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

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

Authors:

Alex J. Mullins[1], James. A. H. Murray[1], Matthew J. Bull[1], Matthew Jenner[2], Cerith Jones[1,6], Gordon Webster[1], Angharad E. Green[3], Daniel R. Neill[3], Thomas R. Connor[1], Julian Parkhill[4], Gregory L. Challis[2,5] and Eshwar Mahenthiralingam[1]

Author affiliations and footnotes

[1]Microbiomes, Microbes and Informatics Group, Organisms and Environment Division, School of Biosciences, Cardiff University, Cardiff CF10 3AX, United Kingdom

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

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

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

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

[6]Current 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 potential pathogenic risk. A panel of 64 *B. ambifaria* strains demonstrated significant antimicrobial activity against priority plant pathogens. Genome sequencing, specialized metabolite biosynthetic gene cluster mining and metabolite analysis revealed an armoury of known and unknown pathways within *B. ambifaria*. The biosynthetic 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 biopesticidal protection and overall 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. 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[1,2]. 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[2].

Multiple species within the *Burkholderia cepacia* complex group were characterised or used as biological control agents[2]. They are highly active in their specialized metabolism, for example, producing antifungal compounds including pyrrolnitrin, occidiofungin, cepafungin and burkholdines; antibacterial bactobolins

and enacyloxin IIa; and broader spectrum agents such as the cepacins[3,4]. Outside of the *B. cepacia* complex, other *Burkholderia* species also produce a range of antagonistic compounds. The bactobolins are also produced by *Burkholderia thailandensis and Burkholderia pseudomallei* and exhibit potent activity against Gram-positive bacteria, including methicillin-resistant *Staphylococcus aureus*[5]. *Burkholderia gladioli* also produces multiple antimicrobials including the anti-mycobacterial macrolide gladiolin[6], its isomer lagriene[7], the cytotoxic azapteridine toxoflavin[8], and the polyyne caryoynencin[9]. There is no consensus on the distribution or co-occurrence of antimicrobial specialized metabolite encoding gene clusters in biopesticidal *B. ambifaria*, nor a holistic understanding of strain bioactivity and antimicrobial compound efficacy against priority bacterial[10,11], fungal and oomycetal plant pathogens[2]. 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[2]. 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 (Supplementary Table 1). 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; Supplementary Table 1). 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 pathway mutants in non-sterile soil biocontrol models. Since curing of the *B. cepacia* complex third genomic replicon (c3) is possible in these multireplicon bacteria[12], 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:

B. ambifaria genomics and *in silico* definition of specialized metabolite biosynthetic gene clusters

To understand the genome-encoded potential of *B. ambifaria* as a biopesticide, phylogenomic and pan-genomic analyses were applied (see Methods, Supplementary Figure 1, Supplementary Table 2 and Supplementary Notes). The three replicon genomic structure was present in 63 of the *B. ambifaria* strains analysed, while strain BCC1105 naturally lacked the third replicon. Contigs were scaffolded to one of three reference genomes to assemble complete genomes. The 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 (Supplementary Table 2). 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 greater consistency in size (Supplementary Table 2). 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* 3784 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* analyses of the 64 *B. ambifaria* genomes (see Methods, Supplementary Table 3 and Supplementary Notes) detected 1272 specialized metabolite biosynthetic gene clusters (BGCs), that were de-replicated into 38 distinct BGCs after Kmer-matching and gene topology comparisons (Figure 2; Supplementary Table 3). Network analysis was used to graphically summarise multiple attributes of the *B. ambifaria* BGCs including their biosynthetic diversity, strain distribution, and core or accessory nature within the species (Figure 2).

Of the 38 distinct BGCs, 13 were previously characterised, and seven known to encode compounds with antimicrobial activity (Supplementary Table 4). Pyrrolnitrin[13] 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[3] 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[3]. The more widely distributed burkholdine[14] BGC was absent from all members of clade 2 and strain BCC1105, but all other strains possessed the anti-fungal biosynthetic locus. Bactobolin[5] 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 the known antimicrobial properties of *B. ambifaria* (Figure 1B).

The silent nature of certain antimicrobial BGCs which are not expressed in standard laboratory cultures, including those in *Burkholderia*, is well established[3,4]. 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[3,15]. Six known antimicrobial metabolites were detected by LC-MS (Supplementary Figure 2 and 3), 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[16], was also detected in *B. ambifaria* J82 (BCC0191) by analytical analyses (see Methods) and subsequently correlated to a BGC (not recognised by antiSMASH v3)[17] 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[10] and human pathogens (Supplementary Table 5) was evaluated as described[3]. 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, *Burkholderia multivorans* and Alphaproteobacteria, *Rhizobium 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 (Supplementary Table 6). The additional anti-Gram-negative activity correlated to the presence of the hybrid NRPS-PKS encoding BGC for bactobolin (Figure 1b). Anti-fungal (against *Candida albicans, Fusarium solani* and *Alternaria alternata*) and anti-Gram-positive (against *Staphylococcus aureus*, *Enterococcus faecalis* and *Bacillus subtilis*) 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, 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[3,4] and manipulation of this regulatory system has also been harnessed for *Burkholderia* metabolite discovery[18]. A search of *luxR* homologues within the *B. ambifaria* genomes was performed (see Supplementary Notes) and downstream of an uncharacterised LuxRI system (encoded by 22 of 64 strains) (Figure 3) was a conspicuous BGC encoding fatty acid desaturases, a beta-ketoacyl synthase and an acyl carrier protein (Figure 4a). Insertional mutagenesis of a fatty acyl-adenosine monophosphate (AMP) ligase-encoding gene, *ccnJ*, within this BCG 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 Supplementary Figure 4). High resolution mass spectrometry of metabolite extracts from strain BCC0191 identified ions with *m/z* = 271.0964, corresponding to a predicted molecular formula of $C_{16}H_{14}O_4$ consistent with cepacin A, a historically described *Burkholderia* polyyne[16] of un-defined biosynthetic origin. Direct comparison with extracts from the originally reported cepacin A and cepacin B producer strain, "*B. cepacia*" ATCC 39356[16] (taxonomically reclassified as a *Burkholderia diffusa* strain), confirmed that *B. ambifaria* BCC0191 produces cepacin A (Supplementary Figure 5). Cepacin A was absent in the *B. ambifaria* BCC0191::*ccnJ* cepacin insertional mutant (Figure 4c), confirming that the uncharacterised 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% and 56% presence in clade 3 and clade 2 strains, respectively (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[2]. However, the metabolites and/or BGCs 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[2]. 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 plant survival rates by more than 60% dependent on *B. ambifaria* seed coat inoculum level (10^5, 10^6 and 10^7 cfu/seed; Figure 5a). No biological control was observed when 10^5 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 replicon is not essential and c3 deletion mutants lose virulence and antifungal phenotypes[12]. The cepacin BGC is located on the second c2 replicon of *B. ambifaria* and its biosynthesis was maintained when a third replicon deletion mutant, BCC0191Δc3, was constructed. Despite the loss of >1 Mb of DNA, the BCC0191Δc3 derivative 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*), c3 knockout mutant (Δc3), and combined cepacin-c3 mutation (::*ccnJ*Δc3) were tested further to understand the wider effect of these mutations on strain fitness (see Supplementary Notes). Antimicrobial activity against Gram-positive bacteria and *Pythium* was lost in the cepacin deficient mutant (Supplementary Figure 6). The BCC0191 c3 mutant lost antifungal activity but had a 2-fold increase in cepacin production enhancing its anti-Gram-positive antagonism (Supplementary Figure 6). Rhizocompetence was similar for the BCC0191 WT and BCC0191::*ccnJ,* but

the third replicon deletion mutant colonised the pea rhizosphere at a significantly lower rate of 8.5 x 10^6 cfu/g root ($p = 0.027$; Supplementary Table 7).

A lack of understanding of safety and human pathogenicity were key reasons the US EPA placed a moratorium on *B. cepacia* complex biopesticides[2]. 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[12], we assessed the pathogenicity of *B. ambifaria* BCC0191 and its c3 deletion mutant. In the *Galleria mellonella* wax-moth larvae model[12], the deletion of the third replicon did not attenuate the virulence (Supplementary Figure 7a), 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[19,20], the persistence of *B. ambifaria* BCC0191 was low and loss of the third replicon BCC0191Δc3 further reduced persistence in the lung (Supplementary Figure 7b and Supplementary Figure 7c). At an infective dose of 2 x 10^6 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 (Supplementary Figure 8). 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 (Supplementary Figure 8).

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 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 BGCs, efficacy in targeting key plant pathogens with specific antimicrobial metabolites, and defining the population biology of historically applied *B. ambifaria* biopesticides (see Supplementary Discussion). 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[12], we have highlighted this as an attenuation strategy for developing potentially safe biopesticide strains which retain biotechnological efficacy.

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[3] and bactobolin[5] BGCs, LuxR regulation of other specialised metabolite BGCs has been further described within and outside the genus. *Burkholderia thailandensis* synthesises the quorum sensing regulated cytotoxic compound malleilactone[21], and pyocyanin production in *Pseudomonas aeruginosa* is controlled by a hierarchical QS network[22]. 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 A BGC.

Cepacins A and B were initially described as metabolites of *Burkholderia cepacia*, formally *Pseudomonas cepacia*[16,23], with the original producer strain now classified as *B. diffusa*[24]. Both polyyne metabolites displayed strong anti-*Staphylococci* activity, while cepacin A showed weak anti-Gram-negative activity[16]. 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 (Supplementary Figure 9). The *C. fungivorans* cluster K has been linked to the biosynthesis of the anti-fungal polyyne collimomycin[25] whose BGC organisation (Supplementary Figure 9) and chemical formula ($C_{16}H_{18}O_4$) resemble cepacin A ($C_{16}H_{14}O_4$). Recent characterisation of this *C. fungivorans* strain showed it produces a range of polyynes, collectively designated the collimonins[26], with collimonin A showing the most structural similarity to cepacin A. Several key differences in gene content were identified between the cepacin and collimonin BGCs in the regions flanking the core biosynthetic genes (Supplementary Figure 9). *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* (Supplementary Figure 9). Similar proteins to those identified in the cepacin BGC are listed in Supplementary Table 8.

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* [27] 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[27].

Other studies have highlighted the *in vitro* antimicrobial activity, and presence of the corresponding specialized metabolite BGCs or protective effects in field trials in *Bacillus*[28] and *Streptomyces*[29], 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).

There has been considerable discussion on whether *Burkholderia* species known to cause opportunistic infections can be safely exploited for environmental benefit[30]. Multiple species in the recently defined *Paraburkholderia* genus have not been associated with infection, are generally environmental, and mediate plant-beneficial interactions[30]. 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 alternative means to facilitate their biotechnological exploitation. Third replicon deletion in *B. ambifaria* BCC0191 led to loss of persistence in the murine lung infection model (Supplementary Figure 7b and 7c), and hence provides an unmarked means of attenuating pathogenicity but preserving biopesticidal potential in this strain (see Supplementary Discussion, Figure 5). In addition, the BCC0191Δc3 mutant also showed a reduced root colonisation after 14 days compared to the wild type (see Supplementary Discussion, Supplementary Table 7), 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 a 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[31]. A 2017 survey of *Burkholderia* infections in 361 UK CF patients did not find *B. ambifaria* at all[32]. This epidemiological data combined with the low murine respiratory persistence of *B. ambifaria* (Supplementary Figure 7b and Supplementary Figure 7c) compared to virulent pathogens such as *P. aeruginosa*[19,20], 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.

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 using genome mining and *in vitro* antimicrobial screening to define BGCs that contribute to biocontrol, and enable their use in the rational design of future bacterial biopesticides. The potential of cepacin-producing *B. ambifaria* in protecting economically relevant crop species from attack by oomycete pathogens has been demonstrated. It is clear that *B.*

ambifaria has accumulated multiple plant protective BGCs that underpin its historical exploitation as a biopesticide[2]. 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

Genomes used in this study were either sequenced as part of this study or downloaded from public databases. 125-nucleotide and 150-nucleotide paired-end reads were generated for 60 *B. ambifaria* genomes (Supplementary Table 1) using an Illumina HiSeq 2000 and HiSeq X Ten, respectively. Illumina adaptors were trimmed, read quality assessed and contigs assembled as described in the Supplementary Notes. Genomic contigs were re-arranged and scaffolded into replicons by mapping the contigs against three reference genomes using CONTIGuator v2.7.4[33]. 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[34]. 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[35]. Two sequenced strains from this study (BCC1630 and BCC1638) and one publically available strain (RZ2MS16) were excluded from the dataset, using a 95% ANI species threshold[36]. The remaining 64 *B. ambifaria* strains along with mutant derivatives used in this study are listed in the supplementary data (Supplementary Table 1).

Genome mining and specialized metabolite BGC network analysis

All bioinformatics analyses were performed using the Cloud Infrastructure for Microbial Bioinformatics (CLIMB) computing resource[37]. Scaffolded replicons and non-scaffolded contigs were annotated using Prokka v1.12-beta[38]. Bioinformatics tool antiSMASH v.3.0.5[17] 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+[39]. BGCs were dereplicated by clustering nucleotide sequences using pairwise Kmer-matching software Mash v1.1.1[40]; 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[41], 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.

Genomic analysis and phylogenomics

The core and accessory genome of the collective 64 *B. ambifaria* strains was calculated using Roary v3.7.0[42] 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[43]. The root position was determined by including the outgroup species *Burkholderia vietnamiensis* G4 (PRJNA10696) (Supplementary Figure 10). Once the root branch point was defined, a second tree was constructed using RAxML v8.2.11[44] with General Time Reversible (GTR) substitution and a GAMMA model of rate heterogeneity supported by 100 bootstraps.

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 14 plant, animal and human pathogens, and other reference species (Supplementary Table 5). A standard overlay assay was used to screen for antimicrobial activity, as previously described[3], with the amendment of using 400 µl overnight (O/N) culture of the susceptibility-testing organism per 100 ml half-strength iso-sensitest agar. A strain was considered antagonistic if the zone of inhibition was >10 mm in diameter. Antagonism assays with *A. alternata* and *F. solani* involved the re-suspension of mycelia in 1 ml PBS from 9-day old cultures grown in 50 mm petri dishes on potato dextrose agar; 320 µl of mycelial resuspension was used per 100 ml half-strength iso-sensitest agar before pouring as an overlay. 10x10 cm square Petri dishes containing basal salts medium[15] supplemented with 4 g/l glycerol (BSM-G)[3] 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 (Supplementary Table 5). 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 BGC-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 (Supplementary Table 9), 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[45] 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 PCR. Comparative-HPLC and antimicrobial activity screens between the wild-type and insertional mutant confirmed the disruption of the cepacin BGC.

Construction and phenotypic analysis of a *B. ambifaria* BCC0191Δc3 mutant

The *B. ambifaria* BCC0191Δc3 mutant was constructed following the methods outlined in Agnoli *et al.* 2012[12]. 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[12]. 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) (Supplementary Table 9). The virulence of *B. ambifaria* BCC0191 and its Δc3 mutant was evaluated in *G. mellonella* wax moth larvae[12] and murine chronic lung[19,20] infection models as described below.

Detection of LuxR homologues as specialized metabolite regulators

LuxR-encoding gene homologues were identified by replicating the systematic *in silico* approach previously described[46]. 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[47] and a phylogenetic tree generated using FastTree v2.1.9[43]. 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 *B. diffusa* strain LMG 24093. 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 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* Trow var. *ultimum* (MUCL 16164) was developed from the methodology proposed in Toda *et al.* 2015[48]. 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 hours) 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 10^9 cfu/ml), and either applied at the neat concentration or diluted to achieve the desired inoculum levels (10^8 and 10^7 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. Significant differences between BCC0191 wild-type and mutant derivatives was assessed using two tailed t-test or Welch's two tailed t-test. Two tailed t-test assumptions were normally distributed data (Shapiro-Wilk test) and equal variances (Bartlett test); Welch's two tailed t-test did not assume equal variances.

Metabolite extraction and HPLC analysis

To confirm the absence of cepacin A in the BGC mutants constructed for the different *B. ambifaria* strain backgrounds (BCC0191, BCC1252, BCC1241, BCC0477, BCC1259 and BCC1218; Supplementary Figure 3), each strain was grown on BSM-G agar for three days at 22°C, the bacterial growth removed and an equal amount of agar (a 2 cm disc) extracted with 0.5 ml ethyl acetate for 2 hours at room temperature. This was fractionated on an acetonitrile gradient (5% to 95%) using a Waters® AutoPurification™ HPLC System fitted with a reverse phase analytical column (Waters® XSelect CSH C18, 4.6 x 100 mm, 5 µm). Metabolites eluted from the column were detected by a photodiode array and the peak corresponding to cepacin A identified by its retention time in relation to the LC-MS confirmed presence of the polyyne in this fraction.

In vivo killing assay using *Galleria* wax moth larvae

Galleria mellonella wax moth larvae were sourced from BioSystems Technology Ltd TruLarv (Exeter, UK). Bacterial cultures of BCC0191 wild-type and BCC0191Δc3 were grown overnight (approximately 18 hours) in TSB broth at 37°C, washed and re-suspended in 1x volume PBS before being adjusted to approximately $1x10^6$ cfu/ml. Larvae were injected in the hindmost proleg on the right-side of the abdomen. Each larva was injected with 10 µl aliquots of *B. ambifaria* BCC0191 wild-type and BCC0191Δc3 bacterial suspensions with a 25G x 1" needle (BD Microlance 3) using a 1705TLL, 50 µL syringe (Hamilton). PBS injections were included as controls. Each condition included ten larvae, and the experiment was performed in triplicate over three days. The larvae were incubated at 37°C for 72 hours, and their survival status was monitored at 18, 24, 42, 48, 66 and 72 hours post-inoculation. Larvae were recorded as dead when they failed to respond to physical agitation. The total viable count of the bacterial suspensions was calculated during the first and third replicate via drop-count. The mean average cfu/ml of BCC0191 wild-type and BCC0191Δc3 suspensions were $1.5x10^6$ cfu/ml and $1.0x10^6$ cfu/ml, respectively.

Murine chronic lung infection model. A murine respiratory infection model as used for modelling *P. aeruginosa* pathogenicity was applied to *B. ambifaria* essentially as described[19,20]. All experiments were carried out at the University of Liverpool with approval from the UK Home Office and University ethics committee. Randomisation of mice to cages (experimental groups) was performed by animal unit staff with no role in study design as described[19]. An initial dosing experiment demonstrated good tolerance of an infectious dose of 10^6 *B. ambifaria,* equivalent to that used for *P. aeruginosa* in previous studies[19,20]. Sample size was calculated to give 95% power at alpha 0.05 to detect a >4-fold difference in lung CFU between BCC0191 and BCC0191Δc3, assuming mean CFU of 100 in BCC0191 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, UK) were infected intranasally under anaesthesia (O_2/isolfuorane), achieving an actual dose of 2 x 10^6 *B. ambifaria* BCC0191 or its third chromosome replicon mutant BCC0191Δc3. Symptoms of disease were monitored, and mice culled at 24 hours, 48 hours and 5 days post infection. Researchers were not blinded to the experimental groupings. The nasopharyngeal tissue, lungs and spleens were removed, homogenised in 3 ml PBS, serial dilutions prepared and plated onto *Burkholderia cepacia* selective agar (Oxoid, UK) for enumeration of surviving *B. ambifaria*. No animals were excluded from the analysis. Colonies from the tissue of mice carrying infection were confirmed to be strain *B. ambifaria* BCC0191 by Random Amplified Polymorphic DNA typing (RAPD; see Supplementary Figure 8).

Rhizocompetence of *B. ambifaria* BCC0191 WT and derivatives

The rhizocompetence of *B. ambifaria* BCC0191 and its derived mutants were evaluated essentially as described[49]. Liquid cultures of BCC0191, BCC0191::*ccnJ* and BCC01911Δc3 were grown in TSB broth overnight at 37°C (50 µg/ml trimethoprim for BCC0191::*ccnJ*), then washed and re-suspended in 1x

volume PBS and adjusted to $1x10^9$ cfu/ml. *Pisum sativum* (pea) seeds were coated in the bacterial suspensions and planted in a potting mix composed of 5:1 multi-purpose compost to sand. The seeds were germinated and grown at 22°C with a 16 hour light - 8 hour dark photoperiod (70% RH) for 14 days and watered as required. After 14 days the plants were removed, and excess soil shaken from the root systems. The 1^{st} – 2^{nd} cm of root was excised, macerated in 1 ml PBS, and serially diluted. Dilutions of the root suspension were plated onto *Burkholderia cepacia* selective agar (BCSA; Oxoid UK) and incubated at 37°C for 24 hours to calculate total viable counts. Three plants were used per strain derivative, and three un-inoculated seeds included as a control. Any growth from the control plants was subject to RAPD PCR profiling to distinguish the bacteria from *B. ambifaria*. Non-*B. ambifaria* growth on the control plates was discounted from the total viable counts. Total viable counts were standardised to 1 g fresh weight of root.

Data Availability

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. 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. Bacterial strains and constructs will be made available upon written request to the corresponding authors and after signing a Material Transfer Agreement. We are restricted in re-distributing certain bacterial strains such as those from recognised culture collections, but such requests will be re-directed to the appropriate source.

Code Availability

The publicly available software and codes used for genome sequence determination, phylogenomics, mass spectrometry and general statistical analysis are described in the appropriate Method sections.

Correspondence:

Please address correspondence to: Eshwar Mahenthiralingam; email: MahenthiralingamE@cardiff.ac.uk or Alex J. Mullins, email: MullinsA@cardiff.ac.uk. Requests for materials should be addressed to Eshwar Mahenthiralingam.

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 and MB acknowledge funding support from the Medical Research Council's Cloud Infrastructure for Microbial Bioinformatics (MR/L015080/1) which provided the computational resources to undertake the analyses for this work. 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 the 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, and enacyloxin MIC analysis, 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.

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**		**Cepacins[2]**	
	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)	+	+	+	+	-	-	-	-	-	-	+	+
Other strains:												
BCC1100 (1a)	+	+	+	+	-	-	+	+	-	-	-	-
BCC1105 (1c)	+	+	-	-	-	-	-	-	-	-	-	-
BCC1220 (2)	+	+	-	-	+	+	-	-	-	-	+	-

Footnotes:

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

[2]The metabolite cepacin was detected prior to identification of its biosynthetic gene cluster.

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 3784 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 Figure 10). 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[50]. **(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. Strains MEX-5 and IOP40-10 were not available for the antimicrobial production assay.

Figure 2. Specialized metabolite BGC network analysis of 64 *B. ambifaria* strains. A total of 1,272 BGCs were detected across the 64 strains, and dereplication indicated these represented 38 distinct BGCs (38 distinct network clusters). Nucleotide sequences were clustered using Mash and visualized with Cytoscape. This network analysis was used to provide a visual summary of the breadth of *B. ambifaria* BGCs including their biosynthetic diversity, strain distribution, and core or accessory nature within the species. Each node represents a specialized metabolite BGC extracted from a single *B. ambifaria* strain. 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 Supplementary Figure 9. 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. *n* = 3 biological replicates. 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); *n* = 3 independent LC-MS analyses of WT and mutant cultures. **(d)** Structure of cepacin A, the identity of which was confirmed by comparison to an authentic standard from a known producer (Supplementary Figure 3).

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 10^7, 10^6 and 10^5 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. 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-sided t-test or Welch's two-sided t-test (* = $p < 0.05$; ** = $p < 0.01$); significant difference (left to right) p = 0.002, p = 0.03, p = 0.002 with 95% confidence interval. *n* = 10 seeds per condition (seed coat) and dosage (cfu/seed), repeated in triplicate. Centre bar represents mean, and error bars represent standard error. **(b)** Pea germination (14 days) in *P. ultimum* infested soil observed for groups of 10 seeds coated with 10^7, 10^6 and 10^5 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 (left to right: p = 0.22, p = 0.22, p = 0.16), as determined by two-sided t-test with 95% confidence interval, in plant survival between BCC0191 WT and BCC0191Δc3 at all inoculum levels. *n* = 10 seeds per condition (seed coat) and dosage (cfu/seed), repeated in triplicate. Centre bar represents mean, and error bars represent standard error.

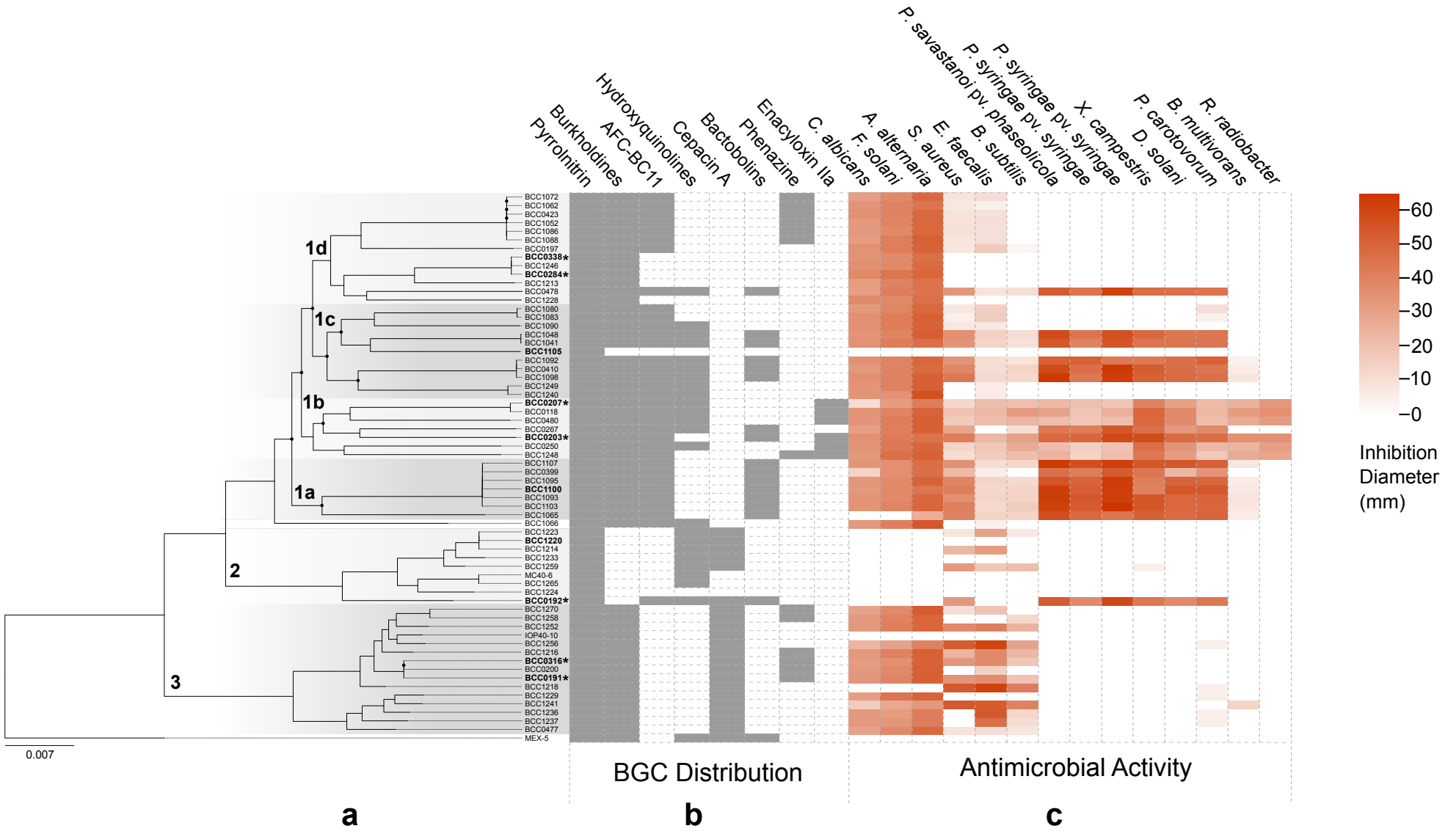
References:

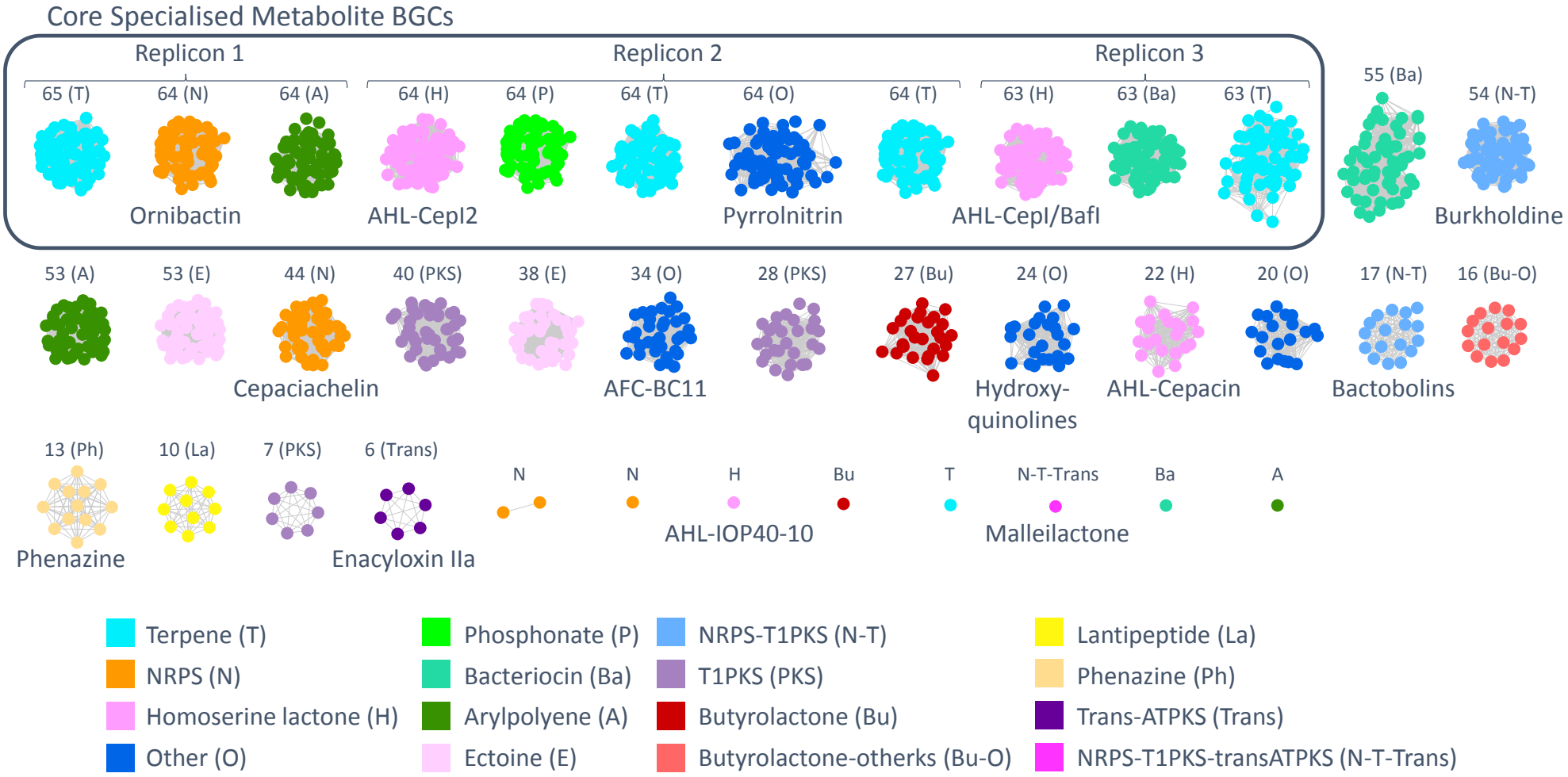
1. Depoorter, E. *et al. Burkholderia*: an update on taxonomy and biotechnological potential as antibiotic producers. *Appl. Microbiol. Biotechnol.* **100,** 5215–5229 (2016).
2. Parke, J. L. & Gurian-Sherman, D. Diversity of the *Burkholderia cepacia* complex and implications for risk assessment of biological control strains. *Annu. Rev. Phytopathol.* **39,** 225–258 (2001).
3. Mahenthiralingam, E. *et al.* Enacyloxins are products of an unusual hybrid modular polyketide synthase encoded by a cryptic *Burkholderia ambifaria* genomic island. *Chem. Biol.* **18,** 665–677 (2011).
4. Masschelein, J., Jenner, M. & Challis, G. L. Antibiotics from Gram-negative bacteria: a comprehensive overview and selected biosynthetic highlights. *Nat. Prod. Rep.* **34,** 712–783 (2017).
5. Seyedsayamdost, M. R. *et al.* Quorum-sensing-regulated bactobolin production by *Burkholderia thailandensis* E264. *Org. Lett.* **12,** 716–719 (2010).
6. Song, L. *et al.* Discovery and biosynthesis of gladiolin: A *Burkholderia gladioli* antibiotic with promising activity against *Mycobacterium tuberculosis*. *J. Am. Chem. Soc.* **139,** jacs.7b03382

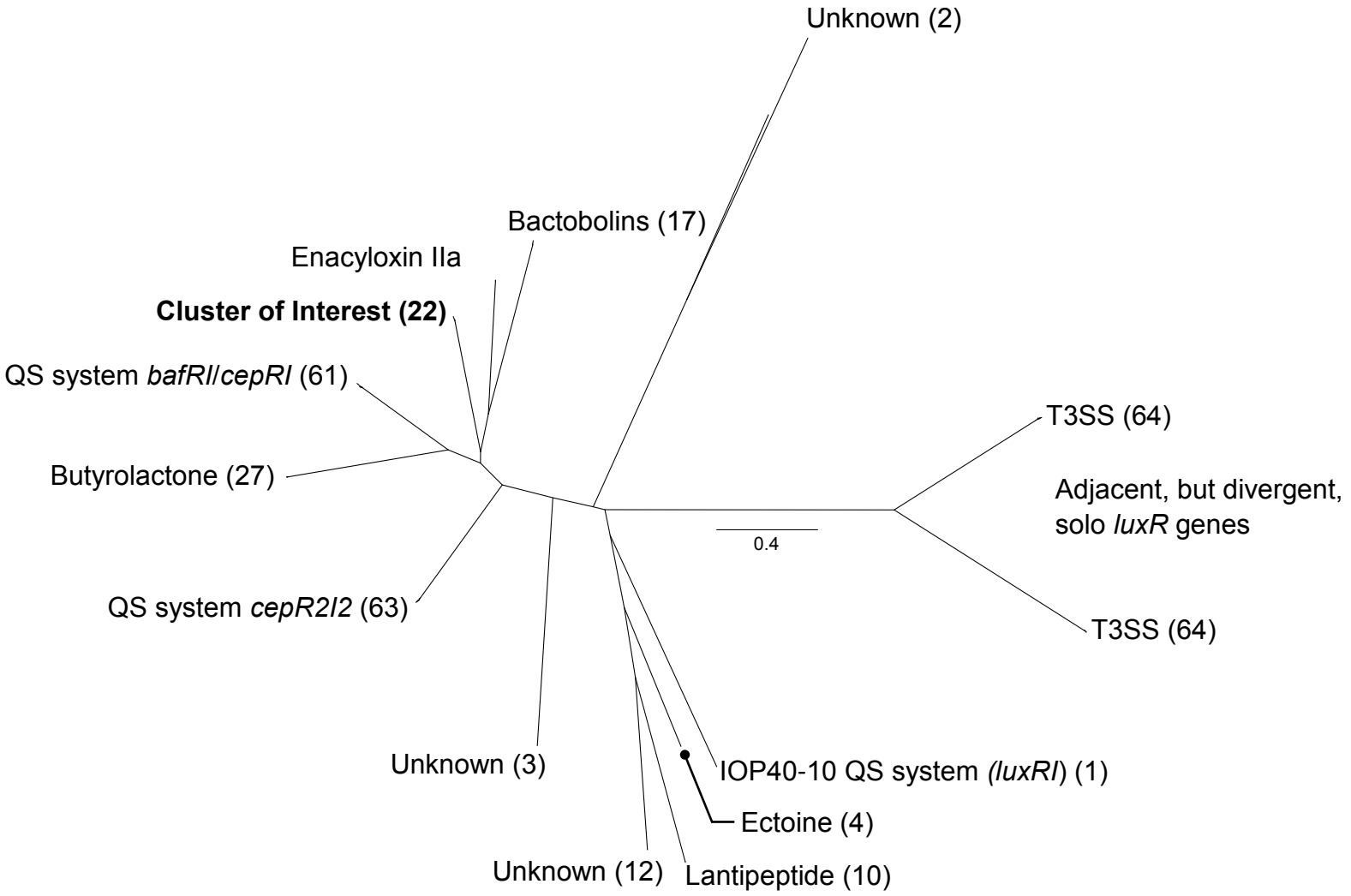
(2017).
7. Flórez, L. V. *et al.* Antibiotic-producing symbionts dynamically transition between plant pathogenicity and insect-defensive mutualism. *Nat. Commun.* **8,** 15172 (2017).
8. Kim, J. *et al.* Quorum sensing and the LysR-type transcriptional activator ToxR regulate toxoflavin biosynthesis and transport in *Burkholderia glumae*. *Mol. Microbiol.* **54,** 921–934 (2004).
9. Ross, C., Scherlach, K., Kloss, F. & Hertweck, C. The molecular basis of conjugated polyyne biosynthesis in phytopathogenic bacteria. *Angew. Chemie - Int. Ed.* **53,** 7794–7798 (2014).
10. Mansfield, J. *et al.* Top 10 plant pathogenic bacteria in molecular plant pathology. *Mol. Plant Pathol.* **13,** 614–629 (2012).
11. Howden, A. J. M., Rico, A., Mentlak, T., Miguet, L. & Preston, G. M. *Pseudomonas syringae* pv. *syringae* B728a hydrolyses indole-3-acetonitrile to the plant hormone indole-3-acetic acid. *Mol. Plant Pathol.* **10,** 857–865 (2009).
12. Agnoli, K. *et al.* Exposing the third chromosome of *Burkholderia cepacia* complex strains as a virulence plasmid. *Mol. Microbiol.* **83,** 362–378 (2012).
13. Schmidt, S. *et al.* Production of the antifungal compound pyrrolnitrin is quorum sensing-regulated in members of the *Burkholderia cepacia* complex. *Environ. Microbiol.* **11,** 1422–1437 (2009).
14. Tawfik, K. A. *et al.* Burkholdines 1097 and 1229, potent antifungal peptides from *Burkholderia ambifaria* 2.2N. *Org. Lett.* **12,** 664–666 (2010).
15. Hareland, W. A., Crawford, R. L., Chapman, P. J. & Dagley, S. Metabolic function and properties of 4-hydroxyphenylacetic acid 1-hydroxylase from *Pseudomonas acidovorans*. *J. Bacteriol.* **121,** 272–85 (1975).
16. Parker, W. L. *et al.* Cepacin A and cepacin B, two new antibiotics produced by *Pseudomonas cepacia*. *J. Antibiot. (Tokyo).* **37,** 431–440 (1984).
17. Weber, T. *et al.* antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* **43,** W237-43 (2015).
18. Ishida, K., Lincke, T., Behnken, S. & Hertweck, C. Induced Biosynthesis of Cryptic Polyketide Metabolites in a *Burkholderia thailandensis* Quorum Sensing Mutant. *J. Am. Chem. Soc.* **132,** 13966–13968 (2010).
19. Fothergill, J. L., Neill, D. R., Loman, N., Winstanley, C. & Kadioglu, A. *Pseudomonas aeruginosa* adaptation in the nasopharyngeal reservoir leads to migration and persistence in the lungs. *Nat. Commun.* **5,** 4780 (2014).
20. Bricio-Moreno, L. *et al.* Evolutionary trade-offs associated with loss of PmrB function in host-adapted *Pseudomonas aeruginosa*. *Nat. Commun.* **9,** 2635 (2018).
21. Duerkop, B. A. *et al.* Quorum-sensing control of antibiotic synthesis in *Burkholderia thailandensis*. *J. Bacteriol.* **191,** 3909–3918 (2009).
22. Lee, J. & Zhang, L. The hierarchy quorum sensing network in *Pseudomonas aeruginosa*. *Protein Cell* **6,** 26–41 (2015).
23. Yabuuchi, E. *et al.* Proposal of *Burkholderia* gen. nov. and transfer of seven species of the genus *Pseudomonas* homology group II to the new genus, with the type species Burkholderia cepacia

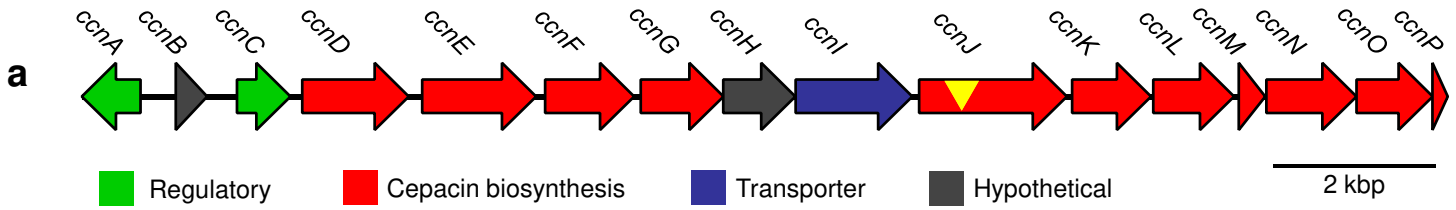
(Palleroni and Holmes 1981) comb. nov. *Microbiol. Immunol.* **36,** 1251–75 (1992).
24. Vanlaere, E. *et al. Burkholderia latens* sp. nov., *Burkholderia diffusa* sp. nov., *Burkholderia arboris* sp. nov., *Burkholderia seminalis* sp. nov. and *Burkholderia metallica* sp. nov., novel species within the *Burkholderia cepacia* comple. *Int. J. Syst. Evol. Microbiol.* **58,** 1580–1590 (2008).
25. Fritsche, K. *et al.* Biosynthetic genes and activity spectrum of antifungal polyynes from *Collimonas fungivorans* Ter331. *Environ. Microbiol.* **16,** 1334–1345 (2014).
26. Kai, K., Sogame, M., Sakurai, F., Nasu, N. & Fujita, M. Collimonins A–D, Unstable Polyynes with Antifungal or Pigmentation Activities from the Fungus-Feeding Bacterium *Collimonas fungivorans* Ter331. *Org. Lett.* **20,** 3536–3540 (2018).
27. Haas, D. & Keel, C. Regulation of antibiotic production in root-colonizing *Pseudomonas* spp. And relevance for biological control of plant disease. *Annu. Rev. Phytopathol.* **41,** 117–153 (2003).
28. Palazzini, J. M., Dunlap, C. A., Bowman, M. J. & Chulze, S. N. *Bacillus velezensis* RC 218 as a biocontrol agent to reduce Fusarium head blight and deoxynivalenol accumulation: Genome sequencing and secondary metabolite cluster profiles. *Microbiol. Res.* **192,** 30–36 (2016).
29. Law, J. W.-F. *et al.* The potential of *Streptomyces* as biocontrol agents against the Rice Blast Fungus, Magnaporthe oryzae (Pyricularia oryzae). *Front. Microbiol.* **8,** 3 (2017).
30. Eberl, L. & Vandamme, P. Members of the genus *Burkholderia*: good and bad guys. *F1000Research* **5,** (2016).
31. LiPuma, J. J. The Changing Microbial Epidemiology in Cystic Fibrosis. *Clin. Microbiol. Rev.* **23,** 299–323 (2010).
32. Kenna, D. T. D. *et al.* Prevalence of *Burkholderia* species, including members of *Burkholderia cepacia* complex, among UK cystic and non-cystic fibrosis patients. *J. Med. Microbiol.* **66,** 490–501 (2017).
33. Galardini, M., Biondi, E. G., Bazzicalupo, M. & Mengoni, A. CONTIGuator: a bacterial genomes finishing tool for structural insights on draft genomes. *Source Code Biol. Med.* **6,** 11 (2011).
34. Li, H. Circlator: automated circularization of genome assemblies using long sequencing reads. **16,** (2013).
35. Pritchard, L. *et al.* Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal. Methods* **8,** 12–24 (2016).
36. Richter, M. & Rossello-Mora, R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci.* **106,** 19126–19131 (2009).
37. Connor, T. R. *et al.* CLIMB (the Cloud Infrastructure for Microbial Bioinformatics): an online resource for the medical microbiology community. *Microb. Genomics* **2,** (2016).
38. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30,** 2068–2069 (2014).
39. Morgulis, A. *et al.* BLAST+: architecture and applications. *Bioinformatics* **24,** 1757–1764 (2008).
40. Ondov, B. D. *et al.* Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* **17,** 132 (2016).
41. Shannon, P. *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13,** 2498–2504 (2003).

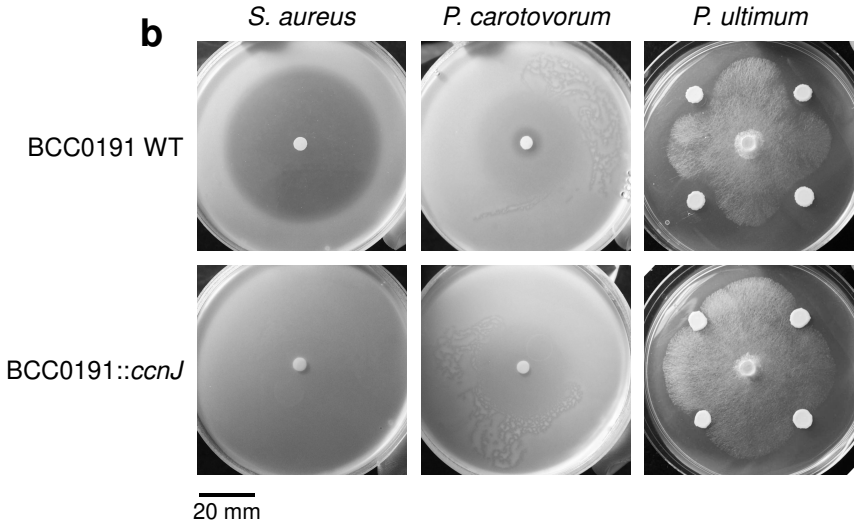
42. Page, A. J. *et al.* Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31,** 3691–3693 (2015).
43. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS One* **5,** e9490 (2010).
44. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30,** 1312–3 (2014).
45. Flannagan, R. S., Aubert, D., Kooi, C., Sokol, P. A. & Valvano, M. A. *Burkholderia cenocepacia* requires a periplasmic HtrA protease for growth under thermal and osmotic stress and for survival *in vivo*. *Infect. Immun.* **75,** 1679–1689 (2007).
46. Gan, H. M. *et al.* 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. *Front. Cell. Infect. Microbiol.* **4,** 188 (2014).
47. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30,** 3059–66 (2002).
48. Toda, T., Iwasa, A., Fuji, S. & Furuya, H. Widespread occurrence of *Pythium arrhenomanes* pathogenic to rice seedlings around Japanese rice fields. *Plant Dis.* **99,** 1823–1831 (2015).
49. Vidal-Quist, J. C. *et al. Arabidopsis thaliana* and *Pisum sativum* models demonstrate that root colonization is an intrinsic trait of *Burkholderia cepacia* complex bacteria. *Microbiology* **160,** 373–384 (2014).
50. Hadfield, J. *et al.* Phandango: an interactive viewer for bacterial population genomics. *Bioinformatics* **34,** 292–293 (2018).

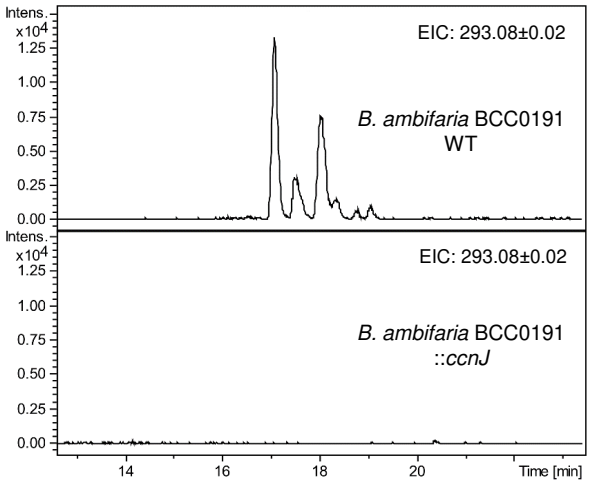

Burkholdines
Pyrrolnitrin
AFC-BC11
Hydroxyquinolines
Cepacin A
Bactobolins
Phenazine
Enacyloxin IIa
C. albicans
A. alternaria
F. solani
S. aureus
E. faecalis
B. subtilis
P. savastanoi pv. phaseolicola
P. syringae pv. syringae
P. syringae pv. syringae
X. campestris
P. carotovorum
D. solani
B. multivorans
R. radiobacter
1d
1c
1b
1a
2
3
BCC1072
BCC1062
BCC0423
BCC1052
BCC1086
BCC1088
BCC0197
BCC0338*
BCC1246
BCC0284*
BCC1213
BCC0478
BCC1228
BCC1080
BCC1083
BCC1090
BCC1048
BCC1041
BCC1105
BCC1092
BCC0410
BCC1098
BCC1249
BCC1240
BCC0207*
BCC0118
BCC0480
BCC0267
BCC0203*
BCC0250
BCC1248
BCC1107
BCC0399
BCC1095
BCC1100
BCC1093
BCC1103
BCC1065
BCC1066
BCC1223
BCC1220
BCC1214
BCC1233
BCC1259
MC40-6
BCC1265
BCC1224
BCC0192*
BCC1270
BCC1258
BCC1252
IOP40-10
BCC1256
BCC1216
BCC0316*
BCC0200
BCC0191*
BCC1218
BCC1229
BCC1241
BCC1236
BCC1237
BCC0477
MEX-5
0.007
BGC Distribution
Antimicrobial Activity
60
50
40
30
20
10
0
Inhibition Diameter (mm)
a
b
c

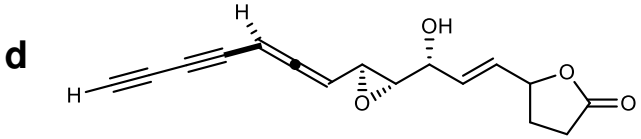

Core Specialised Metabolite BGCs
Replicon 1
Replicon 2
Replicon 3
65 (T)
64 (N)
64 (A)
64 (H)
64 (P)
64 (T)
64 (O)
64 (T)
63 (H)
63 (Ba)
63 (T)
55 (Ba)
54 (N-T)
Ornibactin
AHL-CepI2
Pyrrolnitrin
AHL-CepI/BafI
Burkholdine
53 (A)
53 (E)
44 (N)
40 (PKS)
38 (E)
34 (O)
28 (PKS)
27 (Bu)
24 (O)
22 (H)
20 (O)
17 (N-T)
16 (Bu-O)
Cepaciachelin
AFC-BC11
Hydroxy-quinolines
AHL-Cepacin
Bactobolins
13 (Ph)
10 (La)
7 (PKS)
6 (Trans)
N
N
H
Bu
T
N-T-Trans
Ba
A
Phenazine
Enacyloxin IIa
AHL-IOP40-10
Malleilactone
Terpene (T)
NRPS (N)
Homoserine lactone (H)
Other (O)
Phosphonate (P)
Bacteriocin (Ba)
Arylpolyene (A)
Ectoine (E)
NRPS-T1PKS (N-T)
T1PKS (PKS)
Butyrolactone (Bu)
Butyrolactone-otherks (Bu-O)
Lantipeptide (La)
Phenazine (Ph)
Trans-ATPKS (Trans)
NRPS-T1PKS-transATPKS (N-T-Trans)


Unknown (2)
Bactobolins (17)
Enacyloxin IIa
Cluster of Interest (22)
QS system *bafRI*/*cepRI* (61)
Butyrolactone (27)
QS system *cepR2I2* (63)
T3SS (64)
Adjacent, but divergent, solo *luxR* genes
0.4
T3SS (64)
Unknown (3)
IOP40-10 QS system *(luxRI*) (1)
Ectoine (4)
Unknown (12)
Lantipeptide (10)









Cepacin A

Chemical Formula: $C_{16}H_{14}O_4$

Exact Mass: 270.0892

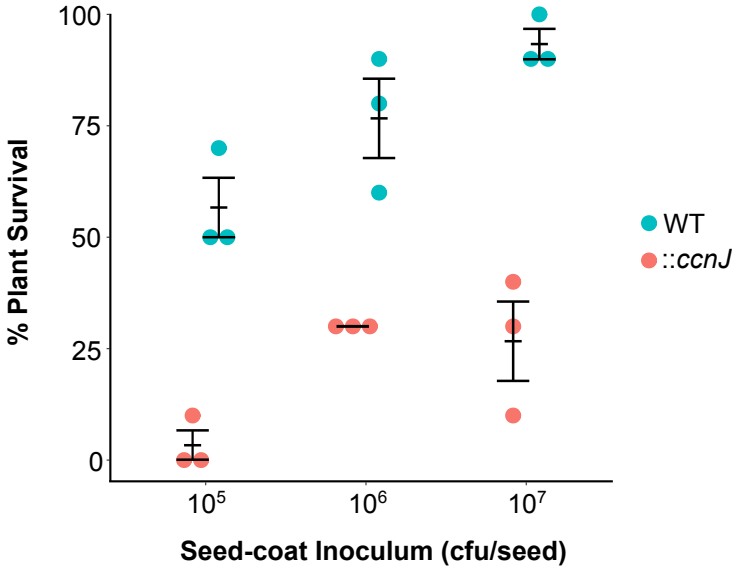
$[M+Na]^+$: 293.0784

a) ***B. ambifaria*** **WT vs ::*ccnJ***

BCC0191 WT

BCC0191::*ccnJ*

% Survival of *P. sativum* plants



b) ***B. ambifaria*** **WT vs Δc3**

BCC0191 WT

BCC0191Δc3

% Survival of *P. sativum* plants