**Title:** Genome mining identifies cepacin as a key plant-protective metabolite of the biopesticidal bacterium *Burkholderia* *ambifaria*

**Short title:** *Burkholderia ambifaria*: Genomics and biology of a biopesticide

**Authors:**

Alex J. Mullins1, James. A. H. Murray1, Matthew J. Bull1, Matthew Jenner2, Cerith Jones1,6, Gordon Webster1, Angharad E. Green3, Daniel R. Neill3, Thomas R. Connor1, Julian Parkhill4, Gregory L. Challis2,5 and Eshwar Mahenthiralingam1

**Author affiliations and footnotes**

1School of Biosciences, Cardiff University, Cardiff CF10 3AX, United Kingdom

2Department of Chemistry and Warwick Integrative Synthetic Biology Centre, University of Warwick, Coventry CV4 7AL, United Kingdom

3Institute of Infection and Global Health, University of Liverpool, Liverpool L69 7BE, United Kingdom

4Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, United Kingdom

5Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Monash University, Clayton, VIC 3800, Australia

6Current address: Faculty of Computing, Engineering and Science, University of South Wales, Pontypridd, CF37 1DL, United Kingdom

**Corresponding author**

Submission and handling: Prof. Eshwar Mahenthiralingam, Microbiomes, Microbes and Informatics Group, Organisms and Environment Division, School of Biosciences, Cardiff University, Sir Martin Evans Building, Museum Avenue, Cardiff CF10 3AX, United Kingdom; Tel. +44 (0)29 20875875; Fax. +44 (0)29 20874305;

Email: MahenthiralingamE@cardiff.ac.uk (ORCID: 0000-0001-9014-3790)

**Joint correspondence:** Alex J. Mullins, Microbiomes, Microbes and Informatics Group, Organisms and Environment Division, School of Biosciences, Cardiff University, Sir Martin Evans Building, Museum Avenue, Cardiff CF10 3AX, United Kingdom; Tel. +44 (0)29 20874648

Email: MullinsA@cardiff.ac.uk (ORCID: 0000-0001-5804-9008)

**Key words:** *Burkholderia*, biopesticides, phylogenomics, genome mining, specialized metabolites

**Abstract**

**Beneficial microorganisms are widely used in agriculture for control of plant pathogens but a lack of efficacy and safety information has limited the exploitation of multiple promising biopesticides. We applied phylogeny-led genome mining, metabolite analyses and biological control assays to define the efficacy of *Burkholderia ambifaria*, a naturally beneficial bacterium with proven biocontrol properties, but a potential pathogenic risk*.* A panel of 64 *B. ambifaria* strains demonstrated significant antimicrobial activity against priority plant pathogens. Genome sequencing, specialized metabolite gene cluster mining and metabolite analysis revealed an armoury of known and unknown biosynthetic pathways within *B. ambifaria*. The gene cluster responsible for the production of the metabolite, cepacin, was identified and directly shown to mediate protection of germinating crops against *Pythium* damping-off disease. *B. ambifaria* maintained this biopesticidal protection and fitness in soil after deletion of its third replicon, a non-essential plasmid associated with virulence in *B. cepacia* complex bacteria. Removal of the third replicon reduced *B. ambifaria* persistence in a murine respiratory infection model. Here we show that by using interdisciplinary phylogenomic, metabolomic and functional approaches, the mode of action of natural biological control agents related to pathogens can be systematically established to facilitate their future exploitation.**

**Main**

Numerous bacterial and fungal species have been recognised for their biological control abilities and plant growth-enhancing properties (Mishra *et al.*, 2015). Pesticides conventionally used in agriculture are under increasing scrutiny regarding their bioaccumulation and toxicity, which includes their fatal impact on pollinator species. Concern over chemical pesticides has reinvigorated research into biological control agents and their secreted bioactive compounds as viable natural alternatives for agriculture. One feature common to most biopesticidal species is their ability to secrete antimicrobial compounds into the environment and inhibit pathogenic microbes from causing crop disease. Bacteria within the genus *Burkholderia* are particularly diverse in their specialized metabolism and have a documented ability to produce a range of potent anti-bacterial, anti-nematodal and anti-fungal compounds (Parke and Gurian-Sherman, 2001; Depoorter *et al.*, 2016). They have demonstrated excellent promise as biological control agents with multiple strains used commercially as biopesticides until 1999. In common with other biological control genera such as *Bacillus*, *Pseudomonas* and *Stenotrophomonas*, certain *Burkholderia* species may also cause human, animal and plant infections. Therefore, in 1999 the US Environmental Protection Agency (EPA) placed a moratorium on new registrations of *Burkholderia* biopesticides unless such agents were defined as safe in terms of their risk of opportunistic infection (Parke and Gurian-Sherman, 2001).

Multiple species within the *Burkholderia cepacia* complex group of *Burkholderia* were characterised or used as biological control agents (Parke and Gurian-Sherman, 2001). They are highly active in their specialized metabolism, for example, producing the broad spectrum anti-fungal pyrrolnitrin with antagonistic activity against fungi or fungal-like oomycetes(Schmidt *et al.*, 2009). Three related anti-fungal compounds: xylocandins (Bisacchi *et al.*, 1987), occidiofungin (Lu *et al.*, 2009) and burkholdines (Tawfik *et al.*, 2010) were characterised in *Burkholderia cepacia* ATCC39277, *Burkholderia contaminans* MS14 and *B. ambifaria* 2.2N, respectively. Collectively, these specialized metabolites have exhibited antagonistic activity against the animal and plant pathogens *Candida albicans, Rhizoctonia solani, Alternaria alternata, Phytophthora infestans* and *Pythium* species (Bisacchi *et al.*, 1987; Lu *et al.*, 2009; Tawfik *et al.*, 2010). Additional antimicrobial and cytotoxic activities exhibited by *Burkholderia* include; the anti-Gram-negative polyketide enacyloxin IIa (Mahenthiralingam *et al.*, 2011), the anti-fungal polyyne cepacin (Parker *et al.*, 1984) and the anti-fungal lipopeptide AFC-B11 (Kang *et al.*, 1998). Outside of the *B. cepacia* complex, other *Burkholderia* species have also been shown to produce a range of antagonistic compounds. The bactobolins are a collection of structurally similar anti-bacterial compounds characterised in several *Burkholderia thailandensis and Burkholderia pseudomallei* strains (Seyedsayamdost *et al.*, 2010). These chlorinated compounds exhibit potent activity against Gram-positive bacteria, including methicillin-resistant *Staphylococcus aureus* (Duerkop *et al.*, 2009; Seyedsayamdost *et al.*, 2010). *Burkholderia* *gladioli* also produces multiple antimicrobials including the anti-mycobacterial macrolide gladiolin (Song *et al.*, 2017), its isomer lagriene (Flórez *et al.*, 2017), the cytotoxic azapteridine toxoflavin (Levenberg and Linton, 1966; Kim *et al.*, 2004), and the polyyne caryoynencin (Ross *et al.*, 2014).

There is no consensus on the distribution or co-occurrence of antimicrobial specialized metabolite encoding gene clusters in *B. ambifaria*, nor a holistic understanding of strain bioactivity and antimicrobial compound efficacy against priority bacterial and fungal plant pathogens (Howden *et al.*, 2009; Mansfield *et al.*, 2012). Biopesticidal activities have been descriptively characterised for individual producer strains against a panel of target organisms, or multiple producer strains against relatively few target organisms (Parke and Gurian-Sherman, 2001). Previous studies have analysed the overall antagonistic properties of *B. ambifaria*, but none have examined the role of specific antimicrobial compounds in mediating biocontrol in natural soil microcosm models. To establish a biotechnological platform for biopesticidal use of *B. ambifaria* that considers its efficacy and safety*,* we systematically defined the genomic basis and functional efficacy of antimicrobial metabolites in 64 strains including 8 previously characterised biocontrol strains (Table S1). The strain collection examined included 58 environmental isolates recovered from multiple sources (soil, the maize, pea and grass rhizosphere, and leaves) and 6 strains isolated from the sputum of people with cystic fibrosis (CF). Collectively, it also represented strains recovered from various geographic origins (USA, Australia and Italy; Table S1). The *B. ambifaria* core and accessory genome was revealed, and gene cluster network analyses were combined with antimicrobial activity assays to rationally understand the biopesticidal activity against crop pathogens. The role of individual antimicrobial metabolites in mediating crop protection was investigated using biosynthetic pathwaymutants in non-sterile soil biocontrol models. Since curing of the *B. cepacia* complex third genomic replicon is possible in these multireplicon bacteria (Agnoli *et al.*, 2012), a *B. ambifaria* c3 mutant was constructed and shown to have reduced virulence in a murine respiratory infection model, yet retained its plant-protective properties. This work provides a foundation for developing targeted biological control agents and effective biocontrol products for reducing agricultural crop losses from bacterial, fungal and oomycete pathogens.

**Results:**

**Establishing a holistic genomic understanding of *B. ambifaria’s* biological control potential**

To understand the genome-encoded potential of *B. ambifaria* as a biopesticide, a phylogenomic and pan-genomic analysis was applied (see Methods). The three replicon genomic structure (Lessie *et al.*, 1996) was present in 63 of the *B. ambifaria* strains analysed, while strain BCC1105 naturally lacked the third replicon (c3). Contigs were scaffolded to one of three reference genomes to assemble complete genomes. These assembled genome sizes varied across the 64 strains, from 6.13 Mbp (BCC1105) to 8.03 Mbp (BCC1248), with a mean of 7.34 Mbp (Table S2). Assembled replicons c1, c2 and c3 possessed a mean of 3.47 Mbp, 2.74 Mbp and 1.15 Mbp, respectively. Replicon c3 possessed the greatest variation in sequence capacity, whereas replicons c2 and c1 displayed a higher consistency in size (Table S2). A large *B. ambifaria* pan-genome was identified (22,376 distinct genes) of which 3784 genes comprised the core genome. The pan-genome represented a collection of genes approximately 3.4-fold greater than the mean *B. ambifaria* genome (6,546 genes). A large proportion of the accessory genome, 78.1% (14,582 genes), was shared by less than 15% of the *B. ambifaria* strains. Exclusion of the strain which lacked the third replicon, BCC1105, from the core genome analysis resulted in the *B. ambifaria* core genome increasing from 3784 to 4166 genes. Three major clades were identified in the *B. ambifaria* 3,784 core-gene phylogeny (Figure 1A), and this established the evolutionary framework onto which the antimicrobial properties of each strain were overlaid using *in silico* and bioactivity approaches.

***In silico* definition of *B. ambifaria* specialized metabolite biosynthetic clusters**

AntiSMASH analysis (Weber *et al.*, 2015) and BLAST (Morgulis *et al.*, 2008) searches for known specialized metabolite biosynthetic gene clusters (BGCs) detected a total of 1,272 BGCs across 64 *B. ambifaria* strains, defining a mean of 20 pathways per genome. Eighteen classes of metabolites were identified as well as one distinct BGC that was not recognised as encoding a previously reported metabolite class by antiSMASH, but was recognised as having the potential to direct the production of a specialized metabolite (there were 20 examples of this BGC in the genome set; see Figure 2). A combination of Kmer-matching and gene topology comparisons enabled de-replication of the 1,272 BGCs to 38 distinct putative BGCs (Figure 2; Table S3). Of the 38 distinct BGCs, three singleton pathways were detected in contigs that had no significant homology to the reference sequences, and could not be incorporated into the assembled replicons c1, c2 and c3 (Figure 2; Table S3). One of these BGCs encoded a hybrid non-ribosomal peptide synthetase *trans*-acyltransferase polyketide synthase (NRPS-*trans*-AT PKS), and showed 92% sequence similarity to the malleilactone biosynthetic gene cluster (Biggins, Ternei and Brady, 2012; Franke, Ishida and Hertweck, 2012). The two remaining BGCs with unknown genomic locations were an uncharacterised NRPS in strain MEX-5, and a LuxRI system in strain IOP40-10. The most frequently detected specialized metabolite classes were terpenes (5/38) and NRPS (4/38). Eleven of the distinct BGCs were encoded by all 64 *B. ambifaria* strains, and included four terpene synthase BGCs and two LuxRI systems (Figure 2). Eight clusters were detected in less than 5% of the *B. ambifaria* strains examined (<3 strains), three of which were type 1 modular PKS gene clusters (Figure 2).

The specialized metabolite encoding capacity of the replicons relative to their size (density) varied significantly (Figure S1). The largest replicon (c1) possessed the lowest specialized metabolite BGC density, whereas the smallest replicon (c3) possessed the largest density. The metabolite BGC density of the replicons was not static across the *B. ambifaria* dataset, with replicon c3 varying in BGC density from less than 10% to above 30% (mean 19.4%; Figure S1). Replicons c1 and c2 displayed a lower mean BGC density and density variance compared to c3. *B. ambifaria* strains encoded 3-6 BGCs on replicon c1, with eight distinct BGCs identified on the replicon. Replicon c2 encoded 8-11 BGCs with 17 distinct BGCs; and c3 encoded 4-9 BGCs with eleven distinct BGCs detected.

Of the 38 distinct pathways, 13 were previously characterised, and seven known to encode compounds with antimicrobial activity (Table S4). Pyrrolnitrin (Schmidt *et al.*, 2009) was the only BGC for an antimicrobial metabolite found in all 64 *B. ambifaria* strains, whereas the BGC for the anti-Gram-negative metabolite enacyloxin IIa (Mahenthiralingam *et al.*, 2011) was the least common known antimicrobial BGC (Figure 2). Pyrrolnitrin and phenazine BGCs were encoded on replicon c2 and the remaining antagonistic compounds were encoded by BGCs on replicon c3. No known antimicrobial BGCs were identified on replicon c1. Barring a few exceptions, multiple antimicrobial encoding BGCs were associated with distinct clades within the *B. ambifaria* core-gene phylogeny (Figure 1). Six of the seven clade 1b strains encoded the pathway responsible for enacyloxin IIa biosynthesis (Mahenthiralingam *et al.*, 2011). The more widely distributed burkholdine BGC (Tawfik *et al.*, 2010) was absent from all members of clade 2, and strain BCC1105, but all other strains possessed the anti-fungal biosynthetic locus. Bactobolin (Seyedsayamdost *et al.*, 2010) BGCs were concentrated in clade 1 and less frequently encountered in clade 2 and 3 strains. Two strains, BCC1105 and BCC1224, only encoded the core anti-fungal metabolite pyrrolnitrin, and lacked any additional antimicrobial BGCs (Figure 1B). No single strain encoded all 7 previously known antimicrobial BGCs, however, approximately 59% of strains encoded four or more BGCs reflecting *B. ambifaria’s* known antimicrobialproperties (Figure 1B).

The silent nature of certain antimicrobial BGCs which are not expressedin standard laboratory cultures, including those in *Burkholderia*, is well established (Mahenthiralingam *et al.*, 2011; Masschelein, Jenner and Challis, 2017). We therefore correlated *in vitro* metabolite production with BGC distribution. Ten *B. ambifaria* strains representing the seven characterised biocontrol strains and three additional strains from the broader species phylogeny (Figure 1A) were screened for metabolite production on agar growth media BSM-G (Hareland *et al.*, 1975; Mahenthiralingam *et al.*, 2011). Six known antimicrobial metabolites were detected by LC-MS (Figure S2 and S3), five of which could be directly correlated to the presence of predicted BGCs (Table 1). Under these screening conditions the majority of BGCs (22 of 25) were biosynthetically active and produced the corresponding metabolite; individual strains encoding pyrrolnitrin, burkholdines and hydroxyquinolines BGCs were exceptions to this trend (Table 1). A sixth known metabolite, cepacin A (Parker *et al.*, 1984), was also detected in *B. ambifaria* J82 (BCC0191) by analytical analyses (see methods) and subsequently correlated to a BGC (not recognised by antiSMASH v3) identified by searching for quorum sensing (QS) regulated BGCs (see below).

**Mapping direct antimicrobial activity against plant and animal pathogens**

Having established the presence of BGCs (Figures 1b and 2) and corresponding metabolites (Table 1), antagonism activity of the 64 *B. ambifaria* strains against priority plant (Mansfield *et al.*, 2012) and human pathogens (Table S5) was evaluated (Mahenthiralingam *et al.*, 2011). The *in vitro* bioactivity was aligned against the core-gene phylogeny to map antagonism across *B. ambifaria* as a species (Figure 1c). A total of six strains lacked observable antimicrobial activity (Figure 1c). Clade 1a, 1b and 1c strains exhibited substantial bioactivity against Gram-negative pathogens, while only two strains outside these clades exhibited similar activity (Figure 1c). Clade 1b strains exhibited additional strong antagonistic activity towards Betaproteobacteria, *B. multivorans* and Alphaproteobacteria, *R. radiobacter* (Figure 1c), an activity not observed in other anti-Gram-negative *B. ambifaria* strains. The extended antimicrobial antagonism of clade 1b *B. ambifaria* correlated to the presence of the *trans*-AT PKS BGC for enacyloxin IIa (Figure 1b); all screened Gram-negative pathogens were susceptible to purified enacyloxin IIa with MICs ranging from 3.2 to 50 µg/ml (Table S6). The additional anti-Gram-negative activity correlated to the presence of the hybrid NRPS-PKS encoding BGC for bactobolin (Figure 1b). Anti-fungal and anti-Gram-positive activity was more widespread than anti-Gram-negative activity in *B. ambifaria*, with 82% and 69% of tested strains (62 of 64) exhibiting these activities (>10 mm diameter against at one or more organisms,), respectively (Figure 1). Clade 2 *B. ambifaria* strains exhibited the least antimicrobial activity (5 of 9 lacking any *in vitro* observable activity) despite encoding BGCs for pyrrolnitrin, hydroxyquinolines, and cepacin (Figure 1b).

**A search for QS-regulated BGCs reveals the biosynthetic locus for the potent anti-oomycetal cepacin**

Multiple specialized metabolite BGCs are QS controlled (Duerkop *et al.*, 2009; Schmidt *et al.*, 2009; Seyedsayamdost *et al.*, 2010; Mahenthiralingam *et al.*, 2011) and manipulation of this regulatory system has also been harnessed for *Burkholderia* metabolite discovery (Ishida *et al.*, 2010). The availability of the extensive *B. ambifaria* genome and BGC datasets enabled interrogation of QS regulatory genes with a focus on LuxR-encoding genes as the key BGC pathway regulators. We detected 356 *luxR* homologues across the 64 *B. ambifaria* strains, representing 14 distinct protein phylogenetic clades (Figure 3). These clades included the following *B. ambifaria* LuxRI quorum sensing (QS) systems: the *bafRI* system (Aguilar *et al.*, 2003) (63 of 64 strains), the *cepR2I2* system (Chapalain *et al.*, 2017) (61 of 64 strains), one functionally uncharacterised system (22 of 64 strains), and a QS system present only in *B. ambifaria* IOP40-10. Six LuxR clades were associated with BGCs encoding the following compounds or compound classes (Figure 3): ectoine, lantipeptide, butyrolactone, enacyloxin IIa, bactobolins and a putative BGC that was identified as directing cepacin biosynthesis as described below. The remaining LuxR clades flanked membrane transporter genes, a type 3 secretion system and four clades were associated with genes of unknown collective functions (Figure 3).

Downstream of an uncharacterised LuxRI system (encoded by 22 of 64 strains) (Figure 3) was a conspicuous BGC of interest encoding fatty acid desaturases, a beta-ketoacyl synthase and an acyl-carrier protein (with a phosphopantetheine-binding domain) (Figure 4a). Insertional mutagenesis of a fatty acyl-adenosine monophosphate (AMP) ligase-encoding gene, *ccnJ*, within this cluster was carried out in six *B. ambifaria* strain backgrounds (BCC0191, BCC1252, BCC1241, BCC0477, BCC1259 and BCC1218; Figure 1). The resulting mutants lacked anti-Gram-positive activity and the weak anti-Gram-negative activity, and showed considerably diminished growth inhibition of the oomycete *Pythium ultimum* (Figure 4b and S4). High resolution mass spectrometry of metabolite extracts from strain BCC0191 identified ions with *m/z* = 271.0964, corresponding to a predicted molecular formula of C16H14­O4 consistent with cepacin A, a historically described *Burkholderia* polyyne (Parker *et al.*, 1984) of un-defined biosynthetic origin. Direct comparison with extracts from the originally reported cepacin A and cepacin B producer strain, “*B. cepacia*” ATCC 39356 (Parker *et al.*, 1984) (taxonomically reclassified as a *Burkholderia diffusa* strain), confirmed that *B. ambifaria* BCC0191produces cepacin A (Figure S5). Cepacin A was absent in the *B. ambifaria* BCC0191::*ccnJ* cepacin insertional mutant (Figure 4c), confirming that the novel LuxRI-associated BGC was responsible for the biosynthesis of this known *Burkholderia* metabolite. The cepacin A BGC is located on the second replicon of 22 *B. ambifaria* strains, with 100% presence in clade 3 strains and 56% encoding in clade 2 strains (Figure 1b).

**Cepacin A is a key mediator of *B. ambifaria* biocontrol of *Pythium* *ultimum* damping-off disease**

*B. ambifaria* has been observed to inhibit *P. ultimum* and application to prevent crop damping-off diseases was a key trait in its historical biopesticide use(Parke and Gurian-Sherman, 2001). However, the metabolites and/or biosynthetic pathways which drive *Burkholderia* crop protection against *Pythium*-mediated damping-off have not been defined in a relevant biopesticide model, such as bacterial seed coating and planting in pathogen infested soil (Parke and Gurian-Sherman, 2001). The cepacin-producer *B. ambifaria* BCC0191 exhibited strong biopesticidal activity when introduced as a *P. sativum* (pea) seed-coat to a *P. ultimum­* biocontrol model in non-sterile soil (Figure 5a). Disruption of the cepacin BGC and application of the BCC0191 cepacin mutant as a seed coat reduced pea plantsurvival rates by more than 60% dependent on *B. ambifaria* seed coat inoculum level (105, 106 and 107 cfu/seed; Figure 5a). No biological control was observed when 105 cfu/seed of BCC0191 cepacin mutant was applied (<10% survival), compared to >50% protection mediated by the wild type at this level (Figure 5a). A unique feature of the *B. cepacia* complex multi-replicon genome is that the third c3 replicon is not essential and c3 deletion mutants lose virulence and antifungal phenotypes (Agnoli *et al.*, 2012). The cepacin BGC is located on the second c2 replicon of *B. ambifaria* and its biosynthesis was unaffected when a third replicon deletion mutant, BCC0191Δc3, was constructed. Despite the loss of >1 Mb of DNA, the BCC0191Δc3 remained competitive and biopesticidal in the *Pythium-*infested soil microbial community, protecting peas from damping-off at a rate marginally below that of the wild type (Figure 5b; the difference was not significant for a given inoculation size).

The phenotypes of *B. ambifaria* BCC0191, its cepacin-deficient derivative (::*ccnJ*), a third replicon c3 knockout mutant (Δc3), and combined mutation (::*ccnJ*Δc3) were tested further to understand the wider effect of these mutations on strain fitness. Antimicrobial activity against the panel of plant and animal pathogenic bacteria and fungi, as well as further reference strains, was examined. The loss of cepacin A production in BCC0191::*ccnJ* resulted in loss of anti-Gram-positive activity against *Staphylococcus aureus*, *Enterococcus* *faecalis* and *Bacillus* *subtilis* (Figure S6), in addition to the loss of *Pythium* inhibition and the weak activity against certain Gram-negative bacteria (Figure 4b; Figure S4). Deletion of the third replicon resulted in loss of antagonism against the fungal species *Candida* *albicans*, *Fusarium* *solani* and *Alternaria* *alternata*, but enhanced anti-Gram positive activity which correlated with a 2-fold increase in cepacin production seen in the BCC0191Δc3 mutant (Figure S6). The double mutant, BCC0191::*ccnJ*Δc3, lost all the antimicrobial phenotypes observed for the wild-type *B. ambifaria* BCC0191. The rhizocompetence of *B. ambifaria* BCC0191 WT, BCC0191::*ccnJ* and BCC0191Δc3 mutants was also evaluated to assess whether their effect on biological control of *Pythium* (Figure 5) was a consequence of reduced root colonisation (see Supplementary Methods). After 14 days of growth in the pea biological control model system (not containing *P. ultimum*), the wild-type and BCC0191::*ccnJ* colonised the rhizosphere at equivalent levels (Table S7). The BCC0191Δc3 mutant colonised the pea rhizosphere at a rate significantly lower (p < 0.05), but within 0.5 log of the mean wild-type level of 2.8 x 107 cfu/g root (Table S7).

A lack of understanding of safety and human pathogenicity were keys reasons the US EPA placed a moratorium on new registration of *B. cepacia* complex biopesticides (Parke and Gurian-Sherman, 2001). Since the BCC0191Δc3 mutant had retained its biopesticidal ability (Figure 5b), yet loss of this replicon is associated with reduced virulence in multiple infection models (Agnoli *et al.*, 2012), we assessed the pathogenicity of *B. ambifaria* BCC0191 and its replicon deletion mutant. In the *Galleria* *melonella* wax-moth larvae model (Agnoli *et al.*, 2012), the deletion of the third replicon did not attenuate the virulence (Figure S7a), showing that genes encoding significant insecticidal pathogenicity were not encoded on c3 in *B. ambifaria* strain BCC0191. In contrast, using a murine respiratory infection model relevant to chronic cystic fibrosis lung infections (Fothergill *et al.*, 2014; Bricio-Moreno *et al.*, 2018), the virulence of *B. ambifaria* BCC0191 was low and loss of the third replicon BCC0191Δc3 further reduced persistence in the lung (Figure S7b and Figure S7c). At an infective dose of 2 x 106 bacteria, the BCC0191 wild type persisted in the nasopharynx for the duration of the 5 day experiment but, was cleared from lungs of 4 out of 6 mice by day 5. In contrast, the c3 mutant was rapidly cleared from both nasopharynx and lungs of mice (Figure S8). Low numbers of the parental BCC0191 strain (<50 colonies) were detected in the lungs of mice after 5 days of infection, but BCC0191Δc3 was cleared within 48 hours. *B. ambifaria* (wild-type or c3 mutant) was not detected within the spleens of infected mice and no visible disease signs were observed throughout. Genotyping by PCR demonstrated that the low numbers of colonies recovered from the mouse infection model were either the administered *B. ambifaria* BCC0191 or BCC0191Δc3, respectively (Figure S8).

**Discussion:**

Harnessing the potential of naturally biopesticidal bacteria is an important consideration if we are to keep pace with agricultural intensification and global food security. With increasing regulatory and environmental scrutiny of pesticides, the properties of natural agents will also have to be systematically defined before widespread use. Our in-depth genomic analysis of the *intra*-species diversity of *B. ambifaria* as a biopesticide and direct linkage of its specific metabolite, cepacin, to antagonism of *Pythium* and prevention of crop damping-off disease, sets a new precedent on the mode of action of *Burkholderia* biopesticides. We have developed a holistic understanding of biopesticidal *B. ambifaria,* determining their pan-genomic content, extensive library of antimicrobial biosynthetic pathways, efficacy in targeting key plant pathogens with specific antimicrobial metabolites, and defining the population biology of historically applied *B. ambifaria* biopesticides. We have shown that biological control of damping-off disease in a relevant soil model is critically mediated by cepacin A, encoded on the second replicon of *B. ambifaria*. Since effective biological control of *Pythium* also occurs in the absence of the third replicon, which has been characterised as a *Burkholderia* virulence plasmid (Agnoli *et al.*, 2012), we have highlighted this as an attenuation strategy for developing potentially safe biopesticide strains which retain biotechnological efficacy.

**Pan-genomics and extensive specialized metabolite diversity**

Genomic analysis across 64 *B. ambifaria* genomes revealed substantial diversity in gene content and predicted specialized metabolite BGCs; branching into five defined clades. As a mean, 56% of the BGCs were encoded by all studied *B. ambifaria* strains, the remaining BGCs constitute the accessory specialized metabolite potential of *B. ambifaria*. The large core and accessory metabolite potential in *B. ambifaria* compliments previous studies of the wider *Burkholderia* genus (Depoorter *et al.*, 2016). Similar in-depth, species-focussed BGC distribution analyses were conducted in three *Salinospora* species with previously unknown metabolite potential, but these were restricted to NRPS, PKS and fatty acid synthase pathways (Udwary *et al.*, 2007). A more limited analysis (6 strains) was also conducted with *Streptomyces albus* genomes (Seipke, 2015). Pan-genomic analysis of nine strains has been performed on the biocontrol species *Pseudomonas putida* (Udaondo *et al.*, 2016), but a detailed correlation to their specialized metabolite biosynthetic capacity and protective properties, such as that reported in this study, was lacking.

**Discovery of cepacin biosynthetic gene cluster**

Mining and phylogenetically clustering the LuxR protein sequences from 64 *B. ambifaria* genomes revealed multiple solo and *luxI-*associated *luxR* genes, and these were linked with both known and uncharacterised specialized metabolite BGCs. In addition to the *B. ambifaria* encoded enacyloxin (Mahenthiralingam *et al.*, 2011) and bactobolin (Seyedsayamdost *et al.*, 2010) BGCs, LuxR regulation of metabolites has been further described within and outside the genus. *Burkholderia thailandensis* synthesises the quorum sensing regulated cytotoxic compound malleilactone (Duerkop *et al.*, 2009), and pyocyanin production in *Pseudomonas aeruginosa* is controlled by a hierarchical QS network (Lee and Zhang, 2015). This approach was initially intended to understand the role of quorum sensing regulation in *B. ambifaria* biopesticidal specialized metabolism, but serendipitously led to the identification of the cepacin BGC.

Cepacins A and B were initially described as metabolites of *Burkholderia cepacia*, formally *Pseudomonas cepacia* (Parker *et al.*, 1984; Yabuuchi *et al.*, 1992), with the original producer strain now classified as *B. diffusa* (Vanlaere *et al.*, 2008). Both polyyne metabolites displayed strong anti-*Staphylococci* activity, while cepacin A showed weak anti-Gram-negative activity (Parker *et al.*, 1984). Cluster K, a gene cluster with 76.9% homologous nucleotide similarity spanning 12.9 kbp (in addition to 8.4 kbp of non-homologous regions) to the *B. ambifaria* cepacin A BGC was identified in *Collimonas fungivorans* Ter331 using nucleotide BLAST (Figure S9). The *C. fungivorans* cluster K has been linked to the biosynthesis of the anti-fungal polyyne collimomycin (Fritsche *et al.*, 2014) whose BGC organisation (Figure S9) and chemical formula (C16H18O4) resemble cepacin A (C16H14O4). Recent characterisation of this *C. fungivorans* strain showed it produces a range of polyynes, collectively designated the collimonins (Kai *et al.*, 2018), with collimonin A showing the most structural similarity to cepacin A. Several key differences in gene content were identified between the cepacin and collimonin BGC in the regions flanking the core biosynthetic genes (Figure S9). *C. fungivorans* cluster K contains genes encoding four additional hypothetical proteins, a major-facilitator superfamily transporter and fatty acid desaturase; whereas the *B. ambifaria* cepacin BGC encoded one extra hypothetical protein. An unusual feature of the cepacin and collimonin BGC variants is the substitution of the associated regulatory genes that comprise a QS-associated *luxRI* in *Burkholderia* and a *lysR* regulator in *Collimonas* (Figure S9). Similar proteins to those identified in the cepacin BGC are listed in Table S8.

**Cepacin A is a major component of *B. ambifaria* biological control**

We have demonstrated the importance of cepacin A in the context of biological control of the major crop family Fabaceae (*Pisum sativum*). Disruption of the cepacin BGC in *B. ambifaria* significantly reduced plant survival beyond germination and emergence compared to the wild-type (Figure 5a; p < 0.05). The contribution of specific metabolites to biocontrol has been studied extensively in *Pseudomonas* (Haas and Keel, 2003) relative to other characterised biocontrol genera. The anti-fungal properties of pyrrolnitrin and 2,4-diacetylphloroglucinol have been evidenced as important metabolites in the biological control of several fungal pathogens, on a diversity of crops, in a range of *Pseudomonas* species and strain backgrounds (Haas and Keel, 2003). Other studies have highlighted the *in vitro* antimicrobial activity, and presence of the corresponding specialized metabolite pathways or protective effects in field trials in *Bacillus* (Palazzini *et al.*, 2016) and *Streptomyces* (Law *et al.*, 2017), but fail to define the impact of distinct metabolites in a biocontrol system. This study establishes the role of cepacin A as the major bioactive component of the *B. ambifaria* armoury in the biocontrol of damping-off disease by *P. ultimum* in a relevant non-sterile soil model. The reduced protection against *P. ultimum* of the cepacin A-deficient mutant compared to the wild-type *B. ambifaria* also indirectly confirms the expression of the cepacin A BGC *in planta* (Figure 5b).

For *B. ambifaria* strains such as AMMD, deletion of the third genomic replicon rendered the derived mutant, AMMDΔc3, attenuated in multiple virulence models and resulted in loss of anti-fungal activity (Agnoli *et al.*, 2012), but retained the ability to colonise the rhizosphere (Vidal-Quist *et al.*, 2014). Our *B. ambifaria* BCC0191Δc3 mutant exhibited a non-significant reduction in biocontrol efficacy against *P. ultimum* (Figure 5b; p > 0.05), but exhibited a significantly reduced rhizocompetence, with the number of viable bacteria per g of root an order of magnitude below that of the wild-type (Table S7; p < 0.05). In contrast to AMMDΔc3 (Agnoli *et al.*, 2012), *B. ambifaria* BCC0191Δc3 retained virulence in a *Galleria* wax-moth larvae virulence assay (Figure S7a), suggesting in strain BCC0191 *Galleria* virulence is mediated by factors encoded on the first and second replicons. However, corroborating the reduced pathogenicity observed for c3 replicon mutants in vertebrate models (Agnoli *et al.*, 2012), *B. ambifaria* BCC0191Δc3 was not able to persist within the lungs of infected mice (Figure S8c). At the equivalent infective dose (2 x106 bacteria), highly virulent respiratory pathogens such as *Pseudomonas* *aeruginosa* persist in the same murine infection model at log-fold higher densities in both nasopharynx and lung (Fothergill *et al.*, 2014; Bricio-Moreno *et al.*, 2018). In addition, no *B. ambifaria* was detected within the spleens of infected mice suggesting that the capability of this species for invasive infection, observed during fatal “cepacia syndrome” CF infection (Parke and Gurian-Sherman, 2001), is low.

There has been considerable discussion on whether *Burkholderia* species known to cause opportunistic infections can be safely exploited for environmental benefit (Eberl and Vandamme, 2016). Multiple species in the new genus *Paraburkholderia* have not been associated with infection, are generally environmental, and mediate plant-beneficial interactions (Eberl and Vandamme, 2016). Transfer of biopesticidal properties such as the *B. ambifaria* cepacin BGC to *Paraburkholderia* species is a potential route to future safe usage. Attenuation of pathogenicity in biopesticidal strains is an alternate means to facilitate their biotechnological exploitation. Third replicon deletion in *B. ambifaria* BCC0191 led to loss of persistence in the murine lung infection model (Figure S7b and Figure S7c), and hence provides an unmarked means of attenuating pathogenicity but preserving biopesticidal potential in this strain (Figure 5). In addition the BCC0191Δc3 mutant also showed a reduced root colonisation after 14 days compared to the wild type (Table S7), suggesting it has less potential for bioaccumulation, which is another desirable trait for a biopesticide. Whether c3 deletion is sufficient to render *B. ambifaria* as a species completely avirulent remains to be fully determined. *B. ambifaria* is rarely found in CF lung infections, with survey of US patients from 1997 to 2007 implicating it collectively with several other *B. cepacia* complex species as causing <3% of all *Burkholderia* cases (LiPuma, 2010). A 2017 survey of *Burkholderia* infections in 361 UK CF patients did not find *B. ambifaria* at all (Kenna *et al.*, 2017). This epidemiological data combined with the low murine respiratory persistence of *B. ambifaria* (Figure S7b and Figure S7c) compared to virulent pathogens such as *P. aeruginosa* (Fothergill *et al.*, 2014; Bricio-Moreno *et al.*, 2018), suggests that *B. ambifaria* has low pathogenicity. From this start point, attenuation of virulence using unmarked c3 deletion as performed herein, combined with further essential gene mutation strategies as used to construct live bacterial vaccines, could also provide a route towards the development of safe *B. ambifaria* biopesticides.

**Multifaceted analysis of previously characterised biocontrol strains**

We analysed the *in vitro* antimicrobial activity, specialized metabolite BGCs, and metabolite profiles of ten strains spanning the *B. ambifaria* core-gene phylogeny, including seven of the eight previously characterised biological control strains (Table S1): AMMD (BCC0207), BC-F (BCC0203), J82 (BCC0191), M54 (BCC0316), Ral-3 (BCC0192), ATCC 53267 (BCC0284) and ATCC 53266 (BCC0338). Most of these biocontrol strains were initially characterised by agricultural and biotechnology companies as potent agents capable of suppressing multiple plant pathogens. Variation was observed in the antimicrobial pathway contents of the ten strains. The presence of multiple pathways encoding broad spectrum antimicrobials suggests a “built-in” redundancy to their biocontrol ability to suppress plant pathogens. Despite encoding antimicrobial biosynthetic pathways, there were several examples of non-producing clusters under the experimental conditions used. The absence of these metabolites in encoding strains may be due to multiple factors such as the presence of BGC or regulatory mutations within these strains, the need for specific growth conditions to prime biosynthesis, or that BGC expression is activated by complex interactions such as contact with a competing organism or interkingdom signalling from the host plant.

**Conclusion:**

Biological control agents have been applied to crops with success in the past, but no in-depth genomic or analytical chemistry analyses have been conducted on individual species to assess their biocontrol potential. This study demonstrated the benefits of genome mining and *in vitro* antimicrobial screening for defining major contributing biocontrol BGCs, can facilitate future rational design and formulation of bacterial biocontrol products. The broad host range of biocontrol illustrates the potential of cepacin-producing *B. ambifaria* in protecting economically relevant crop species from attack by bacterial, fungal and oomycete pathogens. It is clear that *B. ambifaria* has naturally evolved to accumulate multiple plant protective BGCs that underpin its historical exploitation as a biopesticide (Parke and Gurian-Sherman, 2001). With an urgent need to sustain crop protection and agricultural production, yet reduce use of environmentally persistent chemical pesticides, systematically repurposing natural biological control agents such as *B. ambifaria* for biotechnology is a timely alternative solution.

**Methods:**

**Genome sequencing and replicon assembly**

Most genomes were sequenced as part of this study, and the remainder downloaded from public databases. 125-nucleotide paired-end reads were generated for 60 *B. ambifaria* genomes using an Illumina HiSeq 2000. Illumina adaptors were trimmed and read quality assessed using the wrapper script Trim Galore v0.4.2 (Krueger, 2016), which utilises Cutadapt v1.12 (Martin, 2011) and FastQC v0.10.1 (Andrews, 2009), respectively. FLaSH v1.2.11 (Magoč and Salzberg, 2011) was used to merge overlapping short read pairs to improve contiguity of assembled genomes. The resulting overlapped and paired-end reads output from FLASH were assembled into contigs using the assembler SPAdes v3.9.1 (Bankevich *et al.*, 2012). Misassembled contigs were identified and corrected with Pilon v1.21 (Walker *et al.*, 2014). Contig sequences representing contaminating DNA were identified with Kraken v0.10.5-beta (Wood and Salzberg, 2014) using the Minikraken database, and removed prior to genomic analyses. Genome sequence quality assessment and statistics were calculated using QUAST v4.4 (Gurevich *et al.*, 2013). Genomic contigs were re-arranged and scaffolded into replicons by mapping the contigs against three reference genomes using CONTIGuator v2.7.4 (Galardini *et al.*, 2011). The option to fill gaps using strings of “N” was disabled. Reference genomes were *B. ambifaria* AMMD (SAMN02598309) and *B. ambifaria* MC40-6 (SAMN02598385), both obtained from the European Nucleotide Archive; the third reference, *B. ambifaria* BCC0203, was generated using Pacific Biosciences single molecule real time sequencing. The replicons were manually assessed for any scaffolding errors and corrected when necessary. Completed replicons (c1, c2 and c3) were re-circularised based on genes *dnaA*, *parA* and *parB*, respectively, using the software Circlator v1.2.1 (Li, 2013). The species validity of *B. ambifaria* dataset was defined by calculating the average nucleotide identity (ANI) shared between all available *B. ambifaria* genomes using PyANI v0.2.1 (Pritchard *et al.*, 2016). Two sequenced strains from this study (BCC1630 and BCC1638) and one publically available strain (RZ2MS16) were excluded from the dataset, using a 95% species threshold (Richter and Rossello-Mora, 2009). The remaining 64 *B. ambifaria* strains along with mutant derivatives used in this study are listed in the supplementary data (Table S1).

**Genome mining and specialized metabolite cluster network analysis**

All bioinformatics analyses were performed using the Cloud Infrastructure for Microbial Bioinformatics (CLIMB) computing resource (Connor *et al.*, 2016). Scaffolded replicons and non-scaffolded contigs were annotated using Prokka v1.12-beta (Seemann, 2014). Bioinformatics tool antiSMASH v.3.0.5 (Weber *et al.*, 2015) detected specialized metabolite BGCs in both scaffolded and non-scaffolded contig sequences. Known pathways that were not detected by antiSMASH were identified with nucleotide-nucleotide BLAST v2.6.0+ (Morgulis *et al.*, 2008). BGCs were dereplicated by clustering nucleotide sequences using pairwise Kmer-matching software Mash v1.1.1 (Ondov *et al.*, 2016); reporting a maximum p-value and maximum distance of 1 and 0.05, respectively. The resulting distance matrix was visualized with Cytoscape v3.4.0 (Shannon *et al.*, 2003), applying the Jaccard index, p-value and Mash distance (estimated mutation rate between sequences) as edge attributes. Duplicated edges between nodes and self-loops were removed from the network analysis. The resulting cluster network was further refined by comparing the gene topologies of pathway representatives from network clusters of the same specialized metabolite class; and splitting or merging network clusters where necessary. Comparisons of the specialized metabolite BGC potential of each replicon were visualised using the web tool BoxPlotR (Spitzer *et al.*, 2014).

**Genomic analysis and phylogenomics**

The core and accessory genome of the collective 64 *B. ambifaria* strains was calculated using Roary v3.7.0 (Page *et al.*, 2015) using a 95% minimum percentage identity for blastp, and core gene threshold of 99% occurrence across the 64 strains. The core gene alignment generated by Roary was used to construct an approximate-maximum-likelihood core-gene phylogeny with double-precision FastTree v2.1.9 (Price, Dehal and Arkin, 2010). The root position was determined by a secondary tree containing the outgroup species *Burkholderia* *vietnamiensis* G4 (PRJNA10696) (Figure S5). Once the root branch point was defined, a second tree was constructed using RAxML v8.2.11 (Stamatakis, 2014) with General Time Reversible (GTR) substitution and a GAMMA model of rate heterogeneity supported by 100 bootstraps. Sequence comparison figures were constructed using the Python application Easyfig (Sullivan, Petty and Beatson, 2011).

**Culture conditions and antimicrobial activity screens**

All *B. ambifaria* strains were grown in tryptic soy broth (TSB) at 37°C and aerated overnight, unless stated otherwise. The 64 *B. ambifaria* strains were screened for production of antimicrobials with antagonistic activity against ten plant and animal and human pathogens (Table S5). A standard overlay assay was used to screen for antimicrobial activity, as previously described (Mahenthiralingam *et al.*, 2011), with the amendment of using 400 μl overnight (O/N) culture of the susceptibility-testing organism per 100 ml half-strength iso-sensitest agar. The soft agar inoculum with *A. alternata* and *F. solani* involved the resuspension of mycelia in 1ml PBS from 9-day old cultures grown in 50 mm petri dishes on potato dextrose agar; and 320 μl of mycelial resuspension used per 100 ml half-strength iso-sensitest agar. 10x10 cm square Petri dishes containing basal salts medium (Hareland *et al.*, 1975) supplemented with 4 g/l glycerol (BSM-G) (Mahenthiralingam *et al.*, 2011) were used to screen six isolates concomitantly, and a replicator used to transfer the bacterial culture onto the agar surface from 96-well plates. *B. ambifaria* strains were grown on BSM-G for three days at 30°C. Each 96-well plate stored six strains tested for antimicrobial activity, with 200 μl O/N culture grown in TSB broth (Oxoid) per used well, and DMSO at a final concentration of 8% for -80°C storage. Following chloroform-killing, approximately 25 mL of antimicrobial susceptibility test organism-seeded half-strength iso-sensitest agar was poured over each 10x10 cm square Petri dish. The plates were incubated at the optimum temperatures of each susceptibility-test organism (Table S5). The heatmap of pathogen antagonism was created in statistics software R v3.2.3 via RStudio v0.99.484.

**Confirmation of the *Burkholderia* cepacin A cluster-metabolite link**

Insertional mutagenesis was used to disrupt the expression of the cepacin A pathway. Primers were designed to amplify a 649 bp region of the fatty AMP ligase-encoding gene (Table S9), yielding a final product of 707 bp. The product was amplified using the Taq PCR Master Mix Kit (Qiagen), and ligated into the suicide vector pGp-omega-Tp (Flannagan *et al.*, 2007) following restriction with *XbaI* and *EcoRI* (NEB). The resulting construct was transformed into competent *Escherichia coli* SY327 via heat-shock (maintained by trimethoprim selection; 50 μg/mL), and subsequently introduced into *B. ambifaria* via tri-parental mating with *E. coli* HB101 carrying the helper plasmid pRK2013 (kanamycin selection, 50 μg/mL). The transconjugants were selected using trimethoprim 150 μg/mL and polymyxin 600 U/mL. The presence and correct location of the insertional vector was confirmed by whole-genome sequencing. Comparative-HPLC and antimicrobial activity screens between the wild-type and insertional mutant confirmed the disruption of the cepacin BGC.

**Constructing and confirming *B. ambifaria* BCC0191Δc3 mutant**

The *B. ambifaria* BCC0191Δc3 mutant was constructed following the methods outlined in Agnoli *et al.* (2012). This involved using the pMiniC3 vector, a 12.6 kb plasmid constructed from the origin of replication of the *B. cenocepacia* H111 third replicon and containing its *repA*, *parB*, and *parA* genes, and trimethoprim resistance and sucrose counter selection cassettes (Agnoli *et al.*, 2012). In brief, the pMiniC3 vector was mated into the BCC0191 wild-type to displace the native c3 replicon via a tri-parental mating involving the donor *E. coli* MC1061 pMiniC3, recipient *B. ambifaria* BCC0191 and *E. coli* HB101 carrying the helper plasmid pRK2013. *B. ambifaria* BCC0191 pMinic3 clones were subsequently cured of pMinic3 by sucrose counter-selection. *B. ambifaria* BCC0191Δc3 clones were screened for the absence of both replicon c3 and plasmid pMinic3 by PCR using the DreamTaq Green PCR Master Mix (2X) (Thermo Scientific) (Table S9).

**Detection of LuxR homologues as specialized metabolite regulators**

LuxR-encoding gene homologues were identified by replicating the systematic *in silico* approach previously described (Gan *et al.*, 2014). In brief, a Hidden Markov Model was built to identify potential LuxR-encoding genes; and these candidates were annotated for encoded protein domains. Candidate genes were considered LuxR-encoding if all four conserved LuxR protein domains were detected. The extracted amino acid sequences were aligned using MAFFT v7.305b (Katoh *et al.*, 2002) and a phylogenetic tree generated using FastTree v2.1.9 (Price, Dehal and Arkin, 2010). The regulatory function of the *luxR* gene was inferred either from the literature or genes with putative functions starting within 5 kbp upstream and downstream of the *luxR* gene.

**Culture conditions, extraction protocol and high resolution mass spectrometry**

All *B. ambifaria* strains were grown at 30°C on BSM-G. The original cepacin producer, strain ATCC 395396, was obtained from the Belgium Coordinated Collection of Microorganisms where it is deposited as strain LMG 24093, and has been reclassified taxonomically as *Burkholderia diffusa*. Single plates were extracted by addition of 4 mL of acetonitrile for 2hrs, 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 (N2) at 1.6 bar; dry gas (N2) at 8 L min−1; dry temperature at 180 °C. Ion transfer conditions were: ion funnel 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.

**Biocontrol of *Pythium ultimum* using a *Pisum sativum* model**

The infestation of soil with *Pythium ultimum* Trowvar. *ultimum* (MUCL 16164) was developed from the methodology proposed in Toda *et al.* (2015). Plugs of *P. ultimum* were grown at approximately 22°C on potato dextrose agar (PDA) for three days. Infested soil was produced by mixing the surface mycelia from the PDA agar plates into a 5:1 compost:sand mixture (one-90 mm PDA petri per 120 g soil), and incubating the soil at approximately 22°C for three days. Potting mix was composed of 1% (w/w) *Pythium*-infested soil in unsterilised non-infested soil, or 100% unsterilised non-infested soil for a non-infested control. Unsterilised *P. sativum* seeds (Early Onward cultivar) were coated with *B. ambifaria* before planting. *B. ambifaria* coating suspension was produced as follows: an overnight (approximately 18 hrs) TSB culture of *B. ambifaria* was washed in sterile 1x volume phosphate buffer solution (PBS) and re-suspended in sterile 0.5x volume PBS. The suspension was adjusted to 10x 0.5 OD at 600 nm (approximately 109 cfu/ml), and either applied at the neat concentration or diluted to achieve the desired inoculum levels (108 and 107 cfu/ml). Control *P. sativum* seeds lacking the *B. ambifaria* coating were dipped in PBS. *P. sativum* plants were grown at 22°C with a 16-h light/8-h dark photoperiod (70% RH) for 14 days, and watered as required. Groups of ten seeds per inoculum and seed coat organism were assayed per experiment, and the experiment was performed in triplicate (total of 30 seeds assayed per data point). The rhizocompetence of *B. ambifaria* BCC0191 and its derived mutants under the biological control model conditions was evaluated essentially as described (Vidal-Quist *et al.*, 2014); full details are provided in the Supplementary Methods. Significant differences between BCC0191 wild-type and mutant derivatives was assessed using two sample t-test or Welch’s two sample t-test. Two sample t-test assumptions were normally distributed data (Shapiro-Wilk test) and equal variances (Bartlett test); Welch’s two sample t-test did not assume equal variances.

**Acknowledgements:**

AM is funded by a Biotechnology and Biological Sciences Research Council (BBSRC) South West doctoral training partnership award (BY1910 7007). EM, GC, TC and JP acknowledge additional support for genome mining from BBSRC award BB/L021692/1; CJ and MJ were funded by this award. M.J is currently the recipient of a BBSRC Future Leader Fellowship (BB/R01212/1). The Bruker maXis II UHPLC-ESI-Q-TOF-MS system used in this research was funded by the BBSRC (BB/M017982/1). GW was supported by awards to EM from the Life Sciences Bridging Fund and Wellcome Trust Institutional Strategic Support Fund held at Cardiff University. TRC acknowledges support from the Medical Research Council’s Cloud Infrastructure for Microbial Bioinformatics (MR/L015080/1). DN and AG acknowledge funding from a Wellcome Trust and Royal Society Sir Henry Dale Fellowship awarded to DN (Grant Number 204457/Z/16/Z). GC is the recipient of a Wolfson Research Merit Award from the Royal Society (WM130033). We thank Leo Eberl and Kirsty Agnoli for provision of mini-C3 used for third replicon deletion.

**Author contributions:**

The initial study to characterise the genomes of *B. ambifaria* as a biopesticide was conceived by EM, with additional aspects of study design added by AM, GC and JM. AM performed all aspects of the study with the exception of the LC-MS profiling, and was assisted by specific contributions from the following: datasets and input for genome sequencing and mining, EM, GC, JP and TC; genome assembly, phylogenomics, cluster mining and de-replication, MB; LuxR mining, EM; generation of a cepacin insertional mutant and antimicrobial activity screening, CJ; extraction, identification and fractionation of *Burkholderia* metabolites by HPLC, GW; LC-MS identification and confirmation of *B. ambifaria* antimicrobial metabolites, MJ and GC; biocontrol modelling EM, GW, and JM; evaluation and analysis of plant models, JM; *Galleria* virulence assays, GW and CJ; and murine infection modelling and analysis, AG and DN. AM and EM developed the first draft of the manuscript and all authors read and contributed towards finalisation of the study.

**Data Availability Statement:**

Sequence data that support the genomic findings of this study have been deposited in the European Nucleotide Archive with the accession/bioproject codes listed in Supplementary Table 1 (Table S1). The data that support the antimicrobial production, *P. sativum* and *G. mellonella* survival, and murine infection model findings of this study are available from the corresponding authors upon request.

**Competing interests:**

The authors do not have any competing interest to declare.

**Table 1. Correlation of BGC presence and *in vitro*metabolite production in *B. ambifaria*1**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ***B. ambifaria* strain (clade)** | **Pyrrolnitrin** | **Burkholdines** | **Hydroxyquinolines** | **Bactobolins** | **Enacyloxin IIa** | **Cepacins2** |
|  | BGC | Metabolite | BGC | Metabolite | BGC | Metabolite | BGC | Metabolite | BGC | Metabolite | BGC | Metabolite |
| **Characterised biocontrol strains** |  |  |  |  |  |  |  |  |  |  |  |  |
| 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) | + | + | - | - | + | + | - | - | - | - | + | - |

Footnotes:

1Grey cells highlight BGCs in specific strains where the corresponding metabolite was not detected.

2The metabolite cepacin was detected prior to identification of its biosynthetic gene cluster. Phylogenetic clade of strains indicated in parentheses.

**Figure legends**

**Figure 1. Core-gene phylogeny of 64 *B. ambifaria* strains (a) aligned with presence/absence grid of known antimicrobial specialized metabolite BGCs (b) and antimicrobial activity heatmap (c). (a)** The phylogenetic tree was constructed based on 3785 core genes identified and aligned using the software Roary. The root was determined using a secondary tree containing an outgroup species, *Burkholderia* *vietnamiensis* G4 (supplementary data). 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. The evolutionary distance scale bar represents the number of base substitutions per site. **(b)** The presence of the eight 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 (Hadfield *et al.*, 2018). **(c)** The antimicrobial activity of 62 *B. ambifaria* strains were defined by measuring the diameter of the zones of inhibition (mm) in duplicate and calculating the mean. Strains MEX-5 and IOP40-10 were not available for the antimicrobial production assay. The zone of clearing diameter measurements for each strain are ordered by phylogenetic position.

**Figure 2. Specialized metabolite cluster network analysis of 64 *B. ambifaria* strains.** Each node represents a specialized metabolite BGC extracted from a single *B. ambifaria* strain. A total of 1,272 BGCs were detected across the 64 strains, and dereplication indicated these represent 38 distinct BGCs (38 distinct network clusters). Nucleotide sequences were clustered using MinHash and visualized with Cytoscape. Node colours represent specialized metabolite classes, and numbers correspond to the number BGC examples (nodes) of each distinct BGC (network cluster). Core BGCs were defined as BGCs that occurred in >98% of *B. ambifaria* strains. Characterised BGCs known in the literature are labelled. BGCs responsible for pyrrolnitrin, AFC-BC11 and hydroxyquinolines biosynthesis are classified as Other (O) by antiSMASH but represent different metabolite classes not recognised by antiSMASH.

**Figure 3. Unrooted phylogeny of LuxR protein homologues extracted 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 5 kbp 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.

**Figure 4. Organization of the cepacin A biosynthetic gene cluster, LC-MS analysis of cepacin A production and antimicrobial screening of *B. ambifaria* BCC0191 wild-type (WT) and cepacin A deficient derivative (::*ccnJ*)*.* (a)** Organisation and putative function of genes within the cepacin A BGC; further annotation details are provided in Figure S9. 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 BCC0191 WT and BCC0191::*ccnJ*. Scale bar represents 20 mm. Images were converted to greyscale, brightness decreased by 20%, and contrast increased by 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). **(d)** Structure of cepacin A, the identity of which was confirmed by comparison to an authentic standard from a known producer (Figure S3).­­

**Figure 5. Biological control of *Pythium* damping-off disease is mediated by *B. ambifaria* cepacin. (a)** Pea germination (14 days) in *P. ultimum* infested soil observed for groups of 10 seeds coated with 107, 106 and 105 cfu, respectively, of BCC0191 wild-type (WT) and BCC0191*::ccnJ.* The overall percentage survival of germinating peas treated with the WT and BCC0191::*ccnJ B. ambifaria* strains is shown on the right of panel A. Data points represent survival of 10 seeds per replicate; with three replicate experiments, and error bars represent standard error of the mean. Survival was assessed as plants that had stems >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 sample t-test or Welch’s two sample t-test (\* = p < 0.05; \*\* = p < 0.01). **(b)** Pea germination (14 days) in *P. ultimum* infested soil observed for groups of 10 seeds coated with 107, 106 and 105 cfu, respectively, of BCC0191 WT and BCC0191Δc3*.* The overall percentage survival of germinating peas treated with BCC0191 WT and BCC0191Δc3 is shown on the right of panel B. No significant difference (p > 0.05) was observed in plant survival between BCC0191 WT and BCC0191Δc3 at the inoculum levels examined using two sample t-test.

**References:**

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. doi: 10.1111/j.1365-2958.2011.07937.x.

Aguilar, C. *et al.* (2003) ‘Identification of quorum-sensing-regulated genes of *Burkholderia cepacia*’, *Journal of bacteriology*. American Society for Microbiology (ASM), 185(21), pp. 6456–62. doi: 10.1128/jb.185.21.6456-6462.2003.

Andrews, S. (Babraham B. (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).

Bankevich, A. *et al.* (2012) ‘SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing’, *Journal of computational biology*. Mary Ann Liebert, Inc., 19(5), pp. 455–77. doi: 10.1089/cmb.2012.0021.

Biggins, J. B., Ternei, M. A. and Brady, S. F. (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. doi: 10.1021/ja3052156.

Bisacchi, G. S. *et al.* (1987) ‘Xylocandin: A new complex of antifungal peptides. II. Structural studies and chemical modifications’, *The Journal of Antibiotics*. Japan Antibiotics Research Association, 40(11), pp. 1520–1529. doi: 10.7164/antibiotics.40.1520.

Bricio-Moreno, L. *et al.* (2018) ‘Evolutionary trade-offs associated with loss of PmrB function in host-adapted *Pseudomonas aeruginosa*’, *Nature Communications*. Nature Publishing Group, 9(1), p. 2635. doi: 10.1038/s41467-018-04996-x.

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*. Frontiers, 8, p. 1021. doi: 10.3389/fmicb.2017.01021.

Connor, T. R. *et al.* (2016) ‘CLIMB (the Cloud Infrastructure for Microbial Bioinformatics): an online resource for the medical microbiology community’, *Microbial Genomics*. Microbiology Society, 2(9). doi: 10.1099/mgen.0.000086.

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

Duerkop, B. A. *et al.* (2009) ‘Quorum-sensing control of antibiotic synthesis in *Burkholderia thailandensis*’, *Journal of Bacteriology*, 191(12), pp. 3909–3918. doi: 10.1128/JB.00200-09.

Eberl, L. and Vandamme, P. (2016) ‘Members of the genus *Burkholderia*: good and bad guys’, *F1000Research*. Faculty of 1000 Ltd, 5. doi: 10.12688/f1000research.8221.1.

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. doi: 10.1128/IAI.01581-06.

Flórez, L. V. *et al.* (2017) ‘Antibiotic-producing symbionts dynamically transition between plant pathogenicity and insect-defensive mutualism’, *Nature Communications*, 8, p. 15172. doi: 10.1038/ncomms15172.

Fothergill, J. L. *et al.* (2014) ‘*Pseudomonas aeruginosa* adaptation in the nasopharyngeal reservoir leads to migration and persistence in the lungs’, *Nature Communications*. Nature Publishing Group, 5(1), p. 4780. doi: 10.1038/ncomms5780.

Franke, J., Ishida, K. and Hertweck, C. (2012) ‘Genomics-driven discovery of burkholderic acid, a noncanonical, cryptic polyketide from human pathogenic *Burkholderia* species’, *Angewandte Chemie International Edition*, 51(46), pp. 11611–11615. doi: 10.1002/anie.201205566.

Fritsche, K. *et al.* (2014) ‘Biosynthetic genes and activity spectrum of antifungal polyynes from *Collimonas fungivorans* Ter331’, *Environmental Microbiology*, 16(5), pp. 1334–1345. doi: 10.1111/1462-2920.12440.

Galardini, M. *et al.* (2011) ‘CONTIGuator: a bacterial genomes finishing tool for structural insights on draft genomes’, *Source code for biology and medicine*. BioMed Central, 6, p. 11. doi: 10.1186/1751-0473-6-11.

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*. Frontiers Media SA, 4, p. 188. doi: 10.3389/fcimb.2014.00188.

Gurevich, A. *et al.* (2013) ‘QUAST: quality assessment tool for genome assemblies’, *Bioinformatics*. Oxford University Press, 29(8), pp. 1072–1075. doi: 10.1093/bioinformatics/btt086.

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*. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA, 41(1), pp. 117–153. doi: 10.1146/annurev.phyto.41.052002.095656.

Hadfield, J. *et al.* (2018) ‘Phandango: an interactive viewer for bacterial population genomics’, *Bioinformatics*. Edited by J. Kelso, 34(2), pp. 292–293. doi: 10.1093/bioinformatics/btx610.

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).

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. doi: 10.1111/j.1364-3703.2009.00595.x.

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. doi: 10.1021/ja105003g.

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*. American Chemical Society, 20(12), pp. 3536–3540. doi: 10.1021/acs.orglett.8b01311.

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*. American Society for Microbiology (ASM), 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*. Oxford University Press, 30(14), pp. 3059–66. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12136088 (Accessed: 30 October 2017).

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. doi: 10.1099/jmm.0.000458.

Kim, J. *et al.* (2004) ‘Quorum sensing and the LysR-type transcriptional activator ToxR regulate toxoflavin biosynthesis and transport in *Burkholderia glumae*’, *Molecular Microbiology*. Blackwell Science Ltd, 54(4), pp. 921–934. doi: 10.1111/j.1365-2958.2004.04338.x.

Krueger, F. (2016) *Trim Galore!*, *Babraham Bioinformatics*. Available at: https://www.bioinformatics.babraham.ac.uk/projects/trim\_galore/ (Accessed: 7 April 2017).

Law, J. W.-F. *et al.* (2017) ‘The potential of *Streptomyces* as biocontrol agents against the Rice Blast Fungus, Magnaporthe oryzae (Pyricularia oryzae)’, *Frontiers in Microbiology*. Frontiers, 8, p. 3. doi: 10.3389/fmicb.2017.00003.

Lee, J. and Zhang, L. (2015) ‘The hierarchy quorum sensing network in *Pseudomonas aeruginosa*’, *Protein & cell*. Springer, 6(1), pp. 26–41. doi: 10.1007/s13238-014-0100-x.

Lessie, T. G. *et al.* (1996) ‘Genomic complexity and plasticity of *Burkholderia cepacia*’, *FEMS microbiology letters*, 144(2–3), pp. 117–28. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8900054 (Accessed: 27 March 2018).

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).

Li, H. (2013) ‘Circlator: automated circularization of genome assemblies using long sequencing reads’, 16(294). doi: 10.1186/s13059-015-0849-0.

LiPuma, J. J. (2010) ‘The Changing Microbial Epidemiology in Cystic Fibrosis’, *Clinical Microbiology Reviews*, 23(2), pp. 299–323. doi: 10.1128/CMR.00068-09.

Lu, S. E. *et al.* (2009) ‘Occidiofungin, a unique antifungal glycopeptide produced by a strain of *Burkholderia contaminans*’, *Biochemistry*. American Chemical Society, 48(35), pp. 8312–8321. doi: 10.1021/bi900814c.

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. doi: 10.1093/bioinformatics/btr507.

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. doi: 10.1016/j.chembiol.2011.01.020.

Mansfield, J. *et al.* (2012) ‘Top 10 plant pathogenic bacteria in molecular plant pathology’, *Molecular Plant Pathology*, 13(6), pp. 614–629. doi: 10.1111/j.1364-3703.2012.00804.x.

Martin, M. (2011) ‘Cutadapt removes adapter sequences from high-throughput sequencing reads’, *EMBnet.journal*, 17(1), p. 10. doi: 10.14806/ej.17.1.200.

Masschelein, J., Jenner, M. and Challis, G. L. (2017) ‘Antibiotics from Gram-negative bacteria: a comprehensive overview and selected biosynthetic highlights’, *Natural Product Reports*. Royal Society of Chemistry, 34(7), pp. 712–783. doi: 10.1039/C7NP00010C.

Mishra, J. *et al.* (2015) ‘Biopesticides: Where We Stand?’, in *Plant Microbes Symbiosis: Applied Facets*. New Delhi: Springer India, pp. 37–75. doi: 10.1007/978-81-322-2068-8\_2.

Morgulis, A. *et al.* (2008) ‘BLAST+: architecture and applications’, *Bioinformatics*. BioMed Central, 24(16), pp. 1757–1764. doi: 10.1093/bioinformatics/btn322.

Ondov, B. D. *et al.* (2016) ‘Mash: fast genome and metagenome distance estimation using MinHash’, *Genome biology*, 17(1), p. 132. doi: 10.1186/s13059-016-0997-x.

Page, A. J. *et al.* (2015) ‘Roary: rapid large-scale prokaryote pan genome analysis’, *Bioinformatics*, 31(22), pp. 3691–3693. doi: 10.1093/bioinformatics/btv421.

Palazzini, J. M. *et al.* (2016) ‘*Bacillus velezensis* RC 218 as a biocontrol agent to reduce Fusarium head blight and deoxynivalenol accumulation: Genome sequencing and secondary metabolite cluster profiles’, *Microbiological Research*. Urban & Fischer, 192, pp. 30–36. doi: 10.1016/J.MICRES.2016.06.002.

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*. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA, 39(1), pp. 225–258. doi: 10.1146/annurev.phyto.39.1.225.

Parker, W. L. *et al.* (1984) ‘Cepacin A and cepacin B, two new antibiotics produced by *Pseudomonas cepacia*’, *The Journal of Antibiotics*. Japan Antibiotics Research Association, 37(5), pp. 431–440. doi: 10.7164/antibiotics.37.431.

Price, M. N., Dehal, P. S. and Arkin, A. P. (2010) ‘FastTree 2 - Approximately maximum-likelihood trees for large alignments’, *PLoS ONE*. Edited by A. F. Y. Poon. Public Library of Science, 5(3), p. e9490. doi: 10.1371/journal.pone.0009490.

Pritchard, L. *et al.* (2016) ‘Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens’, *Anal. Methods*. The Royal Society of Chemistry, 8(1), pp. 12–24. doi: 10.1039/C5AY02550H.

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. doi: 10.1073/pnas.0906412106.

Ross, C. *et al.* (2014) ‘The molecular basis of conjugated polyyne biosynthesis in phytopathogenic bacteria’, *Angewandte Chemie - International Edition*, 53(30), pp. 7794–7798. doi: 10.1002/anie.201403344.

Schmidt, S. *et al.* (2009) ‘Production of the antifungal compound pyrrolnitrin is quorum sensing-regulated in members of the *Burkholderia cepacia* complex’, *Environmental Microbiology*. Blackwell Publishing Ltd, 11(6), pp. 1422–1437. doi: 10.1111/j.1462-2920.2009.01870.x.

Seemann, T. (2014) ‘Prokka: rapid prokaryotic genome annotation’, *Bioinformatics*, 30(14), pp. 2068–2069. doi: 10.1093/bioinformatics/btu153.

Seipke, R. F. (2015) ‘Strain-Level Diversity of Secondary Metabolism in *Streptomyces albus*’, *PLOS ONE*. Edited by P. Hoskisson. Public Library of Science, 10(1), p. e0116457. doi: 10.1371/journal.pone.0116457.

Seyedsayamdost, M. R. *et al.* (2010) ‘Quorum-sensing-regulated bactobolin production by *Burkholderia thailandensis* E264’, *Organic Letters*. American Chemical Society, 12(4), pp. 716–719. doi: 10.1021/ol902751x.

Shannon, P. *et al.* (2003) ‘Cytoscape: A software environment for integrated models of biomolecular interaction networks’, *Genome Research*. Cold Spring Harbor Laboratory Press, 13(11), pp. 2498–2504. doi: 10.1101/gr.1239303.

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. doi: 10.1021/jacs.7b03382.

Spitzer, M. *et al.* (2014) ‘BoxPlotR: a web tool for generation of box plots’, *Nature methods*. Europe PMC Funders, 11(2), pp. 121–2. doi: 10.1038/nmeth.2811.

Stamatakis, A. (2014) ‘RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies’, *Bioinformatics (Oxford, England)*. Oxford University Press, 30(9), pp. 1312–3. doi: 10.1093/bioinformatics/btu033.

Sullivan, M. J., Petty, N. K. and Beatson, S. A. (2011) ‘Easyfig: a genome comparison visualizer.’, *Bioinformatics (Oxford, England)*. Oxford University Press, 27(7), pp. 1009–10. doi: 10.1093/bioinformatics/btr039.

Tawfik, K. A. *et al.* (2010) ‘Burkholdines 1097 and 1229, potent antifungal peptides from *Burkholderia ambifaria* 2.2N’, *Organic Letters*. American Chemical Society, 12(4), pp. 664–666. doi: 10.1021/ol9029269.

Toda, T. *et al.* (2015) ‘Widespread occurrence of *Pythium arrhenomanes* pathogenic to rice seedlings around Japanese rice fields’, *Plant Disease*. Plant Disease, 99(12), pp. 1823–1831. doi: 10.1094/PDIS-01-15-0124-RE.

Udaondo, Z. *et al.* (2016) ‘Analysis of the core genome and pangenome of *Pseudomonas putida*’, *Environmental Microbiology*, 18(10), pp. 3268–3283. doi: 10.1111/1462-2920.13015.

Udwary, D. W. *et al.* (2007) ‘Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*’, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 104(25), pp. 10376–81. doi: 10.1073/pnas.0700962104.

Vanlaere, E. *et al.* (2008) ‘*Burkholderia latens* sp. nov., *Burkholderia diffusa* sp. nov., <i>Burkholderia arboris>/i> sp. nov., *Burkholderia seminalis* sp. nov. and *Burkholderia metallica* sp. nov., novel species within the *Burkholderia cepacia* comple’, *International Journal of Systematic and Evolutionary Microbiology*, 58(7), pp. 1580–1590. doi: 10.1099/ijs.0.65634-0.

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. doi: 10.1099/mic.0.074351-0.

Walker, B. J. *et al.* (2014) ‘Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement.’, *PloS one*. Public Library of Science, 9(11), p. e112963. doi: 10.1371/journal.pone.0112963.

Weber, T. *et al.* (2015) ‘antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters’, *Nucleic acids research*. Oxford University Press, 43(W1), pp. W237-43. doi: 10.1093/nar/gkv437.

Wood, D. E. and Salzberg, S. L. (2014) ‘Kraken: ultrafast metagenomic sequence classification using exact alignments’, *Genome Biology*. BioMed Central, 15(R46). doi: 10.1186/1471-2105-12-385.

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).