Burkholderia kirstenboschensis sp. nov. nodulates papilionoid legumes indigenous to South Africa

Emma T. Steenkamp[a], Elritha van Zyl[a], Chrizelle W. Beukes[a], Juanita R. Avontuur[a], Wai Yin Chan[a], Marike Palmer[a], Lunghile S. Mthombeni[a], Francina L. Phalane[b], T. Karabo Sereme[a], Stephanus N. Venter[a]

[a]Department of Microbiology and Plant Pathology, Faculty of Natural and Agricultural Sciences, Forestry and Agricultural Biotechnology Institute, University of Pretoria, Pretoria, South Africa

[b]Agricultural Research Council (ARC), Plant Protection Research Institute, Pretoria, Gauteng, South Africa

Email addresses:

Emma.steenkamp@up.ac.za

Elritha.vanzyl@fabi.up.ac.za

Chrizelle.beukes@fabi.up.ac.za

Juanita.avontuur@fabi.up.ac.za

Annie.chan@fabi.up.ac.za

Marike.duplessis@fabi.up.ac.za

MthombeniL@dwa.gov.za

PhalaneF@arc.agric.za

Theodorah.Sereme@ZA.nestle.com

Fanus.venter@up.ac.za

Corresponding author: Stephanus N. Venter (Fanus.venter@up.ac.za)

Tel: +27 12 420 4100

Postal address: Department of Microbiology and Plant Pathology, Faculty of Natural and Agricultural Sciences, University of Pretoria, Lunnon Road, Hillcrest, Pretoria, Republic of South Africa, 0002

Abstract

Despite the diversity of *Burkholderia* species known to nodulate legumes in introduced and native regions, relatively few taxa have been formally described. For example, the Cape Floristic Region of South Africa is thought to represent one of the major centres of diversity for the rhizobial members of *Burkholderia*, yet only five species have been described from legumes occurring in this region and numerous are still awaiting taxonomic treatment. Here, we investigated the taxonomic status of 12 South African root-nodulating *Burkholderia* isolates from native papilionoid legumes (*Hypocalyptus coluteoides*, *H. oxalidifolius*, *H. sophoroides* and *Virgilia oroboides*). Analysis of four gene regions (16S rRNA, *recA*, *atpD* and *rpoB*) revealed that the isolates represent a genealogically unique and exclusive assemblage within the genus. Its distinctness was supported by all other aspects of the polyphasic approach utilized, including the genome-based criteria DNA-DNA hybridization (≥70.9%) and average nucleotide identities (≥96%). We accordingly propose the name *B. kirstenboschensis* sp. nov. for this taxon with isolate Kb15^T (=LMG 28727^T; =SARC 695^T) as its type strain. Our data showed that intraspecific genome size differences (≥0.81 Mb) and the occurrence of large DNA regions that are apparently unique to single individuals (16-23% of an isolate's genome) can significantly limit the value of data obtained from DNA-DNA hybridization experiments. Substitution of DNA-DNA hybridization with whole genome sequencing as a prerequisite for the description of *Burkholderia* species will undoubtedly speed up the pace at which their diversity are documented, especially in hyperdiverse regions such as the Cape Floristic Region.

Keywords: *Hypocalyptus*; *Virgilia*; beta-rhizobia; DNA-DNA hybridization; average nucleotide identity (ANI); *Burkholderia*

All sequences generated in this study were submitted to the European Nucleotide Archive (https://www.ebi.ac.uk) and received the following accession numbers: *atpD* (LN795810-LN795821); *rpoB* (LN795797-LN795808); *atpD* for *B. aspalathi* LMG 27731^T (LN795809); *rpoB* for *B. rhynchosiae* LMG 27174^T (LN795788), *B. phytofirmans* LMG 22487^T (LN795789), *B. phenazinium* LMG 2247^T (LN798790), *B. sartisoli* LMG 24000^T (LN795791), *B. ginsengisoli* LMG 24044^T (LN795792), *B. aspalathi* LMG 27731^T (LN795793), *B. bryophila* LMG 23644^T (LN795794), *B. megapolitana* LMG 23650^T (LN795795) and *B. terrae* LMG 23368^T (LN795796).

1 Introduction

Since the initial discovery that *Burkholderia tuberum* is capable of nodulating several species of the South African legume *Cyclopia* [25], many additional *Burkholderia* species have been demonstrated to efficiently establish the nitrogen-fixing symbiosis with a variety of legumes. To date, formal descriptions have been provided for 15 of these rhizobial or root-nodulating *Burkholderia* species [10, 11, 13, 29, 46, 64, 65, 76, 79]. Among them, only five species (including *B. tuberum*) have been described as symbionts of South African legumes. The other four species also nodulate legumes in the subfamily Papilionoideae, where both *B. sprentiae* and *B. dilworthii* were isolated from *Lebeckia ambigua* [22, 24], *B. rhynchosiae* originated from *Rhynchosia ferulifolia* [23] and *B. aspalathi* was isolated from the root nodules of *Aspalathus abietina* [47].

In comparison to the knowledge available for rhizobial genera in the Alphaproteobacteria, which have been studied for more than a century [80], our understanding of the diversity, distribution and evolution of the legume symbionts in the genus *Burkholderia* (class Betaproteobacteria, family Burkholderiaceae [30]) is limited. Research during the last decade has shown that legumes are often associated with a large diversity of *Burkholderia* symbionts, particularly where the root-nodulating bacteria of indigenous or endemic host species were explored [4, 5, 7, 8, 12, 53]. Phylogenetic analyses of various protein-coding genes have also shown that the association of these legumes with their *Burkholderia* symbionts likely represents an ancient partnership with a long and complex evolutionary history [5, 7]. For example, the South African Cape Floristic Region (CFR) with its indigenous legumes, some of which trace their origins to the early Oligocene [62], represent a unique centre of diversity for the rhizobial species of this genus [5, 42].

This study investigated the taxonomic status of 12 *Burkholderia* isolates that were originally isolated from the root nodules of the papilionoid hosts *Hypocalyptus coluteoides*, *H. oxalidifolius*, *H. sophoroides* and *Virgilia oroboides* growing in various locations in the CFR of South Africa [5]. Based on sequence analyses of the 16S ribosomal RNA (rRNA) and recombinase A (*recA*) genes, these isolates were suggested to represent a distinct group nested within the so-called "environmental clade" of *Burkholderia* [5]. This clade was recently suggested to represent a new genus (i.e., "*Caballeronia*" as proposed by Gyaneshwar et al. [36] or "*Paraburkholderia*" as proposed by Sawana et al. [61]; neither proposal have been formally implemented and a type species for this new taxon has not been identified). This proposed generic taxon includes most of the described rhizobia, as well as the plant associated species, but

excludes all clinically important *Burkholderia* species. The specific aims of this study were to (i) evaluate the conspecificity of the 12 South African isolates, (ii) evaluate their uniqueness within the overall genus and (iii) to provide a formal description of the taxon thus delineated. To achieve these aims, we utilized a polyphasic approach, where a range of phenotypic characters, phylogenetic information using Multi-Locus Sequence Analysis (MLSA) and genome-based criteria such as DNA-DNA hybridization and average nucleotide identities (ANI) [33, 39, 59] were used. To accomplish the latter, we also determined the whole genome sequences for two isolates of our newly recognized taxon.

2 Materials and Methods

2.1 Bacterial strains and growth conditions

The 12 *Burkholderia* isolates all originated from locations in the Western and Eastern Cape Provinces of South Africa. Three isolates (HC1.1be, HC1.1bc and HC1.1a2) associated with *H. sophoroides* were recovered from soil collected in Old du Toit's Kloof Pass, two isolates (RAU2b and RAU2d2) were recovered from *H. coluteoides* growing at the Storms River Bridge, and two isolates originated from *H. oxalidifolius* (RAU6.4d and RAU6.4f) occurring in the Fernkloof Nature Reserve. Five isolates (Kb2, Kb13, Kb14, Kb16 and Kb15^T) were recovered from nodules of *V. oroboides* occurring in the Kirstenbosch National Botanical Gardens. The capacity of these isolates to nodulate their original hosts and/or other hosts such as *Vigna unguiculata* (cowpea) or *Macroptilium atropurpureum* (siratro) has been demonstrated previously [5]. Isolate Kb15^T is also available from the Belgian Coordinated Collections of Microorganisms (Universiteit Gent, Belgium; LMG 28727^T) or from the South African Rhizobium

Collection (Pretoria, South Africa; SARC 695^T). For comparative purposes, *B. caledonica* strain LMG 19076^T, *B. fungorum* strain LMG 16225^T, *B. megapolitana* strain LMG 23650^T and *B. dilworthii* strain LMG 27173^T were also included in the study.

All isolates were routinely grown at 28 °C on Tryptone Yeast Extract Agar (TYA) enriched with $CaCl_2.2H_2O$ (0.088 g/l) or on Yeast Mannitol Agar (YMA). Isolates were preserved at -70 °C in 20% glycerol as part of the University of Pretoria's Rhizobium Culture Collection.

2.2 16S rRNA gene analysis and MLSA

This study utilized the previously determined sequences for the 16S rRNA and *recA* genes [5], together with those for the genes *atpD* and *rpoB* that respectively encode the beta subunits for ATP synthase and RNA polymerase. The *atpD* and *rpoB* gene fragments were amplified with primers *atpDF* (5' GAT CGT ACA GTG CAT CGG 3') and *atpDR* (5' ATC GTG CCG ACC ATG TAG 3') [3] and primers *RpoB*-1394F (5' TGG CGG AAA ACC AGT TCC GCG 3') and *RpoB*-2430R (5' AGC CGT TCC ACG GCA TGA ACG 3'), respectively. The latter primer set was designed based on publicly available *rpoB* sequences by making use of BioEdit v7.0.5.3 [37] and Primer3 [60]. These two primer sets targeted, respectively, 1200 base pair (bp) and 1000 bp regions of the *atpD* and *rpoB* genes (see Table S1 for details regarding PCR and cycling conditions).

PCR products were cleaned using polyethylene glycol precipitation [69] and sequenced in both directions using the original PCR primers for *rpoB* and specific sequencing primers for *atpD* [3]. For this purpose the ABI PRISM Big Dye Terminator version 3.1 Cycle Sequencing Kit (Applied Biosystems, CA) and an ABI 377 Automated Capillary

DNA sequencer (Applied Biosystems) were utilized. The *atpD* and *rpoB* sequences were then manually curated using ChromasLite v2.01 (Technelysium, Queensland, Australia) and BioEdit.

For 10 of the *Burkholderia* isolates examined in this study, single gene nucleotide datasets were compiled for the *atpD* and *rpoB* sequences generated here and the 16S rRNA and *recA* sequences determined previously (see Beukes et al. [5] for accession numbers). For isolates Kb15^{T} and RAU2d2, the relevant gene sequences used in the nucleotide datasets were obtained from the genome sequences of these bacteria (see below). These datasets also included the sequences for the recognized type strains of described *Burkholderia* species as they appear in the List of Prokaryotic Names with Standing in Nomenclature (LPSN) [26, 54] (www.bacterio.net). The 16S rRNA dataset was aligned with the online version of MAFFT (Multiple Alignment using Fast Fourier Transformation; http://mafft.cbrc.jp/alignment/server/) using the Q-INS-I strategy that takes secondary structure into account [38]. The nucleotide datasets for the protein coding genes were manually aligned according to the inferred amino acid sequences. A concatenated dataset was generated with SequenceMatrix [74].

The 16S rRNA dataset was analyzed using CLC bio Main Workbench v7.03 (QIAGEN, Århus, Denmark) to determine the percentage sequence similarity among the various taxa included. The four single gene datasets and the concatenated dataset were subjected to maximum parsimony and maximum-likelihood phylogenetic analyses. For the maximum parsimony analyses, MEGA6 [72] was utilized to determine the most parsimonious tree(s) by making use of the heuristic tree search algorithm with 1000 random addition-sequence replicates and Tree Bisection-Reconnection (TBR) branch

swapping. Branch lengths were calculated by means of the average pathway method [50]. For the maximum-likelihood analyses, PhyML v3.1 [35] or MEGA6 [72] were utilized with the best-fit substitution model parameters as indicated by MEGA5 or jModelTest v0.1.01 [28, 35, 56, 71]. The 16S rRNA data used the model of Tamura and Nei [70], the *atpD* data used the TIM2 "transitional" model [56] while the *rpoB*, *recA* and concatenated datasets all utilized the General Time Reversible model (GTR) [73]; in all cases these models incorporated gamma (G) correction of among site rate variation and a proportion of invariable sites (I). In these analyses the best of the Nearest-Neighbour Interchange (NNI) and Subtree-Pruning-Regrafting (SPR) search algorithms were used for tree searches, which were initiated using the best of five random starting trees. For both the maximum parsimony and maximum-likelihood analyses, branch support was estimated using non parametric bootstrap analysis [27] based on 1000 pseudoreplicates.

2.3 Phenotypic characterization

Growth characteristics were studied on TYA, YMA, Tryptone Soy Agar (TSA; Oxoid, England) and MacConkey Agar without crystal violet (MA; Merck, South Africa). Isolates were incubated aerobically at 28 °C for 3-4 days, except for the experiments with MA, in which case isolate growth was evaluated at 29 and 37 °C, both after 3-4 days of incubation [58]. Growth of the isolates at 28 °C in Tryptone Yeast Broth (Oxoid, England), containing either 0.5% (wt/vol) or 1% (wt/vol) NaCl, was also evaluated. Cell morphology, size and motility were determined using Zeiss Stereo and compound microscopes and Auxiovision version 4.8 software. The metabolic profiles for the various isolates were determined using API 20NE and API 20E strips (bioMérieux), as well as the Biolog GN2 MicroPlate system (Biolog). All of these tests were performed

according to the manufacturers' instructions, where incubation was performed at 28 °C and results recorded after 48 h. Colonies used for inoculating the various test substrates included in these tests were taken from 24 h to 48 h old cultures on TSA at 28 °C.

2.4 DNA-DNA hybridization

The DNA-DNA hybridization experiments included nine isolates (HC1.1a2, HC1.1be, RAU2b, RAU2d2, RAU6.4d, RAU6.4f, Kb2, Kb13 and Kb15^T) to span the diversity within the new taxon delineated here. The experiments also included the type strains for the species *B. fungorum* (LMG 16225^T), *B. caledonica* (LMG 19076^T), *B. megapolitana* (LMG 23650^T), *B. rhynchosiae* (LMG 27174^T) and *B. dilworthii* (LMG 27173^T) which were selected based on their high 16S rRNA similarity to the nine isolates (Fig 1; File S2). High-molecular-weight DNA was prepared from 48 h TSA cultures using the method of Pitcher et al. [55] with an additional RNase step as described by Cleenwerck et al. [16]. DNA-DNA hybridizations utilized a modified version (see File S1) of the protocol described by Goris et al. [33].

2.5 Genome sequencing of isolates Kb15^T and RAU2d2

High quality DNA for isolates Kb15^T and RAU2d2 was subjected to whole genome shotgun sequencing at the Central Analytical Facilities (CAF) of Stellenbosch University. For this purpose, 300 bp chemistry and the 318-chip Ion Torrent™ PGM (Thermo Fisher Scientific) platform were used to generate single-ended reads. For isolate RAU2d2, pair-ended read data with 7000 bp inserts were generated. Following removal of adaptor sequences and filtering out of reads and regions of low quality, using FASTQ Quality Trimmer (version 1.0.0) [6] and an in-house python script, the remaining data were

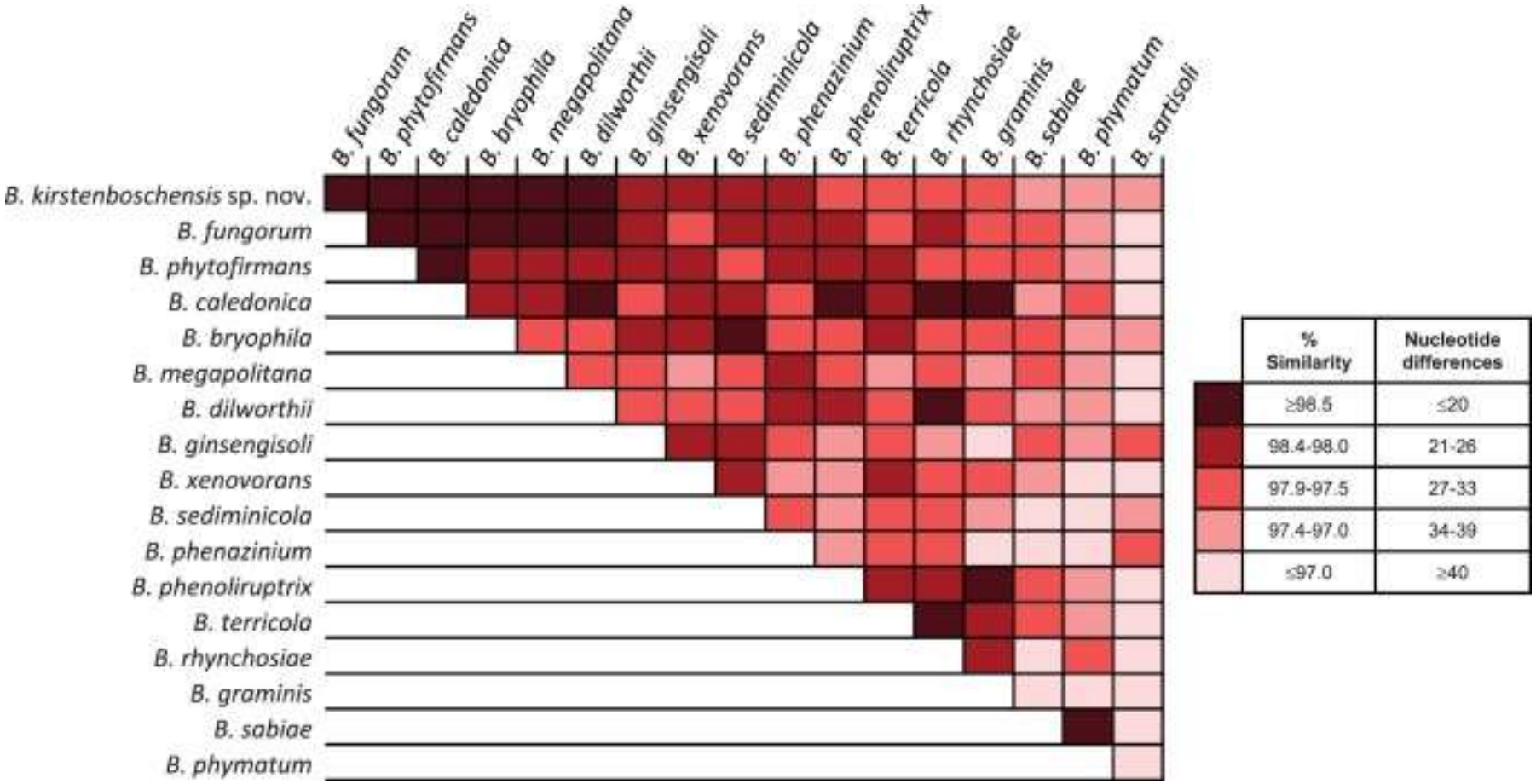


Figure 1. Percentage 16S rRNA gene sequence similarity between *B. kirstenboschensis* sp. nov. and the type strains of other species of *Burkholderia*. The analysis utilized the near complete sequence for the 16S rRNA gene where the alignment consisted of 1338 nucleotides and included only those species with ≥97% similarity to one or more of the *B. kirstenboschensis* isolates. The 12 isolates representing *B. kirstenboschensis* share 99.3-100% with an average of 99.7% 16S rRNA gene sequence similarity (see File S2 for the similarity matrix).

assembled. This was done with Newbler version 2.9 [45] using infrastructure at the Bioinformatics and Computational Biology Unit of the University of Pretoria.

The genomic regions shared between isolates Kb15^T and RAU2d2 were estimated using Mauve version 2.3.1 [20]. This analysis employed the progressive Mauve function, where a Match Seed weight of 11 bp with the default scoring matrix HOXD [14] was used. From the Mauve progressive backbone output file, the location of each segment that is homologous between the two genomes were extracted, and the sizes of the shared homologous region were summed. Open reading frame (ORF) predictions for the draft genomes were done using the Rapid Annotation using Subsystem Technology (RAST) server [2, 52]. For comparative purposes, the latter also included the genome sequences for *B. dilworthii* (LMG 27173^T) and *B. caledonica* (LMG 19076^T). Using RAST, ORFs were then annotated using the FIGfams database [2, 49].

2.6 Average nucleotide identities

The program, JSpecies [59], was used to calculate ANI-values for the genomes of the sequenced isolates and those of closely related species. Pair-wise comparisons between genomes were made by artificially sectioning genomes into fragments consisting of 1020 nucleotides and then comparing these sections to each other [33, 59] using the BLAST algorithm [1]. Only fragments that had more than 30% identity, as calculated across the entire segment, that aligned over more than 70% of the length of the fragments were considered homologous and taken into consideration [33]. The percentage identity values obtained for individual fragments were then averaged across all of the fragments to obtain the ANI-values between genomes [33].

3 Results

3.1 16S rRNA gene analysis and MLSA

Comparison of the 16S rRNA gene sequences revealed the presence of zero to nine nucleotide differences among the 12 *Burkholderia* isolates examined in this study (File S2). This corresponded to 16S rRNA gene sequence similarities ranging from 99.3 to 100%, where the average similarity among the 12 isolates was 99.7%. Based on the 16S rRNA data, these isolates were most similar to *B. fungorum* (98.9%), *B. phytofirmans* (98.8%), *B. caledonica* (98.7%), *B. bryophila* (98.7%), *B. megapolitana* (98.7%) and *B. dilworthii* (98.5%) (Fig. 1, and File S2). However, the 16S rRNA gene sequences for our 12 isolates also shared ≥97% similarity with those of the type strains of *B. ginsengisoli*, *B. xenovorans*, *B. sediminicola*, *B. phenazinium*, *B. phenoliruptrix*, *B. terricola*, *B. rhynchosiae*, *B. graminis*, *B. sabiae*, *B. phymatum* and *B. sartisoli* (Fig. 1, and File S2). The 16S rRNA gene sequences for the 12 *Burkholderia* isolates examined here were on average 95% similar to that of the type strain (ATCC 25416) of *B. cepacia* (type species of the genus).

Phylogenetic analysis of the respective 16S rRNA, *recA*, *atpD* and *rpoB* datasets allowed recovery of the 12 isolates examined in this study as a single clade (Figs. S1-S4). The concatenated four-gene dataset also clustered these isolates into a single well supported (100% maximum parsimony and maximum likelihood bootstrap support) clade (Fig. 2). The data clearly showed that this clade never included any of the species with which the 12 isolates share apparently high levels of 16S rRNA gene sequence similarity (Fig. 2; Figs. S1-S4). We therefore recognized the clade containing these 12 isolates as a new taxon in the genus (we proposed the name *B. kirstenboschensis* sp. nov.

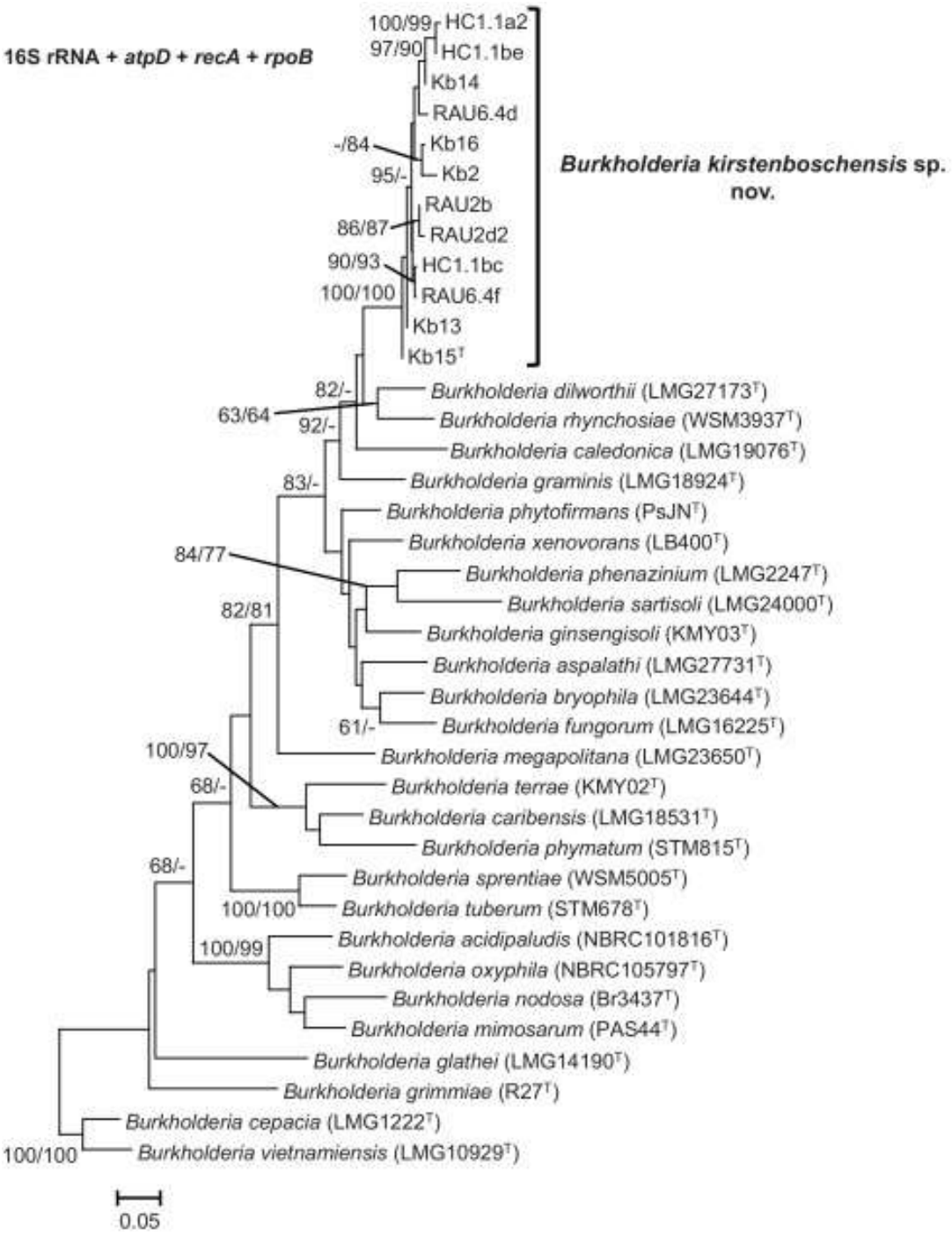


Figure 2. The maximum-likelihood phylogeny inferred from the concatenated data for the genes 16S rRNA, *atpD*, *recA* and *rpoB* of *Burkholderia kirstenboschensis* sp. nov. and related species. The sequences for *B. cepacia* and *B. vietnamiensis* were used for outgroup purposes. Similar groupings were obtained using maximum parsimony analysis. Bootstrap values (≥ 60%) are indicated at the branches in the order maximum-likelihood/parsimony. The type strains for the respective species are indicated in parentheses, and the scale bar indicates the number of nucleotide changes per site. See File S3 for GenBank accession numbers, and Figs. S1-S4 for the individual gene trees.

for the taxon [see the Discussion] and for simplicity sake we used this designation to refer to the clade in the sections below). Within the larger genus, the new taxon represents a sister taxon of the two South African species, *B. dilworthii* and *B. rhynchosiae*, while it is only distantly related to the other South African species (i.e., *B. sprentiae*, *B. tuberum* and *B. aspalathi*).

3.2 Phenotypic characterization

All 12 of the isolates examined represented Gram-negative, rod-shaped, motile bacteria (*ca*. 0.7- 2.3 µm). All isolates displayed growth on TYA, TSA and YMA after 24-48 h of incubation at 28 °C. However, colonies were more mucoid on TYA and YMA than on TSA. Only four isolates displayed growth on MA after 4 days of incubation at 29 °C, while growth on MA was observed for only one isolate (RAU2d2) following 4 days of incubation at 37 °C.

Metabolic profile analysis using API 20NE and API 20E strips and the Biolog GN2 MicroPlate system under standardized conditions, allowed differentiation of *B. kirstenboschensis* sp. nov. from the type strains of other *Burkholderia* species (Table 1), particularly its nearest phylogenetic neighbours *B. rhynchosiae* and *B. dilworthii*. These analyses did not include *B. cepacia* as it is only distantly related to our new taxon. All isolates displayed a positive reaction for the assimilation of adipic acid and a negative reaction for arginine dihydrolase which distinguished *B. kirstenboschensis* sp. nov. from *B. rhynchosiae* [24]. A positive reaction for all isolates for acetoin production and negative reactions for arginine dihydrolase, tryptophan deaminase and the oxidation of

Table 1. Phenotypic traits that differentiate the isolates of *B. kirstenboschensis* sp. nov. from the type strains of related *Burkholderia* species.

Phenotypic traits [a]	***Burkholderia* species** [b]												
	1	2	3	4	5	6	7	8	9	10	11	12	13
Isolation source	RN	RN	RH	RN	RN	RN	RN	RN	RN	RN	RN	RN	RN
Nitrate reduction	V	+	+	+	V	+	+	-	+	-	+	-	+
Activity of:													
Arginine dihydrolase	+	+ [w]	-	-	-	ND	-	+ [w]	-	+	-	-	-
Tryptophan deaminase	+	+	-	ND	ND	ND	-	-	ND	+ [w]	ND	-	ND
Urease	V	-	-	+	V	+	-	-	+	-	+	-	-
β-Galactosidase	V	+ [w]	+	+	-	+	+	+	+	+	+	+	+
Assimilation of:													
Adipic acid	+	+	-	-	-	+	-	-	-	-	-	-	ND
Capric acid	V	-	+	-	-	+	+	+ [w]	+	-	-	+	+
d-Glucose	+	+	+	+	+	+	+	+	+	+	+	+	+
d-Mannitol	+	+	+	+	+	+	+	+	+	+	+	+	+
d-Mannose	+	+	+	+	+	+	+	+	+	+	+	+	+
l-Arabinose	+	+	+	+	V	+	+	+	+	+	+	+	+
N-Acetylglucosamine	+	+	+	+	ND	+	+	+	+	+	+	+	+
Trisodium citrate	V	+	-	+	-	+	-	+	+	+ [w]	+	ND	+
Fermentation/oxidation of:													
d-Glucose	-	+ [w]	-	-	-	-	-	-	-	-	-	-	-

d-Mannitol	-	-	-	+	+	+	+	-	+	-	+	ND	ND
d-Sorbitol	-	-	-	+	+	+	+	-	+	-	+	ND	ND
Inositol	-	-	-	ND	+	+	ND	-	+ [w]	-	+	ND	ND
l-Arabinose	+ [w]	+ [w]	-	+	+	+	ND	-	+	-	+	ND	ND
Acetoin production	+	-	ND	ND	ND	ND	ND	+ [w]	ND	ND	ND	ND	ND
DNA G+C content (mol%)	61.8	61.4-61.9	62.0	63-65	64.8	62.8	62.1	61.2	64.5	61.6	64.2-65.7	62.8	60.1

[a] The results of the various properties were coded as follows: RN, root nodule; RH, rhizosphere; + all strains positive; +[w], weak positive; − all strains negative; V, variable; ND, no data available.

[b] Isolates and strains are numbered as follows: 1, *B. kirstenboschensis* sp. nov. (n=12; this study); 2, *B. dilworthii* LMG 27173^T [24]; 3, *B. caledonica* LMG 19076^T [24]; 4, *B. diazotrophica* LMG 26031^T [64]; 5, *B. mimosarum* LMG 23256^T [11, 13]; 6, *B. nodosa* LMG 23741^T [11]; 7, *B. phymatum* LMG 21445^T [10, 76]; 8, *B. rhynchosiae* WSM3937^T [24]; 9, *B. sabiae* LMG 24235^T [10]; 10, *B. sprentiae* WSM5005^T [22]; 11, *B. symbiotica* JPY345^T [65]; 12, *B. tuberum* LMG 21444^T [24]; and 13, *B. aspalathi* LMG27731^T [47].

D-glucose distinguished *B. kirstenboschensis* sp. nov. from *B. dilworthii* [24]. The complete results of the phenotypic tests are presented in the species description.

3.3 DNA-DNA hybridization

The taxa selected for the DNA-DNA hybridization experiments were selected based on the results for the 16S rRNA sequence comparisons, combined with the MLSA and ANI results (see below). This included three (*B. caledonica*, *B. megapolitana* and *B. fungorum*) of the five species with the highest 16S rRNA sequence similarity (≥98.7%) to *B. kirstenboschensis* sp. nov. (i.e., *B. phytofirmans* and *B. bryophila* were not included as they shared <92% ANI with *B. kirstenboschensis* sp. nov.). In addition to these three species, the DNA-DNA hybridization experiments also included *B. rhynchosiae* and *B. dilworthii*, which appears to be the closest known relatives of *B. kirstenboschensis* sp. nov.

Values between 70.9 (±3.7)% and 101.4 (±7.2)% were obtained for the intra-species DNA-DNA hybridization experiments conducted with the various isolates of *B. kirstenboschensis* sp. nov. (Fig. 3). The only exception was for the hybridizations involving isolate RAU2d2, specifically where the DNA from this isolate was immobilized on the microplate. This caused the variation observed for reciprocal reactions to fall outside the accepted limits of the method [34]. For example, the respective hybridization reactions between Kb15^T and RAU2d2 generated an average value of 73% when Kb15^T was immobilized on the microplate and RAU2d2 used as the probe, and an average value of 61.5% for the reciprocal reaction (i.e., when RAU2d2 was immobilized and Kb15^T used as the probe). The average similarity for this pair of strains was therefore 67.2%, with a difference of 11.5% between reciprocals. This discrepancy

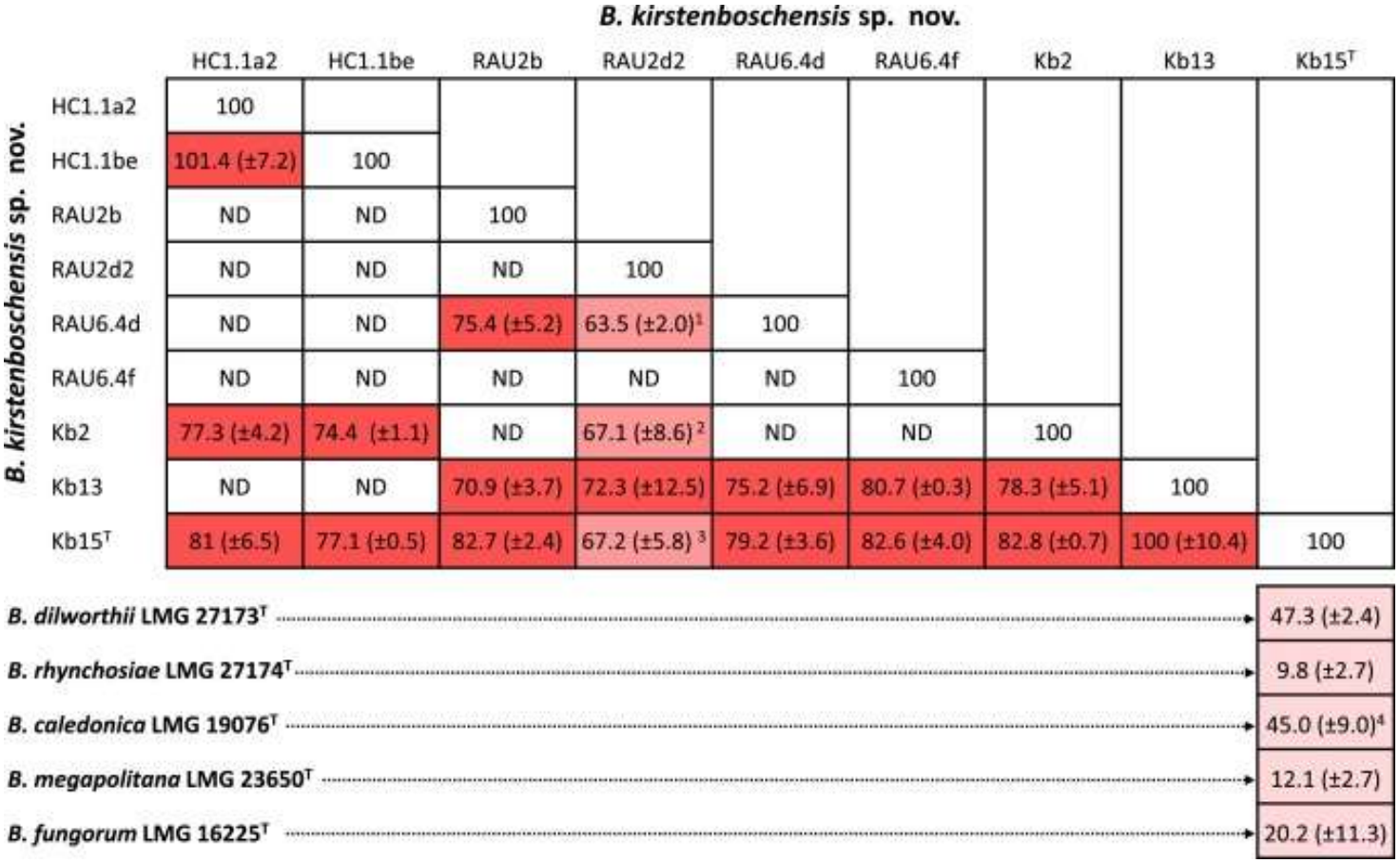


Figure 3. Percentage DNA-DNA hybridization values amongst isolates of *B. kirstenboschensis* sp. nov. and the type strains of other *Burkholderia* species. The values in parentheses indicate the difference between the reciprocal reactions divided by 2. Where RAU2d2 with its larger genome was used as the probe in these experiments, the values were typically higher (experiments with such reciprocal reactions values are indicated with Arabic numerals where [1] = 58.5 and 75.6%; [2] = 59.7 and 84.8%; [3] = 61.5 and 73.0%). The same scenario was encountered when Kb15^T was used as probe against the smaller genome of isolate LMG 19076^T of *B. caledonica* ([4] = 36.0 and 54.0%).

between the reciprocal reactions also generated average values below 70% when RAU2d2 was hybridized with RAU6.4d and Kb2. In all of these cases, however, the use of RAU2d2 as the probe generated hybridization values exceeding 70%. The hybridization experiments involving isolates RAU6.4d and Kb13 followed the same trend. Values below 50% were obtained for the experiments involving *B. kirstenboschensis* sp. nov. isolate Kb15^T and the type strains for any one of *B. fungorum*, *B. caledonica*, *B. megapolitana*, *B. rhynchosiae* and *B. dilworthii* included in the comparison.

3.4 Genome sequences for isolates Kb15^T and RAU2d2

A total of 5 010 566 single-ended reads were generated for *B. kirstenboschensis* sp. nov. isolate Kb15^T. After application of the quality filter, 4 190 410 reads with an average length of 219 bases were assembled into a draft genome consisting of 8 342 886 bases. The draft genome assembly of isolate Kb15^T consisted of 973 contigs with 109x coverage, and it had an N50 of 15 997 bases and G + C content of 61.84 mol%. The draft genome assembly sequence is available from the National Centre for Biotechnology Information (NCBI: http://www.ncbi.nlm.nih.gov/) under the accession number JRZC00000000.

A total of 4 733 415 reads were generated for *B. kirstenboschensis* sp. nov. isolate RAU2d2. After quality filtering, the remaining 3 199 998 pair-ended reads (average length of 214 bases) were assembled into a draft genome consisting of 9 150 776 bases. The draft genome sequence of isolate RAU2d2 consisted of 1 182 contigs with 74x coverage, N50 of 14 028 bases and G + C content of 61.64 mol%. The draft genome

sequence for the isolate was submitted to the NCBI database under the accession number JRTU00000000.

Analysis of the shared genomic regions between isolates Kb15^T and RAU2d2 showed that each strain contained long stretches of apparently unique DNA. Of the 8 342 886 bases in Kb15^T, only about 84% aligned to the genome of isolate RAU2d2 with the remaining 16% (approximately 1.3 Mb) not shared with RAU2d2. Similarly, only 77% of the RAU2d2 genome sequence aligned to that of Kb15^T, with the remaining 23% (approximately 2.1 Mb) unique to RAU2d2.

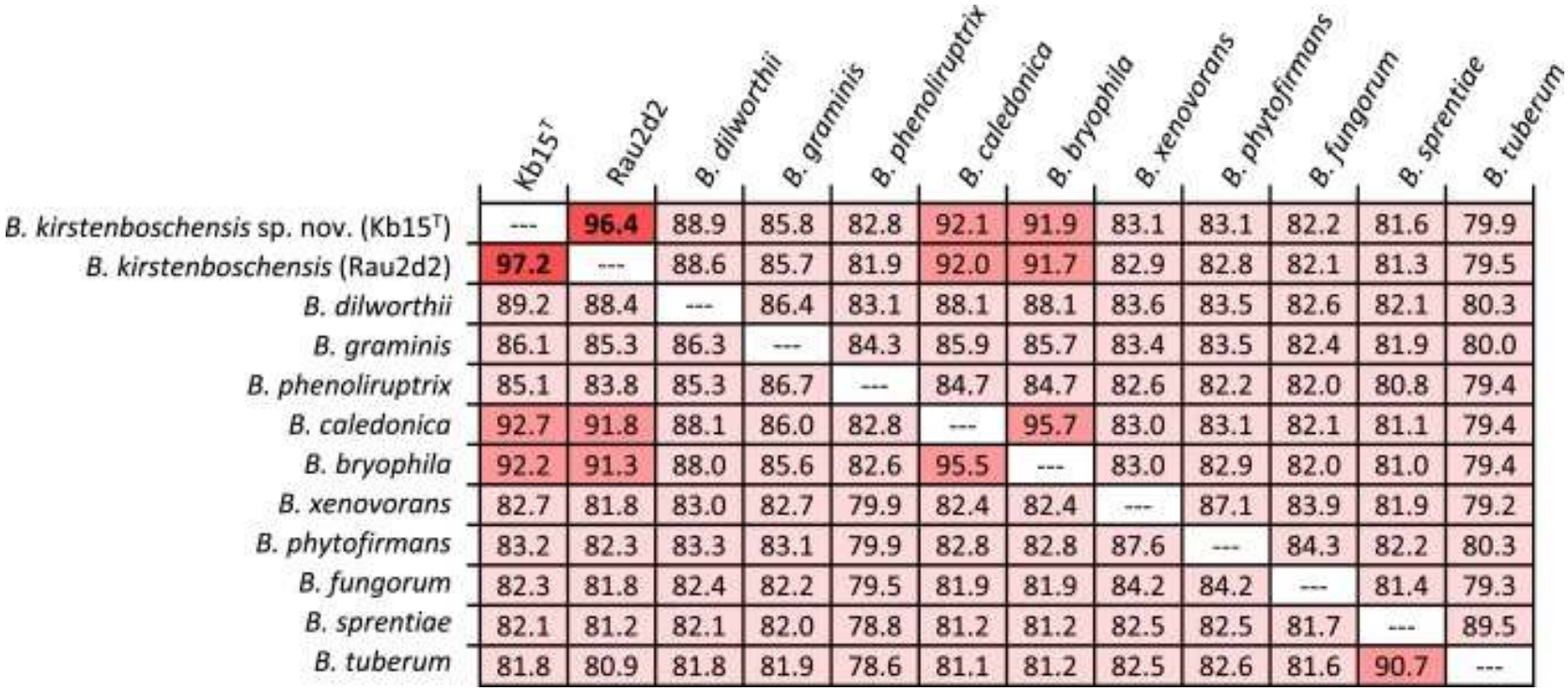
The genomes of isolates Kb15^T and RAU2d2 were predicted to respectively encode 8688 and 9869 ORFs. In both cases, about a third of the ORFs could be analysed using the FIGfams database (the remaining ORFs were either annotated as "hypothetical genes" or had annotations for which suitable Subsystem categories are not available; see Fig. S5). Comparison of the annotations for the ORFs of isolates Kb15^T and RAU2d2 revealed putative differences in the metabolic capabilities and functional pathways they encode (Table 2). These differences included the presence of the MazEF toxin-antitoxin system in RAU2d2 with its larger genome [82]. Comparison of the FIGfam Subsystems categories identified from the genomes of *B. kirstenboschensis* sp. nov., *B. dilworthii* and *B. caledonica* further allowed identification of processes or pathways potentially unique to *B. kirstenboschensis* sp. nov. (Table 2). For example, isolates Kb15^T and RAU2d2 both appear to encode the full pathway for tricarballylate utilization [43].

Table 2. FIGfams Subsystem categories predicted to be differentially encoded by the genomes of *B. kirstenboschensis* sp. nov., *B. dilworthii* LMG 27173^T and *B. caledonica* LMG 19076^T.

Subsystem categories [a]	***B. kirstenboschensis* sp. nov.**		***B. dilworthii***	***B. caledonica***
	Kb15^T	**RAU2d2**		
Polymorphic categories in B. kirstenboschensis sp. nov.				
Arsenic resistance repressor	-	+	-	-
Toluene degradation				
- Toluenesulfonate zinc-independent alcohol dehydrogenase	-	+	-	-
Toxin-antitoxin MT1 system				
- Toxin 1, PIN domain	-	+	-	-
- *vapC*, Toxin-antitoxin replicon stabilization systems	-	+	-	-
MazEF-like toxin-antitoxin system (All genes)				
- *chpS*, programmed cell death antitoxin	-	+	-	-
- *pemI*, programmed cell death antitoxin	-	+	-	-
- *chpB*, programmed cell death toxin	-	+	-	-
Succinate dehydrogenase				
- Fumarate reductase flavoprotein subunit	+	+	-	-
- Flavoprotein(N-terminal) FAD-dependent oxidoreductase	+	-	-	-
Categories that distinguish B. kirstenboschensis sp. nov.				
Tricarballylate utilization (full pathway)				
- *tcuA*	+	+	-	-
- *tcuB*	+	+	-	-

- *tcuC*	+	+	-	-
- *tcuR*	+	+	-	-
Cinnamic acid degradation				
- *mphA*	+	+	-	-
- *mphE*	+	+	-	-
- *mph* operon transcriptional activator	+	+	-	-
Tellurite resistance				
- *tehA*	+	+	-	+
Zinc resistance				
- Response regulator of zinc σ-54-dependent two-component system	+	+	-	+
Denitrification				
- Respiratory nitrate reductase alpha chain	-	-	+	-
- Respiratory nitrate reductase beta chain	-	-	+	-
- Respiratory nitrate reductase delta chain	-	-	+	-
- Respiratory nitrate reductase gamma chain	-	-	+	-
Cyanate hydrolysis				
- Cyanate hydratase	-	-	-	+
- *Cyn* operon transcriptional activator	-	-	-	+
Dye-decolorizing peroxidase DyP	-	-	-	+

[a] The FIGfams Subsystem categories was described by Meyer et al. [49]. FIGfams represent sets of homologous protein sequences where each set likely represent isofunctional homologs. Each Subsystem category reflects a process/phenotype (e.g., Tricarballylate utilization) that is underpinned by a set of genes (e.g., *tcuA*, *tcuB*, *tcuC* and *tcuD*) whose products contribute to, or, make up the process/ phenotype; + genes present; - genes absent.



	Kb15^T	Rau2d2	B. dilworthii	B. graminis	B. phenoliruptrix	B. caledonica	B. bryophila	B. xenovorans	B. phytofirmans	B. fungorum	B. sprentiae	B. tuberum
B. kirstenboschensis sp. nov. (Kb15^T)	---	**96.4**	88.9	85.8	82.8	92.1	91.9	83.1	83.1	82.2	81.6	79.9
B. kirstenboschensis (Rau2d2)	**97.2**	---	88.6	85.7	81.9	92.0	91.7	82.9	82.8	82.1	81.3	79.5
B. dilworthii	89.2	88.4	---	86.4	83.1	88.1	88.1	83.6	83.5	82.6	82.1	80.3
B. graminis	86.1	85.3	86.3	---	84.3	85.9	85.7	83.4	83.5	82.4	81.9	80.0
B. phenoliruptrix	85.1	83.8	85.3	86.7	---	84.7	84.7	82.6	82.2	82.0	80.8	79.4
B. caledonica	92.7	91.8	88.1	86.0	82.8	---	95.7	83.0	83.1	82.1	81.1	79.4
B. bryophila	92.2	91.3	88.0	85.6	82.6	95.5	---	83.0	82.9	82.0	81.0	79.4
B. xenovorans	82.7	81.8	83.0	82.7	79.9	82.4	82.4	---	87.1	83.9	81.9	79.2
B. phytofirmans	83.2	82.3	83.3	83.1	79.9	82.8	82.8	87.6	---	84.3	82.2	80.3
B. fungorum	82.3	81.8	82.4	82.2	79.5	81.9	81.9	84.2	84.2	---	81.4	79.3
B. sprentiae	82.1	81.2	82.1	82.0	78.8	81.2	81.2	82.5	82.5	81.7	---	89.5
B. tuberum	81.8	80.9	81.8	81.9	78.6	81.1	81.2	82.5	82.6	81.6	90.7	---

Figure 4. Pairwise comparison of the percentage Average Nucleotide Identity (ANI) shared amongst *B. kirstenboschensis* sp. nov. and related *Burkholderia* species for which whole genome sequences are available. In all these cases, the analyzed genomes represented those for the type strains of the various species available from public domain databases (GenBank Accession numbers GCA_000472525.1, GCA_000172415.1, GCA_000416445.1, GCA_000685095.1, GCA_000756045.1, GCA_000020125.1, GCA_000685055.1 and GCA_000473465.1, and Gold Project ID Gp0001007). The only exceptions were *B. kirstenboschensis* isolate RAU2d2 (this study; JRTU00000000) and *B. bryophila* for which the sequence for strain BR3495a were utilized (GCA_000383275.1). The analyses were conducted with JSpecies [59], which involved artificial shearing of the respective genomes into 1020-nucleotide fragments that were then compared using the BLAST algorithm [33, 59]. Reciprocal search results are indicated above and below the diagonal.

3.5 Average nucleotide identities

The ANI-values for *B. kirstenboschensis* sp. nov. isolates Kb15^T and RAU2d2 exceeded 96% (Fig. 4). The ANI-values estimated from all experiments involving comparisons with other *Burkholderia* species (including the closely related *B. dilworthii*) were all well below 90% (i.e., the accepted cut off value for species is ≥96%; 33, 59). The only exceptions were the comparisons involving both isolates Kb15^T and RAU2d2 with *B. caledonica* and *B. bryophila*, which generated values of 91.3-92.7%.

4 Discussion

The results of this study indicated that the 12 *Burkholderia* isolates examined represent a unique and exclusive group of rhizobial individuals. This group was unlike any known *Burkholderia* species and was recovered as monophyletic using all four of the housekeeping genes, irrespective of whether the gene sequences were concatenated or analyzed individually. Such genealogical exclusivity and concordance among multiple independent loci are both hallmarks of a species [19, 66]. The 12 *Burkholderia* isolates examined were therefore regarded as conspecific (i.e., they represent members of the same species), because all of the isolates were more closely related to each other than to any other known isolate or species of *Burkholderia*. We accordingly recognized the group of 12 *Burkholderia* isolates examined here as a new species in the genus.

The existence of the newly proposed species taxon was supported by all other aspects of the polyphasic approach utilized to delineate this taxon. These included intra-specific DNA-DNA hybridization and ANI-values that exceeded 70% and 96% respectively. All comparisons involving isolates of the new taxon with any other *Burkholderia* species

yielded values much lower than those typically used as the lower limits for demarcating species [31, 39, 59, 68]. The high level of genetic cohesion observed among isolates of the new taxon was also reflected at the phenotypic level, as they exhibited similar metabolic profiles. Several of the metabolic properties investigated also allow differentiation of the new taxon from closely related South African species capable of nodulating legumes (i.e., *B. rhynchosiae* and *B. dilworthii*) [24]. The name *B. kirstenboschensis* sp. nov. was accordingly proposed for this new assemblage of root-nodule symbionts from *Hypocalyptus* spp. and *V. oroboides* growing in the CRF of South Africa.

Despite the obligatory inclusion of 16S rRNA sequence information in most descriptions of bacterial species [31, 68], the value of this gene for taxonomically defining new or unique *Burkholderia* species is exceedingly limited. As currently circumscribed, the genus includes roughly 88 species of which 17 have 16S rRNA sequences that are ≥97% similar to one or more of the *B. kirstenboschensis* sp. nov. isolates examined in this study. These 17 species include ecologically diverse taxa ranging from root-nodulaters (e.g., *B. dilworthii* and *B. rhynchosiae*) [23, 24] and moss-associated species such as *B. megapolitana* and *B. bryophila* [77] through to environmental species such as *B. xenovorans* [32]. In fact, the species with 16S rRNA sequences most similar to that of *B. kirstenboschensis* sp. nov. is *B. fungorum*, which in addition to having been reported from root nodules has also been recovered from fungal and clinical specimens [17, 18]. Even in combination with ecological data, 16S rRNA sequences thus have little value in diagnosing these species. Even based on a higher similarity value of ≥98.7% [67], five of these species were still more similar to one or more of the *B. kirstenboschensis* sp. nov. isolates. The limited potential of this marker for resolving *Burkholderia* species has

been noted before [75] and our data support suggestions that the absolute necessity of 16S rRNA information for describing species should be revised and/or replaced with more robust and taxonomically meaningful data [15, 51, 59, 75].

The results of our study showed that the members of *B. kirstenboschensis* sp. nov. can differ markedly in terms of the size of their genomes. Although the genomes of the two isolates examined differed by approximately 0.81 Mb, such intra-specific differences are not uncommon among bacteria [41, 44, 63]. However, both the *B. kirstenboschensis* sp. nov. isolates appeared to be characterized by relatively large amounts of seemingly unique DNA. As much as 2.1 Mb (23%) of the DNA of isolate RAU2d2 did not align to the genome sequence of isolate Kb15^T, while 1.3 Mb (16%) of the Kb15^T genome did not align to that of RAU2d2. This lack of conservation in genome content at the intraspecific level thus suggest that the unique genomic regions represent part of the taxon's accessory or dispensable genome [40, 48]. For these isolates, the remaining 77-84% of their genome content likely comprises the core genome of the species [40, 48], which is similar to what has been reported for species such as *Streptococcus pyogenes* [41], *Neisseria meningitides* [63] and *Pantoea ananatis* [21].

Our findings further indicate that genome size differences can have a significant effect on the outcome of DNA-DNA hybridization experiments. This applies specifically to the associated differences between reciprocal reactions, which is consistent with what has been suggested before [9]. Although most of the DNA-DNA hybridization values estimated for *B. kirstenboschensis* sp. nov. were within the range of what is expected for well-delineated species, the values obtained with isolate RAU2d2 with its large 9.1 Mb genome were typically problematic (see Fig. 3). For example, hybridizations using DNA

from isolate RAU2d2 as probe yielded higher values than when a smaller genome was used as the probe in the reverse reaction. Likewise, Goris et al. [33] reported a significant linear correlation between the difference in reciprocal DNA-DNA hybridization values and the difference in the percentage of conserved DNA between two genomes. Although they utilized an extensive set of species and isolates, none of their experimental sets were directly comparable to those reported for *B. kirstenboschensis* sp. nov. in terms of genome size differences within the same species. We therefore concur with their conclusion that in most cases "DNA-DNA hybridization is too coarse a method (i.e. the experimental error is too high) to reveal subtle differences in genome size between strains", but emphasize that large genome differences among conspecifics can significantly influence the outcome of DNA-DNA hybridization experiments.

Taken together the findings presented in this study suggest that the criteria for delineating and describing new *Burkholderia* species require considerable refinement and/or revision. This echoes previous calls for an overhaul of the current standards for species descriptions [57, 78]. By making use of *B. kirstenboschensis* sp. nov. we have clearly illustrated the consequences of substantial genome size differences for DNA-DNA hybridization experiments. The prerequisite inclusion of DNA-DNA hybridization data therefore has little scientific merit in this case, especially in view of the current explosion in genome information for *Burkholderia* and other bacteria (e.g., see Whitman et al. [81] and the Genomes Online Database [https://gold.jgi-psf.org/]). In fact, our application of ANI as a genomic metric based on shared genome content, was more useful for showing genomic cohesion in *B. kirstenboschensis* sp. nov. This is in agreement with the fact that ANI is widely regarded to robustly reveal genomic

coherence among members of the same species [39, 51, 59]. Although current gene ontology databases are still somewhat limited (e.g., functions have not been linked to all known gene sequences and new genomes often contain gene sequences new to Science), our data suggests that the analysis of the genomic components unique to species and specific isolates likely holds valuable clues regarding their repertoires of biological capabilities under natural conditions. Substitution of DNA-DNA hybridization data with the use of genome sequence data as a prerequisite for the description of *Burkholderia* species will thus not only speed up the pace at which bacterial diversity is documented, but also allow access to the information that provides the genetic "blue print" for the species.

4.1 Description of *B. kirstenboschensis* sp. nov.

Burkholderia kirstenboschensis (kir.sten.bosch.en'sis. N.L. fem. adj. *kirstenboschensis* of Kirstenbosch, a botanical garden in South Africa, from where the type strain was isolated).

Cells are Gram-negative, motile, straight rods (*ca*. 0.7 x 2.3 μm). Positive for oxidase and catalase. Colonies are round, creamy white to yellowish and 1-2 mm in size on TYA. Isolates from *Virgilia oroboides* have simple rough surfaces, while other isolates have smooth surfaces. Growth occurs on TYA, TSA and YMA at 28 °C. Positive for growth in Tryptone Yeast Broth with 0.5% and 1% NaCl. Positive for the production of acetoin and arginine dihydrolase and the hydrolysis of Tween 40 and Tween 80. Positive for the assimilation of D-glucose, L-arabinose, D-mannose, D-mannitol, *N*-acetyl-glucosamine, potassium gluconate, adipic acid, malic acid and phenylacetic acid. Additionally, positive reactions are recorded for the assimilation of adonitol, D-arabitol, D-fructose, L-fructose, D-galactose, α-D-glucose, L-rhamnose, D-sorbitol, D-trehalose, methyl

pyruvate, mono-methyl-succinate, acetic acid, *cis*-aconitic acid, formic acid, D-gluconic acid, D-glucosaminic acid, α-hydroxy butyric acid, ρ-hydroxy phenylacetic acid, D,L-lactic acid, propionic acid, quinic acid, D-saccharic acid, succinic acid, bromo succinic acid, succinamic acid, D-alanine, L-alanine, L-aspargine, L-aspartic acid, L-glutamic acid, L-proline, L-pyroglutamic acid, D-serine, L-serine and L-threonine. Negative reactions are recorded for arginine dihydrolase, lysine decarboxylase, ornithine decarboxylase, H_2S production, tryptophan deaminase, indole production, β-glucosidase activity, the assimilation of D-maltose and the fermentation/oxidation of D-glucose, the oxidation of D-mannitol, inositol, D-sorbitol, D-sucrose and amygdalin. Additional negative reactions are recorded for the assimilation of α-cyclodextrin, dextrin, glycogen, *N*-acetyl-D-galactosamine, i-erythritol, gentiobiose, α-D-lactose, maltose, D-melibiose, β-methyl-D-glucoside, D-psicose, turanose, xylitol, itaconic acid, α-keto valeric acid, glycyl-L-aspartic acid, L-ornithine, uridine, thymidine, phenyethylamine, putrescine, 2,3-butanediol and glucose-1-phosphate.

The type strain, Kb15^T (=LMG 28727^T; =SARC 695), was isolated from root nodules of *Virgilia oroboides* from the Western Cape of South Africa. The DNA G + C content of the type strain is 61.8 mol% as determined by whole genome sequencing.

Acknowledgements

We dedicate this paper to the late Ms. Jacomina Bloem; who was responsible for the collection of the *Virgilia* isolates from Kirstenbosch National Botanical Garden in her capacity as researcher at the Agricultural Research Council of South Africa. We are grateful to the South African National Research Foundation (NRF) and the Department of Science and Technology for funding through their Centre of Excellence programme.

We would also like to acknowledge the Bioinformatics and Computational Biology Unit of the University of Pretoria, for the use of their computational infrastructure.

References

[1] Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389-3402.

[2] Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., Zagnitko, O. (2008) The RAST Server: Rapid Annotations using Subsystems Technology. BMC Genomics 9, 75. doi:10.1186/1471-2164-9-75.

[3] Baldwin, A., Mahenthiralingam, E., Thickett, K.M., Honeybourne, D., Maiden, M. C., Govan, J.R., Speert, D.P., LiPuma, J.J., Vandamme, P., Dowson, C.G. (2005) Multilocus sequence typing scheme that provides both species and strain differentiation for the *Burkholderia cepacia* complex. J. Clin. Microbiol. 43, 4665-4673.

[4] Barrett, C.F., Parker, M.A. (2006) Coexistence of *Burkholderia*, *Cupriavidus*, and *Rhizobium* sp. nodule bacteria on two *Mimosa* spp. in Costa Rica. Appl. Environ. Microbiol. 72, 1198–1206.

[5] Beukes, C.W., Venter, S.N., Law, I.J., Phalane, F.L., Steenkamp, E.T. (2013) South African Papilionoid legumes are nodulated by diverse *Burkholderia* with unique nodulation and nitrogen-fixation loci. PLoS ONE 8(7): e68406. doi:10.1371/journal.pone.0068406

[6] Blankenberg, D., Gordon, A., Von Kuster, G., Coraor, N., Taylor, J., Nekrutenko, A. and the Galaxy Team (2010) Manipulation of FASTQ data with Galaxy. Bioinformatics 26, 1783-1785.

[7] Bontemps, C., Elliott, G.N., Simon, M.F., Dos Reis Júnior, F.B., Gross, E., Lawton, R.C., Neto, N.E., de Fátima Loureiro, M., de Faria, S.M., Sprent, J.I., James, E.K., Young, J.P.W. (2010) *Burkholderia* species are ancient symbionts of legumes. Mol. Ecol. 19, 44–52.

[8] Bournaud, C., de Faria, S.M., dos Santos, J.M.F., Tisseyre, P., Silva, M., Chaintreuil, C., Gross, E., James, E.K., Prin, Y., Moulin, L. (2013) *Burkholderia* species are the most common and preferred nodulating symbionts of the Piptadenia group (Tribe Mimoseae). PLoS ONE 8(5): e63478. doi: 10.1371/journal.pone.0063478

[9] Brenner, D.J. (1978) Characterization and clinical identification of Enterobacteriaceae by DNA hybridization. Prog. Clin. Pathol. 7,71-117

[10] Chen, W-M., de Faria, S.M., Chou, J-H., James, E.K., Elliott, G.N., Sprent, J.I., Bontemps, C., Young, J.P.W., Vandamme, P. (2008) *Burkholderia sabiae* sp. nov., isolated from root nodules of *Mimosa caesalpiniifolia*. Int. J. Syst. Evol. Microbiol. 58, 2174-2179.

[11] Chen, W-M., de Faria, S.M., James, E.K., Elliott, G.N., Lin, K-Y., Chou, J-H., Sheu, S-Y., Cnockaert, M., Sprent, J.I., Vandamme, P. (2007) *Burkholderia nodosa* sp. nov., isolated from root nodules of the woody Brazilian legumes *Mimosa bimucronata* and *Mimosa scabrella*. Int. J. Syst. Evol. Microbiol. 57, 1055-1059.

[12] Chen, W-M., de Faria, S.M., Straliotto, R., Pitard, R.M., Simões-Araùjo, J.L., Chou, J-H., Chou, Y-J., Barrios, E., Prescott, A.R., Elliott, G.N., Sprent, J.I., Young, J.P.W., James, E.K. (2005) Proof that *Burkholderia* strains form effective symbioses with legumes: a study of novel *Mimosa*-nodulating strains from South America. Appl. Environ. Microbiol. 71, 7461–7471.

[13] Chen, W-M., James, E.K., Coenye, T., Chou, J-H., Barrios, E., de Faria, S.M., Elliott, G.N., Sheu, S-Y., Sprent, J.I., Vandamme, P. (2006) *Burkholderia mimosarum* sp. nov., isolated from root nodules of *Mimosa* spp. from Taiwan and South America. Int. J. Syst. Evol. Microbiol. 56, 1847-1851.

[14] Chiaromonte, F., Yap, V.B., Miller, W. (2002) Scoring pairwise genomic sequence alignments. Pac. Symp. Biocomput. 115-126.

[15] Chun, J., Rainey, F.A. (2014) Integrating genomics into the taxonomy and systematics of the Bacteria and Archaea. Int. J. Syst. Evol. Microbiol. 64, 316–324.

[16] Cleenwerck, I., Vandemeulebroecke, K., Janssens, D., Swings, J. (2002) Re-examination of the genus *Acetobacter*, with descriptions of *Acetobacter cerevisiae* sp. nov. and *Acetobacter malorum* sp. nov. Int. J. Syst. Evol. Microbiol. 52, 1551–1558.

[17] Coenye, T., Laevens, S., Willems, A., Ohlén, M., Hannant, W., Govan, J.R.W., Gillis, M., Falsen, E., Vandamme, P. (2001) *Burkholderia fungorum* sp. nov. and *Burkholderia caledonica* sp. nov., two new species isolated from the environment, animals and human clinical samples. Int. J. Syst. Evol. Microbiol. 51, 1099–1107.

[18] Coenye, T., Vandamme, P. (2003) Diversity and significance of *Burkholderia* species occupying diverse ecological niches. Environ. Microbiol. 5, 719–729.

[19] Coyne, J.A., Orr, H.A. (2004) Speciation. Sinauer Associates, Inc. Sunderland, Massachusettes.

[20] Darling, A.C., Mau, B., Blattner, F.R., Perna, N.T. (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome research 14, 1394–1403.

[21] De Maayer, P., Chan, W.Y., Rubagotti, E., Venter, S.N., Toth, I.K., Birch, P.R., Coutinho, T.A. (2014) Analysis of the *Pantoea ananatis* pan-genome reveals factors underlying its ability to colonize and interact with plant, insect and vertebrate hosts. BMC Genomics 15, 404.

[22] De Meyer, S.E., Cnockaert, M., Ardley, J.K., Maker, G., Yates, R., Howieson, J.G., Vandamme, P. (2013a) *Burkholderia sprentiae* sp. nov., isolated from *Lebeckia ambigua* root nodules. Int. J. Syst. Evol. Microbiol. 63, 3950-3957.

[23] De Meyer, S.E., Cnockaert, M., Ardley, J.K., Trengove, R.D., Garau, G., Howieson, J.G., Vandamme, P. (2013b) *Burkholderia rhynchosiae* sp. nov., isolated from *Rhynchosia ferulifolia* root nodules. Int. J. Syst. Evol. Microbiol. 63, 3944-3949.

[24] De Meyer, S.E., Cnockaert, M., Ardley, J.K., Van Wyk, B-E., Vandamme, P.A., Howieson, J.G. (2014) *Burkholderia dilworthii* sp. nov., isolated from *Lebeckia ambigua* root nodules. Int. J. Syst. Evol. Microbiol. 64, 1090-1095.

[25] Elliott, G.N., Chen, W-M., Bontemps, C., Chou, J-H., Young, J.P.W., Sprent, J.I., James, E.K. (2007) Nodulation of *Cyclopia* spp. (Leguminosae, Papilionoideae) by *Burkholderia tuberum*. Ann. Bot. 100, 1403-1411.

[26] Euzéby, J.P. (1997) List of bacterial names with standing in nomenclature: a folder available on the internet. Int. J. Syst. Bacteriol. 47, 591-592.

[27] Felsenstein, J. (1985) Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39, 783-791.

[28] Felsenstein, J. (2005) PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle (USA). http://evolution.genetics.washington.edu/phylip.html.

[29] Ferreira, P.A.A., Bomfeti, C.A., Soares, B.L., de Souza Moreira, F.M. (2012) Efficient nitrogen-fixing *Rhizobium* strains isolated from amazonian soils are highly tolerant to acidity and aluminium. World J. Microbiol. Biotechnol. 28, 1947-1959.

[30] Garrity, G.M., Bell, J.A., Linburn, T. (2005). Family I. *Burkholderiaceae.* In *Bergey's Manual of Systematic Bacteriology,* 2nd ed., vol. 2, pp. 438-475. Edited by D.J. Brenner, N.R. Krieg, J.T. Staley & G.M. Garrity. New York: Springer.

[31] Gevers, D., Cohan, F.M., Lawrence, J.G., Spratt, B.G., Coenye, T., Feil, E.J., Stackebrandt, E., Van de Peer, Y., Vandamme, P., Thompson, F.L., Swings, J. (2005) Opinion: re-evaluating prokaryotic species. Nat. Rev. Microbiol. 3, 733–739.

[32] Goris, J., De Vos, P., Caballero-Mellado, J., Park, J., Falsen, E., Quensen, J. F., III, Tiedje, J. M. & Vandamme, P. (2004) Classification of the biphenyl- and polychlorinated biphenyl-degrading strain LB400^T and relatives as *Burkholderia xenovorans* sp. nov. Int. J. Syst. Evol. Microbiol. 54, 1677–1681.

[33] Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P., Tiedje, J. M. (2007) DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int. J. Syst. Evol. Microbiol. 57, 81-91.

[34] Goris, J., Suzuki, K., De Vos, P., Nakase, T., Kersters, K. (1998) Evaluation of a microplate DNA-DNA hybridization method compared with the initial renaturation method. Can. J. Microbiol. 44, 1148-1153.

[35] Guindon, S., Gascuel, O. (2003) A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. Syst. Biol. 52, 696-704.

[36] Gyaneshwar P., Hirsch, A.M., Moulin, L., Chen, W-M., Elliott, G.N., Bontemps, C., Estrada-de los Santos, P., Gross, E., dos Reis, Jr., F.B., Sprent, J.I., Young, J.P.W., James, E.K. (2011) Legume-nodulating betaproteobacteria: diversity, host range, and future prospects. Mol. Plant-Microbe Interact. 24, 1276–1288.

[37] Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41, 95-98.

[38] Katoh, K., Toh, H. (2008) Improved accuracy of multiple ncRNA alignment by incorporating structural information into a MAFFT-based framework. BMC Bioinf. 9:212. doi: 10.1186/1471-2105-9-212

[39] Konstantinidis, K.T., Tiedje, J.M. (2006) Toward a more robust assessment of intraspecies diversity, using fewer genetic markers. Appl. Environ. Microbiol. 72, 7286-7293.

[40] Lan, R., Reeves, P.R. (2000) Intraspecies variation in bacterial genomes: the need for a species genome concept. Trends Microbiol. 8, 396-401.

[41] Lefébure, T., Stanhope, M.J. (2007) Evolution of the core and pan-genome of *Streptococcus*: positive selection, recombination, and genome composition. Genome Biol. 8, 71.

[42] Lemaire, B., Dlodlo, O., Chimphango, S., Stirton, C., Schrire, B., Boatwright, J.S., Honnay, O., Smets, E., Sprent, J., James, E.K., Muasya, A.M. (2014) Symbiotic diversity, specificity and distribution of rhizobia in native legumes of the Core Cape Subregion (South Africa). FEMS Microbiol. Ecol. doi: http://dx.doi.org/10.1093/femsec/fiu024

[43] Lewis, J. A., Horswill, A.R., Schwem, B.E., Escalante-Semerena, J.C. (2004) The tricarballylate utilization (*tcu*RABC) genes of *Salmonella enterica* serovar Typhimurium LT2. J. Bacteriol. 186, 1629-1637.

[44] Lukjancenko, O., Wassenaar, T., Ussery, D. (2012) Comparison of 61 sequenced *Escherichia coli* genomes. Microb. Ecol. 60, 708-720.

[45] Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y-J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L., Jarvie, T.P., Jirage, K.B., Kim, J-B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T.,

Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F., Rothberg, J.M. (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437, 376-380.

[46] Martínez-Aguilar, L., Salazar-Salazar, C., Díaz Méndez, R., Caballero-Mellado, J., Hirsch, A.M., Vásquez-Murrieta, M.S., Estrada-de los Santos, P. (2013) *Burkholderia caballeronis* sp. nov., a nitrogen fixing species isolated from tomato (*Lycopersicon esculentum*) with the ability to effectively nodulate *Phaseolus vulgaris*. Antonie van Leeuwenhoek 104, 1063-1071.

[47] Mavengere, N.R., Ellis, A.G., Le Roux, J.J. (2014) *Burkholderia aspalathi* sp. nov., isolated from root nodules of the South African legume *Aspalathus abietina* Thunb. Int. J. Syst. Evol. Microbiol. 64, 1906-1912.

[48] Medini, D., Donati, C., Tettelin, H., Masignani, V., Rappuoli, R. (2005) The microbial pan-genome. Curr. Opin. Genet. Dev. 15, 589-94.

[49] Meyer, F., Overbeek, R., Rodriguez, A. (2009) FIGfams: yet another set of protein families. Nucleic Acids Res. 37, 6643–6654. http://doi.org/10.1093/nar/gkp698

[50] Nei, M., Kumar, S. (2000) Molecular evolution and phylogenetics. Oxford University Press, New York.

[51] Ormeño-Orrillo, E., Servín-Garcidueñas, L.E., Rogel, M.A., González, V.,Peralta, H., Mora, J., Martínez-Romero, J., Martínez-Romero, E. (2015) Taxonomy of rhizobia and agrobacteria from the *Rhizobiaceae* family in light of genomics. Syst. Appl. Microbiol. doi: dx.doi.org/10.1016/j.syapm.2014.12.002

[52] Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A.R., Xia, F., Stevens, R. (2014) The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res. 42, D206-214. doi: 10.1093/nar/gkt1226

[53] Parker, M.A., Wurtz, A.K., Paynter, Q. (2007) Nodule symbiosis of invasive *Mimosa pigra* in Australia and in ancestral habitats: a comparative analysis. Biol. Invasions 9, 127–138.

[54] Parte, A.C. (2013) LPSN – list of prokaryotic names with standing in nomenclature. Nucleic Acids Res. 42, D613- D616. doi: 10.1093/nar/gkt1111

[55] Pitcher, D.G., Saunders, N.A., Owen, R.J. (1989) Rapid extraction of bacterial genomic DNA with guanidium thiocyanate. Lett. Appl. Microbiol. 8, 151–156. doi: 10.1111/j.1472-765X.1989.tb00262.x

[56] Posada, D. (2008) jModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25, 1253-1256.

[57] Ramasamy, D., Mishra, A.K., Lagier, J-C., Padhmanabhan, R., Rossi, M., Sentausa, E., Raoult, D., Fournier, P-E. (2014) A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. Int. J. Syst. Evol. Microbiol. 64, 384-391.

[58] Reis, V.M., Estrada de los Santos, P., Tenorio-Salgado, S., Vogel, J., Stoffels, M., Guyon, S., Mavingui. P., Baldani, V.L.D., Schmid, M., Baldani, J.I., Balandreau, J., Hartmann, A., Caballero-Mellado, J. (2004) *Burkholderia tropica* sp. nov., a novel nitrogen-fixing, plant-associated bacterium. Int. J. Syst. Evol. Microbiol. 54, 2155-2162.

[59] Richter, M., Rosselló-Móra, R. (2009) Shifting the genomic gold standard for the prokaryotic species definition. Proc. Natl. Acad. Sci. U.S.A. 106, 19126-19131.

[60] Rozen, S., Skaletsky, H.J. (1999) Primer3 on the www for General Users and for Biologist Programmers. Meth. Mol. Biol. 132, 365-386.

[61] Sawana, A., Adeolu, M., Gupta, R.S. (2014) Molecular signatures and phylogenomic analysis of the genus *Burkholderia*: proposal for division of this genus into the emended genus *Burkholderia*

containing pathogenic organisms and a new genus *Paraburkholderia* gen. nov. harboring environmental species. Front. Genet. 5, 429. doi:10.3389/fgene.2014.00429

[62] Schnitzler, J., Barraclough, T.G., Boatwright, J.S., Goldblatt, P., Manning, J.C., Powell, M.P., Rebelo, T., Savolainen, V. (2011) Causes of plant diversification in the Cape biodiversity hotspot of South Africa. Syst. Biol. 60, 343-357.

[63] Schoen, C., Tettelin, H., Parkhill, J., Frosch, M. (2009) Genome flexibility in *Neisseria meningitidis*. Vaccine 27, B103-B111.

[64] Sheu, S-Y., Chou, J-H., Bontemps, C., Elliott, G.N., Gross, E., dos Reis Junior, F.B., Melkonian, R., Moulin, L., James, E.K., Sprent, J.I., Young, J.P.W., Chen, W-M. (2013) *Burkholderia diazotrophica* sp. nov., isolated from root nodules of *Mimosa* spp. Int. J. Syst. Evol. Microbiol. 63, 435-441.

[65] Sheu, S-Y., Chou, J-H., Bontemps, C., Elliott, G.N., Gross, E., James, E.K., Sprent, J.I., Young, J.P.W., Chen, W-M. (2012) *Burkholderia symbiotica* sp. nov., isolated from root nodules of *Mimosa* spp. native to north-east Brazil. Int. J. Syst. Evol. Microbiol. 62, 2272-2278.

[66] Sites, J.W. Jr., Marshall, J.C. (2004) Operational criteria for delimiting species. Annu. Rev. Ecol. Evol. Syst. 35, 199–229.

[67] Stackebrandt, E., Ebers, J. (2006) Taxonomic parameters revisited: tarnished gold standards. Microbiol. Today 33, 152 – 155.

[68] Stackebrandt, E., Goebel, B.M. (1994) Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int. J. Syst. Bacteriol. 44, 846–849.

[69] Steenkamp, E.T., Wright, J., Baldauf, S.L. (2006) The protistan origins of animals and fungi. Mol. Biol. Evol. 23, 93-106.

[70] Tamura, K., Nei, M. (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10, 512-526.

[71] Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S. (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol. Biol. Evol. 28, 2731-2739.

[72] Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol. Biol. Evol. 30, 2725-2729.

[73] Tavaré, S. (1986) Some probabilistic and statistical problems in the analysis of DNA sequences. In: Miura, R.M. (Ed.), Some mathematical questions in biology-DNA sequence analysis. Rhode Island: American Mathematical Society, pp. 57-86.

[74] Vaidya, G., Lohman, D.J., Meier, R. (2011) SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics 27, 171-180. Accessible at: http://dx.doi.org/10.1111/j.1096-0031.2010.00329.x

[75] Vandamme, P., Dawyndt, P. (2011) Classification and identification of the *Burkholderia cepacia* complex: past, present and future. Syst. Appl. Microbiol. 34, 87–95.

[76] Vandamme, P., Goris, J., Chen, W-M., de Vos, P., Willems, A. (2002) *Burkholderia tuberum* sp. nov. and *Burkholderia phymatum* sp. nov., nodulate the roots of tropical legumes. Syst. Appl. Microbiol. 25, 507-512.

[77] Vandamme, P., Opelt, K., Knöchel, N., Berg, C., Schönmann, S., De Brandt, E., Eberl, L., Falsen, E., Berg, G. (2007) *Burkholderia bryophila* sp. nov. and *Burkholderia megapolitana* sp. nov., moss-associated species with antifungal and plant-growth-promoting properties. Int. J. Syst. Evol. Microbiol. 57, 2228–2235.

[78] Vandamme, P., Peeters, C. (2014) Time to revisit polyphasic taxonomy. Antonie van Leeuwenhoek 106, 57–65. doi:10.1007/s10482-014-0148-x

[79] Vermis, K., Coenye, T., LiPuma, J.J., Mahenthiralingam, E., Nelis, H.J., Vandamme, P. (2004) Proposal to accommodate *Burkholderia cepacia* genomovar VI as *Burkholderia dolosa* sp. nov. Int. J. Syst. Evol. Microbiol. 54, 689-691.

[80] Willems, A. (2006) The taxonomy of rhizobia: an overview. Plant and Soil 287, 3–14

[81] Whitman, W.B., Woyke, T., Klenk, H-P., Zhou, Y., Lilburn, T.G., Beck, B.J., De Vos, P., Vandamme P., Eisen, J.A., Garrity, G., Hugenholtz, P., Kyrpides, N.C. (2015) Genomic Encyclopedia of Bacterial and Archaeal Type Strains, Phase III: the genomes of soil and plant-associated and newly described type strains. Standards in Genomic Sciences 10.26, 1-6. doi:10.1186/s40793-015-0017-x

[82] Yamaguchi, Y., Park, J-H., Inouye, M. (2011) Toxin-antitoxin systems in Bacteria and Archaea. Annu. Rev. Genet. 45, 61-79.