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

Alex J. Mullins1,*, James A. H. Murray1, Matthew J. Bull1, Matthew Jenner2, Cerith Jones3,1,6, Gordon Webster1, Angharad E. Green3, Daniel R. Neill3, Thomas R. Connor1, Julian Parkhill4, Gregory L. Challis2,5 and Eshwar Mahenthiralingam1,6*

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 the soil after deletion of its third replicon, a non-essential plasmid associated with virulence in *Burkholderia 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.

Numerous bacterial and fungal species have been recognized for their biological control abilities and plant growth-enhancing properties. Man-made pesticides conventionally used in agriculture are under increasing scrutiny regarding their bioaccumulation and toxicity, which includes their fatal effect 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 microorganisms 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 antibacterial, antinematodal and antifungal compounds. 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. Thus, in 1999, the US Environmental Protection Agency placed a moratorium on new registrations of *Burkholderia* biopesticides unless such agents were defined as safe in terms of their risk of opportunistic infection.

Multiple species within the *Burkholderia cepacia* complex group were characterized or used as biological control agents. They are highly active in their specialized metabolism, for example, producing antifungal compounds including pyrrolnitrin, occidiofungin, cepac- fungin and burkholdines; antibacterial bacterolins and encaplaxin Ila; and broader-spectrum agents such as the cepacins and Outside of the *B. cepacia* complex, other *Burkholderia* species also produce a range of antagonistic compounds. The bacterolins are also produced by *Burkholderia thailandensis* and *Burkholderia pseudomallei* and exhibit potent activity against Gram-positive bacteria, including methicillin-resistant *Staphylococcus aureus*. *Burkholderia gladioli* also produces multiple antimicrobials, including the antimycobacterial macrolide gladiolin, its isomer lagriene, the cytotoxic azapteridine toxoflavin and the polyyne caryoynencin. There is no consensus on the distribution or co-occurrence of antimicrobial specialized metabolite-encoding gene clusters in biopesticidal *Burkholderia ambifaria*, nor a holistic understanding of strain bioactivity and antimicrobial compound efficacy against priority bacterial, fungal and oomycetal plant pathogens. Biopesticidal activities have been descriptively characterized for individual producer strains against a panel of target organisms, or multiple producer strains against relatively few target organisms. Previous studies have analysed the overall antagonistic properties of *B. ambifaria*, but none has 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 characterized...
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. Collectively, it also represented strains recovered from various geographical origins (the United States, Australia and Italy; Supplementary Table 1). The \textit{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. As curing of the \textit{B. cepacia} complex third genomic replicon (\textit{c3}) is possible in these multireplicon bacteria\cite{12}, a \textit{B. ambifaria} \textit{c3} mutant was constructed and shown to have reduced virulence in a murine respiratory infection model, yet retained its genome-encoded potential of \textit{B. ambifaria} as a biopesticide, phylogenomic and pan-genomic analyses were applied (see Methods, Supplementary Fig. 1, Supplementary Table 2 and Supplementary Notes). The three replicon genomic structure was present in 63 of the \textit{B. ambifaria} strains analysed, whereas 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 Mb (BCC1105) to 8.03 Mb (BCC1248), with a mean of 7.34 Mb (Supplementary Table 2). Assembled replicons c1, c2 and c3 possessed a mean of 3.47 Mb, 2.74 Mb and 1.15 Mb, 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 \textit{B. ambifaria} pan-genome was identified (22,376 distinct genes), of which 3,784 genes comprised the core genome. The pan-genome represented a collection of genes approximately 3.4-fold greater than the mean \textit{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 \textit{B. ambifaria} strains. Exclusion of the strain BCC1105, which lacked the third replicon, from the core genome analysis resulted in the \textit{B. ambifaria} core genome increasing from 3,784 to 4,166 genes. Three major clades were identified in the \textit{B. ambifaria} 3,784 core gene phylogeny (Fig. 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 \textit{B. ambifaria} genomes (see Methods, Supplementary Table 3 and Supplementary Notes) detected 1,272 specialized metabolite biosynthetic gene clusters (BGCs), that were dereplicated into 38 distinct BGCs after Kmer-matching and gene topology comparisons (Fig. 2 and Supplementary Table 3). Network analysis was used to graphically summarize multiple attributes of the \textit{B. ambifaria} BGCs, including their biosynthetic diversity, strain distribution and core or accessory nature within the species (Fig. 2).

Of the 38 distinct BGCs, 13 were previously characterized and 7 were known to encode compounds with antimicrobial activity (Supplementary Table 4). Pyrrolnitrin\(^a\) was the only BGC for an antimicrobial metabolite found in all 64 \textit{B. ambifaria} strains, whereas the BGC for the anti-Gram-negative metabolite enacyloxin II\(^a\) was the least common known antimicrobial BGC (Fig. 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 \textit{B. ambifaria} core gene phylogeny (Fig. 1). Six of the seven clade 1b strains encoded the pathway responsible for enacyloxin II\(^a\) biosynthesis\(^1\). The more widely distributed burkholdine\(^a\) BGC was absent from all members of clade 2 and strain BCC1105, but all other strains possessed this antifungal biosynthetic locus. Bactobolin\(^b\) BGCs were concentrated in clade 1 and were less frequently encountered in clades 2 and 3. Two strains, BCC1105 and BCC1224, only encoded the core antifungal metabolite pyrrolnitrin and lacked any additional antimicrobial BGCs (Fig. 1b). No single strain encoded all seven previously known antimicrobial BGCs; however, approximately 59% of strains encoded four or more BGCs reflecting the known antimicrobial properties of \textit{B. ambifaria} (Fig. 1b).

The silent nature of certain antimicrobial BGCs that are not expressed in standard laboratory cultures, including those in \textit{Burkholderia}, is well established\(^6\). Thus, we correlated in vitro metabolite production with BGC distribution. Ten \textit{B. ambifaria} strains representing the seven characterized biocontrol strains and three additional strains from the broader species phylogeny (Fig. 1a) were screened for metabolite production on agar growth media BSM-G (basal salts medium supplemented with 4 g l\(^{-1}\) glycerol\(^4\)). Six known antimicrobial metabolites were detected by liquid chromatography–mass spectrometry (LC–MS) (Supplementary Figs. 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 containing pyrrolnitrin, burkholdine and hydroxyquinoline BGCs were exceptions to this trend (Table 1). A sixth known metabolite, cepacin \(^{16}\), was also detected in \textit{B. ambifaria} J82 (BCC0191) by LC-MS analyses (see Methods) and subsequently correlated to a BGC (not recognized by anti-SMASH \(^v3\))\(^{17}\) identified by searching for quorum sensing-regulated gene clusters (see below).

**Mapping direct antimicrobial activity against plant and animal pathogens.** Having established the presence of BGCs (Figs. 1b and 2) and corresponding metabolites (Table 1), antagonism activity of the 64 \textit{B. ambifaria} strains against priority plant\(^10\) and human pathogens (Supplementary Table 5) was evaluated as described\(^1\). The in vitro bioactivity was aligned against the core gene phylogeny to map antagonism across \textit{B. ambifaria} as a species (Fig. 1c). A total of six strains lacked observable antimicrobial activity (Fig. 1c). Clade 1a, 1b and 1c strains exhibited substantial bioactivity against Gram-negative pathogens, whereas only two strains outside these clades exhibited similar activity (Fig. 1c). Clade 1b strains exhibited additional strong antagonistic activity towards Betaproteobacteria, \textit{Burkholderia multivorans}, and Alphaproteobacteria, \textit{Rhizobium radiobacter} (Fig. 1c), an activity that was not observed in other anti-Gram-negative \textit{B. ambifaria} strains. The extended antimicrobial antagonism of clade 1b \textit{B. ambifaria} correlated to the presence of the hybrid \textit{cis}-acyltransferase (AT) PKS/\textit{trans-AT} PKS nonribosomal peptide synthase (NRPS) BGC for enacyloxin I\(^a\) (Fig. 1b); all screened Gram-negative pathogens were susceptible to purified enacyloxin I\(^a\) with minimum inhibitory concentrations ranging from 3.2 to 50 µg ml\(^{-1}\) (Supplementary Table 6). The additional anti-Gram-negative activity correlated to the presence of the hybrid NRPS-PKS-encoding BGC for bactobolin (Fig. 1b). Antifungal (against \textit{Candida albicans}, \textit{Fusarium solani} and \textit{Alternaria alternata}) and anti-Gram-positive (against \textit{S. aureus}, \textit{Enterococcus faecalis} and \textit{Bacillus subtilis}) activity was more widespread than anti-Gram-negative activity in \textit{B. ambifaria}, with 82%...
Cepacin A is a key mediator of *B. ambifaria* biocontrol of *P. 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. However, the metabolites and/or BGCs that 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. 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 (Fig. 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%, depending on the *P. ambifaria* seed coat inoculum level (10^5, 10^6 and 10^7 colony-forming units (c.f.u.) per seed; Fig. 5a). No biological control was observed when 10^5 c.f.u. per seed of BCC0191 cepacin mutant was applied (<10% survival), compared to >50% protection mediated by the wild type (WT) at this level (Fig. 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. 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 >1Mb 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 WT (Fig. 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 (::ccnJΔc3) 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 Fig. 6). The BCC0191 c3 mutant lost antifungal activity but had a twofold increase in cepacin production, enhancing its anti-Gram-positive antagonism (Supplementary Fig. 6). Rhizocompetence was
similar for the BCC0191 WT and the BCC0191::cenl strains, but the third replicon deletion mutant colonized the pea rhizosphere at a significantly lower rate of $8.5 \times 10^4$ c.f.u. per g of root ($P=0.027$; Supplementary Table 7).

A lack of understanding of safety and human pathogenicity were key reasons that the US Environmental Protection Agency placed a moratorium on B. cepacia complex biopesticides. As the BCC0191Δc3 mutant had retained its biopesticidal activity (Fig. 5b), yet loss of this replicon is associated with reduced virulence in multiple infection models, we assessed the pathogenicity of B. ambifaria BCC0191 and its c3 deletion mutant. In the Galleria mellonella wax moth larvae model, the deletion of the third replicon did not attenuate the virulence (Supplementary Fig. 7a), showing that genes encoding significant insecticidal pathogenicity were not encoded on c3 in B. ambifaria strain BCC0191. By contrast, using a murine respiratory infection model relevant to chronic cystic fibrosis lung infections, the persistence of B. ambifaria BCC0191 was low and loss of the third replicon in the BCC0191Δc3 mutant further reduced persistence in the lung (Supplementary Fig. 7b,c). At an infective dose of $2 \times 10^4$ bacteria, the BCC0191 WT strain persisted in the nasopharynx for the duration of the 5-d experiment but was cleared from the lungs of 4 out of 6 mice by day 5. By contrast, the c3 mutant was rapidly cleared from both the nasopharynx and the lungs of mice (Supplementary Fig. 8). Low numbers of the parental BCC0191 strain (<50 colonies) were detected in the lungs of mice after 5d of infection, but BCC0191Δc3 was cleared within 48 h. B. ambifaria (WT 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 number of colonies recovered from the mouse infection model were either the administered B. ambifaria BCC0191 or BCC0191Δc3 strains (Supplementary Fig. 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 A, to the 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 antimicrobials 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*. As effective biological control of *Pythium* also occurs in the absence of the third replicon, which has been characterized as a *Burkholderia* virulence plasmid, we have highlighted this as an attenuation strategy for developing potentially safe biopesticide strains that retain biotechnological efficacy.

**Discovery of cepacin BGC.** 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 uncharacterized specialized metabolite BGCs. In addition to the *B. ambifaria* encoded enacylocin and bacterolin BGCs, LuxR regulation of other specialized metabolite BGCs has been further described within and outside the genus. *B. thailandensis* synthesizes the quorum sensing-regulated cytotoxic compound malleilactone, and pyocyanin production in *Pseudomonas aeruginosa* is controlled by a hierarchical quorum sensing network. 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 *B. cepacia*, formally *Pseudomonas cepacia*, with the original producer strain now classified as *B. diffusa*. Both polyene metabolites displayed strong anti-staphylococci activity, whereas cepacin A showed weak anti-Gram-negative activity. Cluster K is a gene cluster with 76.9% homologous nucleotide similarity spanning 12.9 kb (in addition to 8.4 kb of non-homologous regions) to the *B. ambifaria* cepacin A BGC was identified in *Collimonas fungivorans* Ter331 using nucleotide BLAST (Supplementary Fig. 9). The *C. fungivorans* cluster K has been linked to the biosynthesis of the antifungal polyene collimycin, whose BGC organization (Supplementary Fig. 9) and chemical formula ($C_{69}H_{127}O_{16}$) resemble cepacin A ($C_{34}H_{64}O_{4}$). Recent characterization of this *C. fungivorans* strain showed that it produces a range of polyynes, collectively designated the collimonins, with collimoin A showing the most structural similarity to cepacin A. Several key differences in gene content were identified between the cepacin and collimoin BGCs in the regions flanking the core biosynthetic genes (Supplementary Fig. 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 collimoin BGC variants is the substitution of the associated regulatory genes that comprise a quorum sensing-associated luxIR in *Burkholderia* and a lysR regulator in *Collimonas* (Supplementary Fig. 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 for pathogens of the major crop family Fabaceae (*P. sativum*). Disruption of the cepacin BGC in *B. ambifaria*
establishes the role of cepacin A as the major bioactive component of impact of distinct metabolites in a biocontrol system. This study have been evidenced as important metabolites in the biological con-
antifungal properties of pyrrolnitrin and 2,4-diacetylphloroglucinol
route to future safe usage. Attenuation of pathogenicity in biopesti-
to the WT
B.
B.
B.
B.

exploited for environmental benefit 30. Multiple species in the species known to cause opportunistic infections can be safely
to the development of safe 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

significantly reduced plant survival beyond germination and emergence compared to the WT (Fig. 5a; P < 0.05). The contribution of specific metabolites to biocontrol has been studied extensively in Pseudomonas relative to other characterized biocontrol genera. The antifungal properties of pyrrolnitrin and 2,4-diacylphloroglucinol 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. Other studies have highlighted the in vitro antimicrobial activity and the presence of the corresponding specialized metabolite BGCs or protective effects in field trials in Bacillus and Streptomyces, 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 armory 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 WT B. ambifaria also indirectly confirms the expression of the cepacin A BGC in planta (Fig. 5b).

There has been considerable discussion on whether Burkholderia species known to cause opportunistic infections can be safely exploited for environmental benefit. Multiple species in the recently defined Paraburkholderia genus have not been associated with infection, are generally environmental and mediate plant-beneficial interactions. 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 biopesti-
cial routes to facilitate their biotechnological exploitation. Third replicon deletion in B. ambifaria BCC0191 led to the loss of persistence in the murine lung infection model (Supplementary Fig. 7b,c), and hence provides an unmarked means of attenuating pathogenicity but preserving biopesticidal potential in this strain (see Supplementary Discussion and Fig. 5). In addition, the BCC0191Δc3 mutant also showed a reduced root colonization after 14d compared to the WT strain (see Supplementary Discussion and Supplementary Table 7), suggesting that 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 cystic fibrosis lung infections, with a survey of US patients from 1997 to 2007 implicating that it collectively, with several other B. cepacia complex species, causes <3% of all Burkholderia cases. A 2017 survey of Burkholderia infections in 361 UK patients with cystic fibrosis did not find B. ambifaria at all. These epidemiological data combined with the low murine respiratory persistence of B. ambifaria (Supplementary Fig. 7b,c) compared to virulent pathogens, such as P. aeruginosa, indicate 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.
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 \textit{B. ambifaria} in protecting economically relevant crop species from attack by oomycete pathogens has been demonstrated. It is clear that \textit{B. ambifaria} has accumulated multiple plant-protective BGCs that underpin its historical exploitation as a biopesticide\textsuperscript{2}. With an urgent need to sustain crop protection and agricultural production, yet reduce the use of environmentally persistent synthetic pesticides, systematically repurposing natural biological control agents, such as \textit{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. For 60 \textit{B. ambifaria} genomes (Supplementary Table 1), 125-nucleotide and 150-nucleotide paired-end reads were generated using Illumina HiSeq 2000 and HiSeq X Ten, respectively. Illumina adaptors were trimmed, read quality was assessed and contigs were assembled as described in the Supplementary Notes. Genomic contigs were rearranged and scaffolded into replicons by mapping the contigs against three reference genomes using CONTIGuator v2.7.4 (ref. \textsuperscript{19}). The option to fill gaps using strings of ‘N’ was disabled. Reference genomes were \textit{B. ambifaria} AMMD (SAMN02598309) and \textit{B. ambifaria} MC40-6 (SAMN02598385), both obtained from the European Nucleotide Archive; the third reference, \textit{B. ambifaria} BCC2083, 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 recircularized based on the genes dnaA, parA and parH, respectively, using the software Circlator v1.2.1 (ref. \textsuperscript{20}). The species validity of the \textit{B. ambifaria} data set was defined by calculating the average nucleotide identity shared between all available \textit{B. ambifaria} genomes using PyANI v0.2.1 (ref. \textsuperscript{21}). Two sequenced strains from this study (BCC1630 and BCC1638) and one publicly available strain (RZ2MS16) were excluded from the data set, using a 95% average nucleotide identity species threshold\textsuperscript{22}. The remaining 64 \textit{B. ambifaria} strains along with mutant derivatives used in this study are listed in 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\textsuperscript{23}. Scaffolds replicons and non-scaffolded contigs were annotated using Prokka v1.12-beta\textsuperscript{24}. The bioinformatics tool antiSMASH v3.0.5 (ref. \textsuperscript{25}) 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+\textsuperscript{26}. BGCs were dereplicated by clustering nucleotide sequences using pairwise
Kmer-matching software Mash v1.11 (ref. 20), 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 (ref. 21), applying the Jaccard index, P value and Mash distance (heat map) as similarity metrics. 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 (ref. 22) using a 95% minimum percentage identity for blastp and a 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 (ref. 23). The root position was determined by including the outgroup species Burkholderia vietnamiensis G4 (PRNA10696) (Supplementary Fig. 10). Once the root branch point was defined, a second tree was constructed using RAxML v8.2.11 (ref. 24) with the general time reversible 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 TSB at 37°C and aerated overnight, unless stated otherwise. The 64 B. ambifaria strains were screened for the production of antimicrobials with antagonistic activity against 14 plant, animal and human pathogens, and other antagonistic activity against 14 plant, animal and human pathogens, and other antagonistic activity against 14 plant, animal and human pathogens, and other...
Articles Nature Microbiology

2. Parke, J. L. & Gurian-Sherman, D. Diversity of the pathovar Burkholderia cepacia. Nature Microbiology, 2, 716–721 (2017).

3. Sowers, S. & Quarantil, M. R. et al. Quorum-sensing-regulated bacitracin production by Burkholderia thailandensis E264. J. Bacteriol., 191, 7959–7972 (2009).

4. Song, L. et al. Discovery and biosynthesis of gladiolins: a Burkholderia gladioli strain with a promising activity against Mycobacterium tuberculosis. J. Am. Chem. Soc., 139, 7974–7981 (2017).

5. Flórez, L. V. et al. Antibiotic-producing symbionts dynamically transition between plant pathogenicity and insect-defensive mutualism. Nat. Commun., 8, 15172 (2017).

6. Kim, J. et al. Quorum sensing and the LysR-type transcriptional activator ToX1 regulate toloxacin biosynthesis and transport in Burkholderia glauca. Mol. Microbiol., 54, 921–934 (2004).

7. Ross, C. et al. Scherlach, K., Kloss, F. & Hetreck, C. The molecular basis of conjugated polyyne biosynthesis in phytopathogenic bacteria. Angew. Chem. Int. Ed. Engl., 53, 7794–7798 (2014).

8. Mansfield, J. et al. Top 10 plant pathogenic bacteria in molecular plant pathology. Mol. Plant Pathol., 13, 614–629 (2012).

9. Howden, A. J. M., Rico, A., Mentlak, T., Miquel, L. & Preston, G. M. Burkholderia cepacia pv. syringae B728a hydrolyses indole-3-acetonitrile to the plant hormone indole-3-acetic acid. Mol. Plant Pathol., 10, 857–865 (2009).

10. Agnoli, K. et al. Exposing the third chromosome of Burkholderia cepacia complex strains as a virulence plasmid. Mol. Microbiol., 83, 362–378 (2012).

11. 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).

12. Tawfik, K. A. et al. Burkholderia 1097 and 1229, potent antifungal peptides from Burkholderia ambifaria. Mol. Microbiol., 12, 664–666 (2010).

13. Harel, J. W. F. et al. Cepacin A and cepacin B, two new antibiotics produced by Pseudomonas cepacia. J. Antimicrob. Chemother., 43, W23–W24 (2015).

14. Ishida, K. et al. Cepacin and cepacin B, two new antibiotics produced by Pseudomonas cepacia. J. Antimicrob. Chemother., 53, 26–41 (2015).

15. Fritsche, K. et al. Production of quorum-sensing-regulated bactobolin in the virulence plasmid of Burkholderia cepacia. J. Bacteriol., 191, 7390–7398 (2009).

16. Lee, J. & Zhang, L. The quorum sensing network in Pseudomonas aeruginosa. Protein Cell, 6, 26–41 (2015).

17. Yabuchi, E. et al. Burkholderia cepacia gen. nov. and transfer of seven species of the genus Pseudomonas homology group II to the new genus, with the type species Burkholderia cepacia (Fallonri and Holmes 1981) comb. nov. J. Bacteriol., 166, 1251–1257 (1992).

18. Vanlaere, E. et al. Burkholderia latens sp. nov., Burkholderia diversa sp. nov., Burkholderia avigata sp. nov., Burkholderia variabilis sp. nov., and Burkholderia aggregata sp. nov., novel species within the Burkholderia cepacia complex. Int. J. Syst. Evol. Microbiol., 58, 1580–1590 (2008).

19. Pritsche, K. et al. Biosynthetic genes and activity spectrum of antifungal polyynes from Collomnatrium fungivorans Ter331. Environ. Microbiol., 16, 1334–1345 (2014).

20. Kai, K., Sogame, M., Sakurai, F., Nao, N. & Fujita, M. Collomnatrium A-D, unstable polyynes with antifungal or pigmentation activities from the fungus-feeding bacterium Collomnatrium fungivorans Ter331. Org. Lett., 20, 3536–3540 (2018).

21. 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).

22. Palazzini, J. M., Dunlap, C. A., Bowman, M. J. & Chulze, S. N. Bacillus subtilis var. natto, a Bacillus strain able to reduce Fusarium head blight and deoxynivalenol accumulation: genome sequencing and secondary metabolite profile. Microbiol. Res., 192, 30–36 (2016).

23. Law, J. W.-F. et al. The potential of Streptomyces to control rice blast fungus, Magnaporthe oryzae (Pyricularia oryzae). Front. Microbiol., 3, 7 (2017).

24. Eberl, L. & Vandamme, P. Members of the genus Burkholderia: good and bad guys. F1000Res., 5, 1007 (2016).

25. LiPuma, J. J. The changing microbial epidemiology in cystic fibrosis. Clin. Microbiol. Rev., 23, 399–423 (2010).
Articles

32. Kenna, D. T. D. J. 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. Galardi, M., Biendi, E. G., Bazzicalupo, M. & Mengoni, A. CONTIGuator: a bacterial genomes finishing tool for structural insights on draft genomes. Source Code Biol. Med. 11 (2011).

34. Hunt, M. et al. Circulator: automated circularization of genome assemblies using long sequencing reads. Genome Biol. 16, 294 (2015).

35. Richter, M. & Rossello-Mora, R. Shifting the genomic gold standard for the prokaryotic species definition. Proc. Natl Acad. Sci. USA 106, 19126–19131 (2009).

36. Connor, T. R. et al. CLIMB (the Cloud Infrastructure for Microbial Bioinformatics): an online resource for the medical microbiology community. Microb. Genomics 2, e000086 (2016).

37. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 15, 1766–1773 (2005).

38. Kenna, D. T. D. J. 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).

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

40. Ondov, B. D. et al. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol. 17, 132 (2016).

41. Kenna, D. T. D. J. 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).

42. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313 (2014).

43. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 15, 1766–1773 (2005).

44. Ondov, B. D. et al. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol. 17, 132 (2016).

45. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 15, 1766–1773 (2005).

46. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 15, 1766–1773 (2005).

47. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 15, 1766–1773 (2005).

48. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 15, 1766–1773 (2005).

49. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 15, 1766–1773 (2005).

Articles

Acknowledgements

AJM is funded by a Biotechnology and Biological Sciences Research Council (BBSRC) South West doctoral training partnership award (BY1910 7007). E.M., G.L.C., T.R.C. and J.P. acknowledge additional support for genome mining from BBSRC award BB/L021692/1; C.J. and M.J. were funded by this award. M.J. is currently the recipient of a BBSRC Future Leader Fellowship (BB/R01212/1). The Bruker maXis U HUPLC-ESI-Q-TOF-MS system used in this research was funded by the BBSRC (BB/M017982/1). G.W. was supported by awards to E.M. from the Life Sciences Bridging Fund and Wellcome Trust Institutional Strategic Support Fund held at Cardiff University. T.R.C. and M.J.B. 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. D.R.N. and A.E.G. acknowledge funding from a Wellcome Trust and Royal Society Sir Henry Dale Fellowship awarded to D.R.N. (grant number 204457/Z/16/Z). G.L.C. is the recipient of a Wolfson Research Merit Award from the Royal Society (WM130033). We thank L. Eberl and K. Agnoli for provision of the mini-c3 used for the third replicon deletion.

Author contributions

The initial study to characterize the genomes of B. ambifaria as a biopesticide was conceived by E.M., with additional aspects of the study design added by A.J.M., G.L.C. and J.A.H.M. A.J.M. performed all aspects of the study with the exception of the LC-MS profiling, and was assisted by specific contributions from the following: data sets and input for genome sequencing and mining: E.M., G.L.C., J.P. and T.R.C.; genome assembly, phylogenomics, cluster mining and de-replication: M.J.B.; LuxR mining: E.M.; generation of a cepacin insertion mutant and antimicrobial activity screening: C.I.; extraction, identification and fractionation of Burkholderia metabolites by HPLC and enacyloxin minimum inhibitory concentration analysis: G.W.; LC-MS identification and confirmation of B. ambifaria antimicrobial metabolites: M.J. and G.L.C.; biocontrol modelling: E.M., G.W. and J.A.H.M.; evaluation and analysis of plant models: J.A.H.M.; Galleria virulence assays: G.W. and C.I.; and murine infection modelling and analysis: A.E.G. and D.R.N. A.J.M. and E.M. developed the first draft of the manuscript, and all authors read and contributed towards finalization of the study.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41564-019-0383-z.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to A.J.M. or E.M.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a | Confirmed
---|---
☐ | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐ | An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐ | The statistical test(s) used AND whether they are one- or two-sided
☐ | Only common tests should be described solely by name; describe more complex techniques in the Methods section.
☐ | A description of all covariates tested
☐ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐ | A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐ | For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted. Give P values as exact values whenever suitable.
☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐ | Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated
☐ | Clearly defined error bars
☐ | State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection

The Illumina HiSeq software suite (HCS 2.2.68) was used for collection of genome sequence data. The Bruker maXis ll software was used for collection of mass spectrometry data and Waters Mass Lynx software used for HPLC analysis. Microsoft office suite 2016 was used to collect text and numerical data for the manuscript.

Data analysis

Trim Galore v0.4.2 and Cutadapt v1.12 - trim illumina adaptors. FastQC v0.10.1 - assess read quality. FlaSH v1.2.11 - merge overlapping short read pairs. SPAdes v3.9.1 - assembly contigs. Pilon v1.21 - identify misassembled contigs. Kraken v0.10.5-beta - identify potential contaminant sequences. QUAST v4.4 - genome assembly statistics. CONTIGuator v2.7.4 - scaffold contigs to reference. Circlator v1.2.1 - recirculate replicons. PyANI v0.2.1 - calculate average nucleotide identity of genomes. Prokka v1.12-beta - annotate genomes. antiSMASH v.3.0.5 - predict secondary metabolite biosynthetic gene clusters (BGCs). Mash v1.1.1 - de-replicate BGCs. BLAST v2.6.0+ - identify homologous sequences. Cytoscape v3.4.0 - visualise distance matrix generated from Mash. BoxPlotR - generate boxplots of replicon biosynthetic capacity. Roary v3.7.0 - define core and accessory genome of collection. RaxML v8.2.11 - generate core-genome phylogeny. Easyfig - visualise sequence comparisons. R v3.2.3 via RStudio v0.99.484 - Generate heatmap and 2-way ANOVA statistics. HMMER 3.1b2, Prodigal-2.6.3 and interproscan-5.22-61.0 - generate HMM, predict coding sequences and annotate encoding sequences with encoded protein signatures for luxR gene identification. MAFFT v7.305b and FastTree v2.1.9 - align protein sequences and generate phylogeny of LuxR proteins. Bruker Compass Data Analysis 4.1 - Analysis of UHPLC-HR-MS data.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.
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 findings of this study are available from the corresponding author 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.

### Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- [x] Life sciences
- [ ] Behavioural & social sciences
- [ ] Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

### Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| Sample size | No sample-size calculations were conducted for use of bacteria in this study; the bacterial strains available for analysis dictated the sample size. For use of animals see Supplementary Methods, Murine Chronic Lung Infection for sample size calculation. |
|---|---|
| Data exclusions | No data were excluded from the analyses conducted in this manuscript. No animals were excluded as stated in the Supplementary Methods. |
| Replication | Multiple experiment in this study required replication to verify reproducibility. Antimicrobial production of the *Burkholderia ambifaria* strains against plant pathogens was performed in duplicate. Biological control assays of pea plant survival were performed in triplicate. Galleria kill assays were performed in triplicate. Minimum inhibitory concentration assays of enacyloxin I1a against plant pathogens was performed in triplicate. |
| Randomization | Mice were randomly assigned to a cage (experimental group) on arrival at the unit by staff with no role in study design. Reported in Supplementary methods: "Murine infection model". Covariates were controlled during experiments by including the necessary no treatment conditions. |
| Blinding | Blinding was not used in this study. Consistent and objective thresholds were defined by investigators for data recording during experiments to prevent subjective recording of results and investigator bias. |

### Reporting for specific materials, systems and methods

#### Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| [ ] | Unique biological materials |
| [x] | Antibodies |
| [x] | Eukaryotic cell lines |
| [x] | Palaeontology |
| [x] | Animals and other organisms |
| [x] | Human research participants |

#### Methods

| n/a | Involved in the study |
|---|---|
| [ ] | ChIP-seq |
| [ ] | Flow cytometry |
| [x] | MRI-based neuroimaging |

### Unique biological materials

Policy information about availability of materials

Obtaining unique materials | 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.
Obtaining unique materials

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.

Animals and other organisms

Policy information about studies involving animals: ARRIVE guidelines recommended for reporting animal research

| Laboratory animals | Mice; BALB/c; 6-8 weeks old |
|--------------------|-----------------------------|
| Wild animals       | The study did not involve wild animals |
| Field-collected samples | The study did not involve samples collected from the field. |