Taxon K, a complex within the *Burkholderia cepacia* complex, comprises at least two novel species, *Burkholderia contaminans* sp. nov. and *Burkholderia lata* sp. nov.

Elke Vanlaere,¹ Adam Baldwin,² Dirk Gevers,¹ Deborah Henry,³
 Evie De Brandt,¹ John J. LiPuma,⁴ Eshwar Mahenthiralingam,⁵
 David P. Speert,³ Chris Dowson² and Peter Vandamme¹

Correspondence

Peter Vandamme

Peter.Vandamme@UGent.be

¹Laboratorium voor Microbiologie, Universiteit Gent, K. L. Ledeganckstraat 35, B-9000 Gent, Belgium

²Department of Biological Sciences, Warwick University, Coventry CV4 7AL, UK

³Department of Pediatrics, University of British Columbia, Vancouver, British Columbia, Canada

⁴Department of Pediatrics, University of Michigan, 1150 W. Medical Centre Drive, 8323 MRSB III, Box 0646, Ann Arbor, MI 48109-0646, USA

⁵Cardiff School of Biosciences, Cardiff University, Cardiff CF10 3TL, UK

The aim of the present study was to re-examine the taxonomic position and structure of taxon K (also known as group K) within the *Burkholderia cepacia* complex (Bcc). For this purpose, a representative set of strains was examined by a traditional polyphasic taxonomic approach, by multilocus sequence typing (MLST) analysis and by analysis of available whole-genome sequences. Analysis of the *recA* gene sequence revealed three different lineages, designated *recA*-I, *recA*-II and *recA*-III. DNA–DNA hybridization experiments demonstrated that *recA*-I and *recA*-II isolates each represented a single novel species. However, DNA–DNA hybridization values of *recA*-II strains towards *recA*-III strains and among *recA*-III strains were at the threshold level for species delineation. By MLST, *recA*-I isolates were clearly distinguished from the others and represented a distinct lineage referred to as MLST-I, whereas *recA*-II and *recA*-III isolates formed a second MLST lineage referred to as MLST-II. A divergence value of 3.5 % was obtained when MLST-I was compared with MLST-II. The internal level of concatenated sequence divergence within MLST-I and MLST-II was 1.4 and 2.7 %, respectively; by comparison with the level of concatenated sequence divergence in established Bcc species, these data demonstrate that the MLST-I and MLST-II lineages represent two distinct species within the Bcc. The latter conclusion was supported by comparison of the whole-genome average nucleotide identity (ANI) level of MLST-I and MLST-II strains with strains of established Bcc species and by a whole-genome-based phylogenetic analysis. We formally propose to classify taxon K bacteria from the MLST-I and MLST-II lineages as *Burkholderia contaminans* sp. nov. (with strain J2956^T = LMG 23361^T = CCUG 55526^T as the type strain) and *Burkholderia lata* sp. nov. (with strain 383^T = ATCC 17760^T = LMG 22485^T = CCUG 55525^T as the type strain), respectively. The MLST approach was confirmed as a valuable instrument in polyphasic taxonomic studies; more importantly, the cumulative data for about 1000 Bcc isolates analysed demonstrate that the 3 % concatenated sequence divergence level correlates with the 70 % DNA–DNA hybridization or 95 % whole-genome ANI threshold levels for species delineation.

Abbreviations: ANI, average nucleotide identity; Bcc, *Burkholderia cepacia* complex; MLST, multilocus sequence typing; RFLP, restriction fragment length polymorphism; ST, sequence type.

The GenBank/EMBL/DDBJ accession number for the 16S rRNA gene sequence of strain R-15816 is AM905038, and those of the *recA* sequences of strains R-23139, R-15816, R-18442, R-20938, R-9896, R-18428, LMG 16227, LMG 23255 and LMG 23253 are respectively AM905032–AM905037 and AM922301–AM922303.

ANI values, G + C contents, DNA–DNA hybridization results and biochemical characteristics of the novel isolates and details of single-copy core genes used in tree reconstruction are available as supplementary material with the online version of this paper.

INTRODUCTION

The *Burkholderia cepacia* complex (Bcc) comprises a group of closely related organisms that are very versatile (Coenye & Vandamme, 2003). The organisms can be exploited for biocontrol, bioremediation and plant-growth-promotion purposes, but their capacity as opportunistic bacteria to cause human infections, in particular in cystic fibrosis patients, hampers their use in these biotechnological applications (Parke & Gurian-Sherman, 2001). The taxonomy of the Bcc has evolved dramatically in the last few years (Coenye *et al.*, 2001). Until recently, the Bcc consisted of nine species with validly published names, but the recognition of *Burkholderia ubonensis* and the description of an additional five novel species (Vanlaere *et al.*, 2008) has increased the number of Bcc species to 15. Recent multicentre studies revealed several clusters of isolates that do not group with one of these 15 Bcc species. One diverse group of unclassified Bcc isolates is known as taxon K or group K (Baldwin *et al.*, 2005; Dalmastri *et al.*, 2005, 2007; Mahenthiralingam *et al.*, 2006; Payne *et al.*, 2005; Vermis *et al.*, 2002) and includes bacteria from human and environmental sources isolated worldwide. These sources include sputum and blood samples of debilitated patients such as cystic fibrosis patients, reservoir water, river water and sediments, soil, roots, animals, industrial and pharmaceutical product contaminants and domestic and personal care products (Mahenthiralingam *et al.*, 2006, 2008; P. Vandamme and A. Baldwin, unpublished data). Moreover, several cases of epidemic transmission of taxon K strains among cystic fibrosis patients have been reported (Campana *et al.*, 2005; Cunha *et al.*, 2007). The whole-genome sequences of two taxon K strains, namely *Burkholderia* sp. 383 (=LMG 22485) (GenBank accession numbers NC_007509, NC_007510 and NC_007511 for the three chromosomes) and the *Burkholderia* sp. metagenome SAR-1 (Venter *et al.*, 2004) (NC_000028), are publicly available. In an earlier study, two taxon K isolates were classified as *B. cepacia*, based on DNA–DNA hybridization experiments (Vermis *et al.*, 2002). However, *recA* gene and multilocus sequence analyses, which are known to correlate well with Bcc species status (Baldwin *et al.*, 2005; Mahenthiralingam *et al.*, 2000b), allocated these and other taxon K strains in a distinct clade, suggesting that they may represent a novel Bcc taxonomic group.

The aim of the present study was to re-examine the taxonomic position and structure of taxon K. For this purpose, a set of strains was examined by a traditional polyphasic taxonomic approach, by multilocus sequence typing (MLST) analysis and by analyses of the available whole-genome sequences.

METHODS

Bacterial strains and growth conditions. In our ongoing diversity studies of Bcc bacteria, about 150 isolates have been identified as taxon K by means of *HaeIII*-based *recA* gene restriction fragment

length polymorphism (RFLP) PAGE or *recA* gene sequence analysis. Twenty of these isolates were selected and used in the present study (Table 1). All strains were routinely cultured on trypticase soy agar (TSA) and incubated at 28 °C unless indicated otherwise.

DNA preparation. For PCR and RFLP experiments, DNA was prepared by alkaline lysis as described previously (Storms *et al.*, 2004). For DNA–DNA hybridization experiments and determination of the DNA base composition, DNA was prepared as described by Pitcher *et al.* (1989).

Bcc-specific *recA* gene PCR. The *recA* gene (1040 bp) was amplified using primers BCR1 and BCR2 as described by Mahenthiralingam *et al.* (2000b).

***recA* RFLP and *recA* gene sequence analysis.** Amplified *recA* fragments were subjected to *HaeIII*-RFLP analysis (Mahenthiralingam *et al.*, 2000b). Electrophoretic separation of the restriction fragments was performed by PAGE as described by Vanlaere *et al.* (2005). The restriction patterns were analysed using the BioNumerics 4.5 software package (Applied Maths) and compared with those of Bcc reference strains. *recA* sequence analysis was performed as described previously (Mahenthiralingam *et al.*, 2000b). Multiple alignment was performed by using the CLUSTAL_X program (Thompson *et al.*, 1997). The aligned sequences were analysed phylogenetically using the BioNumerics 4.5 software. Distances were calculated using the Jukes–Cantor algorithm. Phylogenetic trees based on the neighbour-joining method were constructed with bootstrap values of 1000 replications.

16S rRNA gene sequence analysis. 16S rRNA gene sequence analysis was performed as described previously (Coenye *et al.*, 2001).

Determination of the DNA base composition. DNA was enzymically degraded into nucleosides as described by Mesbah & Whitman (1989). The nucleoside mixture obtained was separated using a Waters Breeze HPLC system and XBridge Shield RP18 column thermostabilized at 37 °C. The solvent was 0.02 M NH₄H₂PO₄ (pH 4.0) with 1.5 % (v/v) acetonitrile. Non-methylated lambda phage DNA (Sigma) and *Escherichia coli* LMG 2093 DNA were used as calibration reference and control, respectively.

DNA–DNA hybridization experiments. Hybridization reactions were performed with photobiotin-labelled probes in microplate wells as described before (Ezaki *et al.*, 1989; Goris *et al.*, 1998), using an HTS7000 Bio Assay Reader (Perkin-Elmer) for the fluorescence measurements. The hybridization temperature was 50 °C.

MLST analysis. MLST analysis was performed as described before (Baldwin *et al.*, 2005). A phylogenetic tree of concatenated sequences (2773 bp), including fragments of seven genes [*atpD* (443 bp), *gltB* (400 bp), *gyrB* (454 bp), *recA* (393 bp), *lepA* (397 bp), *phaC* (385 bp) and *trpB* (301 bp)] from each isolate was constructed based on the neighbour-joining method using MEGA software package version 3 (Baldwin *et al.*, 2005; Kumar *et al.*, 2004). The significance of branching within the trees was evaluated by bootstrap analysis of 1000 computer-generated trees. The program DnaSP version 4.1 (<http://www.ub.es/dnasp/DnaSP32Inf.html>) was used to calculate the mean number of nucleotide substitutions per site (i.e. the percentage divergence of concatenated allele sequences) between populations based on a Jukes–Cantor method (Naser *et al.*, 2007; Rozas *et al.*, 2003). The standard deviation was calculated to show how widely the values are spread in the dataset. Evidence for clonal or recombining populations was estimated by measuring the extent of linkage [using the standardized index of association (*sI_A*)] between alleles at different loci around the chromosome as described previously (Baldwin *et al.*, 2005; Haubold & Hudson, 2000;

Table 1. Bcc taxon K isolates examined

ATCC, American Type Culture Collection, Manassas, VA, USA; LMG, BCCM/LMG Bacteria Collection, Laboratorium voor Microbiologie, Universiteit Gent, Gent, Belgium; CCUG, Culture Collection, University of Göteborg, Department of Clinical Bacteriology, Göteborg, Sweden; NCIMB, National Collections of Industrial, Marine and Food Bacteria, Bucksburn, Aberdeen, Scotland, UK; NCTC, National Collection of Type Cultures, London, UK; R, Research Collection, Laboratorium voor Microbiologie, Universiteit Gent; other strain numbers are personal designations of the depositor. CF, Cystic fibrosis; ND, not determined.

Strain	Other strain designation(s)	Source	Depositor*	<i>HaeIII</i> <i>recA</i> RFLP PAGE type	<i>recA</i> sequence cluster	ST/MLST group
<i>Burkholderia con-</i> <i>taminans</i> sp. nov.						
LMG 16227	CCUG 34411	CF patient (Sweden)	E. Falsen	Se27	<i>recA</i> -I	362/MLST-I
LMG 23253	VERONA 53	CF patient (Italy)	S. Campana	AT	<i>recA</i> -I	102†/MLST-I
LMG 23255	467, BCC0823	CF patient (Czech Republic)	P. Drevinek	AT	<i>recA</i> -I	102†/MLST-I
LMG 23361 ^T	J2956 ^T , CCUG 55526 ^T	Sheep with mastitis, milk (Spain)	J. R. W. Govan	AT	<i>recA</i> -I	102†/MLST-I
R-9896	CEP0097, 90-530, BCC0282	CF patient, blood (USA)	Own isolate (D. P. S.)	K	<i>recA</i> -I	123/MLST-I
R-13528	BCC0340, CEP1056	CF patient, sputum (Canada)	Own isolate (D. P. S.)	K	<i>recA</i> -I	100/MLST-I
R-18428	PC21	Patient, blood (Brazil)	V. Magalhaes	K	<i>recA</i> -I	344/MLST-I
R-18442	PC66	Patient, blood (Brazil)	V. Magalhaes	Se27	<i>recA</i> -I	–‡/MLST-I
R-20938	03174009	CF patient (Sweden)	A. Lagerqvist	Se27	<i>recA</i> -I	362/MLST-I
<i>Burkholderia lata</i> sp. nov.						
LMG 6860	NCIB 9091, ATCC 17771	Forest soil (Trinidad)	NCIMB	AY	<i>recA</i> -II	ND
LMG 6863	CEP0085, ATCC 17460, BCC0278, CDC B2320, NCIB 9690	Maraval river (Trinidad)	NCIMB	K	<i>recA</i> -II	99/MLST-II
LMG 6992	CCUG 2857, ATCC 17769, BCC0174, FC0436, NCIB 9088	Soil (Trinidad)	CCUG	K	<i>recA</i> -II	239/MLST-II
LMG 6993	DSM 50180, ATCC 17770, NCIB 9089, CCUG 2858	Forest soil (Trinidad)	CCUG	Se35	<i>recA</i> -II	–/MLST-II
LMG 22485 ^T	383 ^T , ATCC 17760 ^T , CCUG 55525 ^T	Forest soil (Trinidad)	T. Lessie	K	<i>recA</i> -II	101/MLST-II
LMG 14095	290/74, FC0450, BCC0213	CF patient (UK)	NCTC	K	<i>recA</i> -III	300/MLST-II
R-3211	CEP0088, BCC0217	Patient, cerebrospinal fluid (USA)	Own isolate (D. P. S.)	K	<i>recA</i> -III	98/MLST-II
R-9940	BCC0335, CEP1061	CF patient (Canada)	Own isolate (D. P. S.)	J	<i>recA</i> -III	103/MLST-II
R-15816	323	Flower (Brazil)	P. Hebbar	K	<i>recA</i> -III	–/MLST-II
R-18628	BCC0402, B5	River water (USA)	L. Leff	J	<i>recA</i> -III	119/MLST-II
R-23139	AU6262	CF patient (USA)	Own isolate (J. J. L.)	K	<i>recA</i> -III	ND

*Depositors (in alphabetical order): S. Campana, Cystic Fibrosis Center, Meyer Hospital, University of Florence, Florence, Italy; E. Falsen, CCUG; J. R. W. Govan, Cystic Fibrosis Microbiology Laboratory and Strain Repository, Medical Microbiology Division, University of Edinburgh Medical School, Edinburgh, UK; P. Hebbar, USDA-ARS, Plant Science Institute, Alternate Crops Systems Lab, Beltsville, MD, USA; A. Lagerqvist, Institute for Clinical Science, Clinical Bacteriology, Uppsala University, Uppsala, Sweden; L. Leff, Department of Biological Sciences, Kent State University, Kent, OH, USA; T. Lessie, Department of Microbiology, University of Washington Medical School, Seattle, WA, USA; V. Magalhaes, Centro de Pesquisa Experimental, Instituto de Ensino e Pesquisa Albert Einstein, São Paulo, Brazil; P. Drevinek, Cardiff School of Biosciences, Cardiff University, Cardiff, UK.

†ST102 gives a perfect match with metagenome SAR-1.

‡–, No ST identifier due to incomplete allelic profile.

Maynard Smith *et al.*, 1993). An sI_A not significantly greater than 0 after 1000 computer randomizations suggests that a single species population (monophyletic) is in linkage equilibrium (freely recombining), while a population with an sI_A significantly greater than 0 ($P < 0.001$) is considered to be in linkage disequilibrium

(clonal) (Baldwin *et al.*, 2005; Haubold & Hudson, 2000). Nucleotide sequences of each allele and allelic profile and sequence types for all strains analysed in this study are available from the Bcc MLST website (<http://pubmlst.org/bcc/>) developed by Keith Jolley and sited at the University of Oxford (Jolley *et al.*, 2004).

Determination of shared gene content and average nucleotide identity (ANI) values. At the time of writing, 12 Bcc strains have been included in whole-genome sequencing projects, of which seven whole genomes were used in the present study (unfinished genomes are indicated by asterisks): *Burkholderia multivorans* ATCC 17616* (=LMG 17588), *Burkholderia cenocepacia* IIIA J2315^T (=LMG 16656^T), *Burkholderia vietnamiensis* G4 (=LMG 22486), *Burkholderia dolosa* AU0158* (=R-3200), *Burkholderia ambifaria* AMMD^T (=LMG 19182^T) and *Burkholderia* sp. LMG 22485 and the *Burkholderia* sp. metagenome SAR-1 (Supplementary Table S1, available in IJSEM Online). In addition, six genomes of other *Burkholderia* strains and the genome of *Cupriavidus metallidurans* CH34^T were included for comparison with Bcc genomes. All genomes are publicly available and were obtained from the FTP site of the National Center for Biotechnology Information (NCBI) in their version of 25 August 2007, except for J2315^T, which was produced at the Sanger Institute and which was obtained from <ftp://ftp.sanger.ac.uk/pub/pathogens/bc> (Supplementary Table S1). The number of protein-coding (annotated) genes in each Bcc genome ranged from 5441 to 7717, with a mean gene length of 992.16 bp.

ANI values were calculated as described by Konstantinidis & Tiedje (2005a). Whole-genome sequences present an opportunity to produce a more robust and resolved species tree using concatenation of core gene sequences. The set of single-copy orthologous core genes was determined by building gene families using OrthoMCL (Li *et al.*, 2003). Protein sequence alignments were generated in MUSCLE (Edgar, 2004), and poorly conserved regions were automatically trimmed using Gblocks (Castresana, 2000). Maximum-likelihood trees with 100 bootstrap replicates were obtained for a subset of 20 000 randomly selected amino acid positions from the concatenated single-copy core gene alignment, using the PhyML algorithm (Guindon & Gascuel, 2003) with a JTT amino acid substitution matrix (Jones *et al.*, 1992) and a discrete gamma model.

Biochemical characterization. A biochemical characterization was performed as described previously (Henry *et al.*, 1997, 2001).

RESULTS AND DISCUSSION

recA RFLP and *recA* gene sequence analysis

Different *Hae*III *recA* RFLP patterns, designated K, AT, Se27, AY, Se35 and J, were found among the 20 isolates studied (Table 1). Pattern J has also been observed in *B. cenocepacia*, *Burkholderia stabilis* and *Burkholderia arboris* strains (Mahenthiralingam *et al.*, 2000a; Vanlaere *et al.*, 2008). Therefore, these isolates were further examined by *Mn*II *recA* RFLP PAGE, generating unique profiles for each species (data not shown). The 20 strains examined were subjected to *recA* sequence analysis. Similarity of the *recA* sequences towards established Bcc species was greater than 94 %, and the *recA* sequences of taxon K isolates constituted one cluster, including *Burkholderia metallica*, within the Bcc. Within taxon K, several distinct clusters, referred to as *recA*-I, *recA*-II and *recA*-III, were observed that showed sequence similarities of more than 97 % (Fig. 1).

16S rRNA gene sequencing and phylogenetic analysis

The 16S rRNA gene sequences of a representative strain from each *recA* cluster were examined (LMG 16227, LMG

22485 and R-15816 for clusters *recA*-I, *recA*-II and *recA*-III, respectively). Comparison of these 16S rRNA gene sequences with those of other Bcc strains revealed similarity levels above 98 % (data not shown). These values were similar to those obtained between strains of other Bcc species. Similarity levels of less than 97.5 % were calculated towards representatives of other *Burkholderia* species (data not shown).

DNA–DNA hybridization and determination of the DNA base composition

DNA–DNA hybridization experiments were performed between representatives of the three *recA* clusters, *B. metallica* LMG 24274 and *B. cepacia* LMG 1222^T (Supplementary Table S2). DNA–DNA hybridization experiments were also performed between *recA*-I strain LMG 23253 and type strains of the established members of the Bcc, including *B. ubonensis*. The mean value obtained among *recA*-I strains LMG 23361^T, LMG 23253 and LMG 23255 was 88 % and the value between *recA*-II strains LMG 22485^T and LMG 6863 was 79 %, whereas the value obtained between the *recA*-III strains LMG 14095 and LMG 24274 was 68 %. Moderate to high values, in the range of 59 to 64 %, were obtained between *recA*-I and *recA*-II strains. Moderate to high values of 60 and 66 % were obtained between *recA*-I and *recA*-III strains. Somewhat higher values, in the range 66–68 %, were obtained between *recA*-II and *recA*-III strains. Values in the range of 61 to 66 % were found between strains of the different *recA* clusters and *B. cepacia* LMG 1222^T and *B. metallica* LMG 24274 (Supplementary Table S2). Moderate values (42–62 %) were obtained between *recA*-I strain LMG 23253 and established Bcc members (data not shown). All strains investigated had G + C contents of 66–68 mol%, which is typical for Bcc species (Coenye *et al.*, 2001) (Supplementary Table S2). The DNA–DNA hybridization values demonstrate that strains of the three *recA* clusters are closely related, but may constitute distinct novel species within the Bcc (Wayne *et al.*, 1987). However, cluster *recA*-II and *recA*-III strains showed DNA–DNA hybridization values at the threshold for species delineation. Additionally, the two cluster *recA*-III strains examined also exhibited a DNA–DNA hybridization value near the 70 % threshold.

MLST analysis

Eighteen of 20 taxon K strains selected for the present study were analysed by MLST. Three of these strains consistently yielded incomplete MLST profiles (five genes for R-15816 and R-18442; six genes for LMG 6993), probably due to a sequence variation disrupting the binding site of the amplification primers. Nevertheless, analysis of the remaining loci identified them as taxon K strains (Supplementary Table S3). Data for an additional 21 taxon K strains examined previously and for the *Burkholderia* sp. metagenome SAR-1 (Baldwin *et al.*, 2005; Mahenthiralingam *et al.*, 2006; A. Baldwin, unpublished results) strain were also

Fig. 1. Phylogenetic tree derived from analysis of *recA* gene sequences of established Bcc species and taxon K strains. Different *B. cenocepacia* *recA* lineages are designated IIIA, IIIB, IIIC and IIID (Vandamme *et al.*, 2003). The tree was constructed using the neighbour-joining method (Jukes & Cantor, 1969). Accession numbers are given in parentheses. Bootstrap numbers ($\geq 50\%$) are shown for 1000 replicates. The sequence of *Burkholderia xenovorans* LMG 21463^T was used as outgroup. Bar, 1% sequence dissimilarity.

included in the present analyses (Supplementary Table S3). Phylogenetic analysis of concatenated sequences demonstrated that the isolates occupied a single clade in the MLST tree which consisted of two clusters, designated MLST-I and MLST-II, supported by bootstrap values of 100 and 92%, respectively (Fig. 2). When these results were compared with results from the *recA*-based sequence analysis, cluster MLST-I comprised all *recA*-I strains examined, while cluster MLST-II comprised all *recA*-II and *recA*-III strains examined. Although the *recA*-II strains formed a distinct lineage supported by a 100% bootstrap value within MLST-II, they grouped among *recA*-III strains. The intra- and intertaxon divergence of concatenated nucleotide sequences was calculated for both MLST clusters and for the other members of the Bcc (Fig. 3). Comparison of the percentage divergence of concatenated allele sequences between established Bcc members demonstrated that an average divergence of 3% can clearly delineate Bcc species. The diversity within Bcc species varies from 0.43% among *B. vietnamiensis* strains (based on analysis of 29 strains) to 2.89% within *Burkholderia anthina* strains (based on analysis of nine strains). Although part of these differences can be attributed to the number of isolates examined for each species, it is clear that the genomes of different Bcc species do not evolve at

the same rate. The internal diversity observed within cluster MLST-I was 1.4%, while the value obtained within MLST-II was 2.7%. A divergence value of 3.5% was obtained when MLST-I was compared with MLST-II.

The MLST data were also used to assign allele types to sequences and allelic profiles and sequence types (STs) to all isolates (Baldwin *et al.*, 2005). Isolates with an identical allelic profile were assigned the same ST identifier and considered to be isogenic, as they were indistinguishable at all seven loci (Baldwin *et al.*, 2005). The 37 taxon K isolates examined for which a full MLST dataset was obtained were resolved into 27 STs, with ST-102 occurring seven times, ST-98 occurring three times and ST-119 and ST-362 occurring twice. MLST-I and MLST-II strains shared no identical allele types; identical alleles were mainly found within a species and have been observed between the different *B. cenocepacia* *recA* subgroups (Baldwin *et al.*, 2005). However, interspecies recombination events were recently reported (Waine *et al.*, 2007). The *s*_I_A value estimating clonality within the clusters MLST-I and MLST-II was 0.180 ($P < 0.01$) and 0.440 ($P < 0.01$), respectively (Baldwin *et al.*, 2005; Haubold & Hudson, 2000). The *s*_I_A values were thus significantly greater than zero, and MLST-I and MLST-II are therefore considered clonal populations. Nevertheless, the values are quite low, which may indicate

Fig. 2. Phylogenetic analysis of concatenated nucleotide sequences from seven loci, using the neighbour-joining method (performed using MEGA version 3 software). Bootstrap values greater than 70% are shown for 1000 replicates. Species names of the established species are given and the numbers of STs included in the analysis are indicated in parentheses. Different *B. cenocepacia* lineages are designated IIIA, IIIB, IIIC and IIID (Vandamme *et al.*, 2003). The different STs observed among the taxon K isolates examined are shown. *Burkholderia pseudomallei* was used as outgroup.

that some recombination is occurring. The sl_A values found were comparable with values observed for *B. cepacia* strains (Baldwin *et al.*, 2005). Studies of the Bcc have revealed that some species (e.g. *B. vietnamiensis* and *B. cenocepacia*) recombine more than others (e.g. *B. ambifaria*) (Waine *et al.*, 2007). Nevertheless, the number of STs for each species is small and, therefore, a much larger sample is required for a more accurate comparison of mechanisms of evolution for each Bcc species.

Determination of shared gene content and ANI values

The availability of complete genome sequences of many bacteria provides new possibilities for comprehensive demarcation of species (Konstantinidis & Tiedje, 2005b). In the present study, two genome sequences of taxon K strains were available. Mahenthiralingam *et al.* (2006) reported previously that the *Burkholderia* sp. metagenome SAR-1 was essentially complete and represented taxon K, with the closest sequenced relative being *Burkholderia* sp. LMG 22485, a strain of cluster MLST-II. The SAR-1 metagenome belongs to ST102 and was genetically identical at multiple random loci when compared to cultivable ST102 isolates (Mahenthiralingam *et al.*, 2006) which reside in MLST-I (Supplementary Table S3). The 100% nucleotide identity indicated that the genomic sequence of the ST102 SAR-1 metagenome may be used as a surrogate whole-genome sequence for cultivable ST102 isolates such as *Burkholderia* sp. strain LMG 23361. A whole-genome-based comparison between the cluster MLST-I and MLST-II strains was therefore carried out and compared with genomes of other available Bcc species, non-Bcc *Burkholderia* species and members of related genera (listed in Supplementary Table S1).

Goris *et al.* (2007) demonstrated that the recommended cut-off of 70% DNA–DNA reassociation for species delineation corresponds to 95% ANI, a measure for evolutionary relatedness based on sequence similarity between orthologous genes (Goris *et al.*, 2007). The similarity matrix of ANI values of the strains studied is shown in Supplementary Table S1. The ANI values obtained between different Bcc species ranged from 85.04 to 89.92%, whereas the ANI values between Bcc and other *Burkholderia* species ranged from 75.33 to 81.68%. The ANI value between the *Burkholderia* sp. metagenome SAR-1 and the whole-genome sequence of strain *Burkholderia* sp. LMG 22485 was 92.71%. This high value confirms the similarity between the two sequences, but is clearly lower than 95%, which corresponds to the established threshold for species delineation (Goris *et al.*, 2007).

A set of 420 single-copy core genes, i.e. orthologous genes present in exactly one copy (no duplicates) in each studied genome, was used to estimate the phylogeny of these bacteria by maximum-likelihood (Supplementary Table S4). The phylogenetic analysis (Fig. 4) revealed that the metagenome SAR-1 and *Burkholderia* sp. strain 383 were

Fig. 3. Concatenated nucleotide sequence divergence of MLST loci. Shaded bars represent the divergence within a species. Open bars represent the divergence between a Bcc species and its closest neighbour (listed to the right) as defined by percentage similarity between concatenated sequences (similar to taxon gap as described by Naser *et al.*, 2007). The number of strains included in the calculations for each species is given in parentheses. *B. cenocepacia* strains represent *recA* lineages IIIA (44), IIIB (57), IIIC (4) and IIID (2). For each bar, the SD is indicated by error bars. The cut-off value of 3% divergence for species demarcation is marked.

closest neighbours, showing distances that were comparable with those observed between other Bcc species.

Biochemical characterization

In general, species of the Bcc are phenotypically nearly indistinguishable, making their differentiation quite difficult, even with an extended panel of biochemical tests (Henry *et al.*, 2001). The variable nature of biochemical results for Bcc strains may be due to their unusual genomes, which are among the largest seen in Gram-negative bacteria (6–9 Mb) and which are divided into a minimum of two chromosomal replicons (Lessie *et al.*, 1996; Mahenthiralingam & Drevinek, 2007). The genomes are rich in insertion sequences, allowing the recruitment of

foreign genes (Kenna *et al.*, 2006). Mobile elements such as plasmids, transposons and bacteriophages are also frequently observed (Mahenthiralingam & Drevinek, 2007). The high foreign and mobile DNA content, in combination with the multireplicon structure, contributes to considerable plasticity and diversity in their genomes which, when differentially expressed in isolates, probably results in their variable biochemical phenotypic profiles.

Biochemical characteristics were determined for seven MLST-I strains and 11 MLST-II strains. A comparison of the biochemical characteristics revealed that these taxa could be distinguished by testing growth at 42 °C and haemolysis, for which strains of cluster MLST-I were mainly positive. Supplementary Table S5 lists the characteristics that are helpful in distinguishing the two taxa from each other and from established Bcc species. The biochemical characteristics of the novel species are given in the species descriptions below.

Conclusions

The aim of the present study was to re-examine the taxonomic position and structure of Bcc taxon K. For this purpose, a set of strains was examined by a traditional polyphasic taxonomic approach, by MLST analysis and by analysis of available whole-genome sequences.

Analysis of the *recA* gene sequences of 20 isolates selected from a collection of over 150 revealed that they represented three different lineages, designated *recA*-I, *recA*-II and *recA*-III (Fig. 1). Analysis of 16S rRNA genes and G + C contents of isolates representing the three groups confirmed that these bacteria are typical members of the Bcc, but was not helpful in discriminating the different lineages or in distinguishing them from other Bcc bacteria. DNA–DNA hybridization experiments among and between representatives of each lineage demonstrated that *recA*-I isolates

Fig. 4. Whole-genome-based maximum-likelihood tree for Bcc species and other *Burkholderia* species. The tree was derived from analysis of concatenated single-copy core genes, with *C. metallidurans* CH34^T as an outgroup. Bootstrap analysis indicated that the tree was completely resolved.

represented a single novel species (Supplementary Table S2). *recA*-II strains represented a second novel species, but these strains showed DNA–DNA hybridization values towards *recA*-III strains that were at the threshold level for species delineation, i.e. around 70 % (Wayne *et al.*, 1987). Similarly, the two *recA*-III strains examined exhibited DNA–DNA hybridization near the same threshold (Supplementary Table S2).

MLST analysis of 40 isolates, including 18 that were also analysed by *recA* sequence analysis, proved very helpful. *recA*-I isolates were again clearly distinguished from the others and represented a distinct lineage referred to as MLST-I (Fig. 2). *recA*-II and *recA*-III isolates formed a second MLST lineage referred to as MLST-II (Fig. 2). The level of concatenated sequence divergence within MLST-I and MLST-II was below 3 % (Fig. 3); by comparison with the level of concatenated sequence divergence in established Bcc species, these data demonstrate that the MLST-I and MLST-II lineages represent two distinct species within the Bcc. The latter conclusion was supported by comparison of the whole-genome ANI of MLST-I and MLST-II strains and of strains of established Bcc species (Supplementary Table S1) and by a whole-genome-based phylogenetic analysis (Fig. 4).

It is clear that the taxonomy and the genomic structure of these bacteria, which typically have genomes in the range of 8–9 Mb organized in multiple replicons, is extremely complex and that the classification of bacteria such as taxon K requires care. Nevertheless, they represent some of the most common members of the Bcc and require a formal classification and nomenclature. We feel that the data from the present study indicate that these bacteria are best classified into two novel species corresponding with the two MLST lineages. As for most Bcc bacteria, species-level identification is cumbersome by biochemical characteristics but can be accomplished readily through DNA-based procedures. In particular, sequence analysis of the *recA* gene, or of individual or all genes of the MLST scheme, generally yields a straightforward identification result. Nevertheless, in spite of the current 17 established species within the Bcc, several unidentified strains form additional clusters which probably represent novel Bcc species (unpublished data). Pending the availability of a reasonable set of such isolates and a full characterization thereof, the identification of some of these strains may be problematic. We propose to classify taxon K bacteria from the MLST-I and MLST-II lineages as *Burkholderia contaminans* sp. nov. and *Burkholderia lata* sp. nov., respectively. The MLST approach was confirmed as a valuable instrument in polyphasic taxonomic studies. More importantly, the cumulative data for some 971 Bcc isolates analysed (Mahenthiralingam *et al.*, 2008; A. Baldwin, unpublished data) demonstrate that the 3 % concatenated sequence divergence correlates with the 70 % DNA–DNA hybridization or 95 % whole-genome ANI thresholds for species delineation.

Description of *Burkholderia contaminans* sp. nov.

Burkholderia contaminans [con.ta'mi.nans. L. part. adj. *contaminans* contaminating, polluting, referring to the metagenome that was recovered from the Sargasso Sea, but which probably represented a sample contaminant; ST102 isolates and other Bcc isolates grew very poorly in seawater, suggesting that the open ocean is not a natural habitat of Bcc species (Mahenthiralingam *et al.*, 2006)].

Cells are Gram-negative, aerobic, non-sporulating rods. Colonies are moist, showing a metallic sheen. All known strains grow on MacConkey agar. Known strains grow on *B. cepacia*-selective agar and some strains turn that medium alkaline. Growth is observed at 30, 37 and 42 °C. Most strains are yellow pigmented and are haemolytic, a characteristic not commonly observed among Bcc species. The strains assimilate glucose, L-arabinose, D-mannose, D-mannitol, N-acetylglucosamine, D-gluconate, caprate, adipate, L-malate, citrate and phenylacetate, but not maltose. Acidification of glucose, maltose, lactose, xylose, sucrose and adonitol is observed. Nitrate reduction is strain dependent, but mostly not present. Activities of oxidase, β -galactosidase, aesculin hydrolase and lysine decarboxylase and gelatin liquefaction are present, but no ornithine decarboxylase, tryptophanase, arginine dihydrolase or urease activity. The G + C content is 67 mol%. Strains have mostly been isolated from clinical samples, but also from environmental samples. Strains have been involved in a widespread outbreak in the USA due to a contaminated nasal spray (Mahenthiralingam *et al.*, 2006) and as contaminants in a water reservoir supplying a renal dialysis machine in Brazil (Souza *et al.*, 2004).

The type strain is strain J2956^T (=LMG 23361^T =CCUG 55526^T) and was recovered from milk of a sheep with mastitis in Spain during an outbreak in 1999–2000 (Berriatua *et al.*, 2001). The *recA* RFLP type and a false-positive PCR test first suggested that this isolate belonged to *B. cenocepacia*. The type strain is identical in 14 allele sequences to the metagenome SAR-1 and can be used as a cultured surrogate for research using the SAR-1 metagenome (Mahenthiralingam *et al.*, 2006). Phenotypic and biochemical characteristics are as described above for the species.

Description of *Burkholderia lata* sp. nov.

Burkholderia lata (la'ta. L. fem. adj. *lata* broad, wide, because it is a common Bcc species worldwide).

Cells are Gram-negative, aerobic, non-sporulating rods. Colonies are moist. All known strains grow on MacConkey agar. Strains grow on *B. cepacia*-selective agar and some strains turn that medium alkaline. Growth is observed at 30 and 37 °C but not at 42 °C (except R-18628). Some strains are yellow pigmented or yellow–purple pigmented. No haemolysis (except R-15816). The strains assimilate D-glucose, D-mannose, D-mannitol, N-acetylglucosamine, D-

gluconate, adipate, L-malate and citrate, whereas the assimilation of maltose, L-arabinose, caprate and phenylacetate is strain dependent. Acidification of D-glucose, maltose, lactose and xylose is observed. Acidification of sucrose and adonitol is strain dependent. Nitrate reduction is strain dependent. Activities of oxidase and lysine decarboxylase are present, but no tryptophanase, arginine dihydrolase or urease activity. Gelatin liquefaction, ornithine decarboxylase, aesculin hydrolase and β -galactosidase are strain dependent. The variability observed in the β -galactosidase test is rather uncommon among Bcc species, which mostly show β -galactosidase activity. The G+C content is 67 mol%. Most strains have been isolated from diverse environmental and industrial samples, with some isolated from clinical samples. Remarkably, strains identified as *recA* cluster K-II have only been recovered from environmental samples in Trinidad.

The type strain is strain 383^T (=ATCC 17760^T =LMG 22485^T =CCUG 55525^T) and was originally recovered from forest soil in Trinidad in 1958 (Stanier *et al.*, 1966). Phenotypic and biochemical characteristics are as described above for the species.

ACKNOWLEDGEMENTS

E. V. is indebted to the Special Research Council of Ghent University. P. V. and D. G. are indebted to the Fund for Scientific Research Flanders (Belgium) for research grants and a post-doctoral fellowship, respectively. A. B., C. D. and E. M. are indebted to the Cystic Fibrosis Trust UK (grant PJ535) for funding their research. J. J. L. is supported by the Cystic Fibrosis Foundation (USA). D. S. and D. H. are supported by the Canadian Cystic Fibrosis Foundation. We are grateful to all depositors of strains who contributed to this study. We thank Jean Euzéby for his advice on the nomenclature of the novel species *B. lata*.

REFERENCES

- Baldwin, A., Mahenthalingam, 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.
- Berriatua, E., Ziluaga, I., Miguel-Virto, C., Uribarren, P., Juste, R., Laevens, S., Vandamme, P. & Govan, J. R. (2001). Outbreak of subclinical mastitis in a flock of dairy sheep associated with *Burkholderia cepacia* complex infection. *J Clin Microbiol* **39**, 990–994.
- Campana, S., Taccetti, G., Ravenni, N., Favari, F., Cariani, L., Sciacca, A., Savoia, D., Collura, A., Fiscarelli, E. & other authors (2005). Transmission of *Burkholderia cepacia* complex: evidence for new epidemic clones infecting cystic fibrosis patients in Italy. *J Clin Microbiol* **43**, 5136–5142.
- Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* **17**, 540–552.
- Coenye, T. & Vandamme, P. (2003). Diversity and significance of *Burkholderia* species occupying diverse ecological niches. *Environ Microbiol* **5**, 719–729.
- Coenye, T., Vandamme, P., Govan, J. R. & LiPuma, J. J. (2001). Taxonomy and identification of the *Burkholderia cepacia* complex. *J Clin Microbiol* **39**, 3427–3436.
- Cunha, M. V., Pinto-de-Oliveira, A., Meirinhos-Soares, L., Salgado, M. J., Melo-Cristino, J., Correia, S., Barreto, C. & Sá-Correia, I. (2007). Exceptionally high representation of *Burkholderia cepacia* among *B. cepacia* complex isolates recovered from the major Portuguese cystic fibrosis center. *J Clin Microbiol* **45**, 1628–1633.
- Dalmastri, C., Pirone, L., Tabacchioni, S., Bevivino, A. & Chiarini, L. (2005). Efficacy of species-specific *recA* PCR tests in the identification of *Burkholderia cepacia* complex environmental isolates. *FEMS Microbiol Lett* **246**, 39–45.
- Dalmastri, C., Baldwin, A., Tabacchioni, S., Bevivino, A., Mahenthalingam, E., Chiarini, L. & Dowson, C. (2007). Investigating *Burkholderia cepacia* complex populations recovered from Italian maize rhizosphere by multilocus sequence typing. *Environ Microbiol* **9**, 1632–1639.
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797.
- Ezaki, T., Hashimoto, Y. & Yabuuchi, E. (1989). Fluorometric deoxyribonucleic acid-deoxyribonucleic acid hybridization in microdilution wells as an alternative to membrane filter hybridization in which radioisotopes are used to determine genetic relatedness among bacterial strains. *Int J Syst Bacteriol* **39**, 224–229.
- 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.
- 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.
- Guindon, S. & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**, 696–704.
- Haubold, B. & Hudson, R. R. (2000). LIAN 3.0: detecting linkage disequilibrium in multilocus data. Linkage analysis. *Bioinformatics* **16**, 847–848.
- Henry, D. A., Campbell, M. E., LiPuma, J. J. & Speert, D. P. (1997). Identification of *Burkholderia cepacia* isolates from patients with cystic fibrosis and use of a simple new selective medium. *J Clin Microbiol* **35**, 614–619.
- Henry, D. A., Mahenthalingam, E., Vandamme, P., Coenye, T. & Speert, D. P. (2001). Phenotypic methods for determining genomovar status of the *Burkholderia cepacia* complex. *J Clin Microbiol* **39**, 1073–1078.
- Jolley, K. A., Chan, M.-S. & Maiden, M. C. (2004). mlstdbNet – distributed multi-locus sequence typing (MLST) databases. *BMC Bioinformatics* **5**, 86.
- Jones, D. T., Taylor, W. R. & Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* **8**, 275–282.
- Jukes, T. H. & Cantor, C. R. (1969). Evolution of protein molecules. In *Mammalian Protein Metabolism*, vol. 3, pp. 21–132. Edited by H. N. Munro. New York: Academic Press.
- Kenna, D. T., Yesilkaya, H., Forbes, K. J., Barcus, V. A., Vandamme, P. & Govan, J. R. (2006). Distribution and genomic location of active insertion sequences in the *Burkholderia cepacia* complex. *J Med Microbiol* **55**, 1–10.
- Konstantinidis, K. T. & Tiedje, J. M. (2005a). Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci USA* **102**, 2567–2572.

- Konstantinidis, K. T. & Tiedje, J. M. (2005b). Towards a genome-based taxonomy for prokaryotes. *J Bacteriol* **187**, 6258–6264.
- Kumar, S., Tamura, K. & Nei, M. (2004). MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* **5**, 150–163.
- Lessie, T. G., Hendrickson, W., Manning, B. D. & Devereux, R. (1996). Genomic complexity and plasticity of *Burkholderia cepacia*. *FEMS Microbiol Lett* **144**, 117–128.
- Li, L., Stoeckert, C. J., Jr & Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* **13**, 2178–2189.
- Mahenthalingam, E. & Drevinek, P. (2007). Comparative genomics of *Burkholderia* species. In *Burkholderia: Molecular Microbiology and Genomics*, pp. 53–79. Edited by T. Coenye & P. Vandamme. Wymondham, UK: Horizon Bioscience.
- Mahenthalingam, E., Coenye, T., Chung, J. W., Speert, D. P., Govan, J. R., Taylor, P. & Vandamme, P. (2000a). Diagnostically and experimentally useful panel of strains from the *Burkholderia cepacia* complex. *J Clin Microbiol* **38**, 910–913.
- Mahenthalingam, E., Bischof, J., Byrne, S. K., Radomski, C., Davies, J. E., Av-Gay, Y. & Vandamme, P. (2000b). DNA-based diagnostic approaches for identification of *Burkholderia cepacia* complex, *Burkholderia vietnamiensis*, *Burkholderia multivorans*, *Burkholderia stabilis*, and *Burkholderia cepacia* genomovars I and III. *J Clin Microbiol* **38**, 3165–3173.
- Mahenthalingam, E., Baldwin, A., Drevinek, P., Vanlaere, E., Vandamme, P., LiPuma, J. J. & Dowson, C. G. (2006). Multilocus sequence typing breathes life into a microbial metagenome. *PLoS One* **1**, e17.
- Mahenthalingam, E., Baldwin, A. & Dowson, C. G. (2008). *Burkholderia cepacia* complex bacteria: opportunistic pathogens with important natural biology. *J Appl Microbiol* **104**, 1539–1551.
- Maynard Smith, J., Smith, N. H., O'Rourke, M. & Spratt, B. G. (1993). How clonal are bacteria? *Proc Natl Acad Sci U S A* **90**, 4384–4388.
- Mesbah, M. & Whitman, W. B. (1989). Measurement of deoxyguanosine/thymidine ratios in complex mixtures by high-performance liquid chromatography for determination of the mole percentage guanine + cytosine of DNA. *J Chromatogr* **479**, 297–306.
- Naser, S. M., Dawyndt, P., Hoste, B., Gevers, D., Vandemeulebroecke, K., Cleenwerck, I., Vancanneyt, M. & Swings, J. (2007). Identification of lactobacilli by *pheS* and *rpoA* gene sequence analyses. *Int J Syst Evol Microbiol* **57**, 2777–2789.
- Parke, J. L. & Gurian-Sherman, D. (2001). Diversity of the *Burkholderia cepacia* complex and implications for risk assessment of biological control strains. *Annu Rev Phytopathol* **39**, 225–258.
- Payne, G. W., Vandamme, P., Morgan, S. H., LiPuma, J. J., Coenye, T., Weightman, A. J., Jones, T. H. & Mahenthalingam, E. (2005). Development of a *recA* gene-based identification approach for the entire *Burkholderia* genus. *Appl Environ Microbiol* **71**, 3917–3927.
- 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.
- Rozas, J., Sanchez-Delbarrio, J. C., Messegue, X. & Rozas, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496–2497.
- Souza, A. V., Moreira, C. R., Pasternak, J., Hirata, M. L., Saltini, D. A., Caetano, V. C., Ciosak, S., Azevedo, F. M., Severino, P. & other authors (2004). Characterizing uncommon *Burkholderia cepacia* complex isolates from an outbreak in a haemodialysis unit. *J Med Microbiol* **53**, 999–1005.
- Stanier, R. Y., Palleroni, N. J. & Doudoroff, M. (1966). The aerobic pseudomonads: a taxonomic study. *J Gen Microbiol* **43**, 159–271.
- Storms, V., Van Den Vreken, N., Coenye, T., Mahenthalingam, E., LiPuma, J. J., Gillis, M. & Vandamme, P. (2004). Polyphasic characterisation of *Burkholderia cepacia*-like isolates leading to the emended description of *Burkholderia pyrocinia*. *Syst Appl Microbiol* **27**, 517–526.
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**, 4876–4882.
- Vandamme, P., Holmes, B., Coenye, T., Goris, J., Mahenthalingam, E., LiPuma, J. J. & Govan, J. R. (2003). *Burkholderia cenocepacia* sp. nov. – a new twist to an old story. *Res Microbiol* **154**, 91–96.
- Vanlaere, E., Coenye, T., Samyn, E., Van Den Plas, C., Govan, J., De Baets, F., De Boeck, K., Knoop, C. & Vandamme, P. (2005). A novel strategy for the isolation and identification of environmental *Burkholderia cepacia* complex bacteria. *FEMS Microbiol Lett* **249**, 303–307.
- Vanlaere, E., LiPuma, J. J., Baldwin, A., Henry, D., De Brandt, E., Mahenthalingam, E., Speert, D., Dowson, C. & Vandamme, P. (2008). *Burkholderia latens* sp. nov., *Burkholderia diffusa* sp. nov., *Burkholderia arboris* sp. nov., *Burkholderia seminalis* sp. nov. and *Burkholderia metallica* sp. nov., novel species within the *Burkholderia cepacia* complex. *Int J Syst Evol Microbiol* **58**, 1580–1590.
- Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., Wu, D., Paulsen, I., Nelson, K. & other authors (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66–74.
- Vermis, K., Coenye, T., Mahenthalingam, E., Nelis, H. J. & Vandamme, P. (2002). Evaluation of species-specific *recA*-based PCR tests for genomovar level identification within the *Burkholderia cepacia* complex. *J Med Microbiol* **51**, 937–940.
- Waine, D. J., Henry, D. A., Baldwin, A., Speert, D. P., Honeybourne, D., Mahenthalingam, E. & Dowson, C. G. (2007). Reliability of multilocus sequence typing of the *Burkholderia cepacia* complex in cystic fibrosis. *J Cyst Fibros* **6**, 215–219.
- Wayne, L. G., Brenner, D. J., Colwell, R. R., Grimont, P. A. D., Kandler, O., Krichevsky, M. I., Moore, L. H., Moore, W. E. C., Murray, R. G. E. & other authors (1987). International Committee on Systematic Bacteriology. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int J Syst Bacteriol* **37**, 463–464.