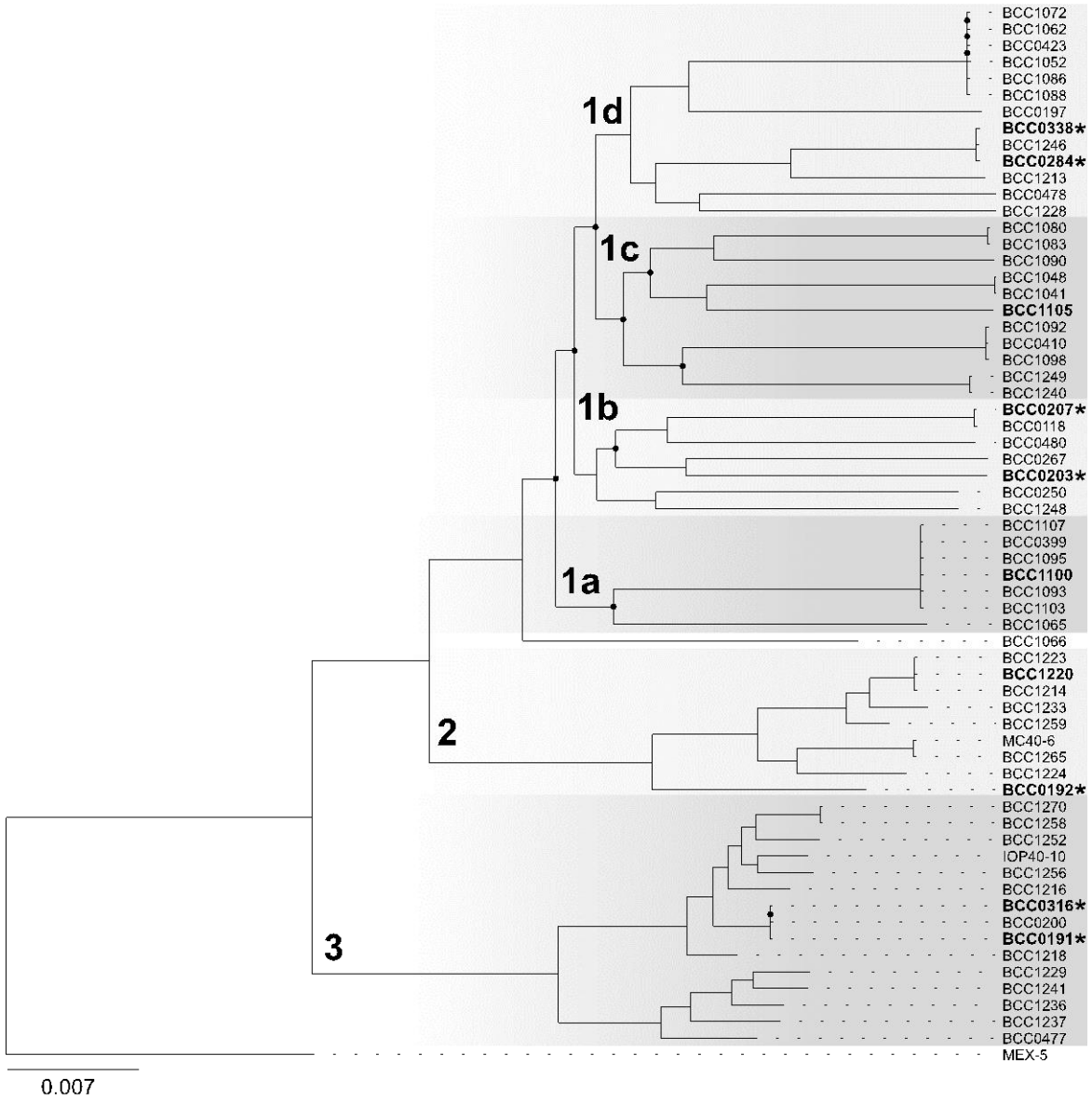


Figure 5. Core-gene phylogeny of 64 *B. ambifaria* genomes.
A core-gene alignment of 3,784 genes was constructed with Roary, and the phylogeny created using RaxML (100 bootstraps). Nodes with bootstrap values of <70% are indicated with black circles. Three major clades were identified that encompassed 62 of the 64 genomes. Strains that were analysed for antimicrobial metabolites via LC-MS are indicated in bold. Strains that represent previously identified biological control strains are indicated with an asterisk. Scale bar represents the number of substitutions per base position.

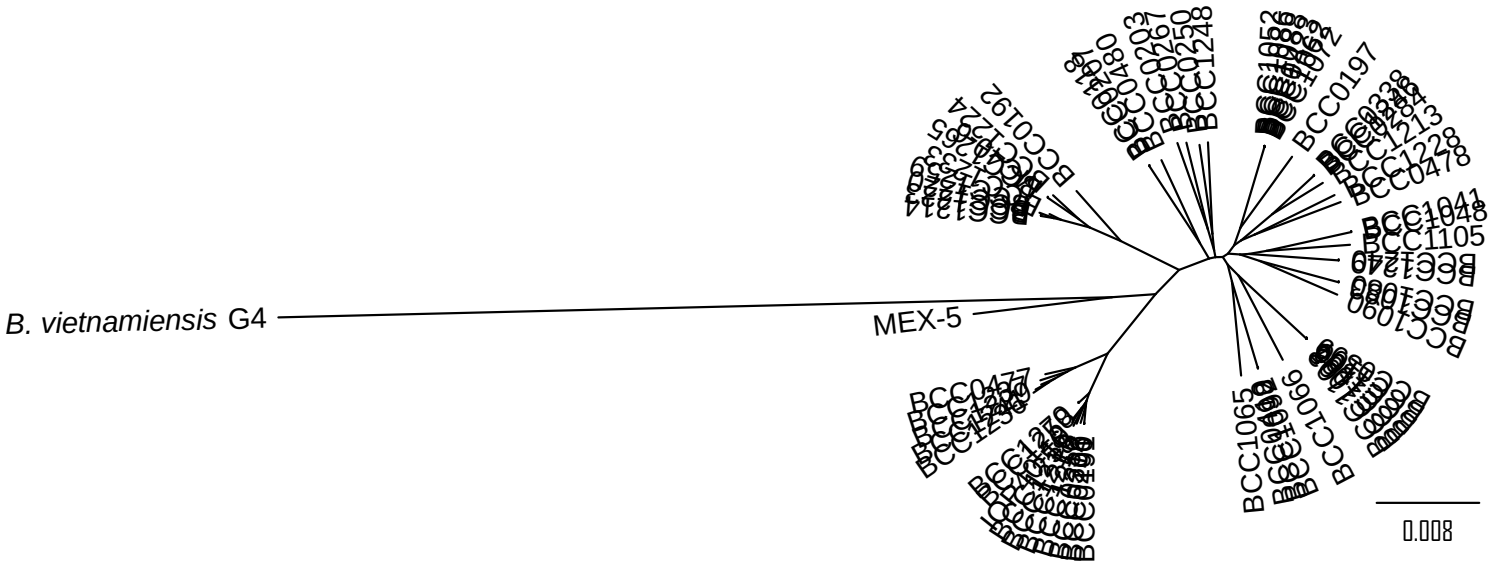
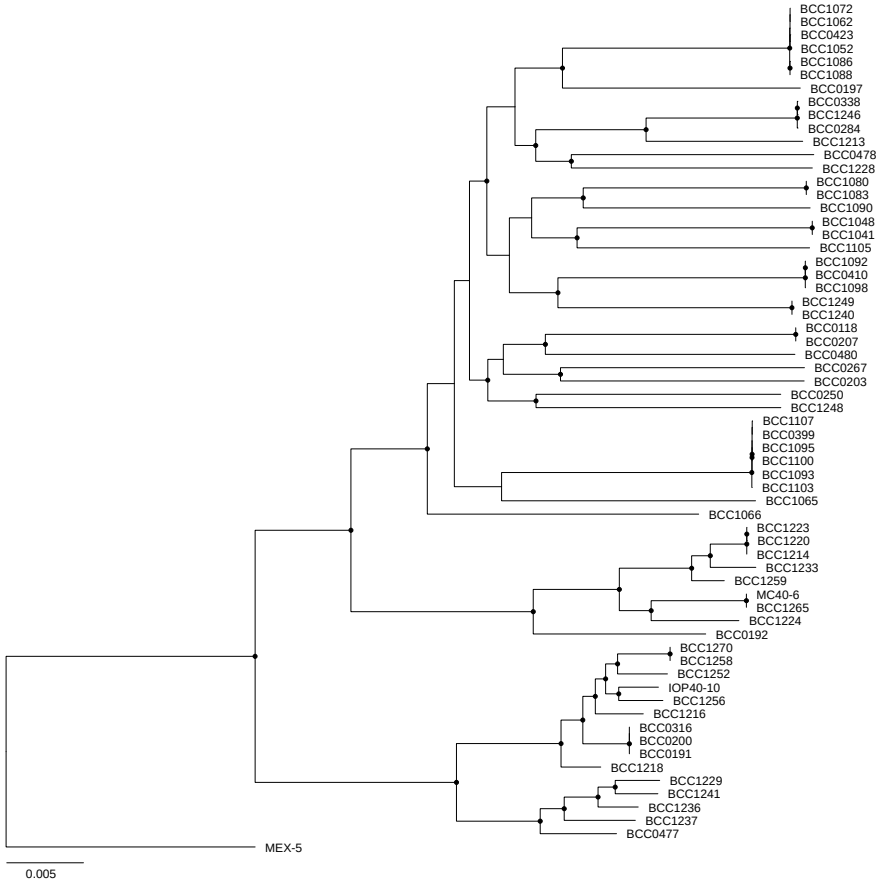
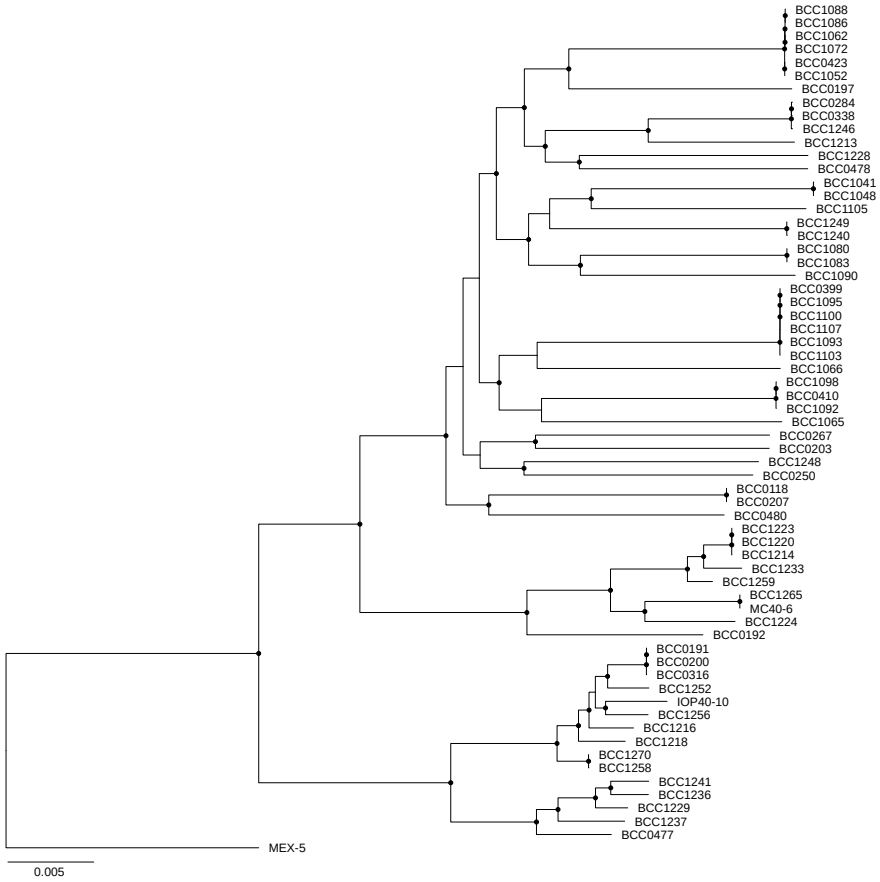


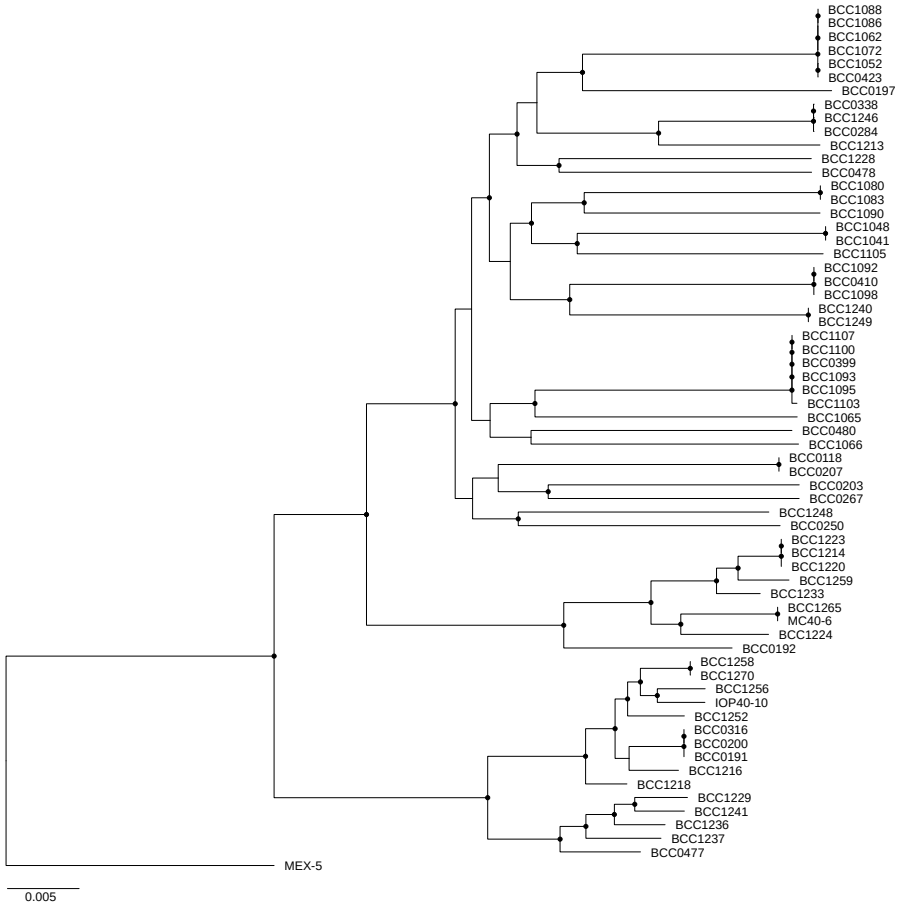
Figure 6. Rooted core-gene phylogeny of *B. ambifaria*.
Core-genes shared between the 64 *B. ambifaria* genomes and the outgroup *B. vietnamiensis* G4 were extracted and a core-gene alignment of 1,594 genes generated. *B. vietnamiensis* G4 branches closest to *B. ambifaria* MEX-5, indicating MEX-5 as the most distant *B. ambifaria* strain, and best suited outgroup to root the *B. ambifaria*-only phylogeny. The approximate-maximum-likelihood phylogeny was constructed with FastTreeMP using the generalised time reversible substitution model. The evolutionary distance scale bar represents the number of base substitutions per site.



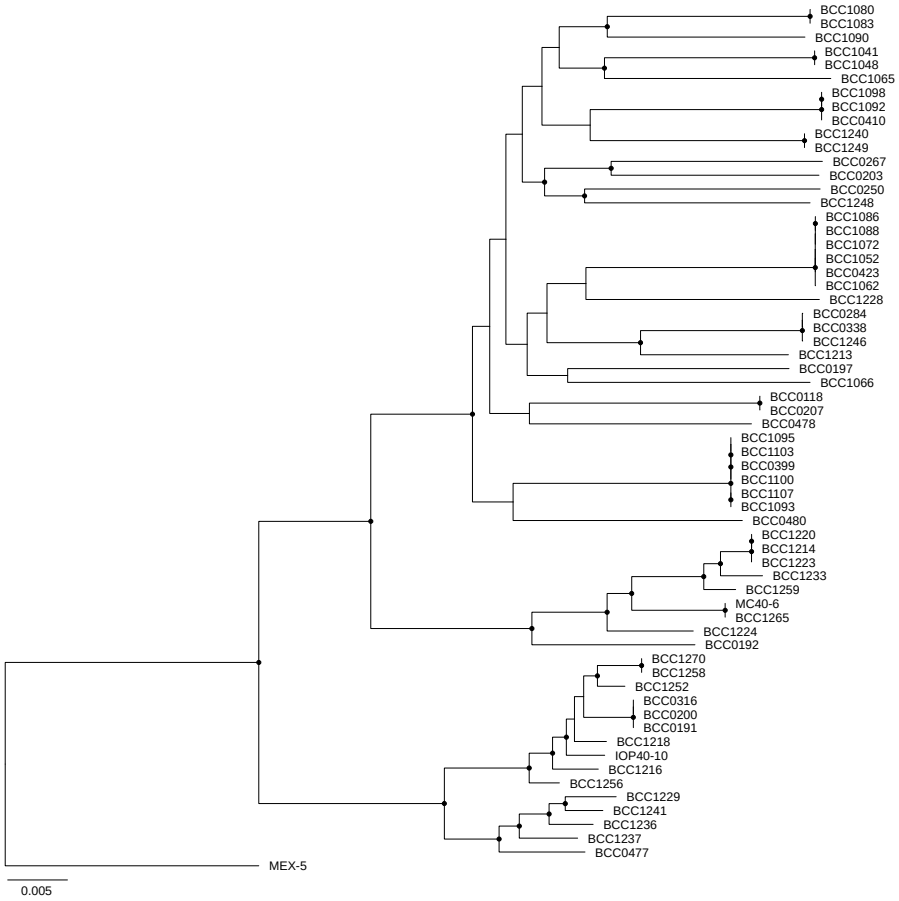
a) Genome-wide core gene phylogeny



b) c1 core gene phylogeny



c) c2 core gene phylogeny



d) c3 core gene phylogeny

Figure 7. Phylogenies constructed from whole-genome and replicon specific core genes.
(a) Whole-genome core-gene phylogeny; (b) replicon c1 core-gene phylogeny; (c) replicon c2 core-gene phylogeny; and (d) replicon c3 core-gene phylogeny. Nodes indicated with black circles represent >75% bootstrap support. Scale bar indicates number of substitutions per base position.

3.2.7. Virulence factor and AMR gene distribution in *B. ambifaria*

Understanding the antimicrobial resistance and virulence potential of biological control microorganisms is a necessity to understand the health risks (plant and animal) associated with their enrichment in the environment (Parke and Gurian-Sherman 2001). Several bioinformatics software exist that predict the presence of the genes associated with these undesirable phenotypes by comparing coding sequences within a genome to a curated database. These curated databases were populated with genes that have experimentally proven AMR or virulence functions. To capture a broader understanding of the AMR and virulence genes associated with *B. ambifaria* as a species, the high throughput screening tool Abricate (https://github.com/tseemann/abricate) was employed.

3.2.7.1. Antimicrobial gene detection and distribution

The 64 *B. ambifaria* genomes were screened against the four AMR databases used by Abricate (CARD, ARG-ANNOT, NCBI and ResFinder) and the outputs combined. A total of 12 genes associated with AMR were predicted following de-duplication. However, the presence of two genes (*amrAB*) associated with a resistance nodulation cell division (RND) efflux pump prompted the manual addition of a *B. ambifaria* homologue of the gene encoding the corresponding outer membrane protein, *orpM*. The addition of the *B. ambifaria oprM* gene resulted in 13 distinct genes associated with *B. ambifaria* antimicrobial resistance. In addition to the core *amrAB-oprM* encoded by all 64 strains, a second core RND efflux pump (64 out of 64) was predicted, *ceoAB-opcM.* Collectively, both efflux pumps provide resistance to aminoglycosides, macrolides, tetracyclines, cethromycin, chloramphenicol, fluoroquinolones and trimethoprim (Podnecky *et al.* 2015). The AMR genes *omp38* and *penA* encoding an outer membrane protein and class A *beta*-lactamase, respectively were the remaining core genes possessed by 100% of *B. ambifaria* strains. The *omp38* gene has been experimentally observed to enhance resistance to Penicillin G, cefoxitin, ceftazidime and imipenem (Aunkham *et al.* 2014), while the *penA* gene provides resistance to *beta*-lactam antibiotics (Randall *et al.* 2015). The remaining five distinct AMR genes, ***qacEΔ1, strA***, *strB, aac(6)-Ib*$_7$ and *sul1* were possessed by a single *B. ambifaria* strain, BCC0478. Aminoglycoside resistance genes *strA* and *strB* encode phosphotransferase modification enzymes, while *aac(6)-Ib*$_7$ encodes an acetyltransferase acetyltransferase (Ramirez and Tolmasky 2010). The *qacEΔ1* gene encodes a small multidrug resistance (SMR) transporter, and *sul1* encodes a dihydropteroate synthase, these proteins enhance resistance to quaternary ammonium compounds and sulfonamide, respectively (Bay *et al.* 2008; Byrne-Bailey *et al.* 2009).

To determine the genomic location of these AMR genes, the Abricate tool (https://github.com/tseemann/abricate) was applied to three *B. ambifaria* genomes (BCC0207, BCC0192 and BCC0191) scaffolded into replicons, in addition to BCC0478 due to its extended antimicrobial resistance gene content. These genomes represent the three major clades observed in the *B. ambifaria* phylogeny. The *amrAB-orpM* RND efflux pump was the only AMR function located on replicon c1. The remaining core AMR genes, *ceoAB-opcM, omp38* and *penA*, were positioned on replicon c2. Antimicrobial resistance genes unique to *B. ambifaria* BCC0478 were located within a 63.7 kbp region toward the edge of a contig scaffolded into replicon c3. Based on plasmid predictions (section 3.2.9.1) there was evidence that the region of BCC0478 replicon c3 possessing these AMR genes was, in fact, an erroneously scaffolded plasmid.

3.2.7.2. Virulence gene detection and distribution

Following an initial screen of all 64 genomes with the default virulence database associated with Abricate (vfdb) it was necessary to customise the database to include *Burkholdria cepacia* complex examples of several virulence factors. Instances where Abricate predicted a fraction of the genes associated with a specific virulence feature through homology with *B. pseudomallei*, such as the genes involved in a T3SS, the *B ambifaria* homologue was identified and added to the database. Any genes associated with the given virulence feature that Abricate failed to identify were also identified in *B. ambifaria* and used to supplement the original database. The customised database was used to re-evaluate the presence and distribution of virulence genes across the 64 genomes. This iterative process was repeated until all partial hits of the original database screen were investigated for an expanded dataset and optimised for *B. ambifaria.*

A set of core virulence genes was apparent with the custom vfdb search and could be assigned to specific virulence systems based on the overarching functions of motility, communication and secretion (Figure 8 and Figure 9). Three major virulence gene loci predicted within the dataset were associated in motility: eleven genes of the *Burkholderia* chemotaxis operon, and 36 genes associated with flagellum formation (Angus *et al.* 2014), 34 of which were present in all *B. ambifaria* screened (*flgL* and *flgM* were identified in 36 and 11 strains, respectively) (Figure 9). Evidence of the type IV pilus biosynthesis operon, *pil* was found (Imam *et al.* 2011), with 13 of the 16 genes encoded by at least 95% of strains screened (61 of 64). The remaining genes of the type IV pilus locus were either sporadically encoded by a subset of strains or clade-restricted in their presence. A single gene, *cdpA*, encoding a cyclic di-GMP phosphodiesterase was also identified as a core virulence gene (Lee *et al.* 2010). The quorum

sensing LuxRI system *bspRI* characterised in *B. pseudomallei* (Kiratisin and Sanmee 2008) was found in all 64 *B. ambifaria* genomes (Figure 9). In addition to this cognate LuxRI system a second quorum sensing related gene, *bpsR2* was predicted in 17 of the 64 strains. The distribution was sporadic through the core-gene phylogeny except for complete coverage of clade 1c. *B. ambifaria*, in contrast to *B. pseudomallei* where the system was originally characterised, lacks the counterpart homoserine lactone synthase-encoding gene, *BpsI2.* Finally, there was evidence of a partial capsular polysaccharide I (cps I) biosynthesis locus in 29 of the 64 *B. ambifaria* genomes, concentrated in clades 1a, clade 1d, and clade 3 (Figure 9). The majority of the 29 strains encoded four genes from the lipid biosynthesis portion of the gene cluster (Cuccui *et al.* 2012).

Multiple secretion systems were predicted within the *B. ambifaria* dataset, most of which were previously described in other *Burkholderia* species but have not been formally identified in *B. ambifaria*. Abricate predicted genes associated with 8 distinct secretion systems, 6 of which were type VI secretions systems (T6SS) (Figure 9). All 64 strains encoded the 17 genes associated with the T6SS-1 (Shalom *et al.* 2007; Spiewak *et al.* 2019); the second most widespread system was T6SS-7 (or T6SSa depending on nomenclature) (Shaloim *et al.* 2007; Spiewak *et al.* 2019) possessed by 55 of the 64 strains. There was no clade specific absence of the T6SS-7 system; *B. ambifaria* strains lacking the system spread throughout the phylogeny. Approximately 28% of *B. ambifaria* strains (18 of 64) possessed the 20 genes that comprise the T6SS-6 system (Shalom *et al.* 2007) with localisation within clades 1d and clade 2, although some of these clade members lacked the secretion system. The T6SS-3 (Shalom *et al.* 2007) (14 of the 19 genes) was present in 14 of the 64 strains, and localised to a monophyletic sub-clade of clade 3, as well as several members of clade 1b and strain MEX-5 (Figure 9). The T6SS-2 was predicted in 10 strains concentrated within a monophyletic sub-clade of clade 3 and a paraphyletic distribution through clade 2. There was limited evidence for a T3SS (Vander Broek and Stevens 2017) in *B. ambifaria* strain BCC1248 possessing genes *prgH* and *prgI*, and strain IOP40-10 possessing *spaP*. The final two secretion systems that the custom vfdb database predicted were the T6SS-5 (18 genes) and T3SS-3 (6 genes) (Shalom *et al.* 2007; Vander Broek and Stevens 2017), both of which were only found within strain BCC1248.

The locations of the multiple virulence factors were determined by screening representative *B. ambifaria* scaffolded genomes, with the addition of a hybrid genome assembly of strain BCC1248. The hybrid assembly provided further genomic architecture to identify virulence factors on extra-chromosomal DNA (plasmids) as opposed to scaffolded and unscaffolded contigs. Due to the large number and varied phylogenetic distribution of virulence factors. The

Abricate analysis using the vfdb database identified genes related to 14 distinct virulence functions, the majority of which were possessed by the largest replicon, c1. Two of the secretion systems, T6SS-1 and T6SS-2, in addition to genes responsible for chemotaxis, flagellum biosynthesis, capsule formation, and the signalling molecule, cyclic diguanylic acid, were found on replicon c1. The quorum sensing LuxRI genes *bpsI1-bpsR1* and secretions systems T6SS-6 and T6SS-7 were located on replicon c2. The smallest replicon c3 possessed the fewest vfdb database-predicted virulence features: the orphan *bpsR2* quorum sensing regulator and type IV pilus biosynthesis genes. Finally, the BCC1248-specific T6SS-3 and T6SS-5 were revealed to be plasmid-borne features contained on one of the largest plasmids (286 kbp) identified in any *B. ambifaria* genome. All core virulence loci (100% of strains): *bpsR2-bpsI2, cdpA*, chemotaxis operon, flagellum biosynthesis operon, and T6SS-1 were found on replicons c1 and c2; however, one of the widespread loci (61 out of 64 strains), type IV pilus biosynthesis, was located on replicon c3.

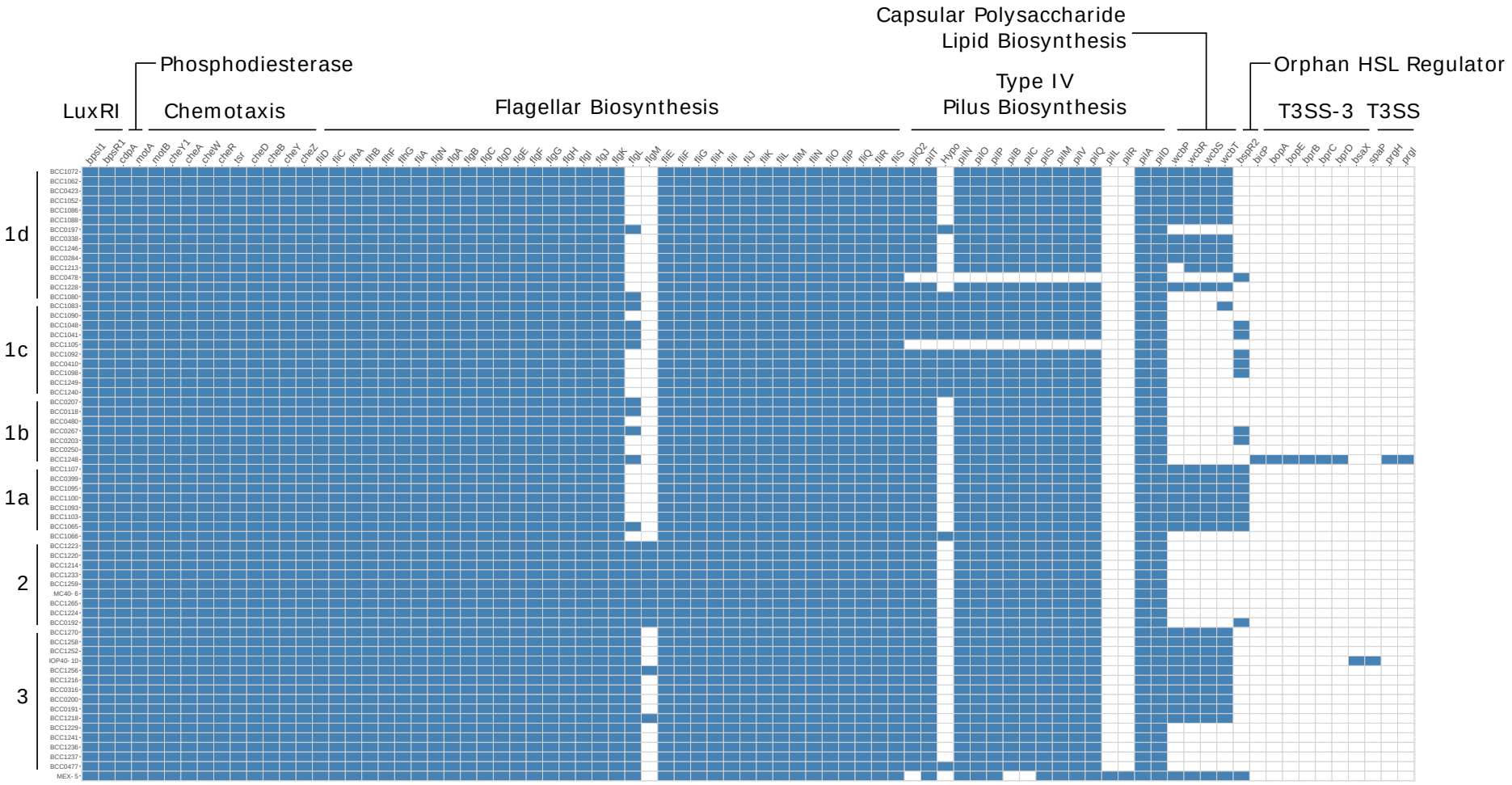


Figure 8. Presence and distribution of virulence factors across *B. ambifaria*.
Chemotaxis, flagellar biosynthesis, type IV pilus biosynthesis, cdpA phosphodiesterase, and a *luxRI* system were core virulence to all 64 *B. ambifaria* strains. Genes involved in the lipid biosynthesis of capsule production, the T3SS-3 and orphan homoserine lactone regulator bspR2 were accessory virulence factors.

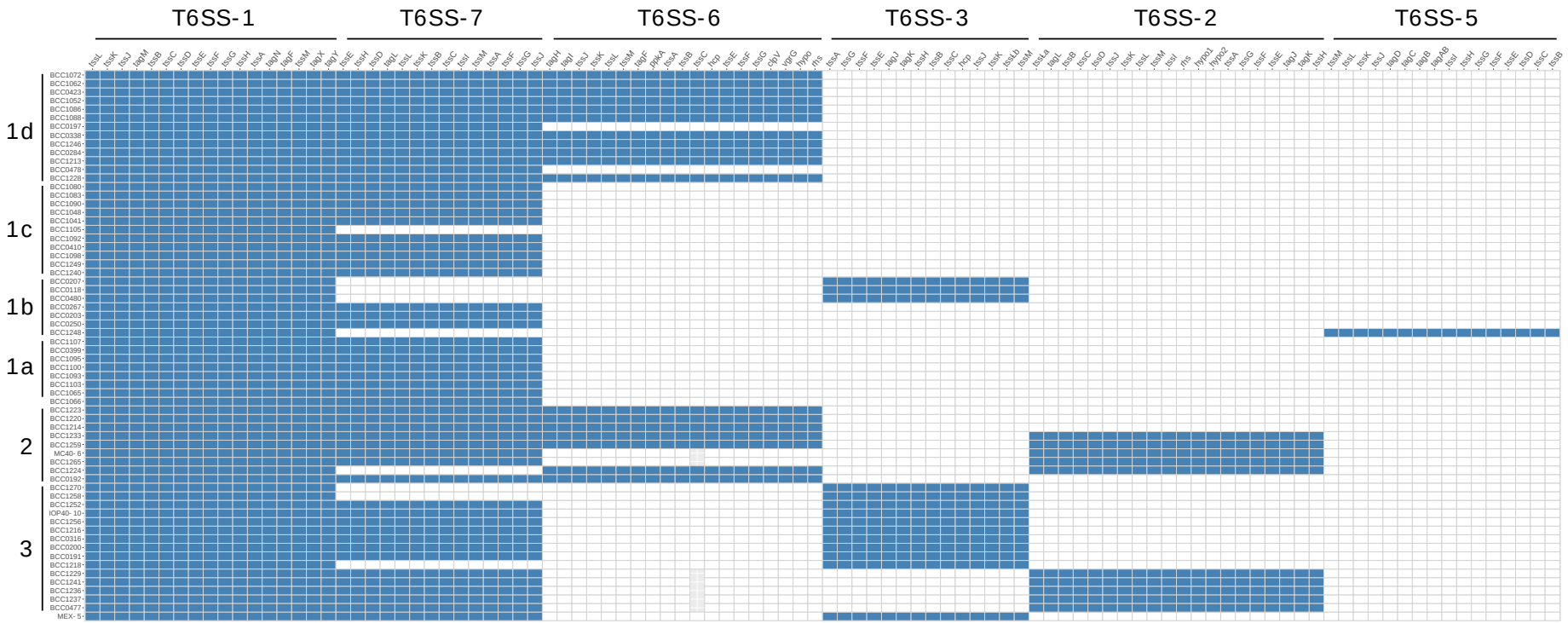


Figure 9. Presence and distribution of type VI secretion systems (T6SS) across *B. ambifaria*.
The use of Abricate supplemented with *B. ambifaria* gene homologues identified six T6SS across the 64 *B. ambifaria* genomes. *B. ambifaria* MC40-6 possessed a plasmid-borne duplicate of the T6SS-2 system. Individual genes for each secretion system are indicated, and the y-axis of strains was aligned against the core-gene phylogeny with the major clades highlighted.

3.2.8. MinION run and hybrid assembly statistics

Six *B. ambifaria* strains were selected for long-read Oxford Nanopore Technology (ONT) sequencing; which included strains that represented the *B. ambifaria* species diversity and strains with evidence of uncharacterised plasmids of interest. *B. ambifaria* BCC0477, BCC1086 and BCC1224 were chosen for long-read sequencing to provide completed genomes in other clades of the *B. ambifaria* core-gene phylogeny in addition to the existing completed genomes BCC0203, BCC0207 and MC40-6 (BCC1212). These strains also represented potential genomic diversity in their un-scaffolded contigs: 380 Mbp, 199 MB and 85 Mbp in BCC0477, BCC1086 and BCC1224, respectively. The final *B. ambifaria* strain sequenced by Oxford Nanopore technology was the model biological control strain BCC0191. The Oxford Nanopore software MinKNOW called 8.33 GB of bases consisting of 2.22 million reads. Reads were divided into two groups based on their average base quality scores: reads with average quality scores greater than Q7 (pass) and reads with average quality scores less than Q7 (Fail). Approximately 1.66 million reads passed this quality score threshold, with a mean read length of 4,203 bp, a maximum length of 403,370 bp, and a read length N50 equal to 6,864 bp (Figure 10a). The number of reads recorded every 10 minutes increased sharply over the first two hours, reaching a peak of approximately 16,000 reads per 10 minutes. Following this maximum throughput there was a steady decline towards zero throughput at the final time point or 48 hours (Figure 10b). The read quality increased slightly over the first 15 hours followed by a gradual decrease up to 42 hours, and finally a more pronounced decrease up to the 48 hour time point (Figure 10c).

A comparison of the total genome size between the Illumina and hybrid assemblies revealed that four out of six genomes were larger in the hybrid assembly (Table 9). The hybrid assemblies produced replicons (c1, c2 and c3) larger than the scaffolded contigs from the Illumina-only assemblies for all six strains, with the exception of BCC0478 c3. The third replicon of BCC0478 was larger in the scaffolded Illumina assembly due to the erroneous assembly of the plasmid sequence into a c3-borne contig (Table 10). A similar number of CDS were predicted for both Illumina-only and hybrid assemblies, except for BCC1086 and BCC1248 whose Illumina-only assemblies varied by over 100 CDS compared to their respective hybrid genome assemblies. The hybrid assembly outputs yielded six complete rRNA operons (18 genes per genome) in contrast to the Illumina-only assemblies that output between 3-8 genes covering partial rRNA operons (Table 9).

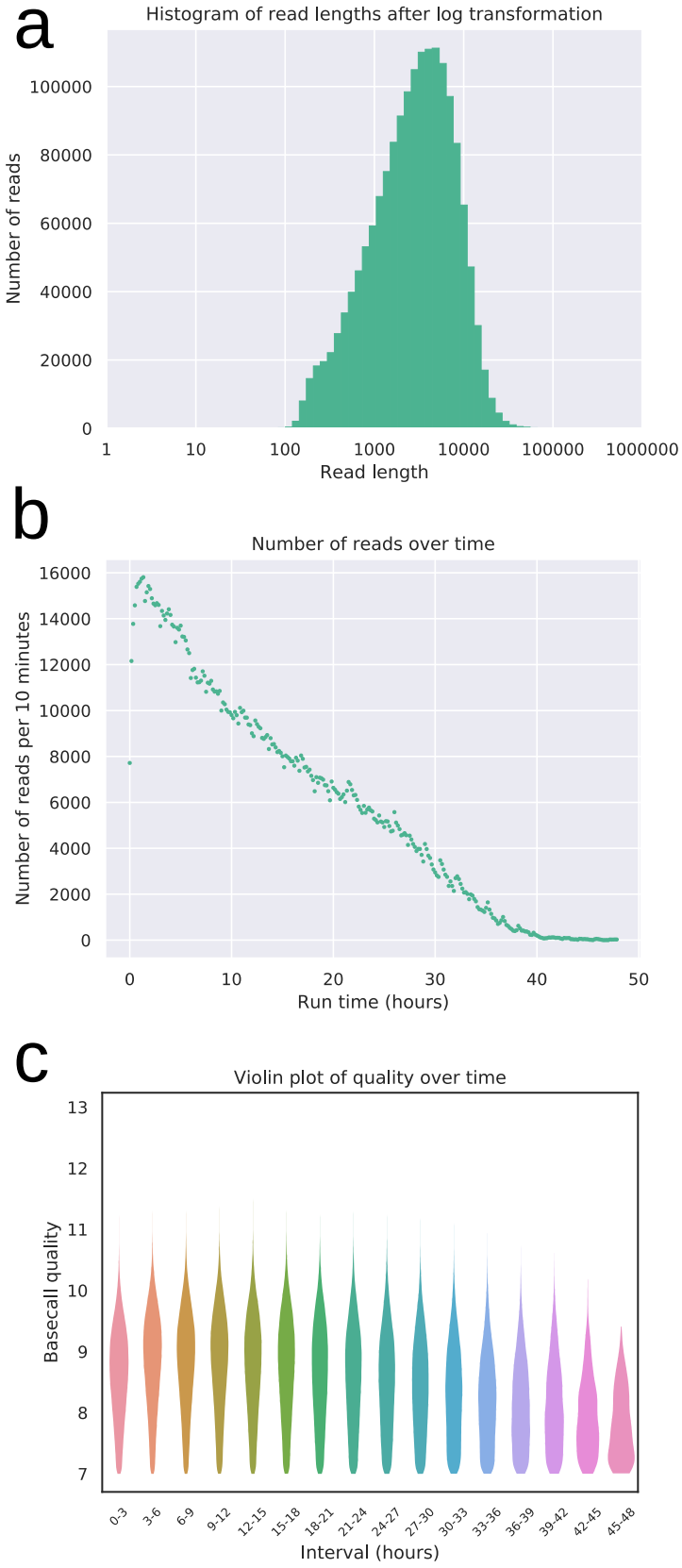


Figure 10. MinION run performance statistics based on passed reads.
(a) Histogram of number of reads vs read length. (b) Number of reads recorded every 10 minutes over time. (c) Violin plot of base call quality over time.

Table 9. Comparison of Illumina-only and hybrid assembly genome statistics.

	Illumina					Hybrid				
Strain	**Contigs**	**Total length (Mbp)**	**Predicted CDS**	**Predicted Genes**	**Predicted rRNA genes**	**Contigs**	**Total length (Mbp)**	**Predicted CDS**	**Predicted Genes**	**Predicted rRNA genes**
BCC0191	97	7.58	6633	6712	6	3	7.62	6633	6729	18
BCC0477	118	7.81	6881	6961	5	4	7.86	6895	6993	18
BCC0478	89	7.24	6338	6421	8	4	7.25	6334	6432	18
BCC1086	302	7.60	6652	6727	8	4	7.46	6542	6634	18
BCC1224	152	7.27	6480	6554	8	3	7.24	6448	6540	18
BCC1248	273	8.03	7083	7169	3	10[a]	8.23	7259	7369	18

[a]The hybrid assembly failed to resolve a circular replicon for *B. ambifaria* BCC1248 c2, but a linear single-path through four nodes was observed on the

Table 10. Comparison of scaffolded and assembled replicons and plasmids between Illumina-only and hybrid assemblies.

	Illumina			Hybrid					
	Replicon Length (Mbp)			**Replicon Length (Mbp)**			**Plasmid Length (kbp)**		
Strain	**c1**	**c2**	**c3**	**c1**	**c2**	**c3**	**p1**	**p2**	**p3**
BCC0191	3.52	2.93	1.09	3.56	2.95	1.11	/	/	/
BCC0477	3.44	2.88	1.12	3.48	2.89	1.13	364.4	/	/
BCC0478	3.36	2.72	1.13	3.38	2.72	1.09	62.0	/	/
BCC1086	3.55	2.69	1.16	3.58	2.71	1.17	3.3	/	/
BCC1224	3.46	2.66	1.07	3.50	2.67	1.08	/	/	/
BCC1248	3.55	2.72	1.23	3.58	2.93	1.31[a]	286.1	125.6	3.5

[a]The sequence lengths associated with the four nodes were combined to provide a replicon size.

3.2.9. Prevalence of plasmids in *B. ambifaria*

All three complete *B. ambifaria* genomes (BCC0203, MC40-6 and AMMD) are known to harbour plasmids. The prevalence, distribution, diversity and content of plasmids within the expanded *B. ambifaria* genome collection was therefore investigated. Plasmid contig assembly and identification was attempted for the 59 Illumina-sequenced strains (this excluded AMMD, BCC0203, IOP40-10, MC40-6 and MEX-5), of which, contigs were successfully detected for 25 strains. The un-scaffolded contigs for *B. ambifaria* strains MEX-5 and IOP40-10, and complete plasmid sequences for AMMD, BCC0203 and MC40-6, were included at this stage for further analyses. The un-scaffolded contigs of *B. ambifaria* BCC1086 were also included due to known contamination of the sequenced sample with *Serratia* DNA, resulting in the erroneous identification of *Serratia* derived contigs as contigs of plasmid origin.

3.2.9.1. Detecting evidence of plasmids in draft *B. ambifaria* genomes

Candidate *parAB* gene homologues were extracted from 13 of the 25 strains with plasmidSPAdes assembled contigs. The plasmidSPAdes assembled contigs were queried against their respective scaffolded genomes and unscaffolded contigs by creating local blastn databases. This identified plasmidSPAdes assembled contigs that were potentially chromosomal DNA. Strains BCC0267 and BCC0478 encoded *parAB* homologues in contigs with high nucleotide similarity to their respective genome assemblies. The BCC0267 *parA* gene clustered with the c3 *parA* genes, and was consequently discounted as a plasmid. The BCC0478 *parA* gene was separate from c1, c2 and c3 derived *parA* sequences, despite the encoding contig possessing high nucleotide homology to the BCC0478 c3 replicon. A review of the c3 scaffolded contigs highlighted an approximately 70 kbp region with no homology to the complete BCC0207 c3 reference; a similarly sized region to the proposed 62 kbp plasmid contig. This suggests an erroneously assembled genomic contig that was scaffolded into the BCC0478 c3 replicon.

Only *parA* genes with an adjacent *parB* counterpart, and *vice versa*, were extracted for phylogenetic analysis (Figure 11). Candidate homologues in MEX-5 and IOP40-0 were an exception where all identified candidate genes were extracted for further analysis due to the highly fragmented genomes and potential for contig splits between these adjacent genes. Unpaired *parA* and *parB* genes were subject to blastn searches against the nucleotide collection database. MEX-5 and IOP40-10 un-scaffolded contigs possessed one unpaired *parA* candidate, both of which possessed homology to *Burkholderia* chromosomal genes. MEX-5 and IOP40-10 un-scaffolded contigs encoded two *parB* candidates, one IOP40-10 *parB* was homologous to chromosomal genes and the second possessed homology to one

complete plasmid sequence which lacked a paired *parAB*. One MEX-5 *parB* candidate lacked high homology to any gene in the database, and the second gene possessed homology to a *Burkholderia cenocepacia* 842 plasmid pBcn842-1 gene. The candidate *parB* genes were extracted from the pBcn842-1 plasmid and incorporated into the *parB* gene phylogeny. Candidate MEX-5 *parB* gene clustered with the confirmed false positive *parB* in pBcn842-1. A confirmed false positive was defined as a candidate *parA* or *parB* gene that lacked the respective *par* gene, and occurred in a plasmid or replicon that possessed paired *parAB* genes.

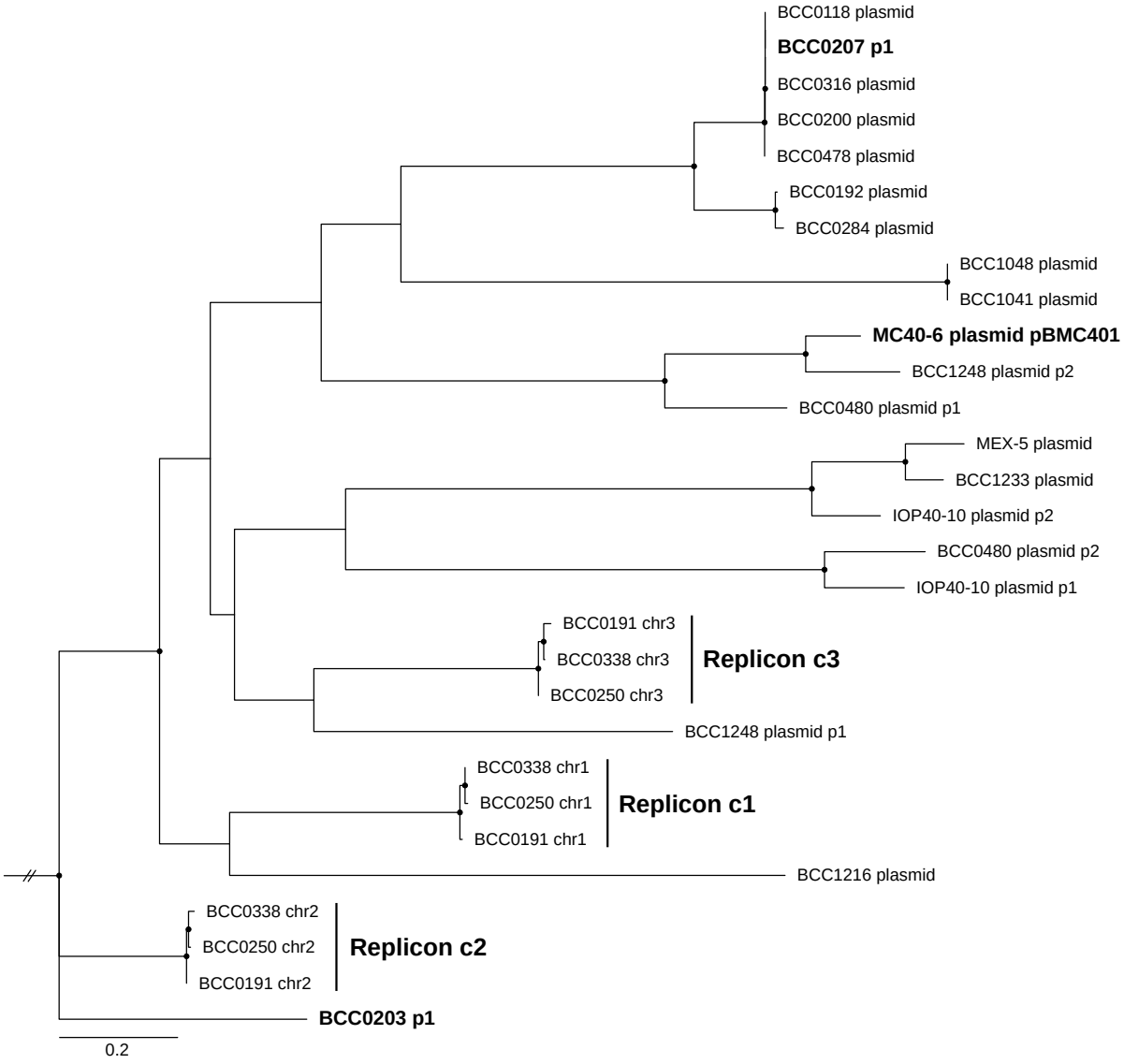


Figure 11. Evolutionary history of *B. ambifaria* plasmids and replicons.
The *parB* gene was extracted from 20 putative plasmids derived from 17 *B. ambifaria* strains. Characterised *B. ambifaria* plasmids are indicated in bold and representative c1, c2 and c3 replicon *parB* genes are highlighted. Nodes with bootstrap values >75% are highlighted with black circles. Scale bar represents number of substitutions per base position.

3.2.10. Hybrid assembly of AMR and virulence plasmids

Of the six *B. ambifaria* strain genomes completed by long-read sequencing, six plasmids were identified, varying between 3.3 to 364 kbp in length. Two plasmids were chosen for further analysis: BCC0478 p1 due to evidence of multiple AMR genes, and BCC1248 p1 due to evidence of virulence gene presence. Similar to *B. ambifaria* BCC1248 c3, the hybrid assembly failed to resolve a circular replicon for one of the BCC1248 plasmids (p2), but a circular, single path connecting four nodes (125,232 bp, 159 bp, 151 bp and 85 bp) was observed on the Bandage graph (data not shown). Long read mapping identified a 39.6 kbp read that spanned the smaller nodes and the end regions of the largest node. This evidence was used to justify merging these nodes into a single contig to represent the completed plasmid sequence.

3.2.10.1. *B. ambifaria* BCC0478 p1 AMR plasmid and similar NCBI plasmids

The *parB* gene was predicted and extracted from *B. ambifaria* BCC0478 p1, and similar *parB* genes identified via BLAST against the public nucleotide collection (nr/nt) with a threshold of 85% sequence identity and 90% alignment coverage. Removal of partial and synthetic sequences from the BLAST search resulted in 69 accessions that met these criteria. The 69 sequences were downloaded and their *parB* gene sequences extracted to generate a phylogeny (Figure 12). The *parB* phylogeny formed two major clades representing incompatibility groups IncP-1β1 and IncP-1β2 (Sen *et al.* 2013), with BCC0478 p1 falling into the former incompatibility group. Of the 70 IncP-1β representatives, including BCC0478 p1, only 25 plasmids possessed AMR genes contained within the CARD and NCBI databases. These plasmids varied in their AMR-carrying capacity from a solitary AMR gene in *Proteus mirabilis* plasmid R772 (KF743817.1) to the multi-AMR gene-carrying plasmid pKPC-038c in *Aeromonas* sp. ASNIH1 (CP026230.1) that possessed eleven of such genes. There was also evidence of gene duplication or multiple acquisitions of the same gene on several plasmids based on the Abricate gene predictions. *B. ambifaria* BCC0478 p1 carried five distinct AMR genes distributed across two regions: *strA* and *strB* at one site, and *sul1*, *qacEΔ1*, and *aac(6)-Ib7* at the second site.

3.2.10.2. Identifying previously unknown *Burkholderia* IncP-1β plasmids

In addition to the *B. ambifaria* BCC0478 p1 plasmid the BLAST search identified AMMD p1 plasmid as a member of the IncP-1β1 clade. This prompted the question of how many additional plasmids of the same incompatibility family resided within the publically available *Burkholderia* sequence data but remained, as of yet, undetected. To address this query all available paired-end Illumina reads labelled as *Burkholderia* were downloaded and assembled into draft genomes, and all existing *Burkholderia* genome assemblies were also downloaded.

A screen of these genome collections, representing 4,479 and 3,001 assemblies, respectively, via BLAST produced multiple alignments to BCC0478 p1. Based on a minimum threshold of 5 kbp high-scoring-pair alignment 41 assemblies, excluding BCC0478, from the locally assembled genomes and six from the downloaded assemblies were candidates for carrying IncP-1 clade plasmids. Combining and de-duplicating the outputs from both genome collections produced 45 candidate assemblies. The contig with a minimum 5 Kbp alignment to BCC0478 p1 were extracted from these assemblies and screened for the *parB* gene. We detected 27 *parB* genes in 24 contigs; 23 of these *parB* genes clustered within the IncP-1β1 clade, representing five species: *B. ambifaria*, *B. vietnamiensis*, *B. cenocepacia*, *B. cepacia* and *B. gladioli* (Figure 13).

The *parB* gene extracted from *B. vietnamiensis* Bp7801 grouped phylogenetically with the IncP-1α pBS228 reference plasmid from *P. aeruginosa* (Figure 13). A screen for AMR genes using the CARD and NCBI databases through Abricate detected candidate genes in 9 of the 45 contigs, three of which were potential IncP-1β1 plasmids, and one represented the potential IncP-1α plasmid. The remaining AMR-encoding contigs either possessed non-IncP-1 *parB* genes and therefore excluded from the phylogeny, or lacked a detectable *parB* homologues altogether. A custom database was used in Abricate to screen for the presence of the seven-gene mercury resistance operon *merRTPFADE* detected in BCC0478 p1. Only six of the *mer* genes, *merRTPADE*, were predicted in 18 of the 45 contigs, while *B. gladioli* BCC1809 possessed five of these genes. Interestingly, the locally assembled Illumina-only version of BCC0478 also lacked the *merF* gene, when compared to the hybrid assembly. The co-occurrence of AMR genes and the *mer* operon was found in four genomic contigs, in addition to BCC0478 p1. The IncP-1β1 plasmid-carrying *Burkholderia* were *B. cenocepacia* BCC0506, *B. cenocepacia* BCC0048, and *B. gladioli* BCC1763, and these contigs encoded six, one and four AMR genes, respectively. The IncP-1α plasmid carried by *B. vietnamiensis* Bp7801 possessed four AMR genes, but lacked the *strAB* genes associated with the AMR 1β1 plasmids.

3.2.10.3. *B. ambifaria* BCC1248 p1 virulence plasmid and similar plasmids

In comparison to BCC0478 p1, the BLAST search of the *B. ambifaria* BCC1248 p1 *parB* gene against the public nucleotide collection (nr/nt) produced no results. To identify similar *parB* genes, the default BLAST algorithm megablast was replaced with the less stringent blastn algorithm. The use of blastn reported 47 alignments, the 24 highest scoring alignments were of *Burkholderia* origin and represented both formerly identified plasmids and chromosomal DNA. A *parB* gene phylogeny revealed the *Burkholderia* sp. KK1 plasmid pkk3 as an outlier of the clade. To address this, the pkk3 plasmid *parB* gene was searched against the nr/nt database using blastn with the aim of identifying other closely related sequences. The search returned 21 bacterial sequences which, following de-duplication with the existing sequences, identified a *Paraburkholderia* sp. 7MH5 plasmid. Additional members of the plasmid family were identified in the Illumina-only locally assembled genome collection and downloaded ENA assemblies by a pairwise k-mer matching approach. This method identified two extra members of the plasmid family based on *parB* gene sequence: *B. multivorans* BCC0031 and *B. cepacia* Bp7551 (Figure 14). A virulence gene screen of the *B. ambifaria* BCC1248 p1 plasmid confirmed the plasmid-borne nature of the T6SS and T3SS predicted by the initial virulence factor analysis (section 3.2.7.2). Secretion systems were not detected on any of the remaining 16 plasmids. The BCC1248 p1 plasmid possessed a considerable number of insertion sequence transposases and integrases representing 117 of the 287 coding sequences. Gene annotations combined with protein domain searches identified putative collagenase and thermolabile haemolysin-encoding genes. A comparison of genes between BCC1248 p1 and the nearest plasmid based on the *parB* phylogeny, *B. glumae* BGR1 plasmid bglu_3p, revealed a distinct lack of shared gene content. Only four genes were common to both plasmids, which included transposase, integrase and IstB ATP-binding proteins, highlighting the unique nature and contents of the BCC1248 p1 plasmid.

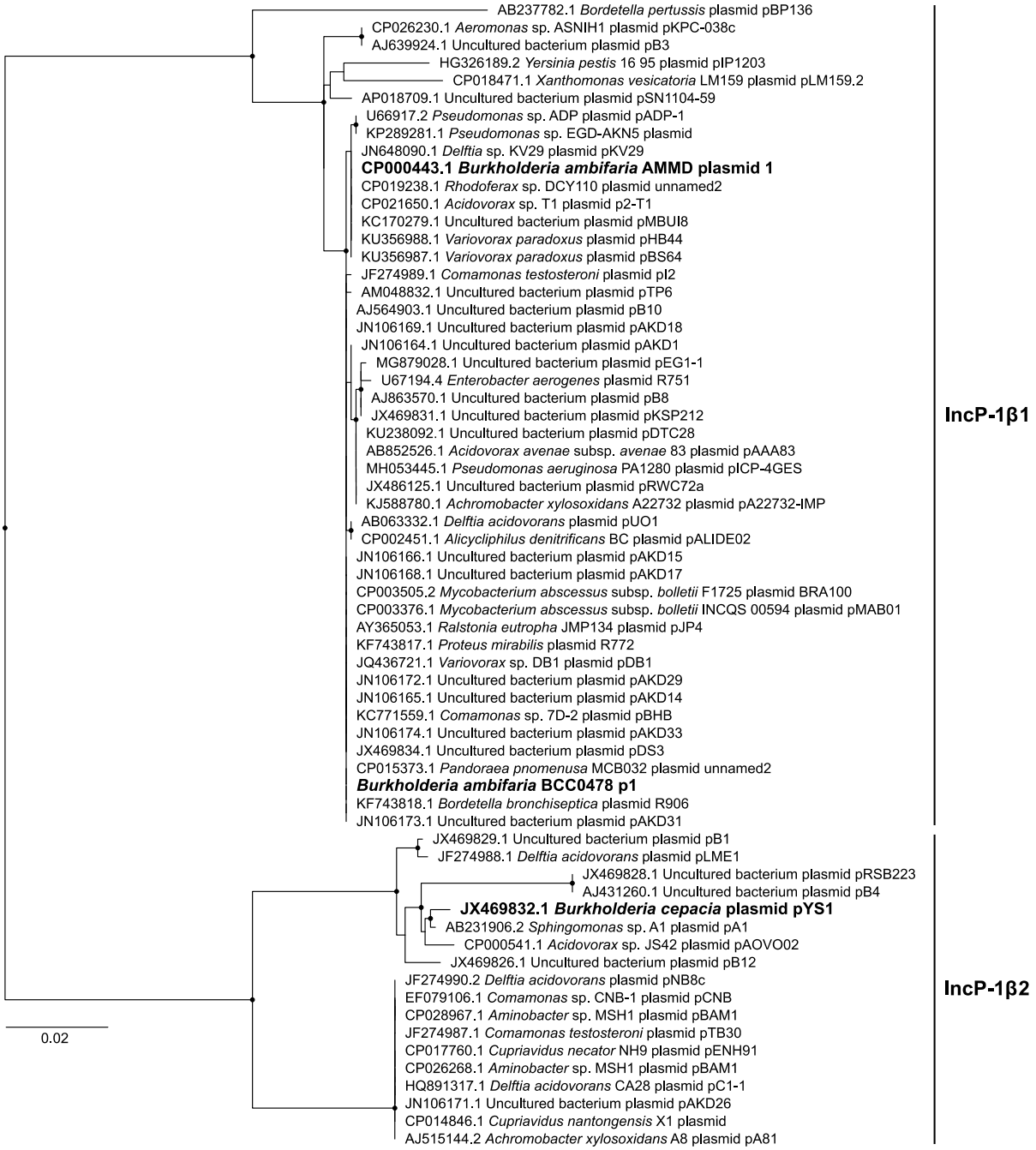


Figure 12. Phylogeny of NCBI-sourced *parB* genes with similarity to *B. ambifaria* BCC0478 p1. A non-exhaustive collection of *parB* genes with similarity to *B. ambifaria* BCC0478 p1 highlighted the phylogenetic split between IncP-1β1 and IncP-1β2 clades. *Burkholderia*-derived *parB* genes are indicated in bold font. Nodes with bootstrap values >60% are highlighted with black circles. Scale bar represents number of substitutions per base position.

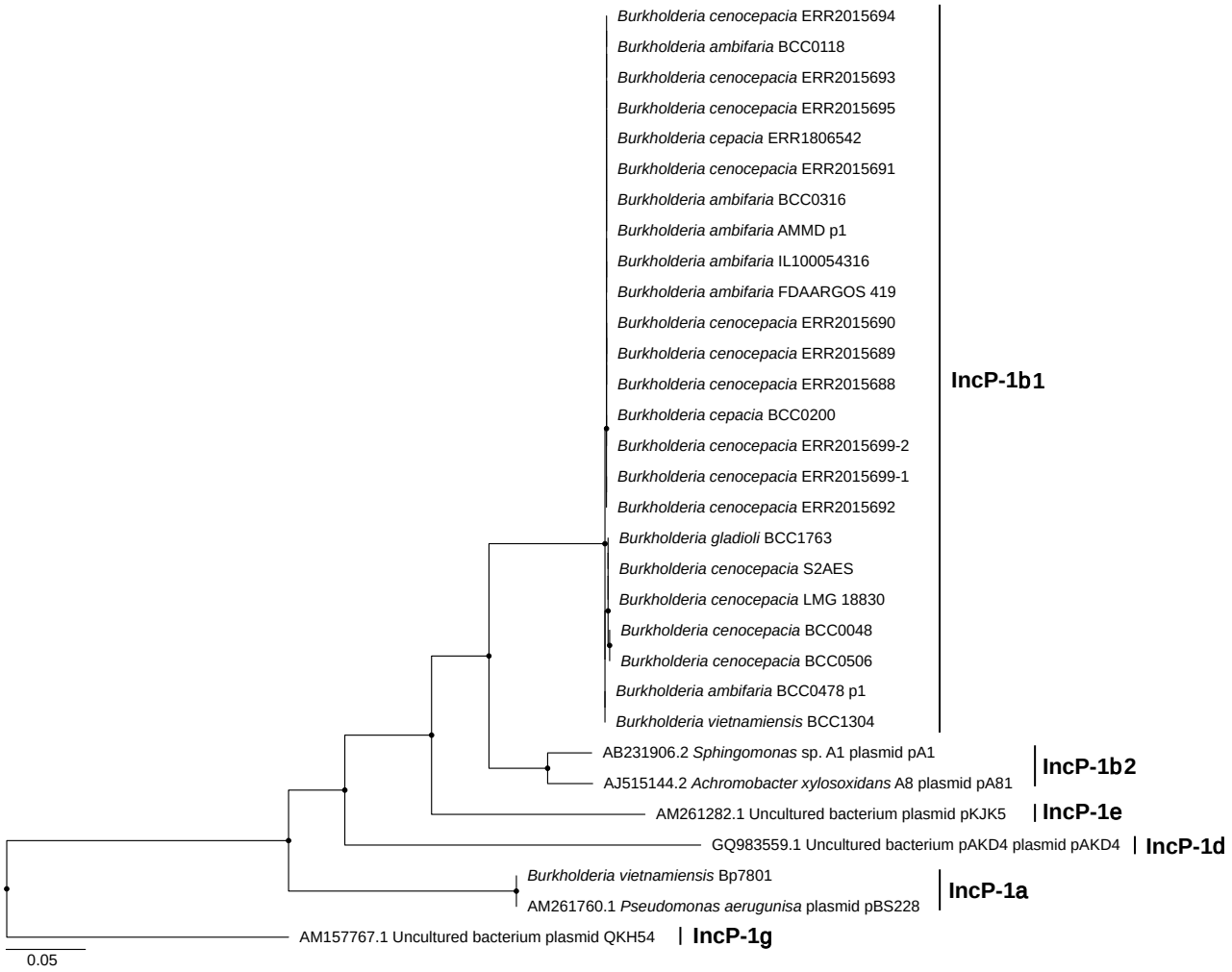


Figure 13. Phylogeny of incompatibility group IncP-1 plasmid *parB* genes.
Multiple previously unknown *Burkholderia* IncP-1 plasmids were identified based on the reference *B. ambifaria* BCC0478 p1 plasmid. Representatives of neighbouring IncP-1 sub-groups are included to provide context for the *Burkholderia* plasmids. Nodes with bootstrap values >60% are highlighted with black circles. Scale bar represents number of substitutions per base position.

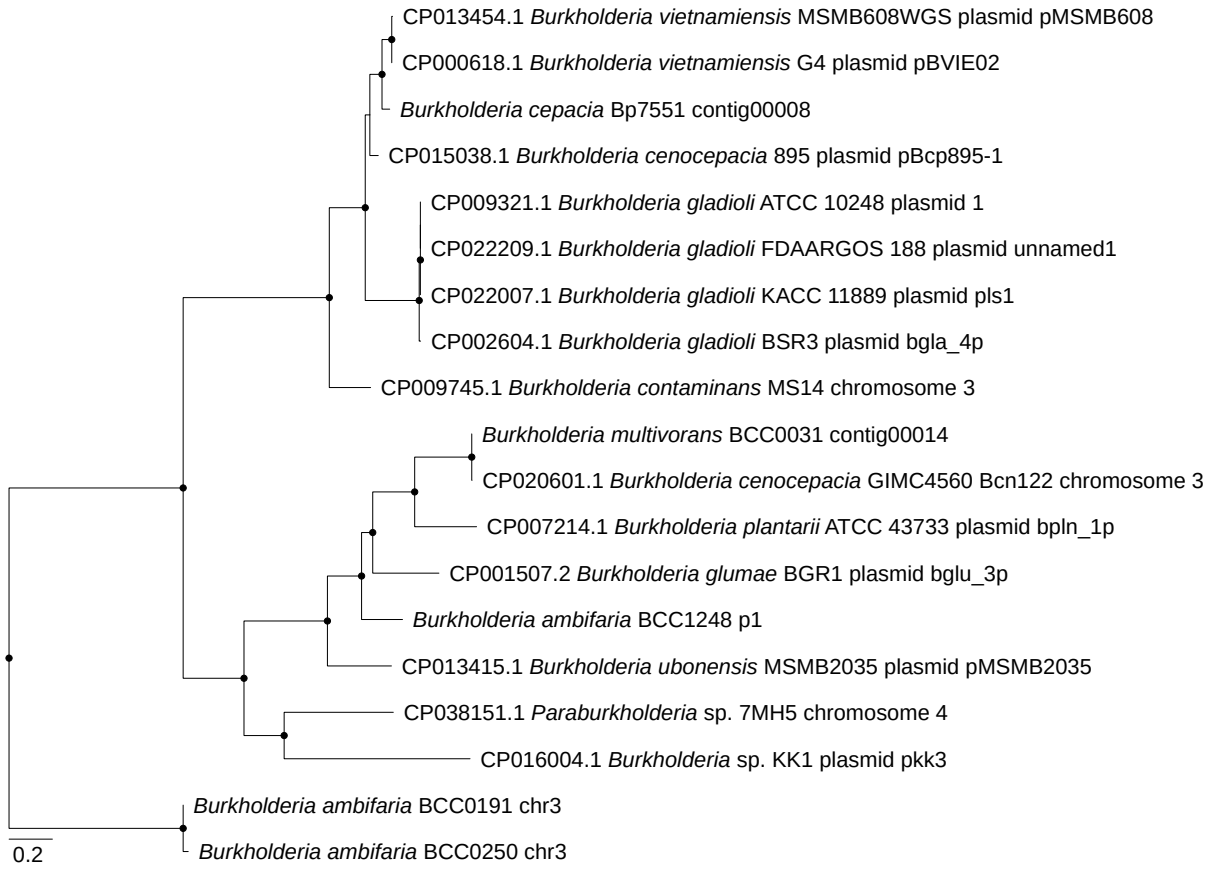


Figure 14. Phylogeny of *Burkholderia* and *Paraburkholderia* restricted plasmid *parB* genes related to BCC1248 plasmid p1.
A search of NCBI-deposited genomes and locally assembled *Burkholderia* genomes identified 16 related *parB* sequences from previously identified plasmids and chromosomal sequences. Two major clades were identified, with the *B. ambifaria* BCC1248 p1 plasmid branching within the second clade. The phylogeny was rooted with replicon c3-derived *parB* sequences. Nodes with bootstrap values >75% are highlighted with black circles. Scale bar represents number of substitutions per base position.

3.3. Discussion

Using a combination of short and long-read sequencing technologies a collection of 61 *B. ambifaria* genomes was assembled and, combined with public database genomes, provided the foundation for genomic analyses in this study. Average nucleotide identity enabled *B. ambifaria* species confirmation from genomic sequences, and genomes representing different species to be excluded from further analysis. Population biology was interrogated through high-resolution core-gene phylogenomics, which identified three major evolutionary clades supported by ANI. The core and accessory genome of *B. ambifaria* was evaluated, and AMR and virulence factors predicted across all genomes. Finally, long-read ONT sequencing supported the closure of six *B. ambifaria* genomes providing an insight into the plasmid diversity of the species and their contribution to AMR and virulence.

3.3.1. Flexible virulence factor profile observed in *B. ambifaria*

The use of vfdb via Abricate for the prediction of virulence factors in *B. ambifaria* highlighted the dominance of *B. pseudomallei* virulence factors, and the absence of *B. cepacia* complex gene representatives. The lack of database genes possessing high homology to *B. ambifaria* resulted in fragmented prediction of virulence loci due to low percentage sequence similarity and coverage. This required the manual curation of the database with *B. ambifaria* representative genes based on virulence locus prediction through partial matches to *B. pseudomallei*. Compared to the predictions of AMR genes there was considerably diversity in the occurrence of virulence genes. Interestingly, most virulence loci were found on replicons c1 and c2, and these represented both core and accessory virulence loci. In comparison to c3 which encoded two widespread loci with 86% and 95% occurrence.

The smallest replicon c3 was exposed as a large virulence (Agnoli *et al.* 2012) and stress tolerance encoding (Agnoli *et al.* 2014) plasmid, and therefore not essential to the viability of *B. cepacia* complex species. This research would imply that the c3 replicon possessed a larger share of the virulence loci. However, detailed analysis of *B. ambifaria* demonstrated that most virulence loci predicted *in silico* occurred on the essential replicons c1 and c2. Species-to-species variation in virulence was observed in *G. mellonella* when the wild-type was compared to the third-replicon knockout mutant (Agnoli *et al.* 2012). This supports the *intra*-species disparity in the distribution of *B. ambifaria* virulence factors; suggesting that complete virulence in a given pathogenicity model might not be completely attenuated by third replicon knockouts due to the variation in virulence loci on the remaining replicons.

The genomic regions of replicon c3 responsible for virulence have been investigated in *B. cenocepacia* H111 and attributed to the lipopetide *afc* cluster, LysR-type transcriptional regulator *shvR*, nematocidal toxin-encoding gene *aidA*, and metalloprotease *zmpA zmpA* (Agnoli *et al.* 2017). These virulence factors were absent in the default vfdb database. The secretion system T6SS-1 was initially characterised as a virulence factor in eukaryotic models, contributing to persistence *in vitro* (Hunt *et al.* 2004). However, more recent research has shown the secretion system to be widespread in *Burkholderia* and important in bacterial competition rather than virulence (Spiewak *et al.* 2019). Fundamental differences exist between the two studies in terms of the infection model applied, namely the use of chronic rat lung infection models in Hunt *et al.* (2004), whereas zebrafish, *G. mellonella,* and C*. elegans* were assessed in Spiewak et al (2019).

There is evidence of virulence associated with type IV pilus formation in *B. pseudomallei* (Allwood *et al.* 2011) and flagellum biosynthesis, although the presence of flagellum biosynthesis genes, chemotaxis and type IV pilus biosynthesis genes may not be exclusively involved in virulence and may contribute to colonisation and interactions with plant roots. A comparison of the rhizo-competence between BCC0191 and BCC0191Δc3 revealed a significant difference in the recoverable *Burkholderia* colony forming units per cm of root (see Section 6.2.9). The reduced root colonisation or persistence of the third replicon knockout could be partially attributable to the deletion of the widespread type IV pilus. Indeed, type IV pili have been suggested as the means of adhesion for the beneficial *Paraburkholderia phytofirmans* PsJN (Mitter *et al.* 2013).

3.3.2. Stability of the *B. ambifaria* AMR gene profile

With the exception of the broad host range AMR plasmid identified in strain BCC0478, all AMR genes identified on the replicons c1 and c2 were detected in all 64 *B. ambifaria* strains. Six of these AMR genes were responsible for the synthesis of two intrinsic efflux pumps, *amrAB-oprM* and *ceoAB-opcM* conferring elevated resistance against a range of antimicrobials (Podnecky *et al.* 2015), and represent a genetically stable AMR feature of the species. Most of the *B. ambifaria* genomes analysed were isolated from environmental sources, potentially explaining the lack of AMR gene variation due to an absence of selection. In contrast, of the six clinically sourced *B. ambifaria* strains, one of them possessed the AMR gene-carrying plasmid. This has implications regarding *B. ambifaria*, and more broadly *Burkholderia,* infections and consequent antibiotic treatment strategies. The presence of this plasmid in a clinical isolate potentially represents a survival adaptation and environmental selection. Other CF pathogens, such as *Pseudomonas aeruginosa*, possess greater AMR profile variation

(Freschi *et al.* 2019). However, this is likely due to a bias in the available genomic assemblies, which is considerably higher for *P. aeruginosa*, and the greater virulence of *P. aeruginosa* resulting in a higher number of clinical manifestations compared to *B. ambifaria*.

3.3.3. Hybrid genome assembly-mediated analyses

Illumina-sequenced genomes have proven instrumental in discovering the genomic diversity of the naturally beneficial soil organism *B. ambifaria*. Combining these Illumina-assemblies with reference sequences has provided an insight into the variation in gene content and distribution within the context of the three-replicon genomes. However, the extent of this variation has prevented the complete elucidation of the genomic architecture and existence of extrachromosomal elements resulting in multiple unscaffolded contigs, ranging in size from several kbps to over 350 kbps. The long reads afforded by MinION sequencing has provided the necessary scaffolding to complete a selection of *B. ambifaria* genomes. While the hybrid-assemblies did not differ considerably in the predicted gene content for most strains compared to Illumina-only assemblies, the long read-mediated assemblies were pivotal in our confirmation and extraction of complete plasmid sequences – two of which were shown to carry AMR genes and virulence genes. The extended antimicrobial resistance and virulence gene profiles of *B. ambifaria* BCC0478 and BCC1248, respectively, coupled to evidence provided by plasmidSPADES assemblies that these additional genes were plasmid-borne earmarked these strains for MinION sequencing. The initial prediction of plasmids was possible through plasmidSPADES assemblies and subsequent extraction of *parA* and *parB* genes, but this failed to output distinct plasmids, only contigs of likely plasmid origin, preventing the high-resolution analysis of virulence, AMR and specialised metabolite loci. This approach also failed to predict a plasmid in *B. ambifaria* BCC0477, while the hybrid assembly produced a circularised 364 Kbp megaplasmid; in addition to several minor plasmids across the MinION sequenced strains (Table 10).

3.3.4. *B. ambifaria* BCC0478 AMR plasmid

One of the 64 *B. ambifaria* genomes analysed possessed a plasmid carrying multiple antimicrobial resistance genes. Further analysis revealed this plasmid to be a member of the broad host range IncP-1β group (Sen *et al.* 2013). An interrogation of publically available plasmids with high similarity to BCC0448 p1 revealed that less than 40% of the non-exhaustive list possessed AMR genes. The number of AMR genes detected in this subset also varied considerably highlighting the role of transposons in conferring AMR properties on IncP-1β plasmids. The distribution of IncP-1β plasmids was not restricted to Gram-negative organisms, with AMR derivatives detected in the Gram-positive bacterium *Mycobacterium abscessus*

(2017). While *B. ambifaria* BCC1248 was isolated from cystic fibrosis sputum in the USA, other IncP-1β plasmids have been isolated from *Aeromonas* sp. cultured from manhole wastewater near a hospital (Weingarten *et al.* 2018), *Acidovorax avenae* isolated from diseased grain plants (Yoshii *et al.* 2012), *M. abscessus* from human surgical sites (Leão *et al.* 2013), and *Bordetella bronchiseptica* from agricultural animals (Hedges *et al.* 1974). IncP-1β plasmids also possessed a global distribution encompassing North America, South America, Europe, and Asia.

Screening the locally assembled *Burkholderia* genomes identified two strains with evidence of IncP-1β AMR plasmids, *B. cenocepacia* BCC0506 and *B. gladioli* BCC1763. These strains, like *B. ambifaria* BCC0478, were also isolated from sputum samples of cystic fibrosis patients in US clinics. All three IncP-1β representatives carried a different collection of AMR genes, once again demonstrating the variability of this plasmid incompatibility group and importance of transposons in transferring AMR genes into these broad host range mobile genetic elements. Multiple other potential *Burkholderia* IncP-1β1 plasmids were detected in previously uncharacterised strains, but lacked AMR genes. A disadvantage of screening draft genomes for plasmid content is the possibility of plasmid sequences splitting across multiple contigs. Consequently, the plasmid origin of replication could be located on a different contig to the AMR genes, thereby underestimating the AMR carriage rate of the IncP-1β plasmids and overestimating the chromosomal AMR capacity. Understanding the clinical and agricultural impact of these AMR plasmids emphasises the importance of genomic data in characterising potential biological control strains to mitigate the dissemination of AMR in the environment.

A limitation of antimicrobial and metal resistance gene prediction is the comprehensiveness and accurate curation of the related databases. A combination of all four relevant AMR gene databases (CARD, ARG-ANNOT, NCBI, and ResFinder) provided a thorough analysis of the BCC0478 p1 plasmid. However, the use of a metal and biocide specific database (BacMet) produced multiple false positive results that prevented the effective prediction of metal resistance genes. A subset of IncP-1β plasmids were shown to carry genes conferring mercury resistance, the *mer* operon. To detect the presence of these genes, they were manually added to the vfdb database in a similar fashion to the secretion system genes discussed in section 3.3.1. A discrepancy between the number of *mer* genes predicted in the Illumina-only (six genes) and hybrid assembly (seven genes) of the BCC0478 genome illustrated the subtle yet potentially important differences in apparent gene content that are influenced by the sequencing technology.

3.3.5. *B. ambifaria* BCC1248 p1 virulence plasmid

Following plasmid prediction and hybrid genomic assembly a large plasmid was discovered in strain BCC1248 designated plasmid p1. BCC1248 p1 was confirmed to carry multiple virulence factors unique to this *B. ambifaria* strain compared to the remainder of the collection. These virulence factors included the *Burkholderia* T6SS-5 and T3SS-3 systems; the malleilactone biosynthetic gene cluster was also identified in BCC1248, but the hybrid assembly positioned the gene cluster on the c2 replicon. The T6SS-5 system was originally characterised in *B. pseudomallei* but a homologous system exists in *B. mallei* and *B. thailandensis* (Lennings *et al.* 2019). The exclusive secretion of the effector VgrG by T6SS-5 has been demonstrated as essential to the formation of multinucleated giant cells during infection, enabling the cell-cell dissemination of the pathogen without entry into the extracellular matrix (Schwarz *et al.* 2014; Lennings *et al.* 2019). Similar to the T6SS, the second secretion system, T3SS-3, was also characterised in *B. pseudomallei* with a homologous system in *B. mallei* and *B. thailandensis* (Vander Broek and Stevens 2017). A review of the literature by Vander Broek and Stevens (2017) illustrated the essential nature of this system for eukaryotic virulence in murine models. T3SS-3 is necessary to escape the endocytic vesicle and also supports infection in the lungs and subsequent dissemination (Vander Broek and Stevens 2017).

A search for similar plasmids revealed the large genetic distances between the closest *parB* genes indicated by the phylogeny scale bar and lack of shared genes between candidate plasmids. Both secretions systems were absent from all candidate plasmids with BCC1248 p1 *parB* homology; as such, this *B. ambifaria* plasmid is unique in its acquisition of chromosomally-encoded virulence factors from a biosafety level 3 organism. However, there is no evidence that these plasmid-borne secretion systems or malleilactone gene cluster are actively expressed, or contribute to the virulence of the *B. ambifaria* host.

3.4. Conclusions

The main conclusions following the pangenomic analysis, including AMR gene profiling, virulence factor detection and plasmid characterisation in *B. ambifaria* were as follows:

1) Considerable genomic diversity existed within *B. ambifaria*, the full extent of which was revealed following a systematic breakdown of replicon and extrachromosomal gene-carrying capacity.

2) The AMR profile of *B. ambifaria* is highly stable with the exception of a single AMR plasmid found in 1 of the 64 strains analysed; in contrast, a substantial variation in both secretion systems and other virulence factors was evident throughout the *B. ambifaria* genome.

3) Hybrid genomic assembly aided by long-read Oxford Nanopore Technology was essential in understanding the role of plasmids in contributing both AMR and virulence genes to *B. ambifaria* genomic diversity.

4. Specialised metabolites and antimicrobial activity of *Burkholderia ambifaria*

The data and results shown in this chapter represent my own independent work/investigation, except where otherwise stated. Research and methods pertaining to this chapter have been published in part in the following manuscript. Contributions from the authors are acknowledged where appropriate, and wording has been adapted where necessary.

Mullins AJ, Murray JAH, Bull MJ, Jenner M, Jones C, Webster G *et al.* Genome mining identifies cepacin as a plant-protective metabolite of the biopesticidal bacterium *Burkholderia ambifaria*. *Nature Microbiology* 2019; 4: 996–1005. DOI: 10.1038/s41564-019-0383-z

4.1. Introduction

Specialised metabolites are products, highly diverse in structure and function, that are not essential to an organism's growth (primary metabolism), but enhance survival under different environmental conditions (Demain and Fang 2000). These natural products may act as defensive metabolites against other species, inter- and intra-domain communication during symbioses formation (Demain and Fang 2000), and virulence factors during infections. The *Burkholderia* genus is rich in secondary metabolites which is often attributed to their large, multi-replicon genomes encoding a plethora of biosynthetic gene clusters (BGCs) (Depoorter *et al.* 2016). A limited number of characterised *Burkholderia* secondary metabolites lack antimicrobial activity, for example, ornibactin (Stephan *et al.* 1993), pyochelin (Sokol 1986) and cepaciachelin (Barelmann *et al.* 1996) are examples of siderophores; while the polyketide thailandamide possesses moderate anti-proliferative activity (Ishida *et al.* 2010). Antimicrobial and cytotoxic compounds represent the dominant secondary metabolite class characterised in *Burkholderia* species (Masschelein *et al.* 2017).

The first antimicrobial characterised in *Burkholderia* was the broad range anti-fungal pyrrolnitrin in *Burkholderia pyrrocinia* (*Pseudomonas pyrrocinia*) (Arima *et al.* 1964), and closely followed by cytotoxic azapteridine toxoflavin in *B. gladioli* (Levenberg and Linton 1966). Xylocandins (Bisacchi *et al.* 1987), cepacidines (Lim *et al.* 1994), occidiofungin (Lu *et al.* 2009) and burkholdines (Tawfik *et al.* 2010) are a series of structurally similar or identical metabolites characterised in *Burkholderia cepacia* ATCC39277, *Burkholdeia cepacia* (*P. cepacia*) AF 2001, *Burkholderia contaminans* MS14 and *B. ambifaria* 2.2N, respectively. These glycopeptides have exhibited antagonistic activity against *Candida albicans, Rhizoctonia solani, Alternaria alternata, Phytophthora infestans* and *Pythium* species (Bisacchi *et al.* 1987;

Lu *et al.* 2009; Tawfik *et al.* 2010). Additional antimicrobials described in the earlier literature include the anti-bacterial polyynes cepacin (Parker *et al.* 1984) and caryoynencin (Kusumi *et al.* 1987), and anti-fungal lipopeptide AFC-B11 (Kang *et al.* 1998). Most recently a collection of antibacterial compounds have been identified, such as the chlorinated anti-Gram-positive metabolites bactobolins (Seyedsayamdost *et al.* 2010), enacyloxin IIa that exhibited anti-Gram-negative activity (Mahenthiralingam *et al.* 2011), and the anti-mycobacterial macrolide gladiolin (Song *et al.* 2017) along with its isomer lagriene (Flórez *et al.* 2017).

Many of the BGCs responsible for antimicrobial metabolite synthesis in *Burkholderia*, such as pyrrolnitrin, toxoflavin and occidiofungin (Table 11), were identified via random transposon mutagenesis. Transposon mutagenesis is performed by mating an *E. coli* donor carrying a plasmid encoding the desired transposon with the target organism, resulting in a single transposon integrating randomly into each target bacterial cell. Following conjugation, the resulting transconjugant colonies are screened for the loss of a specific phenotype, such as antimicrobial production (Mahenthiralingam *et al.* 2011) and antimicrobial resistance (Ramage *et al.* 2017). Notable exceptions to transposon mutagenesis-based BGC discovery in *Burkholderia* were the site-directed (targeted) mutagenesis of genes hypothesised to be involved in the regulation or biosynthesis of the antimicrobials bactobolins (Duerkop *et al.* 2009) and gladiolin (Song *et al.* 2017), respectively (Table 11). The BGC associated with cepacin biosynthesis was identified via the adjacent LuxRI QS system, and subsequent targeted mutagenesis of biosynthetic genes resulting in loss of antimicrobial activity (Mullins *et al.* 2019). Identifying the genomic origin of these antimicrobial metabolites contributes to the holistic understanding of bacterial biosynthetic machinery, and the resulting knowledge to manipulate this machinery to synthesise novel derivatives.

Targeted mutagenesis of *Burkholderia* regulatory systems has been exploited to activate silent BGCs (Ishida *et al.* 2010), disrupt BGC gene expression to confirm the BGC origin of metabolites (Duerkop *et al.* 2009), and identify BGCs influenced by global regulatory systems (Schmidt *et al.* 2009b; Mao *et al.* 2017). Global and pathway-specific quorum sensing regulatory networks are associated with multiple antimicrobial BGCs including pyrrolnitrin, bactobolins and enacyloxin. The standard model of quorum sensing requires a cognate *luxR* and *luxI* gene system. The *luxR* gene encodes a regulatory protein, while the *luxI* gene encodes an acyl-homoserine lactone synthase that synthesises a diffusible signalling molecule, acyl-homoserine lactones (AHLs); both of which are present at low levels in the bacterial cell. At sufficiently high cell densities, the AHL passes a threshold concentration in the cell, enabling the AHL to bind the LuxR regulatory protein. The resulting complex is capable of binding promotors containing a LuxR specific binding motif (Lux box), consequently altering

gene expression. However, non-canonical QS systems exist where the LuxR regulatory protein lacks the autoinducer-binding domain, and instead binds an exogenous factor, such as the LuxR proteins encoded in the burkholdine BGC (Gu *et al.* 2009a; Chapalain *et al.* 2013). Prediction of these regulatory systems in bacterial genomes provides an alternative method of identifying BGCs, which has yielded results with the identification of the previously unknown genetic origin of the polyyne cepacin (Mullins *et al.* 2019). *Burkholderia* also possess a quorum sensing system independent of AHLs, known as the *Burkholderia* diffusible signal factor (BDSF) which is a fatty acid, *cis*-2-dodecenoic acid (Boon *et al.* 2008). At low cell densities the high intracellular cyclic-di-guanylate (c-di-GMP) level prevents the RpfR-GtrR complex binding to promotors and regulating gene expression (Yang *et al.* 2017). Increasing cell densities results in elevated BDSF concentrations that interact with RpfR and enhances RpfR phosphodiesterase activity, converting c-di-GMP to GMP, and consequently allowing the RpfR-GtrR complex to bind target promoters and regulate gene expression (Yang *et al.* 2017). Based on RNA-sequencing experiments in *B. cenocepacia* there was overlap in the regulons of both AHL and BDSF quorum sensing systems, but the two systems appeared to function in parallel rather than hierarchical (Schmid *et al.* 2012).

There is no consensus on the prevalence or co-occurrence of antimicrobial BGCs in *B. ambifaria*, nor a complete understanding of pathogen antagonism exhibited by individual strains. Literature searches have highlighted that previous research on *B. ambifaria* focusses on individual strain antimicrobial activity against multiple pathogens, or multiple *B. ambifaria* strains exhibiting activity against a limited pathogen collection (Parke and Gurian-Sherman 2001). The lack of a holistic understanding of antimicrobial distributions represents a gap in our knowledge of *B. ambifaria*.

Table 11. *Burkholderia* metabolite discovery and genetic method of identifying the corresponding BGC.

Metabolite	Discovery	Pathway	Method
Pyrrolnitrin	(Arima *et al.* 1964)	(Hammer *et al.* 1997)	Tn5 mutagenesis of gain-of-function cosmid
Toxoflavin	(Levenberg and Linton 1966)	(Suzuki *et al.* 2004)	Transposon mutagenesis often lost their ability to rot rice seedlings (phenotype)
Xylocandins	(Meyers *et al.* 1987) (Bisacchi *et al.* 1987)	/	/
Cepacidines	(Lee *et al.* 1994) (Lim *et al.* 1994)	/	/
Occidiofungin	(Lu *et al.* 2009)	(Gu *et al.* 2009a) (Gu *et al.* 2009b)	Transposon mutagenesis
Burkholdines	(Tawfik *et al.* 2010)	/	/
Bcc toxin	(Thomson and Dennis 2012)	(Thomson and Dennis 2012)	Plasposon mutagenesis
Cepacin	(Parker *et al.* 1984)	(Mullins *et al.* 2019)	LuxR mining and site-directed mutagenesis
Caryoynencin	(Kusumi *et al.* 1987)	(Ross *et al.* 2014)	Transposon mutagenesis
Lipopeptide AFC-B11	(Kang *et al.* 1998)	(Kang *et al.* 1998)	Transposon mutagenesis
Bactobolins	(Duerkop *et al.* 2009)	(Duerkop *et al.* 2009)	Site-directed (targeted) mutagenesis
Enacyloxin	(Mahenthiralingam *et al.* 2011)	(Mahenthiralingam *et al.* 2011)	Transposon mutagenesis
Gladiolin	(Song *et al.* 2017)	(Song *et al.* 2017)	Site-directed (targeted) mutagenesis
Lagriene	(Flórez *et al.* 2017)	(Flórez *et al.* 2017)	BGC homology and compound structural relatedness
Sinapigladioside	(Flórez *et al.* 2017)	/	/
Cepafungins	(Shoji *et al.* 1990)	(Schellenberg *et al.* 2007) (glidobactin)	PCR (nested degenerate from peptide sequence) screened λ library and targeted mutagenesis
Cepaciamides	(Jiao *et al.* 1996)	/	/
Phenazines	(Cartwright *et al.* 1995)	(Mavrodi *et al.* 1998)	Transposon mutagenesis
Hydroxyquinolines	(Moon *et al.* 1996)	(Gallagher *et al.* 2002)	Transposon mutagenesis
Capistruin	(Knappe *et al.* 2009)	(Knappe *et al.* 2009)	Metabolite profiling and genome mining
Icosalide	(Jenner *et al.* 2019)	(Jenner *et al.* 2019)	Metabolite profiling and genome mining

4.1.1. Aims and objectives

This chapter aims to exploit the 64 available *B. ambifaria* genomes and corresponding 62 strains to investigate the distribution and diversity of biosynthetic gene clusters, and spectrum of antimicrobial activity through the following objectives:

1) To define the antimicrobial bioactivity across the *B. ambifaria* phylogeny against bacterial, fungal and oomycete plant pathogens.

2) To define the specialised metabolite potential of *B. ambifaria* by predicting the encoded biosynthetic gene clusters.

3) To identify the specialised metabolite regulatory systems *in silico*, and interrogate the LuxR regulation of *B. ambifaria*.

4.2. Results

4.2.1. Establishing a curated collection of *B. ambifaria* BGCs

4.2.1.1. *In silico* prediction of *B. ambifaria* secondary metabolite biosynthetic clusters

AntiSMASH analysis (Weber *et al.* 2015) of the 64 *B. ambifaria* genomes identified 1251 BGCs, representing 17 recognised metabolite classes, and included 127 BGCs that could not be classified by antiSMASH, and were subsequently labelled as “other”. Literature searches highlighted the AFC-BC11 lipopeptide BGC as a gene cluster not recognised by antiSMASH. A BLAST (Morgulis *et al.* 2008) search of the genomes detected 34 examples of the AFC-BC11 BGC. These BGCs were extracted from the *B. ambifaria* genomes and added to the antiSMASH predicted collection, resulting in 1,285 BGCs.

4.2.1.2. De-replication of the predicted *B. ambifaria* BGCs

The combination of antiSMASH and BLAST analyses predicted 1,285 BGCs, but lacked information about the relatedness of the BGCs to one another. A pairwise k-mer matching approach permitted the comparison of BGCs to each other, and subsequent clustering of homologous BGCs. A range of Mash-maximum distances (0.2, 0.1, 0.07, 0.05 and 0.03) were tested to identify the optimum threshold for defining distinct BGCs. Mash-distances of 0.2 and 0.03 were immediately discounted; the 0.2 threshold resulted in interlinking of clearly distinct network clusters of BGCs, in contrast to the 0.03 threshold that began to erroneously split network clusters of distinct BGCs. Mash-distance thresholds of 0.1, 0.07 and 0.05 generated the most reliable networks regarding network clusters. Despite the 0.1 distance threshold generating a network of clusters that were divided correctly by distinct BGC; the 0.05 distance threshold was favoured due to the separation of MEX-5 and IOP40-10 un-positioned BGCs into singleton BGCs. This was preferred as it allowed the manual curation of these antiSMASH predicted partial gene clusters into known BGCs (described below); and prevented the overestimation of predicted BGCs in the *B. ambifaria* MEX-5 and IOP40-10 genomes.

4.2.1.3. Manual curation of the de-replicated BGCs

The genome sequences available in the public databases for *B. ambifaria* MEX-5 and IOP40-10 were assembled from the historical pyrosequencing (454) sequencing platform, resulting in a lower coverage compared to the remaining 62 genomes sequenced by the Illumina platform during this study. The lower coverage of the MEX-5 and IOP40-10 genomes has consequently impacted the scaffolding of contigs, producing a lower N50 value that has affected the ability of antiSMASH to detect complete BGCs. AntiSMASH predicted 26 BGCs in *B. ambifaria* MEX-5 (16 assigned to replicons; 10 of unknown genomic position) and 25 BGCs in *B. ambifaria* IOP40-10 (14 assigned to replicons; 11 of unknown genomic position). Multiple un-positioned

BGCs occurred as singletons or doublets following de-replication. A combination of BLAST and gene synteny comparisons to other gene clusters predicted by antiSMASH enabled the re-assigning of BGCs with unknown genomic locations.

Of the 10 MEX-5 BGCs with unknown locations, four BGCs (2, 3, 8 and 9) possessed high gene synteny and nucleotide sequence similarity to the burkholdines BGC; and two unknown location BGCs (5 and 7) possessed similarities to the bactobolins BGC. These gene clusters were subsequently deleted and replaced by representative burkholdine and bactobolin BGCs for downstream bioinformatics analyses. One of the unknown location BGCs (BGC 10) represented a partial ornibactin BGC, while remaining ornibactin BGC sequence was already positioned on a replicon, as such; the unmapped cluster was deleted from the dataset. Unknown location BGCs 4 and 6 possessed homology to a T1PKS and arylpolyene BGC, respectively, already localised to replicons in other *B. ambifaria* strains, and were subsequently assigned an encoding replicon. The final non-positioned MEX-5 BGC (BGC 1) possessed low homology to all other antiSMASH predicted BGCs in the 64 *B. ambifaria* genomes, and thus remained with an unknown genomic position. Following the manual curation of the antiSMASH predicted MEX-5 BGCs, 21 gene clusters were predicted; 20 of these gene clusters could be assigned an encoding replicon.

B. ambifaria IOP40-10 encoded 11 antiSMASH predicted BGCs with unknown genomic locations. Of these BGCs, the gene clusters labelled 1, 3, 4, 8, 9, 10 and 11 possessed high sequence identity to the burkholdine BGC, which was already scaffolded onto the third replicon in a mapped BGC (BGC 13). Consequently, all seven un-positioned BGCs were removed from the dataset. Similar to MEX-5, the ornibactin BGC in IOP40-10 was mapped to the first replicon (c1), but also possessed un-positioned fragments (BGCs 6 and 7) which could be removed from the BGC dataset. The un-positioned BGC 5 possessed high sequence identity to the arylpolyene BGC encoded on the first replicon (c1); and the final un-positioned BGC encoded a homoserine lactone synthase with no homology to other LuxRI systems located on the *B. ambifaria* genome. Following these manual adjustments, the *B. ambifaria* IOP40-10 genome encoded 16 BGCs, of which one could not be assigned an encoding replicon.

In addition to the manual curation of the MEX-5 and IOP40-10 BGCs, the *B. ambifaria* BCC1105 antiSMASH predicted bacteriocin-ectoine BGC, upon further analysis, actually represented the ectoine 2 BGC encoded on the second replicon, and a novel *B. ambifaria* bacteriocin BGC. Following the addition of the manual edits into the cluster prediction, a refined total of 1,272 BGCs were predicted across the 64 *B. ambifaria* genomes. Using the 0.05 distance threshold resulted in the separation of un-positioned BGCs into singletons, allowing

manual curation and correction for BGC overestimation (Section 4.2.1.2). However, this more stringent distance threshold caused two distinct BGCs network clusters to split into smaller network clusters. Replicon 1 (c1) NRPS 2 was split into two network clusters with 39 and five BGC homologues; and the replicon 2 (c2) phenazine network cluster was split into two clusters with six and seven BGC homologues. These BGCs were known to be homologous following BLAST nucleotide sequence comparisons and gene synteny comparisons. These network clusters were merged into two clusters: c1 NRPS 2 with 44 BGC representatives and c2 phenazine with 13 BGC representatives. Following the manual curation of the BGC network, 38 distinct BGCs were identified (Table 12; Figure 15).

Table 12. De-replicated secondary metabolite gene clusters of *B. ambifaria*.

Cluster Type (antiSMASH prediction)	Prevalence in *B. ambifaria* (out of 64 genomes)	Average length (kbp)[a]
Replicon c1		
Terpene 1	65[b]	20.9
NRPS 1	64	54.5
Arylpolyene 1	64	41.2
NRPS 2	44	46.8
PKS 1	28	47.6
Lantipeptide 1	10	27.1
Butyrolactone 1	1	11.0
Terpene 2	1	21.0
Replicon c2		
Homoserine lactone 1	64	20.6
Phosphonate	64	41.7
Terpene 3	64	21.1
Other 1	64	41.1
Terpene 4	64	24.0
Bacteriocin 1	55	13.1
Arylpolyene 2	53	44.9
Ectoine 1	53	10.4
Ectoine 2	38	10.4
Homoserine lactone (cepacin-associated) 2	22	20.7
Other 2	20	43.0
Butyrolactone-otherKS	16	32.5
Phenazine	13	20.4
PKS 2	7	44.9
NRPS 3	2	52.8
Bacteriocin 2	1	25.9
Arylpolyene 3	1	41.1
Replicon c3		
Homoserine lactone 3	63	20.6
Bacteriocin 3	63	10.8
Terpene 5	63	22.0
NRPS-T1PKS 1	54[c]	85.5 or 117.6[d]
PKS 3	40	43.9 and 63.9[e]
Other 3	34	27.0
Butyrolactone 2	27	10.7
Other 4	24	43.8
NRPS-T1PKS 2	17	60.9
Trans-AT-PKS	6	91.2
Unknown genomic location		
NRPS 4	1	26.8
Homoserine lactone 4	1	20.1
NRPS-T1PKS-*trans*-AT-PKS	1	69.8

Footnotes:

[a] Average length was calculated from predicted antiSMASH clusters or extracted sequences identified using BLAST; partial clusters were excluded (BGCs from strains MEX-5 and IOP40-10). Other strains were excluded if the clusters were manually split from the antiSMASH predicted cluster.

[b] *B. ambifaria* strain BCC1249 encoded a duplicate terpene 1 cluster.

[c] *B. ambifaria* strains MEX-5, IOP40-10 and BCC1065 either possess low quality genomes preventing the reconstruction of the burkholdine NRPS-T1PKS pathway or encode a partial cluster.

[d] The published NRPS-T1PKS pathway (85.5 kbp) was encoded by 36 *B. ambifaria* strains, while the larger 117.6 kbp pathway was encoded by 15 *B. ambifaria* strains and consists of the published pathway with additional biosynthetic genes.

[e] The 43.9 kbp version of the pathway was encoded by 26 *B. ambifaria* strains, while the 63.9 kbp version was encoded by 14 *B. ambifaria* strains.

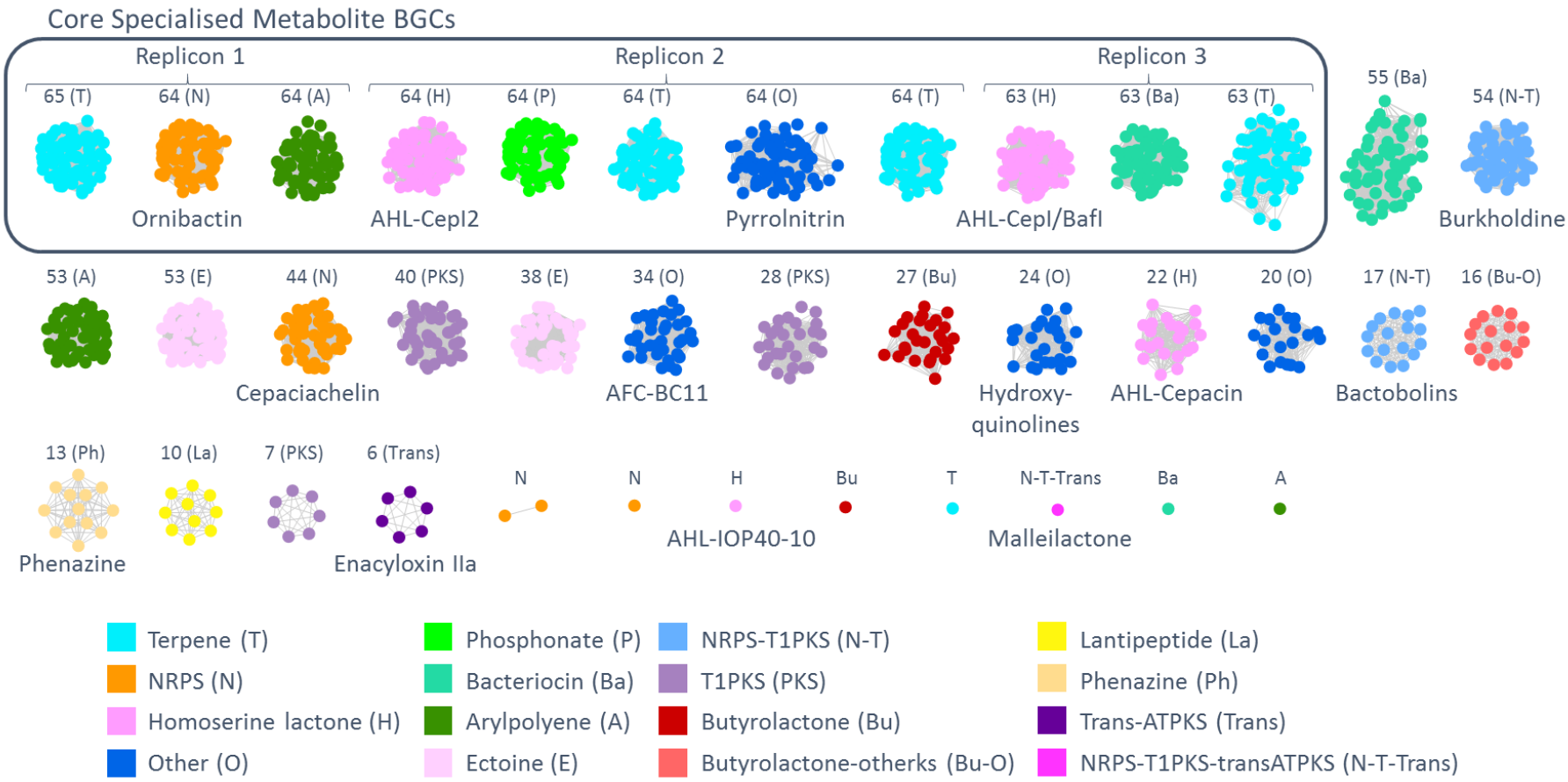


Figure 15. Specialised metabolite BGC network analysis of 64 *B. ambifaria* strains.
The specialized metabolite BGC dataset was graphically summarized to highlight the prevalence of BGCs, specialized metabolite diversity and the core or accessory nature of the BGCs. A total of 1,272 BGCs were detected across the 64 strains, and de-replication indicated these represent 38 distinct BGCs. Each node represents a secondary metabolite biosynthetic gene cluster (BGC) extracted from a single *B. ambifaria* strain Nucleotide sequences were clustered using bioinformatics software MinHash and visualized with Cytoscape. Node colours represent secondary metabolite classes, and numbers correspond to the number of BGC representatives of each distinct network cluster. Core BGCs were defined as BGCs that occurred in >98% of *B. ambifaria* strains. Characterised BGCs known in the literature were labelled. BGCs responsible for pyrrolnitrin, AFC-BC11 and hydroxyquinolines biosynthesis were classified as Other (O) by antiSMASH but represent different metabolite classes not recognised by antiSMASH.

4.2.1.4. Identifying known BGCs within the *B. ambifaria* cluster network

A literature search uncovered a plethora of characterised secondary metabolites and corresponding BGCs in *B. ambifaria* and the wider *Burkholderia* genus. Where available, the gene maps of these BGCs were compared to representatives of each of the 38 distinct BGCs predicted in *B. ambifaria* during this study. Thirteen characterised BGCs were identified in the literature: ornibactin, pyrrolnitrin, burkholdines, cepaciachelin, AFC-BC11, hydroxyquinolines, bactobolins, phenazine, enacyloxin IIa and malleilactone; and included the three QS systems: CepR2I2, CepRI/BafRI and IOP40-10 specific LuxRI system (Table 12; Figure 15). The remaining 25 distinct BGCs are uncharacterised and represent un-exploited biosynthetic diversity in *B. ambifaria*.

4.2.1.5. Distribution of BGCs in *B. ambifaria* multi-replicon genome

Of the 1,272 BGCs predicted across the 64 *B. ambifaria* strains, each genome encoded a mean of 20 BGCs. Eighteen secondary metabolite classes were identified, fifteen recognised by antiSMASH, in addition to pyrrolnitrin, AFC-BC11 and hydroxyquinolines whose BGCs were characterised as "other" by antiSMASH. One "other" cluster with 20 BGC examples was not characterised in the literature and could represent a novel secondary metabolite class. Of the 38 distinct BGCs, three singleton pathways were detected in contig sequences that possessed no significant homology to the reference sequences used as scaffolds; and thus could not be incorporated into replicons c1, c2 or c3. One of the three BGCs with unknown genomic locations was the hybrid non-ribosomal peptide synthetase *trans*-acyltransferase polyketide synthase (NRPS-*trans*-AT PKS) with 92% nucleotide sequence identity to the BGC responsible for malleilactone biosynthesis (Duerkop *et al.* 2009). The two remaining BGCs with unknown genomic locations were a NRPS encoded by MEX-5 and the characterised cognate LuxRI QS system encoded by IOP40-10. Terpenes (5/38) and NRPS (4/38) were the most frequently predicted secondary metabolite classes in the *B. ambifaria* de-replicated BGC collection. Eleven of the 39 distinct BGCs were encoded by 63-64 *B. ambifaria* strains and were consequently classified as core secondary metabolite biosynthetic pathways; these included the antimicrobial encoding pyrrolnitrin BGC, two cognate LuxRI systems and the ornibactin BGC. A significant proportion (21%) of the distinct BGCs were encoded by less than 5% of the *B. ambifaria* strains; that is, the BGCs occurred as singletons or doublets.

The distribution of BGCs varied across the *B. ambifaria* replicons c1, c2 and c3, in both capacity and density (Figure 16). Each *B. ambifaria* strain encoded 3-6 BGCs on replicon c1, with eight distinct BGCs identified on the replicon. Replicon c2 encoded between 8-11 BGCs, with 17 distinct BGCs; and replicon c3 encoded 4-9 BGCs across the 64 *B. ambifaria* genomes,

with 11 distinct BGCs. The biosynthetic density of the replicons (total replicon BGC sequence per replicon size) was highest in replicon c3 (mean 19.4%), and lowest in replicon c1 (mean 5.0%); replicon c2 encoded a mean of 9.2% BGCs. The biosynthetic density of each replicon was not static across the 64 *B. ambifaria* strains; depending on the strain, replicon c3 encoded between 9.6% (MC40-6) and 30.1% (BCC0203) biosynthetic density. The variability of replicons c1 and c2 biosynthetic density was lower: c1 varied between 2.1% (IOP40-40) and 6.2%, and c2 varied between 7.6% (BCC1252) and 11.7% (BCC0480).

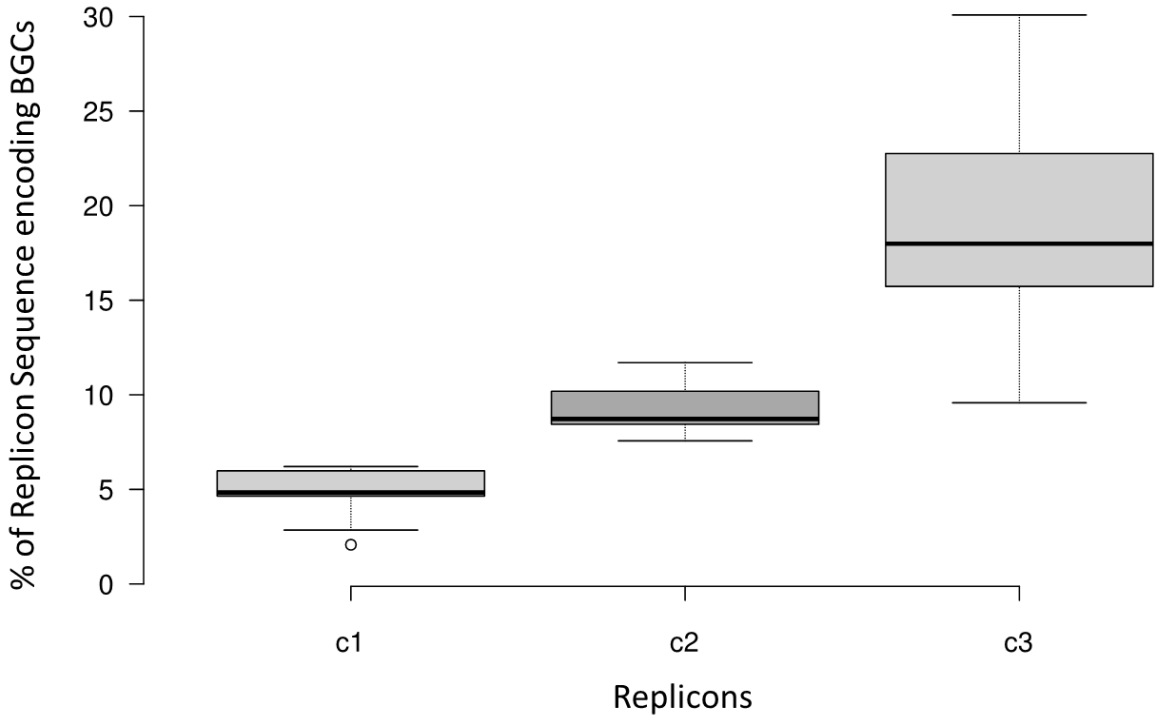


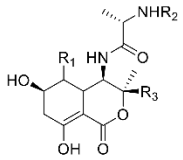
Figure 16. Distribution of secondary metabolite biosynthetic potential across the three replicons of 64 *B. ambifaria* strains.
Centre lines represent the median; box limits indicate the 25th and 75th percentiles; and whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles. Only 63 strains were represented by the box plot of replicon c3 due to the lack of a third replicon in *B. ambifaria* BCC1105.

4.2.1.6. Antimicrobial BGCs distribution

As noted previously, 13 of the 38 distinct *B. ambifaria* BGCs have been characterised in the literature. Of the 13 known BGCs, seven are responsible for the synthesis of compounds with antimicrobial activity. Pyrrolnitrin (Schmidt *et al.* 2009b) was the only antimicrobial-encoding BGC that was predicted in all 64 *B. ambifaria* strains. The least common antimicrobial-encoding BGC was enacyloxin IIa (Mahenthiralingam *et al.* 2011), detected in only six strains. Apart from the pyrrolnitrin and phenazine BGCs encoded on replicon c2, the remaining five antagonistic compounds were synthesised by BGCs encoded on replicon c3; no known antimicrobial-encoding BGCs were encoded on replicon c1. Aligning the presence of these BGCs against the core-gene phylogeny of *B. ambifaria* revealed that several antagonistic compounds were associated with distinct phylogenetic clades. Six of the seven clade 3 strains encoded the enacyloxin IIa BGC. All *B. ambifaria* strains encoded the burkholdine (Tawfik *et al.* 2010) except for the strains in clade 4 and BCC1105 that naturally lacked replicon c3. The bactobolin BGC (Seyedsayamdost *et al.* 2010) did not correlate strictly with specific core-phylogeny clades, but the bactobolin BGC was concentrated in clade 1, and less frequently encountered in other clades. Despite *B. ambifaria* as a species having the capacity to encode seven BGCs encoding antimicrobial compounds, no individual strain encoded all seven BGCs. However, 59% of strains (38 of 64) encoded four or more antimicrobial-encoding BGCs. Two strains, BCC1105 and BCC1224, did not possess any antimicrobial-associated BGCs, except the core anti-fungal pyrrolnitrin BGC.

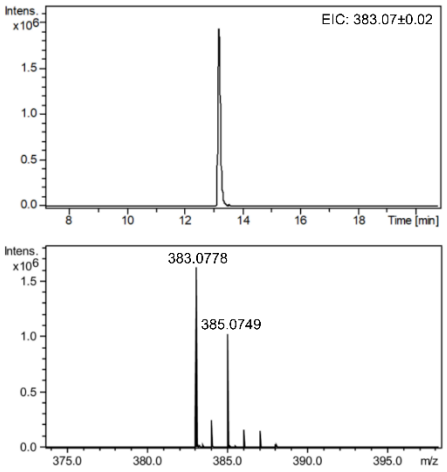
4.2.2. Expression of antimicrobial-encoding BGCs

To address the issue of silent gene clusters, which has been previously documented in *Burkholderia* (Mahenthiralingam *et al.* 2011; Masschelein *et al.* 2017), metabolite extracts from a selection of *B. ambifaria* strains were separated and analysed by UHPLC-ESI-Q-TOF-MS. LC-MS analysis was performed by Dr Matthew Jenner at the University of Warwick. Detection of these antimicrobial metabolites enabled the correlation of *in vitro* synthesis to BGC prediction. Ten *B. ambifaria* strains were selected, representing the seven previously characterised biocontrol strains in the collection, in addition to three strains from phylogenetically separate clades of the core-gene phylogeny to capture *B. ambifaria* intra-species diversity. Six known antimicrobial metabolites were detected: pyrrolnitrin, burkholdines, bactobolins, enacyloxin IIa, hydroxyquinolines and cepacin A (Table 13); however, only five metabolites could be correlated to a BGC, the cepacin BGC was unknown. Among the ten *B. ambifaria* strains there were 25 BGC representatives of the five known metabolites; the majority of which (22 of 25) were biosynthetically active and synthesised the corresponding antimicrobial compound. The three silent BGCs were representatives of pyrrolnitrin and burkholdine BGCs present in strain ATCC 53266 (BCC0338), and a hydroxyquinoline BGC in strain BC-F (BCC0207). Cepacin was detected in *B. ambifaria* J82 (BCC0191), but was not detected in the remaining nine strains analysed for *in vitro* metabolite production.

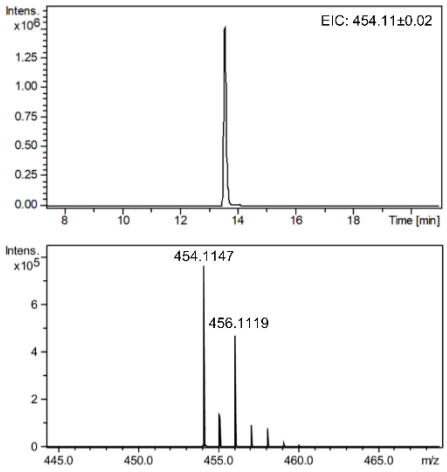


Bactobolins	Chemical Formula	Exact Mass
A: R_1=OH, R_2=H, R_3=$CHCl_2$	$C_{14}H_{20}Cl_2N_2O_6$	382.0698
B: R_1=OH, R_2=L-Ala, R_3=$CHCl_2$	$C_{17}H_{25}Cl_2N_3O_7$	453.1070
C: R_1=H, R_2=H, R_3=$CHCl_2$	$C_{14}H_{20}Cl_2N_2O_5$	366.0749
D: R_1=H, R_2=L-Ala, R_3=$CHCl_2$	$C_{17}H_{25}Cl_2N_3O_6$	437.1120
E: R_1=OH, R_2=L-Ala-L-Ala, R_3=$CHCl_2$	$C_{20}H_{30}Cl_2N_4O_8$	524.1441
G: R_1=H, R_2=H, R_3=CH_2Cl	$C_{14}H_{21}ClN_2O_5$	332.1139
H: R_1=H, R_2=L-Ala, R_3=CH_2Cl	$C_{17}H_{26}ClN_3O_6$	403.1510

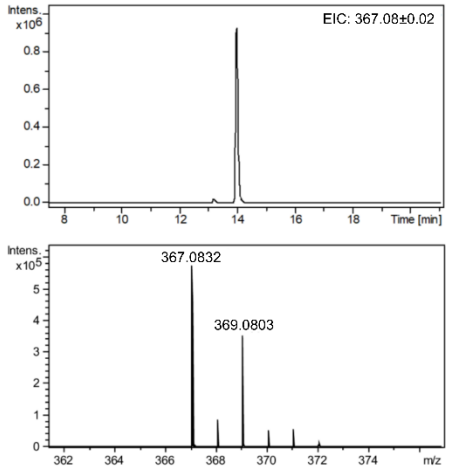
Bactobolin A
[M+H]$^+$: 383.0776



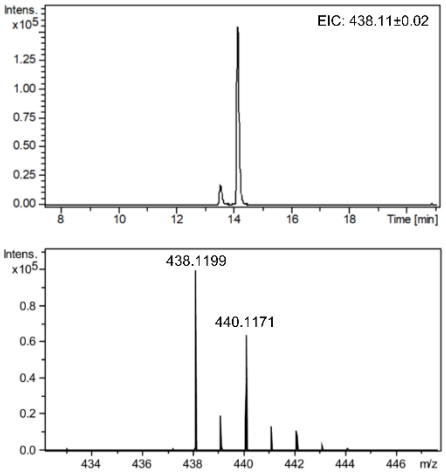
Bactobolin B
[M+H]$^+$: 454.1148



Bactobolin C
[M+H]$^+$: 367.0827



Bactobolin D
[M+H]$^+$: 438.1198



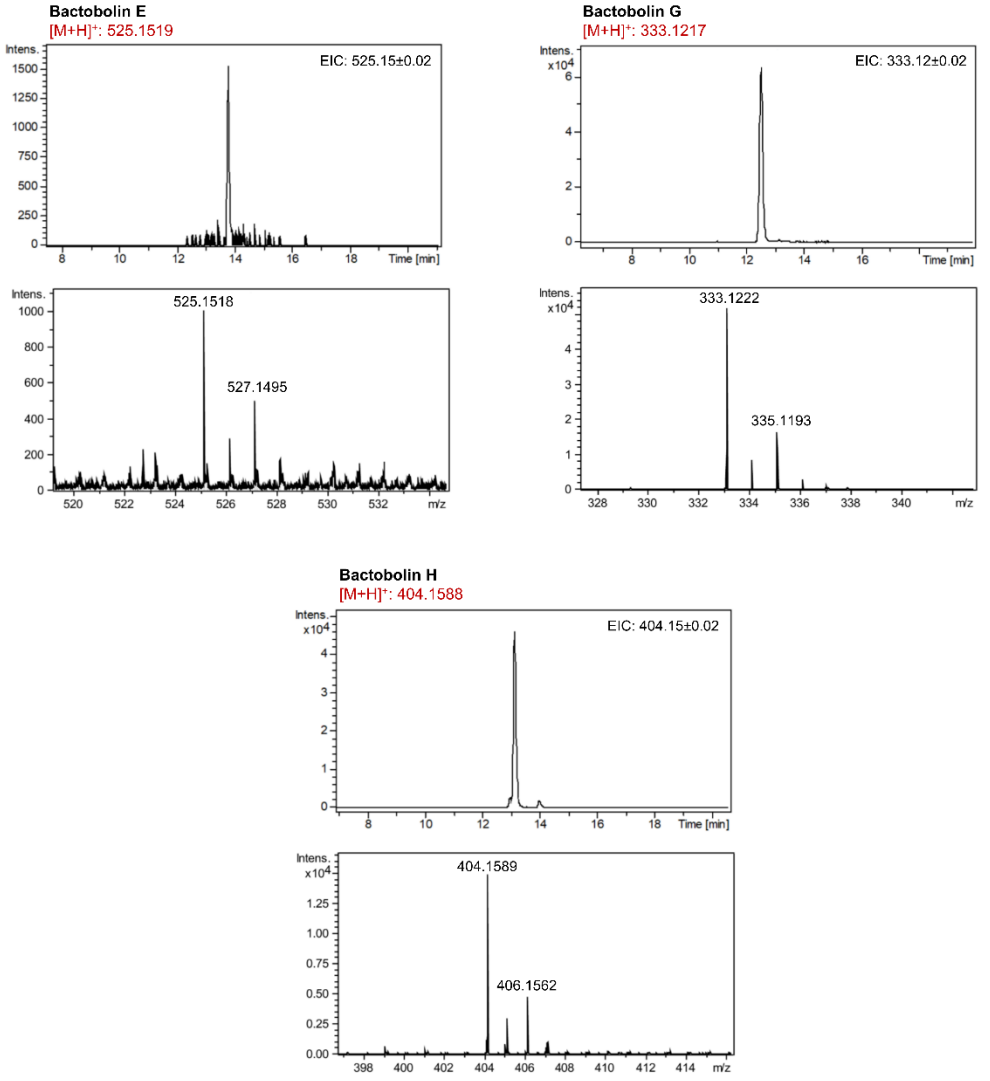
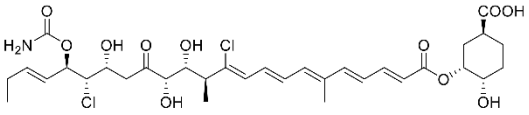
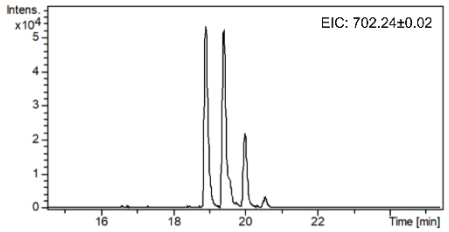


Figure 16 (a). Bactobolins (A-E, G and H)



Enacyloxin IIa
Chemical Formula: $C_{33}H_{45}Cl_2NO_{11}$
Exact Mass: 701.2370
$[M+H]^+$: 702.2447



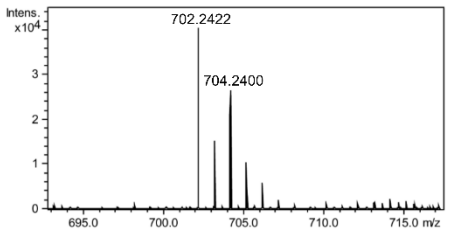
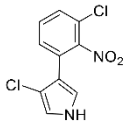
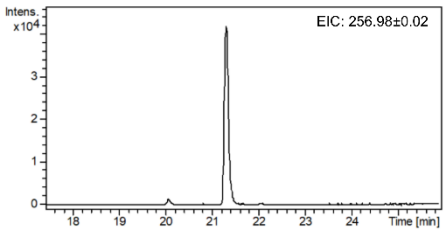


Figure 16 (b). Enacyloxin IIa



Pyrrolnitrin
Chemical Formula: $C_{10}H_6Cl_2N_2O_2$
Exact Mass: 255.9806
$[M+H]^+$: 256.9884



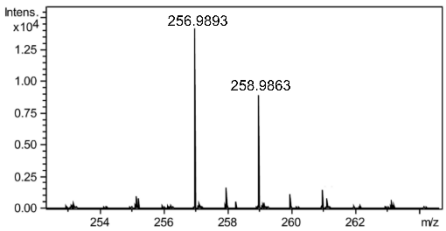


Figure 16 (c). Pyrrolnitrin

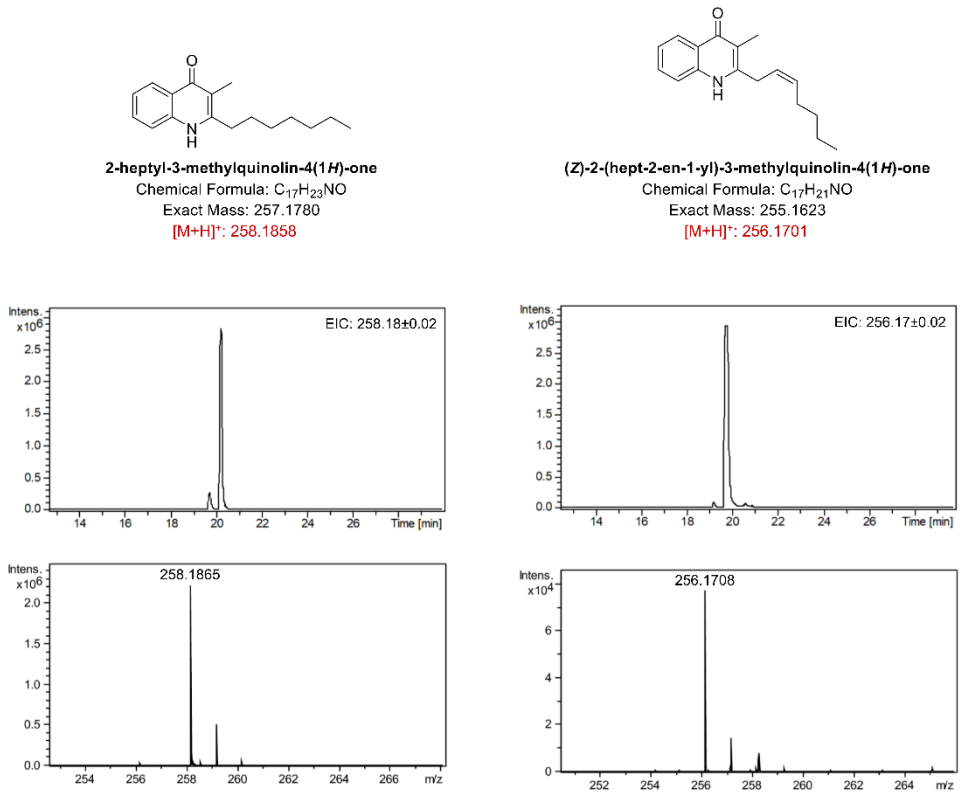


Figure 16 (d). Hydroxyquinolines

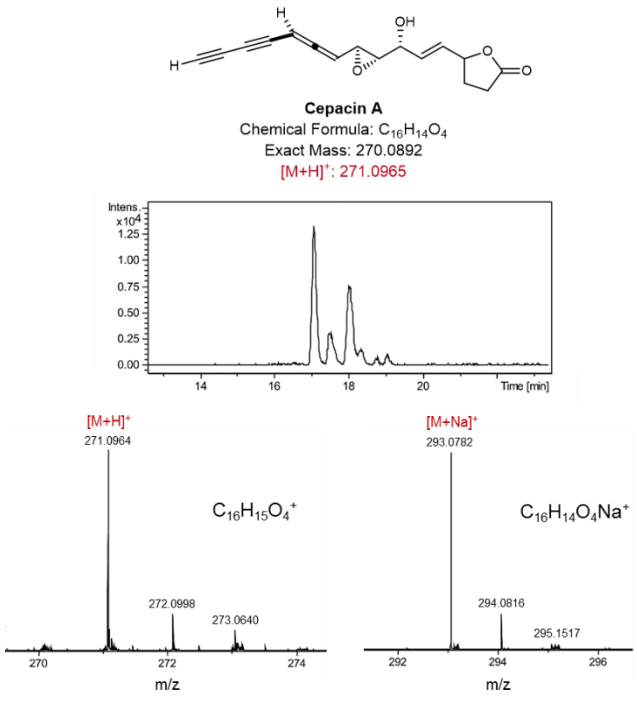
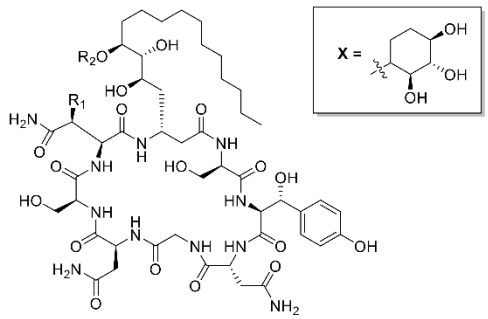
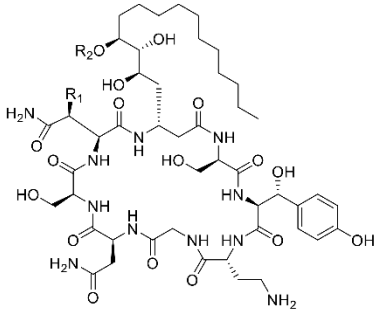


Figure 16 (e). Cepacin A

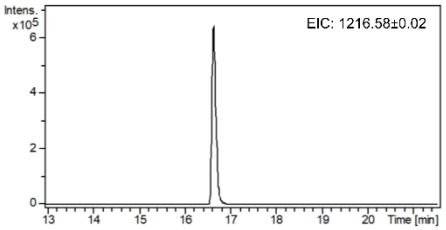


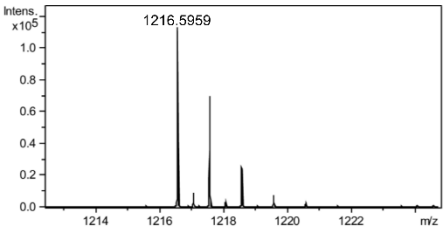


Burkholdine	Chemical Formula	Exact Mass
Bk-1229: R_1=OH, R_2=β-**X**	$C_{52}H_{83}N_{11}O_{23}$	1229.5663
Bk-1097: R_1=OH, R_2=H	$C_{47}H_{75}N_{11}O_{19}$	1097.5240
Bk-1213: R_1=H, R_2=β-**X**	$C_{52}H_{83}N_{11}O_{22}$	1213.5714

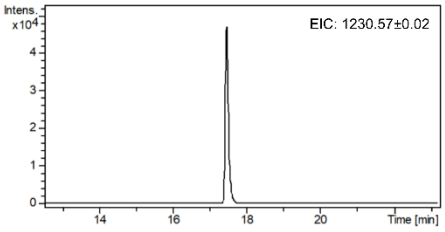
Burkholdine	Chemical Formula	Exact Mass
Bk-1215: R_1=OH, R_2=α-**X**	$C_{52}H_{85}N_{11}O_{22}$	1215.5870
Bk-1199: R_1=H, R_2=β-**X**	$C_{52}H_{85}N_{11}O_{21}$	1199.5921

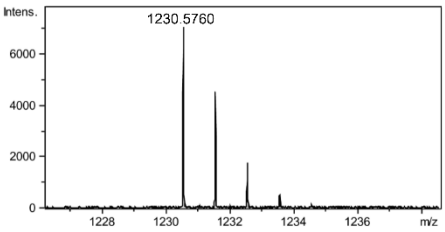
Bk-1215
[M+H]⁺: 1216.5948



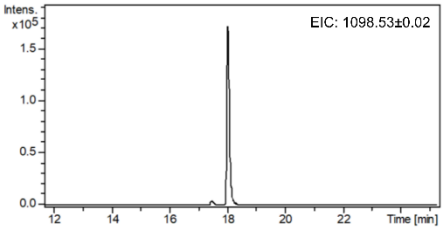


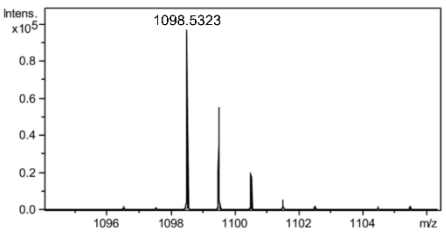
Bk-1229
[M+H]⁺: 1230.5741



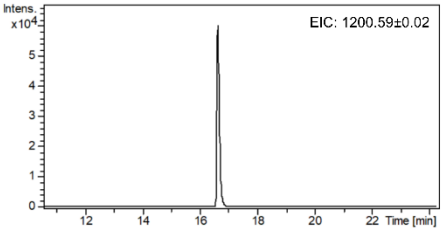


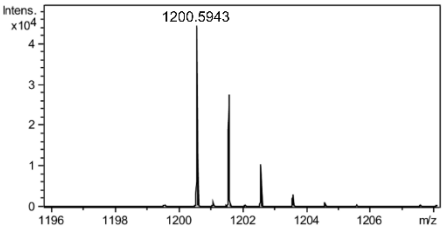
Bk-1097
[M+H]⁺: 1098.5318





Bk-1199
[M+H]⁺: 1200.5999





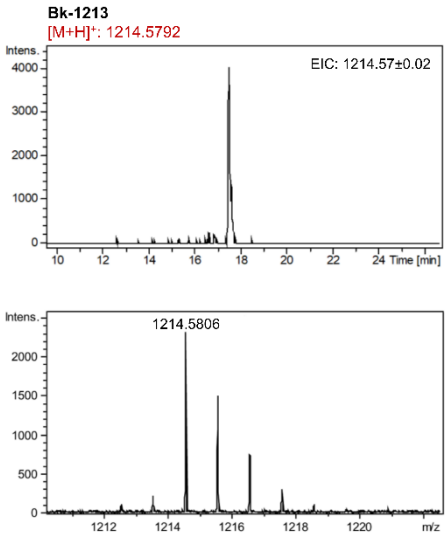


Figure 16 (f). Burkholdines (Bk-1229, Bk-1097, Bk-1215, Bk-1199 and Bk-1213)

Figure 17. LC-MS data for observed and theoretical mass(es), proposed structure based on observed mass ions, and mass spectrum for each antimicrobial detected in the screened *B. ambifaria* strains.

(a) Bactobolins; **(b)** Enacyloxin IIa **(c)** Pyrrolnitrin **(d)** Hydroxyquinolines **(e)** Cepacin A **(f)** Burkholdines. UHPLC-ESI-Q-TOF-MS analyses were performed by Dr Matthew Jenner and Prof. Gregory Challis at the University of Warwick.

Table 13. Correlation of biosynthetic gene cluster (BGC) presence and metabolite production in *B. ambifaria*[a].

***B. ambifaria* strain (clade)**	**Pyrrolnitrin**		**Burkholdines**		**Hydroxyquinolines**		**Bactobolins**		**Enacyloxin IIa**		**Cepacins[b]**	
	BGC	Metabolite	BGC	Metabolite	BGC	Metabolite	BGC	Metabolite	BGC	Metabolite	BGC	Metabolite
ATCC 53267 / BCC0284 (1d)	+	+	+	+	-	-	-	-	-	-	-	-
ATCC 53266 / BCC0338 (1d)	+	-	+	-	-	-	-	-	-	-	-	-
BC-F / BCC0203 (1b)	+	+	+	+	-	-	+	+	+	+	-	-
AMMD / BCC0207 (1b)	+	+	+	+	+	-	-	-	+	+	-	-
Ral-3 / BCC0192 (2)	+	+	-	-	+	+	+	+	-	-	+	-
J82 / BCC0191 (3)	+	+	+	+	-	-	-	-	-	-	+	+
M54 / BCC0316 (3)	+	+	+	+	-	-	-	-	-	-	+	-
Novel strains:												
BCC1100 (1a)	+	+	+	+	-	-	+	+	-	-	-	-
BCC1105 (1c)	+	+	-	-	-	-	-	-	-	-	-	-
BCC1220 (2)	+	+	-	-	+	+	-	-	-	-	+	-

[a] Grey cells highlight examples of biosynthetically inactive BGCs.

[b] The metabolite cepacin was detected prior to identification of its biosynthetic gene cluster. Phylogenetic clade of strains indicated in parentheses.

4.2.3. Regulatory systems of *B. ambifaria* BGCs

4.2.3.1. Quorum sensing regulation of BGCs in *B. ambifaria*

QS is a common pleiotropic and pathway specific regulatory system, and has been extensively described in *Burkholderia* (Duerkop *et al.* 2009; Schmidt *et al.* 2009b; Seyedsayamdost *et al.* 2010; Mahenthiralingam *et al.* 2011). Manipulation of QS associated genes in *Burkholderia* has resulted in the activation of previously silent BGCs and the consequent discovery of a novel metabolite, thailandamide lactone in *Burkholderia thailandensis* (Ishida *et al.* 2010). Access to an extensive collection of *B. ambifaria* genomes has enabled the in-depth interrogation of the QS regulatory networks within the species. This analysis focused on detecting authentic LuxR-encoding genes, *luxR*, and searching the flanking DNA sequence to identify potential *cis*-regulated operons and BGCs. Searching the *B. ambifaria* genomes for genes encoding the four co-occurring protein signatures described in the methods (Section 2.16.1), 356 *luxR* homologues were detected. Protein sequence alignment and phylogenetic analysis provided a means of de-replicating the *luxR* genes into 14 distinct phylogenetic clades (Figure 18).

4.2.3.2. Assigning BGCs or regulatory functions to luxR phylogenetic clades

Modifying the approach implemented for the de-replicated BGCs, a literature search identified known *B. ambifaria* BGCs that were described as QS regulated, and the distinct *luxR* phylogenetic clades subsequently associated with these known BGCs and cognate LuxRI systems (Figure 18). A combination of solo/orphan *luxR* homologues (10 of 14) and cognate LuxRI systems (4 of 14) were detected across the 64 *B. ambifaria* genomes. Orphan *luxR* genes were associated with the known BGCs responsible for bactobolins (17 of 64) (Seyedsayamdost *et al.* 2010) and enacyloxin IIa (6 of 64) (Mahenthiralingam *et al.* 2011). Uncharacterised BGCs identified by antiSMASH that encoded an adjacent *luxR* gene were predicted to synthesise butyrolactone (27 of 64), lantipeptide (10 of 64) and ectoine (4 of 64) metabolites. Three of the remaining orphan *luxR* gene clades could not be associated with an obvious flanking BGC or collective function of genes, *luxR* homologues from one of these unknown clades (2 of 64) were encoded in extrachromosomal DNA. The final two orphan *luxR* gene clades were positioned adjacent to each other, but divergently transcribed, and flanked several genes with predicted type 3 secretion system (T3SS) functions (Figure 18).

Homoserine lactone synthase-encoding genes, *luxI*, flanked four of the *luxR* phylogenetic clades, composing cognate LuxRI systems. Three of these systems have been previously characterised, and were highlighted in the de-replicated secondary metabolite BGCs: *cepRI*/*bafRI* (63 of 64) (Aguilar *et al.* 2003), *cepR2I2* (61 of 64) (Chapalain *et al.* 2017) and the

IOP40-10 specific *luxRI* system (1 of 64) (Figure 18). The discrepancy between the number of predicted homoserine lactone synthase *cepI2* BGCs (64 of 64) and the number of *cepR2* regulatory genes (61 of 64) in the cognate *cepR2I2* system was due to premature stop codons splitting the *cepR2* into two coding sequences in *B. ambifaria* BCC0480 and BCC1265. In both BCC0480 and BCC1265 the autoinducer-binding domain was encoded in one gene, and the remaining three protein signatures: winged helix-like DNA-binding domain, signal transduction response regulator and transcription regulator LuxR (Section 2.16.1) were encoded in an adjacent gene. Similar to BCC0480 and BCC1265, *B. ambifaria* BCC1259 encoded the autoinducer-binding domain in one gene, but the remaining protein signatures were not predicted in a coding sequence, despite being detected in the flanking DNA sequence. The lack of an annotated gene encoding the remaining three protein domains likely reflected erroneous annotation, but the lack of a single gene encoding all four domains was likely a genuine feature of the BCC1259 *cepR2I2* system caused by a premature stop codon.

The burkholdine BGC (Tawfik *et al.* 2010) was initially characterised in *B. contaminans* as the occidiofungin BGC (Gu *et al.* 2009b A; Gu *et al.* 2009a B). The occidiofungin BGC encoded two convergently expressed *luxR* genes both of which have been experimentally proven to regulate the synthesis of the antimicrobial metabolite (Gu *et al.* 2009a). However, prediction of encoded protein domains revealed an absence of the autoinducer-binding domain in both LuxR homologues (Gu *et al.* 2009a). As such, neither burkholdine LuxR protein was identified during the LuxR mining exercise of *B. ambifaria* genomes, which required the presence of this signature protein domain for a positive hit.

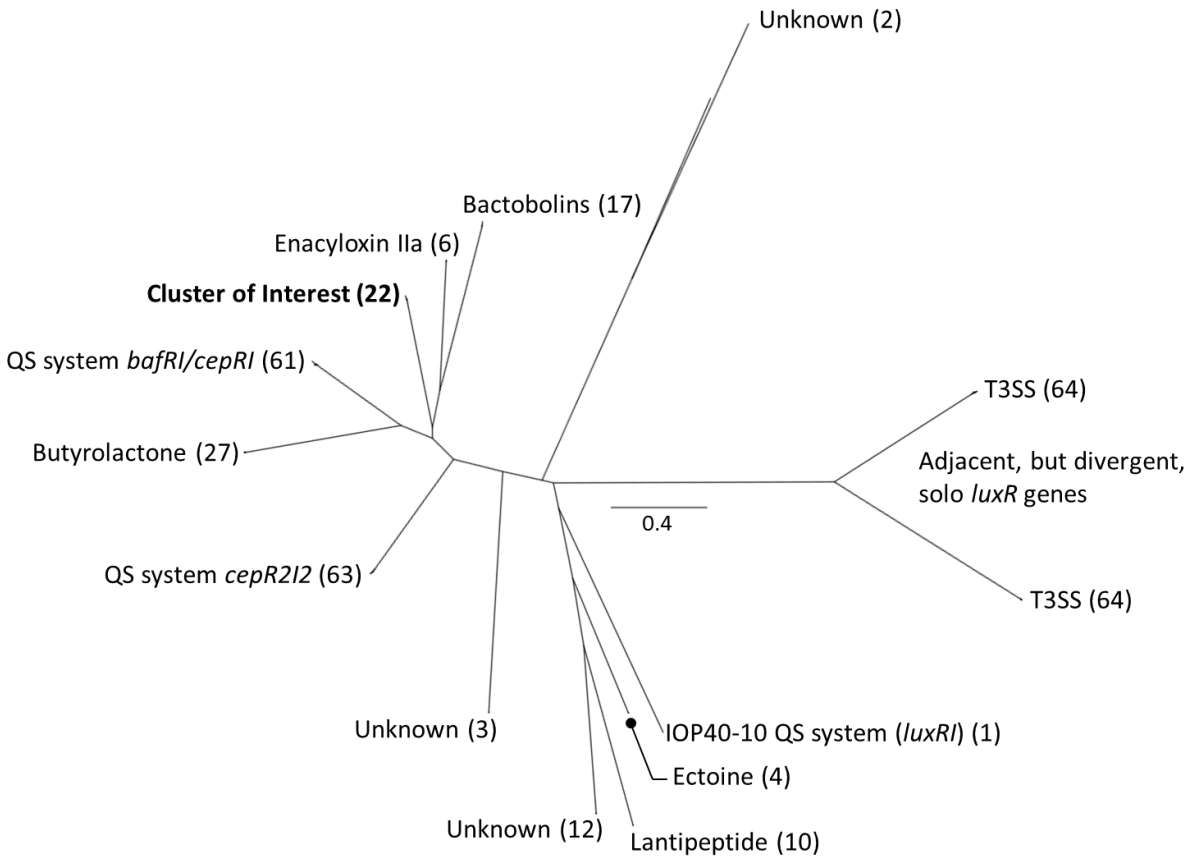


Figure 18. Unrooted phylogeny of LuxR protein homologues from 64 *B. ambifaria* strains.

Branches were labelled with characterised quorum sensing systems or putative/confirmed LuxR regulatory functions based on the literature and annotated flanking genes starting within 5kbp upstream and/or downstream of the *luxR* gene. The number of strains encoding distinct LuxR homologues is indicated in brackets. A total of 356 homologues were identified across the 64 strains, representing 14 distinct LuxR protein clades. FastTree was used to construct the approximate-maximum-likelihood phylogeny using the generalised time reversible substitution model. The evolutionary distance scale bar represents the number of base substitutions per site.

4.2.3.3. Alternative regulation of *B. ambifaria* BGCs

Excluding BGCs with previously characterised LuxR quorum sensing regulation, and cognate LuxRI systems predicted by antiSMASH (6 of 38), the remaining BGCs sequences were analysed for encoded alternative regulatory genes (Table 14). Of the 32 remaining BGCs, three uncharacterised BGCs possess LuxR-encoding genes: the fourth LuxRI system, lantipeptide BGC and butyrolactone BGC. The lantipeptide and butyrolactone BGCs also encode other regulator families, such as LysR and TetR, respectively. LysR-like family regulatory genes were the most prevalent regulatory class predicted in the *B. ambifaria* BGCs, excluding known *luxR*-regulated BGCs (15 of 32); seven of which encoded more than one LysR-like regulator. Other regulatory families predicted in the 32 BGCs included MarR, AbrB, GntR, AraC, NrdR, Xre, NtrC, MerR, TetR, IclR, NarL, Fis, Crp and two-component systems. Two BGCs lacked any detectable or known regulatory families: c1 butyrolactone 1 BGC and the NRPS 4 BGC of unknown genomic location. Rarer regulatory mechanisms detected included putatively BGC specific sigma factors in c1 NRPS 1 and c2 terpene 4; along with a putative histone-like (Hns family) encoded in the c2 arylpolyene 2 BGC.

Table 14. Regulatory genes predicted in *B. ambifaria* BGCs.

Cluster Type (antiSMASH prediction)	Local regulator	Prevalence in *B. ambifaria* (out of 64 genomes)
Replicon c1		
Terpene 1	MarR-type OmpR/EnvZ 2CS	65[c]
NRPS 1	AbrB, SF Ecf	64
Arylpolyene 1	GntR-type Unorthodox 2CS	64
NRPS 2	AraC-type	44
PKS 1	AraC, NrdR	28
Lantipeptide 1	3x LysR (LuxR)	10
Butyrolactone 1	/	1
Terpene 2	LysR, Xre, NtrC	1
Replicon c2		
Phosphonate	LysR, MerR SF RpoE	64
Terpene 3	GntR, LysR, TetR	64
Other 1	3x TetR, 2x LysR, IclR	64
Terpene 4	Unclassified SF, Classic 2CS, NarL	64
Bacteriocin 1	AraC-type	55
Arylpolyene 2	Histone Hns family	53
Ectoine 1	GntR	53
Ectoine 2	LysR	38
Other 2	2x LysR, IclR, Xre	20
Butyrolactone-otherKS	2x LysR	16
Phenazine	AraC, LysR	13
KS 2	2x LysR, GntR, AraC, Xre	7
NRPS 3	3x AraC, 2x LysR, Fis	2
Bacteriocin 2	LysR, Xre	1
Arylpolyene 3	Crp, TetR, Xre	1
Replicon c3		
Bacteriocin 3	GntR	63
Terpene 5	LysR, 2x TetR	63
PKS 3	2CS, LysR	40
Other 3	LysR	34
Butyrolactone 2	LuxR, TetR	27
Other 4	2x LysR, AraC	24
Unknown genomic location		
NRPS 4	/	1
Homoserine lactone 4	LuxR	1
(NRPS-T1PKS-*trans*-AT-PKS	2x AraC, NarL	1

4.2.4. Antimicrobial activity of *B. ambifaria* against plant and animal pathogens

Having established the presence of antimicrobial-encoding BGCs and biosynthesis of the corresponding metabolites (Table 13), antagonism activity of the 64 *B. ambifaria* strains against priority plant (Mansfield *et al.* 2012) and human pathogens (Table 3) was evaluated (Mahenthiralingam *et al.* 2011). The *in vitro* bioactivity assayed via zone of inhibition was aligned against the core-gene phylogeny to map antagonism across *B. ambifaria* and identify correlations in activity with BGC presence (). Of the 62 strains available for *in vitro* antagonism testing, 56 strains exhibited inhibitory activity against at least one of the pathogens screened; six strains lacked any observable antagonism (BCC1105, BCC1220, BCC1224, BCC1233, BCC1259 and MC40-6. Anti-fungal activity was the most widespread antimicrobial activity in *B. ambifaria*, with 81% of strains (50 of 62) exhibiting *C. albicans* antagonism. The lack of anti-fungal activity was localised to clade 2 of the core-gene phylogeny, with the exception of four strains. This correlated to the absence of the burkholdine BGC in all members of clade 2. Clade 2 possessed the least antimicrobial activity (5 of 9 strains lacked observable *in vitro* bioactivity) despite these strains encoding the pyrrolnitrin, hydroxyquinoline or cepacin BGCs. Anti-Gram-positive activity followed in prevalence with 69% of strains (43 of 62) possessing antagonism against *S. aureus*. The majority of strong anti-Gram-positive antagonism was correlated with the presence of the bactobolin BGC, the remaining strong activity was partially correlated with cepacin biosynthesis.

The majority of plant and animal pathogens screened were Gram-negative organisms. Anti-Gram-negative activity was the least prevalent antagonism observed in *B. ambifaria* with 34% of strains (21 of 62) possessing observable antimicrobial activity. Clade 1a, 1b and 1c strains exhibited substantial antagonism against at least six of the Gram-negative pathogens, with only two strains outside of these clades exhibiting similar activity: BCC0478 in clade 1d and BCC0192 in clade 2. The presence of the hybrid NRPS-PKS bactobolin and *trans*-AT PKS enacyloxin IIa BGCs correleated with the observed anti-Gram-negative activity. A distinction in activity spectrum was apparent between the bactobolin BGC and enacyloxin IIa BGC encoding strains. Bactobolin producing strains exhibited antimicrobial activity toward Betaproteobacteria pathogens, with more potent activity against *P. savastanoi* pv. *phaseolicola* and *P. syringae* pv. *syringae*. In contrast, enacyloxin IIa producing strains exhibited bioactivity against Alphaproteobacteria (*B. multivorans* and *R. radiobacter*) and Betaproteobacteria pathogens, but exhibited less potent antimicrobial activity overall.

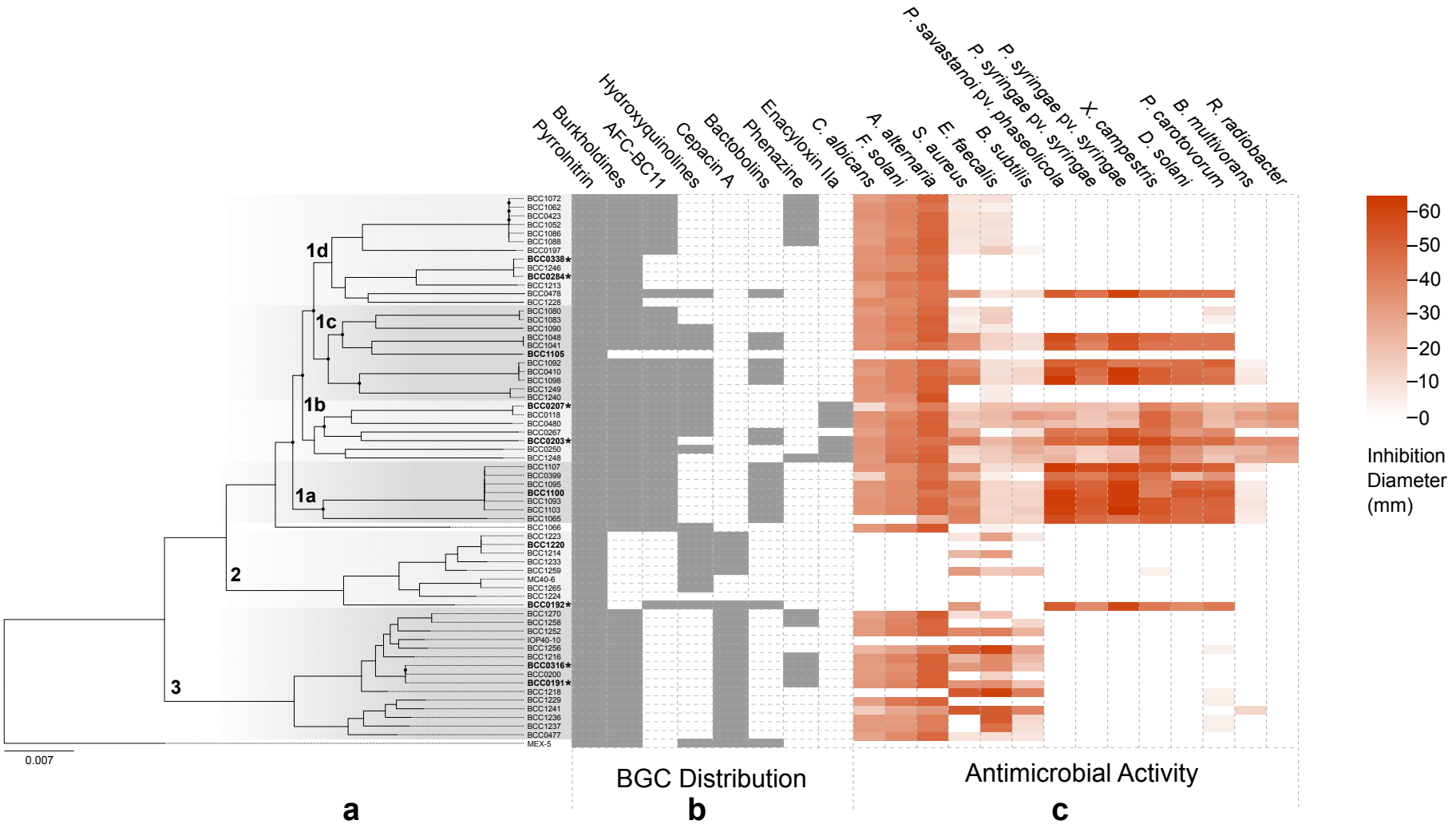


Figure 19. Core-gene phylogeny of 64 *B. ambifaria* strains with presence of antimicrobial-associated BGCs and *in vitro* pathogen antagonism.
(a) The phylogenetic tree was constructed based on 3,784 core genes identified and aligned using the software Roary. The root was determined using a secondary tree containing an outgroup species, *Burkholderia vietnamiensis* G4 (Figure 6). Six clades were defined in the phylogeny, however, strains BCC1066 and MEX-5 branched outside these clades. Strains subject to further LC-MS analysis are highlighted in bold; strains with historical biocontrol usage are indicated with an asterisk. RAxML was used to construct the maximum-likelihood phylogeny using the generalised time reversible (GTR) model with a GAMMA substitution (100 bootstraps). Nodes with bootstrap values <70% are indicated with black circles. Evolutionary distance scale bar represents the number of base substitutions/site. **(b)** The presence of the 8 characterised anti-fungal and antibiotic gene clusters: pyrrolnitrin, burkholdine, AFC-BC11, hydroxyquinolines, cepacin A, bactobolins, phenazine and enacyloxin IIa in the 64 *B. ambifaria* strains are ordered by phylogenetic position. Matrix generated using Phandango. **(c)** The antimicrobial activity of 62 *B. ambifaria* strains were defined by measuring the diameter of the zones of inhibition (mm); *n* = 2 overlays of each *B. ambifaria* strain against each susceptibility organism. Heatmap shows mean zone of inhibition. MEX-5 and IOP40-10 were not available for the antagonistism assay.

4.2.4.1. Minimum inhibitory concentration of enacyloxin IIa against plant and animal pathogens

To assess the specific sensitivities of the Gram-negative plant and animal pathogens to one of the well characterised *B. ambifaria* antimicrobials, enacyloxin IIa, broth-based minimum inhibitory concentrations (MICs) were determined (Table 15). MIC testing was performed by Dr Gordon Webster at Cardiff University. All Gram-negative pathogens possessed MICs ranging between 3.2 and 50 µg/ml. The most susceptible pathogens were *R. radiobacter*, *P. carotovorum*, *P. syringae* pv. *syringae*, and *B. multivorans* with MICs equal to or below 6.3 µg/ml. In contrast, *P. savastanoi* pv. *phaseolicola*, *P. syringae* pv. *tomato* and *X. campestris* pv. *campestris* were more tolerant of enaxyloxin IIa, all with an MIC of 50.0 µg/ml. *D. solani* possessed an intermediate MIC of 12.5 µg/ml.

Table 15. Minimum Inhibitory Concentration (MIC) of enacyloxin IIa against plant and animal pathogens[a].

Organism	MIC (µg/ml)
Rhizobium radiobacter	3.2
Pseudomonas savastanoi pv. *phaseolicola*	50.0
Dickeya solani	12.5
Pectobacterium carotovorum	6.3
Pseudomonas syringae pv. *tomato*	50.0
Xanthamonas campestris pv. *campestris*	50.0
Pseudomonas syringae pv. *syringae*	6.3
Burkholderia multivorans	6.3

[a] MICs were performed in a microbroth dilution assay using iso-sensitest broth or TSB broth with doubling-dilutions of enacyloxin IIa between 100 and 0.098 µg/ml. Bacteria were grown for 18-24 hrs and optical density measurements were taken at 600 nm. The MIC was determined by calculating the enacyloxin IIa concentration required to produce an 80% knockdown in optical density compared to the organism grown in the absence of the antibiotic.

4.3. Discussion

In this chapter the *in vitro* antimicrobial activity, specialised metabolite biosynthetic pathway potential, regulation and BGC distribution of 64 *B. ambifaria* strains was analysed. The *in vitro* expression of antimicrobial BGCs was assessed by LC-MS metabolite profiling of ten *B. ambifaria* strains, including seven previously characterised biological control strains: AMMD (BCC0207), BC-F (BCC0203), J82 (BCC0191), M54 (BCC0316), Ral-3 (BCC0192), ATCC 53267 (BCC0284) and ATCC 53266 (BCC0338). These data showed that there is considerable biosynthetic diversity and widespread antimicrobial activity within *B. ambifaria* as a species. A variety of regulatory systems are used for BGCs, and this highlighted the role of quorum sensing LuxRI regulation in antimicrobial production.

4.3.1. Widespread antimicrobial activity in *B. ambifaria*

Prior to the formal classification of *B. ambifaria* as a novel species (Coenye *et al.* 2001a), there were multiple examples in the literature of '*Pseudomonas cepacia*' strains, subsequently re-classified as *B. ambifaria*, with *in vitro* antifungal activity. Antagonism against plant fungal pathogen *Sclerotinia sclerotiorum* was observed in the characterised biocontrol *B. ambifaria* strain J82 (BCC0191) (McLoughlin *et al.* 1992). The corn rhizosphere-isolated strain ATCC 53267 (BCC0284) was screened for *in vitro* against an extensive collection of fungal pathogens including *Macrophomina phaseolina*, *Colletotrichum lindemuthianum*, *Rhizoctonia solani*, *S. sclerotiorum*, and multiple *Fusarium* spp.: *Fusarium moniliforme*, *Fusarium oxysporum* and *Fusarium graminearum* (Hebbar *et al.* 1992). *B. ambifaria* ATCC 53267 exhibited medium/strong inhibition toward most challenge fungi, with weaker activity against *F. oxysporum* and *C. lindemuthianum* (Hebbar *et al.* 1992).

The publication formally recognising *B. ambifaria* as a novel species also highlighted multiple strains that were biological control agents (Coenye *et al.* 2001a). Following *B. ambifaria* species recognition, Groenhagen *et al.* (2013) has been the only publication to analyse the antimicrobial properties of more than one *B. ambifaria* strain against multiple plant pathogens. Volatile *B. ambifaria* metabolite antagonism against *R. solani* and *A. alternata* was investigated in three strains: rhizosphere isolates LMG 17828 (BCC0338) and LMG 19182 (BCC0207) and cystic fibrosis isolate LMG 19467 (BCC0267) (Groenhagen *et al.* 2013). *B. ambifaria* Bc-F (BCC0203) was the most widely studied strain *in vitro*, with evidence of anti-fungal activity against *M. incognita, P. ultimum, R. solani, P. capsici, and F. oxysporum f.sp. lycopersici*; and anti-nematodal activity against *Meloidogyne incognita* (Li *et al.* 2002a; Roberts *et al.* 2005; Roberts *et al.* 2014). Anti-bacterial activity was only addressed for *B. ambifaria* AMMD with the discovery of enacyloxin IIa, establishing antagonism against other

Burkholderia species: *Burkholderia dolosa* and *Burkholderia multivorans*, as well as multi-drug resistant *Acinetobacter baumannii* (Mahenthiralingam *et al.* 2011). The most recent study of *B. ambifaria* focussed on *Fusarium* spp. antagonism by an Argentinian barley plant rhizosphere strain T16 (Simonetti *et al.* 2018).

The anti-fungal activity of several key *B. ambifaria* biological control, or environmental strains: BCF (BCC0203), AMMD (BCC0203), J82 (BCC0191), LMG 17828 (BCC0338) and LMG 19467 (BCC0267); in addition to recently identified *B. ambifaria* strain T16, has been extensively investigated. However, there is no holistic understanding of the shared anti-fungal antagonism shared by these agriculturally important strains, or the wider *B. ambifaria* species; and no substantial clarification of the anti-bacterial properties of *B. ambifaria*. By combining both extensive anti-bacterial screening of plant pathogens with a large, diverse collection of *B. ambifaria* strains, a broader understanding of the intrinsic antimicrobial repertoire of the species was established.

The presence of anti-fungal activity was not limited to the previously characterised *B. ambifaria* strains, but represented a widespread phenotype driven by the presence of several anti-fungal metabolites. Initially the anti-fungal screening was performed using *C. albicans*, however, the panel was expanded to include the agriculturally relevant fungi *F. solani* and *A. alternata*, which required development of a novel mycelial-based overlay method. *Burkholderia* antagonism toward *R. solani*, *A. alternata* and *C. albicans* has been attributed to the anti-fungal glycopeptide natural products burkholdines (Bisacchi *et al.* 1987; Lu *et al.* 2009; Tawfik *et al.* 2010). This is supported in the *B. ambifaria* fungal inhibition screen where anti-fungal activity was observed for all strains encoding the burkholdine BGC, with the exceptions of BCC1218, which lacked all anti-fungal activity and BCC1065, which possessed *A. alternata* antagonism. *B. ambifaria* strains lacking the burkholdine BGC, mainly clade 2 representatives, lacked all anti-fungal activity. Deletion of the third replicon c3, and consequently the burkholdine BGC, from BCC0191 resulted in a complete loss of anti-fungal activity against the three fungal pathogens screened (Figure 19a), congruent with the genetic disruption of the BGC (Gu *et al.* 2009a).

The anti-bacterial activity of *B. ambifaria* was assessed in detail for the first time, and revealed the antagonism was driven mainly by several key metabolites: bactobolins, cepacin A and enacyloxin IIa. The antimicrobial activity of enacyloxin IIa was explored in *B. ambifaria* AMMD (BCC0203) (Mahenthiralingam *et al.* 2011), however, this current work highlights the broad activity spectrum of the compound against priority plant pathogens (Mansfield *et al.* 2012). The anti-bacterial antagonism was divided into anti-Gram-positive and anti-Gram-negative.

Enacyloxin IIa-encoding strains displayed strong antagonism towards Gram-negative pathogens, as observed previously (Mahenthiralingam *et al.* 2011). Bactobolin-encoding strains possessed activity against both bacterial groups; whereas cepacin A conveyed anti-Gram-positive activity as well as anti-fungal activity, congruous with the original cepacin publication publication (Parker *et al.* 1984). Similar to the initial use of *C. albicans* as a representative fungal pathogen, *S. aureus* was used as a measure of anti-Gram-positive activity; however, the same caveats applied to *S. aureus*. The Gram-positive susceptibility panel was expanded by the inclusion of environmentally-relevant *B. subtilis* and multi-drug resistant human pathogen *E. faecalis*. Between these key anti-bacterial metabolites, in addition to the remaining antimicrobial secondary metabolites with minor bacterial antagonism, there was a widespread distribution of anti-bacterial activity within *B. ambifaria*.

A limited number of antimicrobial secondary metabolites encoded by biological control strains have been purified and their minimum inhibitory concentration determined. This is likely due to the application of the organism directly to the soil or crop seed of interest, in place of the purified compound. Cepacin A and B are the only *B. ambifaria*-encoded antimicrobials to have been subject to MIC testing, although the metabolites were extracted from a *B. diffusa* strain. (Parker *et al.* 1984). This work provides MIC data for enacyloxin IIa against all seven bacterial pathogens screened in the overlay assay. Interestingly, there were significant differences in MIC values between *Pseudomonas* spp.; with *P. savastanoi* pv. *phaseolicola* and *P. syringae* pv. *tomato* displaying an eight-fold higher MIC compared to *P. syringae* pv. *syringae*. Despite the high MIC value of several plant pathogens *in vitro*, the concentrations of these metabolites *in planta*, and the additive/synergistic interactions of the multiple antimicrobial metabolites, are unknown.

4.3.2. Secondary metabolite BGCs in *B. ambifaria*

This section discusses the systematic approach to predicting and de-replicating secondary metabolite BGCs in *B. ambifaria*, the identification of known and novel pathways, and their distribution and occurrence across the *B. ambifaria* pan-genome.

4.3.2.1. Approaches to BGC mining and dataset de-replication

Publically available informatics-based methods of predicting secondary metabolite BGCs have been established since 2003, with software capable of detecting polyketide synthase domains (Yadav *et al.* 2003). Predictive software for secondary metabolites has become widespread in the academic community, with tools available for identifying specific classes of secondary metabolites, such as bacteriocins (de Jong *et al.* 2006), Type-1 PKS (Zierep *et al.* 2017),

Type-2 PKS (Kim and Yi 2012), Type-3 PKS (Vijayan *et al.* 2011), and lassopeptides (Tietz *et al.* 2017). The concept of predicting secondary metabolite pathways has also expanded into fungal (Vesth *et al.* 2016) and plant (Kautsar *et al.* 2017) genomics. A widely applied bacterial secondary metabolite pathway prediction tool that combines multiple hidden Markov models (HMMs) and support vector machines (SVMs) to identify a variety of metabolite classes is antiSMASH, currently in its fourth iteration (Blin *et al.* 2017). During the course of this study, antiSMASH version 3 (Weber *et al.* 2015) was applied to the *B. ambifaria* genomic analysis.

The application of antiSMASH to 64 *B. ambifaria* genomes generated a large specialised metabolite BGC collection that necessitated a de-replication strategy. Previously used methods of de-replicating/clustering BGCs with high sequence similarity rely on using sequence homology as a direct comparator. Biosynthetic Genes Similarity Clustering and Prospecting Engine (BiG-SCAPE) is a publically available software (https://git.wageningenur.nl/medema-group/BiG-SCAPE) that extracts and compares the protein domains encoded in genes of BGCs, allowing the user to apply a sequence similarity threshold to separate distinct BGC clusters. This strategy has been applied to de-replicate BGCs and define the secondary metabolite potential of the genus *Rhodococcus* (Ceniceros *et al.* 2017). Clustering BGCs based on protein sequence similarity has been applied to a collection of *Pseudovibrio* genomes to generate similarity networks, and included previously characterised BGCs to identify several BGC clusters (Naughton *et al.* 2017). The de-replication strategy in this study differed from previously published studies by clustering BGCs based on k-mers generated from nucleotide sequences. The use of k-mers as opposed to direct sequence (protein or nucleotide) reduced the computing time required to generate similarity networks, enabling a faster optimisation of cut-off values.

4.3.2.2. *B. ambifaria* BGC novelty and genomic distribution

The majority of distinct BGCs identified in the 64 *B. ambifaria* genomes were uncharacterised in the literature, and included multiple PKS and NRPS pathways. This is congruent with the *Pseudovibrio* (Naughton *et al.* 2017) and *Rhodococcus* (Ceniceros *et al.* 2017) studies, which had even greater numbers of unknown BGCs; highlighting the vast unexploited secondary metabolite biosynthetic potential across multiple genera. The existence of several completed *B. ambifaria* genomes permitted the scaffolding of the remaining *B. ambifaria* genomes into replicons. This provided an insight into the genomic distribution, variability and density of BGCs across the multi-replicon *B. ambifaria* genome, a rarely encountered aspect of BGCs in the literature with regard to other genera and species. The metabolites bactobolins have not previously been described as an antimicrobial associated with *B. ambifaria*. The bactobolins

BGC was initially discovered in *B. thailandensis*, with homologous clusters detected in closely related species *B. pseudomoallei* (Tawfik *et al.* 2010). The remaining *Burkholderia* metabolites encoded by *B. ambifaria* strains were either discovered in the species or discovered elsewhere and homologous BGCs identified in *B. ambifaria*. Variation was observed in the antimicrobial pathway content of the seven previously characterised biological control *B. ambifaria* strains. All *B. ambifaria* biocontrol strains possessed two-to-four antimicrobial pathways. Approximately 59% of the 64 genome collection encoded four or more antimicrobial-related BGCs reflecting a plausible redundancy in biological control function and a multi-layered mechanism of inhibiting plant pathogens.

4.3.3. Regulatory systems of BGCs

A plethora of regulatory genes and systems were identified across the 38 distinct *B. ambifaria* BGCs (Table 14). Understanding the regulation of uncharacterised secondary metabolite BGCs can provide a valuable insight into inducing or enhancing expression, and has been previously demonstrated in *Burkholderia* (Ishida *et al.* 2010). Detecting regulatory genes can also be exploited to identify unconventional BGCs, such as the cepacin BGC that was discovered by association with a *luxR* regulatory gene (Mullins *et al.* 2019). Understanding the diversity of LuxR-mediated regulation only became apparent through multi-genomic analyses. This technique has been exploited previously to analyse the genus *Burkholderia* (Choudhary *et al.* 2013), the family Sphingomonadaceae (Gan *et al.* 2014), and bacteria as a whole (Hudaiberdiev *et al.* 2015; Subramoni *et al.* 2015).

4.3.4. Detecting antimicrobial metabolites *in vitro*

Coupling genomic prediction of BGCs to metabolite detection *in vitro* provided information about the expression conditions of the *B. ambifaria* antimicrobial compounds. However, difficulties arise if the metabolite of interest lacks a known molecular and structural formula, preventing the use of mass spectrometry for detection. This issue was encountered with the AFC-BC11 lipopeptide. Acetonitrile and ethyl acetate were successful in capturing the *B. ambifaria* specialised metabolites; however, depending on the polarity of the metabolite, the use of a variety of solvents may be necessary to capture the entire secondary metabolic diversity of a given strain. Mass spectrometry of multiple strains identified several examples of antimicrobial BGCs that appeared biosynthetically inactive under growth conditions that stimulated production in other strains. This highlights the variation in the regulatory system background of each strain, and the importance of understanding that these regulation differences can precipitate variable expression profiles for the same BGC.

4.4. Conclusions

The main conclusions of the systematic analysis of secondary metabolite BGCs and *in vitro* antimicrobial activity were as follows:

1) Significant and widespread antimicrobial activity existed across the 62 *B. ambifaria* strains, exhibiting antagonism toward multiple bacterial, fungal and oomycete plant pathogens.

2) Pan-genomic analysis of secondary metabolite BGCs highlighted the previously unknown large secondary metabolite potential of the species, in addition to the distribution of known antimicrobial-encoding BGCs across the species phylogeny.

3) The prediction of LuxR regulatory proteins across the 64 *B. ambifaria* genomes led to the discovery of the biosynthetic origin of antimicrobial metabolites cepacins.

5. Distribution and diversity of bacterial polyyne BGCs

The data and results shown in this chapter represent my own independent work/investigation, except where otherwise stated. Research and methods pertaining to this chapter have been published in part in the following manuscript. Contributions from the authors are acknowledged where appropriate, and wording has been adapted where necessary.

Mullins AJ, Murray JAH, Bull MJ, Jenner M, Jones C, Webster G *et al.* Genome mining identifies cepacin as a plant-protective metabolite of the biopesticidal bacterium *Burkholderia ambifaria*. *Nature Microbiology* 2019; 4: 996–1005. DOI: 10.1038/s41564-019-0383-z

5.1. Introduction

Polyynes are a class of compound characterised by the alternating single-triple carbon bond moiety (polyalkyne chain) within the structure. Natural polyyne compounds have been described characterised in plants, fungi, and marine corals, in addition to a single example from an insect (Shi Shun and Tykwinski 2006). A terminal alkyne bond within polyyne metabolites has been suggested as the source of their bioactive properties (Yamaguchi *et al.* 1995). However, more recent experiments have demonstrated that antimicrobial activity is retained when the terminal alkyne bond is converted into a triazole moiety to stabilise the structure (Ross *et al.* 2014). The polyalkyne chain is created from the consecutive desaturation of an aliphatic chain, first with desaturases to form carbon-carbon double bonds, and followed by acetylenases to form the triple bonds (Ross *et al.* 2014).

The first bacterial polyynes were discovered in *Pseudomonas cepacia* (Figure 20), later re-classified as *Burkholderia diffusa*, however, the biosynthetic origin of the polyynes cepacins A and B remained unknown (Parker *et al.* 1984). This discovery was soon followed by caryoynencins in *Pseudomonas caryophylli* (re-classified as *Trinickia caryophylli*) (Kusumi *et al.* 1987) and Sch 31828 in *Microbiospora* sp. (Patel *et al.* 1988), both of which lacked an associated genetic basis. The polyyne collimomycin synthesised by *Collimonas fungivorans* was the first bacterial polyyne to have a biosynthesis gene cluster linked to its production (Kai *et al.* 2018). The biosynthetic logic of bacterial polyyne biosynthesis was not understood until the identification of the biosynthetic gene cluster responsible for caryoynencin (Ross *et al.* 2014). More recently, the *C. fungivorans*-derived polyynes have been structurally characterised and unified under the name collimonins (Kai *et al.* 2018).

Recent developments in bacterial polyynes have involved genomics and searches for orthologous BGCs. Despite sequence similarity between characterised BGCs and purportedly orthologous BGCs, these pathways represent distinct BGCs that synthesise structurally different metabolites. Following the identification of the BGC responsible for cepacin biosynthesis in *B. ambifaria*, we sought to understand the distribution of polyyne BGCs within *Burkholderia* and more broadly in bacteria. In addition, the distribution and diversity of other polyyne-like BGCs was explored in the publically available genomic sequences.

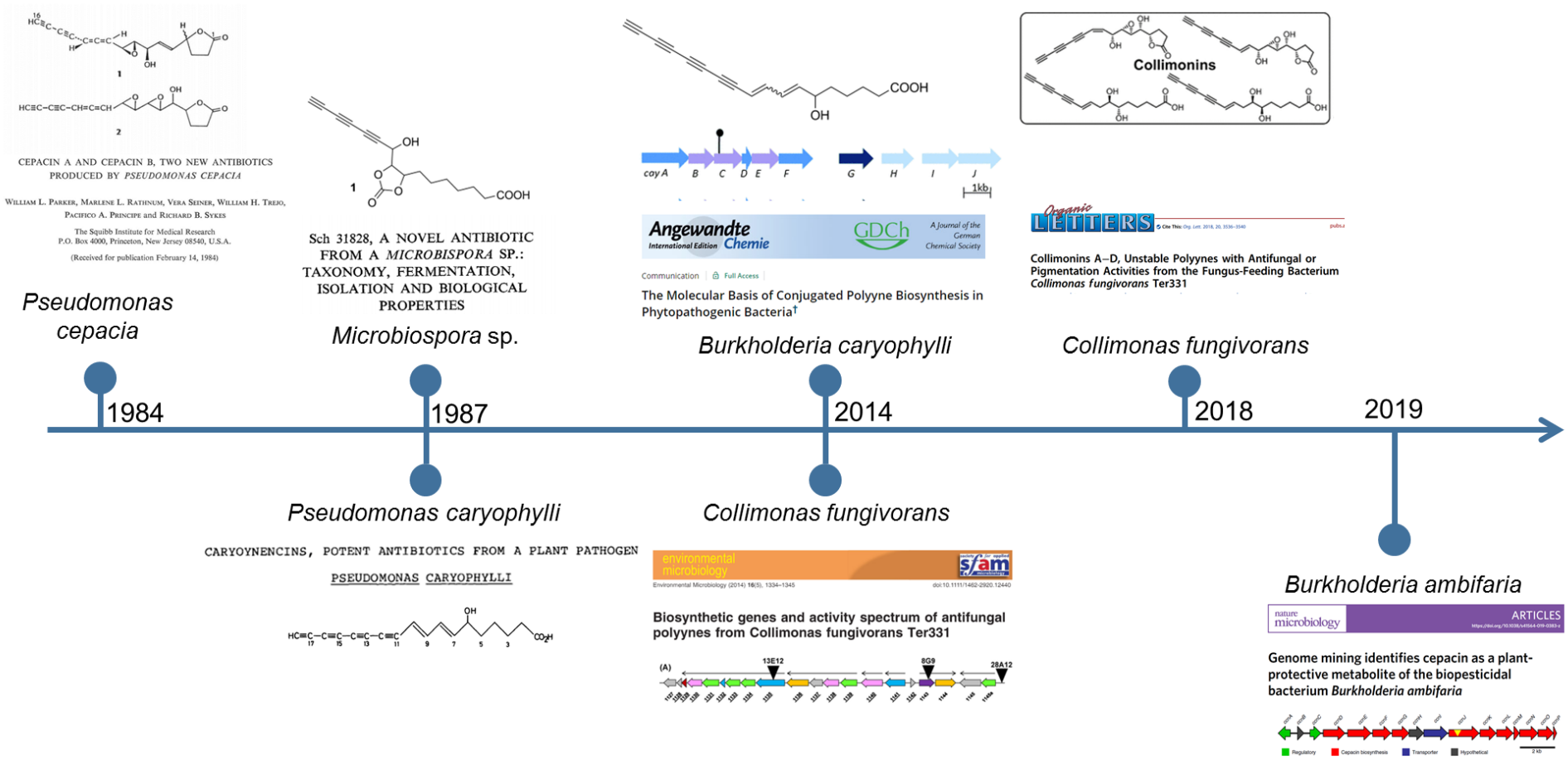


Figure 20. Timeline displaying publication related to the discovery of bacterial polyyne metabolites and associated BGCs.

The first publication on bacterial polyynes were the metabolites cepacins A and B in *B. dffusa* (formerly *Pseudomonas cepacia*) in 1984. A total of four bacterial polyyne metabolites and their respective BGCs were characterised between 1984 and 2019.

5.1.1. Aims and objectives

This chapter aims to characterise the *B. ambifaria* polyyne cepacin, and understand the wider distribution of bacterial polyyne BGCs within *Burkholderia* and other bacterial genera through the following objectives:

1) Characterise the genetic locus responsible for cepacin biosynthesis.

2) Elucidate the structure of cepacin using mass spectrometry.

3) Determine the influence of temperature on the regulation of cepacin biosynthesis.

4) Characterise the antimicrobial activity associated with cepacin production.

5) Determine the distribution of the cepacin BGC and other bacterial polyynes, and compare their genetic loci.

5.2. Results

5.2.1. Identifying an uncharacterised *B. ambifaria* BGC

5.2.1.1. Uncharacterised LuxRI system adjacent to previously unknown *Burkholderia* BGC

In addition to the three characterised cognate LuxRI systems *cepRI*/*bafRI* (Flórez *et al.* 2017), *cepR2I2* (Chapalain *et al.* 2017) and the IOP40-10 specific *luxRI* system; a fourth system was highlighted in both the secondary metabolite BGC de-replication (Figure 15) and LuxR mining analysis (Figure 18). This fourth LuxRI system was predicted in 22 of the 64 *B. ambifaria* strains and was encoded upstream of a conspicuous BGC not directly detected by antiSMASH 3.0, but indirectly via the homoserine lactone synthase. The BGC measured 16.6 kbp and encoded 16 genes, including the LuxRI regulatory system and several enzymes of interest, such as multiple fatty acid desaturases, a beta-ketoacyl synthase and a PKS synthase (Figure 21a).

5.2.1.2. Insertional mutagenesis to determine antimicrobial properties of uncharacterised *Burkholderia* BGC

The largest, central gene in the BGC, the fatty acyl-adenosine monophosphate (AMP) ligase-encoding gene, *ccnJ*, was targeted for insertional mutagenesis to disrupt the expression of the BGC (Figure 21a). Six *B. ambifaria* strain backgrounds (BCC0191, BCC0477, BCC1252, BCC1241, BCC1259 and BCC1218) were used during insertional mutagenesis. The wild-type and cepacin-deficient derivatives (::*ccnJ*) of each strain background were tested in the traditional antimicrobial overlay assay against *S. aureus*, *C. albicans* and *P. carotovorum* (Figure 21b). No difference was observed in the zone of inhibition diameters against *C. albicans*, *F. solani* and *A. alternata*, however, the mutation abolished strong anti-Gram-positive activity and weak anti-Gram-negative activity in all strain backgrounds (Figure 21b). All strain backgrounds were screened for activity against the fungus-like oomycete *P. ultimum*. Antagonistic activity against *P. ultimum* was considerably reduced in the insertional mutants compared to the wild-type backgrounds.

5.2.1.3. Identifying the metabolite synthesised by the unknown *Burkholderia* BGC

Comparative high resolution mass spectrometry of metabolite extracts of *B. ambifaria* BCC0191 WT and BCC0191::*ccnJ* identified ions with *m/z* = 271.0964 with a predicted molecular formula of $C_{16}H_{14}O_4$ in the wild-type that were absent in BCC0191::*ccnJ* metabolite extract (Figure 21c). The molecular formula $C_{16}H_{14}O_4$ and activity spectrum of BCC0191 were indicative of the historically described *Burkholderia* polyyne, cepacin A (Figure 21d) (Parker *et al.* 1984). Literature searches failed to identify a defined BGC for cepacin A or B. The cepacin BGC was encoded on the second replicon of 22 *B. ambifaria* strains, with 100% and 56% presence in clade 3 and clade 2, respectively (Figure 19b).

5.2.1.4. Comparison of original cepacin-producer to *B. ambifaria*

The original characterised producer of cepacin A and cepacin B, *P. cepacia* SC 11783 (Parker *et al.* 1984) was obtained from the Belgian Co-ordinated Collection of Microorganisms (BCCM). The strain has since been taxonomically re-classified as *Burkholderia cepacia*, and finally as *Burkholderia diffusa* (LMG 24093; ATCC 39356). Direct comparison of metabolites extracted from LMG 24093 and BCC0191 confirmed the identity of cepacin A in *B. ambifaria* (Figure 23). The only sequence data available for the strain was a partial *hisA* gene, and was likely used for the re-classification of *P. cepacia*. The genome of LMG 24093 was therefore sequenced to confirm species classification, and to include the original cepacin producer in further analyses of the polyyne pathway.

5.2.1.5. *B. ambifaria* cepacin temperature-dependent regulation

Cepacin-encoding *B. ambifaria* strains were phylogenomically confined to the core-gene phylogeny clades 2 and 3, with the exception of MEX-5 (Figure 19b). Of the 22 cepacin-encoding strains, MEX-5 and IOP40-10 were not available for phenotypic testing. Biosynthetically active cepacin BGCs were assessed through the presence of strong anti-Gram-positive activity via the classical overlay assay. *B. ambifaria* was incubated at the standard temperature of 30°C, in addition to the lower, more environmentally relevant temperature of 22°C. *B. ambifaria* strains BCC1220, BCC1229 and BCC1233 lacked any anti-Gram-positive activity at both 22°C and 30°C (Figure 24). The majority of biosynthetically active cepacin-producer strains (12 of 17) exhibited larger zones of clearing (10+ mm increased diameter) against *S. aureus* at 22°C compared to 30°C (Figure 24). The remaining five strains exhibited zones of clearing within +/-9 mm diameters at both temperatures. No obvious phylogenetic strain distribution of this temperature-dependent trait existed in the core-gene phylogeny, with closely related strains exhibiting different responses in cepacin production with temperature change as observed through zones of inhibition.

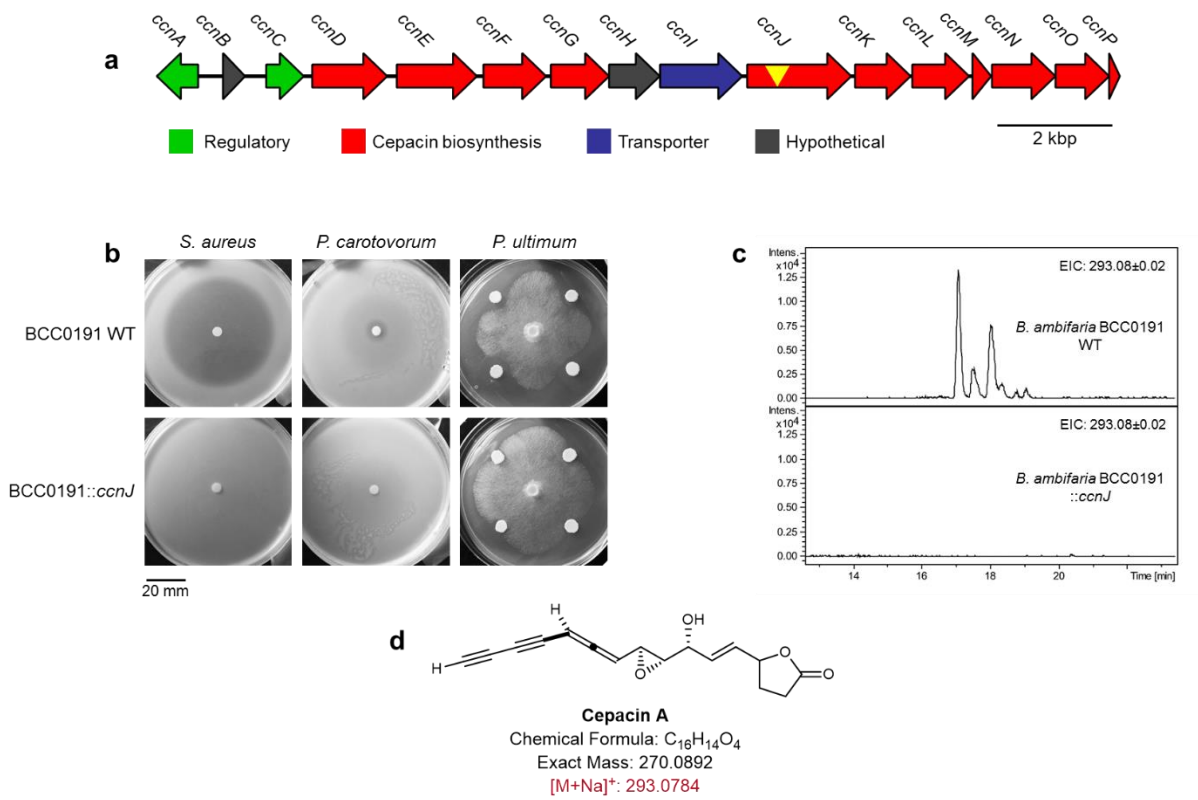


Figure 21. Gene architecture of cepacin A BGC, antimicrobial screening of *B. ambifaria* wild-type and cepacin-deficient derivative, and LC-MS analysis.
(a) Organisation and identity of the genes within the cepacin A pathway. The insertion site of the vector used during mutagenesis is highlighted by the inverted yellow triangle. **(b)** Zones of inhibition against S. aureus NCTC 12981, *P. carotovorum* LMG 2464 and *P. ultimum* Trow var. *ultimum* MUCL 16164 by cepacin producer *B. ambifaria* BCC0191 WT and insertional mutant (::*ccnJ*). Scale bar represents 20 mm. All inhibition images were processed by decreasing brightness 20% and increasing contrast 20%. **(c)** Extracted ion chromatograms at m/z = 293.08 ± 0.02, corresponding to [M + Na]+ for cepacin A, from LC-MS analyses of crude extracts from agar-grown cultures of BCC0191 WT (top) and the BCC0191::*ccnJ* mutant (bottom); *n* = 3 independent LC-MS analyses of WT and mutant cultures. **(d)** Structure of cepacin A confirmed by comparison of mass ions between *B. ambifaria* and authentic standard derived from *B. diffusa* LMG 24093. LC-MS analyses were performed by Dr Matthew Jenner and Prof. Gregory Challis at the University of Warwick.

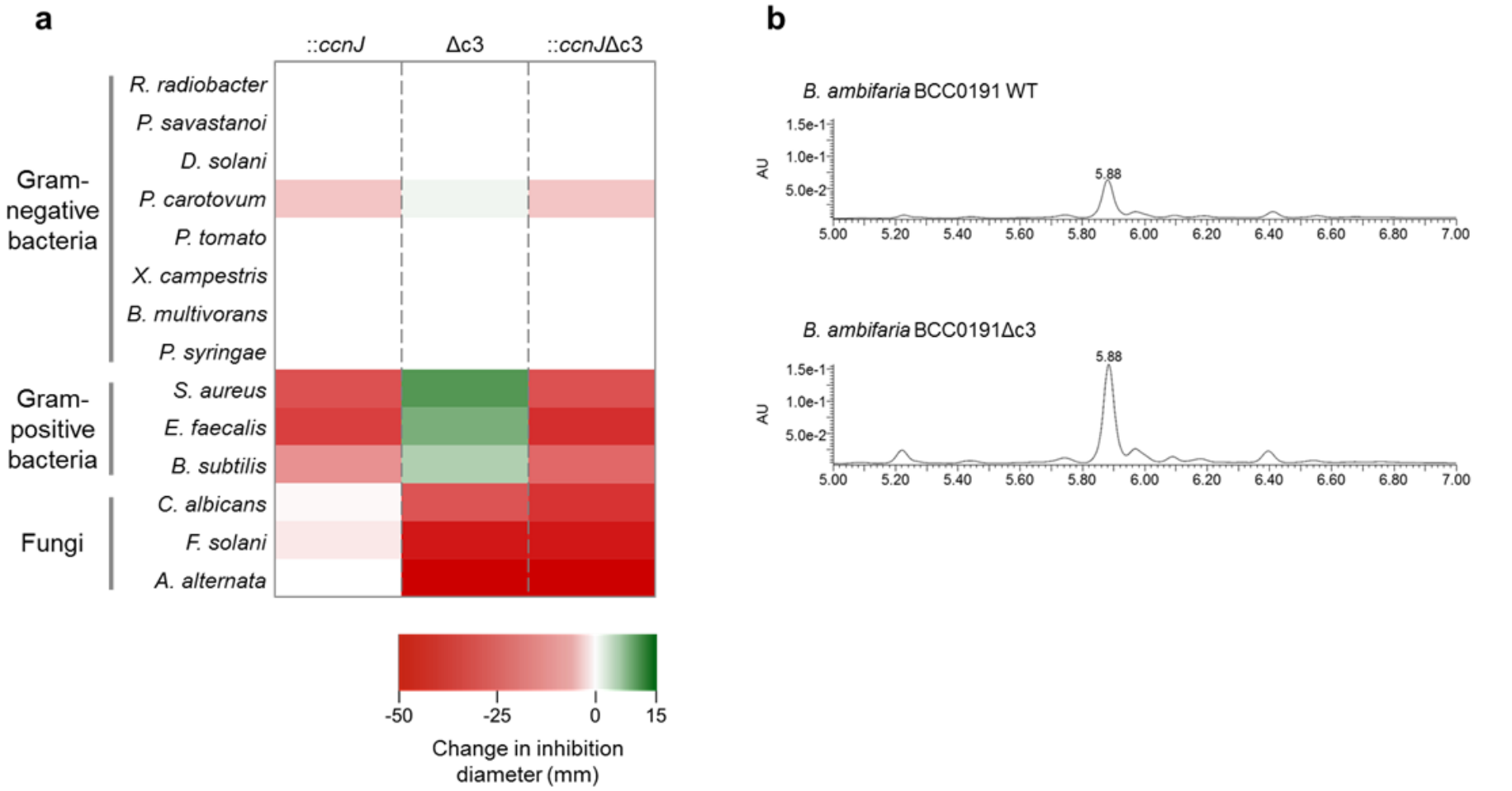


Figure 22. Impact of mutations on *B. ambifaria* BCC0191 antimicrobial activity and cepacin production.
(a) Three mutants were compared to BCC0191 wild-type (WT): cepacin-deficient derivative (::ccnJ), third replicon knockout (Δc3), and a combined mutation (::ccnJΔc3). n = 2 independent overlays per condition, and the mean average used to generate the heat map. Scale represents change in zone of inhibition diameter (mm) compared to BCC0191 WT (red = reduced zone, white = no change, green = increased zone). (b) HPLC chromatograms at 260 nm of *B. ambifaria* BCC0191 WT and BCC0191Δc3 (n = 6 independent HPLC analyses per strain) highlighting the impact of third replicon deletion on cepacin production. HPLC analyses performed by Dr Gordon Webster at the Cardiff University.

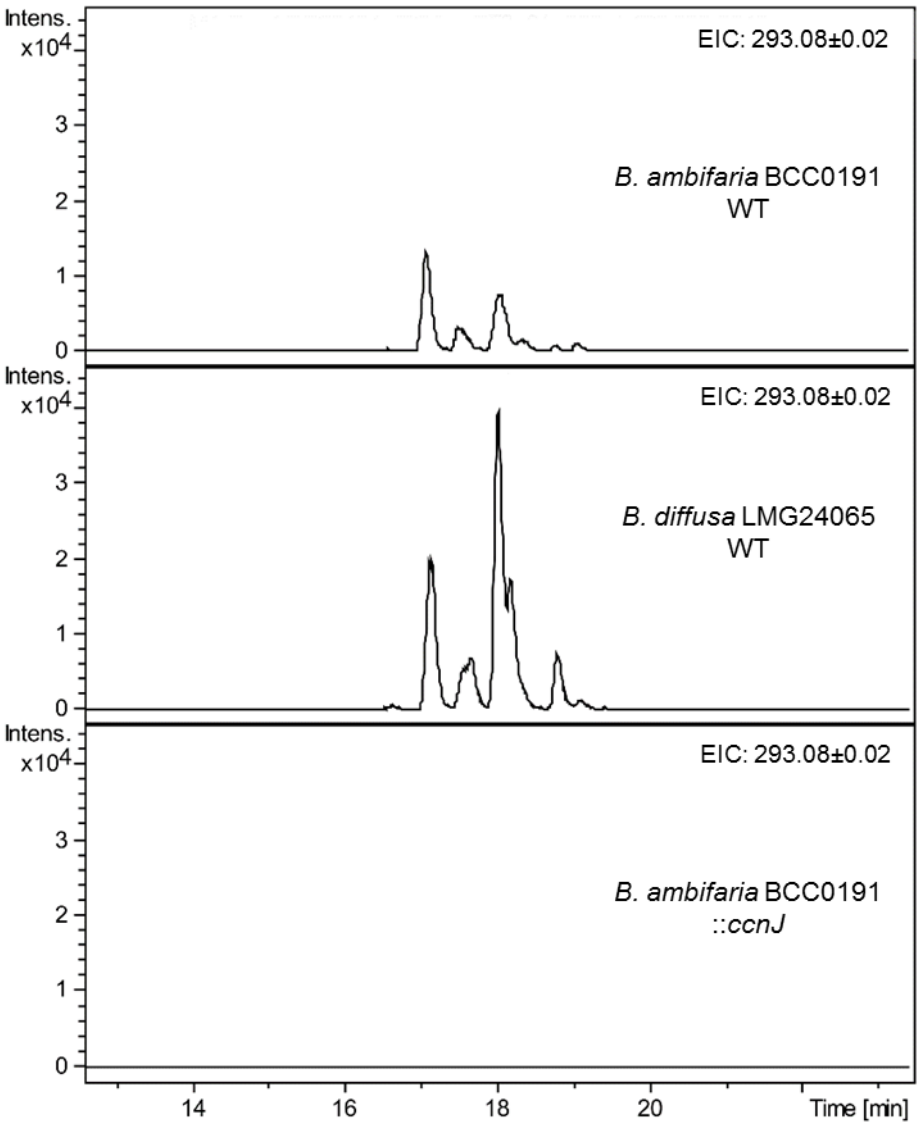


Figure 23. LC-MS comparison of original characterised cepacin-producer *B. diffusa* LMG 29043, and *B. ambifaria* BCC0191 WT and ::*ccnJ* derivative.
Extracted ion chromatograms at m/z 293.08 ± 0.02, corresponding to the $[M + Na]^+$ ion of cepacin A, from LC-MS analyses of crude extracts from agar-grown cultures of *B. ambifaria* BCC0191 wild-type (top), *B. diffusa LMG 29043* (middle) and *B. ambifaria* BCC0191::*ccnJ* (bottom). Cepacin A was produced by both *B. ambifaria* BCC0191 and *B. diffusa LMG 29043*, but was absent in the insertional mutant BCC0191::*ccnJ,* as noted previously. LC-MS analyses were performed by Dr Matthew Jenner at the University of Warwick.

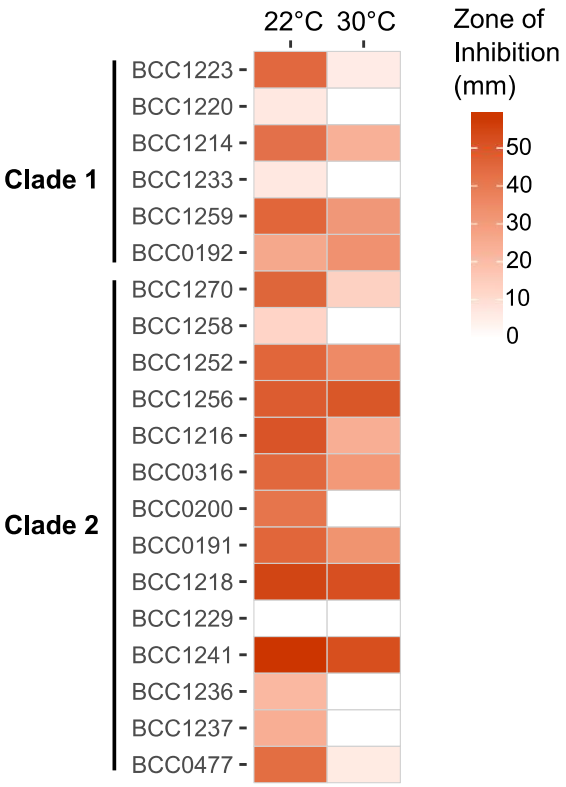


Figure 24. Temperature-dependent antagonism towards Gram-positive bacteria by *B. ambifaria* strains with cepacin BGC.

B. ambifaria strains possessing the cepacin BGC were grown at 22°C and 30°C, and overlaid with soft agar seeded with *S. aureus*. The resulting zones of inhibition were measured (n = 3).

5.2.2. *Burkholderia* polyyne BGCs

5.2.2.1. Comparison of cepacin BGC to similar polyyne BGCs

Two other known polyyne BGC were characterised previously in bacteria: collimomycin (Fritsche *et al.* 2014) in *Collimonas fungivorans* and caryoynencin (Ross *et al.* 2014) in *Burkholderia caryophylli*. The collimomycin BGC possesses the highest similarity to the cepacin BGC, whereas only a collection of purportedly core biosynthetic genes were shared between the caryoynencin BGC and cepacin BGC. The collimomycin BGC spanned 21.3 kbp, but only encoded 12.9 kbp of homologous sequence at 76.9% nucleotide similarity; the remaining 8.4 kbp encoded non-homologous regions (Figure 25). Twenty genes were encoded in the collimomycin BGC, but only thirteen genes possessed homology to those encoded in the cepacin BGC. The regulatory system associated with each pathway was one of the key differences between the two BGCs; the cepacin BGC encoded a cognate LuxRMI system upstream of the biosynthetic genes, whereas the collimomycin BGC encoded a LysR-type transcriptional regulator near the start of the pathway. Other genes encoded by the *Collimonas* BGC but absent in the *Burkholderia* BGC were four hypothetical genes, a major-facilitator superfamily transporter and fatty acid desaturase (Figure 25). Similar proteins to those identified in the cepacin BGC were identified in other *Burkholderia* species (Table 16). *C. fungivorans* Ter331 has since been further characterised and shown to synthesise multiple polyyne compound derivatives, collectively known as collimonins (Kai *et al.* 2018).

The second polyyne pathway with similarities to the *Burkholderia* cepacin BGC was the caryoynencin BGC, synthesised and encoded historically by *B. caryophylli* (Kusumi *et al.* 1987; Ross *et al.* 2014); but more recently characterised in *B. gladioli* (Flórez *et al.* 2017). Seven genes exhibited nucleotide homology between the pathways, including three desaturase-type genes, now known to mediate bacterial polyyne biosynthesis (Ross *et al.* 2014), a fatty acyl-AMP ligase, acyl-carrier protein and rubredoxin-encoding gene (Figure 26). No *cis*-regulatory genes were predicted in the caryoynencin gene cluster, compared to the proposed QS regulation of the cepacin BGC.

5.2.2.2. Polyyne BGC species distribution and horizontal gene transfer

To determine the prevalence of polyyne BGCs a BLAST analysis for the cepacin, collimomycin, caryoynencin and *Pseudomonas* polyyne BGCs in the publicly available GenBank *Burkholderia*, *Paraburkholderia*, *Caballeronia, Collimonas, Pseudomonas* and *Mycobacterium* genomes, and in-house assembled *Burkholderia* and *Paraburkholderia* genomes was undertaken, and predicted multiple examples of the pathways. Homology searches identified 387 similar pathways in a least 17 named species, and multiple other unclassified species within *Burkholderia* and *Pseudomonas*. To understand the relationship between the different polyyne BGCs a phylogenetic tree was constructed using the conserved fatty acyl-AMP ligase protein sequence (Figure 27). The protein sequence was extracted from 388 complete or partial BGCs spanning six genera. Distinct clades were observed for each proposed polyyne BGC. The tree topology and branching pattern suggested the cepacin and collimoin BGCs separated evolutionarily from the caryoynencin BGC, which was in agreement with the broader gene organisation and synteny observed between the BGCs (Figure 26). The caryoynencin BGC was evolutionarily closer to the proposed *Pseudomonas* polyyne BGC, despite all caryoynencin encoding strains originating from the *Burkholderia* or *Paraburkholderia* genera. Similarly, the cepacin BGC was encoded by *Burkholderia*, *Paraburkholderia* and *Caballeronia* species, but possessed greater sequence similarity to the collimonin BGC, rather than the caryoynencin BGC (Figure 26).

Table 16. Proteins with similarity to the cepacin A biosynthetic gene cluster.

Gene Name	Length bp/aa	Similar Proteins (excluding *B. ambifaria*)	% aa Sequence Identity
ccnA	714/237	LuxR family transcriptional regulator [*Burkholderia* vietnamiensis] LuxR family transcriptional regulator [*Burkholderia* contaminans] LuxR family transcriptional regulator [*Burkholderia* sp. USM B20]	82% 83% 82%
ccnB	384/127	Hypothetical Protein [*Burkholderia* sp. LA-2-3-30-S1-D2] Hypothetical Protein [*Burkholderia* sp. RF2-non_BP3] Hypothetical Protein [*Burkholderia* sp. RF4-BP95]	90% 87% 88%
ccnC	654/217	GNAT family N-acetyltransferase [*Burkholderia* sp. LA-2-3-30-S1-D2] GNAT family N-acetyltransferase [*Burkholderia* sp. RF4-BP95] GNAT family N-acetyltransferase [*Burkholderia* contaminans]	88% 87% 87%
ccnD	1293/430	beta-ketoacyl-ACP synthase II [*Burkholderia* sp. LA-2-3-30-S1-D2] beta-ketoacyl-ACP synthase II [*Burkholderia* sp. RF4-BP95] beta-ketoacyl-ACP synthase II [*Burkholderia* sp. RF2-non_BP3]	95% 94% 94%
ccnE	1386/461	NAD(P)/FAD-dependent oxidoreductase [*Burkholderia* contaminans] NAD(P)/FAD-dependent oxidoreductase [*Burkholderia* vietnamiensis] NAD(P)/FAD-dependent oxidoreductase [*Burkholderia* vietnamiensis]	88% 85% 84%
ccnF	1083/360	fatty acid desaturase [*Burkholderia* sp. LA-2-3-30-S1-D2] fatty acid desaturase [*Burkholderia* sp. RF4-BP95] fatty acid desaturase [*Burkholderia* vietnamiensis]	92% 92% 89%
ccnG	1005/334	aromatic ring-hydroxylating dioxygenase subunit alpha [*Burkholderia* contaminans] aromatic ring-hydroxylating dioxygenase subunit alpha [*Burkholderia* sp. LA-2-3-30-S1-D2] Rieske (2Fe-2S) domain protein [*Burkholderia* vietnamiensis G4]	95% 94% 94%
ccnH	876/291	hypothetical protein [*Burkholderia* contaminans] hypothetical protein [*Burkholderia* sp. LA-2-3-30-S1-D2] hypothetical protein [*Burkholderia* sp. RF2-non_BP3]	86% 82% 81%
ccnI	1419/472	MFS transporter [*Burkholderia* sp. LA-2-3-30-S1-D2] MFS transporter [*Burkholderia* sp. RF4-BP95] MFS transporter [*Burkholderia* sp. RF2-non_BP3]	90% 90% 90%
ccnJ	1797/598	AMP-dependent synthetase [*Burkholderia* contaminans] AMP-dependent synthetase [*Burkholderia* sp. RF4-BP95] AMP-dependent synthetase [*Burkholderia* vietnamiensis]	88% 86% 85%
ccnK	960/319	acyl-CoA desaturase [*Burkholderia* contaminans] acyl-CoA desaturase [*Burkholderia* sp. RF4-BP95] acyl-CoA desaturase [*Burkholderia* sp. LA-2-3-30-S1-D2]	92% 93% 92%
ccnL	984/327	acyl-CoA desaturase [*Burkholderia* contaminans] acyl-CoA desaturase [*Burkholderia* sp. RF4-BP95] acyl-CoA desaturase [*Burkholderia* sp. RF2-non_BP3]	96% 95% 94%

ccnM	321/106	polyketide synthase [*Burkholderia* sp. RF2-non_BP3] polyketide synthase [*Burkholderia* sp. LA-2-3-30-S1-D2] polyketide synthase [*Burkholderia* vietnamiensis]	96% 95% 95%
ccnn	1098/365	fatty acid desaturase [*Burkholderia* contaminans] fatty acid desaturase [*Burkholderia* stagnalis] fatty acid desaturase [*Burkholderia* stagnalis]	94% 92% 91%
ccnO	924/307	delta-12-desaturase [*Burkholderia* sp. LA-2-3-30-S1-D2] delta-12-desaturase [*Burkholderia* vietnamiensis] delta-12-desaturase [*Burkholderia* contaminans]	87% 86% 87%
ccnP	186/61	rubredoxin [*Burkholderia* vietnamiensis] rubredoxin [*Burkholderia* contaminans] rubredoxin [*Burkholderia* stagnalis]	95% 93% 92%

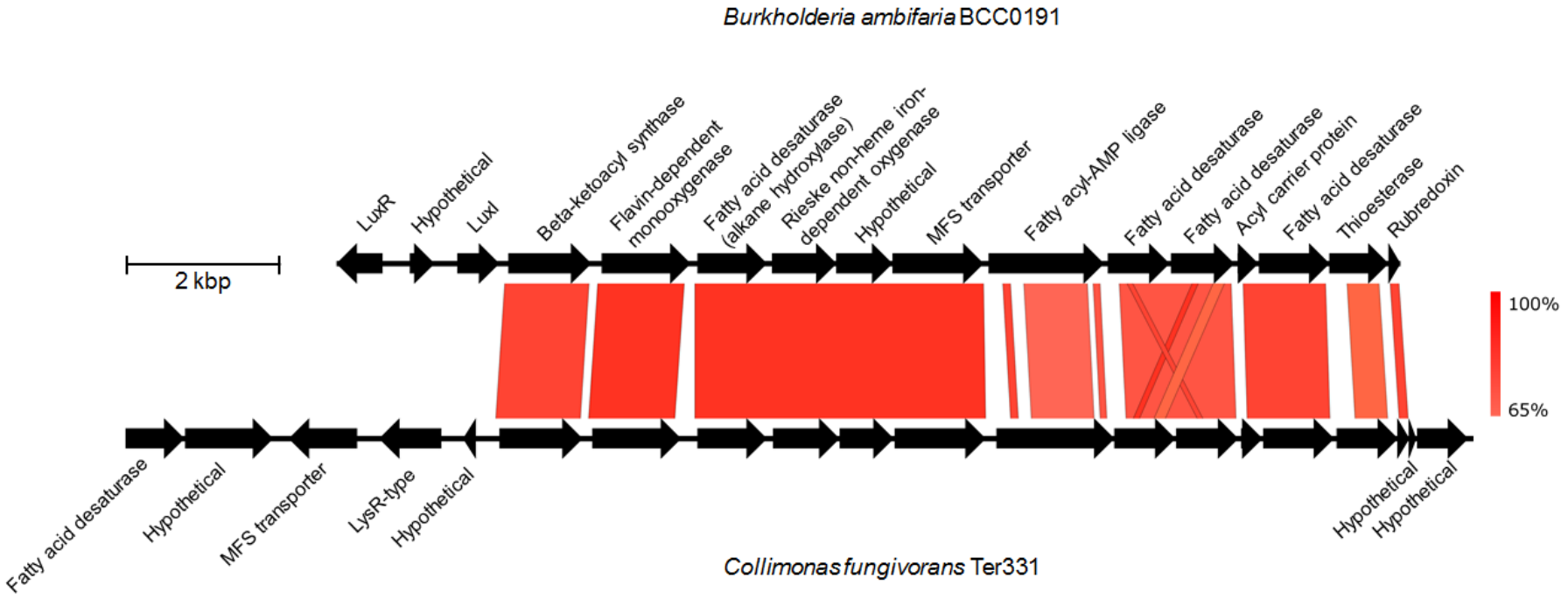


Figure 25. Comparison of the gene organisation between *Burkholderia* cepacin A BGC and *Collimonas* collimomycin BGC.

The predicted core biosynthesis genes are highly conserved between the BGCs. The *Collimonas* BGC possessed several additional genes, mostly of hypothetical function, and the regulatory component upstream of the core biosynthetic genes differs between the two species.

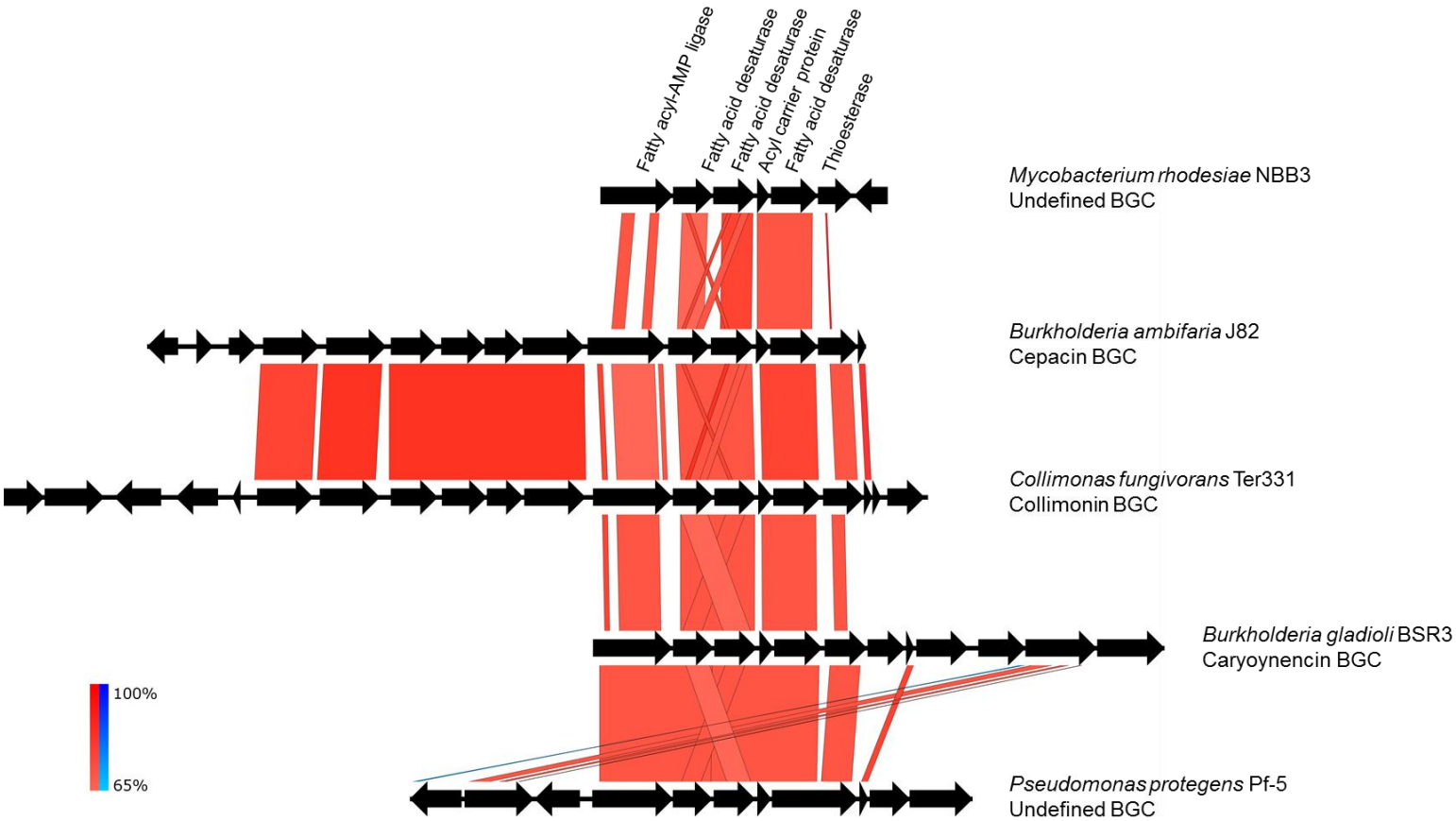


Figure 26. Comparison of the gene organisation between characterised and putative polyyne BGCs.

Polyyne BGCs have been described in *B. gladioli*, *B. ambifaria* and *C fungivorans*, while putative polyyne BGCs with sequence homology to the aforementioned BGCs have been identified in *M. rhodesiae* and *P. protegens*. Core biosynthetic genes shared by almost all polyyne BGCs are highlighted at the top of the comparison figure.

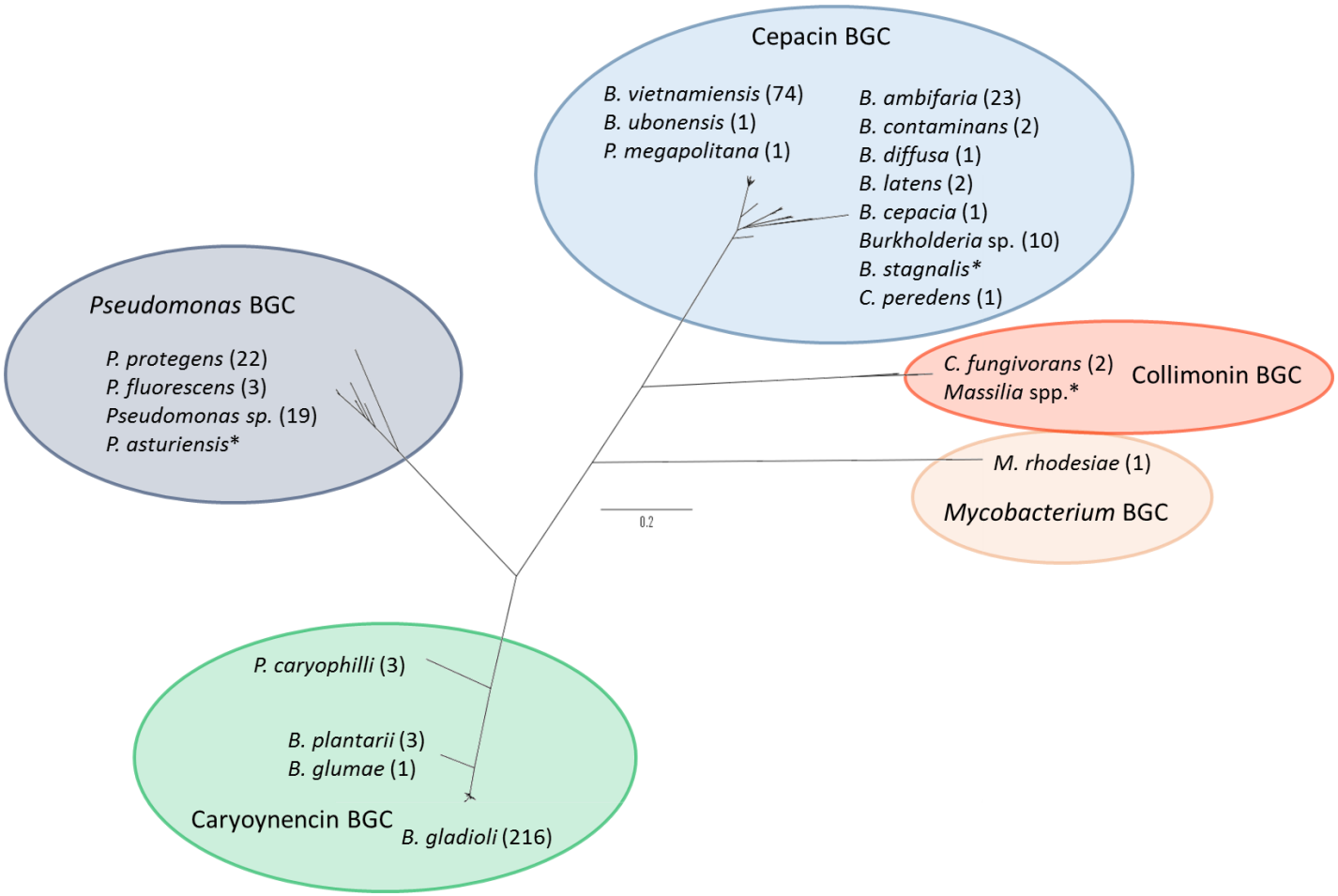


Figure 27. Protein phylogeny of the conserved polyyne fatty acyl-AMP ligase.

The phylogeny was constructed with 388 proteins derived from five characterised or putative polyyne BGCs. Species that encoded partial BGCs are highlighted with an asterisk (*). Alignment length = 681 aa.

5.3. Discussion

Regulatory gene mining led to the identification of a previously uncharacterised biosynthetic gene cluster in *B. ambifaria*. Insertional mutagenesis confirmed the association between the BGC and antimicrobial activity against Gram-positive bacteria, and to a lesser extent, Gram-negative bacteria. High resolution mass spectrometry elucidated the identity of the compound as the *Burkholderia* polyyne cepacin. Comparison of the extracted ion chromatograms from *B. ambifaria* BCC0191 with the original producer *Burkholderia diffusa* LMG 24093 (formerly *Pseudomonas cepacia*) confirmed that the metabolite was the same as the one described in 1984 (Parker *et al.* 1984). The distribution and diversity of bacterial polyyne BGCs was explored through phylogenetic analysis of shared genes between different polyyne BGCs.

5.3.1. The role of cepacin and polyynes in *Burkholderia*

Cepacins A and B are historical *Burkholderia* metabolites originally identified in *B. diffusa* (formally *P. cepacia*), with anti-bacterial (Gram-positive and Gram-negative) and anti-fungal activity (Parker *et al.* 1984). In comparison to most *Burkholderia* antimicrobial secondary metabolites (Table 11), the biosynthetic origin of cepacins remained unknown. An interesting contrast to the aforementioned specialised metabolite BGCs (Table 11) was the identification of the cepacin BGC initially through association with *luxRI* regulatory genes. The cepacin BGC was later confirmed to be the source of antimicrobial activity with site-directed mutagenesis, similar to bactobolins (Seyedsayamdost *et al.* 2010) and gladiolin (Song *et al.* 2017). Antimicrobial profile comparisons between wild-type and insertional mutant derivatives of cepacin-encoding *B. ambifaria* strains highlighted the contribution of cepacin to the *in vitro* antimicrobial capacity of the species. Construction of these cepacin deficient mutants, coupled to the development of a pea plant-pathogen model, provided a means to assess the contribution of a specific metabolite in an applied biological control crop model (Chapter 6).

5.3.2. Distribution of polyyne pathways in *Burkholderia*

A limited number of bacterial polyynes have been characterised; cepacins were discovered in *Burkholderia diffusa* (formally *Pseudomonas cepacia*) (Parker *et al.* 1984), caryoynencins isolated from *Burkholderia caryophylli* (formally *Pseudomonas caryophylli*) (Kusumi *et al.* 1987), Sch 31828 from *Microbispora* sp. (Patel *et al.* 1988), and most recently collimonins (originally collimomycin) in *Collimonas fungivorans* (Fritsche *et al.* 2014; Kai *et al.* 2018). Of the four defined bacterial polyynes, biosynthetic pathways have been elucidated for caryoynencins (Ross *et al.* 2014), collimonins (Kai *et al.* 2018), and as a result of this study, cepacins (Mullins *et al.* 2019). The high degree of pathway similarity between the caryoynencins, collimonins and cepacins was explained by the elucidated biosynthetic logic

of bacterial polyyne synthesis, giving rise to the characteristic alternating single-triple carbon bond moiety (Ross *et al.* 2014). Fritsche *et al.* (2014) identified orthologous regions to the collimonins BGC in *Burkholderia* sp. during a search for further examples of the collimonins BGC, describing them as partial versions of the collimonins BGC. These 'partial' *Burkholderia*-derived collimonin BGCs were in fact the origins of caryoynencin and cepacin biosynthesis, and were fully described as such by Ross *et al.* (2014) and Mullins *et al.* (2019), respectively.

Three of the four described bacterial polyyne BGCs were characterised in the order Burkholderiales, two of which, caryoynencins and cepacins, were identified in *Burkholderia*. Limited genomic screens have been performed to identify the distribution of these polyynes pathways following the identification of the collimonins and caryoynencins BGCs. As mentioned previously, orthologous genes to both the caryoynencin and collimonin BGCs were identified in several *Burkholderia* spp., *Pseudomonas protegens*, *Mycobacterium rhodesiae* and *Streptomyces* sp. Mg1 (Fritsche *et al.* 2014; Ross *et al.* 2014). The collimonins publication focussed on the structural characterisation of collimonins, as opposed to the distribution of the pathway (Kai *et al.* 2018). During the discovery of caryoynencins and collimonins (Fritsche *et al.* 2014; Ross *et al.* 2014; Kai *et al.* 2018), the authors erroneously referred to the first description of cepacin biosynthesis as the product of *B. cepacia*, which was taxonomically re-defined as *B. diffusa*. However, this was likely due to the intermediate definition of *P. cepacia* as *B. cepacia*, and the listing of the strain as *B. cepacia* in the American Type Culture Collection (ATCC). The existence of 4,479 locally assembled *Burkholderia* genomes has permitted the interrogation of polyyne biosynthesis dissemination in the genus, and expanded on the limited previous knowledge of polyyne pathway distribution. Based on the phylogenetic positioning of the polyyne BGCs (Figure 27), we can propose the direction of acquisition of the polyynes by the genus *Burkholderia* and closely related genera. The distant relationship between the caryoynencin and cepacin BGCs, despite their occurrence in the same genera, could have resulted from one or two main HGT acquisition events into the genus *Burkholderia*. In the event of a single major horizontal gene transfer (HGT) event, an ancestral *Burkholderia* would obtain the archetypal polyyne BGC, followed by divergence and subsequent HGT events to other species and genera. The alternative hypothesis is the divergence of the archetypal polyyne BGC in an unknown origin species, followed by the independent acquisition of the diverged caryoynencin and cepacin BGCs.

All three previously described polyyne metabolites displayed anti-Gram-positive activity in the form of *Staphylococcus* antagonism, while Sch 31828 and collimonins also possessed anti-fungal activity (Patel *et al.* 1988; Fritsche *et al.* 2014; Ross *et al.* 2014). Antagonism toward Gram-negative organisms was also observed with polyynes Sch 31828 and caryoynencins, but not investigated for the collimonin-synthesising *Collimonas* spp. The anti-Gram-positive and anti-oomycetal activity was congruent with the observed antagonism of cepacin-encoding *B. ambifaria*, however, no change in *C. albicans* inhibition zone was observed in cepacin-deficient mutants, in contrast to the anti-candidal activity of Sch 31828 (Patel *et al.* 1988). Despite the high sequence similarity between the collimonins and cepacins BGCs, there was variation in the oomycete bioactivity. Collimonins selectively targeted oomycete pathogens, exhibiting antagomism towards *Saprolegnia parasitica* and *Phytophthora megakarya*, but failed to inhibit *P. ultimum* (Fritsche *et al.* 2014). Cepacin-encoding *B. ambifaria* were noticeably active against *P. ultimum*, to such an extent that *P. ultimum* became the model crop pathogen in the biological control experiments described in Chapter 6.

5.3.3. Temperature-dependent expression of cepacin BGC

Many *Burkholderia* metabolites are extracted from cultures grown at 30°C (Mahenthiralingam *et al.* 2011; Flórez *et al.* 2017; Song *et al.* 2017) as an optimal production temperature. However, cepacin A derived antagonism was greater at reduced temperatures (Figure 24), approximately 22°C, in the majority of *B. ambifaria* strains; this was confirmed by integrated HPLC-peak area as a proxy for metabolite concentration. The phenomenon of enhanced secondary metabolite yields at lower incubation temperatures has also been observed in *B. thailandensis* rhamnolipid biosynthesis (Funston *et al.* 2016). Higher catalytic activity of the rhamnolipid biosynthesis enzymes at the lower temperature was proposed as a possible explanation in *B. thailandensis* (Funston *et al.* 2016). While this explanation would purportedly explain the phenomenon observed in the majority of cepacin-encoding *B. ambifaria* strains, a minority of strains maintain their cepacin-associated antimicrobial activity, or even exhibit elevated antagonism such as strain BCC1256 (Figure 24). Additionally, *B. diffusa* LMG 29043 exhibited similar antimicrobial activity at both 22°C and 30°C. An alternative explanation is the existence of temperature-dependent regulatory systems that increase gene expression of the cepacin BGC in the majority of *B. ambifaria* strains. Multiple examples of enhanced gene expression at sub-optimal growth temperatures have been observed in plant- and fish-pathogenic bacteria (Guijarro *et al.* 2015).

5.4. Conclusions

The main conclusions following insertional mutagenesis of the cepacin BGC, antimicrobial production screening and phylogenetic distribution analyses were as follows:

1) Insertional mutagenesis confirmed the biosynthetic origin of the historical *Burkholderia* polyyne cepacin in *B. ambifaria*.

2) Cepacin possesses considerable activity against anti-Gram-positive bacteria and the oomycete *P. ultimum*.

3) Lower temperatures were correlated with higher cepacin-associated antimicrobial activity in the majority of *B. ambifaria* cepacin-producing strains.

4) Polyyne-based antimicrobials were encoded by multiple *Burkholderia* species, and other bacterial genera, and possessed a highly conserved biosynthetic origin.

6. Biological control and virulence of *B. ambifaria*

The data and results shown in this chapter represent my own independent work/investigation, except where otherwise stated. Research and methods pertaining to this chapter have been published in part in the following manuscript. Contributions from the authors are acknowledged where appropriate, and wording has been adapted where necessary.

Mullins AJ, Murray JAH, Bull MJ, Jenner M, Jones C, Webster G *et al.* Genome mining identifies cepacin as a plant-protective metabolite of the biopesticidal bacterium *Burkholderia ambifaria*. *Nature Microbiology* 2019; 4: 996–1005. DOI: 10.1038/s41564-019-0383-z

6.1. Introduction

Several biological control products registered with the US Environmental Protection Agency (EPA) during the 1990s contained live *Burkholderia* suspensions. These products included Deny® and Blue Circle® manufactured by Stine Microbial Products, and Intercept® manufactured by Agrium (Parke and Gurian-Sherman 2001). At the time of registration they were described as containing *B. cepacia* isolates M54, J82 and M36; the former two isolates have now been classified as *B. ambifaria*, and the latter strain identified as *B. cenocepacia* (Parke and Gurian-Sherman 2001). Multiple publications illustrate the protective abilities of *Burkholderia* strains on a variety of crops against a plethora of plant pathogenic fungi and oomycetes (Parke *et al.* 1991; Hebbar *et al.* 1992; McLoughlin *et al.* 1992; King and Parke 1993; Mao *et al.* 1998). These biological control functions were verified at different experimental scales from laboratory-based microcosms, to greenhouses and field-based trials (Parke and Gurian-Sherman 2001). In addition to the registered biological control strains, many other *Burkholderia* strains were experimentally confirmed to function as biocontrol agents, many of which have since been identified as *B. ambifaria* (Coenye *et al.* 2001a).

The *B. ambifaria* type strain AMMD (BCC0207) was one of the earliest to feature in a biocontrol publication, demonstrating its ability to protect pea cultivars against pathogenic oomycetes *Pythium* spp. and *Aphanomyces wuteiches* via seed coats in field tests (Parke *et al.* 1991; King and Parke 1993). Fungal plant pathogens *Fusarium, Sclerotinia* and *Rhizoctonia* were inhibited by *B. ambifaria* strains J82 (BCC0191) (McLoughlin *et al.* 1992), ATCC53267 (BCC0284) (Hebbar *et al.* 1992) and BC-F (BCC0203) (Mao *et al.* 1998) in various biocontrol models featuring important crop species including maize, tomato, and pepper, in addition to sunflower. A precedent was developing for the use of biological control agents in agriculture, either as substitutes or as supplementary treatment alongside traditional chemical pesticides.

However, almost in parallel to the *Burkholderia* biological control research these bacteria were emerging as pathogens of immunocompromised patients, especially in cystic fibrosis (CF) treatment centres (Parke and Gurian-Sherman 2001). Several studies observed increasing numbers of CF patients contracting *Burkholderia* infections (formerly *Pseudomonas cepacia*), noting its high antimicrobial resistance and associated rapid progressive decline in lung function leading to death in a subgroup of patients (Isles *et al.* 1984; Thomassen *et al.* 1985). An environmental reservoir was proposed as the source of *Burkholderia* in CF clinics, however, up to 50% of the *B. cepacia* burden in some US clinics derived from the same strain by genotypic tests supporting patient-to-patient transmission of specific (Smith *et al.* 1993). Epidemic strains of *B. cepacia* also emerged in the UK causing infections across regional CF clinics (Govan *et al.* 1993), and intercontinental transmission (Sun *et al.* 1995).

Increasing concern over the rise of *B. cepacia* infections in CF patients and the use of similar bacteria in agriculture as biocontrol agents led the US EPA to develop a risk assessment structured on the existing knowledge of the opportunistic pathogen (US Environmental Protection Agency 1999b). This report was reviewed at a scientific advisory meeting involving both clinical and environmental researchers (US Environmental Protection Agency 1999a). Insufficient knowledge was highlighted in regards to the background level of different *Burkholderia* species*,* the fate of introduced *Burkholderia* as biocontrol agents, interaction between artificially introduced and background environmental *Burkholderia*, and the mode of infection in CF patients (US Environmental Protection Agency 1999a). The scientific panel subsequently agreed that a highly conservative approach to *Burkholderia* biocontrol agents was necessary, resulting in a moratorium on such products (US Environmental Protection Agency 2004). The moratorium was followed by the voluntary withdrawal of existing *Burkholderia*-containing biocontrol products.

Research into the biological control abilities of *Burkholderia* bacteria diminished during the early 2000s, but a resurgence in both research and commercial interest has occurred in recent years. *Burkholderia rinojensis* A396 was recently characterised as possessing insecticidal and mitocidal properties as both live and heat treated formulations (Cordova-Kreylos *et al.* 2013). Marrone Bio Innovations® marketed this strain in multiple heat-killed biological control products, such as turf protector Zelto®, insecticide Venerate®, and nematicide Majestine®.

Advancements in the understanding of *Burkholderia* evolutionary history, supported by high-throughput sequencing technologies, have been accompanied by improved *Burkholderia* biology knowledge that the third replicon c3 in *Burkholderia cepacia* complex species functions as a virulence plasmid (Agnoli *et al.* 2012). Both of these factors have reinvigorated research into *Burkholderia* biological control and provided researchers with the necessary tools to address the concerns raised at the scientific advisory meeting chaired in 1999. The research in this thesis harnessed these advances to systematically define the factors required for successful biological control using *B. ambifaria*.

6.1.1. Aims and objectives

This chapter focused on exploiting *B. ambifaria* as a biological control agent and understanding the contribution of antimicrobials to this effect by addressing the following aims:

1. Develop a reliable biological control model to understand the role of *Burkholderia* metabolites in crop protection.

2. Explore the importance of cepacin A in biological control of *P. ultimum* through construction and testing of pathway mutants.

3. Identify high cepacin A producer as an optimal *B. ambifaria* biocontrol strain against *P. ultimum*.

4. Develop a seed exudate-based media to test for the induction of antimicrobial metabolites in *B. ambifaria*.

5. Create a *B. ambifaria* third-replicon deletion mutant and test in both insect and murine models for reduced virulence.

6. Interrogate the impact of third replicon loss on *B. ambifaria* rhizocompetence.

6.2. Results

6.2.1. Differential inhibition of pea germination and growth by bacterial plant pathogens

A scoping experiment (*n* = 3, one replicate) was performed to determine if any of the bacterial plant pathogens could be incorporated into a seed germination and seedling growth biocontrol assay. Following soil drenching with the pathogens, the growth of the seedlings was monitored and the result recorded at day 14-post inoculation. Minimal or subtle differences in growth were observed for *P. carotovorum*, *P. syringae* pv. *syringae*, *P. savastanoi* pv. *phaseolicola*, *R. radiobacter* and *X. campestris*. In contrast, shoot height was considerably reduced for seeds growing in soil inoculated with *P. syringae* pv. *tomato*, while seeds exposed to *D. solani* either failed to germinate or the seedlings were severely stunted.

Based on the scoping experiment, plant pathogens *D. solani* and *P. syringae* pv. *tomato* were pursued further as potential biocontrol antagonist organisms. The previous experiment was adapted to investigate the impact of pathogen contact with the seed on survival and growth. The *P. sativum* seeds were exposed to the pathogens by soil drenching or seed coating (*n*=5 per condition, one replicate). No impact on growth or survival was observed when seeds were exposed to biocontrol agent *B. ambifaria* BCC0191 (Figure 28). A reduction in shoot height was observed in seedlings, exposed to *D. solani*, and no impact on plant survival was observed compared to the control. In contrast, *P. syringae* pv. *tomato* caused a reduction in plant survival in both seed coat and soil drenching exposure methods, and a large variation in shoot height in the soil drench method (Figure 28).

6.2.2. Low *P. ultimum* infestation levels prove highly virulent to *P. sativum*

To develop the biological control model, the fastest growing mycelial plant pathogen, *P. ultimum*, was chosen as the focus of the assay. The susceptibility of wheat, soybean, pea, alfalfa, and broad bean were assessed by exposure of the germinating seeds to a 20% *Pythium* soil infestation level. No effect of this *P. ultimum* pathovar was observed on alfalfa, wheat, broad bean or the legume soybean, while a significant knock-down in survival was observed for pea seed germination and emergence. Based on the high susceptibility of *P. sativum* to *P. ultimum*, the Early Onward pea variety was chosen as the model crop species in the biological control model.

A scoping experiment was conducted to determine the infective dose of *P. ultimum* to use in the biological control assay. Challenging pea seed germination in soil infested with 1%, 5%, 10% or 20% *P. ultimum* provided an insight into the highly virulent nature of *P. ultimum*. An infective dose of 1% *P. ultimum* was sufficient to prevent the germination of 100% of seeds in the scoping experiment (*n* = 3 seeds per condition, one replicate) (Figure 29). *B. ambifaria* seed coats were applied as a crude high cell density suspension, as described in section 2.24.4, to assess the biocontrol ability of the organism. *B. ambifaria* BCC0191 provided protection to *P. sativum* against all *P. ultimum* infestation levels, however, increasing the *P. ultimum* infective dose resulted in a mean decrease in surviving plant shoot height up to 10% *P. ultimum* (Figure 29). Based on this experiment an infective dose of 1% *P. ultimum* was chosen as the optimum condition to progress in a biological control model. All three plants survived at 1% *P. ultimum* infestation and the shoot heights were closest to the control, in comparison to any other infective dose; thereby providing the best distinction between the *P. ultimum* kill virulence and *B. ambifaria* protection.

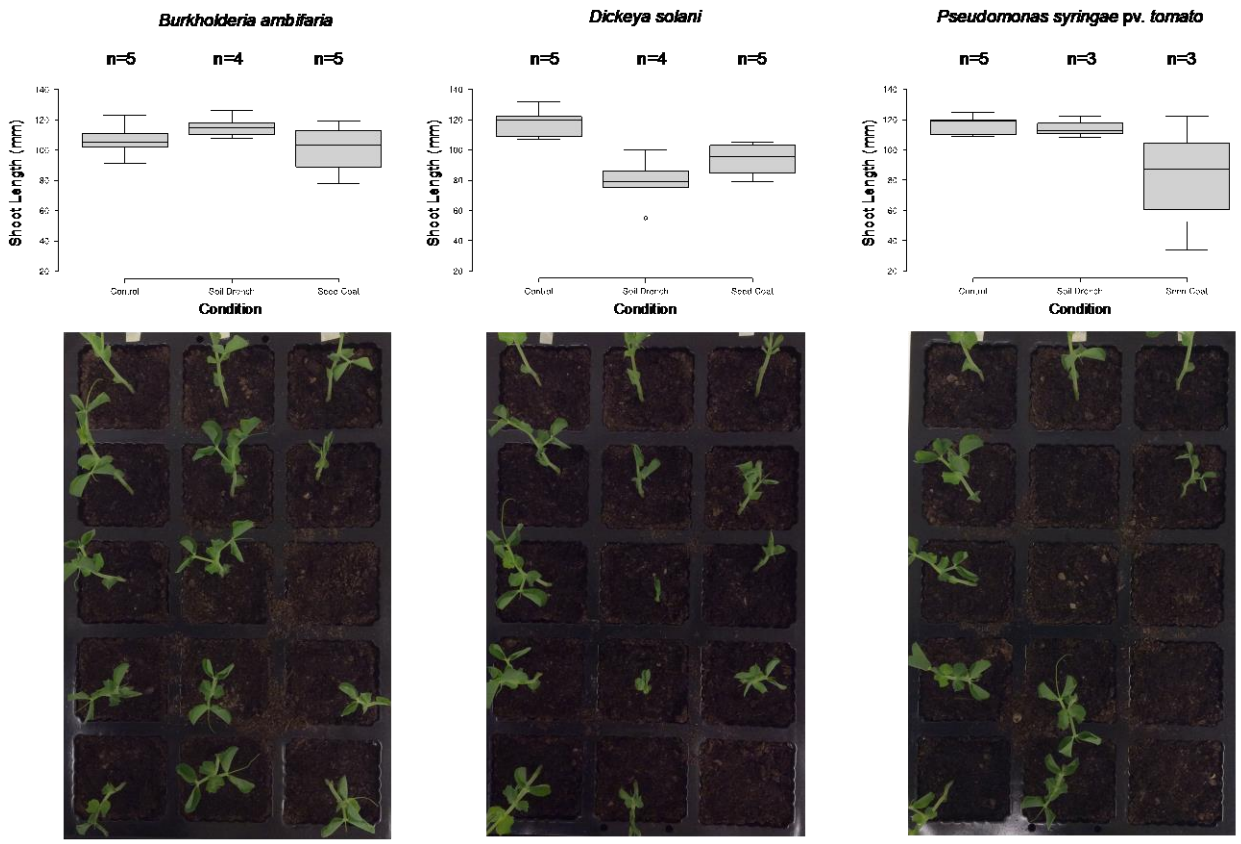


Figure 28. Influence of bacterial biocontrol agent and plant pathogens on *Pisum sativum* germination and growth.

The effect of biological control agent *B. ambifaria* and plant pathogens *D. solani* and *P. syringae* pv. *tomato* on *P. sativum* germination and growth when inoculated by either seed coat or soil drenching. Five seeds were used per condition. Germination success is indicated by the *n* value, and growth was assessed by measuring shoot height. *P. sativum* seeds were allowed to grow for 14 days prior to recording survival and shoot height.

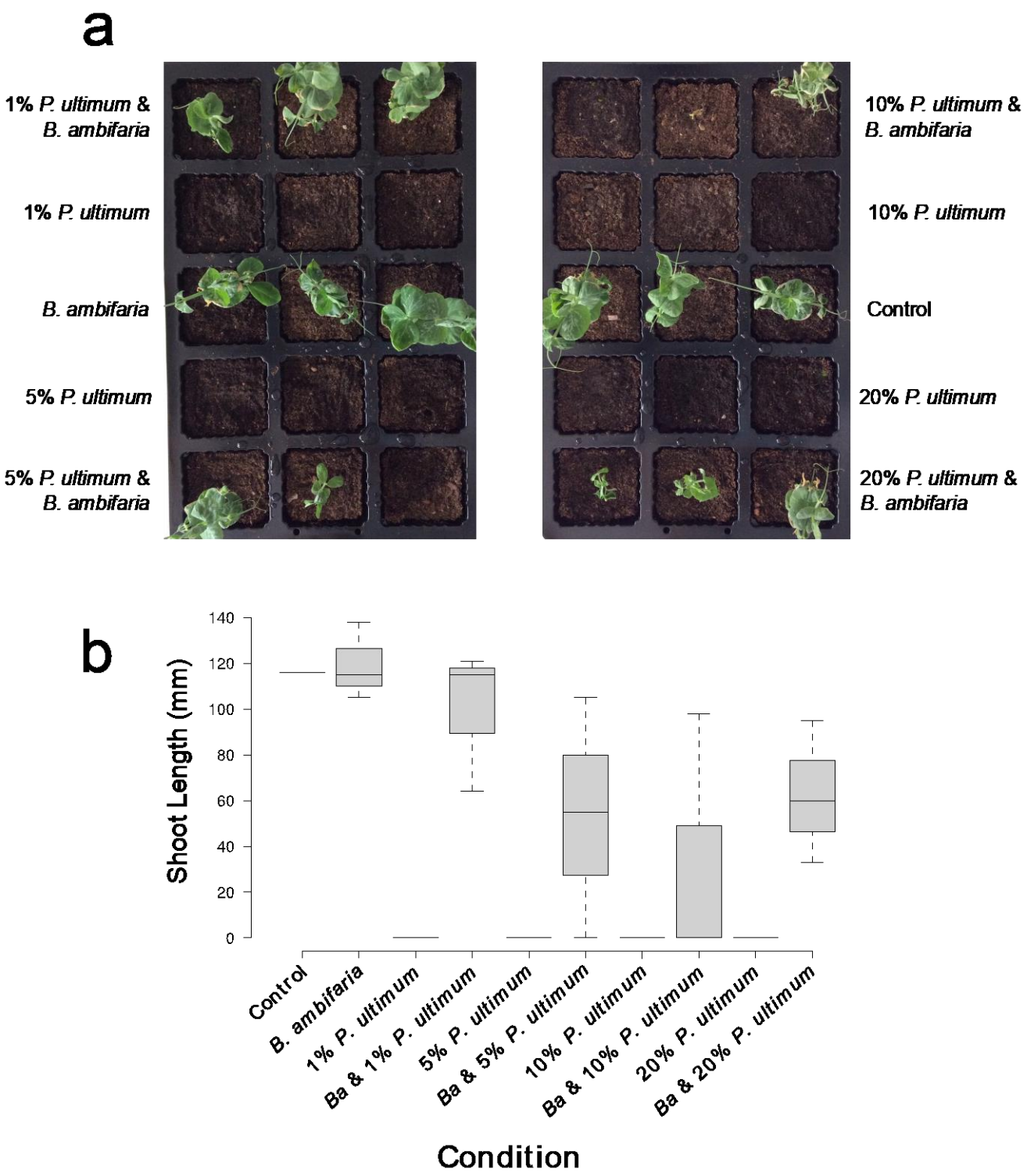


Figure 29. *B. ambifaria* biocontrol against varying levels of *P. ultimum* infestation.

(a) Survival of *P. sativum* (*n*=3 per condition) against *P. ultimum* at different infestation rates in the presence and absence of biocontrol strain *B. ambifaria* BCC0191 (*Ba*). (b) The growth of *P.sativum* seedlings was assessed by measuring shoot height following 14 days of incubation in 18:6 light:dark photoperiod at 22°C and 70% RH.

6.2.3. Seed coat carrying capacity linked to initial cell suspension density

A dilution series of *B. ambifaria* was prepared and used to coat *P. sativum* seeds to understand the relationship between the recoverable viable count per seed and the initial cell suspension. A distinct dose response was observed with a 100-fold reduction in cfu/seed in comparison to the initial cell suspension, implying that each *P. sativum* seed retained approximately 10 µl of the cell suspension (Table 17). This information enabled us to consistently and accurately control the cell density of biocontrol agent per seed.

Table 17. *B. ambifaria* cell suspension and recoverable colony forming units per seed.

Initial cell suspension magnitude[a]	Average recovered cfu/seed
10^8	3.5×10^6
10^7	1.2×10^5
10^6	2.1×10^4
10^5	1.3×10^3
10^4	2.3×10^2

[a]The original cell suspension of *B. ambifaria* was quantified at 5.3×10^8 cfu/ml.

6.2.4. Cepacin A as a key mediator of biocontrol in *B. ambifaria* against *P. ultimum* damping-off disease

To progress the discovery of the biosynthetic origin of the *Burkholderia* polyyne metabolite cepacin A, described in Chapter 5, we investigated the biological control propeties of the metabolite. *B. ambifaria* BCC0191 WT, and BCC0191::*ccnJ*, a cepacin-deficient derivative, were applied as seed coats to *P. sativum* seeds at varying cell densities to assess the limits of protection against a standardised *P. ultimum* soil infestation. Strong protection of *P. sativum* seed germination and emergence against *P. ultimum*-infested soil was observed with the addition of *B. ambifaria* BCC0191 WT as a seed coat. This protection ranged between 50-70% and 90-100% with inoculum cell densities of approximately $1x10^5$ cfu/seed and $1x10^7$ cfu/seed, respectively (Figure 30). In contrast to *B. ambifaria* BCC0191::*ccnJ* that displayed statistically significant loss in the protective phenotype relative to the wild-type at inoculums 10^5, 10^6, and 10^7 cfu/seed ($P = 0.002$, $P = 0.03$ and $P = 0.002$, respectively) (Figure 30). Even at the highest applied seed coat cell density of ~$1x10^7$ cfu/seed the mutant only afforded a maximum of 40% plant survival. A mean of less than 5% plant survival was recorded for BCC0191::*ccnJ* coated seeds at $1x10^5$ cfu/seed, highlighting the significant protective phenotype imparted by cepacin A biosynthesis.

6.2.5. Loss of c3 replicon does not affect in biocontrol in *B. ambifaria* BCC0191

In addition to *B. ambifaria* BCC0191::*ccnJ*, the third replicon deletion mutant BCC0191Δc3 was also screened as a biocontrol agent against *P. ultimum* infested soil. The discovery that the third replicon of *B. cepacia* complex species was a non-essential genomic entity representing a large virulence plasmid provided an avenue for reductive engineering of the historical biopesticide *B. ambifaria*. We hypothesised the loss of virulence but maintenance of anti-*P. ultimum* bioactivity demonstrated in the biological control model. At all three BCC0191Δc3 seed coat inoculums we observed a non-significant difference in protection compared to the BCC0191 wild-type ($P = 0.22$, $P = 0.22$ and $P = 0.16$) (Figure 31). The loss of the c3 replicon was not as detrimental as the disruption of cepacin A biosynthesis. To understand the basis of this reduced plant protection we performed HPLC analysis of the third-replicon deletion mutant grown on BSM-G to assess the production of cepacin A. Unexpectedly, the production of cepacin A doubled in comparison to wild-type BCC0191, despite this mutant being associated with a lower level of protection (Figure 22).

To further investigate the impact of third replicon deletion on cepacin A production we assessed the antimicrobial properties of several *B. ambifaria* BCC0191 mutants. In addition to BCC0191 wild-type, BCC0191::*ccnJ* and BCC0191Δc3, an additional derivative was generated to aid the disentanglement of antimicrobial activity and presence of specialised metabolite gene clusters. The third derivative combined both previously mentioned mutations and was constructed by re-introducing the insertional mutation into the BCC0191Δc3 background (BCC0191::ccnJΔc3). Loss of the third replicon abolished the anti-fungal activity against all three fungal pathogens: *C. albicans*, *F. solani* and *A. alternata*, but enhanced the observed anti-Gram-positive activity (Figure 22). The observed increase in Gram-positive antimicrobial inhibition against *S. aureus*, *E. faecalis* and *B. subtilis* corroborated the HPLC result showing increased cepacin A production (Figure 22).

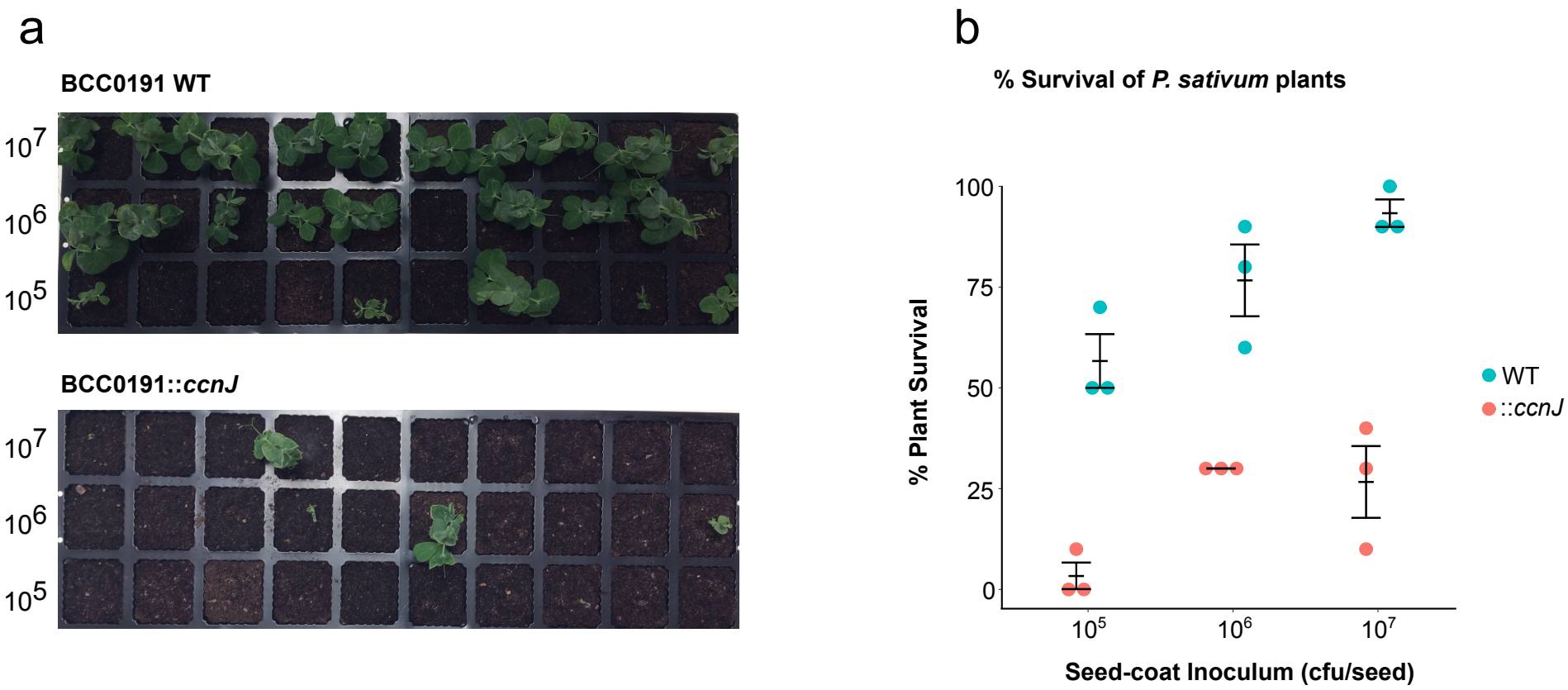


Figure 31. Biological control assay illustrating the role of cepacin A in preventing *P. ultimum*-mediated damping-off disease.

(a) *P. sativum* plants germinated and grown (14 days) in soil infested with *P. ultimum*. Seeds were observed in groups of 10 coated with 107, 106 and 105 cfu, respectively, of BCC0191 WT and BCC0191*::ccnJ*. (b) The overall percentage survival of germinating peas treated with the *B. ambifaria* WT and BCC0191*::ccnJ* (*n*= 3). Survival was assessed as plants with had stems of >30 mm in height after 14 days. Plant survival was significantly different at every inoculum level between BCC0191 WT and BCC0191*::ccnJ,* as indicated by two-sided t-test or Welch's two-sided t-test; significant difference (left to right) $P = 0.002$, $P = 0.03$, $P = 0.002$ with 95% confidence interval. The centre bar represents the mean, and the error bars represent the standard error.

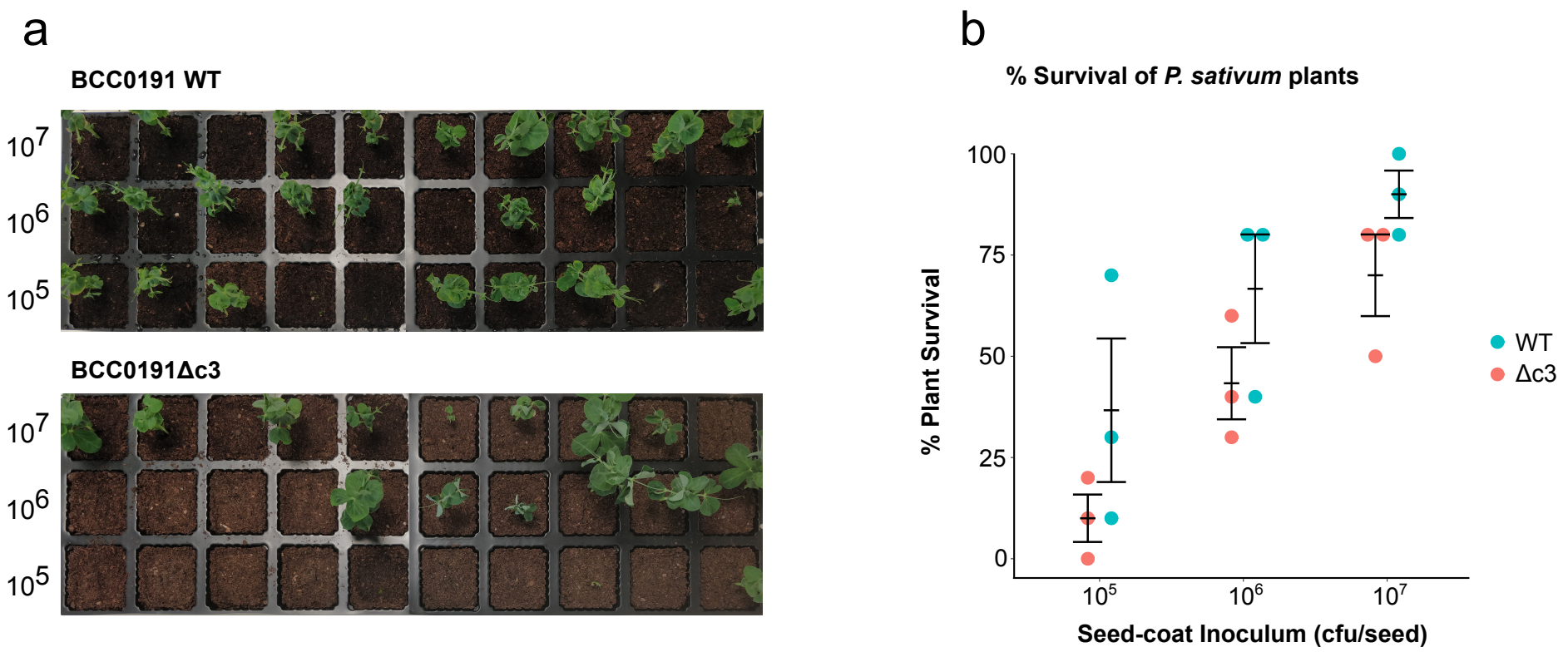


Figure 31. Biological control assay illustrating the role of the third replicon c3 in preventing *P. ultimum*-mediated damping-off disease.

(a) *P. sativum* plants germinated and grown (14 days) in soil infested with *P. ultimum*. Seeds were observed in groups of 10 coated with 107, 106 and 105 cfu, respectively, of BCC0191 WT and BCC0191Δc3. (b) The overall percentage survival of germinating peas treated with the *B. ambifaria* WT and BCC0191Δc3 (*n*= 3). Survival was assessed as plants with had stems of >30 mm in height after 14 days. No significant difference (left to right: $P = 0.22$, $P = 0.22$, $P = 0.16$) was observed, as determined by two-sided t-test with 95% confidence interval, in plant survival between BCC0191 WT and BCC0191Δc3 at all inoculum levels. The centre bar represents the mean, and the error bars represent the standard error.

6.2.6. Seed-exudate as a specific elicitor of polyyne cepacin A biosynthesis

Following the observation of cepacin-mediated plant protection against the damping-off pathogen *P. ultimum*, we decided to explore the ability of pea exudate to induce antimicrobial metabolites in *B. ambifaria*. Two strains of *B. ambifaria* were used to assess the induction of multiple specialised metabolite biosynthetic gene clusters: BCC0191 for cepacin A induction, and BCC0203 for bactobolins and enacyloxin IIa induction.

The resulting antimicrobial activity and HPLC analyses obtained from pea exudate agar were compared to the minimal BSM-G media and nutrient-rich TSA media. The antimicrobial profile for *B. ambifaria* BCC0203 was similar between TSA and BSM-G, both of which displayed Gram-negative, Gram-positive and fungal inhibition. Pea exudate induced anti-Gram-positive and anti-fungal activity to a lesser degree, but almost abolished all observable anti-Gram-negative activity (Table 18). Similar antimicrobial antagonism profiles were observed for BCC0191 grown on pea exudate agar and BSM-G. Growth of BCC0191 on both pea exudate and BSM-G resulted in antagonism against *S. aureus,* while all three media types induced activity against *C. albicans* (Table 18 and Figure 32). HPLC analysis of BCC0191 extracts confirmed the production of cepacin A at comparable levels between the BSM-G and pea exudate medium; however, only BSM-G increased pyrrolnitrin production with reduced levels detected for TSA and pea exudate (Figure 32).

Table 18. *B. ambifaria* zones of inhibition (mm) induced by different media types[a].

	BCC0203			BCC0191		
Media Type	***C. albicans***	***P. carotovorum***	***S. aureus***	***C. albicans***	***P. carotovorum***	***S. aureus***
TSA	36	32	25	31	3	4
BSM-G	39	35	25	32	9	46
Pea Exudate	21	7	31	29	7	48

[a] Experiments performed in triplicate

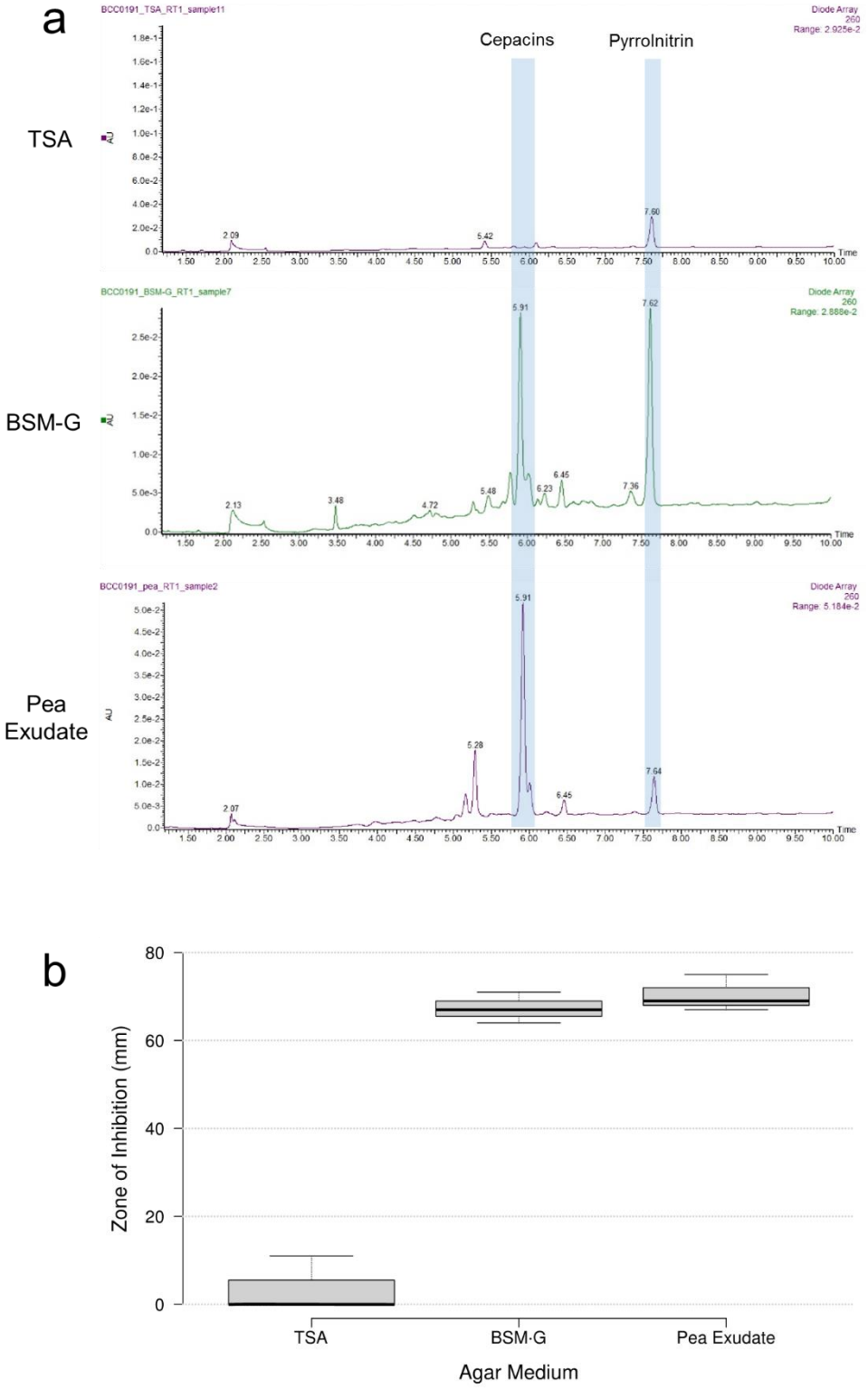


Figure 32. Influence of agar composition on antimicrobial production and bioactivity of *B. ambifaria* BCC0191.

B. ambifaria BCC0191 was grown on TSA, BSM-G and Pea Exudate agar for 3 days at 22°C. (a) HPCL traces at 260 nm from metabolite extracts obtained from agar plugs. (b) Zones of inhibition against *S. aureus* (*n*=3 independent overlay assays). Dr Gordon Webster performed the HPLC analyses.

6.2.7. Loss of replicon c3 failed to attenuate *B. ambifaria* BCC0191 virulence *G. mellonella* pathogenicity model

We used the model organism *Galleria mellonella*, as a well characterised infection model (Agnoli *et al.* 2012), to assess the virulence of *B. ambifaria* BCC0191 wild-type versus the third replicon deletion mutant. The survival of *G. mellonella* larvae were assessed for three days post infection. There was no statistically significant difference observed between the wild-type and mutant at any time-point over the three-day experiment (Figure 33).

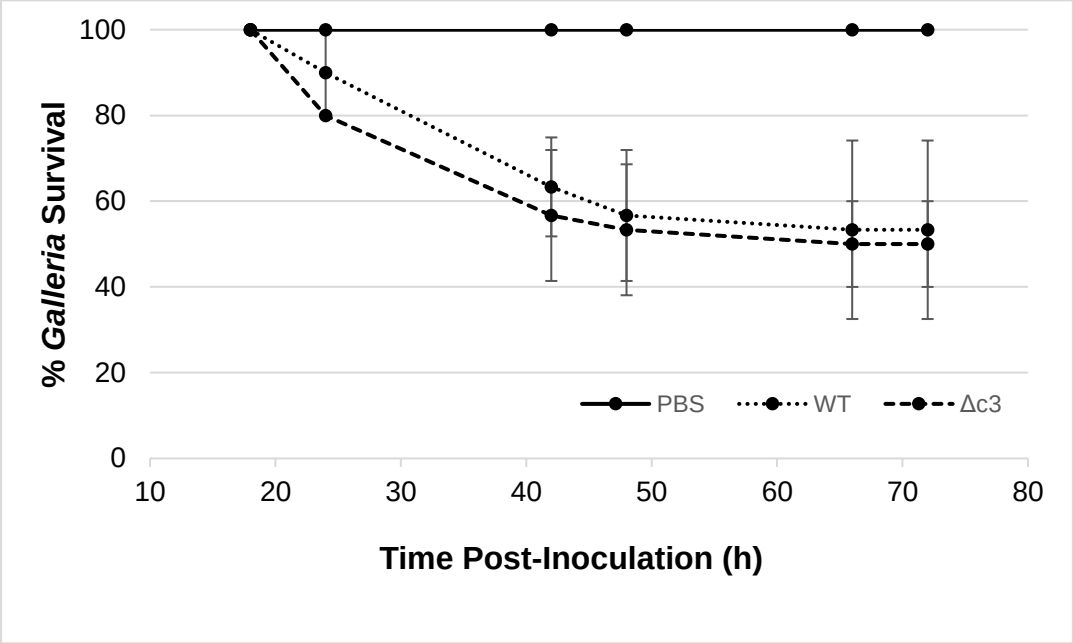


Figure 33. Pathogenicity assay of *B. ambifaria* in *G. mellonella* wax-moth larvae.

G. mellonella wax-moth larvae were injected in their last right hind proleg with approximately 1 x10^6 cfu of PBS, *B. ambifaria* BCC0191 wild-type, or BCC0191Δc3. Larvae were monitored for 72 hours post infection (n = 10 larvae per condition, 3 experimental replicates). Dr Gordon Webster assisted with the experimental design and set-up of the *G. mellonella* pathogenicity assay.

6.2.8. Reduced murine virulence of *B. ambifaria* third-replicon deletion mutant

To interrogate further the virulence of the BCC0191Δc3 derivative we conducted a murine respiratory inhalation model in collaboration with the University of Liverpool. Each mouse was exposed to approximately $1x10^6$ cfu of either *B. ambifaria* BCC0191 wild-type or BCC0191Δc3 and the cfu determined in the nasopharynx, lungs and spleen at one, two and five days by sacrificing mice. A key observation from the analysis was the low persistence of *B. ambifaria* wild-type in the model over the five-day experiment. *B. ambifaria* persisted at levels below $1x10^3$ cfu per mouse nasopharynx over the five-day experiment, and was cleared from the lungs of four mice by day 5 (Figure 34). The BCC0191Δc3 mutant was rapidly cleared from the lungs and nasopharynx, with no detectable *B. ambifaria* in the lungs after 48 hours, and only one out of six mice possessing detectable *B. ambifaria* in the nasopharynx after five days. *B. ambifaria* was undetectable in the spleens of mice over the five days, and no mice displayed visible disease symptoms throughout. Random amplified polymorphic DNA profiling confirmed that all bacterial colonies recovered from the mouse model on *B. cepacia* selective agar (BCSA) possessed the same genotypic-profile as the *B. ambifaria* BCC0191 wild-type and BCC0191Δc3 mutant (Figure 34). The same PCR banding pattern was observed from genetic fingerprinting of the isolated *Burkholderia* colonies and inoculum (Figure 35). This confirmed that the same *B. ambifaria* background used to infect the mice was recovered following the experiment, and that no other *Burkholderia* was contributing to the viable count.

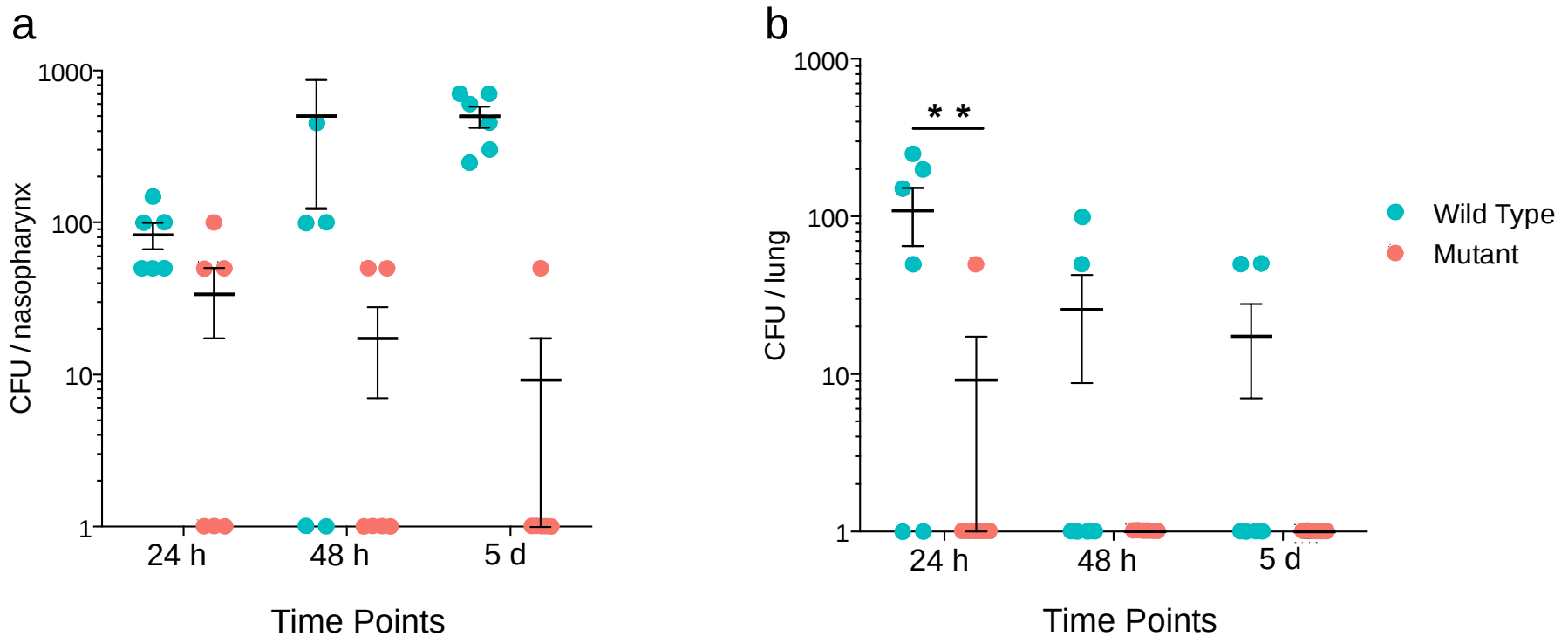


Figure 34. Comparison of *B. ambifaria* BCC0191 wild-type and BCC0191Δc3 persistence in murine respiratory inhalation model.

Mice were infected with approximately 2 x10^6 cfu of either *B. ambifaria* BCC0191 wild-type (WT) or BCC0191Δc3 via an inhalation model. The nasopharynx (a) and lungs (b) of mice (*n* = 6 mice per timepoint) were extracted at 24 hours, 48 hours or 5 days, and the cfu per tissue enumerated. The centre bar represents the mean, and the error par represents the standard error. Statistical significance was inferred by a two-way ANOVA with Sidak's multiple comparisons test (** = *P* = 0.0038). Dr Angharad Green and Dr Daniel Neill at the University of Liverpool performed the murine infection experiments and statistics.

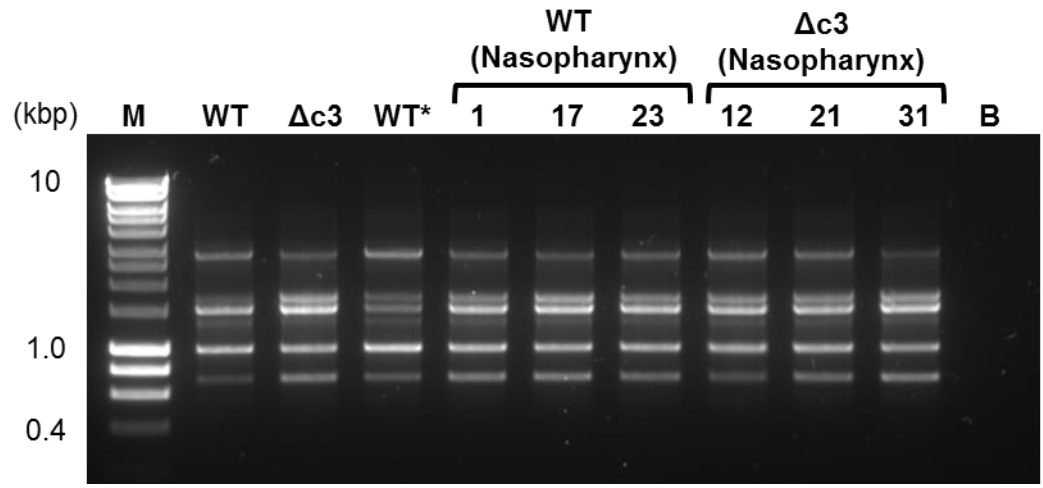


Figure 35. PCR genotyping of bacterial colonies isolated from murine respiratory infection model and grown on *Burkholderia* selective media (BCSM).

Total DNA was extracted from bacterial colonies (*n* = 31 lung and nasopharynx samples) using chelix-100 resin. Random amplified polymorphic DNA (RAPD) fingerprinting PCR was performed on total DNA (*n* = 1 PCR per sample) to validate genetic identity, and products separated by gel electrophoresis. Lanes are as follows: M, DNA ladder; WT, *B. ambifaria* BCC0191 wild-type; Δc3, *B. ambifaria* BCC0191Δc3; WT*, high quality DNA extract of *B. ambifaria* BCC0191 wild-type; 1, 17 and 23, BCC0191 wild-type colonies isolated from nasopharynx of mice at 24 h, 48 h, and 5 d, respectively; 12, 21, and 31, BCC0191Δc3 colonies isolated from nasopharynx of mice at 24 h, 48 h, and 5 d, respectively; B, PCR blank negative control.

6.2.9. Loss of third replicon reduces rhizocompetence in *P. sativum* model

The rhizocompetence of the *B. ambifaria* BCC0191 wild-type and the two derivatives: BCC0191::*ccnJ* and BCC0191Δc3 was assessed as a potential factor influencing the results observed in the biological control assay. The wild-type colonised the pea root with a mean of $2.8x10^7$ cfu per 1 g fresh root, which was comparable to the BCC0191::*ccnJ* mutant at $7.8x10^7$ cfu per 1 g fresh root, with no statistical significance. However, there was a reduced level of colonisation between the wild-type and third replicon mutant BCC0191Δc3 that colonised with a mean of $8.5x10^6$ cfu per 1 g fresh root ($P = 0.027$, df = 4, t = -4.40, 95% confidence interval).

6.2.10. Refinement of the biological control model

We sought to refine the biological control model to enhance throughput, reduce set-up time and reduce soil usage. To achieve this multiple seeds were planted in one weighed unit of soil, in contrast to the original model of one seed per unit of soil. By planting 8 seeds per 200 g potting mix as opposed to 1 seed per 60 g potting mix, we were able to reduce set up time. The use of 2 x 200 g aliquots per condition (16 seeds) in the refined model increased the number of seeds per replicate, compared to the use of 10 seeds per condition in the original model.

The *P. ultimum* dosage required refinement due to the presence of multiple seeds in close proximity contributing to the spread of the pathogen between seedlings. While the refined model better mimics the natural spread of the pathogen, the model increased the perceived virulence of *P. ultimum* and exhibited an apparent reduced the biological control phenotype of *B. ambifaria* compared to the original model. Several soil inoculum levels <1% were tested for suitability with and without *B. ambifaria* BCC0191 WT (Table 19).

Table 19*. P. sativum* seedling survival in refined biological control model with varying *P. ultimum* inoculums[a].

% Inoculum	No *B. ambifaria*	*B. ambifaria*
0%	94-100%	94%
0.1%	31%	94%
0.25%	13%	69%
0.5%	6%	38%
1.0%	0%	19%

[a] Experiments with and without *B. ambifaria* (10^7 cfu/seed) experiments were conducted separately. 16 seeds were used per condition with the exception of 'No *B. ambifaria* 0% inoculum' which possessed 8 seeds. *P. sativum* seedling survival was measured 14-days post planting. *n* = 1. Dr Gordon Webster assisted in experimental design and set-up.

6.2.11. *In vitro* polyyne production does not equate to biocontrol function

A high cepacin producing strain of *B. ambifaria* (BCC1241) was identified in Chapter 5 based on *in vitro* antimicrobial properties against *S. aureus* and *P. ultimum* and comparative HPLC analysis. The high producer BCC1241 and its cepacin-deficient derivative (BCC1241::*ccnJ*) were compared to BCC0191 WT and BCC0191::*ccnJ*. In addition, the prototypic cepacin producer *B. diffusa* LMG 29043 and *Collimonas fungivorans* Ter331 were also compared for their biological control properties. Interestingly, there was a smaller disparity between BCC0191 WT and BCC0191::*ccnJ* in the updated model compared to the original model, with plant survival at 44% with the cepacin-deficient BCC0191 seed-coating (Table 20). Surprisingly, despite the strong antagonism observed *in vitro* for BCC1241, the strain failed to surpass BCC0191 in biological control efficacy, offering protection levels equal to that of *B. ambifaria* BCC0191::*ccnJ*. The original cepacin producer *B. diffusa* LMG 29043 protected less than a third of the seedlings, while the *Collimonas* exhibited no protective phenotype (Table 20).

Table 20. Comparing the biocontrol abilities of different polyyne producers[a]

Seed-coat strain	% Survival
Live Control	100%
Kill Control	6%
B. ambifaria BCC0191 WT	69%
B. ambifaria BCC0191::*ccnJ*	44%
B. ambifaria BCC1241 WT	44%
B. ambifaria BCC1241::*ccnJ*	13%
Burkholderia diffusa LMG 29043	31%
Collimonas fungivorans Ter331	0%

[a] Experiment represented a scoping exercise (*n* = 1). Dr Gordon Webster assisted in experimental design and set-up.

6.3. Discussion

The screening of multiple plant pathogens and model crops led to the identification of an optimal biological control system involving oomycete *P. ultimum* and crop *P. sativum*. Following the elucidation of the biosynthetic origins of the historical *Burkholderia* polyyne metabolite cepacin A in Chapter 5, we sought to progress this discovery by understanding the involvement or cepacin A in the biological control of *P. ultimum*. Deletion of the third replicon resulted in no significant change in biocontrol efficacy, and subsequent virulence experiments in a murine respiratory inhalation model illustrated a distinct loss of persistence compared to the wild-type *Burkholderia* strain.

6.3.1. The influence of specific metabolites in biological control

The use of *Pseudomonas* species as biological control agents and the influence of multiple antimicrobials in mediating this phenotype have been well characterised (Haas and Keel 2003). However, despite the promising biocontrol activity demonstrated by *Burkholderia* species, there have been no studies categorically linking specific metabolites to a targeted biological activity. Previous studies have observed *B. ambifaria* AMMD (BCC0207) and BC-F (BCC0203) antagonism towards *P. ultimum* (Parke *et al.* 1991; King and Parke 1993; Mao *et al.* 1998; Heungens and Parke 2000; Li *et al.* 2002b)**,** but all fail to identify the key metabolites that mediate crop protection. Establishing a biological control model with a distinct protection phenotype of *B. ambifaria* versus kill controls provided a platform to dissect the impact of different antimicrobial specialised metabolites on the system. The marked reduction in germination and emergence observed following the disruption of the recently identified cepacin biosynthetic gene cluster categorically linked this metabolite to the mediation of a specific biological control phenotype. Oomycete and fungal-related damping-off diseases are capable of causing significant economic losses to the agricultural industry across multiple continents (Lamichhane *et al.* 2017). Effective biocontrol relies upon an understanding of the biology and specific interaction between the beneficial microbe and pathogen.

6.3.2. Developing a robust biological control model

Initial screening for a suitable pathogen focussed on bacterial plant pathogens; however, the bacteria tested either failed to produce a phenotype or caused weak disease phenotypes. Reasons for the unsuitability of these pathogens were likely due to the host range not extending to the crop models examined, and the mechanism by which these pathogens cause disease. While crops are inoculated from contaminated seeds or soil, the bacteria mainly cause disease symptoms in later crop growth stages, causing soft rot in root vegetables and black rot in foliage (Mansfield *et al.* 2012). Consequently, we investigated the potential of

pathogens that cause damping-off disease, as this would provide a relatively fast read-out of disease and biocontrol efficacy compared to pathogens that require a complete crop growing season to exhibit symptoms. The publication trail associated with *B. ambifaria* and *P. ultimum* (Parke *et al.* 1991; King and Parke 1993; Mao *et al.* 1998; Heungens and Parke 2000; Li *et al.* 2002b) also offered a historical precedent to dissect the well-studied phenomenon with high-resolution genomic analyses and knowledge of the biosynthetic origin of cepacin.

A very low infective dose of 1% *P. ultimum* was required to cause substantial germination inhibition and post-emergence stunted growth. This has been observed previously with *Pythium arrhenomanes* during pathogenicity test with rice seedlings (Toda *et al.* 2015). This inoculum was further reduced during assay refinement to 0.25% to enable the observation of the biological control phenotype. This was potentially the result of planting multiple seeds in close proximity, enabling the efficient transmission of *P. ultimum* among seedlings. While this new experimental set up more closely approximated genuine crop infestation in an agricultural scenario, the model required modification to account for the seed proximity. In the original biocontrol model, the survival of each seed was independent of others due to the use of individual well potting trays, which was not reflective of field conditions where root systems are not physically separated (Figure 30 and Figure 31).

6.3.3. Third replicon deletion and attenuation of *Burkholderia* biocontrol agents

Following the deletion of the third replicon from our prototypic *B. ambifaria* BCC0191 biological control strain we did not observe a reduction of virulence in the *G. mellonella* wax-moth model. This was in contrast to the results of Agnoli *et al.* (2012) who observed the attenuated of virulence in *B. ambifaria* AMMD, in addition to multiple other *Burkholderia cepacia* complex species. Despite using an inoculum 13-fold greater, we observed a higher survival rate at the final 72 hour time point compared to published data (Agnoli *et al.* 2012). We therefore conjectured the existence of insecticidal factors encoded on replicons c1 and/or c2 that remained active despite the removal of the third replicon; as well as the overall lower virulence of strain BCC0191 compared to AMMD. The altogether lower virulence of *B. ambifaria* BCC0191 was substantiated in the murine lung inhalation model (Fothergill *et al.* 2014). *B. ambifaria* BCC0191 wild-type exhibited a low persistence in the murine respiratory model compared to other more virulent respiratory pathogens such as *P. aeruginosa* (Fothergill *et al.* 2014; Bricio-Moreno *et al.* 2018). The deletion of the third replicon in BCC0191 considerably reduced persistence and resulted in higher levels of clearance from the murine lung and nasopharynx. The safety of *Burkholderia* biological control strains regarding the susceptibility of CF patients was a key issue raised by the scientific advisory panel in 1999 (US

Environmental Protection Agency 1999a), and contributed to the decision to implement a moratorium on the registration of live-*Burkholderia* products (US Environmental Protection Agency 2004). By addressing the virulence of *B. ambifaria* BCC0191, and demonstrating attenuation through reductive engineering of the *B. ambifaria* genome in a model of significance to cystic fibrosis, we have consequently begun to address some of these concerns. These findings provide a rationale for progressing safe biocontrol agents through reductive genome engineering.

6.3.4. Replicon c3 contributes to rhizocompetence in *B. ambifaria* BCC0191

Despite the two-fold increase in cepacin biosynthesis with the deletion of replicon c3 in *B. ambifaria* BCC0191 we observed no significant change in biological control function compared to the wild-type (Figure 31). However, the mean survival rate was marginally reduced at all inoculum levels following the loss of the third replicon compared to the wild-type (Figure 31), suggesting additional factors may be influencing biocontrol function. A possible explanation for this lower trend was the loss of fitness traits associated with the third replicon that support the colonisation of root surfaces. This supported the results observed with *B. ambifaria* BCC0191 colonisation of the *P. sativum* root system following seed coat inoculation. The c3-deficient mutant colonised the root surface at a lower cell density compared to the wild-type counterpart. Evidence to support this has been observed in *B. ambifaria* AMMD colonisation of the *Arabidopsis thaliana* root system (Vidal-Quist *et al.* 2014). The AMMD c3-null mutant displayed a significantly reduced ability to colonise the root surface, yielding an approximately 1-log lower cell density per 2 cm root section (Vidal-Quist *et al.* 2014). Interestingly, this phenomenon was not observed in the four other *B. cepacia* complex species examined, suggesting that *B. ambifaria* encodes rhizocompetence fitness traits on the third replicon in contrast to other species. Removal of the replicon triggers the loss of extracellular polymeric substance (EPS) production in *B. ambifaria* AMMD and several other species (Agnoli *et al.* 2012), a trait that could aid root colonisation. There is also clear evidence that replicon c3 confers various stress tolerance traits in *B. cenocepacia* H111 such as heat, oxidative and osmotic stresses (Agnoli *et al.* 2014). Not only does this large plasmid appear to influence root colonisation, but confers an evolutionary advantage to *B. cepacia* complex members with a plethora of stress tolerances and virulence factors to support both environmental survival and host colonisation.

6.3.5. Seed exudates as elicitors of antimicrobial production

The rhizosphere is the zone of interaction between the plant root system and the surrounding soil bacteria. Plant roots are known to exude a plethora of compounds that facilitate communication with both beneficial and pathogenic microorganisms and pests (Massalha *et al.* 2017). A well-characterised plant-microbe interaction is that of legume crops and *Rhizobium* spp. The release of flavonoids by legumes function as attractants of *Rhizobium* spp. and subsequently regulate *Rhizobium nod* genes that in turn orchestrate nodule formation in the plant tissue (Abdel-Lateif *et al.* 2012). Few examples exist in the literature of seed exudates released from imbibing seeds regulating bacterial metabolite production. The lipopeptide amphisin is an antimicrobial synthesised by *Pseudomonas* sp. DSS73 with activity against *Rhizoctonia solani* and *Pythium ultimum* (Koch *et al.* 2002). The *Pseudomonas* strain was isolated from the sugar beet rhizosphere, and growth of *Pseudomonas* sp. DSS73 on media composed of sugar beet seed exudate stimulated amphisin production compared to media lacking seed exudate (Koch *et al.* 2002).

The significant role of cepacin in mediating biological control in a *P. sativum* model prompted further investigation into the specific induction mechanism in *B. ambifaria*. Seed exudate was extracted from pea seeds and incorporated into an agar-based medium, with no further nutrient supplements. The growth of *B. ambifaria* BCC0191 on the pea exudate-based medium triggered the induction of cepacin, similar to that seen with the use of artificial BSM-G medium (Figure 32). The use of nutrient rich TSA medium failed to induce cepacin, suggesting a specific nutrient signal within pea exudate and BSM-G is necessary to influence gene expression in *B. ambifaria* BCC0191. *B. ambifaria* BCC0203 exhibited antimicrobial activity on pea exudate-based medium; however, antimicrobial activity was also observed on TSA and BSM-G and therefore was not specifically elicited by the seed exudate. The natural variation in *B. ambifaria* strains and their response to nutrients and signals from host plants warrants further investigation.

6.4. Conclusions

The establishment of a robust biological control model and subsequent investigation of cepacin-mediated biocontrol, the contribution of replicon c3 to rhizocompetence, and induction of antimicrobial metabolites by seed exudates led to the following conclusions:

1) *P. sativum* was highly susceptible to low inoculum levels of oomycete *P. ultimum*.

2) Cepacin A was the primary metabolite contributing the biological control of *P. ultimum* in *B. ambifaria* BCC0191.

3) Loss of the third replicon caused no significant change in observed biological control.

4) Seed exudates proved effective in eliciting the production of cepacin A, indicating a mechanism of antimicrobial biosynthesis stimulation in the rhizosphere.

5) The deletion of replicon c3 significantly reduced the persistence of *B. ambifaria* BCC0191 in a murine respiratory infection model compared to the wild-type. However, no loss of virulence was observed in a *G. mellonella* wax-moth larvae infection assay.

7. General conclusions, discussion, and future work

7.1. Conclusions

This work sought to develop a comprehensive understanding of the genomics and specialised metabolites associated with the biopesticidal properties of the bacterium *B. ambifaria*. A diverse collection of *B. ambifaria* strains with either known biocontrol properties, or limited to no previous characterisation, were whole-genome sequenced, and the population biology of the species analysed with high-resolution core genome phylogenomics. The specialised metabolite biosynthetic potential of the 64 *B. ambifaria* genomes was determined by applying genome mining tools to the sequence data, and de-replicated through bioinformatics to accurately determine the biosynthetic gene cluster capacity across the species. The biosynthetic origin of the historical *Burkholderia* antimicrobial polyyne metabolite cepacin was elucidated, and the wider distribution and diversity of bacterial polyyne metabolites investigated. The virulence and antimicrobial resistance gene profiles were analysed for both chromosomal and plasmid components of the *B. ambifaria* pan-genome. Each *B. ambifaria* strain was assayed for antimicrobial activity against a broad collection of plant and animal pathogens. A biological control model was developed to determine the role of *B. ambifaria* antimicrobials in plant protection, and cepacin was identified to specifically protect the model crop species *P. sativum* against the damping-off oomycete *P. ultimum*. Finally, the *in vivo* virulence of *B. ambifaria* was assessed in both invertebrate wax-moth larvae and a murine respiratory inhalation infection model. The integration of these results provide a holistic understanding of the bacterium *B. ambifaria*, its potential as a biological control agent, and sets a precedent for harnessing its biotechnological function in the future.

The main conclusions of this study were as follows:

1) Considerable genomic diversity exists within *B. ambifaria* based on high-resolution phylogenomics, tailored virulence gene analysis and plasmid characterisation.

This work represents one of a few examples of the detailed genomic characterisation of a bacterium for biological control purposes, and the first instance of a *Burkholderia* species used commercially for this purpose. Average nucleotide identity confirmed the species designation of the *B. ambifaria* strain collection. The pan-genome of *B. ambifaria*, based on 64 genomes, equalled 22,376 genes, while the average gene content per genome was 6,576. This highlighted the considerable gene-carrying potential of the species with a pan-genome 3.4-fold larger than the average strain genome. A phylogeny based on a 3,784 core-gene alignment revealed three major phylogenetic clades to *B. ambifaria*; deep branches defined two of these clades, while the third was composed of several shallow-branch sub-clades. The mapping of known biocontrol strains revealed their broad distribution throughout the phylogeny, indicating that the biological control potential was not clade restricted. The antimicrobial resistance profile of *B. ambifaria* remained consistent across strains, with the exception of an AMR-carrying plasmid in strain BCC0478. In contrast, this study identified variation within the virulence factors and secretion systems possessed by different *B. ambifaria* strains, and their presence/absence mirrored the core-gene phylogenetic clades.

2) *B. ambifaria* possesses an abundance of specialised metabolite biosynthetic gene clusters, and displays potent broad-range antimicrobial activity.

Genome mining of specialised metabolite biosynthetic gene clusters within multiple strains of a single species exposed the considerable biosynthetic capacity of *B. ambifaria*. Such a detailed genomic analysis has been performed on very few other potential bacterial biocontrol strains, with exception to recent studies on *Bacillus thuringiensis*, and the evolution of its insecticidal Cry toxins and other virulence agents (Zheng *et al.* 2017). The *B. ambifaria* genome collection was initially predicted to encode 1,272 BGCs, subsequently de-replicated to 38 distinct BGCs. Only 34% of the BGCs were linked to known metabolites, which indicated extensive *intra*-species specialised metabolite diversity, and considerable potential for the discovery of novel, and as yet uncharacterised compounds. The genomic distribution of BGCs across the three replicons confirmed the third replicon, c3, as a hot-spot of BGC recombination and capture. Replicon c3 possessed the highest biosynthetic capacity in relation to replicon size compared to replicons c1 and c2. Replicating a previously described *luxR* regulatory gene mining strategy (Gan *et al.* 2014), highlighted the unexpected diversity and accessory nature

of orphan LuxR-encoding genes in *B. ambifaria*. The analysis confirmed that regulatory gene mining is a valid strategy for identifiying BGCs, and in parallel led to the identification of the biosynthetic origin of the historical *Burkholderia* antimicrobial, cepacin.

3) The polyyne class of specialised metabolites possesses a core region of biosynthetic genes, and represents a promiscuous BGC detected in multiple bacterial species.

This study uniquely investigated the antimicrobial properties and wider functions of the polyyne compound cepacin in *B. ambifaria*. The antagonism screening complemented the original documentation of the polyyne where strong anti-Gram negative activity was observed (Parker *et al.* 1984). Further screening revealed considerable *P. ultimum* antagonism, a property previously unreported in other bacterial derived polyynes. This observation was exploited in a biological control model to scrutinise the function of specific metabolites in biopesticidal agents. The discovery of the *B. ambifaria* polyyne BGC also prompted a detailed search for other bacterial polyynes in the literature, and subsequent screen of publically available genomes. A core region of biosynthetic genes was common to all bacterial polyyne BGC examples, which complemented the deduced biosynthetic logic of bacterial polyyne synthesis first described for the *B. gladioli* polyyne caryoynencin (Ross *et al.* 2014). An interesting observation was the substitution of regulatory regions in different polyyne BGCs, suggesting an optimisation to the bacterial host background. *B. ambifaria* represented a highly antagonistic species with antimicrobial activity against Gram-positive bacteria, Gram-negative bacteria, fungi, and oomycetes shown in this study. These antimicrobial activities appeared to be clade specific and mirrored the presence of previously characterised BGCs linked to antimicrobial metabolites.

4) The antimicrobial polyyne cepacin is a major functional component of the biological control properties of *B. ambifaria*.

This study developed a robust model for assaying the contribution of specific antimicrobial metabolites to the biological control phenotype of *B. ambifaria*. The production of the metabolite cepacin was observed to be a significant factor in protecting the germination and emergence of the model crop *P. sativum* against the oomycete *P. ultimum*. Seed exudate-based media also proved capable of stimulating the production of cepacin in *B. ambifaria*, implying an ecological signalling mechanism between the germinating plant host and surrounding *B. ambifaria* in the rhizosphere.

5) Loss of the third replicon reduced *B. ambifaria* rhizosphere fitness, but also reduced persistence in a murine lung inhalational model.

Deletion of the third replicon from *B. ambifaria* strain BCC0191 significantly reduced the persistence of the bacterium on the root surface, potentially due to the removal of genetic factors required for rhizocompetence. This mirrored the loss of rhizocompetence observed for c3-deletion mutant of *B ambifaria* strain AMMD (Vidal-Quist *et al.* 2014), however, the same study also demonstrated no change or a gain in rhizocompetence for other *B. cepacia* complex species. Despite the loss of the third-replicon, the derivative still conferred biological protection levels above the cepacin deficient mutant. The virulence of the c3-deficient derivative was assessed in the invertebrate *G. mellonella* wax-moth larvae model, but failed to exhibit a noticeable difference in virulence compared to the *B. ambifaria* wild-type. This result for strain BCC0191 was in contrast to other *B. ambifaria* strains and other *B. cepacia* complex species with third replicon-deficient derivatives that exhibited reduced virulence (Agnoli *et al.* 2012). The employment of a murine lung respiratory inhalation model revealed a reduction in lung and nasopharynx persistence following c3-deletion in *B. ambifaria* BCC0191 compared to the wild-type. Confirmation of reduced mammalian infective potential, coupled to a maintained biocontrol phenotype following loss of the third replicon offers a potential solution to the concerns of *Burkholderia* virulence toward immunocompromised individuals. This work presented a systematic means of assessing the biological control potential and virulence of potential biopesticidal bacteria.

7.2. Discussion and future work

B. ambifaria was successfully exploited as a biopesticide in several commercially available formulations in the US. However, a lack of genomic knowledge and taxonomic distinction, coupled with concerns over pathogenicity in immunocompromised and vulnerable individuals led to a moratorium on the use of live *Burkholderia* in biopesticidal formulations (US Environmental Protection Agency 1999b; US Environmental Protection Agency 1999a). Research into the biological control properties of *B. ambifaria* subsequently diminished. The discovery of the accessory nature of the third replicon, and consequent reduction in virulence (Agnoli *et al.* 2012) offered a means to improve the safety of *Burkholderia* pesticides. The research summarised in this thesis systematically defined the antimicrobial properties and biosynthetic content of *B. ambifaria*, and investigated the biocontrol properties and virulence of a third replicon deficient mutant. The relevance and importance of these results are discussed below, in addition to future work to progress research into virulence reduction and biocontrol properties of *B. ambifaria*.

7.2.1. Integrated approach required to understand biocontrol and virulence in *B. ambifaria*

The sequencing revolution in the last decade has generated a wealth of genomic data providing an unprecedented insight into the diversity of bacteria. The inundation of public databases with sequence data has afforded the opportunity to interrogate the intricacies of multiple bacterial traits with clinical, environmental and industrial significance. Advancements in the detection and prediction of specialised metabolite biosynthetic gene clusters has permitted the mining of large genomic datasets for uncharacterised BGCs. This research analysed the BGC capacity of *B. ambifaria* and mapped the distribution of antimicrobial-associated BGCs in relation to the population biology (Chapter 4). Similar studies have been performed in *Salinospora* which revealed the majority of BGCs lacked an associated metabolite, offering a considerable untapped resource of compounds for pharmaceutical use (Letzel *et al.* 2017).

Population biology has also been examined in *Klebsiella pneumoniae* to determine the distribution of virulence and AMR traits (Wyres *et al.* 2019). A 2,265 genome collection was analysed for recombination and horizontal gene transfer between drug-resistant and virulent clones, and provided an insight into the evolution and emergence of hyper-virulent, multidrug-resistant *K. pneumoniae* clones (Wyres *et al.* 2019). Initially, the 64 *B. ambifaria* genome collection was intended to understand the population biology of the species, and determine the biosynthetic capacity of individual strains and the whole species, in the context of biological

control. This research expanded the genomic analysis to predict the antimicrobial resistance and virulence factors possessed by *B. ambifaria*. The *in silico* virulence analyses in Chapter 4 were complemented by exploring the *in vivo* implications of large-scale genomic deletions on insect and murine models (Chapter 6).

Unlike most studies that exploit the genomic data in relatively one-dimensional analyses, this research provided a unique opportunity to explore several facets of a single species with both clinical and biotechnological implications. Several antimicrobial metabolites have been characterised in *B. ambifaria* with effective *in vitro* activity against plant pathogens demonstrating the capacity to inhibit undesirable microbes. Historical exploitation of *B. ambifaria* as a biopesticide confirmed the efficacy and commercial value of the bacterium in agriculture as an alternative to synthetic pesticides. From a clinical perspective, decades of research into *Burkholderia* infections in CF patients has unequivocally established the risk of *B. cepacia* species and *B. gladioli* to the CF community (LiPuma 2010). A multifarious approach was therefore required to define the basis of biological control and address the concerns over pathogenicity and the risk of acquiring infections from biopesticide products used in agriculture. An integrated genomic and experimental approach allowed this study to begin addressing the unanswered questions and concerns raised at the scientific advisory panel meeting in 1999 (US Environmental Protection Agency 1999a). Improvements in microbial identification have highlighted the absence of *B. ambifaria* in the UK CF community (Kenna *et al.* 2017); and in other CF populations such as the US its prevalence has remained very low at less than 3% (LiPuma 2010). Advancements in *Burkholderia* biology and genomics have presented techniques such as third replicon deletion to further reduce virulence while maintaining biological control. While these findings support a review of the US moratorium, the issue of releasing genetically modified bacteria into the environment, even genomic reductions lacking selection markers or transgenes, remains a potentially divisive regulatory issue.

7.2.2. Unexploited biosynthetic specialised metabolite diversity in *B. ambifaria*

One of the key objectives of this research was to determine the biosynthetic diversity of *B. ambifaria* beyond the published antimicrobial-associated BGCs (Chapters 4 and 5). Approximately two-thirds of the identified BGCs lacked an associated metabolite, and represent a source of novel compounds. However, most BGCs remain unexpressed under laboratory screening conditions; this represents a limitation of specialised metabolite discovery. An efficient pathway capture and expression mechanism is required to exploit the untapped metabolic diversity of *B. ambifaria*, and other *Burkholderia* species that uncouples

the BGC from its native local regulatory genes. These genetic systems have been established for capturing BGCs from environmental DNA as well as specific BGCs using a yeast-based transformation-associated homologous recombination (TAR) system (Kim *et al.* 2010). However, following homologous recombination in a yeast, TAR relies on *Streptomyces* to express the BGC. This requires the *Streptomyces* possessing the necessary biological mechanisms to assemble the mature proteins and final compound, and resistance to the resulting potentially toxic metabolite. Expressing BGCs in the native species traverses these concerns, and *Burkholderia*-based systems are starting to emerge in the research community. An *in situ* pathway activation system has been developed that exploits a *Burkholderia* recombinase to integrate an inducible promoter at a specific location in the genome, and has proven successful in the broader Burkholderiales order (Wang *et al.* 2018). This system was also capable of insertional inactivation of BGCs and deleting BGCs up to 200 kbp in length (Wang *et al.* 2018). *Burkholderia*-specific CRISPR interference systems have also emerged that efficiently silence gene expression without modifying the genomic sequence (Hogan *et al.* 2019). Advancements in the *Burkholderia* genomic toolkit will undoubtedly contribute to exploiting biosynthetic diversity of *Burkholderia* and related genera.

Mapping the distribution of known BGCs and correlating this to the *in vitro* antimicrobial activity against a broad range of plant pathogens enabled the interrogation of the ecological antagonism role of these metabolites. The correlation of the BGC distribution to pathogen antagonism resulted in the discovery of the biosynthetic origin of the historic *Burkholderia* antimicrobial cepacin. Polyynes possess an unusual metabolite structural feature of an unstable polyalkyne moiety purportedly responsible for their antimicrobial activity. Four bacterial polyynes have been structurally elucidated, three of which are restricted to the order Burkholderiales (Ross *et al.* 2014; Kai *et al.* 2018; Mullins *et al.* 2019), while the fourth was detected in the actinomycete *Microbispora* (Patel *et al.* 1988). Based on previous publications there was evidence of a polyyne BGC in *Pseudomonas* with homology to collimonins and caryoynencin (Ross *et al.* 2014; Kai *et al.* 2018). This research identified a shared biosynthetic region between the Burkholderiales polyyne BGCs, and the hypothetical *Pseudomonas* BGC suggesting a core base structure to the antimicrobial metabolites, that is subsequently tailored. The distribution of the polyyne core in phylogenetically diverse bacteria raises an interesting question surrounding the origin of the bacterial polyyne BGC assuming an ancestral prototypic polyyne BGC diverged into the various metabolites currently observed in bacteria. Understanding the diversification of polyyne BGCs and metabolite tailoring represents an important area of future research given that different bacterial polyyne metabolites possess different spectrums of antimicrobial activity, and the polyyne cepacin contributes to *B. ambifaria* biological control.

This research also explored the silent nature of BGCs by identifying several strains that failed to produce an associated metabolite *in vitro*. LC-MS was used to confirm expression and identity of known antimicrobial metabolites and highlighted an important caveat to genome-restricted analysis that the presence of a biosynthetic gene cluster did not guarantee *in vitro* expression. As a result, *in vitro* antagonism assays and the development of a robust model for assessing the biological control ability of different species and strains was essential to understanding the biopesticidal potential of strains beyond genomic analyses.

7.2.3. The persistence of *Burkholderia* in the rhizosphere

The persistence of both the *B. ambifaria* wild-type and third-replicon deletion mutant was examined on the *P. sativum* root following seed-coat inoculation (Chapter 6). The loss of the third replicon caused a significant reduction in cfu/root section compared to the wild-type over the 14-day experiment. A reduced ability to colonise the root system of *Arabidopsis thaliana* following loss of the third replicon has been demonstrated previously (Vidal-Quist *et al.* 2014). Loss of the c3 replicon also reduced virulence in a murine lung infection model, hinting towards a potential link between the ability to maintain high cell density in the rhizosphere and the ability to colonise the mammalian lung. To understand if these phenotypes share genetic loci a transposon-sequencing or RNA-sequencing strategy could be employed in future research. Applying a transposon library of *B. ambifaria* as a seed coat and identifying the mutants that fail to persist in the rhizosphere, or determining the genes expressed on the rhizosphere and comparing to known virulence genes would begin to elucidate the possible connection. A significant difference between *B. ambifaria* wild-type and third-replicon mutant was already apparent at 14-days post inoculation. However, a longer-term experiment is necessary to understand if the persistence of the c3-mutant stabilises at a lower cell density or continues to decrease over a growth season. One of the main concerns highlighted by the EPA scientific advisory panel in 1999 was a lack of knowledge surrounding the fate of *Burkholderia* biological control agents following application to crops, and the health implications of a potential artificially induced accumulation of *Burkholderia* in the environment (US Environmental Protection Agency 1999a).

7.2.4. Strains and species background potentially influences the biological control properties associated with a specific metabolite

The research presented in Chapter 6 focussed on the *B. ambifaria* strain BCC0191 (J82) which has been historically characterised as a biological control strain in a commercial product marketed by Stine Microbial Products in the 1990s. Based on *in vitro* antimicrobial analyses and HPLC comparisons the strain BCC1241 was identified as the highest producer of cepacin within the *B. ambifaria* collection. However, BCC1241 failed to confer greater protection in the refined biological control model compared to *B. ambifaria* BCC0191 (Table 20). This surprising result implied that the strain background has an impact on the biocontrol potential despite possessing the same antimicrobial-associated BGC and higher *in vitro* production. Consequently, *in vitro* antimicrobial screening for strains exhibiting greater pathogen antagonism does not guarantee efficient biological control against the same pathogen in an ecologically relevant system. This result has implications in biopesticide development and the concept of using *Paraburkholderia* or related genera carrying the same antimicrobial-associated BGCs as safer alternatives to *Burkholderia*. The use of *T. caryophylli* in the refined biological control model failed to exhibit biocontrol properties against *P. ultimum*, despite observed *in vitro* antagonism towards *P. ultimum* and confirmation of caryoynencin production by HPLC. *Paraburkholderia* and related genera may be capable of biological control against other pathogens as previously suggested (Eberl and Vandamme 2016), however, they appear to lack polyyne production under conditions ecologically relevant to agriculture.

To summarise, the research undertaken during this PhD has integrated genomic analyses and biological control systems to understand the fundamental basis of the biopesticidal properties of *B. ambifaria*. By systematically defining the specialised metabolite potential, virulence factors, and *in vitro* antimicrobial activity of *B. ambifaria*, coupled to strategies of reducing virulence in relevant models, this research has begun to address the concerns surrounding *Burkholderia* biopesticides.

References

Abdel-Lateif, K. *et al.* 2012. The role of flavonoids in the establishment of plant roots endosymbioses with arbuscular mycorrhiza fungi, rhizobia and *Frankia* bacteria. *Plant signaling & behavior* 7(6), pp. 636–41. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22580697 [Accessed: 1 July 2019].

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. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22171913 [Accessed: 10 April 2017].

Agnoli, K. *et al.* 2014. The third replicon of members of the *Burkholderia cepacia* Complex, plasmid pC3, plays a role in stress tolerance. *Applied and environmental microbiology* 80(4), pp. 1340–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24334662 [Accessed: 13 June 2019].

Agnoli, K. *et al.* 2017. Use of synthetic hybrid strains to determine the role of replicon 3 in virulence of the *Burkholderia cepacia* complex. *Applied and Environmental Microbiology* 83(13), pp. e00461-17. Available at: https://aem.asm.org/content/83/13/e00461-17.short [Accessed: 13 June 2019].

Aguilar, C. *et al.* 2003. Identification of quorum-sensing-regulated genes of *Burkholderia cepacia*. *Journal of bacteriology* 185(21), pp. 6456–62. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14563881 [Accessed: 22 August 2017].

Allwood, E.M. *et al.* 2011. Strategies for intracellular survival of *Burkholderia pseudomallei*. *Frontiers in microbiology* 2, p. 170. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22007185 [Accessed: 13 June 2019].

Andrews, S. 2009. FastQC: A quality control tool for high throughput sequence data. *Babraham Bioinformatics* . Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ [Accessed: 7 April 2017].

Angus, A.A. *et al.* 2014. Plant-associated symbiotic *Burkholderia* species lack hallmark strategies required in mammalian pathogenesis. van Schaik, W. ed. *PLoS ONE* 9(1), p. e83779. Available at: https://dx.plos.org/10.1371/journal.pone.0083779 [Accessed: 15 February 2019].

Antipov, D. *et al.* 2016. plasmidSPAdes: assembling plasmids from whole genome sequencing data. *Bioinformatics* 32(22), p. btw493. Available at: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btw493 [Accessed: 5 December 2018].

Arima, K. *et al.* 1964. Pyrrolnitrin, a new antibiotic substance, produced by *Pseudomonas*. *Agricultural and Biological Chemistry* 28(8), pp. 575–576. Available at: http://www.tandfonline.com/doi/full/10.1080/00021369.1964.10858275 [Accessed: 10 August 2018].

Aunkham, A. *et al.* 2014. Porin involvement in cephalosporin and carbapenem resistance of *Burkholderia pseudomallei*. van Veen, H. W. ed. *PLoS ONE* 9(5), p. e95918. Available at: http://dx.plos.org/10.1371/journal.pone.0095918 [Accessed: 13 February 2019].

Authority, E.F.S. 2017. Peer review of the pesticide risk assessment of the active substance thiram. *EFSA Journal* 15(7). Available at: http://doi.wiley.com/10.2903/j.efsa.2017.4700 [Accessed: 17 May 2018].

Bailey, K.L. *et al.* 2010. Social and economic drivers shaping the future of biological control: A Canadian perspective on the factors affecting the development and use of microbial biopesticides. *Biological Control* 52(3), pp. 221–229. Available at: https://www.sciencedirect.com/science/article/pii/S1049964409001327 [Accessed: 7

July 2019].

Balandreau, J. *et al.* 2001. *Burkholderia cepacia* genomovar III Is a common plant-associated bacterium. *Applied and environmental microbiology* 67(2), pp. 982–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11157274 [Accessed: 12 July 2019].

Baldwin, A. *et al.* 2005. Multilocus sequence typing scheme that provides both species and strain differentiation for the *Burkholderia cepacia* complex. *Journal of clinical microbiology* 43(9), pp. 4665–73. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16145124 [Accessed: 26 September 2018].

Bankevich, A. *et al.* 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology* 19(5), pp. 455–77. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22506599 [Accessed: 7 April 2017].

Barakat, M. *et al.* 2013. P2RP: a web-based framework for the identification and analysis of regulatory proteins in prokaryotic genomes. *BMC Genomics* 14(1), p. 269. Available at: http://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-14-269 [Accessed: 4 July 2018].

Barelmann, I. *et al.* 1996. Cepaciachelin, A New Catecholate Siderophore From *Burkholderia* (*Pseudomonas*) *cepacia*. *Zeitschrift für Naturforschung C* 51(9–10), pp. 627–630. Available at: https://www.degruyter.com/view/j/znc.1996.51.issue-9-10/znc-1996-9-1004/znc-1996-9-1004.xml [Accessed: 18 May 2018].

Bay, D.C. *et al.* 2008. Small multidrug resistance proteins: A multidrug transporter family that continues to grow. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1778(9), pp. 1814–1838. Available at: https://www.sciencedirect.com/science/article/pii/S000527360700301X [Accessed: 13 February 2019].

Bevivino, A. *et al.* 1998. Characterization of a free-living maize-rhizosphere population of *Burkholderia cepacia*: effect of seed treatment on disease suppression and growth promotion of maize. *FEMS Microbiology Ecology* 27(3), pp. 225–237. Available at: https://www.sciencedirect.com/science/article/pii/S0168649698000695 [Accessed: 2 November 2018].

Biggins, J.B. *et al.* 2012. Malleilactone, a polyketide synthase-derived virulence factor encoded by the cryptic secondary metabolome of *Burkholderia pseudomallei* group pathogens. *Journal of the American Chemical Society* 134(32), pp. 13192–13195. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22765305 [Accessed: 12 November 2018].

Bisacchi, G.S. *et al.* 1987. Xylocandin: A new complex of antifungal peptides. II. Structural studies and chemical modifications. *The Journal of Antibiotics* 40(11), pp. 1520–1529. Available at: http://joi.jlc.jst.go.jp/JST.Journalarchive/antibiotics1968/40.1520?from=CrossRef [Accessed: 5 April 2017].

Blin, K. *et al.* 2017. antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Research* 45(W1), pp. W36–W41. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28460038 [Accessed: 7 December 2017].

Boon, C. *et al.* 2008. A novel DSF-like signal from *Burkholderia cenocepacia* interferes with *Candida albicans* morphological transition. *The ISME Journal* 2(1), pp. 27–36. Available at: http://www.nature.com/articles/ismej200776 [Accessed: 17 September 2019].

Boran, R. and Ugur, A. 2016. Burkholderia multivorans SB6 Lipase as a Detergent Ingredient: Characterization and Stabilization. *Journal of Surfactants and Detergents* 19(1), pp. 39–48. Available at:

http://doi.wiley.com/10.1007/s11743-015-1767-6 [Accessed: 10 July 2019].

Brenner, D.J. *et al.* 1969. Batch procedure for thermal elution of DNA from hydroxyapatite. *Analytical Biochemistry* 28, pp. 447–459. Available at: https://www.sciencedirect.com/science/article/pii/0003269769901997 [Accessed: 26 September 2018].

Bricio-Moreno, L. *et al.* 2018. Evolutionary trade-offs associated with loss of PmrB function in host-adapted *Pseudomonas aeruginosa*. *Nature Communications* 9(1), p. 2635. Available at: http://www.nature.com/articles/s41467-018-04996-x [Accessed: 26 October 2018].

Vander Broek, C.W. and Stevens, J.M. 2017. Type III secretion in the melioidosis pathogen *Burkholderia pseudomallei*. *Frontiers in Cellular and Infection Microbiology* 7, p. 255. Available at: http://journal.frontiersin.org/article/10.3389/fcimb.2017.00255/full [Accessed: 15 February 2019].

Burkholder, W.H. 1950. Sour skin, a bacterial rot of onion bulbs. *Phytopathology* 40, pp. 115–17. Available at: https://www.cabdirect.org/cabdirect/abstract/19500300856 [Accessed: 2 July 2019].

Butt, A.T. and Thomas, M.S. 2017. Iron acquisition mechanisms and their role in the virulence of *Burkholderia* species. *Frontiers in Cellular and Infection Microbiology* 7, p. 460. Available at: https://www.frontiersin.org/article/10.3389/fcimb.2017.00460/full [Accessed: 6 July 2019].

Byrne-Bailey, K.G. *et al.* 2009. Prevalence of sulfonamide resistance genes in bacterial isolates from manured agricultural soils and pig slurry in the United Kingdom. *Antimicrobial agents and chemotherapy* 53(2), pp. 696–702. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19064898 [Accessed: 13 February 2019].

Cartwright, D.K. *et al.* 1995. Pyrrolnitrin and phenazine production by *Pseudomonas cepacia*, strain 5.5B, a biocontrol agent of *Rhizoctonia solani*. *Applied Microbiology and Biotechnology* 43(2), pp. 211–216. Available at: http://link.springer.com/10.1007/BF00172814 [Accessed: 17 May 2018].

Di Cello, F. *et al.* 1997. Biodiversity of a *Burkholderia cepacia* population isolated from the maize rhizosphere at different plant growth stages. *Applied and environmental microbiology* 63(11), pp. 4485–93. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9361434 [Accessed: 9 April 2018].

Ceniceros, A. *et al.* 2017. Genome-based exploration of the specialized metabolic capacities of the genus Rhodococcus. *BMC Genomics* 18(1), p. 593. Available at: http://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-017-3966-1 [Accessed: 23 August 2018].

Chapalain, A. *et al.* 2013. Identification of quorum sensing-controlled genes in *Burkholderia ambifaria*. *MicrobiologyOpen* 2(2), pp. 226–42. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23382083 [Accessed: 21 May 2018].

Chapalain, A. *et al.* 2017. Interplay between 4-Hydroxy-3-Methyl-2-Alkylquinoline and N-Acyl-Homoserine Lactone Signaling in a *Burkholderia cepacia* complex clinincal strain. *Frontiers in Microbiology* 8, p. 1021. Available at: http://journal.frontiersin.org/article/10.3389/fmicb.2017.01021/full [Accessed: 22 August 2017].

Chevrette, M.G. and Currie, C.R. 2019. Emerging evolutionary paradigms in antibiotic discovery. *Journal of Industrial Microbiology & Biotechnology* 46(3–4), pp. 257–271. Available at: http://link.springer.com/10.1007/s10295-018-2085-6 [Accessed: 6 July 2019].

Choudhary, K.S. *et al.* 2013. The organization of the quorum sensing luxI/R family genes in *Burkholderia*. *International journal of molecular sciences* 14(7), pp. 13727–47. Available at:

http://www.ncbi.nlm.nih.gov/pubmed/23820583 [Accessed: 7 September 2017].

Coenye, T. *et al.* 2001a. *Burkholderia ambifaria* sp. nov., a novel member of the *Burkholderia cepacia* complex including biocontrol and cystic fibrosis-related isolates. *International Journal of Systematic and Evolutionary Microbiology* 51(4), pp. 1481–1490. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11491349 [Accessed: 9 April 2018].

Coenye, T. *et al.* 2001b. *Burkholderia fungorum* sp. nov. and *Burkholderia caledonica* sp. nov., two new species isolated from the environment, animals and human clinical samples. *INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY* 51(3), pp. 1099–1107. Available at: http://ijs.microbiologyresearch.org/content/journal/ijsem/10.1099/00207713-51-3-1099 [Accessed: 16 July 2019].

Coenye, T. *et al.* 2001c. Taxonomy and identification of the *Burkholderia cepacia* complex. *Journal of clinical microbiology* 39(10), pp. 3427–36. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11574551 [Accessed: 26 September 2018].

Coenye, T. *et al.* 2002. Characterization of unusual bacteria isolated from respiratory secretions of cystic fibrosis patients and description of *Inquilinus limosus* gen. nov., sp. nov. *Journal of clinical microbiology* 40(6), pp. 2062–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12037065 [Accessed: 16 July 2019].

Coleman, J.J. 2016. The *Fusarium solani* species complex: ubiquitous pathogens of agricultural importance. *Molecular Plant Pathology* 17(2), pp. 146–158. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26531837 [Accessed: 29 October 2018].

Commission, E. 2018. *Final renewal report for the active substance thiram*.

Compant, S. *et al.* 2008. Diversity and occurrence of *Burkholderia* spp. in the natural environment. *FEMS Microbiology Reviews* 32(4), pp. 607–626. Available at: https://academic.oup.com/femsre/article-lookup/doi/10.1111/j.1574-6976.2008.00113.x [Accessed: 14 July 2019].

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). Available at: http://www.microbiologyresearch.org/content/journal/mgen/10.1099/mgen.0.000086 [Accessed: 24 August 2017].

Cordova-Kreylos, A.L. *et al.* 2013. Isolation and characterization of Burkholderia rinojensis sp. nov., a non-Burkholderia cepacia complex soil bacterium with insecticidal and miticidal activities. *Applied and environmental microbiology* 79(24), pp. 7669–78. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24096416 [Accessed: 30 June 2019].

Cuccui, J. *et al.* 2012. Characterization of the *Burkholderia pseudomallei* K96243 capsular polysaccharide I coding region. *Infection and immunity* 80(3), pp. 1209–21. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22252864 [Accessed: 23 February 2019].

Dalmastri *et al.* 1999. Soil Type and Maize Cultivar Affect the Genetic Diversity of Maize Root-Associated *Burkholderia cepacia* Populations. *Microbial ecology* 38(3), pp. 273–284. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10541789 [Accessed: 9 April 2018].

Demain, A.L. and Fang, A. 2000. The natural functions of secondary metabolites. Springer, Berlin, Heidelberg, pp. 1–39. Available at: http://link.springer.com/10.1007/3-540-44964-7_1 [Accessed: 26 April 2018].

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. Available at: http://link.springer.com/10.1007/s00253-016-7520-x [Accessed: 5 April 2017].

Deris, Z.Z. *et al.* 2010. First isolation of Burkholderia tropica from a neonatal patient successfully treated with imipenem. *International Journal of Infectious Diseases* 14(1), pp. e73–e74. Available at: https://www.sciencedirect.com/science/article/pii/S1201971209001404?via%3Dihub [Accessed: 16 July 2019].

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. Available at: http://www.microbiologyresearch.org/content/journal/ijsem/10.1099/ijsem.0.001065 [Accessed: 4 July 2019].

Duerkop, B.A. *et al.* 2009. Quorum-sensing control of antibiotic synthesis in *Burkholderia thailandensis*. *Journal of Bacteriology* 191(12), pp. 3909–3918. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19376863 [Accessed: 5 April 2017].

Eberl, L. and Vandamme, P. 2016. Members of the genus *Burkholderia*: good and bad guys. *F1000Research* 5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27303639 [Accessed: 5 April 2017].

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. Available at: http://www.mdpi.com/2073-4425/9/8/389 [Accessed: 4 July 2019].

Estrada-De Los Santos, P. *et al.* 2001. *Burkholderia* , a genus rich in plant-associated nitrogen fixers with wide environmental and geographic distribution. *Applied and environmental microbiology* 67(6), pp. 2790–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11375196 [Accessed: 14 July 2019].

Finn, R.D. *et al.* 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research* 39(suppl), pp. W29–W37. Available at: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkr367 [Accessed: 1 June 2019].

Fisher, M.C. *et al.* 2012. Emerging fungal threats to animal, plant and ecosystem health. *Nature* 484(7393), pp. 186–194. Available at: http://www.nature.com/articles/nature10947 [Accessed: 7 July 2019].

Flannagan, R.S. *et al.* 2007. *Burkholderia cenocepacia* requires a periplasmic HtrA protease for growth under thermal and osmotic stress and for survival *in vivo*. *Infection and Immunity* 75(4), pp. 1679–1689. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17220310 [Accessed: 8 May 2018].

Fleischmann, R. *et al.* 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269(5223), pp. 496–512. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7542800 [Accessed: 18 July 2019].

Flórez, L. V. *et al.* 2017. Antibiotic-producing symbionts dynamically transition between plant pathogenicity and insect-defensive mutualism. *Nature Communications* 8, p. 15172. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28452358 [Accessed: 16 June 2017].

Fothergill, J.L. *et al.* 2014. *Pseudomonas aeruginosa* adaptation in the nasopharyngeal reservoir leads to migration and persistence in the lungs. *Nature Communications* 5(1), p. 4780. Available at: http://www.nature.com/articles/ncomms5780 [Accessed: 26 October 2018].

Freschi, L. *et al.* 2019. The *Pseudomonas aeruginosa* Pan-Genome Provides New Insights on Its Population Structure, Horizontal Gene Transfer, and Pathogenicity. Martin, B. ed. *Genome Biology and Evolution* 11(1), pp. 109–120. Available at: https://academic.oup.com/gbe/article/11/1/109/5215156 [Accessed: 17 June 2019].

Fritsche, K. *et al.* 2014. Biosynthetic genes and activity spectrum of antifungal polyynes from *Collimonas fungivorans* Ter331. *Environmental Microbiology* 16(5), pp. 1334–1345. Available at: http://doi.wiley.com/10.1111/1462-2920.12440 [Accessed: 4 August 2017].

Funston, S.J. *et al.* 2016. Characterising rhamnolipid production in *Burkholderia thailandensis* E264, a non-pathogenic producer. *Applied Microbiology and Biotechnology* 100(18), pp. 7945–7956. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27147528 [Accessed: 13 December 2017].

Galardini, M. *et al.* 2011. CONTIGuator: a bacterial genomes finishing tool for structural insights on draft genomes. *Source code for biology and medicine* 6, p. 11. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21693004 [Accessed: 6 April 2017].

Gallagher, L.A. *et al.* 2002. Functions required for extracellular quinolone signaling by *Pseudomonas aeruginosa*. *Journal of bacteriology* 184(23), pp. 6472–80. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12426334 [Accessed: 17 May 2018].

Gan, H.M. *et al.* 2014. Whole genome sequencing and analysis reveal insights into the genetic structure, diversity and evolutionary relatedness of *luxI* and *luxR* homologs in bacteria belonging to the Sphingomonadaceae family. *Frontiers in cellular and infection microbiology* 4, p. 188. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25621282 [Accessed: 29 April 2017].

Gerrits, G.P. *et al.* 2005. *Burkholderia fungorum* septicemia. *Emerging infectious diseases* 11(7), pp. 1115–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16022793 [Accessed: 16 July 2019].

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 Bacteriology* 45(2), pp. 274–289. Available at: http://ijs.microbiologyresearch.org/content/journal/ijsem/10.1099/00207713-45-2-274 [Accessed: 2 July 2019].

Ginther, J.L. *et al.* 2015. Identification of *Burkholderia pseudomallei* near-Neighbor species in the Northern Territory of Australia. Yang, R. ed. *PLOS Neglected Tropical Diseases* 9(6), p. e0003892. Available at: https://dx.plos.org/10.1371/journal.pntd.0003892 [Accessed: 3 July 2019].

Glare, T. *et al.* 2012. Have biopesticides come of age? *Trends in Biotechnology* 30(5), pp. 250–258. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0167779912000042 [Accessed: 7 July 2019].

Goodlet, K.J. *et al.* 2019. Successful lung re-transplant in a patient with cepacia syndrome due to *Burkholderia ambifaria*. *Journal of Cystic Fibrosis* 18(1), pp. e1–e4. Available at: http://www.ncbi.nlm.nih.gov/pubmed/30224331 [Accessed: 1 August 2019].

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. Available at: http://ijs.microbiologyresearch.org/content/journal/ijsem/10.1099/ijs.0.63101-0 [Accessed: 16 July 2019].

Goris, J. *et al.* 2007. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology* 57(1), pp. 81–91. Available at:

http://www.microbiologyresearch.org/content/journal/ijsem/10.1099/ijs.0.64483-0 [Accessed: 26 September 2018].

Gould, K. 2016. Antibiotics: from prehistory to the present day. *Journal of Antimicrobial Chemotherapy* 71(3), pp. 572–575. Available at: https://academic.oup.com/jac/article-lookup/doi/10.1093/jac/dkv484 [Accessed: 5 July 2019].

Govan, J.R.. *et al.* 1993. Evidence for transmission of *Pseudomonas cepacia* by social contact in cystic fibrosis. *The Lancet* 342(8862), pp. 15–19. Available at: https://linkinghub.elsevier.com/retrieve/pii/014067369391881L [Accessed: 30 June 2019].

Groenhagen, U. *et al.* 2013. Production of bioactive volatiles by different *Burkholderia ambifaria* strains. *Journal of Chemical Ecology* 39(7), pp. 892–906. Available at: http://link.springer.com/10.1007/s10886-013-0315-y [Accessed: 14 August 2018].

Gu, G. *et al.* 2009a. AmbR1 is a key transcriptional regulator for production of antifungal activity of *Burkholderia contaminans* strain MS14. *FEMS Microbiology Letters* 297(1), pp. 54–60. Available at: https://academic.oup.com/femsle/article-lookup/doi/10.1111/j.1574-6968.2009.01653.x [Accessed: 21 May 2018].

Gu, G. *et al.* 2009b. Biosynthesis of an antifungal oligopeptide in *Burkholderia contaminans* strain MS14. *Biochemical and Biophysical Research Communications* 380(2), pp. 328–332. Available at: https://www.sciencedirect.com/science/article/pii/S0006291X09001156?via%3Dihub [Accessed: 14 May 2018].

Guijarro, J.A. *et al.* 2015. Temperature-dependent expression of virulence genes in fish-pathogenic bacteria. *Frontiers in Microbiology* 6, p. 700. Available at: http://journal.frontiersin.org/Article/10.3389/fmicb.2015.00700/abstract [Accessed: 13 December 2017].

Gurevich, A. *et al.* 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29(8), pp. 1072–1075. Available at: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btt086 [Accessed: 28 March 2018].

Haas, D. and Keel, C. 2003. Regulation of antibiotic production in root-colonizing *Pseudomonas* spp. And relevance for biological control of plant disease. *Annual Review of Phytopathology* 41(1), pp. 117–153. Available at: http://www.annualreviews.org/doi/10.1146/annurev.phyto.41.052002.095656 [Accessed: 20 December 2017].

Hammer, P.E. *et al.* 1997. Four genes from *Pseudomonas fluorescens* that encode the biosynthesis of pyrrolnitrin. *Applied and environmental microbiology* 63(6), pp. 2147–54. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9172332 [Accessed: 2 May 2018].

Hanauske, A.R. *et al.* 1996. Phase II clinical trials with rhizoxin in breast cancer and melanoma. The EORTC Early Clinical Trials Group. *British journal of cancer* 73(3), pp. 397–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8562349 [Accessed: 9 July 2019].

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. 272–85. Available at: http://www.ncbi.nlm.nih.gov/pubmed/234937 [Accessed: 9 April 2018].

Hebbar, K.P. *et al.* 1992. Suppression of *Fusarium moniliforme* by maize root-associated *Pseudomonas cepacia*. *Soil Biology and Biochemistry* 24(10), pp. 1009–1020. Available at: https://www.sciencedirect.com/science/article/pii/003807179290029W [Accessed: 9 August 2018].

Hedges, R.W. *et al.* 1974. Properties of an R Factor from *Bordetella bronchiseptica*. *Journal of General Microbiology* 84(1), pp. 199–204. Available at: http://mic.microbiologyresearch.org/content/journal/micro/10.1099/00221287-84-1-199 [Accessed: 14 June 2019].

Heungens, K. and Parke, J.L. 2000. Zoospore homing and infection events: effects of the biocontrol bacterium *Burkholderia cepacia* AMMDR1 on two oomycete pathogens of pea (*Pisum sativum* L.). *Applied and environmental microbiology* 66(12), pp. 5192–200. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11097889 [Accessed: 3 May 2017].

Ho, C.-C. *et al.* 2011. Novel pan-genomic analysis approach in target selection for multiplex PCR identification and detection of *Burkholderia pseudomallei*, *Burkholderia thailandensis*, and *Burkholderia cepacia* complex species: a proof-of-concept study. *Journal of clinical microbiology* 49(3), pp. 814–21. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21177905 [Accessed: 26 September 2018].

Hogan, A.M. *et al.* 2019. A broad-host range CRISPRi toolkit for silencing gene expression in *Burkholderia*. *bioRxiv* , p. 618413. Available at: https://www.biorxiv.org/content/10.1101/618413v1 [Accessed: 25 July 2019].

Howden, A.J.M. *et al.* 2009. *Pseudomonas syringae* pv. *syringae* B728a hydrolyses indole-3-acetonitrile to the plant hormone indole-3-acetic acid. *Molecular Plant Pathology* 10(6), pp. 857–865. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19849791 [Accessed: 7 April 2017].

Hudaiberdiev, S. *et al.* 2015. Census of solo LuxR genes in prokaryotic genomes. *Frontiers in cellular and infection microbiology* 5, p. 20. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25815274 [Accessed: 7 September 2017].

Hunt, T.A. *et al.* 2004. Identification of *Burkholderia cenocepacia* genes required for bacterial survival in vivo. *Infection and immunity* 72(7), pp. 4010–22. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15213146 [Accessed: 13 June 2019].

Hyatt, D. *et al.* 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11(1), p. 119. Available at: http://www.biomedcentral.com/1471-2105/11/119 [Accessed: 4 July 2018].

Imam, S. *et al.* 2011. Identification of surprisingly diverse type IV pili, across a broad range of Gram-positive bacteria. Donlin, M. J. ed. *PLoS ONE* 6(12), p. e28919. Available at: http://dx.plos.org/10.1371/journal.pone.0028919 [Accessed: 18 February 2019].

Ishida, K. *et al.* 2010. Induced Biosynthesis of Cryptic Polyketide Metabolites in a *Burkholderia thailandensis* Quorum Sensing Mutant. *Journal of the American Chemical Society* 132(40), pp. 13966–13968. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20853892 [Accessed: 28 March 2018].

Isles, A. *et al.* 1984. *Pseudomonas cepacia* infection in cystic fibrosis: An emerging problem. *The Journal of Pediatrics* 104(2), pp. 206–210. Available at: https://www.sciencedirect.com/science/article/pii/S0022347684809932?via%3Dihub [Accessed: 29 June 2019].

Jain, C. *et al.* 2017. High-throughput ANI Analysis of 90K Prokaryotic Genomes Reveals Clear Species Boundaries. *bioRxiv* , p. 225342. Available at: https://www.biorxiv.org/content/early/2017/11/27/225342 [Accessed: 27 July 2018].

Jenner, M. *et al.* 2019. An unusual *Burkholderia gladioli* double chain-initiating nonribosomal peptide synthetase assembles ‘fungal’ icosalide antibiotics. *Chemical Science* 10(21), pp. 5489–5494. Available at:

http://xlink.rsc.org/?DOI=C8SC04897E [Accessed: 6 July 2019].

Jenul, C. *et al.* 2018. Biosynthesis of fragin is controlled by a novel quorum sensing signal. *Nature Communications* 9(1), p. 1297. Available at: http://www.nature.com/articles/s41467-018-03690-2 [Accessed: 1 August 2019].

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. Available at: https://www.sciencedirect.com/science/article/pii/0040403995023429 [Accessed: 17 May 2018].

Jolley, K.A. *et al.* 2012. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. *Microbiology* 158(Pt_4), pp. 1005–1015. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22282518 [Accessed: 26 September 2018].

Jones, P. *et al.* 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics (Oxford, England)* 30(9), pp. 1236–40. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24451626 [Accessed: 4 July 2018].

de Jong, A. *et al.* 2006. BAGEL: a web-based bacteriocin genome mining tool. *Nucleic acids research* 34(Web Server issue), pp. W273-9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16845009 [Accessed: 23 August 2018].

Kai, K. *et al.* 2018. Collimonins A–D, Unstable Polyynes with Antifungal or Pigmentation Activities from the Fungus-Feeding Bacterium *Collimonas fungivorans* Ter331. *Organic Letters* 20(12), pp. 3536–3540. Available at: http://pubs.acs.org/doi/10.1021/acs.orglett.8b01311 [Accessed: 21 August 2018].

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 solani*. *Applied and environmental microbiology* 64(10), pp. 3939–47. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9758823 [Accessed: 5 April 2017].

Katoh, K. *et al.* 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research* 30(14), pp. 3059–66. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12136088 [Accessed: 30 October 2017].

Katoh, K. and Standley, D.M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* 30(4), pp. 772–80. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23329690 [Accessed: 4 July 2018].

Kautsar, S.A. *et al.* 2017. plantiSMASH: automated identification, annotation and expression analysis of plant biosynthetic gene clusters. *Nucleic Acids Research* 45(W1), pp. W55–W63. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28453650 [Accessed: 23 August 2018].

Kenna, D.T.D. *et al.* 2017. Prevalence of *Burkholderia* species, including members of *Burkholderia cepacia* complex, among UK cystic and non-cystic fibrosis patients. *Journal of Medical Microbiology* 66(4), pp. 490–501. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28463663 [Accessed: 29 October 2018].

Kilani-Feki, O. *et al.* 2011. Environmental *Burkholderia cepacia* strain Cs5 acting by two analogous alkyl-Quinolones and a didecyl-phthalate against a broad spectrum of phytopathogens fungi. *Current Microbiology* 62(5), pp. 1490–1495. Available at: http://link.springer.com/10.1007/s00284-011-9892-6 [Accessed: 6 July 2019].

Kim, J. and Yi, G.-S. 2012. PKMiner: a database for exploring type II polyketide synthases. *BMC Microbiology*

12(1), p. 169. Available at: http://bmcmicrobiol.biomedcentral.com/articles/10.1186/1471-2180-12-169 [Accessed: 23 August 2018].

Kim, J.H. *et al.* 2010. Cloning large natural product gene clusters from the environment: Piecing environmental DNA gene clusters back together with TAR. *Biopolymers* 93(9), pp. 833–844. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20577994 [Accessed: 25 July 2019].

King, E.B. and Parke, J.L. 1993. Biocontrol of *Aphanomyces* root rot and *Pythium* damping-off by *Pseudomonas cepacia* AMMD on four pea cultivars. *Plant Disease* 77(12), p. 1185. Available at: http://www.apsnet.org/publications/PlantDisease/BackIssues/Documents/1993Abstracts/PD_77_1185.htm [Accessed: 3 May 2017].

Kiratisin, P. and Sanmee, S. 2008. Roles and interactions of *Burkholderia pseudomallei* BpsIR quorum-sensing system determinants. *Journal of bacteriology* 190(21), pp. 7291–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18757538 [Accessed: 15 February 2019].

Knappe, T.A. *et al.* 2009. Insights into the Biosynthesis and Stability of the Lasso Peptide Capistruin. *Chemistry & Biology* 16(12), pp. 1290–1298. Available at: https://www.sciencedirect.com/science/article/pii/S1074552109004001 [Accessed: 5 July 2019].

Koch, B. *et al.* 2002. Lipopeptide production in *Pseudomonas* sp. strain DSS73 is regulated by components of sugar beet seed exudate via the Gac two-component regulatory system. *Applied and Environmental Microbiology* 68(9), pp. 4509–4516. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12200307 [Accessed: 1 July 2019].

Krueger, F. 2016. Trim Galore! Available at: https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ [Accessed: 7 April 2017].

Kurtz, S. *et al.* 2004. Versatile and open software for comparing large genomes. 5(2), p. 12. Available at: http://www.tigr.org/software/mummer. [Accessed: 26 September 2018].

Kusumi, T. *et al.* 1987. Caryoynencins, potent antibiotics from a plant pathogen pseudomonas caryophylli. *Tetrahedron Letters* 28(34), pp. 3981–3984. Available at: https://www.sciencedirect.com/science/article/pii/S0040403900964372 [Accessed: 1 May 2018].

Lamichhane, J.R. *et al.* 2017. Integrated management of damping-off diseases. A review. *Agronomy for Sustainable Development* 37(2), p. 10. Available at: http://link.springer.com/10.1007/s13593-017-0417-y [Accessed: 30 June 2019].

Lapage, S. *et al.* 1992. Rules of Nomenclature with Recommendations. Available at: https://www.ncbi.nlm.nih.gov/books/NBK8808/ [Accessed: 2 July 2019].

Laslett, D. and Canback, B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Research* 32(1), pp. 11–16. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14704338 [Accessed: 28 September 2018].

Leão, S.C. *et al.* 2013. The Detection and Sequencing of a Broad-Host-Range Conjugative IncP-1β Plasmid in an Epidemic Strain of Mycobacterium abscessus subsp. bolletii. Ahmed, N. ed. *PLoS ONE* 8(4), p. e60746. Available at: https://dx.plos.org/10.1371/journal.pone.0060746 [Accessed: 14 June 2019].

Lee, C.-H. *et al.* 1994. Cepacidine A, a novel antifungal antibiotic produced by *Pseudomonas cepacia*. I. Taxonomy, production, isolation and biological activity. *The Journal of Antibiotics* 47(12), pp. 1402–1405.

Available at: http://joi.jlc.jst.go.jp/JST.Journalarchive/antibiotics1968/47.1402?from=CrossRef [Accessed: 14 May 2018].

Lee, H.S. *et al.* 2010. CdpA is a *Burkholderia pseudomallei* cyclic di-GMP phosphodiesterase involved in autoaggregation, flagellum synthesis, motility, biofilm formation, cell invasion, and cytotoxicity. *Infection and immunity* 78(5), pp. 1832–40. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20194589 [Accessed: 15 February 2019].

Lennings, J. *et al.* 2019. The *Burkholderia* type VI secretion system 5: composition, regulation and role in virulence. *Frontiers in Microbiology* 9, p. 3339. Available at: https://www.frontiersin.org/article/10.3389/fmicb.2018.03339/full [Accessed: 17 June 2019].

Letzel, A.-C. *et al.* 2017. Genomic insights into specialized metabolism in the marine actinomycete *Salinispora*. *Environmental Microbiology* 19(9), pp. 3660–3673. Available at: http://doi.wiley.com/10.1111/1462-2920.13867 [Accessed: 23 July 2019].

Levenberg, B. and Linton, S.N. 1966. On the biosynthesis of toxoflavin, an azapteridine antibiotic produced by *Pseudomonas cocovenenans*. *Journal of Biological Chemistry* 241(4), pp. 846–852. Available at: http://www.jbc.org/content/241/4/846.full.pdf [Accessed: 5 April 2017].

Lewis, T. *et al.* 2010. High-throughput whole-genome sequencing to dissect the epidemiology of *Acinetobacter baumannii* isolates from a hospital outbreak. *Journal of Hospital Infection* 75(1), pp. 37–41. Available at: https://www.sciencedirect.com/science/article/pii/S0195670110000289 [Accessed: 26 September 2018].

Li, H. 2013. Circlator: automated circularization of genome assemblies using long sequencing reads. 16(294). Available at: http://arxiv.org/abs/1303.3997 [Accessed: 7 April 2017].

Li, K. *et al.* 2017. *Burkholderia cepacia* lipase immobilized on heterofunctional magnetic nanoparticles and its application in biodiesel synthesis. *Scientific Reports* 7(1), p. 16473. Available at: http://www.nature.com/articles/s41598-017-16626-5 [Accessed: 9 July 2019].

Li, W. *et al.* 2002a. Broad spectrum anti-biotic activity and disease suppression by the potential biocontrol agent *Burkholderia ambifaria* BC-F. *Crop Protection* 21(2), pp. 129–135. Available at: http://www.sciencedirect.com.abc.cardiff.ac.uk/science/article/pii/S0261219401000746 [Accessed: 3 May 2017].

Li, W. *et al.* 2002b. Broad spectrum anti-biotic activity and disease suppression by the potential biocontrol agent *Burkholderia ambifaria* BC-F. *Crop Protection* 21(2), pp. 129–135. Available at: https://www.sciencedirect.com/science/article/pii/S0261219401000746 [Accessed: 14 August 2018].

Lim, Y. *et al.* 1994. The journal of antibiotics cepacidine a, a novel antifungal antibiotic produced by *Pseudomonas cepacia* II. Physico-chemical properties and structure elucidation. *The Journal of Antibiotics* 47(12), pp. 1406–1416. Available at: https://www.jstage.jst.go.jp/article/antibiotics1968/47/12/47_12_1406/_pdf/-char/en [Accessed: 1 May 2018].

Limmathurotsakul, D. *et al.* 2016. Predicted global distribution of *Burkholderia pseudomallei* and burden of melioidosis. *Nature Microbiology* 1(1), p. 15008. Available at: http://www.nature.com/articles/nmicrobiol20158 [Accessed: 15 July 2019].

LiPuma, J.J. 2010. The changing microbial epidemiology in cystic fibrosis. *Clinical Microbiology Reviews* 23(2), pp. 299–323. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20375354 [Accessed: 2 November 2018].

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. Available at: http://link.springer.com/10.1007/s10482-017-0842-6 [Accessed: 4 July 2019].

Lu, S.E. *et al.* 2009. Occidiofungin, a unique antifungal glycopeptide produced by a strain of *Burkholderia contaminans*. *Biochemistry* 48(35), pp. 8312–8321. Available at: http://pubs.acs.org/doi/abs/10.1021/bi900814c [Accessed: 5 April 2017].

Magoč, T. and Salzberg, S.L. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics (Oxford, England)* 27(21), pp. 2957–63. Available at: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btr507 [Accessed: 6 April 2017].

Mahenthiralingam, E. *et al.* 2000. DNA-Based diagnostic approaches for identification of *Burkholderia cepacia complex, Burkholderia vietnamiensis, Burkholderia multivorans, Burkholderia stabilis*, and *Burkholderia cepacia* genomovars I and III. *Journal of clinical microbiology* 38(9), pp. 3165–73. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10970351 [Accessed: 1 August 2019].

Mahenthiralingam, E. *et al.* 2001. Infection with *Burkholderia cepacia* Complex Genomovars in Patients with Cystic Fibrosis: Virulent Transmissible Strains of Genomovar III Can Replace *Burkholderia multivorans*. *Clinical Infectious Diseases* 33(9), pp. 1469–1475. Available at: https://academic.oup.com/cid/article-lookup/doi/10.1086/322684 [Accessed: 17 July 2019].

Mahenthiralingam, E. *et al.* 2005. The multifarious, multireplicon *Burkholderia cepacia* complex. *Nature Reviews Microbiology* 3(2), pp. 144–156. Available at: http://www.nature.com/articles/nrmicro1085 [Accessed: 8 July 2019].

Mahenthiralingam, E. *et al.* 2008. *Burkholderia cepacia* complex bacteria: opportunistic pathogens with important natural biology. *Journal of Applied Microbiology* 104(6), pp. 1539–1551. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18217926 [Accessed: 2 July 2019].

Mahenthiralingam, E. *et al.* 2011. Enacyloxins are products of an unusual hybrid modular polyketide synthase encoded by a cryptic *Burkholderia ambifaria* genomic island. *Chemistry and Biology* 18(5), pp. 665–677. Available at: http://www.sciencedirect.com/science/article/pii/S1074552111001190 [Accessed: 5 April 2017].

Mansfield, J. *et al.* 2012. Top 10 plant pathogenic bacteria in molecular plant pathology. *Molecular Plant Pathology* 13(6), pp. 614–629. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22672649 [Accessed: 7 April 2017].

Mao, D. *et al.* 2017. Discovery of scmR as a global regulator of secondary metabolism and virulence in *Burkholderia thailandensis* E264. *Proceedings of the National Academy of Sciences of the United States of America* 114(14), pp. E2920–E2928. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28320949 [Accessed: 21 May 2018].

Mao, W. *et al.* 1997. Seed treatment with a fungal or a bacterial antagonist for reducing corn damping-off caused by species of *Pythium* and *Fusarium*. *Plant Disease* 81(5), pp. 450–454. Available at: http://apsjournals.apsnet.org/doi/10.1094/PDIS.1997.81.5.450 [Accessed: 9 April 2018].

Mao, W. *et al.* 1998. Biocontrol of selected soilborne diseases of tomato and pepper plants. *Crop Protection* 17(6), pp. 535–542. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0261219498000556 [Accessed: 3

May 2017].

Mark, G.L. *et al.* 2006. Molecular-based strategies to exploit *Pseudomonas* biocontrol strains for environmental biotechnology applications. *FEMS Microbiology Ecology* 56(2), pp. 167–177. Available at: https://academic.oup.com/femsec/article-lookup/doi/10.1111/j.1574-6941.2006.00056.x [Accessed: 8 July 2019].

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17(1), p. 10. Available at: http://journal.embnet.org/index.php/embnetjournal/article/view/200 [Accessed: 7 April 2017].

Martina, P. *et al.* 2018. *Burkholderia puraquae* sp. nov., a novel species of the *Burkholderia cepacia* complex isolated from hospital settings and agricultural soils. *International Journal of Systematic and Evolutionary Microbiology* 68(1), pp. 14–20. Available at: http://www.microbiologyresearch.org/content/journal/ijsem/10.1099/ijsem.0.002293 [Accessed: 3 July 2019].

Massalha, H. *et al.* 2017. Small molecules below-ground: the role of specialized metabolites in the rhizosphere. *The Plant Journal* 90(4), pp. 788–807. Available at: http://doi.wiley.com/10.1111/tpj.13543 [Accessed: 1 July 2019].

Masschelein, J. *et al.* 2017. Antibiotics from Gram-negative bacteria: a comprehensive overview and selected biosynthetic highlights. *Natural Product Reports* 34(7), pp. 712–783. Available at: http://xlink.rsc.org/?DOI=C7NP00010C [Accessed: 5 April 2018].

Mavrodi, D. V *et al.* 1998. A seven-gene locus for synthesis of phenazine-1-carboxylic acid by *Pseudomonas fluorescens* 2-79. *Journal of bacteriology* 180(9), pp. 2541–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9573209 [Accessed: 17 May 2018].

McLoughlin, T.J. *et al.* 1992. *Pseudomonas cepacia* suppression of sunflower wilt fungus and role of antifungal compounds in controlling the disease. *Applied and Environmental Microbiology* 58(5), pp. 1760–1763. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1377900 [Accessed: 9 August 2018].

Meyers, E. *et al.* 1987. Xylocandin: a new complex of antifungal peptides. I. Taxonomy, isolation and biological activity. *The Journal of antibiotics* 40(11), pp. 1515–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3693121 [Accessed: 14 May 2018].

Miles, A.A. *et al.* 1938. The estimation of the bactericidal power of the blood. *The Journal of hygiene* 38(6), pp. 732–49. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20475467 [Accessed: 24 July 2018].

Mitter, B. *et al.* 2013. Comparative genome analysis of *Burkholderia phytofirmans* PsJN reveals a wide spectrum of endophytic lifestyles based on interaction strategies with host plants. *Frontiers in Plant Science* 4, p. 120. Available at: http://journal.frontiersin.org/article/10.3389/fpls.2013.00120/abstract [Accessed: 13 June 2019].

Mnif, I. and Ghribi, D. 2015. Potential of bacterial derived biopesticides in pest management. *Crop Protection* 77, pp. 52–64. Available at: https://www.sciencedirect.com/science/article/pii/S0261219415300727#bib26 [Accessed: 8 July 2019].

Moon, S.-S. *et al.* 1996. Plant growth promoting and fungicidal 4-quinolinones from *Pseudomonas cepacia*. *Phytochemistry* 42(2), pp. 365–368. Available at: https://www.sciencedirect.com/science/article/pii/0031942295008977 [Accessed: 17 May 2018].

Morganti, S. *et al.* 2019. Complexity of genome sequencing and reporting: Next generation sequencing (NGS)

technologies and implementation of precision medicine in real life. *Critical Reviews in Oncology/Hematology* 133, pp. 171–182. Available at: https://www.sciencedirect.com/science/article/pii/S1040842818304220?via%3Dihub [Accessed: 18 July 2019].

Morgulis, A. *et al.* 2008. BLAST+: architecture and applications. *Bioinformatics* 24(16), pp. 1757–1764. Available at: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btn322 [Accessed: 7 April 2017].

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. Available at: http://www.nature.com/articles/s41564-019-0383-z [Accessed: 1 July 2019].

Naughton, L.M. *et al.* 2017. Identification of secondary metabolite gene clusters in the *Pseudovibrio* genus reveals encouraging biosynthetic potential toward the production of novel bioactive compounds. *Frontiers in Microbiology* 8, p. 1494. Available at: http://journal.frontiersin.org/article/10.3389/fmicb.2017.01494/full [Accessed: 23 August 2018].

O'Sullivan, L.A. and Mahenthiralingam, E. 2005. Biotechnological potential within the genus *Burkholderia*. *Letters in Applied Microbiology* 41(1), pp. 8–11. Available at: http://doi.wiley.com/10.1111/j.1472-765X.2005.01758.x [Accessed: 8 July 2019].

Ondov, B.D. *et al.* 2016. Mash: fast genome and metagenome distance estimation using MinHash. *Genome biology* 17(1), p. 132. Available at: http://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-0997-x [Accessed: 7 April 2017].

Page, A.J. *et al.* 2015. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31(22), pp. 3691–3693. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26198102 [Accessed: 7 April 2017].

Palleroni, N.J. and Holmes, B. 1981. Pseudomonas cepacia sp. nov., nom. rev. *International Journal of Systematic Bacteriology* 31(4), pp. 479–481. Available at: http://ijs.microbiologyresearch.org/content/journal/ijsem/10.1099/00207713-31-4-479 [Accessed: 2 July 2019].

Parke, J.L. *et al.* 1991. Biological control of *Pythium* damping-off and *Aphanomyces* root rot of peas by application of *Pseudomonas cepacia* or *P. fluorescens* to seed. *Plant Disease* 75(10), p. 987. Available at: http://www.apsnet.org/publications/PlantDisease/BackIssues/Documents/1991Abstracts/PD_75_987.htm [Accessed: 3 May 2017].

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. Available at: http://www.annualreviews.org/doi/10.1146/annurev.phyto.39.1.225 [Accessed: 16 June 2017].

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. Available at: http://joi.jlc.jst.go.jp/JST.Journalarchive/antibiotics1968/37.431?from=CrossRef [Accessed: 5 April 2017].

Partida-Martinez, L.P. and Hertweck, C. 2007. A gene cluster encoding rhizoxin biosynthesis in "*Burkholderia rhizoxina*", the bacterial endosymbiont of the fungus *Rhizopus microsporus*. *ChemBioChem* 8(1), pp. 41–45. Available at: http://doi.wiley.com/10.1002/cbic.200600393 [Accessed: 6 July 2019].

Patel, M. *et al.* 1988. Sch 31828, a novel antibiotic from a *Microbispora* sp. Taxonomy, fermentation, isolation and biological properties. *The Journal of Antibiotics* 41(6), pp. 794–797. Available at:

http://joi.jlc.jst.go.jp/JST.Journalarchive/antibiotics1968/41.794?from=CrossRef [Accessed: 21 August 2018].

Paungfoo-Lonhienne, C. *et al.* 2014. A new species of *Burkholderia* isolated from sugarcane roots promotes plant growth. *Microbial biotechnology* 7(2), pp. 142–54. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24350979 [Accessed: 12 July 2019].

Paungfoo-Lonhienne, C. *et al.* 2016. Crosstalk between sugarcane and a plant-growth promoting *Burkholderia* species. *Scientific Reports* 6(1), p. 37389. Available at: http://www.nature.com/articles/srep37389 [Accessed: 14 July 2019].

Peeters, C. *et al.* 2017. Comparative genomics of *Burkholderia multivorans*, a ubiquitous pathogen with a highly conserved genomic structure. *PloS one* 12(4), p. e0176191. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28430818 [Accessed: 26 September 2018].

Pirone, L. *et al.* 2005. Detection of cultured and uncultured *Burkholderia cepacia* complex bacteria naturally occurring in the maize rhizosphere. *Environmental Microbiology* 7(11), pp. 1734–1742. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16232288 [Accessed: 9 April 2018].

Podnecky, N.L. *et al.* 2015. Efflux pump-mediated drug resistance in *Burkholderia*. *Frontiers in microbiology* 6, p. 305. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25926825 [Accessed: 13 February 2019].

Poupin, M.J. *et al.* 2013. Effects of the Plant Growth-Promoting Bacterium *Burkholderia phytofirmans* PsJN throughout the Life Cycle of *Arabidopsis thaliana*. Vinatzer, B. A. ed. *PLoS ONE* 8(7), p. e69435. Available at: http://dx.plos.org/10.1371/journal.pone.0069435 [Accessed: 7 July 2019].

Price, M.N. *et al.* 2010. FastTree 2 - Approximately maximum-likelihood trees for large alignments. Poon, A. F. Y. ed. *PLoS ONE* 5(3), p. e9490. Available at: http://dx.plos.org/10.1371/journal.pone.0009490 [Accessed: 7 April 2017].

Pritchard, L. *et al.* 2016. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal. Methods* 8(1), pp. 12–24. Available at: http://xlink.rsc.org/?DOI=C5AY02550H [Accessed: 10 April 2017].

Quick, J. *et al.* 2014. Seeking the source of *Pseudomonas aeruginosa* infections in a recently opened hospital: an observational study using whole-genome sequencing. *BMJ open* 4(11), p. e006278. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25371418 [Accessed: 18 July 2019].

Quick, J. *et al.* 2016. Real-time, portable genome sequencing for Ebola surveillance. *Nature* 530(7589), pp. 228–232. Available at: http://www.nature.com/articles/nature16996 [Accessed: 18 July 2019].

Ramage, B. *et al.* 2017. Comprehensive arrayed transposon mutant library of *Klebsiella pneumoniae* outbreak strain KPNIH1. *Journal of bacteriology* 199(20), pp. e00352-17. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28760848 [Accessed: 31 July 2019].

Ramette, A. *et al.* 2005. Species abundance and diversity of *Burkholderia cepacia* complex in the environment. *Applied and environmental microbiology* 71(3), pp. 1193–201. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15746318 [Accessed: 12 July 2019].

Ramette, A. and Tiedje, J.M. 2007. Multiscale responses of microbial life to spatial distance and environmental heterogeneity in a patchy ecosystem. *Proceedings of the National Academy of Sciences* 104(8), pp. 2761–2766. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17296935 [Accessed: 9 April 2018].

Ramirez, M.S. and Tolmasky, M.E. 2010. Aminoglycoside modifying enzymes. *Drug Resistance Updates* 13(6), pp. 151–171. Available at: https://www.sciencedirect.com/science/article/pii/S1368764610000385 [Accessed: 13 February 2019].

Randall, L.B. *et al.* 2015. Membrane-bound PenA β-lactamase of *Burkholderia pseudomallei*. *Antimicrobial agents and chemotherapy* 60(3), pp. 1509–14. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26711764 [Accessed: 13 February 2019].

Richardson, J. *et al.* 2002. Diversity of *Burkholderia* isolates from woodland rhizosphere environments. *Journal of Applied Microbiology* 93(4), pp. 616–630. Available at: http://doi.wiley.com/10.1046/j.1365-2672.2002.01722.x [Accessed: 14 July 2019].

Richter, M. and Rossello-Mora, R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences* 106(45), pp. 19126–19131. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19855009 [Accessed: 10 April 2017].

Roberts, D.P. *et al.* 2005. Biocontrol agents applied individually and in combination for suppression of soilborne diseases of cucumber. *Crop Protection* 24(2), pp. 141–155. Available at: https://www.sciencedirect.com/science/article/pii/S0261219404001887 [Accessed: 14 August 2018].

Roberts, D.P. *et al.* 2014. Control of damping-off of organic and conventional cucumber with extracts from a plant-associated bacterium rivals a seed treatment pesticide. *Crop Protection* 65, pp. 86–94. Available at: https://www.sciencedirect.com/science/article/pii/S0261219414002294#bib36 [Accessed: 14 August 2018].

Ross, C. *et al.* 2014. The molecular basis of conjugated polyyne biosynthesis in phytopathogenic bacteria. *Angewandte Chemie - International Edition* 53(30), pp. 7794–7798. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24898429 [Accessed: 16 June 2017].

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. 2972–80. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23587949 [Accessed: 30 July 2019].

Rutledge, P.J. and Challis, G.L. 2015. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nature Reviews Microbiology 2015 13:8* 13(8), p. 509. Available at: http://www.nature.com/articles/nrmicro3496 [Accessed: 1 August 2019].

Ryan, R.P. *et al.* 2011. Pathogenomics of *Xanthomonas*: understanding bacterium–plant interactions. *Nature Reviews Microbiology* 9(5), pp. 344–355. Available at: http://www.nature.com/articles/nrmicro2558 [Accessed: 7 July 2019].

Sánchez, D.A. *et al.* 2018. *Burkholderia cepacia* lipase: A versatile catalyst in synthesis reactions. *Biotechnology and Bioengineering* 115(1), pp. 6–24. Available at: http://doi.wiley.com/10.1002/bit.26458 [Accessed: 9 July 2019].

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 env. *Frontiers in Genetics* 5, p. 429. Available at: http://journal.frontiersin.org/article/10.3389/fgene.2014.00429/abstract [Accessed: 3 July 2019].

Schadt, E.E. *et al.* 2010. A window into third-generation sequencing. *Human Molecular Genetics* 19(R2), pp. R227–R240. Available at: https://academic.oup.com/hmg/article-lookup/doi/10.1093/hmg/ddq416 [Accessed: 18

July 2019].

Schellenberg, B. *et al.* 2007. Identification of genes involved in the biosynthesis of the cytotoxic compound glidobactin from a soil bacterium. *Environmental Microbiology* 9(7), pp. 1640–1650. Available at: http://doi.wiley.com/10.1111/j.1462-2920.2007.01278.x [Accessed: 17 May 2018].

Scherlach, K. *et al.* 2006. Antimitotic rhizoxin derivatives from a cultured bacterial endosymbiont of the rice pathogenic fungus *Rhizopus microsporus*. Available at: https://pubs.acs.org/doi/10.1021/ja062953o [Accessed: 9 July 2019].

Schisler, D.A. *et al.* 1991. Enhancement of disease caused by *Colletotrichum truncatum* in *Sesbania exaltata* by coinoculating with epiphytic bacteria. *Biological Control* 1(4), pp. 261–268. Available at: https://www.sciencedirect.com/science/article/pii/104996449190076C [Accessed: 9 April 2018].

Schmid, N. *et al.* 2012. The AHL- and BDSF-dependent quorum sensing systems control specific and overlapping sets of genes in Burkholderia cenocepacia H111. van Schaik, W. ed. *PLoS ONE* 7(11), p. e49966. Available at: https://dx.plos.org/10.1371/journal.pone.0049966 [Accessed: 17 September 2019].

Schmidt, S. *et al.* 2009a. Production of the antifungal compound pyrrolnitrin is quorum sensing-regulated in members of the *Burkholderia cepacia* complex. *Environmental Microbiology* 11(6), pp. 1422–1437. Available at: http://doi.wiley.com/10.1111/j.1462-2920.2009.01870.x [Accessed: 10 August 2018].

Schmidt, S. *et al.* 2009b. Production of the antifungal compound pyrrolnitrin is quorum sensing-regulated in members of the *Burkholderia cepacia* complex. *Environmental Microbiology* 11(6), pp. 1422–1437. Available at: http://doi.wiley.com/10.1111/j.1462-2920.2009.01870.x [Accessed: 5 April 2017].

Schwarz, S. *et al.* 2014. VgrG-5 Is a *Burkholderia* type VI secretion system-exported protein required for multinucleated giant cell formation and virulence. *Infection and Immunity* 82(4), pp. 1445–1452. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24452686 [Accessed: 17 June 2019].

Seemann, T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30(14), pp. 2068–2069. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24642063 [Accessed: 7 April 2017].

Sen, D. *et al.* 2013. Inferring the evolutionary history of IncP-1 plasmids despite incongruence among backbone gene trees. *Molecular Biology and Evolution* 30(1), pp. 154–166. Available at: https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/mss210 [Accessed: 13 June 2019].

Seo, Y.-S. *et al.* 2015. Comparative genome analysis of rice-pathogenic *Burkholderia* provides insight into capacity to adapt to different environments and hosts. *BMC Genomics* 16(1), p. 349. Available at: http://www.biomedcentral.com/1471-2164/16/349 [Accessed: 26 September 2018].

Seyedsayamdost, M.R. *et al.* 2010. Quorum-sensing-regulated bactobolin production by *Burkholderia thailandensis* E264. *Organic Letters* 12(4), pp. 716–719. Available at: http://pubs.acs.org/doi/abs/10.1021/ol902751x [Accessed: 5 April 2017].

Shalom, G. *et al.* 2007. In vivo expression technology identifies a type VI secretion system locus in *Burkholderia pseudomallei* that is induced upon invasion of macrophages. *Microbiology* 153(8), pp. 2689–2699. Available at: http://mic.microbiologyresearch.org/content/journal/micro/10.1099/mic.0.2007/006585-0 [Accessed: 20 February 2019].

Shannon, P. *et al.* 2003. Cytoscape: A software environment for integrated models of biomolecular interaction

networks. *Genome Research* 13(11), pp. 2498–2504. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14597658 [Accessed: 7 April 2017].

Shi Shun, A.L.K. and Tykwinski, R.R. 2006. Synthesis of Naturally Occurring Polyynes. *Angewandte Chemie International Edition* 45(7), pp. 1034–1057. Available at: http://doi.wiley.com/10.1002/anie.200502071 [Accessed: 27 July 2019].

Shoji, J. *et al.* 1990. Isolation of cepafungins I, II and III from *Pseudomonas* species. *The Journal of Antibiotics* 43(7), pp. 783–787. Available at: http://joi.jlc.jst.go.jp/JST.Journalarchive/antibiotics1968/43.783?from=CrossRef [Accessed: 17 May 2018].

Simonetti, E. *et al.* 2018. A novel *Burkholderia ambifaria* strain able to degrade the mycotoxin fusaric acid and to inhibit *Fusarium* spp. growth. *Microbiological Research* 206, pp. 50–59. Available at: https://www.sciencedirect.com/science/article/pii/S0944501317302082 [Accessed: 14 August 2018].

Smith, D.L. *et al.* 1993. Epidemic of *Pseudomonas cepacia* in an adult cystic fibrosis unit: evidence of person-to-person transmission. *Journal of clinical microbiology* 31(11), pp. 3017–22. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7505294 [Accessed: 29 June 2019].

Sokol, P.A. 1986. Production and utilization of pyochelin by clinical isolates of Pseudomonas cepacia. *Journal of clinical microbiology* 23(3), pp. 560–2. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2937804 [Accessed: 18 May 2018].

Sokol, P.A. *et al.* 1999. Role of Ornibactin Biosynthesis in the Virulence ofBurkholderia cepacia: Characterization of pvdA, the Gene Encoding l-OrnithineN5-Oxygenase. *Infection and Immunity* 67(9), pp. 4443–4455. Available at: https://iai.asm.org/content/67/9/4443.long [Accessed: 6 July 2019].

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), p. jacs.7b03382. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28528545 [Accessed: 16 June 2017].

Spiewak, H.L. *et al.* 2019. *Burkholderia cenocepacia* utilizes a type VI secretion system for bacterial competition. *MicrobiologyOpen* , p. e774. Available at: http://doi.wiley.com/10.1002/mbo3.774 [Accessed: 20 February 2019].

Spitzer, M. *et al.* 2014. BoxPlotR: a web tool for generation of box plots. *Nature methods* 11(2), pp. 121–2. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24481215 [Accessed: 7 April 2017].

Spring-Pearson, S.M. *et al.* 2015. Pangenome Analysis of *Burkholderia pseudomallei*: Genome Evolution Preserves Gene Order despite High Recombination Rates. Planet, P. J. ed. *PLOS ONE* 10(10), p. e0140274. Available at: http://dx.plos.org/10.1371/journal.pone.0140274 [Accessed: 26 September 2018].

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics (Oxford, England)* 30(9), pp. 1312–3. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24451623 [Accessed: 3 April 2018].

Stephan, H. *et al.* 1993. Ornibactins--a new family of siderophores from *Pseudomonas*. *Biometals : an international journal on the role of metal ions in biology, biochemistry, and medicine* 6(2), pp. 93–100. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7689374 [Accessed: 18 May 2018].

Studholme, D.J. *et al.* 2013. Investigating the beneficial traits of Trichoderma hamatum GD12 for sustainable agriculture-insights from genomics. *Frontiers in plant science* 4, p. 258. Available at:

http://www.ncbi.nlm.nih.gov/pubmed/23908658 [Accessed: 7 July 2019].

Subramoni, S. *et al.* 2015. A bioinformatic survey of distribution, conservation, and probable functions of LuxR solo regulators in bacteria. *Frontiers in cellular and infection microbiology* 5, p. 16. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25759807 [Accessed: 7 September 2017].

Sullivan, M.J. *et al.* 2011. Easyfig: a genome comparison visualizer. *Bioinformatics (Oxford, England)* 27(7), pp. 1009–10. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21278367 [Accessed: 11 May 2018].

Sun, L. *et al.* 1995. *The emergence of a highly transmissible lineage of cbr Pseudomonas (Burkholderia) cepacia causing CF centre epidemics in North America and Britain*. Available at: http://www.nature.com/naturemedicine [Accessed: 30 June 2019].

Suzuki, F. *et al.* 2004. Molecular characterization of the tox operon involved in toxoflavin biosynthesis of *Burkholderia glumae*. *Journal of General Plant Pathology* 70(2), pp. 97–107. Available at: http://link.springer.com/10.1007/s10327-003-0096-1 [Accessed: 14 May 2018].

Tange, O. 2018. GNU Parallel 2018. Available at: https://zenodo.org/record/1146014#.WzyseVVKiUk [Accessed: 4 July 2018].

Tawfik, K.A. *et al.* 2010. Burkholdines 1097 and 1229, potent antifungal peptides from *Burkholderia ambifaria* 2.2N. *Organic Letters* 12(4), pp. 664–666. Available at: http://pubs.acs.org/doi/abs/10.1021/ol9029269 [Accessed: 5 April 2017].

Thomassen, M.J. *et al.* 1985. *Pseudomonas cepacia* colonization among patients with cystic fibrosis. A new opportunist. *The American review of respiratory disease* 131(5), pp. 791–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3923882 [Accessed: 29 June 2019].

Thomson, E.L.S. and Dennis, J.J. 2012. A *Burkholderia cepacia* complex non-ribosomal peptide-synthesized toxin is hemolytic and required for full virulence. *Virulence* 3(3), pp. 286–98. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22546908 [Accessed: 16 May 2018].

Tietz, J.I. *et al.* 2017. A new genome-mining tool redefines the lasso peptide biosynthetic landscape. *Nature Chemical Biology* 13(5), pp. 470–478. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28244986 [Accessed: 23 August 2018].

Toda, T. *et al.* 2015. Widespread occurrence of *Pythium arrhenomanes* pathogenic to rice seedlings around Japanese rice fields. *Plant Disease* 99(12), pp. 1823–1831. Available at: http://apsjournals.apsnet.org/doi/10.1094/PDIS-01-15-0124-RE [Accessed: 5 April 2018].

Van Tran, V. *et al.* 1996. *Burkholderia vietnamiensis*, a new nitrogen-fixing species associated with rice roots, isolated from an acid sulphate soil in Vietnam: plant-growth-promoting effects on rice. In: *Biological Nitrogen Fixation Associated with Rice Production*. Dordrecht: Springer Netherlands, pp. 181–190. Available at: http://link.springer.com/10.1007/978-94-015-8670-2_20 [Accessed: 12 July 2019].

Tsuge, T. *et al.* 2013. Host-selective toxins produced by the plant pathogenic fungus *Alternaria alternata*. *FEMS Microbiology Reviews* 37(1), pp. 44–66. Available at: https://academic.oup.com/femsre/article-lookup/doi/10.1111/j.1574-6976.2012.00350.x [Accessed: 29 October 2018].

US Environmental Protection Agency 1999a. *FIFRA Scientific Advisory Panel Meeting*. Available at: https://archive.epa.gov/scipoly/sap/meetings/web/pdf/finlrpt1.pdf [Accessed: 30 June 2019].

US Environmental Protection Agency 1999b. *Risk Assessment of Burkholderia cepaia Based Biopesticides, and Other Bacteria Related to Opportunistic Human Pathogens*. Available at: https://archive.epa.gov/scipoly/sap/meetings/web/pdf/bc_bkgrnd.pdf [Accessed: 30 June 2019].

US Environmental Protection Agency 2004. *Allethrin, Bendiocarb, Burkholderia cepacia, Fenridazon potassium, and Molinate; Tolerance Actions*. Available at: www.epa.gov/edocket. [Accessed: 30 June 2019].

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 Bacteriology* 47(4), pp. 1188–1200. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9336927 [Accessed: 2 July 2019].

Vandamme, P. *et al.* 2000. Identification and population structure of Burkholderia stabilis sp. nov. (formerly Burkholderia cepacia genomovar IV). *Journal of clinical microbiology* 38(3), pp. 1042–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10698993 [Accessed: 2 July 2019].

Vandamme, P. *et al.* 2002. *Burkholderia anthina* sp. nov. and *Burkholderia pyrrocinia* , two additional *Burkholderia cepacia* complex bacteria, may confound results of new molecular diagnostic tools. *FEMS Immunology & Medical Microbiology* 33(2), pp. 143–149. Available at: https://academic.oup.com/femspd/article-lookup/doi/10.1111/j.1574-695X.2002.tb00584.x [Accessed: 2 July 2019].

Vandamme, P. *et al.* 2003. *Burkholderia cenocepacia* sp. nov.—a new twist to an old story. *Research in Microbiology* 154(2), pp. 91–96. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12648723 [Accessed: 2 July 2019].

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. Available at: http://ijs.microbiologyresearch.org/content/journal/ijsem/10.1099/ijs.0.65142-0 [Accessed: 6 July 2019].

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. Available at: http://journal.frontiersin.org/article/10.3389/fmicb.2017.01679/full [Accessed: 26 September 2018].

Vandamme, P. and Peeters, C. 2014. Time to revisit polyphasic taxonomy. *Antonie van Leeuwenhoek* 106(1), pp. 57–65. Available at: http://link.springer.com/10.1007/s10482-014-0148-x [Accessed: 4 July 2019].

Vermis, K. *et al.* 2002. Evaluation of species-specific *recA*-based PCR tests for genomovar level identification within the *Burkholderia cepacia* complex. *Journal of Medical Microbiology* 51(11), pp. 937–940. Available at: http://jmm.microbiologyresearch.org/content/journal/jmm/10.1099/0022-1317-51-11-937 [Accessed: 3 July 2019].

Vermis, K. *et al.* 2004. Proposal to accommodate *Burkholderia cepacia* genomovar VI as *Burkholderia dolosa* sp. nov. *INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY* 54(3), pp. 689–691. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15143009 [Accessed: 2 July 2019].

Vesth, T.C. *et al.* 2016. FunGeneClusterS: Predicting fungal gene clusters from genome and transcriptome data. *Synthetic and Systems Biotechnology* 1(2), pp. 122–129. Available at: https://www.sciencedirect.com/science/article/pii/S2405805X1530017X [Accessed: 23 August 2018].

Vial, L. *et al.* 2008. *Burkholderia pseudomallei*, *B. thailandensis*, and *B. ambifaria* produce 4-hydroxy-2-alkylquinoline analogues with a methyl group at the 3 position that is required for quorum-sensing regulation.

Journal of bacteriology 190(15), pp. 5339–52. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18539738 [Accessed: 17 August 2018].

Vidal-Quist, J.C. *et al.* 2014. *Arabidopsis thaliana* and *Pisum sativum* models demonstrate that root colonization is an intrinsic trait of *Burkholderia cepacia* complex bacteria. *Microbiology* 160(Pt_2), pp. 373–384. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24327425 [Accessed: 24 May 2018].

Vijayan, M. *et al.* 2011. PKSIIIexplorer: TSVM approach for predicting Type III polyketide synthase proteins. *Bioinformation* 6(3), pp. 125–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21584189 [Accessed: 23 August 2018].

Walker, B.J. *et al.* 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS one* 9(11), p. e112963. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25409509 [Accessed: 6 April 2017].

Wang, X. *et al.* 2018. Discovery of recombinases enables genome mining of cryptic biosynthetic gene clusters in Burkholderiales species. *Proceedings of the National Academy of Sciences* 115(18), pp. E4255–E4263. Available at: https://www.pnas.org/content/115/18/E4255 [Accessed: 24 July 2019].

Weber, T. *et al.* 2015. antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic acids research* 43(W1), pp. W237-43. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25948579 [Accessed: 7 April 2017].

Weingarten, R.A. *et al.* 2018. Genomic Analysis of Hospital Plumbing Reveals Diverse Reservoir of Bacterial Plasmids Conferring Carbapenem Resistance. *mBio* 9(1), pp. e02011-17. Available at: http://www.ncbi.nlm.nih.gov/pubmed/29437920 [Accessed: 14 June 2019].

Wong, V.K. *et al.* 2016. An extended genotyping framework for *Salmonella enterica* serovar *Typhi*, the cause of human typhoid. *Nature Communications* 7, p. 12827. Available at: http://www.nature.com/doifinder/10.1038/ncomms12827 [Accessed: 26 September 2018].

Wood, D.E. and Salzberg, S.L. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology* 15(R46). Available at: http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-12-385 [Accessed: 10 April 2017].

Wyres, K.L. *et al.* 2019. Distinct evolutionary dynamics of horizontal gene transfer in drug resistant and virulent clones of *Klebsiella pneumoniae*. Hughes, D. ed. *PLOS Genetics* 15(4), p. e1008114. Available at: http://dx.plos.org/10.1371/journal.pgen.1008114 [Accessed: 23 July 2019].

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. 1251–75. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1283774 [Accessed: 19 December 2017].

Yadav, G. *et al.* 2003. SEARCHPKS: a program for detection and analysis of polyketide synthase domains. *Nucleic Acids Research* 31(13), pp. 3654–3658. Available at: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkg607 [Accessed: 23 August 2018].

Yamaguchi, M. *et al.* 1995. Chemistry and Antimicrobial Activity of Caryoynencin Analogs. *Journal of Medicinal Chemistry* 38(26), pp. 5015–5022. Available at: http://pubs.acs.org/doi/abs/10.1021/jm00026a008 [Accessed: 27 July 2019].

Yang, C. *et al.* 2017. *Burkholderia cenocepacia* integrates cis-2-dodecenoic acid and cyclic dimeric guanosine monophosphate signals to control virulence. *Proceedings of the National Academy of Sciences* 114(49), pp. 13006–13011. Available at: https://www.pnas.org/content/114/49/13006 [Accessed: 17 September 2019].

Yoshii, A. *et al.* 2012. The novel kasugamycin 2'-N-acetyltransferase gene aac(2')-IIa, carried by the IncP island, confers kasugamycin resistance to rice-pathogenic bacteria. *Applied and environmental microbiology* 78(16), pp. 5555–64. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22660700 [Accessed: 14 June 2019].

Zheng, J. *et al.* 2017. Comparative genomics of *Bacillus thuringiensis* reveals a path to specialized exploitation of multiple invertebrate hosts. *mBio* 8(4), pp. e00822-17. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28790205 [Accessed: 9 September 2019].

Ziemert, N. *et al.* 2016. The evolution of genome mining in microbes – a review. *Natural Product Reports* 33(8), pp. 988–1005. Available at: http://xlink.rsc.org/?DOI=C6NP00025H [Accessed: 17 May 2018].

Zierep, P.F. *et al.* 2017. SeMPI: a genome-based secondary metabolite prediction and identification web server. *Nucleic Acids Research* 45(W1), pp. W64–W71. Available at: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkx289 [Accessed: 23 August 2018].