Genomics 114 (2022) 398–408

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Pangenome inventory of *Burkholderia sensu lato*, *Burkholderia sensu stricto*, and the *Burkholderia cepacia* complex reveals the uniqueness of *Burkholderia catarinensis*☆

Evelise Bach [a,*], Fernando Hayashi Sant'Anna [a], Guilherme Dubal dos Santos Seger [b], Luciane Maria Pereira Passaglia [a]

[a] Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil
[b] Departamento Interdisciplinar, Centro de Estudos Costeiros, Limnológicos e Marinhos, UFRGS, Imbé, Rio Grande do Sul, Brazil

ARTICLE INFO

Keywords:
Burkholderia sensu lato
Burkholderia cepacia complex
Burkholderia catarinensis
Pangenome
Functional annotation

ABSTRACT

Here the pangenome analysis of *Burkholderia sensu lato* (s.l.) was performed for the first time, together with an updated analysis of the pangenome of *Burkholderia sensu stricto,* and *Burkholderia cepacia* complex (Bcc) focusing on the Bcc *B. catarinensis* specific features of its re-sequenced genome. The pangenome of *Burkholderia* s.l., *Burkholderia* s.s., and of the Bcc was open, composed of more than 96% of accessory genes, and more than 62% of unknown genes. Functional annotations showed that secondary metabolism genes belonged to the variable portion of genomes, which might explain their production of several compounds with varied bioactivities. Taken together, this work showed the great variability and uniqueness of these genomes and revealed an underexplored unknown potential in poorly characterized genes. Regarding *B. catarinensis* 89^T, its genome harbors genes related to hydrolases production and plant growth promotion. This draft genome will be valuable for further investigation of its biotechnological potentials.

1. Introduction

The genus *Burkholderia* was proposed in 1992 to accommodate seven species of the genus *Pseudomonas* from the ribosomal RNA group II [1]. This Betaproteobacteria genus is known not only by the pathogenicity of some of its members but also by their impressive biotechnological potential [2,3]. There is an ever-growing number of new species descriptions of *Burkholderia sensu lato* (s.l.), which can be found as free-living bacteria in soil or water or as commensals of plants, animals, or fungi, highlighting the wide metabolic versatility of the group [4]. The taxonomy of this group has been revised through phylogenomic studies. Currently, *Burkholderia* s.l. is divided into *Burkholderia sensu stricto* (s.s.) and other six genera named *Paraburkholderia, Caballeronia, Robbsia, Mycetohabitans*, *Trinickia, and Pararobbsia* [5–8]. The genera *Paraburkholderia, Caballeronia,* and *Trinickia* contain plant symbionts, *Robbsia* is phytopathogenic, *Pararobbsia* contains environmental species, and *Mycetohabitans* accommodates fungal endosymbionts. In *Burkholderia* s.s. remains those considered plant pathogens, including *Burkholderia gladioli, Burkholderia glumae, and Burkholderia plantarii* [9], and the opportunistic human and other animal pathogenic species, such as *Burkholderia mallei* and *Burkholderia pseudomallei* [10]. Besides that, among *Burkholderia* s.s., there is a versatile group of 24 closely related species whose contrasting biological features are even more evident, the *B. cepacia* complex (Bcc) [11,12]. Although some members exhibit an important biotechnological potential for biocontrol, bioremediation, and plant growth promotion (PGP), some strains can be pathogens to plant or to immunocompromised humans with cystic fibrosis (CF) [4]. A phylogenomic study of this group suggested the existence of 13 new species within Bcc [13]. Genome metrics are valuable tools to develop an accurate identification of Bcc for clinical risk assessment, since some species are more related to poor prognosis and high patient to patient transmission [14].

Burkholderia s.s. species exhibit an unusual genomic structure formed of multiple replicons, whereas most genomes are composed of two or three chromosomes [10]. The extra chromosome of Bcc species is nonessential and could be considered a megaplasmid that codes for

☆ *Burkholderia catarinensis* 89^T genome sequence data from this article have been deposited with the GenBank Data Libraries under Accession No. MDEQ00000000. The version described in this paper is version MDEQ02000000.

* Corresponding author at: IB-UFRGS, Av. Bento Gonçalves 9500, 91501-970 Porto Alegre, Brazil.
E-mail address: evelise.bach@ufrgs.br (E. Bach).

https://doi.org/10.1016/j.ygeno.2021.11.011
Received 24 December 2020; Received in revised form 12 July 2021; Accepted 10 November 2021
Available online 13 November 2021

genes related to virulence, secondary metabolism, and other accessory roles [15]. Among *Burkholderia* s.l., genome sizes range from 3.28 Mb of the fungal endosymbiont *Mycetohabitans endofungorum* HKI456^T to 11.5 Mb of the soil bacterium *Paraburkholderia hospita* DS64^T. The smallest *Burkholderia* s.s. genome belongs to the pathogen *B. mallei* SAVP (5.23 Mb) and the largest to *Burkholderia contaminans* LMG 23361^T (10.35 Mb) [16,17]. These large genomes might explain in part their impressive metabolic versatility. The increasing amount of genome sequences available enabled new insights about bacterial genomic structures. In this regard, pangenome, accessory, and core genes have been determined for many species [18–21]. The core genome is composed of genes that are present in all species of a determined taxonomic group, accessory genes are those present in some strains, and pangenome consists of all genes present in the evaluated group [22]. For instance, Seo et al. (2015) evaluated the pangenome of the rice pathogens *B. gladioli, B. glumae,* and *B. plantarii* [9] and Bochkareva et al. (2011) analyzed the pangenome of 127 *Burkholderia* s.s. strains, focusing on the animal pathogens *B. mallei* and *B. pseudomallei* [10]. The pangenome of 116 Bcc has been recently investigated, revealing a high level of recombination between the species of this group [23]. Moreover, evolutionary processes namely recombination, inversions, gain or loss of genes, and selective pressures were investigated in *Burkholderia* s.s. genomes, providing insights on structure and adaptation [10,23]. Although many interesting aspects of *Burkholderia* s.l. genomes were investigated previously [24], their pangenome has not been determined yet.

We have recently described an interesting new member of Bcc, *Burkholderia catarinensis* 89^T, using polyphasic taxonomy and genome-based approaches [17]. This strain was isolated from Southern Brazilian native grassland soil and has shown impressive PGP features, namely the production of phytohormone and siderophore, phosphate solubilization, and the ability to control the growth of a wide range of phytopathogenic fungi [25]. When inoculated in apple plants, this strain delayed the development of infection by the fungus *Colletotrichum gloeosporioides* [26]. Moreover, strain 89 forms biofilm and produces several hydrolytic enzymes, which is critical for its competition in soil and also finds biotechnological applications. Since its description as a new species, no other *B. catarinensis* strain was found and phylogenetic reconstructions consistently result in strain 89 occupying a separate cluster [13]. Thus, this may indicate the uniqueness of this Bcc species. To better explore its biotechnological potentials and specific features, here we describe the re-sequencing of *B. catarinensis* 89^T through an Illumina approach and the investigation of its genome characteristics. Besides that, we evaluated for the first time the pangenome of *Burkholderia* s.l. and updated the analysis of the pangenome of *Burkholderia* s.s., and Bcc, especially focusing on *B. catarinensis* unique genes.

2. Results and discussion

2.1. Phylogenomic analysis

A phylogenomic analysis of type strains from the *Burkholderia* s.l. group and of selected strains of Bcc was performed to confirm their taxonomic identification. Running the analysis with three clustering algorithms, Get-Homologues recognized 346 and 1760 orthologs shared by all members of the *Burkholderia* s.l. and Bcc datasets, respectively (Fig. S1). The phylogenetic relationships among *Burkholderia* s.l., *Burkholderia* s.s., and Bcc strains are similar to the ones previously described [13,24] (Figs. 1 and 2). Although it was not the focus of this work, we questioned the identification of some strains [13]. In this regard, we could observe that some species are closely related and difficult to distinguish by applying the threshold of Average Nucleotide Identity (ANI) values commonly used (95–96%) [27]. For instance, in our analysis, *Paraburkholderia insulsa* and *Paraburkholderia fungorum* shared ANI values >98%. This is also the case of *Paraburkholderia phytofirmans* and *Paraburkholderia dipogonis* (96.1%), and *Paraburkholderia steynii* and *Paraburkholderia terrae* (96.6%) (Fig. S2). Further digital DNA:DNA hybridization (dDDH) analysis obtained from pairwise comparisons of genomes of type strains confirmed that not only *P. insulsa* and *P. fungorum*, but also *P. steynii* and *P. terrae* might belong to the same species, since their dDDH values were 85.80 and 71.20%, respectively (Table S4). The dDDH value between *P. phytofirmans* and *P. dipogonis* (66.10%) was below the threshold of 70% applied for species circumscription [27], which corroborated that they belong to different species [28].

Jin et al. [13] observed that genomes of some Bcc strains are misidentified at the NCBI database and suggested the split of this complex into 36 species (BCC01-BCC36), especially due to ANI and dDDH values lower than the threshold for species delimitation. Interestingly, these authors proposed the use of 96.48% as the ANI threshold, which corresponded to 70% of dDDH for Bcc members. Furthermore, they observed that dDDH analysis is more discriminatory for Bcc strains than ANI. Applying this threshold, our data mostly corroborated these authors' observations. Even though the clusters previously named as BCC03 and BCC05-09 are in the same cluster in the phylogenetic tree (Fig. 2, old pink-colored) and have ANI values >94.9% (Fig. S3), they are easily distinguished by dDDH (Table S4). Strains of these clusters may belong to the so-called *Burkholderia cenocepacia* genomovars IIIA (BCC05) and IIIB (BCC08) [29]. Likewise, strains *Burkholderia lata* 383^T, *Burkholderia* sp. FL7-5-30-S1-D0, *Burkholderia* sp. 170816, and strains from the *B. contaminans* cluster shared ANI values of 94.9–95.9%, close to the threshold commonly used. However, dDDH values indicated they belong to different species. A combination of genome metrics with phylogenomic was used to rename those strains misidentified in the NCBI database (Fig. 2, Table S3). Whenever strains differed from the type, they were considered *Burkholderia* sp. BCC01 to BCC47, updating the code proposed by Jin et al. [13]. It is noteworthy that by applying the genome metrics ANI and dDDH, we indicated that the current 24 species of Bcc should be splitted into at least 47 species. Therefore, our results corroborated previous observations that the taxonomy of Bcc members should be revisited [13]. The precise identification of Bcc members is especially critical to differentiate species related to poor outcomes or highly transmissible in CF patients, including *B. cenocepacia* and *Burkholderia multivorans* [14]. Thus, the taxonomy of this group is relevant for risk and safety assessments.

2.2. Pangenome determination

After phylogenomic analysis, we kept 129 genomes of type strains in *Burkholderia* s.l. dataset, 32 in *Burkholderia* s.s., and 113 Bcc genomes (Tables S1-S3). The *Burkholderia* s.l. dataset was mostly comprised of draft genomes, whereas most genomes from the Bcc dataset were complete or of scaffold assembly levels. Except for two genomes, all genomes exhibited completeness above 97.7% and contamination below 3%. Therefore, all genomes showed enough quality to perform pangenome analysis.

One of the interesting information pangenome analysis reveals is the possibility of observing a saturated curve of new genes (closed pangenome) or an ever-growing steep slope (open pangenome) for a determined taxonomic group [22]. This can be calculated by plotting the number of new genes added to the pangenome by the sequential addition of new genomes. Similar analyses are performed to evaluate the number of conserved genes (core) and unique genes that are incorporated into the pangenome. The pattern of the accumulation curves of *Burkholderia* s.l. and *Burkholderia* s.s. were similar (Fig. 3 and Fig. S4), both presenting a steep increase in the number of unique genes and in the overall number of genes included in the pangenome. Both curves did not reach a plateau. The rarefaction curves of Bcc was increasing albeit showing a slight tendency of reaching a plateau (Fig. 4) as obtained previously analyzing 116 Bcc genomes [23].

All three pangenomes were determined as open, in which new genes continue to be added to the pangenome pool each time a new genome is

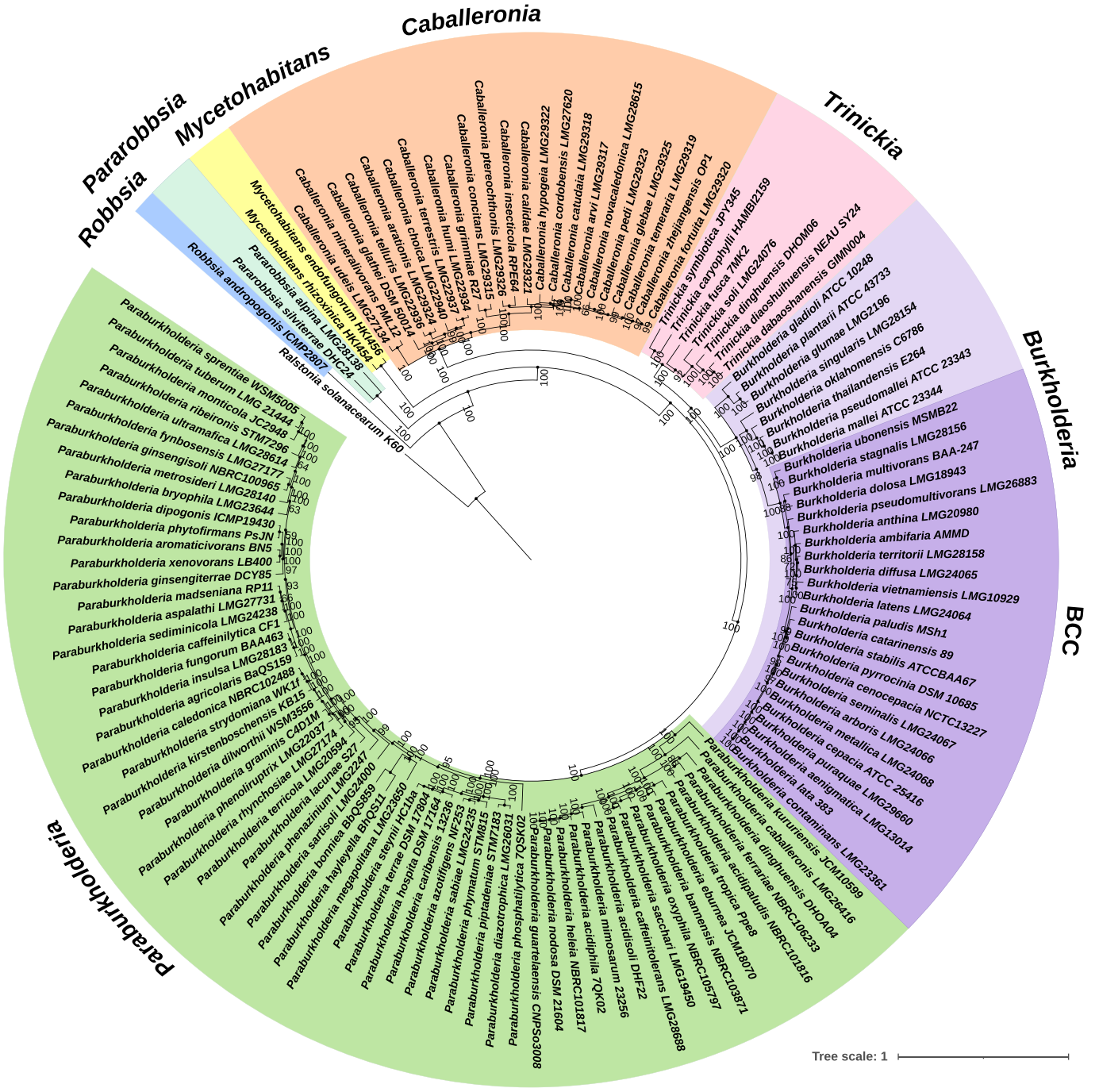

Fig. 1. Phylogenetic reconstruction of *Burkholderia sensu lato* highlighting the genera *Paraburkholderia, Robbsia, Pararobbsia, Mycetohabitans, Caballeronia, Trinickia, Burkholderia sensu stricto*, and the species belonging to the *Burkholderia cepacia* complex (BCC). The phylogeny was reconstructed based on the alignment of 113 orthologous protein sequences recognized in the genome of type strains through three clustering algorithms (bidirectional best-hit, COGtriangles, or OrthoMCL) using the software Get-Homologues. The best tree was estimated through the maximum likelihood approach and IQ-TREE using the software Get-Phylomarkers. *Ralstonia solanacearum* K60^{T} was set as the outgroup. All bootstrap values are shown.

included. Open pangenomes were already observed for other free-living versatile genera including *Pseudomonas, Bacillus,* and *Paenibacillus* [20,21,30]. This curve pattern may indicate the necessity of constant genomic adaptations and diversifications to cope with heterogeneous environments, which corroborates the high number of genomic islands (GIs), insertion sequences (ISs), and recombination rates observed in *Burkholderia* genomes [23]. In contrast to an increasing pangenome, we observed a stabilization on the number of core genes within less than 10 genomes for all three datasets.

BPGA predicted approximately 134,000 gene families in the repertoire of *Burkholderia* s.l. pangenome, 25,000 in *Burkholderia* s.s., and 24,000 in the Bcc (Figs. 3e, S4e, and 4e). Get-Homologues recognized 83,324, 26,163, and 37,824 genes in the pangenome of *Burkholderia* s.l., *Burkholderia* s.s., and Bcc, respectively (Fig. S5). Table 1 shows the number of pangenome genes recognized by each software and their distribution among core, soft-core, shell, and cloud genes, considering 75% of sequence identity level in BLAST pairwise alignments. The absolute numbers varied among softwares, reflecting differences in the algorithms implemented by each tool. As expected, we could clearly observe an increase in the number of conserved (core)

genes predicted in lower taxonomic ranks. *Burkholderia* s.l. pangenome was composed of 178–539 (<0.55%) core genes, whereas *Burkholderia* s. s. shared 1338–1938 (<7.4%), and Bcc 2049–2588 (<5.5%) genes. This highlights the great number of accessory genes (shell and cloud) in these genomes (>96%) and an impressive variability among species of the same genus and among different strains of the same species. It is noteworthy that low values of core genes (<6%) were also obtained when we lowered the sequence identity level of the analysis to 50% (Table S5). This low number of shared conserved genes has also been observed in the pangenome of seven rice-pathogenic *Burkholderia* [9], 127 *Burkholderia* spp. [10], and 116 Bcc [23]. Therefore, *Burkholderia* s.l. genomes are characterized by being highly heterogeneous.

Homologous recombination (HR) among species and within the same species exerts a major role in the evolution of prokaryotic genomes [31]. These authors observed 338 genomes distributed in 54 bacterial and archaeal species and observed different levels of HR according to their lifestyle. Opportunistic pathogens showed the highest levels of HR, followed by obligate pathogens, commensals and free-living, and endosymbionts and intracellular pathogens showed the lowest HR frequencies. In this regard, we investigated the variation in the number of

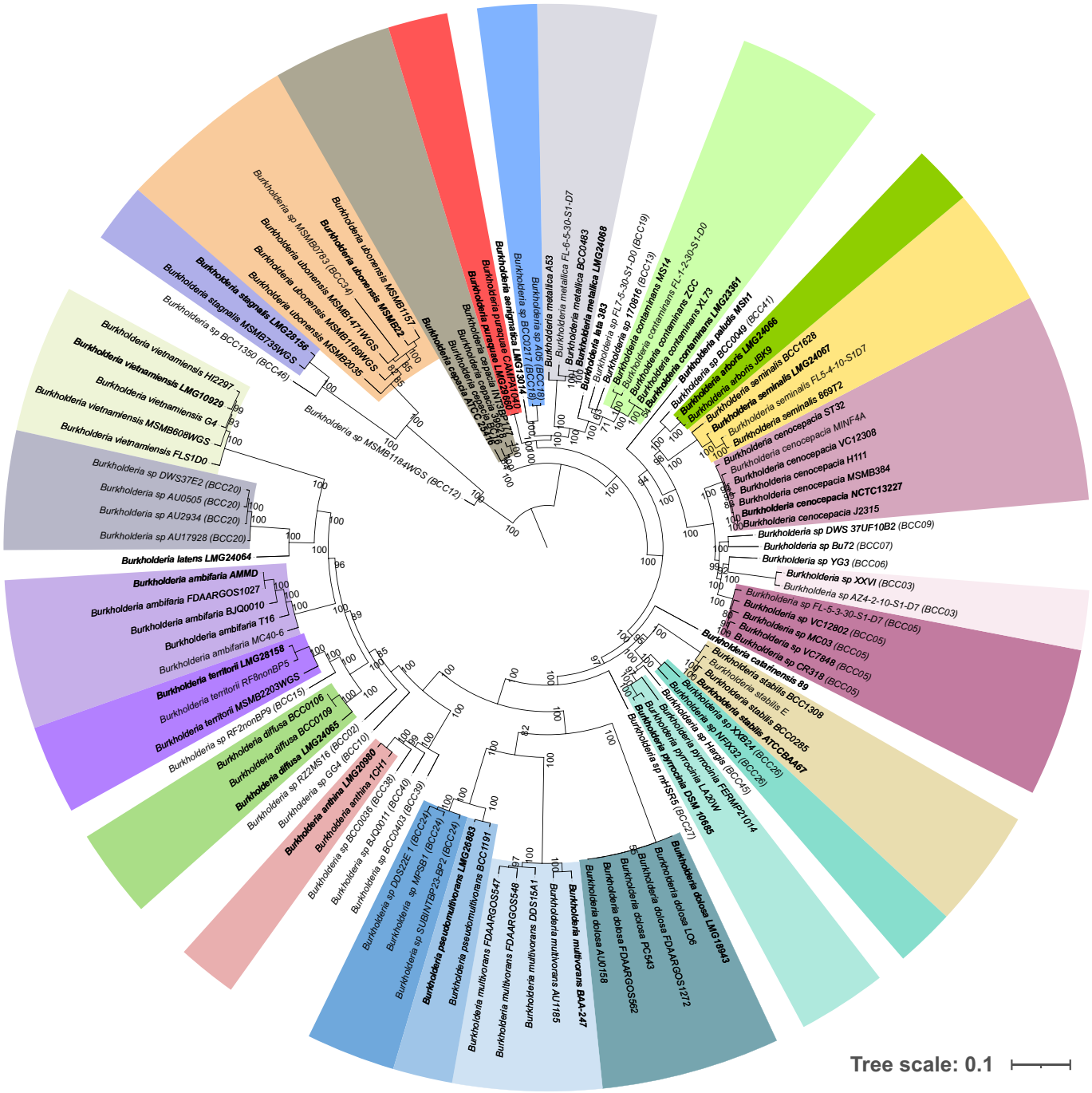

Fig. 2. Phylogenetic reconstruction of the selected strains from the *Burkholderia cepacia* complex analyzed in this work. The phylogeny was reconstructed based on 120 orthologous protein sequences recognized in their genomes through three clustering algorithms (bidirectional best-hit, COGtriangles, or OrthoMCL) using the software Get-Homologues. The best tree was estimated through the maximum likelihood approach and IQ-TREE using the software Get-Phylomarkers. All bootstrap values are shown. Type strains are shown in bold. Background colored clusters contain strains considered of the same species according to ANI and dDDH values. The codes (BCC02-46) were adopted for strains that might belong to new species according to genome metrics.

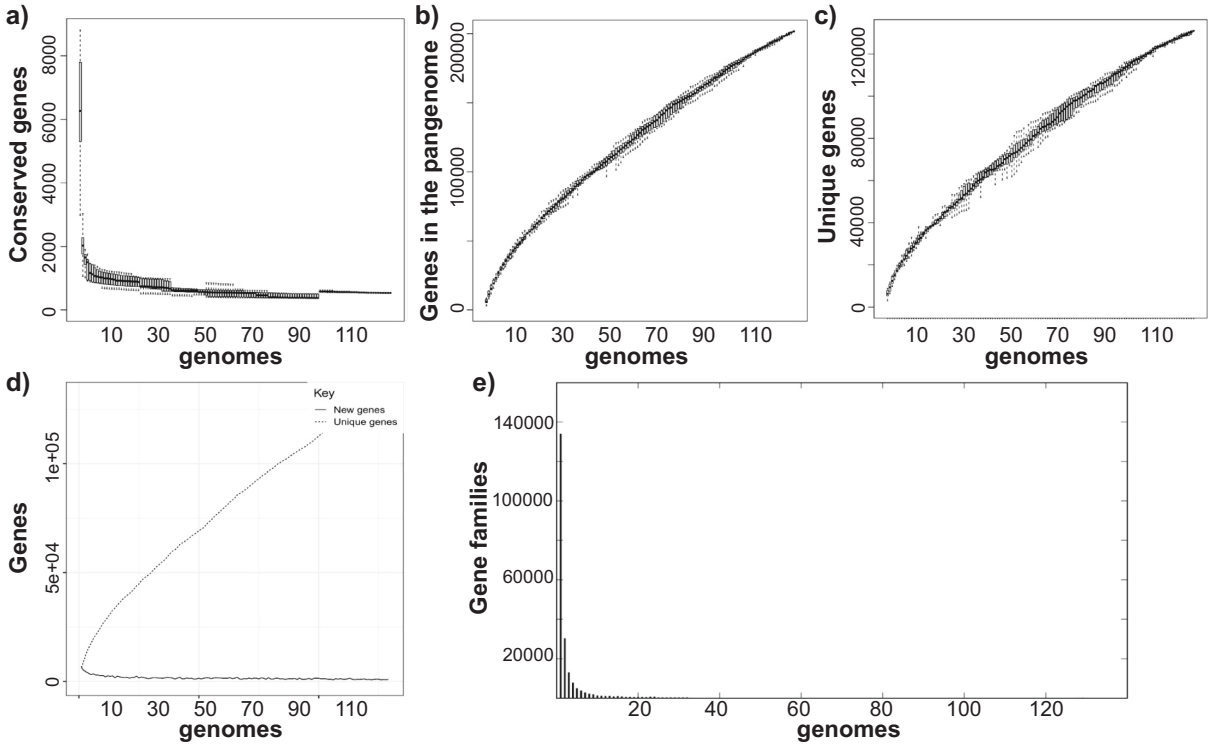

Fig. 3. Pangenome features of 129 *Burkholderia sensu lato* (s.l.) whole-genome sequences. Accumulation curves of conserved genes **(a),** total genes in the pangenome **(b)**, unique genes **(c)**, new and unique genes **(d)**, and distribution of gene families shared by different numbers of *Burkholderia* s.l. genomes **(e)**.

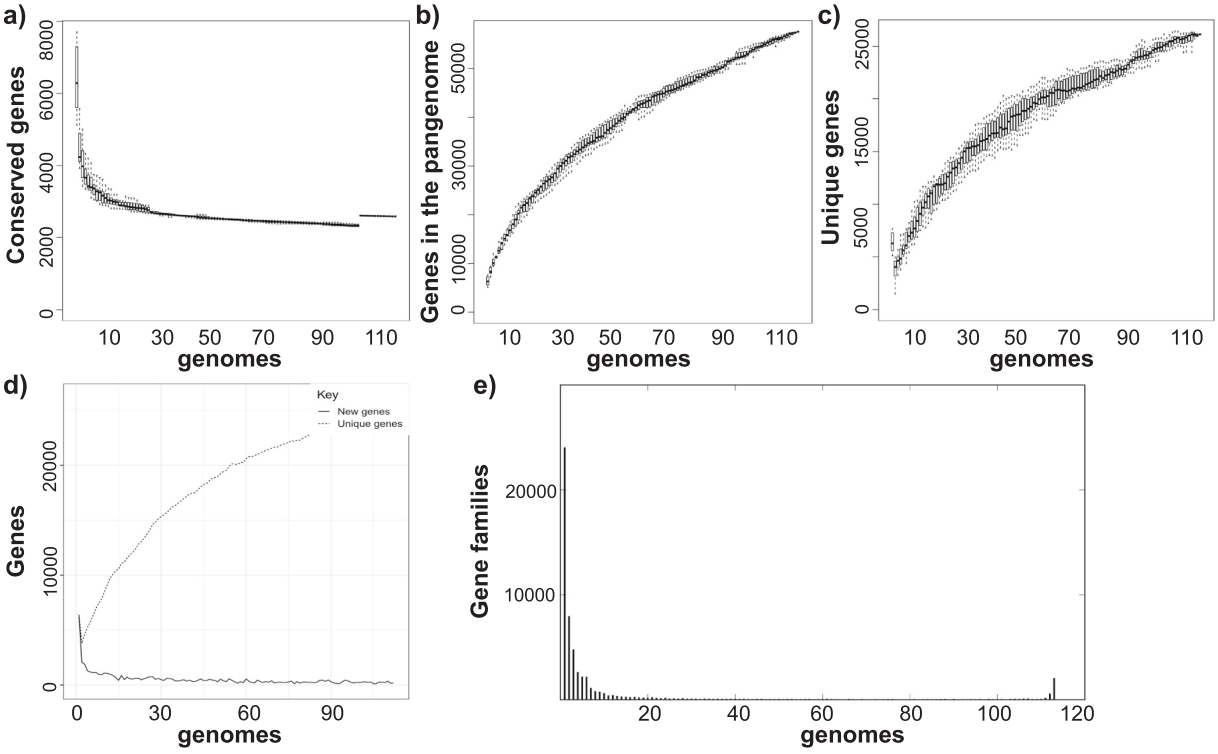

Fig. 4. Pangenome features of 113 whole-genome sequences of strains from the *Burkholderia cepacia* complex (Bcc). Accumulation curves of conserved genes **(a),** total genes in the pangenome **(b)**, unique genes **(c)**, new and unique genes **(d)**, and distribution of gene families shared by different numbers of Bcc genomes **(e)**.

unique genes in different taxonomic ranks looking for patterns that could indicate different genomic plasticities (Fig. 5).

There were 134,011 unique genes in *Burkholderia* s.l. pangenome repertoire and 24,039 in Bcc. The genera *Caballeronia* and *Paraburkholderia* showed the greatest variation in the number of unique genes among *Burkholderia* s.l., 438 to 2139 and 350 to 2734, respectively

Table 1
Distribution of pangenome genes into core, soft-core, shell, or cloud according to BPGA, Get-Homologues, and Roary tools considering 75% of sequence identity level in BLAST pairwise alignments.

Genes	*Burkholderia senso latu*			*Burkholderia senso strictu*			*Burkholderia cepacia* complex		
	BPGA*	Get-Homologues	Roary	BPGA*	Get-Homologues	Roary	BPGA*	Get-Homologues	Roary
core (99–100%)	178 (0.08%)	460 (0.55%)	539 (0.27%)	1338 (3%)	1938 (7.4%)	1831 (4.3%)	2049 (3.6%)	2085 (5.5%)	2588 (4.5%)
soft core (95–99%)		397 (0.47%)	473 (0.24%)		523 (2.0%)	277 (0.6%)		878 (2.3%)	309 (0.54%)
shell (15–95%)	217,605 (99.9%)	9929 (11.9%)	8454 (4.2%)	44,526 (97%)	5574 (21.3%)	8452 (19.6%)	55,238 (96.4%)	10,154 (26.8%)	6167 (10.7%)
cloud (0–15%)		72,538 (87.05%)	192,255 (95,3%)		18,128 (69.3%)	32,517 (75.5%)		24,698 (65.3%)	48,496 (84.2%)
pangenome (total)	217,783	83,324	201,721	45,864	26,163	43,077	57,287	37,824	57,560

* BPGA pangenome genes are classified into core or accessory genes.

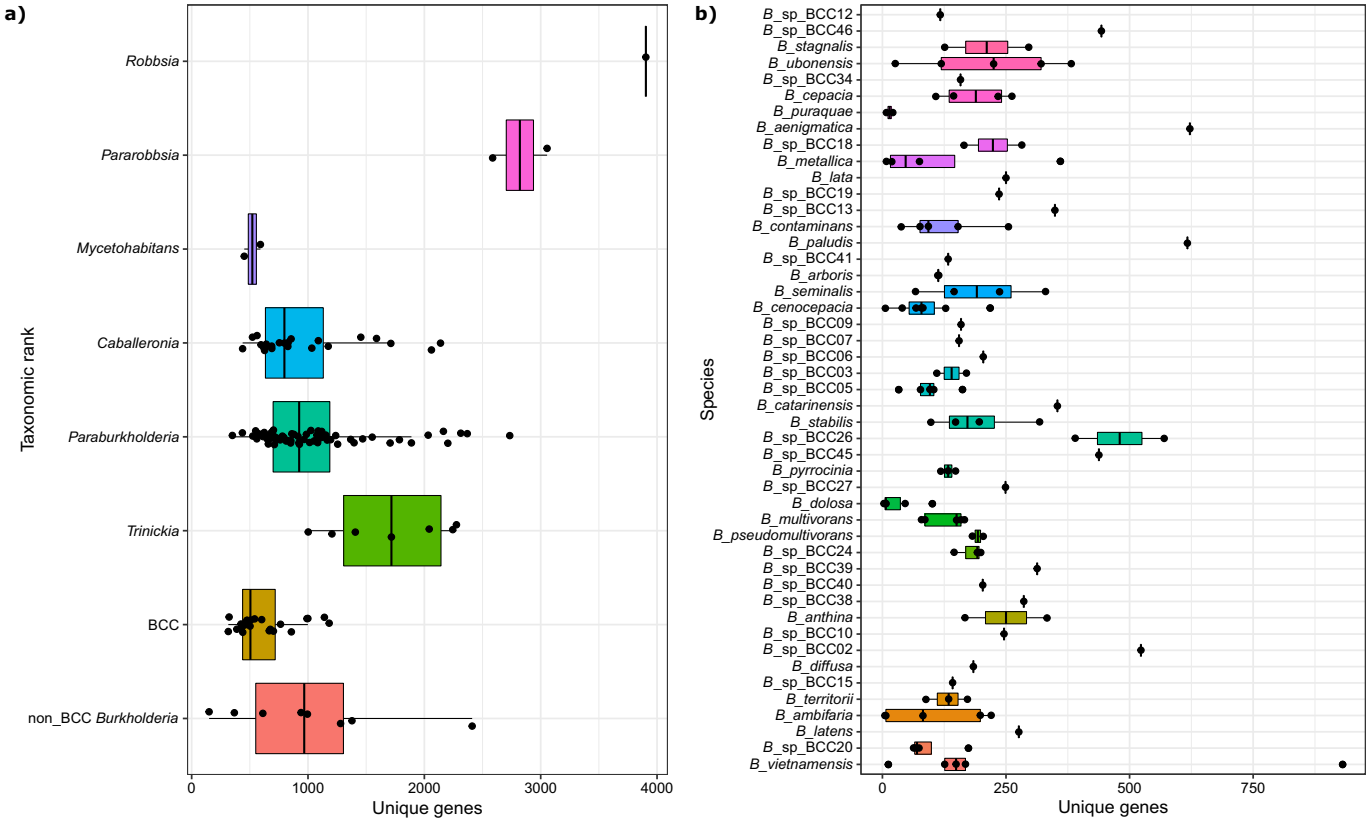

Fig. 5. Box plot showing the variation in the number of unique genes in *Burkholderia sensu lato* genomes **(a)** and in the genomes of Bcc members **(b)** as obtained by BPGA. Putative new species of BCC are named according to Fig. 2.

(Table S1). The genus *Robbsia*, up to now represented by only one species, added the highest number of species-specific genes to *Burkholderia* s.l. pangenome: 3904. *Burkholderia vietnamiensis* G4, *Burkholderia paludis* Msh1, and *Burkholderia aenigmatica* LMG13014 genomes harbor the highest numbers of unique genes among Bcc members, 1265, 983, and 807, respectively (Table S3). In *B. vietnamiensis* G4, this might be a reflection of its large genome organized in three chromosomes and five plasmids [16]. Although some species varied more (*B. vietnamiensis, Burkholderia ubonensis*, and *Burkholderia metallica*) than others (*B. cenocepacia* and *B. multivorans*) in the number of unique genes, we could not clearly correlate with differences in lifestyles. Indeed, this might be difficult to observe in a versatile genus such as *Burkholderia* s.s., since the same species could be commonly isolated not only from the soil but also from the sputum of CF patients [4].

Noteworthily, it is clear that pangenome evaluations strongly depend upon both the number of genomes evaluated and the quality of genome sequences [21,32]. Once more high-quality genomes and genomes representing more strains are available, we would exclude a possible effect of overrepresentation of some species genomes in this type of analysis. In this regard, an analysis of the whole dataset through BPGA also revealed that the genera *Caballeronia* and *Paraburkholderia* are highly heterogeneous in the number of species-specific genes (Fig. S6 and Table S6). These genera might be an underexplored source of new genes and features. Among Bcc, putative new species and those with few genomes sequenced showed a high number of unique genes as well. Interestingly, *B. catarinensis* showed an above-average number of unique genes (456) with the potential to be explored (Table S3). Surprisingly, other putative new Bcc species also displayed a high number of strain-specific genes and composed a separate cluster in the phylogenetic reconstructions. However, except for some outliers, Bcc species contribute

less with unique genes to *Burkholderia* s.l. pangenome when compared to other taxonomic groups.

The heatmaps generated based on the matrix of presence or absence of genes in the *Burkholderia* s.s. and Bcc pangenomes also displayed the variation in the distribution of accessory genes among strains and species (Figs. S7 and S8). Get-Homologues predicted 331 orthologues of IS family transposase in *Burkholderia* s.l. pangenome and 210 within Bcc genes. Both pangenomes also exhibited a great number of orthologues of type 2, 3, and 4 secretion systems and *pga* genes putatively related to biofilm formation. Regarding genes that may encode beneficial traits for plants, we observed *Burkholderia* s.s. and *Paraburkholderia* genomes harboring *iaaM* genes putatively related to the phytohormone auxin production; the widespread presence of *pst, pho,* and *phn* genes for phosphate metabolism; except for *Mycetohabitans* spp., the widespread presence of *acdS* genes for ACC deaminase production that may reduce the stress response of plants; *nodA* genes in *Trinickia symbiotica* and some *Paraburkholderia*; *nodD* genes in some *Burkholderia* s.s. strains, related to plant nodulation factors; a *nif* operon exclusively present in *B. vietnamiensis*, related to its capacity of nitrogen fixation; and Bcc genomes harboring genes related to the production of the volatiles acetoin and 2,3-butanediol, which are related to PGP. Briefly, regarding genes useful for the biocontrol of fungi and insects, *Burkholderia* and *Caballeronia* species present chitinase genes for fungal cell-wall degradation and the *Pseudomonas* insect toxin encoding genes *mcf* and *fitD* were absent in all genomes [33].

To gain further insights into the functional roles of core, accessory, and unique genes, we annotated them according to COG and KEGG categories. The overall distribution of functions among *Burkholderia* s.l., *Burkholderia* s.s., and Bcc were similar (Fig. S9). It is interesting to note that more than 70% of all dataset genes were classified at KEGG category metabolism. Moreover, COG and KEGG category distributions of *Burkholderia* s.s. genes resembled the ones of the subgroup Bcc (Figs. 6 and 7). In general, housekeeping functions belonged to core genes. The housekeeping COG categories C, F, and J were more represented in the core genome of *Burkholderia* s.l. than in *Burkholderia* s.s. and Bcc pangenomes. Unlikely, categories amino acid (E), carbohydrate (G), and lipid metabolism (I), cell motility (N), and signal transduction (T) were more represented in core genes of *Burkholderia* s.s. than in *Burkholderia* s.l.. Most genes from categories carbohydrate (G), lipid (I), inorganic ion (P), and secondary metabolism (Q), transcription (K), cell motility (N), signal transduction (T), and defense mechanisms (V) were not conserved among *Burkholderia* s.l. strains and thus belonged to the variable fraction of genomes. More specifically, genes coding for replication, recombination, and repair (L) were expanded in unique genes, especially among Bcc strains. *Pseudomonas* species share the same pattern of having their impressive secondary metabolism genes within the variable genome portion [21,34]. Furthermore, an extensive transcription regulation might be a requirement for such large genomes of strains living in complex environments, which was also observed in *Paenibacillus* pangenome [30].

According to the KEGG database, categories translation, energy and nucleotide metabolism, and drug resistance were functions highly conserved among *Burkholderia* s.l. species, while cell motility was encoded mostly by accessory and unique genes. The metabolism of lipids, amino acids, terpenoids, and polyketides were enriched in the variable fraction of all pangenomes, which could explain the capacity of *Burkholderia* s.s. strains to produce several molecules with impressive bioactivities [35]. Similarly, the majority of membrane transport, cell

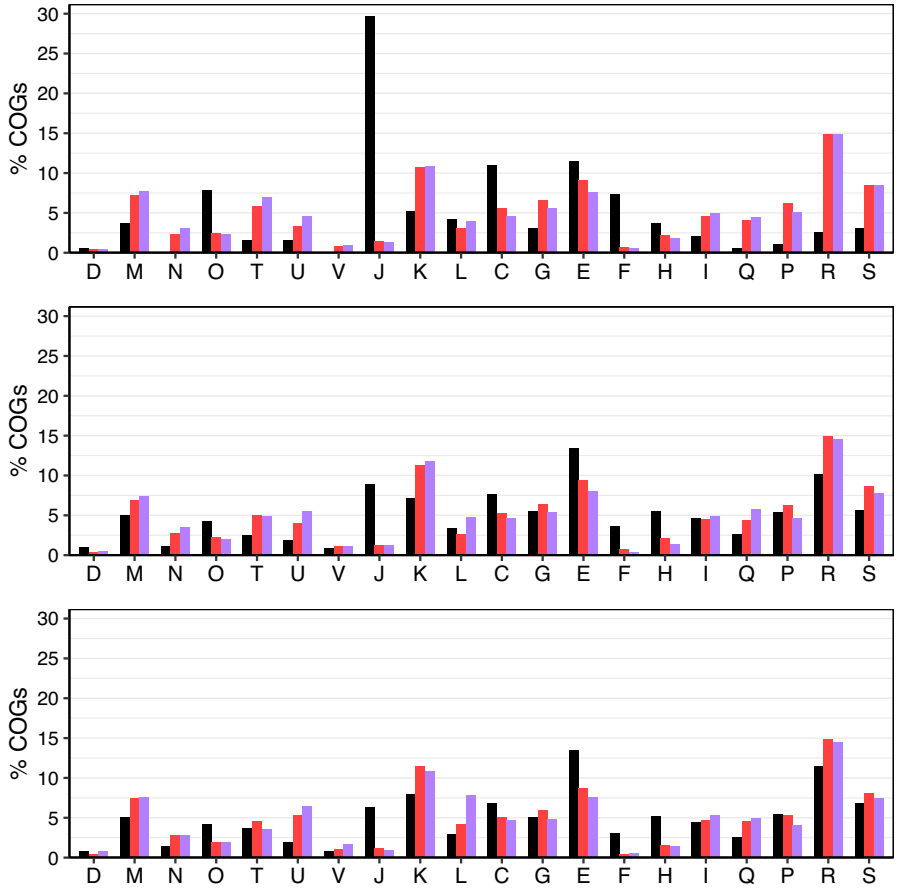

Fig. 6. Functional distribution of core, accessory, and unique genes in the pangenome of *Burkholderia sensu lato* **(a)**, *Burkholderia sensu stricto* **(b)**, and *Burkholderia cepacia* complex **(c)** according to COG categories. Underrepresented categories were removed from the figure.

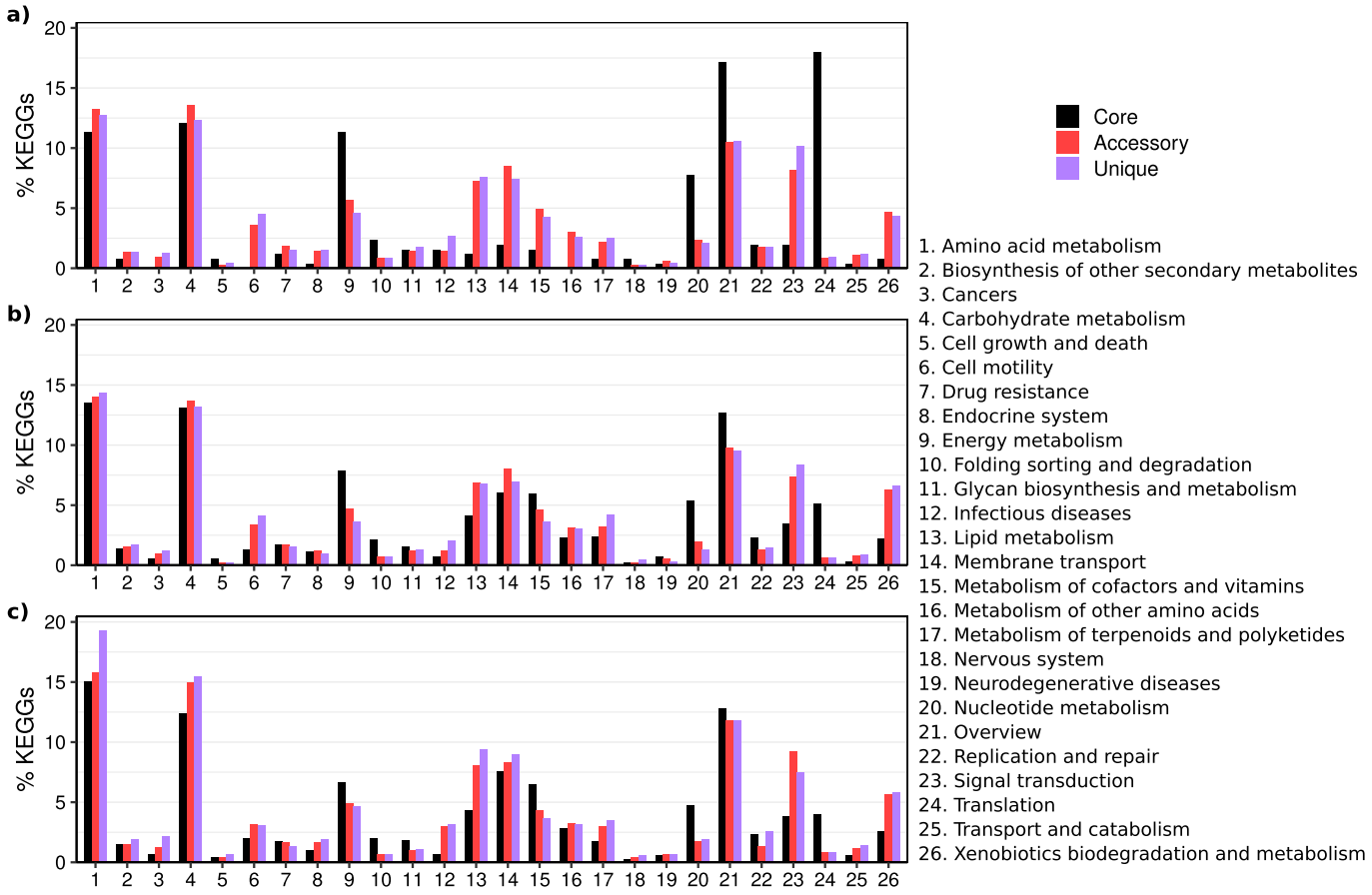

Fig. 7. Functional distribution of core, accessory, and unique genes in the pangenome of *Burkholderia sensu lato* **(a)**, *Burkholderia sensu stricto* **(b)**, and *Burkholderia cepacia* complex **(c)** according to KEGG categories. Underrepresented categories were removed from the figure.

motility, signal transduction, and xenobiotics metabolism genes belonged to the accessory or unique portion of pangenomes. Corroborating COG data, replication and repair functions of Bcc were more located in the strain-specific genes than in *Burkholderia* s.l. species, which could be an indicative of increased recombination. Zhou et al. [23] suggested that Bcc core genomes are under a process of strong positive selection and that recombination is a critical driving force for both the cohesion of this group and their metabolic versatility. Interestingly, these authors observed that recombination was more frequent between species than within the same species, which produces the observed difficulty in determining their taxonomic boundaries. Core genes of *Burkholderia* s.l. were better characterized than *Burkholderia* s.s. (Fig. S9). Moreover, the matrix of absence or presence of genes provided by the Roary tool resulted in 67.8% of pangenome repertoire of *Burkholderia* s.l. being composed by hypothetical proteins, whereas they sum 62.8% in *Burkholderia* s.s. and 67.2% in Bcc. Therefore, we estimate there is still an unknown biotechnological potential in *Burkholderia* s.l. genomes.

2.3. *Burkholderia catarinensis* 89^T genome general features

The first version of the genome sequence of *B. catarinensis* strain 89 has already been published [17] and is composed of 892 contigs. To better explore the genomic features of this strain, here we described a second version of its genome with more contiguous sequences. The draft genome of *B. catarinensis* 89 has a total length of 8,198,227 bp and an estimated G + C content of 66.5%. A5 assembly was chosen due to the highest value of N_{50} (264,500), fewer contigs (85), completeness of 99.95%, and absence of contamination. Table S7 presents some general features of *B. catarinensis* 89^T draft genome. The genome of *B. catarinensis* 89^T is highly represented by genes putatively related to amino acid, carbohydrate, cofactors, vitamins, prosthetic groups, pigments, fatty acids, lipids, isoprenoids, and protein metabolisms (Fig. S10). With regards to the aforementioned 456 unique genes predicted in the genome of strain 89 through pangenome analysis, approximately 80% of them are of unknown functions. Other unique genes were classified in the categories of genetic and environmental information processing, signaling, carbohydrate, amino acid, and nucleotide metabolism (Fig. S11). Interestingly, some of *B. catarinensis* 89 unique genes were also putatively related to xenobiotics degradation, lipid metabolism, and biosynthesis of secondary metabolism, which could be related to its remarkable antifungal activity. Since this strain displays interesting PGP activities and biocontrol of a wide range of phytopathogenic fungi, we further investigated its genome specific features.

The ISfinder database predicted 21 genes of transposable elements belonging to the families IS3 (18 genes), IS256 (2 genes), and Tn31 (1 gene) similar to those from various *Burkholderia* species, but also from Gram-negative bacteria of the genera *Aeromonas, Escherichia, Ralstonia,* and *Xanthomonas* (Table S8). Surprisingly, 222 GIs were predicted by IslandViewer (Table S9). *Burkholderia* spp. genomes are known by the presence of many IS and GIs [36,37]. The presence of GIs indicates the acquisition of foreign DNA through horizontal gene transfer, which might exert a critical role in the evolution of *B. catarinensis* originating this unique species. Genes obtained through these mechanisms often confer antibiotic resistance or some competitive advantages to thrive on the rhizosphere or host environment [38].

Indeed, it was possible to observe a great number of genes related to the type 6 secretion system (T6SS) among the GIs recognized by Island

Viewer. Three clusters predicted both by ISFinder and IslandViewer seem to be comprised of the complete set of T6SS genes [39]. This feature agreed with the pangenome data and with the *Burkholderia* genus, which exhibits many secretion systems with important roles in the interaction of the bacterium with its host and other bacteria. For instance, the type III secretion system (T3SS), which enables bacteria to secrete effector proteins into eukaryotic host cells, has a crucial role in the virulence of several plant and human pathogens [40]. Moreover, the T3SS is also essential for the interaction of *Paraburkholderia terrae* BS001 with the fungus *Laccaria proxima* [41]. Spiewak et al. [39] showed that T6SS present roles in bacteria competition in *B. cenocepacia* and probably in other *Burkholderia* strains. The genome of *B. catarinensis* 89 harbors one gene cluster for the type I secretion system (T1SS), three clusters for type II (T2SS), two for type III (T3SS), three for type VI (T6SS), and three clusters for a combined type II and type IV secretion systems (Table S10).

Bcc members harbor several genes related to antibiotic resistance, which is worrisome in infections of patients with CF [42]. Interestingly, not many resistance genes were predicted by the RGI tool for the genome of strain 89 (Table S11). Five Resistance-Nodulation-Division efflux pumps for the resistance to aminoglycoside, fluoroquinolone, and tetracycline antibiotics were predicted. Besides that, a Major Facilitator Superfamily efflux pump putatively related to tetracycline resistance was also found. ResFinder detected no resistance to antibiotics or disinfectants, even lowering default threshold values. Clustered regularly interspaced short palindromic repeat (CRISPR) gene sequences that encode defense mechanisms against phages and plasmids are also indicative of the ability to deal with the environment. In this sense, CRISPRCasFinder predicted eight of these regions in the genome of strain 89 (Table S12).

B. catarinensis 89 exhibits many features that might be important for rhizosphere survival and the beneficial interaction with plants, including phytohormone, siderophore, and biofilm production, as well as phosphate solubilization [25,26]. The presence of genes related to these activities in the genome of strain 89 corroborated our experimental results. Its genome has several genes putatively involved in the production of the phytohormone auxin, including those related to tryptophan and indole metabolism. There are at least six possible pathways for bacterial auxin biosynthesis, sometimes in redundancy [43], whereas the described auxin production of strain 89 [26] might involve the gene for indoleacetamide hydrolase (*iaaH*, EC 3.5.1) of the Indole-3-Acetamide pathway. Besides that, we could identify a non-ribosomal peptide synthetase gene cluster related to the siderophore production of strain 89. Additionally, several genes related to iron uptake were found in the genome, including the ones related to hydroxamate siderophores (*fhuABC* and *tonB*) and hemin uptake systems (ABC transporters) [44]. Moreover, there are many genes probably involved in phosphate homeostasis in the genome of strain 89, including three genes for polyphosphate kinases; two genes for exopolyphosphatases; one gene for inorganic pyrophosphatase and one gene for alkaline phosphatase production. Besides that, its genome presents the polycistronic *pstSCAB-phoU* operon that encodes proteins of the phosphate-specific transport (Pst) system.

The extracellular matrix that forms bacterial biofilms is composed of polymers like poly-beta-1,6-N-acetyl-D-glucosamine (PGA) [45]. In agreement with pangenome data, we found the entire *pgaABCD* operon in *B. catarinensis* 89's genome. Moreover, many other genes that putatively encode for several types of hydrolases were predicted in this strain's genome by the dbCAN2 server (Fig. S12). This tool predicted a total of 97 enzymes distributed among the families of Glycoside Hydrolases (36), Glycosyl Transferases (46), Carbohydrate Esterases (6), Carbohydrate Binding Modules (3), Polysaccharide Lyases (1), and Auxiliary Activities (5). Among the glycoside hydrolases (GH), were annotated genes belonging to the Carbohydrate Active enZymes (CAZy) subfamilies of cellulases (GH-5, GH-8), amylases (GH-13 and GH-15), and chitinase (GH-23), which corroborated our previous laboratory results [25]. These features could not only be crucial for this bacterium rhizosphere competence but could also have biotechnological potential. Additionally, *B. catarinensis* strain 89 shows an impressive antifungal activity, whose metabolites investigation is underway.

3. Conclusion

This is the first analysis of *Burkholderia* s.l. pangenome. Taken together, our work showed that the pangenome of *Burkholderia* s.l., *Burkholderia* s.s., and of the Bcc was open, and therefore more novel genes might be discovered once more genomes are sequenced. The conserved portion of the pangenomes was small and more than 96% of the pangenome was composed of accessory or unique genes, which in its turn was mostly composed of genes of unknown functions. Our work revealed the great variability in the repertoire of genes harbored in genomes of *Burkholderia* s.l. species and strains, especially in *Caballeronia* and *Paraburkholderia*, that might be an underexplored source of features and metabolites with a variety of biotechnological applications. In this regard, we further investigated the specific features of the draft genome of *B. catarinensis* 89. Our findings corroborated some previous experiments that tested its abilities of PGP, enzyme production, and rhizosphere competence [25] and also revealed the great plasticity of this IS- and GI-rich genome. The pangenome analysis indicated 456 unique genes putatively dedicated to genetic and environmental information processing, signaling, xenobiotics degradation, and biosynthesis of secondary metabolites. Noteworthily, approximately 80% of *B. catarinensis* 89 unique genes are of unknown functions, which indicated that we still have a lot more to learn from this interesting strain and from other *Burkholderia* s.l. strains.

4. Material and methods

4.1. Genome sequencing, assembly, and annotation

Genomic DNA of *Burkholderia catarinensis* 89^T was extracted and used to construct 500 to 1200 bp insert libraries using Nextera XT kit for sequencing in the MiSeq Illumina platforms using the MiSeq Reagent kit v3 (2 × 300). Genome assembly was performed with A5 [46] and Spades tools [47]. Quality assessment of assemblies was evaluated using CheckM [48] and QUAST [49]. Scaffolds were automatically annotated in the RAST server and subsystems were investigated in SEED Viewer [50]. A search for tRNAs and rRNAs was carried out using ARAGORN and Barrnap, respectively [51]. This whole-genome shotgun project has been deposited at GenBank under the accession number MDEQ00000000. The version described in this paper is version MDEQ02000000.

4.2. Core and pangenome analyses

The most complete genomes of type strains of species from the genera *Burkholderia*, *Paraburkholderia, Caballeronia, Robbsia, Pararobbsia, Mycetohabitans*, and *Trinickia* were extracted from the GenBank database [52]. Some species were excluded since they were not validly published following the LPSN database (accessed on May 2021) [53]. Additionally, to analyze the Bcc pangenome, we downloaded genomes presenting the assembly level Complete, Chromosome, or Scaffold, updating the dataset used by Jin et al. [13]. All genomes were obtained on May 2021. To avoid overrepresentation of some species, the top five genomes presenting the highest quality were chosen. The quality parameters completeness and contamination were obtained through CheckM [48]. To standardize the comparisons, all genomes were annotated using Prokka4 [54,55].

Pangenome analyses were performed with three datasets (*Burkholderia* s.l., *Burkholderia* s.s., and Bcc) using three different tools, named Get-Homologues [56], Roary [57], and BPGA [58], considering 75% of sequence identity level in BLAST pairwise alignments. Pan- and

core-genes accumulation curves were generated using the clustering algorithm CD-HIT implemented in Roary and Usearch clustering tool of BPGA. BPGA applies the power-law regression model to determine if the rarefaction curve indicates an open or closed pangenome [22]. A consensus of homologues recognized by Get-Homologues through the clustering algorithms COGtriangles (COGt) and OrthoMCL (OMCL) was produced, which were further classified into cloud, shell, soft-core, or core genes. Core are those genes present in 99–100% of genomes, soft-core in 95–99%, shell (15–95%), and cloud in 0–15% of genomes. The analyses of unique genes were obtained through BPGA and using the R statistical environment [59] based on the matrix of absence or presence of genes generated by Get-Homologues. The functional annotation of genes according to COG and KEGG categories was performed by BPGA.

4.3. Phylogenomics

Single-copy orthologous genes were recognized from the datasets through Get-Homologues using the three clustering algorithms COGt, OMCL, and bidirectional best-hit. The consensus of those orthologs was used for the phylogenetic reconstruction performed following the Get-Phylomarkers pipeline [60]. With this tool, multiple sequence alignments and maximum likelihood phylogenies are computed in parallel and the tree is estimated using IQ-TREE. Get-Homologues was also used to calculate the Average Nucleotide Identity (ANI) among genomes based on BLAST alignments (ANIb). Genomes sharing ANI values above the 96.48% threshold were considered from the same species [13]. Digital DNA:DNA hybridization (dDDH) analyses were performed for all genomes displaying borderline ANI values. This analysis was carried out at http://ggdc.dsmz.de/home.php using the recommended formula 2 [27].

4.4. *B. catarinensis* 89^T genome features

ISfinder [61] database was used for the identification of transposable elements or insertion sequences (IS) using an identity cut-off value of 7×10^{-5}. Genomic Island (GI) was predicted through Island Viewer 4 [62] aligning against chromosomes of *Burkholderia pyrrocinia* DSM 10685^T. Clustered regularly interspaced short palindromic repeat sequences were searched using CRISPRCasFinder [63]. Antibiotic resistance genes were identified using the strict module of the Resistance Gene Identifier software of the Comprehensive Antibiotic Resistance Database [64]. ResFinder-4.1 Server was also used to predict genes related to the resistance of antibiotics and disinfectants. Carbohydrate Active enZymes (CAZy) annotation was performed at dbCAN2 server selecting the enzymes predicted by the three tools HMMER, Diamond, and Hotpep [65].

Declaration of Competing Interest

None declared.

Acknowledgements

The authors thank Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil, for providing EB's postdoctoral scholarship (Grant number 155771/2018-3). We also would like to thank the staff of the Biochemistry and Molecular Biology Department of Universidade Federal do Paraná for library preparation and genome sequencing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2021.11.011.

References

[1] E. Yabuuchi, Y. Kosako, H. Oyaizu, I. Yano, H. Hotta, Y. Hashimoto, T. Ezaki, M. Arakawa, 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 (1992) 1251–1275.
[2] E. Depoorter, M.J. Bull, C. Peeters, T. Coenye, P. Vandamme, E. Mahenthiralingam, Burkholderia: an update on taxonomy and biotechnological potential as antibiotic producers, Appl. Microbiol. Biotechnol. 100 (2016) 5215–5229.
[3] E. Bach, L.M.P. Passaglia, J. Jiao, H. Gross, Burkholderia in the genomic era: from taxonomy to the discovery of new antimicrobial secondary metabolites, Crit. Rev. Microbiol. 1–40 (2021).
[4] T. Coenye, P. Vandamme, Diversity and significance of Burkholderia species occupying diverse ecological niches, Environ. Microbiol. 5 (2003) 719–729.
[5] A. Sawana, M. Adeolu, R.S. Gupta, Molecular signatures and phylogenomic analysis of the genus Burkholderia: proposal for division of this genus into the emended genus Burkholderia containing pathogenic organisms and a new genus Paraburkholderia gen. nov. harboring environmental species, Front. Genet. 5 (2014).
[6] A.P. Dobritsa, M. Samadpour, Transfer of eleven species of the genus Burkholderia to the genus Paraburkholderia and proposal of Caballeronia gen. nov. to accommodate twelve species of the genera Burkholderia and Paraburkholderia, Int. J. Syst. Evol. Microbiol. 66 (2016) 2836–2846.
[7] L. Lopes-Santos, D.B.A. Castro, M. Ferreira-Tonin, D.B.A. Corrêa, B.S. Weir, D. Park, L.M.M. Ottoboni, J.R. Neto, S.A.L. Destéfano, Reassessment of the taxonomic position of Burkholderia andropogonis and description of Robbsia andropogonis gen. nov., comb. nov, Antonie Van Leeuwenhoek (2017), https://doi.org/10.1007/s10482-017-0842-6.
[8] Q. Lin, Y. Lv, Z. Gao, L. Qiu, Pararobbsia silviterrae gen. nov., sp. nov., isolated from forest soil and reclassification of Burkholderia alpina as Pararobbsia alpina comb. nov, Int. J. Syst. Evol. Microbiol. 70 (2020) 1412–1420.
[9] Y.-S. Seo, J.Y. Lim, J. Park, S. Kim, H.-H. Lee, H. Cheong, S.-M. Kim, J.S. Moon, I. Hwang, Comparative genome analysis of rice-pathogenic Burkholderia provides insight into capacity to adapt to different environments and hosts, BMC Genomics 16 (2015) 349.
[10] O.O. Bochkareva, E.V. Moroz, I.I. Davydov, M.S. Gelfand, Genome rearrangements and selection in multi-chromosome bacteria Burkholderia spp, BMC Genomics 19 (2018) 965.
[11] C. Peeters, J.E.A. Zlosnik, T. Spilker, T.J. Hird, J.J. LiPuma, P. Vandamme, *Burkholderia pseudomultivorans* sp. nov., a novel *Burkholderia cepacia* complex species from human respiratory samples and the rhizosphere, Syst. Appl. Microbiol. 36 (2013) 483–489.
[12] E. Depoorter, E. De Canck, C. Peeters, A.D. Wieme, M. Cnockaert, J.E. Zlosnik, J. J. LiPuma, T. Coenye, P. Vandamme, Burkholderia cepacia complex taxon K: where to split? Front. Microbiol. 11 (2020) 1594.
[13] Y. Jin, J. Zhou, J. Zhou, M. Hu, Q. Zhang, N. Kong, H. Ren, L. Liang, J. Yue, Genome-based classification of Burkholderia cepacia complex provides new insight into its taxonomic status, Biol. Direct 15 (2020) 1–14.
[14] D.T.D. Kenna, D. Lilley, A. Coward, K. Martin, C. Perry, R. Pike, R. Hill, J.F. Turton, Prevalence of Burkholderia species, including members of Burkholderia cepacia complex, among UK cystic and non-cystic fibrosis patients, J. Med. Microbiol. 66 (2017) 490–501.
[15] K. Agnoli, S. Schwager, S. Uehlinger, A. Vergunst, D.F. Viteri, D.T. Nguyen, P. A. Sokol, A. Carlier, L. Eberl, Exposing the third chromosome of Burkholderia cepacia complex strains as a virulence plasmid, Mol. Microbiol. 83 (2012) 362–378.
[16] Q. Esmaeel, M. Pupin, N.P. Kieu, G. Chataigné, M. Béchet, J. Deravel, F. Krier, M. Höfte, P. Jacques, V. Leclère, Burkholderia genome mining for nonribosomal peptide synthetases reveals a great potential for novel siderophores and lipopeptides synthesis, Microbiol. Open. 5 (2016) 512–526.
[17] E. Bach, F.H. Sant'Anna, J.F. Magrich dos Passos, E. Balsanelli, V.A. de Baura, F. O. de Pedrosa, E.M. de Souza, L.M.P. Passaglia, Detection of misidentifications of species from the Burkholderia cepacia complex and description of a new member, the soil bacterium Burkholderia catarinensis sp. nov, Pathog. Dis. 75 (2017) ftx076.
[18] Z. Liu, H. Sheng, B.A. Okorley, Y. Li, F.L. Sossah, Comparative genomic analysis provides insights into the phylogeny, resistome, virulome, and host adaptation in the genus Ewingella, Pathogens. 9 (2020) 330.
[19] L. Zhou, T. Zhang, S. Tang, X. Fu, S. Yu, Pan-genome analysis of *Paenibacillus polymyxa* strains reveals the mechanism of plant growth promotion and biocontrol, Antonie Van Leeuwenhoek 113 (2020) 1539–1558.
[20] S. Datta, D. Saha, L. Chattopadhyay, B. Majumdar, Genome comparison identifies different bacillus species in a bast fibre-retting bacterial consortium and provides insights into pectin degrading genes, Sci. Rep. 10 (2020) 1–15.
[21] C. Hesse, F. Schulz, C.T. Bull, B.T. Shaffer, Q. Yan, N. Shapiro, K.A. Hassan, N. Varghese, L.D. Elbourne, I.T. Paulsen, Genome-based evolutionary history of Pseudomonas spp, Environ. Microbiol. 20 (2018) 2142–2159.
[22] H. Tettelin, D. Riley, C. Cattuto, D. Medini, Comparative genomics: the bacterial pan-genome, Curr. Opin. Microbiol. 11 (2008) 472–477.
[23] J. Zhou, H. Ren, M. Hu, J. Zhou, B. Li, N. Kong, Q. Zhang, Y. Jin, L. Liang, J. Yue, Characterization of *burkholderia cepacia* complex core genome and the underlying recombination and positive selection, Front. Genet. 11 (2020).
[24] P. Estrada-de Los Santos, M. Palmer, B. Chávez-Ramírez, C. Beukes, E. T. Steenkamp, L. Briscoe, N. Khan, M. Maluk, M. Lafos, E. Humm, Whole genome analyses suggests that Burkholderia sensu lato contains two additional novel genera (Mycetohabitans gen. nov., and Trinickia gen. nov.): implications for the

evolution of diazotrophy and nodulation in the Burkholderiaceae, Genes 9 (2018) 389.
[25] E. Bach, G.D.D.S. Seger, G.D.C. Fernandes, B.B. Lisboa, L.M.P. Passaglia, Evaluation of biological control and rhizosphere competence of plant growth promoting bacteria, Appl. Soil Ecol. 99 (2016), https://doi.org/10.1016/j.apsoil.2015.11.002.
[26] J.F.M. dos Passos, P.B. da Costa, M.D. Costa, G.R. Zaffari, G. Nava, J.I. Boneti, A.M. R. de Oliveira, L.M.P. Passaglia, Cultivable bacteria isolated from apple trees cultivated under different crop systems: Diversity and antagonistic activity against Colletotrichum gloeosporioides, Genet. Mol. Biol. 37 (2014) 560–572.
[27] F. Hayashi Sant'Anna, E. Bach, R.Z. Porto, F. Guella, E. Hayashi Sant'Anna, L.M. P. Passaglia, Genomic metrics made easy: what to do and where to go in the new era of bacterial taxonomy, Crit. Rev. Microbiol. 45 (2019) 182–200.
[28] A.P. Dobritsa, M. Samadpour, Reclassification of Burkholderia insecticola as Caballeronia insecticola comb. nov. and reliability of conserved signature indels as molecular synapomorphies, Int. J. Syst. Evol. Microbiol. 69 (2019) 2057–2063.
[29] P. Vandamme, B. Holmes, T. Coenye, J. Goris, E. Mahenthiralingam, J.J. LiPuma, J. R. Govan, Burkholderia cenocepacia sp. nov.—a new twist to an old story, Res. Microbiol. 154 (2003) 91–96.
[30] H. Xu, S. Qin, Y. Lan, M. Liu, X. Cao, D. Qiao, Y. Cao, Y. Cao, Comparative genomic analysis of Paenibacillus sp. SSG-1 and its closely related strains reveals the effect of glycometabolism on environmental adaptation, Sci. Rep. 7 (2017) 1–11.
[31] P. González-Torres, F. Rodríguez-Mateos, J. Antón, T. Gabaldón, Impact of homologous recombination on the evolution of prokaryotic core genomes, MBio. 10 (2019).
[32] J.J. Koehorst, J.C. Van Dam, R.G. Van Heck, E. Saccenti, V.A.M. Dos Santos, M. Suarez-Diez, P.J. Schaap, Comparison of 432 Pseudomonas strains through integration of genomic, functional, metabolic and expression data, Sci. Rep. 6 (2016) 1–13.
[33] J.E. Loper, K.A. Hassan, D.V. Mavrodi, E.W. Davis II, C.K. Lim, B.T. Shaffer, L.D. H. Elbourne, V.O. Stockwell, S.L. Hartney, K. Breakwell, Comparative genomics of plant-associated Pseudomonas spp.: insights into diversity and inheritance of traits involved in multitrophic interactions, PLoS Genet. 8 (2012) e1002784.
[34] L. Freschi, A.T. Vincent, J. Jeukens, J.-G. Emond-Rheault, I. Kukavica-Ibrulj, M.-J. Dupont, S.J. Charette, B. Boyle, R.C. Levesque, The Pseudomonas aeruginosa pan-genome provides new insights on its population structure, horizontal gene transfer, and pathogenicity, Genome Biol. Evol. 11 (2019) 109–120.
[35] S. Kunakom, A.S. Eustáquio, Burkholderia as a source of natural products, J. Nat. Prod. 82 (2019) 2018–2037.
[36] M. Mannaa, I. Park, Y.-S. Seo, Genomic features and insights into the taxonomy, virulence, and benevolence of plant-associated Burkholderia species, Int. J. Mol. Sci. 20 (2019) 121.
[37] E. Mahenthiralingam, T.A. Urban, J.B. Goldberg, The multifarious, multireplicon *Burkholderia cepacia* complex, Nat. Rev. Microbiol. 3 (2005) 144–156.
[38] Y. Luo, Y. Cheng, J. Yi, Z. Zhang, Q. Luo, D. Zhang, Y. Li, Complete genome sequence of industrial biocontrol strain *Paenibacillus polymyxa* HY96-2 and further analysis of its biocontrol mechanism, Front. Microbiol. 9 (2018) 1520.
[39] H.L. Spiewak, S. Shastri, L. Zhang, S. Schwager, L. Eberl, A.C. Vergunst, M. S. Thomas, Burkholderia cenocepacia utilizes a type VI secretion system for bacterial competition, MicrobiologyOpen. 8 (2019), e00774.
[40] T.-T. Tseng, B.M. Tyler, J.C. Setubal, Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology, BMC Microbiol. 9 (2009) 1.
[41] I.U. Haq, K. Graupner, R. Nazir, J.D. van Elsas, The genome of the fungal-interactive soil bacterium *Burkholderia terrae* BS001—A plethora of outstanding interactive capabilities unveiled, Genome Biol. Evol. 6 (2014) 1652–1668.
[42] P. Deng, X. Wang, S.M. Baird, K.C. Showmaker, L. Smith, D.G. Peterson, S. Lu, Comparative genome-wide analysis reveals that Burkholderia contaminans MS14 possesses multiple antimicrobial biosynthesis genes but not major genetic loci required for pathogenesis, MicrobiologyOpen. 5 (2016) 353–369.
[43] S. Spaepen, J. Vanderleyden, R. Remans, Indole-3-acetic acid in microbial and microorganism-plant signaling, FEMS Microbiol. Rev. 31 (2007) 425–448.
[44] M.S. Thomas, Iron acquisition mechanisms of the *Burkholderia cepacia* complex, Biometals. 20 (2007) 431–452.
[45] Y. Itoh, J.D. Rice, C. Goller, A. Pannuri, J. Taylor, J. Meisner, T.J. Beveridge, J. F. Preston, T. Romeo, Roles of pgaABCD genes in synthesis, modification, and export of the *Escherichia coli* biofilm adhesin poly-β-1, 6-N-acetyl-D-glucosamine, J. Bacteriol. 190 (2008) 3670–3680.
[46] D. Coil, G. Jospin, A.E. Darling, A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data, Bioinformatics. 31 (2014) 587–589.
[47] A. Bankevich, S. Nurk, D. Antipov, A.A. Gurevich, M. Dvorkin, A.S. Kulikov, V. M. Lesin, S.I. Nikolenko, S. Pham, A.D. Prjibelski, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing, J. Comput. Biol. 19 (2012) 455–477.
[48] D.H. Parks, M. Imelfort, C.T. Skennerton, P. Hugenholtz, G.W. Tyson, CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes, Genome Res. 25 (2015) 1043–1055.
[49] A. Gurevich, V. Saveliev, N. Vyahhi, G. Tesler, QUAST: quality assessment tool for genome assemblies, Bioinformatics. 29 (2013) 1072–1075.
[50] R. Aziz, D. Bartels, A. Best, M. DeJongh, T. Disz, R. Edwards, K. Formsma, S. Gerdes, E. Glass, M. Kubal, F. Meyer, G. Olsen, R. Olson, A. Osterman, R. Overbeek, L. McNeil, D. Paarmann, T. Paczian, B. Parrello, G. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, O. Zagnitko, The RAST server: rapid annotations using subsystems technology, BMC Genomics 9 (2008) 75.
[51] D. Laslett, B. Canback, ARAGORN, a program for the detection of transfer RNA and transfer-messenger RNA genes, Nucleic Acids Res. 32 (2004) 11–16.
[52] E.W. Sayers, M. Cavanaugh, K. Clark, J. Ostell, K.D. Pruitt, I. Karsch-Mizrachi, GenBank, Nucleic Acids Res. 48 (2020) D84–D86.
[53] A.C. Parte, J. Sardà Carbasse, J.P. Meier-Kolthoff, L.C. Reimer, M. Göker, List of prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ, Int. J. Syst. Evol. Microbiol. 70 (2020) 5607.
[54] T. Seemann, Prokka: rapid prokaryotic genome annotation, Bioinformatics. 30 (2014) 2068–2069.
[55] O. Tange, Gnu parallel-the command-line power tool, USENIX Mag. 36 (2011) 42–47.
[56] B. Contreras-Moreira, P. Vinuesa, GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis, Appl. Environ. Microbiol. 79 (2013) 7696–7701.
[57] A.J. Page, C.A. Cummins, M. Hunt, V.K. Wong, S. Reuter, M.T.G. Holden, M. Fookes, D. Falush, J.A. Keane, J. Parkhill, Roary: rapid large-scale prokaryote pan genome analysis, Bioinformatics. 31 (2015) 3691–3693.
[58] N.M. Chaudhari, V.K. Gupta, C. Dutta, BPGA- an ultra-fast pan-genome analysis pipeline, Sci. Rep. 6 (2016) 24373, https://doi.org/10.1038/srep24373.
[59] R Core Team, R, A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2021. http://www.r-project.org/.
[60] P. Vinuesa, L.E. Ochoa-Sánchez, B. Contreras-Moreira, GET_PHYLOMARKERS, a software package to select optimal orthologous clusters for phylogenomics and inferring pan-genome phylogenies, used for a critical geno-taxonomic revision of the genus Stenotrophomonas, Front. Microbiol. 9 (2018) 771.
[61] P. Siguier, J. Pérochon, L. Lestrade, J. Mahillon, M. Chandler, ISfinder: the reference centre for bacterial insertion sequences, Nucleic Acids Res. 34 (2006) D32–D36.
[62] C. Bertelli, M.R. Laird, K.P. Williams, S.F.U.R.C. Group, B.Y. Lau, G. Hoad, G. L. Winsor, F.S. Brinkman, IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets, Nucleic Acids Res. 45 (2017) W30–W35.
[63] D. Couvin, A. Bernheim, C. Toffano-Nioche, M. Touchon, J. Michalik, B. Néron, E. P. Rocha, G. Vergnaud, D. Gautheret, C. Pourcel, CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins, Nucleic Acids Res. 46 (2018) W246–W251.
[64] A.G. McArthur, N. Waglechner, F. Nizam, A. Yan, M.A. Azad, A.J. Baylay, K. Bhullar, M.J. Canova, G. De Pascale, L. Ejim, The comprehensive antibiotic resistance database, Antimicrob. Agents Chemother. 57 (2013) 3348–3357.
[65] H. Zhang, T. Yohe, L. Huang, S. Entwistle, P. Wu, Z. Yang, P.K. Busk, Y. Xu, Y. Yin, dbCAN2: a meta server for automated carbohydrate-active enzyme annotation, Nucleic Acids Res. 46 (2018) W95–W101.