UNIVERSITEIT GENT

FACULTY OF **SCIENCES**

Burkholderia multivorans and beyond: epidemiology, genomics and taxonomy of opportunistic pathogens in cystic fibrosis patients

Charlotte Peeters

Dissertation submitted in fulfillment of the requirements for the degree of Doctor (Ph.D.) of Science: Biochemistry and Biotechnology (Ghent University)

Supervisor:
Prof. Dr. Peter Vandamme

Laboratory of Microbiology (LM-UGent)
Faculty of Sciences
Ghent University

LM-UGent
LABORATORY of MICROBIOLOGY

Peeters, C. (2016). *Burkholderia multivorans* and beyond: epidemiology, genomics and taxonomy of opportunistic pathogens in cystic fibrosis patients. Ph.D. thesis, Ghent University, Belgium.

Printed by University Press, Zelzate, Belgium |http://www.universitypress.be

ISBN-number:

Cover design by Isabelle Muhring.

This work was funded by the Special Research Council of Ghent University.

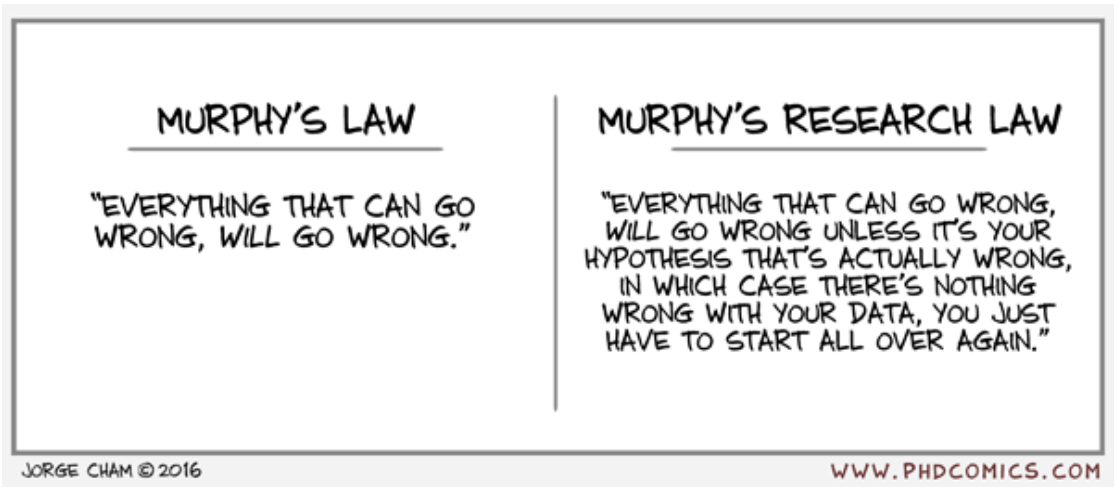
Ph.D. thesis, Faculty of Sciences, Ghent University, Belgium.

Publicly defended in Ghent, Belgium, on September 8, 2016.

Cava

&

rabarbertaart



Examination Committee

Prof. Dr. Savvas SAVVIDES (Chairman)
Laboratory for Protein Biochemistry and Biomolecular Engineering (L-ProBE)
Faculty of Sciences, Ghent University, Ghent, Belgium

VIB Inflammation Research Center
VIB, Ghent, Belgium

Prof. Dr. Peter VANDAMME (Supervisor)
Laboratory of Microbiology (LM-UGent)
Faculty of Sciences, Ghent University, Ghent, Belgium

Prof. Dr. Aurélien CARLIER (Secretary)
Laboratory of Microbiology (LM-UGent)
Faculty of Sciences, Ghent University, Ghent, Belgium

Prof. Dr. Tom COENYE
Laboratory of Pharmaceutical Microbiology (LPM)
Faculty of Pharmaceutical Sciences, Ghent University, Ghent, Belgium

Prof. Dr. Kurt HOUF
Department of Veterinary Public Health and Food Safety
Faculty of Veterinary Medicine, Ghent University, Ghent, Belgium

Prof. Dr. Leo EBERL
Department of Microbiology
Institute of Plant Biology, University of Zürich, Zürich, Switzerland

English Summary

Burkholderia cepacia complex (Bcc) bacteria are significant pathogens in people with cystic fibrosis (CF) because of their high impact on morbidity, mortality and post-lung transplant survival. Epidemiological studies revealed that *Burkholderia multivorans* is the most prevalent Bcc CF pathogen in many countries, including Belgium. The continued emergence of unique *B. multivorans* strains in CF patients suggests acquisition from non-human sources, such as the natural environment. Although semi-selective growth media have been developed for the isolation of Bcc bacteria from environmental samples, *B. multivorans* has thus far only rarely been isolated from such samples and its environmental niche is considered unknown. Therefore, the first goal of the present thesis was to gain a better insight into the epidemiology of *B. multivorans* by examining its environmental niche and by comparing the genomes of clinical and environmental *B. multivorans* isolates.

A first study assessed the environmental occurrence of *B. multivorans* in water and soil samples from Flanders (Belgium) using a fast, cultivation-independent PCR assay. *B. multivorans* was detected in 11 % of the water samples and 92 % of the soil samples, demonstrating that *B. multivorans* DNA is present in water and – to a greater extent – soil samples.

In a second study, cultivation strategies were evaluated and optimized for the isolation of *B. multivorans* from the PCR positive samples from the first study. These included direct plating and liquid enrichment procedures and the use of semi-selective and diluted isolation media, acclimatizing recovery and co-cultivation with amoebae, seedlings, CF sputum and helper strain panels. However, none of these approaches yielded *B. multivorans* isolates from PCR positive water and soil samples. Nonetheless, many non-Bcc *Burkholderia* bacteria, several Gram-negative non-fermenting bacteria (including *Cupriavidus*, *Inquilinus*, *Pandoraea*, *Pseudomonas* and *Stenotrophomonas*) and rapidly growing mycobacteria were all isolated from water and soil samples. The use of Bcc isolation media thus yielded a surprisingly wide array of rare but often clinically relevant CF pathogens, confirming that water and soil are

potential reservoirs for these opportunistic CF pathogens.

Finally, a third study examined to which extent *B. multivorans* isolates with the same multilocus sequence type (ST) but from different origin differ in their genetic potential. Therefore, eight isolates were selected, representing four distinct STs. For each ST, a CF and an environmental isolate were sequenced using the PacBio SMRT sequencing technology, resulting in eight high-quality *B. multivorans* genome assemblies. The genomic structure of *B. multivorans* was shown to be highly conserved and the genomic lineages were defined by their ST. The finding that the ST predicts both phylogeny and gene content of *B. multivorans* isolates corroborates the use of multilocus sequence typing (MLST) for epidemiological surveillance of Bcc bacteria.

Because the CF lung can harbor a wide range of bacteria, accurate identification of CF pathogens at the species level is important to assess the clinical impact of these generally rare opportunistic pathogens. Furthermore, accurate typing of CF pathogens is necessary to identify outbreaks, gain insight into their epidemiology and improve infection control guidelines. Therefore, the second goal of the present thesis was to contribute to the general knowledge of the prevalence and epidemiology of Gram-negative non-fermenting bacteria in CF patients.

Since January 2011 a National Reference Center (NRC) is charged with the surveillance of respiratory infections caused by Gram-negative non-fermenting bacilli in Belgian CF patients. LM-UGent is part of the NRC Bcc and is responsible for the molecular identification and typing of these CF bacteria. The NRC Bcc results of 2011-2015 demonstrated that *B. multivorans* is still the most prevalent Bcc species in Belgian CF patients. Although the *B. multivorans* isolates from Belgian CF patients mostly represented unique strains, three *B. multivorans* strains were found to be present in multiple Belgian CF patients and in multiple CF centers.

The ever-decreasing cost of high-throughput sequencing technologies is revolutionizing prokaryotic taxonomy. The taxonomic work performed in the context of the present thesis led to the formal classification of novel species in the genera *Burkholderia*, *Achromobacter* and *Bordetella*. Throughout the present thesis, an evolution in taxonomic methods can be observed with regards to phylogeny and species delineation. In this genomics era, DNA-DNA hybridization is gradually being replaced by multilocus sequence analysis (MLSA) and whole-genome sequence-based parameters such as digital DNA-DNA hybridization (dDDH). In the present thesis, the application of MLSA led to the formal classification of three novel Bcc and four novel *Achromobacter* species, while the application of the 70 % dDDH threshold

value led to the formal classification of 13 novel *Burkholderia glathei*-like species.

Although we described three novel Bcc species and more than a dozen novel *B. glathei*-like species, there is still an overwhelming number of putative novel *Burkholderia* species awaiting formal classification. In the future, a robust, whole-genome sequence-based species definition will contribute to classify the backlog of diversity awaiting formal description and the huge diversity of bacteria that is yet to be discovered.

Nederlandse Samenvatting

Burkholderia cepacia complex (Bcc) bacteriën zijn belangrijke pathogenen voor mensen met mucoviscidose, een genetische aandoening die ook wel *cystic fibrosis* (CF) genoemd wordt. Bcc longinfecties kunnen een zeer negatieve invloed hebben op de gezondheidstoestand van CF patiënten en zijn een belangrijke doodsoorzaak. Eerdere epidemiologische studies toonden aan dat *Burkholderia multivorans* de meest voorkomende Bcc CF pathogeen is in veel landen, waaronder België. Omdat *B. multivorans* isolaten uit CF patiënten veelal tot unieke stammen behoren, wordt de omgeving als een mogelijke infectiebron beschouwd. In eerdere studies werden reeds verschillende Bcc bacteriën uit omgevingsstalen geïsoleerd met behulp van semi-selectieve isolatiemedia, maar deze waren slechts zelden *B. multivorans* en de niche van dit Bcc species is dus ongekend. Het eerste doel van deze thesis was daarom een beter inzicht te genereren in de epidemiologie van *B. multivorans* door diens omgevingsniche te onderzoeken en door genomen te vergelijken van *B. multivorans* isolaten die geïsoleerd werden uit enerzijds de natuurlijke omgeving en anderzijds CF patiënten.

De eerste studie onderzocht aan de hand van een snelle, cultuur-onafhankelijke PCR analyse de aanwezigheid van *B. multivorans* in water- en grondstalen uit Vlaanderen (België). *B. multivorans* DNA werd gedetecteerd in 11 % van de waterstalen en in 92 % van de grondstalen.

Een tweede studie had als doel om *B. multivorans* te isoleren uit de PCR positieve stalen van de eerste studie. Hiervoor werden bestaande cultivatiemethoden geëvalueerd en geoptimaliseerd, waaronder uitplating op vaste media, vloeibare aanrijking, het gebruik van semi-selectieve en nutriëntarme isolatiemedia en co-cultivatie met amoeben, kiemplantjes, CF sputum en andere CF pathogenen. Hoewel geen enkel experiment *B. multivorans* isolaten opleverde uit PCR positieve water- en grondstalen, werden wel non-Bcc *Burkholderia* bacteriën verkregen. Daarenboven werden ook tal van Gram-negatieve niet-fermenterende bacteriën geïsoleerd (waaronder *Cupriavidus*, *Inquilinus*, *Pandoraea*, *Pseudomonas* en *Stenotrophomonas*), alsook

verschillende mycobacteriën. De toepassing van Bcc isolatiemedia op omgevingsstalen maakte het dus mogelijk om een diverse set van zeldzame, maar klinisch erg relevante CF pathogenen te isoleren en toont aan dat zowel water als grond een potentieel reservoir zijn voor opportunistische CF pathogenen.

In een derde studie tenslotte werd nagegaan of omgevingsisolaten en CF isolaten van *B. multivorans* over hetzelfde genetisch potentieel beschikken, en in welke mate het multilocus sequentie type (ST) de genoominhoud kan voorspellen. Hiervoor werden acht isolaten geselecteerd die tot vier verschillende STs behoorden. Voor elk ST werd van één omgevingsisolaat en van één CF isolaat het genoom gesequeneerd met behulp van de PacBio SMRT technologie, resulterend in acht *B. multivorans* genoomsequenties van hoge kwaliteit. De genoomstructuur van *B. multivorans* bleek erg geconserveerd en de genoominhoud werd gedefinieerd door het ST. De bevinding dat op basis van het ST voorspellingen gemaakt kunnen worden over zowel de fylogenie als de genoominhoud van *B. multivorans* isolaten bevestigt het gebruik van multilocus sequentie typering (MLST) voor epidemiologische studies van Bcc bacteriën.

Omdat de longinfecties in CF patiënten veroorzaakt kunnen worden door een brede variëteit aan bacteriën is een accurate identificatie van CF pathogenen noodzakelijk om de klinische impact van deze opportunistische pathogenen te kunnen onderzoeken. Daarnaast is een accurate typering van CF pathogenen nodig om uitbraken vast te stellen, om inzichten te verkrijgen in de epidemiologie, en om richtlijnen en interventies met oog op infectiecontrole te verbeteren. Het tweede doel van deze thesis bestond er daarom in bij te dragen aan de algemene kennis over de prevalentie en de epidemiologie van Gram-negatieve niet-fermenterende bacteriën in CF patiënten.

Sinds Januari 2011 is een Nationaal Referentie Centrum (NRC) belast met het toezicht op luchtwegeninfecties in Belgische CF patiënten die veroorzaakt worden door Gram-negatieve niet-fermenterende bacillen. LM-UGent maakt deel uit van het NRC Bcc en is verantwoordelijk voor de moleculaire identificatie en typering van deze CF bacteriën. De NRC Bcc resultaten van 2011-2015 toonden aan dat *B. multivorans* nog steeds het meest voorkomende Bcc species is in Belgische CF patiënten. Hoewel de meeste *B. multivorans* isolaten van Belgische CF patiënten tot unieke stammen behoorden, werden ook drie *B. multivorans* stammen teruggevonden die aanwezig waren in meerdere Belgische CF patiënten en in meerdere CF centra.

De steeds afnemende kostprijs van de *high-throughput* sequeneringsmethoden heeft een revolutie veroorzaakt in de taxonomische methodologie. De taxonomische studies die werden uitgevoerd tijdens deze thesis resulteerden in de formele classificatie van nieuwe species in de

genera *Burkholderia*, *Achromobacter* en *Bordetella*. Doorheen deze thesis is een evolutie waar te nemen in de toegepaste taxonomische methoden met betrekking tot fylogenie en species afbakening. In dit tijdperk van genoomanalyses wordt de traditionele DNA-DNA hybridisatie geleidelijk aan vervangen door multilocus sequentie analyse (MLSA) en genoomgebaseerde parameters zoals digitale DNA-DNA hybridisatie (dDDH). De toepassing van MLSA in deze thesis resulteerde in de formele classificatie van drie nieuwe Bcc en vier nieuwe *Achromobacter* species, en de toepassing van de 70 % dDDH grenswaarde leidde tot de formele classificatie van 13 nieuwe species in de *Burkholderia glathei* groep.

Hoewel tijdens deze thesis drie nieuwe Bcc species en 14 nieuwe *B. glathei*-gerelateerde species werden beschreven is er nog steeds een indrukwekkend aantal nieuwe *Burkholderia* species dat nog niet formeel geclassificeerd werd. Een robuuste, genoomgebaseerde species definitie zal het in de toekomst mogelijk maken om de grote, nog gedeeltelijk ongekende, diversiteit aan bacteriën te inventariseren.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

A

AAI	average amino acid identity
AF	alignment fraction
ANI	average nucleotide identity

B

B.	*Burkholderia*
Bcc	*Burkholderia cepacia* complex
BCEM	*Burkholderia cepacia* complex enrichment medium
BCSA	*Burkholderia cepacia* selective agar
BGC	*Burkholderia glathei* clade
BLAST	basic local alignment search tool
bp	base pairs

C

CCUG	culture collection university of Göteborg, Sweden
CDS	protein-coding sequence
CF	cystic fibrosis
CFTR	cystic fibrosis transmembrane conductance regulator
COG	cluster of orthologous groups

D

DDH	DNA-DNA hybridization
dDDH	digital DNA-DNA hybridization
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate

E

ENV	environmental

F

FAME	fatty acid methyl ester

G

GBDP	genome BLAST distance phylogeny
GGDC	genome-to-genome distance calculator
GNNF	Gram-negative non-fermenting

L

LM-UGent	laboratory of microbiology, Ghent university
LMG	BCCM/LMG bacteria collection, laboratory of microbiology, Ghent university, Belgium

M

MALDI-TOF MS	matrix-assisted laser desorption/ionization time-of-flight mass spectrometry
MEGA	molecular evolutionary genetics analysis
MES	4-morpholine-ethane-sulfonic acid
MLSA	multilocus sequence analysis
MLST	multilocus sequence typing
MSMB	Menzies school of health research, Darwin, Australia
MUMi	maximal unique matches index

N

NRC	national reference center

P

PCAT	*Pseudomonas cepacia* azelaic acid tryptamine
pcDNA	percentage of conserved DNA
PCR	polymerase chain reaction
PGPR	plant growth promoting rhizobacteria

Q

qPCR	quantitative PCR

R

RAPD	randomly amplified polymorphic DNA
RFLP	restriction fragment length polymorphism
rMLST	ribosomal MLST
RNA	ribonucleic acid
rpm	rotations per minute
rRNA	ribosomal RNA

S

SMRT	single-molecule real-time
ST	multilocus sequence type

T

T	type
TB-T	trypan blue tetracycline
TSA	tryptone soya agar

U

UPGMA	unweighted pair group method with arithmetic mean

Part I

Introduction

1 | *Burkholderia cepacia* complex bacteria as opportunistic pathogens in cystic fibrosis patients

1.1 Cystic fibrosis lung pathology

Cystic fibrosis (CF) is an autosomal recessive genetic disorder that results from a mutation in the CF transmembrane conductance regulator (CFTR) gene. This gene encodes a membrane protein which is involved in transepithelial ion transport and its defect leads to the secretion of viscous secretions in all organs (Rowe *et al.*, 2005). CF is most common in the northern European population but the birth prevalence varies for different countries and ethnic backgrounds. The incidence of CF in Belgium is estimated at 1:2850 (Farrell, 2008). In total, more than 1500 mutations have been identified but the clinical significance is known only for a very small fraction of mutations. The best characterized mutation is F508del, representing a deletion of phenylalanine at position 508, which accounts for two-thirds of the mutations in northern European and North American populations (O'Sullivan & Freedman, 2009).

The symptoms of CF are mainly evident in the gastrointestinal and respiratory tracts and chronic inflammation and respiratory infections represent the main causes of morbidity and mortality. Several hypotheses exist about how CFTR dysfunction leads to phenotypic symptoms, and possibly all of them contribute to the outcome of recurrent respiratory infections and chronic inflammation. The constant cycle of inflammation and infection leads to permanent lung damage and may eventually result in death by pulmonary insufficiency (O'Sullivan & Freedman, 2009).

The accumulation of mucus in the CF lung creates an environment suitable for colonization

by various opportunistic pathogens, resulting in a complex community consisting of bacteria, fungi and viruses. The recurrent bacterial respiratory infections in CF are caused by a variety of organisms and their prevalence typically varies with the age of the CF patient (Fig. 1.1) (Harrison, 2007). A recent study showed that the microbial community in the lower airways of children with CF is less rich and less even compared with that of non-CF children and that the CF lung microbial community is disrupted early in life (Renwick *et al.*, 2014). Moreover, a longitudinal study demonstrated that the community diversity is decreased in patients with more progressive lung disease and that this decrease in diversity in driven primarily by antibiotic therapy (Zhao *et al.*, 2012). Similarly, two recent studies showed that a loss of microbial diversity is associated with severe lung disease and the presence of *Burkholderia* bacteria (Flight *et al.*, 2015; Stokell *et al.*, 2015).

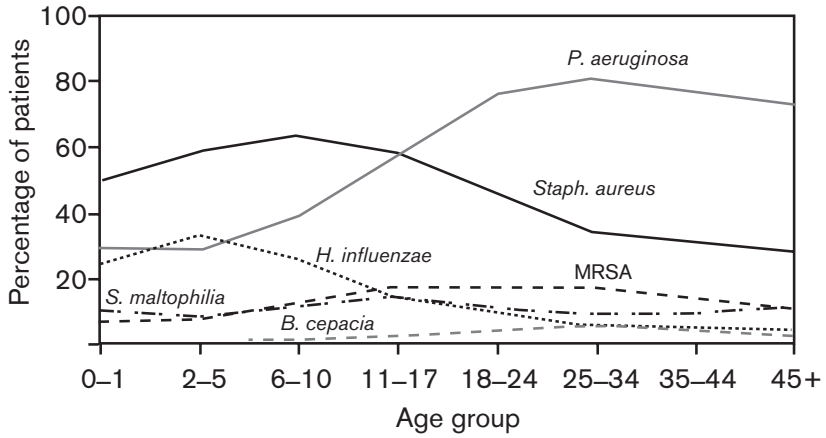


Figure 1.1: Prevalence of bacterial respiratory infections in CF patients by age group. MRSA, meticillin-resistant *Staphylococcus aureus*. (Harrison, 2007)

Because the CF lung can harbor a wide diversity of bacteria, accurate identification is crucial to assess the prevalence and clinical relevance of these generally rare opportunistic pathogens (Coenye *et al.*, 2002a; LiPuma, 2010). One important group of opportunistic pathogens consists of the Gram-negative non-fermenting (GNNF) bacteria, including the genera *Pandoraea*, *Inquilinus*, *Achromobacter*, *Ralstonia*, *Cupriavidus*, *Pseudomonas* and *Stenotrophomonas* (Vandamme & Dawyndt, 2011; Alby *et al.*, 2013). Although the clinical significance of these *B. cepacia*-like organisms is mostly unclear, infection with these GNNF opportunistic pathogens may have a high impact on morbidity because of their multi-resistance to antibiotics and their transmissibility (Coenye & Vandamme, 2003; Jorgensen *et al.*, 2003; Schmoldt *et al.*, 2006; Kalka-Moll *et al.*, 2009).

1.2 *Burkholderia cepacia* complex epidemiology in cystic fibrosis patients

Burkholderia cepacia complex (Bcc) bacteria are significant pathogens in people with CF because of their high impact on morbidity, mortality and post-lung transplant survival (De Boeck *et al.*, 2004; Govan *et al.*, 2007; LiPuma, 2010). Bcc infection in CF is characterized by a highly variable clinical outcome, but generally results in a progressive decline of lung function. In extreme cases, Bcc infection can result in "cepacia syndrome", a necrotizing pneumonia and septicemia resulting in early death. Bcc infections are difficult to eradicate because Bcc bacteria show an innate multiresistance to antibiotics. Finally, the diagnosis of Bcc infection also has a social and psychological impact because the transmissibility of Bcc strains through social contact resulted in stringent infection control guidelines that necessitate patient segregation (Mahenthiralingam *et al.*, 2005).

Bcc bacteria rarely cause infection in healthy (i.e. non-immunocompromised) individuals, except when they are present as contaminants in pharmaceutical products (Mahenthiralingam *et al.*, 2005). Several studies reported on disinfectants, intravenous solutions and contaminated medical devices as sources of nosocomial outbreaks (Weber *et al.*, 2007; Heo *et al.*, 2008; Torbeck *et al.*, 2011; Souza Dias *et al.*, 2013).

Of the 20 formally named species within the Bcc (Fig. 2.1 and 2.2), *Burkholderia multivorans* and *Burkholderia cenocepacia* are generally the most prevalent Bcc species in CF. During the last decade, infection control measures have restricted patient-to-patient transmission, which is mostly associated with *B. cenocepacia* (Mahenthiralingam *et al.*, 2001; Speert *et al.*, 2002; Turton *et al.*, 2003; France *et al.*, 2008). Consequently, *B. multivorans* emerged as the most prevalent Bcc pathogen in many countries, including Belgium, France, Denmark, the United Kingdom, the United States (Fig. 1.2) and New Zealand (Brisse *et al.*, 2004; De Boeck *et al.*, 2004; Govan *et al.*, 2007; LiPuma, 2010; Norskov-Lauritsen *et al.*, 2010; Pope *et al.*, 2010). In Spain, Portugal and Argentina a high prevalence has recently been reported of *B. contaminans* in CF patients (Martina *et al.*, 2013; Coutinho *et al.*, 2015; Medina-Pascual *et al.*, 2015).

Historically, *B. cenocepacia* strains have been responsible for large epidemics within the CF community and are often extremely virulent (Drevinek & Mahenthiralingam, 2010). The ET12 clone is responsible for infecting many CF patients in Canada and Europe (Mahenthiralingam *et al.*, 2002), while the Midwest clone (Coenye & LiPuma, 2002) and the PHDC clone (Chen

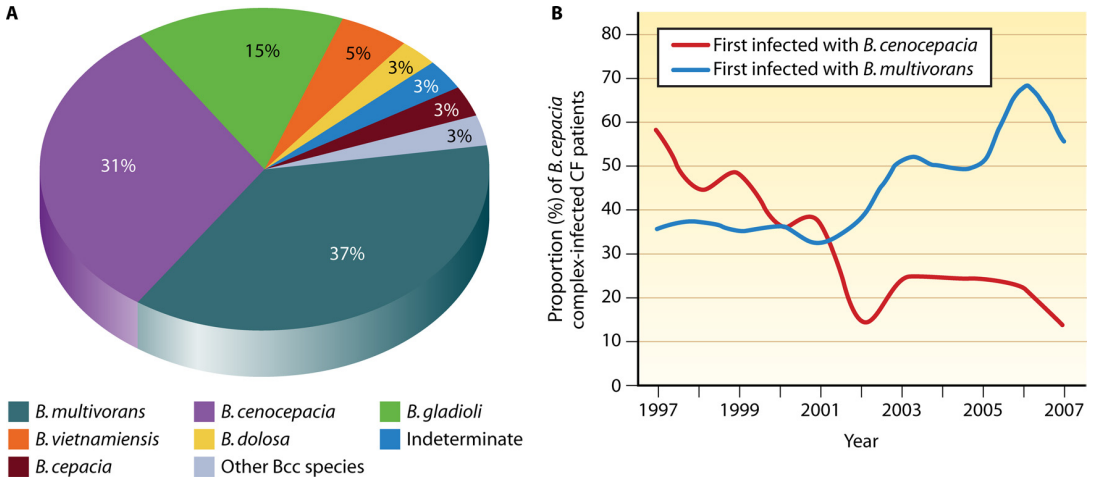


Figure 1.2: Distribution and incidence of *Burkholderia* in U.S. CF patients. **(A)** Distribution of *Burkholderia* species among U.S. CF patients. The proportions of CF patients infected with various *Burkholderia* species are shown. The data are based on 2,024 CF patients who were infected with *Burkholderia* species and whose isolates were referred to the *Burkholderia cepacia* Research Laboratory and Repository (University of Michigan) between 1997 and 2007. "Other Bcc species" indicates patients infected with Bcc species other than those specified in the chart. "Indeterminate" refers to patients infected with strains that phylogenetically are members of the Bcc species but that cannot be definitively placed into one of the 17 defined species in this group. **(B)** Incidence of *B. cenocepacia* and *B. multivorans* infection in U.S. CF patients. The proportions of Bcc-infected CF patients who were first infected with either *B. cenocepacia* (red line) or *B. multivorans* (blue line) in the years indicated are shown. (LiPuma, 2010)

et al., 2001) are dominant strains infecting CF patients in the United States. Although *B. multivorans* is generally considered a lesser virulent Bcc pathogen compared to *B. cenocepacia*, rare cases of cepacia syndrome caused by *B. multivorans* have been reported (Blackburn *et al.*, 2004; Jones *et al.*, 2004). While preoperative Bcc infection is generally associated with a poor prognosis for post-lung transplant survival, mortality is primarily dependent on the Bcc species and strain (Aris *et al.*, 2001; De Soyza *et al.*, 2001; Murray *et al.*, 2008).

The ability to differentiate Bcc strains has been crucial in understanding their epidemiology and improving infection control guidelines for the CF community. Methods for strain differentiation include multilocus restriction typing (Coenye & LiPuma, 2002), pulsed field gel electrophoresis (Coenye *et al.*, 2002b), randomly amplified polymorphic DNA (RAPD) analysis (Mahenthiralingam *et al.*, 1996), BOX-PCR fingerprinting (Coenye *et al.*, 2002b), multilocus sequence typing (MLST) (Baldwin *et al.*, 2005) and single-nucleotide polymorphism analysis of whole-genome sequences (Pallen *et al.*, 2010; Lieberman *et al.*, 2011). MLST is a well-established method for studying the population structure of Bcc organisms and takes

into account the allelic variation of seven housekeeping genes (*atpD*, *gltB*, *gyrB*, *recA*, *lepA*, *phaC* and *trpB*) (Baldwin *et al.*, 2005). As a result, each strain is defined by its unique allelic profile and multilocus sequence type (ST). Only a limited number of *B. multivorans* outbreak strains was described and for all cases of implied patient-to-patient spread, unique STs were found (Baldwin *et al.*, 2008). In the United Kingdom, the strains ST-27 and ST-15 caused hospital outbreaks among CF patients in Glasgow and South Wales, respectively (Whiteford *et al.*, 1995; Millar-Jones *et al.*, 1998). Strains ST-25 and ST-179 were shared by CF patients in the United States (Biddick *et al.*, 2003). Finally, strain ST-419 caused an outbreak among CF patients in France (Segonds *et al.*, 1999).

The low number of outbreaks caused by *B. multivorans* and the fact that *B. multivorans* isolates from CF patients commonly represent unique strains suggest that there is only limited person-to-person transmission and that *B. multivorans* strains are acquired from non-human sources such as the natural environment (Baldwin *et al.*, 2008).

1.3 The natural environment as a reservoir for cystic fibrosis pathogens

Bcc bacteria are characterized by a versatile lifestyle and almost all Bcc species have been isolated from both CF patients and environmental samples (Coenye & Vandamme, 2003; Baldwin *et al.*, 2007). Exceptions are *Burkholderia ubonensis* and *Burkholderia territorii* which have not been isolated from CF patients thus far (http://pubmlst.org/bcc/) (Jolley & Maiden, 2010). Baldwin *et al.* (2008) demonstrated a continued emergence of unique *B. multivorans* strains in the CF population, with several STs being globally distributed and associated with human infection. The same study showed an overlap in strains between water environments, industrial products and human infection, suggesting that the environment may be an important reservoir for infection with *B. multivorans* (Baldwin *et al.*, 2008).

Specific media have been developed for the isolation of Bcc from environmental samples such as *Pseudomonas cepacia* azelaic acid tryptamine (PCAT) (Burbage & Sasser, 1982), trypan blue tetracycline (TB-T) (Hagedorn *et al.*, 1987) and Bcc enrichment medium (BCEM) (Vermis *et al.*, 2003a; Vanlaere *et al.*, 2005). The selectivity of these media is achieved by specific nitrogen and carbon sources, often combined with antimicrobial compounds to inhibit other bacteria and fungi (Vermis *et al.*, 2003b). Although many studies described the isolation of *B. cenocepacia* from rhizosphere samples (Butler *et al.*, 1995; Balandreau *et al.*, 2001;

Fiore *et al.*, 2001; Bevivino *et al.*, 2002; LiPuma *et al.*, 2002; Pirone *et al.*, 2005; Zhang & Xie, 2007; Bevivino *et al.*, 2011; Hall *et al.*, 2015), only few studies reported the isolation of *B. multivorans* from maize rhizosphere (Ramette *et al.*, 2005) and agricultural soil (Lin *et al.*, 2011; Hsueh *et al.*, 2015). Vermis *et al.* (2003b) developed a selective enrichment broth for the isolation of Bcc bacteria from water samples and obtained eight *B. multivorans* isolates from the Schelde river in Belgium. Using the same strategy Vanlaere *et al.* (2005) obtained two *B. multivorans* isolates from soil in a veranda. Finally, a more recent study described the isolation of 48 *B. multivorans* isolates from West Lake, China (Fang *et al.*, 2011). Yet together, the number of environmental *B. multivorans* isolates is very small compared to the number of CF isolates and its true environmental niche is considered unknown.

Environmental pressure can select for traits that confer virulence and natural environments could therefore serve as potential reservoirs of opportunistic pathogens (Coenye & Vandamme, 2003; Berg *et al.*, 2005). Bcc bacteria adopt a wide range of lifestyles (Section 2.2) and the very same mechanisms of adaptation that let these bacteria thrive in different ecological niches may enable them to colonize the CF lung and cause infection (Vial *et al.*, 2011). Therefore, the natural environment may be considered a fitness school for emerging pathogens.

2 | Taxonomy and diversity of the genus *Burkholderia*

2.1 Taxonomy of the genus *Burkholderia*

2.1.1 Historical background

The genus *Burkholderia* originated in the 1940s when plant physiologist Walter Burkholder isolated bacteria from onion bulbs that were suffering from "sour skin" disease and proposed to name these strains *Pseudomonas cepacia* (Burkholder, 1950). Based on rRNA-based methods, the genus *Pseudomonas* was shown to be phylogenetically diverse and to consist of five major clusters, the so called rRNA homology groups (Palleroni *et al.*, 1973). In 1992, the new genus *Burkholderia* was proposed to accommodate RNA homology group II which then comprised *P. cepacia* and six other species (Yabuuchi *et al.*, 1992). During the last two decades, the genus *Burkholderia* expanded dramatically and it became apparent that this genus comprised versatile bacteria which occupy a wide range of ecological niches (Coenye & Vandamme, 2003). In March 2016, the genus comprised 90 validly named species (Euzeby, 1997) and a large number of uncultivated *Candidatus* species (Van Oevelen *et al.*, 2002, 2004; Lemaire *et al.*, 2011a, 2012; Verstraete *et al.*, 2011). However, literature data and analysis of publicly available 16S rRNA gene sequences suggest that many unclassified *Burkholderia* isolates represent additional novel species, showing that the diversity of *Burkholderia* bacteria is still not fully uncovered.

Organisms now known as *Burkholderia cepacia* emerged as important CF pathogens in the 1980s (Isles *et al.*, 1984; LiPuma *et al.*, 1990; Govan *et al.*, 1993). Only after its formal reclassification in the genus *Burkholderia* it became apparent that presumed *B. cepacia*

strains showed a marked heterogeneity (Butler *et al.*, 1995; Gillis *et al.*, 1995). This led to a polyphasic taxonomic study which revealed that presumed *B. cepacia* strains from CF patients in fact belonged to at least five distinct genomic species or genomovars, collectively referred to as the *B. cepacia* complex (Vandamme *et al.*, 1997). Bcc species are closely related and typically show 30-50 % DNA-DNA hybridization (DDH). Only two out of the five closely related genomic species could also be distinguished phenotypically at that time and these were *Burkholderia vietnamiensis* and *B. multivorans*. The term "genomovar" was used to denote groups of strains that are delimited by DDH but that are phenotypically similar so that they cannot be differentiated (Ursing *et al.*, 1995). The used polyphasic approach was based on multiple techniques including whole-cell protein electrophoresis, whole-cell fatty acid methyl ester analysis, DNA-DNA and DNA-rRNA hybridizations and several biochemical tests and set an example for a taxonomic consensus approach to bacterial systematics in which the results of several genotypical and phenotypical tests are integrated (Vandamme *et al.*, 1996).

Driven by the devastating effect of Bcc infections in CF patients, a wide range of alternative identification methods were developed, including species-specific 16S rRNA (LiPuma *et al.*, 1999; Whitby *et al.*, 2000) and *recA* PCR assays (Mahenthiralingam *et al.*, 2000a; Drevinek *et al.*, 2002; Vermis *et al.*, 2002a), amplified fragment length polymorphism analysis of genomic DNA (Coenye *et al.*, 1999b), ribotyping (Brisse *et al.*, 2000), restriction fragment length polymorphism (RFLP) analysis of PCR amplified 16S rRNA and *recA* gene fragments (Mahenthiralingam *et al.*, 2000a; McDowell *et al.*, 2001; Vermis *et al.*, 2002b) and tRNA profiling (Storms *et al.*, 2002). The sensitivity and specificity of these first-generation molecular identification approaches needed re-evaluation each time novel Bcc species were described and several misidentifications were reported (Coenye *et al.*, 1999b; McMenamin *et al.*, 2000; Moore *et al.*, 2002; Cesarini *et al.*, 2009; Drevinek *et al.*, 2010). At present, sequence analysis of *recA* and other housekeeping genes have been established as powerful, objective and portable taxonomic tools that are useful for the accurate identification and classification of Bcc organisms (Section 2.1.3.1) (Vandamme & Dawyndt, 2011). As more CF isolates were examined and techniques with a better resolution became available, additional Bcc species could be distinguished both genotypically and phenotypically and were formally classified. Today (March 2016) the Bcc comprises 20 validly named species (De Smet *et al.*, 2015a).

2.1.2 Taxonomic overview of the genus *Burkholderia* based on the 16S rRNA gene

Based on 16S rRNA gene sequence analysis, the genus *Burkholderia* consists of several deep-branching lineages (Fig. 2.1). The type species *B. cepacia* is part of a first clade of about 30 species that comprises the Bcc, species closely related to the risk class 3 pathogens *Burkholderia mallei* and *Burkholderia pseudomallei* and a group of plant-pathogenic species that includes *Burkholderia gladioli*, *Burkholderia plantarii* and *Burkholderia glumae*. This clade (from here on referred to as 16S rRNA clade 1) comprises the well-known human pathogens in this genus (Section 2.2.3), but also includes strains that show potential for plant-growth promotion and biocontrol (Section 2.2.1).

A second deep-branching *Burkholderia* clade (from here on referred to as 16S rRNA clade 2) comprises *Burkholderia glathei* and 11 validly named *Burkholderia* species (Fig. 2.1). Most species in this clade have been isolated as free-living organisms from soil (Zolg & Ottow, 1975; Vandamme *et al.*, 2013a; Draghi *et al.*, 2014; Baek *et al.*, 2015) but also associations with fungi (Lim *et al.*, 2003), plants (Tian *et al.*, 2013) and insects (Kikuchi *et al.*, 2011) have been described for this clade. Additionally, many uncultivated species adapted a endosymbiotic lifestyle in plant tissue and represent candidate species (Verstraete *et al.*, 2013; Carlier *et al.*, 2016). Exemplary for the versatile metabolism of *Burkholderia* bacteria, several *B. glathei* clade species were isolated from contaminated soil (Vandamme *et al.*, 2013a; Liu *et al.*, 2014) or from a wastewater treatment system (Lu *et al.*, 2012). The *B. glathei* clade thus far includes only 12 formally classified species, yet many unclassified *B. glathei*-like bacteria have been reported (Nogales *et al.*, 2001; Salles *et al.*, 2006b; Pumphrey & Madsen, 2008; Draghi *et al.*, 2014; Verstraete *et al.*, 2014).

A third deep-branching *Burkholderia* clade (from here on referred to as 16S rRNA clade 3) comprises more than 40 species which are primarily plant-associated beneficial and environmental (Suarez-Moreno *et al.*, 2012) (Fig. 2.1). However, *Burkholderia fungorum*, *Burkholderia ginsengisoli*, *Burkholderia tropica* and *Burkholderia xenovorans* represent important exceptions in this clade because they have been isolated from human clinical samples too (Section 2.2.3).

In addition to these three large clades, several *Burkholderia* species represent rather unique 16S rRNA lineages and do not cluster closely with any other *Burkholderia* species (Fig. 2.1). Their 16S rRNA-based phylogenetic position is variable and dependent on the remaining sequences included in the phylogenetic analysis. *Burkholderia rhizoxinica* and *Burkholderia*

endofungorum represent a first lineage that includes bacterial endosymbionts of a plant-pathogenic fungus (Partida-Martinez *et al.*, 2007). A second lineage comprises *Burkholderia caryophylli*, *Burkholderia symbiotica* and *Burkholderia soli* which are known as a plant pathogen, a root nodule endosymbiont of Mimosa species and a soil bacterium, respectively (Yabuuchi *et al.*, 1992; Yoo *et al.*, 2007; Sheu *et al.*, 2012). Finally, *Burkholderia andropogonis* is a plant pathogen (Coenye *et al.*, 2001a) that represents a deep-branching lineage within 16S rRNA clade 2 in the present analysis (Fig. 2.1) but which occupies a distinct position in most 16S rRNA-based phylogenetic trees (Gyaneshwar *et al.*, 2011; Suarez-Moreno *et al.*, 2012; Estrada-de los Santos *et al.*, 2013, 2015)).

2.1.3 *Burkholderia* taxonomy in a changing landscape

2.1.3.1 Gene sequence analysis

While the 16S rRNA gene phylogeny gives a traditional representation of the phylogeny of *Burkholderia* bacteria, this gene only has limited taxonomic resolution for species identification within the genus *Burkholderia*. Within the Bcc, 16S rRNA sequences of different Bcc species typically show more than 98 % similarity, while exhibiting at the same time up to 2 % intraspecies divergence (Vandamme & Dawyndt, 2011). In contrast with the 16S rRNA gene, the *recA* gene is a powerful taxonomic tool for Bcc species identification (Mahenthiralingam *et al.*, 2000a). *RecA* sequences typically show 94-95 % similarity between Bcc species and, mostly, about 98-99 % similarity within Bcc species (Vandamme & Dawyndt, 2011). Remarkably, *B. cenocepacia* (formerly known as genomovar III) comprises four *recA* lineages referred to as IIIA through IIID, with most clinical strains residing in clusters IIIA and IIIB (Vandamme *et al.*, 2003). Another housekeeping gene that proved useful for species identification within the genus *Burkholderia* is *gyrB* (Tayeb *et al.*, 2008). This gene serves as a reliable taxonomic tool for species-level discrimination of both Bcc (Tabacchioni *et al.*, 2008) and non-Bcc *Burkholderia* species (Vandamme *et al.*, 2013a).

Multilocus sequence analysis (MLSA) is a technique that uses a set of genes (usually six or seven) that are universally present within the taxon, occur as a single copy in the genome, are not closely linked and are not subject to high levels of recombination (Gevers *et al.*, 2005). With the advent of new sequencing technologies, MLSA was proposed to replace DDH as a tool for species delineation (Stackebrandt *et al.*, 2002; Gevers *et al.*, 2005). The Bcc MLST scheme was developed for differentiation of Bcc isolates at both strain and species level so that the same sequence data that was used for strain differentiation (Section 1.2) could

also be employed for identification. Phylogenetic analysis of housekeeping gene fragments of *atpD*, *gltB*, *gyrB*, *recA*, *lepA*, *phaC* and *trpB* has a superior taxonomic resolution for species level identification within the Bcc (Baldwin *et al.*, 2005) and PCR primers were subsequently improved to reliably amplify the target loci from both Bcc and non-Bcc *Burkholderia* species and to enable the use of a single primer set for both amplification and sequencing (Spilker *et al.*, 2009). Vanlaere *et al.* (2009) compared the average concatenated allele sequence divergence within and between established Bcc species as delineated by DDH studies and demonstrated that a 3 % concatenated allele sequence divergence level can be used as a threshold value for species delineation within the Bcc, thus replacing the need to perform DDH experiments in this complex.

The ability to carry out both species identification and strain differentiation in a single approach represented a major advantage over previous methods (Baldwin *et al.*, 2005). An additional advantage of these multilocus sequence based methods over previous typing and identification methods was its reproducibility and portability (Jolley & Maiden, 2010). The public database enabled researchers worldwide to analyze and deposit data and as a result the Bcc PubMLST database currently (March 2016) holds information on 1,964 isolates and their 987 associated STs. Fig. 2.2 shows a phylogenetic analysis of the concatenated sequences of the seven loci of all STs in the database and reveals the presence of at least 12 putative novel Bcc species awaiting formal classification.

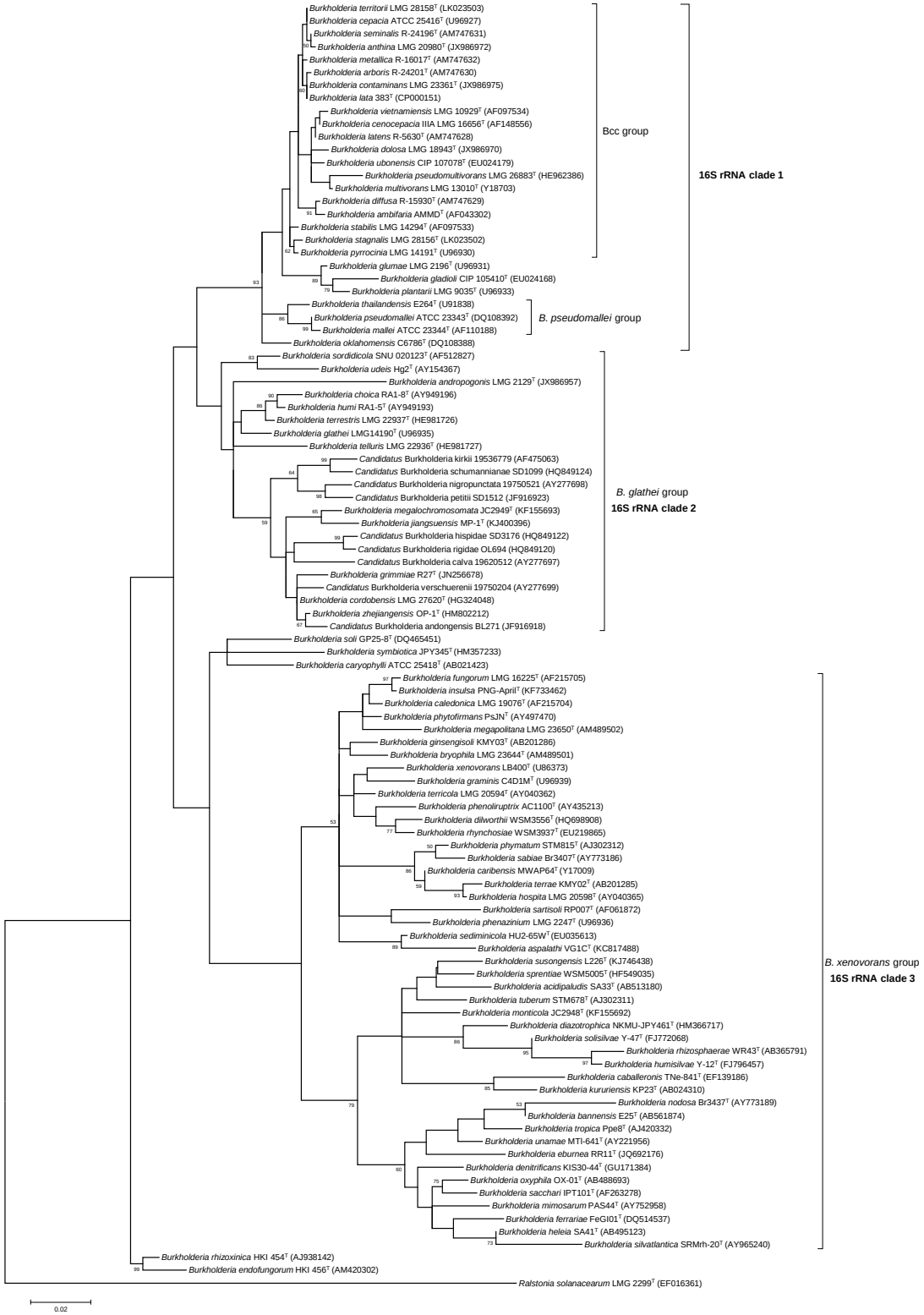

Burkholderia territorii LMG 28158T (LK023503)
Burkholderia cepacia ATCC 25416T (U96927)
Burkholderia seminalis R-24196T (AM747631)
Burkholderia anthina LMG 20980T (JX986972)
Burkholderia metallica R-16017T (AM747632)
Burkholderia arboris R-24201T (AM747630)
Burkholderia contaminans LMG 23361T (JX986975)
Burkholderia lata 383T (CP000151)
Burkholderia vietnamiensis LMG 10929T (AF097534)
Burkholderia cenocepacia IIIA LMG 16656T (AF148556)
Burkholderia latens R-5630T (AM747628)
Burkholderia dolosa LMG 18943T (JX986970)
Burkholderia ubonensis CIP 107078T (EU024179)
Burkholderia pseudomultivorans LMG 26883T (HE962386)
Burkholderia multivorans LMG 13010T (Y18703)
Burkholderia diffusa R-15930T (AM747629)
Burkholderia ambifaria AMMDT (AF043302)
Burkholderia stabilis LMG 14294T (AF097533)
Burkholderia stagnalis LMG 28156T (LK023502)
Burkholderia pyrrocinia LMG 14191T (U96930)
Bcc group
Burkholderia glumae LMG 2196T (U96931)
Burkholderia gladioli CIP 105410T (EU024168)
Burkholderia plantarii LMG 9035T (U96933)
Burkholderia thailandensis E264T (U91838)
Burkholderia pseudomallei ATCC 23343T (DQ108392)
Burkholderia mallei ATCC 23344T (AF110188)
B. pseudomallei group
16S rRNA clade 1
Burkholderia oklahomensis C6786T (DQ108388)
Burkholderia sordidicola SNU 020123T (AF512827)
Burkholderia udeis Hg2T (AY154367)
Burkholderia andropogonis LMG 2129T (JX986957)
Burkholderia choica RA1-8T (AY949196)
Burkholderia humi RA1-5T (AY949193)
Burkholderia terrestris LMG 22937T (HE981726)
Burkholderia glathei LMG14190T (U96935)
Burkholderia telluris LMG 22936T (HE981727)
Candidatus Burkholderia kirkii 19536779 (AF475063)
Candidatus Burkholderia schumannianae SD1099 (HQ849124)
Candidatus Burkholderia nigropunctata 19750521 (AY277698)
Candidatus Burkholderia petitii SD1512 (JF916923)
Burkholderia megalochromosomata JC2949T (KF155693)
Burkholderia jiangsuensis MP-1T (KJ400396)
Candidatus Burkholderia hispidae SD3176 (HQ849122)
Candidatus Burkholderia rigidae OL694 (HQ849120)
Candidatus Burkholderia calva 19620512 (AY277697)
Burkholderia grimmiae R27T (JN256678)
Candidatus Burkholderia verschuerenii 19750204 (AY277699)
Burkholderia cordobensis LMG 27620T (HG324048)
Burkholderia zhejiangensis OP-1T (HM802212)
Candidatus Burkholderia andongensis BL271 (JF916918)
B. glathei group
16S rRNA clade 2
Burkholderia soli GP25-8T (DQ465451)
Burkholderia symbiotica JPY345T (HM357233)
Burkholderia caryophylli ATCC 25418T (AB021423)
Burkholderia fungorum LMG 16225T (AF215705)
Burkholderia insulsa PNG-AprilT (KF733462)
Burkholderia caledonica LMG 19076T (AF215704)
Burkholderia phytofirmans PsJNT (AY497470)
Burkholderia megapolitana LMG 23650T (AM489502)
Burkholderia ginsengisoli KMY03T (AB201286)
Burkholderia bryophila LMG 23644T (AM489501)
Burkholderia xenovorans LB400T (U86373)
Burkholderia graminis C4D1MT (U96939)
Burkholderia terricola LMG 20594T (AY040362)
Burkholderia phenoliruptrix AC1100T (AY435213)
Burkholderia dilworthii WSM3556T (HQ698908)
Burkholderia rhynchosiae WSM3937T (EU219865)
Burkholderia phymatum STM815T (AJ302312)
Burkholderia sabiae Br3407T (AY773186)
Burkholderia caribensis MWAP64T (Y17009)
Burkholderia terrae KMY02T (AB201285)
Burkholderia hospita LMG 20598T (AY040365)
Burkholderia sartisoli RP007T (AF061872)
Burkholderia phenazinium LMG 2247T (U96936)
Burkholderia sediminicola HU2-65WT (EU035613)
Burkholderia aspalathi VG1CT (KC817488)
Burkholderia susongensis L226T (KJ746438)
Burkholderia sprentiae WSM5005T (HF549035)
Burkholderia acidipaludis SA33T (AB513180)
Burkholderia tuberum STM678T (AJ302311)
Burkholderia monticola JC2948T (KF155692)
Burkholderia diazotrophica NKMU-JPY461T (HM366717)
Burkholderia solisilvae Y-47T (FJ772068)
Burkholderia rhizosphaerae WR43T (AB365791)
Burkholderia humisilvae Y-12T (FJ796457)
Burkholderia caballeronis TNe-841T (EF139186)
Burkholderia kururiensis KP23T (AB024310)
Burkholderia nodosa Br3437T (AY773189)
Burkholderia bannensis E25T (AB561874)
Burkholderia tropica Ppe8T (AJ420332)
Burkholderia unamae MTI-641T (AY221956)
Burkholderia eburnea RR11T (JQ692176)
Burkholderia denitrificans KIS30-44T (GU171384)
Burkholderia oxyphila OX-01T (AB488693)
Burkholderia sacchari IPT101T (AF263278)
Burkholderia mimosarum PAS44T (AY752958)
Burkholderia ferrariae FeGI01T (DQ514537)
Burkholderia heleia SA41T (AB495123)
Burkholderia silvatlantica SRMrh-20T (AY965240)
B. xenovorans group
16S rRNA clade 3
Burkholderia rhizoxinica HKI 454T (AJ938142)
Burkholderia endofungorum HKI 456T (AM420302)
Ralstonia solanacearum LMG 2299T (EF016361)
0.02

Figure 2.1: Phylogenetic tree based on partial 16S rRNA gene sequences of *Burkholderia* species. Sequences (1125-1610 bp) were aligned against the SILVA SSU reference database using SINA v1.2.11 www.arb-silva.de/aligner/ (Pruesse *et al.*, 2012). Phylogenetic analysis was conducted using MEGA6 (Tamura *et al.*, 2013). All positions containing gaps and missing data were eliminated, resulting in a total of 1087 positions in the final dataset. The optimal tree (highest log likelihood) was constructed using the maximum likelihood method and Tamura-Nei model (Tamura & Nei, 1993). A discrete gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.3498)) and allowed for some sites to be evolutionarily invariable ([+I], 68.6154 % sites). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches if greater than 50 %. The sequence of *Ralstonia solanacearum* LMG 2299^{T} was used as outgroup. The scale bar indicates the number of substitutions per site. (Depoorter *et al.*, 2016)

2.1.3.2 Genomic taxonomy

In this genomics era, the number of finished or draft bacterial genomes has risen exponentially and phylogenetic analysis has entered the new era of phylogenomics (Eisen & Fraser, 2003; Delsuc *et al.*, 2005). Several tools are available for studying genomic taxonomy and these can be subdivided based on how phylogeny is inferred and what part of the genome is being analyzed (Table 2.1). Methods for inferring evolutionary relationships can generally be classified as either distance-based or character-based (Fig. 2.3). Distance-based methods convert aligned sequences into a distance matrix employing a model of evolution and subsequently use this distance matrix to infer a phylogenetic tree (Hall, 2011; Sleator, 2013). Although effective, these methods have the disadvantage of providing only one tree instead of a consensus tree and changing the order in which the constituent sequences are entered into the analysis may result in different trees (Sleator, 2013). On the other hand, character-based methods infer the most probable tree(s) based on the characters at each position in a multiple sequence alignment (Hall, 2011; Sleator, 2013).

The character-based methods in genomic taxonomy infer phylogenetic trees based on the multiple sequence alignment of conserved protein-coding genes in all genomes under study and represent an extension of the MLSA principle (Table 2.1). This extended MLSA approach has the advantage that it only takes into account the coding part of the genome and is therefore not influenced by non-coding sequences or pseudogenes which might have a different evolutionary history than the rest of the genome. Important disadvantages are the need for correctly annotated genomes and the fact that the set of single-copy orthologous genes may become very small when studying more distantly related organisms or organisms that underwent reductive genome evolution (Vandamme & Dawyndt, 2011). Examples of this approach are the core gene identity (Vanlaere *et al.*, 2009), supermatrix (Ciccarelli *et al.*, 2006)

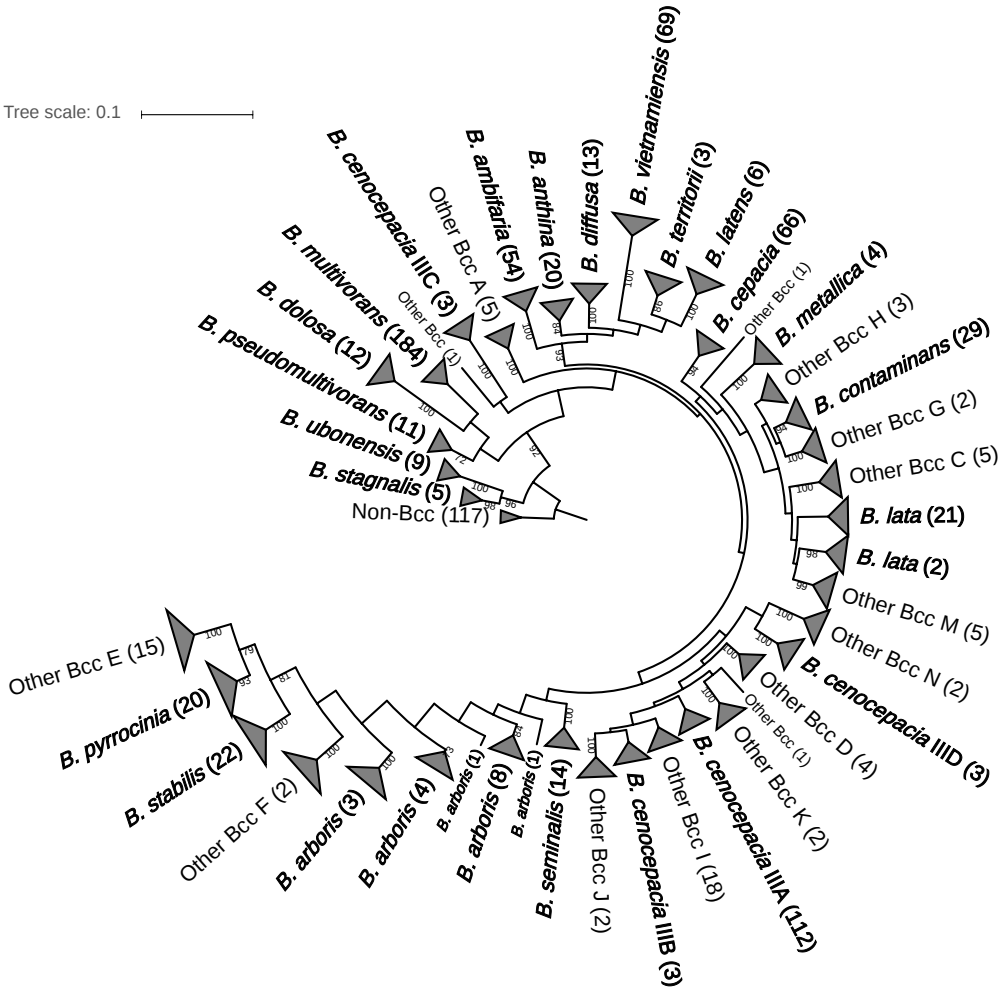


Figure 2.2: Phylogenetic tree based on the concatenated sequences of seven housekeeping gene fragments [*atpD* (443 bp), *gltB* (397-400 bp), *gyrB* (454 bp), *recA* (393 bp), *lepA* (397 bp), *phaC* (385 bp) and *trpB* (301 bp)] of all STs in the Bcc PubMLST database, showing the 20 established Bcc species (in bold type) and putative novel Bcc species (labeled "Other Bcc"). The number of STs per species is given in between brackets. Concatenated allele sequences of all STs (n=987, March 2016) were exported in frame from the Bcc PubMLST database (http://pubmlst.org/bcc) (Jolley & Maiden, 2010). Sequences (2757-2760 bp) were aligned based on their amino acid sequences using Muscle (Edgar, 2004) in MEGA6 (Tamura *et al.*, 2013). The phylogenetic tree was constructed based on the nucleic acid alignment in RAxML version 7.4.2 (Stamatakis, 2014). Rapid bootstrapping and the maximum likelihood search were performed using the general time reversible model with CAT approximation (GTRCAT) and the best scoring maximum likelihood tree (optimized under GTRGAMMA) was annotated in iTOL (Letunic & Bork, 2011). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches if greater than 70 %.

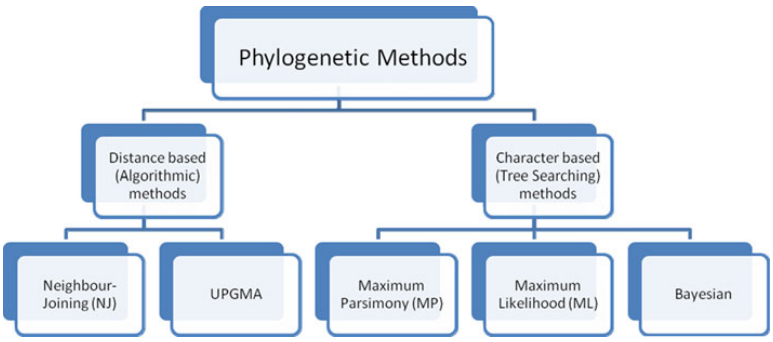


Figure 2.3: Schematic overview of the major analytical approaches to phylogenetic tree building. (Sleator, 2011)

and supertree (Daubin *et al.*, 2001) methods. The supermatrix method infers phylogeny based on the concatenated alignments of all orthologous genes in the genome dataset, while the supertree approach infers phylogeny for all loci separately and combines the trees afterwards (Klenk & Goker, 2010). The core gene identity method is similar to the supermatrix approach by the fact that it also restricts its calculation to the conserved genes in all genomes, but the phylogenetic tree is in this case calculated on a randomly chosen set of positions in the amino acid alignment (Vanlaere *et al.*, 2009).

The ribosomal MLST (rMLST) scheme enables rapid species and strain identification from whole-genome sequences by analyzing the divergence of 53 universal ribosomal protein encoding genes (Jolley *et al.*, 2012). This approach was recently applied to all publicly available *Burkholderia* genomes and provided a robust and high-resolution phylogenetic tree of the genus *Burkholderia* (Fig. 2.4) (Depoorter *et al.*, 2016). The same main clades as in the 16S rRNA phylogeny (Fig. 2.1) were revealed, except that 16S rRNA clade 2 clustered among species belonging to 16S rRNA clade 3 (Section 2.1.2). In the rMLST analysis too, *B. rhizoxinica* occupied a very distinct position.


Burkholderia cenocepacia PT15
Burkholderia cenocepacia S2AES
Burkholderia cenocepacia PC184
Burkholderia cenocepacia MC0-3
Burkholderia cenocepacia AU1054
Burkholderia cenocepacia HI2424
Burkholderia cenocepacia D2AES
Burkholderia cenocepacia J2315T
Burkholderia cenocepacia BC7
Burkholderia cenocepacia K56-2Valvano
Burkholderia cenocepacia H111
Burkholderia cepacia ATCC 25416T
Burkholderia cepacia AU4i
Burkholderia lata 383T
Burkholderia sp. RB-39
Burkholderia ambifaria AMMDT
Burkholderia ambifaria MC40-6
Burkholderia ambifaria IOP40-10
Burkholderia ambifaria MEX-5
Burkholderia vietnamiensis G4
Burkholderia vietnamiensis KJ006
Burkholderia stagnalis CH-67
Burkholderia ubonensis Bu
Burkholderia multivorans CF2
Burkholderia multivorans ATCC 17616
Burkholderia multivorans DWS 42B-1
Burkholderia multivorans CGD1
Burkholderia multivorans CGD2
Burkholderia multivorans LMG 13010T
Burkholderia dolosa AU0158
Burkholderia dolosa PC543
Burkholderia pseudomultivorans TJI49
Bcc group
Burkholderia mallei NCTC 10229
Burkholderia mallei NCTC 10247
Burkholderia mallei ATCC 23344T
Burkholderia mallei SAVP1
Burkholderia pseudomallei K96243
Burkholderia pseudomallei 1710b
Burkholderia pseudomallei 1106a
Burkholderia pseudomallei 668
Burkholderia thailandensis Bt4
Burkholderia thailandensis E264T
Burkholderia thailandensis TXDOH
Burkholderia thailandensis E555
"Burkholderia humptydooensis" MSMB43
"Burkholderia humptydooensis" MSMB121
Burkholderia oklahomensis C6786T
Burkholderia oklahomensis EO147
B. pseudomallei group
Burkholderia gladioli BSR3
Burkholderia gladioli 3848s-5
Burkholderia gladioli ERR353354
Burkholderia gladioli CIP 105410T
Burkholderia gladioli UCD-UG_CHAPALOTE
Burkholderia glumae BGR1
Burkholderia glumae LMG 2196T
Burkholderia glumae 336gr-1
Burkholderia glumae 3252-8
Burkholderia glumae AU6208
Burkholderia phenoliruptrix AC1100T
Burkholderia phenoliruptrix JPY366
Burkholderia phenoliruptrix CCGE1001
Burkholderia phenoliruptrix BR3459a
Burkholderia sp. CCGE1003
Burkholderia sp. WSM2232
Burkholderia sp. WSM2230
Burkholderia graminis URHA0054
Burkholderia bryophila 376MFSha3.1
Burkholderia caledonica LMG 19076T
Burkholderia dilworthii WSM3556T
Burkholderia xenovorans LB400T
Burkholderia sp. Ch1-1
Burkholderia phytofirmans PsJNT
Burkholderia fungorum LMG 16225T
Burkholderia sp. CCGE1002
Burkholderia sp. JPY251
Burkholderia tuberum WSM4176
Burkholderia sprentiae WSM5005T
Burkholderia sp. H160
Burkholderia terrae BS001
Burkholderia terrae BT03
Burkholderia terrae NBRC 100964T
Burkholderia caribensis MBA4
Burkholderia phymatum STM815T
Burkholderia sp. UYPR1.413
B. xenovorans group
Burkholderia cordobensis YI23
Burkholderia zhejiangensis SJ98
Candidatus Burkholderia kirkii UZHbot1
Burkholderia jiangsuensis MP-1T
Burkholderia sp. RPE64
Burkholderia grimmiae R27T
Burkholderia glathei LMG 14190T
B. glathei group
Burkholderia mimosarum LMG 23256
Burkholderia mimosarum NBRC 106338T
Burkholderia mimosarum STM 3621
Burkholderia nodosa DSM 21604T
Burkholderia oxyphila NBRC 105797T
Burkholderia bannensis NBRC 103871T
Burkholderia ferrariae NBRC 106233T
Burkholderia kururiensis M130
Burkholderia kururiensis JCM 10599T
Burkholderia kururiensis NBRC 107107
Burkholderia symbiotica JPY347T
Burkholderia rhizoxinica HKI 454T
Burkholderia andropogonis Ba3549
Burkholderia andropogonis ICMP 2807T
Ralstonia solanacearum PSI07
15
76
69
55
78
10
11
75
48
5
70
0.04

Figure 2.4: *Burkholderia* phylogeny reconstructed from concatenated ribosomal protein gene sequences. Aligned concatenated gene sequences from defined rMLST loci were downloaded from the rMLST database (http://pubmlst.org/rmlst/) (Jolley *et al.*, 2012). Low confidence regions of the alignment were removed with Gblocks (Talavera & Castresana, 2007), resulting in a total of 18,490 positions in the final dataset. A phylogeny was reconstructed with FastTree (Price *et al.*, 2010) using the general time reversible model of nucleotide evolution, and the resulting tree was visualized with FigTree (http://tree.bio.ed.ac.uk/software/figtree). Type strains are indicated in bold type. Node confidence is shown if less than 80 %. The sequence of *Ralstonia solanacearum* PSI07 was used as outgroup. The scale bar indicates the number of substitutions per site. (Depoorter *et al.*, 2016)

A first distance-based approach estimates intergenomic distances or similarities based on the divergence of the conserved DNA or protein-coding genes (Table 2.1). Examples of this approach are average nucleotide identity (ANI) (Konstantinidis & Tiedje, 2005a; Varghese *et al.*, 2015) and average amino acid identity (AAI) (Konstantinidis & Tiedje, 2005b). The ANI and AAI methods estimate the genetic relatedness of a pair of genomes (expressed as a percentage similarity value) based on the average nucleotide and amino acid identity, respectively, of all the conserved genes between the two genomes and this may induce a bias in pairwise similarities when computed on draft genomes (Vandamme & Dawyndt, 2011). Goris *et al.* (2007) introduced a variation on the original ANI method (ANIg) (Konstantinidis & Tiedje, 2005a) in which the nucleotide identity is calculated as the mean identity of all BLAST matches of 1020 nt long DNA fragments in a pair of genomes (ANIb). This artificial fragmentation of genomic sequences was implemented to mimic the genomic DNA fragmentation in DDH experiments, but was later implemented as such in the JSpecies software (Richter & Rossello-Mora, 2009). Another variation on the ANI method that was implemented in the JSpecies software makes use of the MUMmer software (ANIm) instead of BLAST to find the conserved DNA in a pair of genomes and does not require slicing of the genome sequence (Richter & Rossello-Mora, 2009).

A second distance-based approach estimates intergenomic distances or similarities based on the proportion of DNA or orthologous genes that is shared among genomes (Table 2.1). Examples of this approach are the percentage of conserved DNA (pcDNA) (Goris *et al.*, 2007), the maximal unique matches index (MUMi) (Deloger *et al.*, 2009) and the alignment fraction (AF) (Varghese *et al.*, 2015). While both the pcDNA and MUMi approaches are based on finding stretches of nucleotides that are conserved in the genomic DNA under comparison, they differ by the fact that pcDNA calculations are based on inexact BLAST matches, while MUMi is based on exact MUMmer matches between two genomes and was developed to be most sensitive at the intraspecies level (Deloger *et al.*, 2009). Both the pcDNA and MUMi approaches do not require any annotation of the genomes but need (nearly) complete

genome sequences because large stretches of missing DNA in draft genome assemblies can dramatically influence the result (Vandamme & Dawyndt, 2011). On the other hand, the recently introduced AF method considers the fraction of orthologous genes instead of DNA and is thus dependent of annotation (Varghese *et al.*, 2015).
The Genome BLAST Distance Phylogeny (GBDP) approach was introduced as an *in silico* replacement for DDH and calculates pairwise intergenomic distances based on the results of a genome-wide homology search (Henz *et al.*, 2005). Several studies were performed in which different distance formulas, algorithms and statistical models were compared to find the optimal method for *in silico* DDH (Auch *et al.*, 2010; Meier-Kolthoff *et al.*, 2013). As a result, the GBDP method provides a quick and reliable alternative to the wet-lab DDH technique and its improved DDH prediction capability produces classifications which correlate better with the traditional DDH values than do any of the ANI implementations (Meier-Kolthoff *et al.*, 2013). Besides its good mimicking of wet-lab DDH values, this method has the advantages of being independent from genome annotation and being immune against problems caused by incompletely sequenced or low-quality draft genomes. Finally, the GBDP method provides confidence intervals and support values for groupings in the resulting phylogenetic trees (Meier-Kolthoff *et al.*, 2013). As two recent examples, the GBDP method was used to successfully clarify the taxonomic affiliations of the *Bacillus cereus* group (Liu *et al.*, 2014) and to propose a novel method to delineate subspecies, exemplified by the *Escherichia coli* group (Meier-Kolthoff *et al.*, 2014b).
Genomic signatures reflecting di-, tri- or tetranucleotide relative abundances circumvent the need of pairwise alignment of genomic DNA and represent a fast, alignment-free alternative to the approaches described above (Burge *et al.*, 1992; Karlin *et al.*, 1998; Richter & Rossello-Mora, 2009). The dinucleotide relative abundance was shown to be constant within a genome (Karlin *et al.*, 1997; Karlin, 1998) and correlated well with 16S rRNA similarity for closely related species (Coenye & Vandamme, 2004).
For each of these genomic taxonomy tools, cut-offs for species delineation have been proposed that correspond to the recommended cut-off of 70 % DDH (Table 2.1).

2.1.3.3 Routine identification of Bcc bacteria by MALDI-TOF MS

Biochemical tests alone are insufficient to reliably distinguish between Bcc species and commonly cannot differentiate them from members of related genera such as *Ralstonia*, *Pandoraea*, *Achromobacter* or *Cupriavidus* (Henry *et al.*, 2001; Vandamme & Dawyndt, 2011). Therefore, it is important to use both phenotypic and molecular methods for the identification of Bcc bacteria. Now that bioinformatics tools catch up with the ever-advancing

Table 2.1: Overview genomic taxonomy tools

Tool	Principle	Cut-off for species delineation	References
Character-based methods			
Extended MLSA and rMLST	Similarity of conserved protein-coding genes	NA	(Daubin *et al.*, 2001; Ciccarelli *et al.*, 2006; Vanlaere *et al.*, 2009; Jolley *et al.*, 2012)
Distance-based methods			
ANIg	Nucleotide acid similarity of conserved protein-coding genes	94 %, 96.5 %	(Konstantinidis & Tiedje, 2005a; Varghese *et al.*, 2015)
AAI	Amino acid similarity of conserved protein-coding genes	95-96 %	(Konstantinidis & Tiedje, 2005b)
ANIb	Similarity of conserved genomic DNA	95-96 %	(Goris *et al.*, 2007; Richter & Rossello-Mora, 2009)
ANIm	Similarity of conserved genomic DNA	95-96 %	(Richter & Rossello-Mora, 2009)
GBDP	Proportion and similarity of conserved genomic DNA or protein-coding genes	70 % dDDH	(Meier-Kolthoff *et al.*, 2013)
pcDNA	Proportion of conserved genomic DNA	69 %	(Goris *et al.*, 2007)
MUMi	Proportion of conserved genomic DNA	0.33	(Deloger *et al.*, 2009)
AF	Proportion of conserved protein-coding genes	0.6	(Varghese *et al.*, 2015)
TETRA	Genomic nucleotide signature of genomic DNA	99 %	(Richter & Rossello-Mora, 2009)

high-throughput sequencing technologies, taxonomy and identification based on whole-genome sequences are likely to become an everyday practice (Varghese *et al.*, 2015; Whitman, 2015). However, fast, accurate and easy-to-implement tools are needed for the routine identification of large numbers of clinical isolates.

During the last decade, Matrix-Assisted Laser Desorption/Ionization Time-Of-Flight mass spectrometry (MALDI-TOF MS) has become an important tool for routine identification of clinical isolates (Cherkaoui *et al.*, 2010; Van Veen *et al.*, 2010; Carbonnelle *et al.*, 2011). In this technique, intact cells or whole-cell protein extracts are subjected to MALDI-TOF MS which separates ionized peptides and proteins based on their molecular mass and ionization status. The resulting mass spectrum represents a fingerprint that can be compared to reference databases to identify the isolate. Although protein expression is dependent on the growth conditions and this technique requires some level of standardization (Clark *et al.*, 2013), species level identification is unaffected by varying growth media (Wieme *et al.*, 2014a). Several studies reported on the application of MALDI-TOF MS for identification of

GNNF bacteria from CF patients (Degand *et al.*, 2008; Desai *et al.*, 2012; Fernandez-Olmos *et al.*, 2012; Alby *et al.*, 2013; Lambiase *et al.*, 2013). While its overall performance for the identification of GNNF bacteria is generally good, identification of Bcc species requires an extended reference database (Fernandez-Olmos *et al.*, 2012; Alby *et al.*, 2013).

2.2 The good, the bad and the ugly: a tribute to adaptation

The genus *Burkholderia* currently (March 2016) comprises 90 validly named species (Euzeby, 1997) which have a versatile metabolism and occupy very diverse niches (Coenye & Vandamme, 2003). *Burkholderia* organisms have been isolated mainly as free-living organisms from water, soil and the rhizosphere of plants but these bacteria also developed different types of interactions with different hosts including plants, animals, fungi and amoebae (Marolda *et al.*, 1999; Van Borm *et al.*, 2002; Kikuchi *et al.*, 2011; Verstraete *et al.*, 2013; Stopnisek *et al.*, 2016).

2.2.1 Plant growth promotion

Several *Burkholderia* species are plant growth promoting rhizobacteria (PGPR). PGPR are rhizosphere soil bacteria that influence plant growth through a range of interactions. These interactions can be direct through the production of phytohormones or through the enhancement of nutrient availability, or indirect through biological control of plant pathogens (Lugtenberg & Kamilova, 2009). Several diazotrophic (nitrogen-fixing) *Burkholderia* species have been described, including *B. vietnamiensis*, *Burkholderia silvatlantica*, *Burkholderia diazotrophica*, *Burkholderia unamae*, *B. xenovorans* and *B. tropica* (Van *et al.*, 2000; Caballero-Mellado *et al.*, 2004; Perin *et al.*, 2006; Caballero-Mellado *et al.*, 2007; Sheu *et al.*, 2013). *Burkholderia phytofirmans*, *B. unamae* and *B. xenovorans* show 1-aminocyclopropane-1-carboxylic acid deaminase activity and thereby lower the ethylene levels in stressed or developing plants (Sessitsch *et al.*, 2005; Caballero-Mellado *et al.*, 2007). Plant hormone levels are also influenced by the production of indol acetic acid by *B. fungorum*, *Burkholderia graminis*, *Burkholderia kururiensis*, *B. phytofirmans*, *B. unamae* and *B. vietnamiensis* (Bevivino *et al.*, 1994; Suarez-Moreno *et al.*, 2012; Castanheira *et al.*, 2016). Finally, the production of organic acids and siderophores increases the availability of phosphate and iron for the plant and thereby also promotes plant growth (Caballero-Mellado *et al.*, 2007).
While many *Burkholderia* species live in close interactions with plants in the rhizosphere, some rhizosphere bacteria also colonize roots, stems and leaves (Coenye & Vandamme, 2003;

Compant *et al.*, 2008). Many *Burkholderia* species including *B. cepacia*, *B. cenocepacia*, *B. gladioli*, *B. phytofirmans*, *Burkholderia pyrrocinia*, *B. silvatlantica*, *B. tropica*, *B. unamae* and *B. vietnamiensis* have been isolated from plant tissue specimens (Compant *et al.*, 2008). *Burkholderia caledonica* was detected between mesophyll cells in *Fadogia* and *Vangueria* species (family *Rubiaceae*) known to cause gousiekte, a fatal disease of South African ruminants (Verstraete *et al.*, 2011). Recently, *B. multivorans* WS-FJ9 was shown to colonize the rhizosphere and root tissues of poplar seedlings and promoted their growth (Li *et al.*, 2013). Additionally, several *Burkholderia* species including *Burkholderia caribensis*, *Burkholderia mimosarum*, *Burkholderia nodosa*, *Burkholderia phymatum* and *Burkholderia tuberum* are also known for nodulation of plant species of the *Fabaceae* family (Vandamme *et al.*, 2002a; Chen *et al.*, 2005, 2006, 2007). In addition to these free-living plant-beneficial species, several *Candidatus Burkholderia* species have been described. They are obligate endosymbionts in species of the *Rubiaceae* (genera *Pavetta*, *Psychotria* and *Sericanthe*) and *Primulaceae* families (Van Oevelen *et al.*, 2002, 2004; Lemaire *et al.*, 2011a, 2012; Verstraete *et al.*, 2011, 2013). These endosymbionts occur in leaf nodules, are vertically transmitted and represent an obligatory symbiosis which was estimated to originate millions of years ago (Lemaire *et al.*, 2011b).

Besides directly stimulating plant growth through the mechanisms described above, *Burkholderia* organisms can also produce antibacterial and antifungal metabolites which prevent the proliferation of other rhizobacteria and fungi (Compant *et al.*, 2008). By suppressing other organisms in the rhizosphere environment, they do not only compete with other soil bacteria, but also protect plants from soilborne pathogens. Examples of these metabolites are cepacin, quinolones, phenazine and pyrrolnitrin (Parke & Gurian-Sherman, 2001). The best known example is *Burkholderia ambifaria* AMMD^{T}, which protects pea plants from *Aphanomyces* root rot and *Phytium* damping-off disease (King & Parke, 1993). *B. pyrrocinia* produces at least two antifungal compounds including pyrrolnitrin and occidiofungin (Kwak & Shin, 2015; Wang *et al.*, 2016) and an endophytic *B. cenocepacia* strain showed *in planta* biocontrol of soilborne *Fusarium* wilt of banana plants (Ho *et al.*, 2015). Next to the production of antimicrobial compounds, the biocontrol properties of *Burkholderia* bacteria can also be attributed to their capacity to compete for limited resources (e.g. iron, vitamins, amino acids), to utilize root exudates and to activate an induced systemic resistance in plants (Compant *et al.*, 2008; Vial *et al.*, 2011).

2.2.2 Plant pathogens

The best known plant-pathogenic species within the genus *Burkholderia* are *B. andropogonis*, *B. caryophylli*, *B. cepacia*, *B. gladioli*, *B. glumae* and *B. plantarii* (Compant *et al.*, 2008). *B. andropogonis* is an important cause of stripe disease of sorghum and leaf spot of velvet bean and causes considerable loss of carnation (Coenye *et al.*, 2001a). *B. cepacia* causes soft rot of onion leaves and bulbs (Burkholder, 1950). *B. caryophylli* causes wilt in various plant species (Furuya *et al.*, 2000). Finally, *B. gladioli*, *B. glumae* and *B. plantarii* form a phylogenetically coherent group (Fig. 2.1 and 2.4) of notorious rice pathogens (Seo *et al.*, 2015). In addition, *B. gladioli* also induces soft rot in onions (Lee *et al.*, 2005) and *B. glumae* causes wilting symptoms in several field crops (Jeong *et al.*, 2003).

2.2.3 Human and animal pathogens

B. mallei and *B. pseudomallei* are primary pathogens causing glanders in horses, donkeys and mules and melioidosis in humans and animals, respectively. The Bcc contains opportunistic pathogens that pose a threat to patients with an underlying immunocompromising illness such as CF (Chapter 1) and chronic granulomatous disease. Also the plant pathogen *B. gladioli* has been isolated from human clinical samples and is a well-known CF pathogen. *B. fungorum* is a most striking exception among the 16S rRNA clade 3 species (Fig. 2.1) since it has not only been isolated from the environment but also from a wide range of human and animal samples such as cerebrospinal fluid, vaginal secretions, blood, respiratory secretions of CF patients sputum and the nose of mice (Coenye *et al.*, 2001b, 2002a; Gerrits *et al.*, 2005). Other exceptions are *B. tropica* which was isolated from a neonatal patient who developed septicemia (Deris *et al.*, 2010), *B. xenovorans* which was isolated from human blood (Goris *et al.*, 2004) and *B. ginsengisoli* which was isolated from a young man with Crohn's disease who developed bacteraemia (Marks *et al.*, 2016).

2.2.4 Fungal interactions

Interactions between *Burkholderia* bacteria and fungi exist in many forms (Scherlach *et al.*, 2013). *B. rhizoxinica* and *B. endofungorum* represent bacterial endosymbionts of the plant-pathogenic fungus *Rhizopus microsporus* (Partida-Martinez *et al.*, 2007). The phytotoxin used by the plant-pathogenic fungi to cause rice seedling blight is produced by the bacterial endosymbiotic cells (Partida-Martinez & Hertweck, 2005). *B. fungorum* and *Burkholderia sordidicola* have been isolated from the white-rot fungi *Phanerochaete chrysosporium* and *Phanerochaete sordida*, respectively (Coenye *et al.*, 2001b; Lim *et al.*, 2003). Associations between *Burkholderia* and fungal cells have not only been described for plant-pathogenic

fungi, but also for plant-beneficial mycorrhizal fungi. *Burkholderia* species have been found in close association with both ectomycorrhizal (Izumi *et al.*, 2007; Uroz *et al.*, 2012) and arbuscular mycorrhizal (Andrade *et al.*, 1997) fungi. *Burkholderia* are often associated with fungi in soils (Warmink *et al.*, 2011; Nazir *et al.*, 2012b; Uroz *et al.*, 2012) and it was recently demonstrated that this co-occurrence of *Burkholderia* with a wide range of fungi might represent a survival strategy of *Burkholderia* organisms in acidic soils (Stopnisek *et al.*, 2014, 2016).

2.2.5 Insect interactions

Burkholderia bacteria belonging to the *B. glathei* clade have been described as gut endosymbionts of the bean bug *Riptortus pedestris* (Kikuchi *et al.*, 2011). The bacterial endosymbionts are not vertically transmitted but are acquired from soil by the nymphal insect (Kikuchi *et al.*, 2007). The insecticide resistance to fenitrothion in the pest insects is established by the endosymbiotic *Burkholderia* in the gut (Kikuchi *et al.*, 2012) and emerges as a consequence of repeated insecticide use (Tago *et al.*, 2015). Also in the chinch bug *Blissus insularis Burkholderia* bacteria were found in the gut (Boucias *et al.*, 2012; Xu *et al.*, 2016). Furthermore, a survey of *Bartonella* bacteria bed bugs in the United States detected DNA sequences of Bcc bacteria (Saenz *et al.*, 2013). Finally, also ants (Van Borm *et al.*, 2002; Santos *et al.*, 2004) and beetles (Lundgren *et al.*, 2007; Kim *et al.*, 2009) have been reported to live in complex interactions with *Burkholderia* bacteria. In contrast with these beneficial insect interactions, *Burkholderia* strain A396 (named "*Burkholderia rinojensis*") was isolated from soil in Japan and showed insecticidal and miticidal activities (Cordova-Kreylos *et al.*, 2013).

2.2.6 Amoebal interactions

Several Bcc strains survive in *Acanthamoeba* cells and grow on by-products released by the amoebae (Marolda *et al.*, 1999; Landers *et al.*, 2000). *Burkholderia* bacteria were internalized by phagocytosis and survived in acidic vacuoles instead of being digested (Lamothe *et al.*, 2004). Since amoebae are also found in freshwater environments they have been suggested as a natural reservoir for Bcc strains (Marolda *et al.*, 1999). More recently, a cultivation-independent study showed that *Burkholderia* was one of the genera found in free-living amoebae that were isolated from drinking water (Delafont *et al.*, 2013). This ability to survive intracellularly in eukaryotic cells was also demonstrated in human epithelial cells (Burns *et al.*, 1996) and macrophages (Saini *et al.*, 1999), suggesting that the same mechanisms that lead to survival and persistence in the natural environment may also lead to survival and

persistence as opportunistic pathogens in host cells (Valvano *et al.*, 2005; Adiba *et al.*, 2010). More recently, *Burkholderia* were found to colonize *Dictyostelium discoideum* amoebae and to induce the farming phenomenon in which amoebae stably associate with bacteria as a food source (DiSalvo *et al.*, 2015).

2.2.7 Bioremediation

The metabolic versatility of *Burkholderia* allows them to utilize recalcitrant pollutants and makes these bacteria useful for bioremediation applications (Coenye & Vandamme, 2003). *Burkholderia* bacteria are known to degrade antibiotics (Beckman & Lessie, 1979), herbicides (Jacobsen, 1997; Coenye *et al.*, 2004), insecticides (Hong *et al.*, 2007; Kikuchi *et al.*, 2012), crude oil constituents (Castorena *et al.*, 2006; Vanlaere *et al.*, 2008b) and solvents such as trichloroethylene (Mars *et al.*, 1996; Zhang *et al.*, 2000). *B. xenovorans* LB400 was isolated from a polychlorinated biphenyl contaminated landfill in New York (Bopp, 1986) and its genome encodes at least 27 aromatic pathways (Chain *et al.*, 2006). *Burkholderia phenoliruptrix* AC1100 is well-known for its ability to degrade a variety of xenobiotics including the herbicide 2,4,5-trichlorophenoxyacetic acid (Coenye *et al.*, 2004). As a last example, *Burkholderia* sp. strain NF100 was reported more than a decade ago as a degrader of the insecticide fenitrothion (Hayatsu *et al.*, 2000) and more strains with this bioremediation capacity have been reported recently (Hong *et al.*, 2007; Lim *et al.*, 2012; Tago *et al.*, 2015).

2.3 The genetic source code for versatility: *Burkholderia* genomes

Burkholderia genomes vary in size from 2.4 Mb (*Ca.* Burkholderia schumannianae UZHbot8) (Pinto-Carbo *et al.*, 2016) to 11.5 Mb (*Burkholderia terrae* BS001) (Nazir *et al.*, 2012a), are characterized by a high G+C content (60-68 %) and consist of multiple replicons (Lessie *et al.*, 1996; Winsor *et al.*, 2008; Ussery *et al.*, 2009). In general, large prokaryotic genomes are disproportionately enriched in regulation and secondary metabolism, while being depleted in protein translation, DNA replication, cell division and nucleotide metabolism. This may explain why bacteria with large genomes dominate in environments were resources are scarce but diverse, such as soil (Konstantinidis & Tiedje, 2004).
These large genomes encode a multitude of metabolic functions and virulence factors (Fig. 2.5). The first and largest chromosome of *Burkholderia* species harbors most of the genes essential for growth and basic metabolism while the other chromosome(s) encode mostly genes

involved in adaptation and secondary metabolism (Mahenthiralingam *et al.*, 2005; Ussery *et al.*, 2009). Cooper *et al.* (2010) demonstrated that genes were less conserved and evolved more rapidly if they were located on the secondary chromosome and that secondary chromosomes may serve as evolutionary test beds. The multireplicon structure of *Burkholderia* genomes thus allows for long-term segregation of genes by their expression rate and dispensability. Although the third chromosome of Bcc bacteria encodes at least one rRNA operon and was therefore originally considered a chromosome, this replicon is not essential for survival and should be considered a megaplasmid which is involved in virulence and stress response (Agnoli *et al.*, 2012). *Burkholderia* bacteria can also carry one or more plasmids which further foster the diversification and adaptation to variable environments such as soil (Heuer & Smalla, 2012).

Finally, *Burkholderia* genomes contain numerous genomic islands and insertion sequences, representing as much as 9 % of total DNA (Holden *et al.*, 2009). Insertion sequences or transposons are small mobile genetic elements (up to 2500 bp) that can jump from one location in the genome to another, thereby causing mutations or affecting gene expression. Genomic islands are larger regions that are identifiable by their deviating G+C content and are involved in the dissemination of variable genes, including antibiotic resistance and virulence genes as well as catabolic genes (Juhas *et al.*, 2009). Altogether, this genomic diversity and plasticity provides the genetic source code for the diversification and adaptation of *Burkholderia* bacteria to a wide range of ecological niches.

2.4 What's in the name *Burkholderia*?

Burkholderia bacteria interact in different ways with different hosts, making them ubiquitous and versatile bacteria. This versatility and adaptability is both a blessing and a curse. Without their ability to thrive in varying environments, they would never have emerged as useful strains for plant growth promotion, bioremediation and biocontrol. However, this same ability to thrive in varying environments most likely also contributes to their success as opportunistic pathogens. In fact, many properties important for rhizosphere colonization and survival in the natural environment (versatile metabolism, antibiotic resistance and production, quorum sensing, siderophore production and adherence to host cells) may also be considered traits that could contribute to establishing infection in immunocompromised patients (Parke & Gurian-Sherman, 2001).

Especially for the Bcc, their biotechnologically useful features are in marked contrast with the severe and often fatal infections they cause as opportunistic human pathogens (Mahenthiralingam *et al.*, 2008). Increased concerns about the health risks for immunocompromised

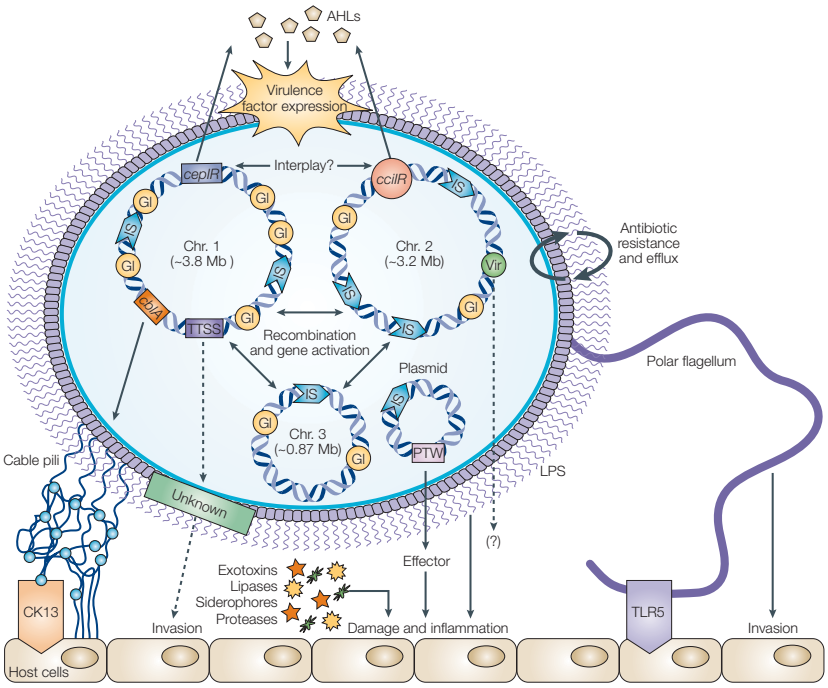


Figure 2.5: Genomic structure and major virulence factors of *B. cenocepacia* ET-12. Features of the genome include the three chromosomal replicons (Chr. 1, Chr. 2 and Chr. 3) and the plasmid, the presence of genomic islands on each of the three large replicons and the potential contribution of insertion sequences to genomic plasticity. The following virulence factors and their roles or interactions with host cells are also shown: cable pili and the 22-kDa adhesin (blue circles) that binds to the intermediate filament protein cytokeratin 13 (CK13), the quorum-sensing systems cepIR and cciIR which secrete and sense acyl-homoserine lactones (AHLs), the type III secretion system (TTSS); flagellin that binds to Toll-like receptor 5 (TLR5) to mediate the inflammatory response, multiple extracellular and surface-associated factors including lipopolysaccharide (LPS), exotoxins, lipases, siderophores and proteases, inherent antibiotic resistance of the cell envelope, and the type IV secretion systems (Vir and PTW). (Mahenthiralingam *et al.*, 2005)

patients by use of these bacteria in biopesticidal and bioremediation applications (Holmes *et al.*, 1998; LiPuma & Mahenthiralingam, 1999) led to a ban of biocontrol applications of Bcc strains by the U.S. Environmental Protection Agency in 2003 (Environmental Protection Agency, 2003; Mahenthiralingam *et al.*, 2008).

The phylogenetic diversity and conflicting biological properties of *Burkholderia* bacteria inspired several researchers to split the genus *Burkholderia* in order to separate the beneficial from the pathogenic strains (Gyaneshwar *et al.*, 2011; Suarez-Moreno *et al.*, 2012; Estrada-de los Santos *et al.*, 2013; Angus *et al.*, 2014; Zuleta *et al.*, 2014; Estrada-de los Santos *et al.*, 2015). Giving some *Burkholderia* species a less controversial name and thus taking away this negative connotation might pave the way for agricultural applications (Estrada-de los Santos *et al.*,

2015). However, there is no phylogenetic subdivision in this genus that distinguishes beneficial from pathogenic strains, as exemplified by the fact that *B. fungorum*, *B. ginsengisoli*, *B. tropica* and *B. xenovorans*, four species belonging to the so-called plant-associated beneficial and environmental group (Suarez-Moreno *et al.*, 2012), have been isolated from human clinical samples too (Coenye *et al.*, 2001b, 2002a; Goris *et al.*, 2004; Gerrits *et al.*, 2005; Deris *et al.*, 2010; Marks *et al.*, 2016). Yet, Sawana *et al.* (2014) recently reclassified all species not belonging to 16S rRNA clade 1 (Fig. 2.1) into a single novel genus, *Paraburkholderia* (Sawana *et al.*, 2014; Oren & Garrity, 2015). The rationale for splitting the genus *Burkholderia* was based on the identification of conserved sequence indels in whole-genome sequences of 45 *Burkholderia* strains representing only 25 formally named species and several unclassified strains. Species belonging to 16S rRNA clade 1 shared six conserved sequence indels. The remaining *Burkholderia* strains representing a subset of species belonging to 16S rRNA clades 2 and 3 shared only two conserved sequence indels. However, the phylogenetic diversity among the clade 2 and 3 species and *B. rhizoxinica* as revealed by 16S rRNA-based divergence and by differences in the distribution of 22 additional conserved sequence indels was ignored and the authors proposed to restrict the name *Burkholderia* to 16S rRNA clade 1 while reclassifying all other species into a single novel genus named *Paraburkholderia* (Sawana *et al.*, 2014). More recently, a similar rationale led to the reclassification of all species of 16S rRNA clade 2 (Fig. 2.1) into the novel genus *Caballeronia* (Dobritsa & Samadpour, 2016).

Part II

Aims and Outline

Aims and Outline

As outlined in Part I, Bcc bacteria are rare but significant pathogens in people with CF. Epidemiological surveys reveal that *B. multivorans* is the most prevalent Bcc CF pathogen in many countries, including Belgium, and the continued emergence of unique *B. multivorans* strains in CF patients suggests acquisition from non-human sources, such as the natural environment. Yet, environmental *B. multivorans* isolates are rare and its environmental niche is therefore considered unknown.
The first goal of the present thesis was to gain better insight into the epidemiology of *B. multivorans* by **(i)** examining its environmental niche and **(ii)** comparing the genomes of clinical and environmental *B. multivorans* isolates. **Chapter 3** describes how a large number of environmental water and soil samples were examined using a cultivation-independent *B. multivorans*-specific PCR assay (Section 3.1). Next, cultivation strategies were evaluated and optimized for the isolation of *B. multivorans* from PCR positive samples (Sections 3.2 and 3.3). Finally, section 3.4 presents a comparative genomics study of clinical and environmental *B. multivorans* isolates. **Chapter 6** provides a general discussion of these results and some future perspectives.

The CF lung can harbor a wide range of bacteria and accurate identification of CF pathogens at the species level is important to assess the clinical impact of these generally rare opportunistic pathogens. Furthermore, accurate typing of CF pathogens is necessary to identify outbreaks, gain insight into their epidemiology and improve infection control guidelines. Since January 2011 a National Reference Center (NRC) is charged with the surveillance of respiratory infections caused by GNNF bacilli in Belgian CF patients (https://nrchm.wiv-isp.be/nl/ref_centra_labo/burkholderia_cepacia_complex) and LM-UGent is responsible for the molecular identification and typing of these CF isolates.
The second goal of the present thesis was to contribute to the general knowledge of the prevalence and epidemiology of GNNF bacteria in CF by **(i)** applying MLST to study the epidemiology of Bcc bacteria, **(ii)** using MLSA for the classification of novel species within

the Bcc and **(iii)** performing polyphasic taxonomic studies for the classification of novel *Burkholderia* species. **Chapter 4** describes how increased knowledge of the diversity of Bcc bacteria and the application of MLSA led to the formal classification of three novel Bcc species. **Chapter 5** presents the application of both a traditional and modern polyphasic approach for studying *Burkholderia* taxonomy and the formal classification of 14 novel species in the *B. glathei* clade. **Chapter 7** provides a general discussion of these results, reflections on the use of MLST for surveillance of Bcc infections in CF and some future perspectives.

Part III

Experimental Work

3 | The environmental niche of *Burkholderia multivorans*: fitness school for a cystic fibrosis pathogen?

The continued emergence of unique *B. multivorans* strains in CF patients suggests acquisition from non-human sources such as the natural environment (Baldwin *et al.*, 2008). Yet, environmental *B. multivorans* isolates are rare and its true environmental niche is considered unknown. Section 3.1 describes the application of a cultivation-independent PCR assay to screen a large number of environmental water and soil samples and to assess the occurrence of *B. multivorans* in the natural environment. Sections 3.2 and 3.3 present the application of multiple cultivation strategies to *B. multivorans* PCR positive water and soil samples. Finally, section 3.4 provides the results of a comparative genomics study that examined to which extent *B. multivorans* isolates with the same multilocus sequence type but from different origins (CF vs environment) differed in genetic potential.

3.1 PCR detection of *Burkholderia multivorans* in water and soil samples

Redrafted from: Charlotte Peeters, Stijn Daenekindt and Peter Vandamme (2016). PCR detection of *Burkholderia multivorans* in water and soil samples. BMC Microbiology, 16, 1-7.

Author contributions: CP and PV conceived the study and wrote the manuscript. CP performed the experiments. CP and SD performed the statistical data analysis. SD proofread the manuscript.

3.1.1 Abstract

Background: Although semi-selective growth media have been developed for the isolation of Bcc bacteria from the environment, thus far *B. multivorans* has rarely been isolated from such samples. Because environmental *B. multivorans* isolates mainly originate from water samples, we hypothesized that water rather than soil is its most likely environmental niche. The aim of the present study was to assess the occurrence of *B. multivorans* in water samples from Flanders (Belgium) using a fast, culture-independent PCR assay.
Results: A nested PCR approach was used to achieve high sensitivity, and specificity was confirmed by sequencing the resulting amplicons. *B. multivorans* was detected in 11 % of the water samples (n=112) and 92 % of the soil samples (n=25) tested. The percentage of false positives was higher for water samples compared to soil samples, showing that the presently available *B. multivorans recA* primers lack specificity when applied to the analysis of water samples.
Conclusion: The results of the present study demonstrate that *B. multivorans* DNA is commonly present in soil samples and to a lesser extent in water samples in Flanders (Belgium).

3.1.2 Background

The Bcc represents a group of closely related (Vandamme & Peeters, 2014; De Smet *et al.*, 2015a) and extremely versatile bacteria that can be applied for a range of bioremediation, plant growth promotion and biocontrol purposes (Parke & Gurian-Sherman, 2001). Yet, these bacteria are simultaneously rare but important opportunistic pathogens in CF patients (De Boeck *et al.*, 2004; LiPuma, 2010). The most prevalent Bcc CF pathogens are *B. cenocepacia* and *B. multivorans*. The low number of outbreaks caused by *B. multivorans*

(Whiteford *et al.*, 1995; Segonds *et al.*, 1999; Biddick *et al.*, 2003), and the fact that *B. multivorans* isolates from CF patients commonly represent unique strains, suggest that there is only limited person-to-person transmission, and that strains are acquired from non-human sources such as the natural environment (Baldwin *et al.*, 2008). Although many studies have described the isolation of *B. cenocepacia* from rhizosphere samples (Mahenthiralingam *et al.*, 2008), only few reported on the occasional isolation of *B. multivorans* from environmental samples (Vermis *et al.*, 2003a; Ramette *et al.*, 2005; Vanlaere *et al.*, 2005; Fang *et al.*, 2011). Consequently, the true environmental niche of *B. multivorans* is considered unknown. The fact that the few environmental *B. multivorans* isolates that have been reported mainly originated from water samples (Vermis *et al.*, 2003a; Fang *et al.*, 2011) suggested that water is the most likely environmental niche of *B. multivorans*.

PCR-based diagnostic tests have been developed for Bcc species identification based on both the 16S rRNA and *recA* genes, resulting in two sets of species-specific primers for *B. multivorans* (LiPuma *et al.*, 1999; Mahenthiralingam *et al.*, 2000a). Miller *et al.* (2002) used the 16S rRNA-based PCR assays for the culture-independent detection of Bcc in soil environments. In the latter study, none of the samples that were Bcc PCR positive yielded Bcc isolates, suggesting that cultivation-dependent methods for the detection of Bcc bacteria may underestimate environmental populations.

In the present study, we used a cultivation-independent *recA*-based PCR assay to assess the presence of *B. multivorans* in water and soil samples in Flanders, Belgium. Our results show that *B. multivorans* DNA could be detected only occasionally in water samples but to a greater extent in soil samples, and that the available *B. multivorans recA* primers lack specificity, especially when applied to the analysis of water samples.

3.1.3 Materials and methods

3.1.3.1 Samples

Water (n=112) and soil (n=27) samples were taken from August to October 2013 in Flanders, Belgium (Supplementary Fig. 3.1). For the water samples, an autoclaved 1 L Duran bottle was opened and filled 10 cm below the water surface. Per sample, three times 150 ml was filtered using a Nalgene vacuum filter funnel and cellulose nitrate membrane filters with 0.45 µm pore size and 47 mm diameter (Thermo Scientific). For the soil samples, samples were taken 2 cm below the soil surface using a sterile spoon, and collected in sterile falcon tubes. Per sample, 1 g of soil was homogenized in 9 ml phosphate-buffered resuspension buffer (0.15 M NaCl, 10 mM EDTA, 0.1 M phosphate buffer, pH 8.0) using a Stomacher blender for 30 sec at 230 rpm. Three times 1 ml of soil suspension was transferred to an Eppendorf tube

and centrifuged for 5 minutes at 13,000 rpm (17,949 g) before removing the supernatant. Filters and soil pellets were stored at −20 °C until DNA extraction.
For each sample, sampling date, address, region (i.e. West-Vlaanderen [WV], Oost-Vlaanderen [OV], Limburg [L], Vlaams-Brabant [VB] or Antwerpen [A]), and class (i.e. swimming or recreational water [SRW], canal-river-stream [CRS] or other [O]) were recorded (Supplementary Table 3.1). Swimming and recreational waters were those under surveillance of the Flemish Environment Agency (www.kwaliteitzwemwater.be). For water samples, pH and temperature were measured on site. For soil samples, pH was measured after dissolving 10 g of soil in 50 ml distilled water and magnetic stirring for 10 min.

3.1.3.2 DNA extraction from environmental samples and quality assessment

Prior to DNA extraction, filters with biological material from water samples were cut into smaller pieces with sterilized scissors. Total DNA was extracted in triplicate from the filters and soil pellets (three per sample) following the protocol for Gram-negative bacteria of Pitcher *et al.* (1989). DNA pellets were dissolved by adding 50 or 100 µl TE buffer depending on the size of DNA pellet, and left to dissolve overnight at 4 °C. RNA was degraded by adding 2.5 or 5 µl RNase (2 mg/ml) for pellets dissolved in 50 or 100 µl TE, respectively, and incubating at 37 °C for 1 hour.
The quality and quantity of the extracted DNA were examined by measuring optical densities (OD) at 234 nm, 260 nm, 280 nm and 320 nm (Miller, 2001; Arbeli & Fuentes, 2007) with a SpectraMax Plus 384 spectrophotometer. DNA was considered of acceptable quality if the $OD_{260/280}$ ratio was higher than 1.7, the $OD_{234/260}$ ratio was smaller than 1 and the $OD_{320/260}$ ratio was smaller than 0.15. If both quality and quantity of the three DNA extraction replicates per sample were similar, these replicates were pooled. DNA fragmentation and RNA contamination was assessed by agarose (1 %) gel electrophoresis and EtBr staining.
To test for the presence of PCR inhibitors, DNA extracts were subjected to a 16S rRNA amplification PCR with universal primers ARI C/T (5'-CTG GCT CAG GAY GAA CGC TG-3') and pH (5'-AAG GAG GTG ATC CAG CCG CA-3'). The PCR mix contained 1x CorelLoad PCR buffer (Qiagen), 0.2 mM dNTP (Applied Biosystems), 0.5 U AmpliTaq (Applied Biosystems) 0.1 µM of both primers and 200 ng/µl BSA (Roche). For each sample, 2 µl DNA was added to 23 µl PCR master mix. *B. multivorans* R-20526 DNA and sterile MQ were used as positive and negative control, respectively. PCR was performed using a MJ Research PTC-100 thermal cycler. Initial denaturation for 5 min at 95 °C was followed by 3 cycles of 1 min at 95 °C, 2 min 15 sec at 55 °C and 1 min 15 sec at 72 °C, another 30 cycles of 35 sec at 95 °C, 1 min 15 sec at 55 °C and 1 min 15 sec at 72 °C, and a final elongation for

7 min at 72 °C. The presence of amplicons was verified via agarose (1 %) gel electrophoresis with SmartLadder (Eurogentec) as molecular size marker and EtBr staining.
If the $OD_{320/260}$ ratio was higher than 0.15 and/or no universal 16S rRNA amplicon could be obtained, an extra purification step using agarose plugs was performed to remove humic acids and other PCR inhibiting contaminants (Moreira, 1998). After purification using plugs, DNA extracts were diluted 5x in TE buffer. Only DNA extracts for which a universal 16S rRNA amplicon could be obtained, were subjected to the *B. multivorans* PCR assay.

3.1.3.3 Preliminary experiment

A preliminary experiment compared the specificity of the *B. multivorans*-specific primers that were based on the *recA* and 16S rRNA gene and were available from previous studies (LiPuma *et al.*, 1999; Mahenthiralingam *et al.*, 2000a). Sampling of two water samples, DNA extraction and quality assessment of the DNA extracts was performed as described above. For the nested *B. multivorans* 16S rRNA PCR assay, PCR products of the first, universal 16S rRNA PCR were used as template in a second PCR in which *B. multivorans*-specific primers BC-GII (5'-AGG CGG TCT GTT AAG ACA-3') and BC-R (5'-AGC ACT CCC GAA TCT CTT-3') were used (LiPuma *et al.*, 1999). The second PCR mix was identical to the first PCR mix, except for a lower BSA concentration (50 ng/µl). For each sample, 2 µl PCR product of the first PCR was added to 23 µl PCR master mix. The positive (*B. multivorans* R-20526) and negative (blank) control of the first PCR (5 µl) were also transferred as template into the second PCR. The thermal cycling program was identical to that of the first PCR. The presence of amplicons was verified via agarose (1 %) gel electrophoresis with SmartLadder (Eurogentec) as molecular size marker and EtBr staining. If an amplicon (445 bp) was visible in the second PCR, it was sequenced using the BC-GII and BC-R primers as described previously (Peeters *et al.*, 2013) to exclude false positive results. The *recA*-based *B. multivorans* PCR assay was performed as described below.

3.1.3.4 Nested *recA* PCR assay for *B. multivorans*

In a first PCR, Bcc-specific *recA* primers recA-01-F (5'-GAT AGC AAG AAG GGC TCC-3') and recA-02-R (5'-CTC TTC TTC GTC CAT CGC CTC-3') were used (Baldwin *et al.*, 2005). The PCR mix contained 1x CorelLoad PCR buffer (Qiagen), 0.25 mM dNTP (Applied Biosystems), 1 U Taq (Qiagen), 0.5 µM of both primers, 1x Q-solution (Qiagen) and 200 ng/µl BSA (Roche). For each sample, 2 µl DNA was added to 23 µl PCR master mix. *B. multivorans* R-20526 DNA and sterile MQ were used as positive and negative control, respectively. This PCR reaction was setup in duplicate for each sample to enable pooling and

to increase sensitivity. PCR was performed using a MJ Research PTC-100 thermal cycler. Initial denaturation for 2 min at 94 °C was followed by 35 cycles of 30 sec at 94 °C, 45 sec at 58 °C and 1 min at 72 °C, and a final elongation for 10 min at 72 °C. PCR product of the duplicate reactions for each sample were pooled using filter tips and used as template in the second PCR, in which *B. multivorans*-specific primers BCRBM1 (5'-CGG CGT CAA CGT GCC GGA T-3') and BCRBM2 (5'-TCC ATC GCC TCG GCT TCG T-3') were used (Mahenthiralingam *et al.*, 2000a). The second PCR mix was identical to the first PCR mix, except for a lower BSA concentration (50 ng/µl) and a higher primer concentration (1 µM). For each sample, 5 µl of pooled PCR product from the first PCR was added to 20 µl PCR master mix. The positive (*B. multivorans* R-20526) and negative (blank) control of the first PCR (5 µl) were also transferred as template into the second PCR. Thermal cycling was identical to the first PCR, except that the annealing temperature was 64 °C instead of 58 °C (Vermis *et al.*, 2002a). The presence of amplicons was verified via agarose (1 %) gel electrophoresis with SmartLadder (Eurogentec) as molecular size marker and EtBr staining. The nested *recA* PCR assay was performed twice for each sample. If the results for the two runs were not the same, the assay was performed a third time. If an amplicon (714 bp) was visible in the second *B. multivorans*-specific PCR, it was sequenced using the BCRBM1 and BCRBM2 primers as described previously (Peeters *et al.*, 2013) to exclude false positive results. Only if at least for one of the replicate runs an amplicon from the second PCR was sequenced that showed at least 97 % similarity to the *recA* sequence of *B. multivorans* ATCC 17616, the sample was considered a true positive for the detection of *B. multivorans*.

To determine the detection limit of this PCR assay, it was applied on serial dilutions of genomic DNA from *B. multivorans* R-20526, a strain for which whole-genome sequencing data is available (BioProject PRJNA234537). The mean weight for an AT and GC base pair is 615.3830 Da and 616.3711 Da, respectively (Dolezel *et al.*, 2003). Given the GC content of this genome of about 68 %, and ignoring the presence of modified nucleotides, the mean relative weight of one base pair of R-20526 is 616.0549 Da or 1.023×10^{-9} pg (1 Da = 1.660539×10^{-24} g). Given the genome size of 6.5 Mb, one genome of R-20526 contains 6.65×10^{-6} ng of DNA, or 1 ng of DNA contains 1.50×10^{5} genome equivalents. The undiluted genomic DNA stock of strain R-20526 contained 225 ng/µl DNA, as measured with the Promega QuantiFluor ONE dsDNA system, or 3.38×10^{7} genomic DNA equivalents per µl. The PCR assay was first applied on a 50x dilution series (in TE buffer) to find the approximate fading range. The highest 50x dilution for which an amplicon could be obtained was then used to make a twofold dilution series to find the limit of detection. The PCR assay was performed three times for each dilution and the detection limit was defined as the highest twofold dilution that tested positive in all three runs. Considering the *recA* gene is

a single-copy gene for *B. multivorans* R-20526, the detection limit was calculated from the measured DNA concentration and calculated DNA content.

3.1.3.5 Statistical data analysis

Statistical data analysis was performed using R version 3.1.2 in RStudio (version 0.98.1091), with the following packages: car, ggplot2, and MASS. Binomial logistic regression was used to test which variables (type, region, class, pH, temperature) were significant predictor variables for the outcome variable, i.e. *B. multivorans* detection (Bm). Type, class, region and Bm were coded as factors, and the level with the most cases was chosen as reference category. Temperature and pH were coded as numeric variables, and centered around the mean to reduce standard error (SE). Backward stepwise model selection was applied to select the best fitting models.

3.1.4 Results

3.1.4.1 Specificity of the 16S rRNA and *recA* PCR assays

Preliminary experiments were performed to compare the specificity of the *B. multivorans*-specific primers that were available from previous studies (De Boeck *et al.*, 2004; LiPuma, 2010). Therefore, PCR assays based on the *recA* and 16S rRNA genes were applied to DNA extracts of two water samples and the resulting amplicons were sequenced. The obtained sequences were analyzed using the NCBI blastn suite (blast.ncbi.nlm.gov) to evaluate specificity. For amplicons of the *recA* PCR assay, hits showed 98 % or more similarity with the *recA* sequence of *B. multivorans* ATCC 17616 (CP000868) (this degree of variability in *recA* gene sequences corresponds with the sequence diversity commonly observed within Bcc species (Vandamme & Dawyndt, 2011)). For the 16S rRNA nested PCR, amplicons showed the highest similarity (96 %) with sequences of *Comamonadaceae* sp. (FM886892), *Ideonella* sp. (FM886860), *Roseatales* sp. (JQ917995), and *Mitsuaria* sp. (JQ659937), all belonging to the order *Burkholderiales*, yet demonstrating that the nested 16S rRNA PCR assay was not *B. multivorans*-specific. Therefore, only the *B. multivorans* nested *recA* PCR assay was further optimized (i.e. PCR mix, BSA concentration, number of cycles, primer concentration and quantity of pooled PCR product being transferred to the second round of PCR; data not shown) and used to analyze all environmental samples.

3.1.4.2 Detection of *B. multivorans* in environmental samples

All 112 water samples yielded PCR-grade DNA extracts without the need of an extra purification step. *B. multivorans* DNA was detected in 12 water samples (11 %), of which one sample (W132) already yielded a visible amplicon in the first round of PCR (showing 97 % similarity to the *recA* sequence of *B. multivorans* ATCC 17616). For the soil samples (n=27), four yielded PCR-grade DNA without the need of an extra purification step, 21 samples needed extra purification using agarose plugs, and for two soil samples (S6 and S14) no PCR-grade DNA extract could be obtained. The latter soil samples were therefore excluded from the dataset. For soil DNA extracts containing 50 ng/µl or more DNA, an $OD_{320/260}$ ratio of 0.15 or higher was a good indicator for the necessity of an extra purification step to obtain PCR-grade DNA. *B. multivorans* DNA was detected in 23 soil samples (92 %), of which none yielded a visible amplicon in the first round of PCR.

3.1.4.3 Statistical data analysis

The first model tested if type (water, soil) and class (SRW, CRS, O) were significant predictors for *B. multivorans* detection. Binomial logistic regression showed that the type of sample (water versus soil) was indeed a significant predictor for *B. multivorans* detection (Table 3.1). Adding class to the model did not significantly improve the model (data not shown). The odds ratio from this model predicted that there was a 96 times ($e^{4.56} = 95.58$) bigger chance to detect *B. multivorans* in soil than in water samples.

Table 3.1: Binomial logistic regression model 1 (water and soil samples)

	Beta (SE)	**Odds ratio (95 % CI)**
Intercept	-2.12 *** (0.31)	0.12 (0.06; 0.21)
Type		
Water (ref)	—	—
Soil	4.56 *** (0.80)	95.83 (24.48; 646.70)
χ^2	65.50 *** (df = 1)	

Binomial logistic regression with Bm as outcome and Type as predictor. Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' . SE, standard error; CI, confidence interval.

To test which characteristics of the water environment affected the presence of *B. multivorans*, a second model was tested in which pH, temperature and class were assessed as predictors for *B. multivorans* detection. Binomial logistic regression showed that pH and class were significant predictors for *B. multivorans* detection in water samples (Table 3.2). Model 2a includes the main effects of all the variables and shows that the probability of detecting *B. multivorans* was higher in water samples with high pH. The results also show that the

probability of detecting *B. multivorans* was higher in streams (CRS) compared to swimming and recreational waters (SRW). Model 2b includes an additional interaction effect between temperature and pH to test whether the effect of pH on detection depends on the temperature. The model fit of model 2b improved and the interaction effect between temperature and pH indicated that the positive effect of pH was stronger when the temperature of the water was higher.

Table 3.2: Binomial logistic regression model 2 (water samples)

	Model 2a	**Model 2b**	
	Beta (SE)	**Beta (SE)**	**Odds ratio (95 % CI)**
Intercept	-2.72 *** (0.59)	-3.51 *** (0.82)	0.03 (0.00; 0.01)
pH	0.97 . (0.52)	1.70 * (0.69)	5.48 (1.52; 23.84)
Temperature	-0.22 (0.16)	-0.19 (0.16)	0.83 (0.59; 1.13)
Class			
SRW (ref)	—	—	—
O	-0.18 (0.92)	0.16 (0.96)	1.17 (0.14; 7.29)
CRS	1.47 . (0.80)	1.95 * (0.94)	7.06 (1.21; 52.19)
Interaction			
pH*Temperature	—	0.39 * (0.20)	1.47 (1.01; 2.25)
χ^2	8.24 . (df = 4)	12.36 * (df = 5)	

Binomial logistic regression with Bm as outcome and pH, Temperature and Class as predictor, without (model 2a) and with (model 2b) interaction effect between pH and Temperature. Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' . SE, standard error; CI, confidence interval. SRW, swimming or recreational water; CRS, canal-river-stream; O, other.

3.1.4.4 Detection limit *recA* PCR assay

The *recA* PCR assay was applied to serial dilutions of genomic DNA from *B. multivorans* R-20526 to determine its detection limit. The highest 50x dilution testing positive was the fourth one, from which a twofold dilution series was made. The highest twofold dilution testing positive in all three replicate runs of the assay was the second one, containing 1.35 genome equivalents/µl. Given that *recA* is a single-copy gene, and that for each dilution 2x 2 µl was added to the first round of PCR (set up in duplicate), the detection limit was determined to be 5.41 *recA* copies.

3.1.5 Discussion

Because the goal of the present study was to gain insight in the environmental niche of *B. multivorans* in a time and cost-effective way, we chose for a conventional PCR method with gel-based detection of the resulting amplicons, instead of quantitative PCR. *B. multivorans*-specific primers based on both 16S rRNA and *recA* sequences were available from previous

studies (LiPuma *et al.*, 1999; Mahenthiralingam *et al.*, 2000a). Although the *recA* gene has a higher taxonomic resolution (Mahenthiralingam *et al.*, 2000a), a 16S rRNA gene based assay is potentially more sensitive because the *B. multivorans* genome contains five copies of the 16S rRNA gene (Stoddard *et al.*, 2015). However, preliminary tests revealed that the 16S rRNA-based assay yielded only false positive results, as shown by sequence analysis of the resulting amplicons. A nested PCR design was used to increase the sensitivity of the *recA* PCR assay (Drevinek *et al.*, 2002), in which the first PCR was carried out in duplicate and the resulting PCR product was pooled before using it as a template in the second PCR. The *in vitro* detection limit of this PCR assay was therefore determined to be 5.41 *recA* copies or genome equivalents.

The specificity of the *recA* primers (Mahenthiralingam *et al.*, 2000a) used was previously evaluated (Vermis *et al.*, 2002a). The latter study demonstrated that non-specific primer binding could be eliminated by raising the annealing temperature for the *B. multivorans* assay from 58 °C to 64 °C, resulting in a specificity of 100 % (Vermis *et al.*, 2002a). Additionally, the combination of two different primer pairs in our nested PCR design (specific for the Bcc and *B. multivorans*) further minimized false positive results. Finally, all resulting amplicons were sequenced to ensure that only true positive results were recorded. Whereas this final check yielded satisfactory results for the soil samples (one false positive versus 23 true positives), it revealed a strikingly high number of false positives for the water samples (20 false positives versus 12 true positives). The false positives included samples for which the amplicon could not be sequenced (one soil sample and 15 water samples) or samples for which the resulting sequence did not yield any BLAST hits (five water samples). These findings demonstrate that this *recA*-based PCR assay in its current form is not suited for routine analysis of water samples, and that one should re-evaluate PCR-based methods in terms of specificity when applying them to environmental samples (Bergmark *et al.*, 2012).

To eliminate false negative results caused by PCR inhibitors such as humic acids in the environmental DNA extracts, all samples were screened for the presence of PCR inhibitors using a universal 16S rRNA PCR. In case this universal PCR was inhibited, the DNA extracts were purified using low-melting point agarose plugs. Re-testing the DNA extracts after this purification step demonstrated that this purification procedure was very efficient in removing the humic acids from the DNA extracts, as shown previously by Moreira (1998). Furthermore, optical density should not only be measured at 260 nm because this can lead to an overestimation of the DNA concentration if humic acids are present, but also at 320 nm to quantify these contaminations and to evaluate the quality of environmental DNA extracts properly (Miller, 2001; Arbeli & Fuentes, 2007).

Because the few environmental *B. multivorans* isolates that have been reported mainly

originated from water samples (Vermis *et al.*, 2003a; Fang *et al.*, 2011), we hypothesized that water rather than soil was the most likely environmental niche of *B. multivorans*. Our study therefore focused primarily on water samples. However, only 12 out of 112 water samples (11 %) examined were *B. multivorans* PCR positive, in contrast to 23 out of 25 soil samples (92 %) examined. Despite the small number of soil samples, binomial logistic regression showed a highly significant effect ($p < .001$) of the type of sample (Model 1, Table 1) and predicted that there was a 96 times bigger chance to detect *B. multivorans* in soil than in water samples. This result also seemed to suggest that, if present, *B. multivorans* can be isolated more easily from water than from soil samples. Accordingly, Miller *et al.* (2002) showed that although a high percentage of soil samples was PCR positive for Bcc, none of these samples yielded Bcc isolates, demonstrating that cultivation-dependent recovery of Bcc bacteria likely underestimates their prevalence in environmental samples. However, the detection of *B. multivorans* DNA does not necessarily imply that viable target organisms were present at the moment samples were taken (Josephson *et al.*, 1993).

Binomial logistic regression showed that the probability of detecting *B. multivorans* in water samples was higher in streams (CRS) compared to swimming and recreational waters (SRW), and in water samples with a higher pH (Model 2, Table 3.2). Since pH is known to be an important predictor of bacterial diversity in soil (Fierer & Jackson, 2006; Rousk *et al.*, 2010), and *Burkholderia* bacteria have been shown to be acid tolerant (Stopnisek *et al.*, 2014), we expected to detect more *B. multivorans* in acidic versus alkaline waters. The opposite findings of the present study again may suggest that water is not the natural reservoir of *B. multivorans*. However, the number of PCR positive water samples was rather small and therefore one should be careful when extrapolating the results of our binomial logistic regression model 2.

Our finding that *B. multivorans* is widely distributed in soil samples contrasts with the results of previous isolation campaigns, but nevertheless agrees with the notion that the soil environment typically harbors large-genome sized organisms (Konstantinidis & Tiedje, 2004; Raes *et al.*, 2007). As *B. multivorans* harbors a ~6.5 Mb genome, it is equipped with the metabolic versatility needed to thrive in a complex, variable environment such as soil. Future research could focus on the genome biology of this organism, and try to infer the lifestyle of this organism based on genome data (Barberan *et al.*, 2014; Livermore *et al.*, 2014).

3.1.6 Conclusions

In summary, we applied a *recA*-based PCR assay that demonstrated that *B. multivorans* DNA is widely distributed in soil samples but only occasionally in water samples in Flanders,

Belgium. As for all Bcc bacteria it is unclear if and how this mere observation should be implemented in infection control guidelines. Our study also demonstrated that the presently available *B. multivorans recA* primers lack specificity when applied to the analysis of water samples.

3.1.7 Acknowledgments

C. P. is indebted to the Special Research Council of Ghent University.

3.1.8 Supplementary material

Supplementary Figure 3.1: Sampling map. Map showing all sampling locations as produced by the ggmap and ggplot2 packages in R (own figure). Bm, *B. multivorans*.

Supplementary Table 3.1: Sample data

Bcc, *B. cepacia* complex PCR positive; Bm, *B. multivorans* PCR positive. NA, not available. [a]WV, West-Vlaanderen; OV, Oost-Vlaanderen; L, Limburg; VB, Vlaams-Brabant; A, Antwerpen. [b]SRW, swimming or recreational water; CRS, canal-river-stream; O, other.

ID	Date	Type	Region[a]	Class[b]	Temperature	pH	Bcc	Bm	Remarks
W40	13/08/2013	water	OV	CRS	19.0	7.69	no	yes	
W41	13/08/2013	water	OV	SRW	20.0	8.41	no	yes	
W42	13/08/2013	water	OV	SRW	21.0	7.86	no	no	
W43	13/08/2013	water	OV	SRW	20.0	8.33	no	no	
W44	14/08/2013	water	WV	O	16.5	7.32	no	no	
W45	14/08/2013	water	WV	O	19.0	9.59	no	no	
W46	14/08/2013	water	WV	O	16.0	5.49	no	no	
W47	14/08/2013	water	WV	O	17.0	6.70	no	no	
W48	14/08/2013	water	WV	O	20.0	8.80	no	yes	
W49	14/08/2013	water	WV	CRS	20.0	7.82	no	yes	
W50	14/08/2013	water	WV	SRW	22.0	8.87	no	no	
W51	14/08/2013	water	WV	SRW	22.0	8.28	no	no	
W52	14/08/2013	water	WV	CRS	22.0	7.80	no	no	
W53	14/08/2013	water	OV	CRS	21.0	7.77	no	no	
W54	20/08/2013	water	A	SRW	19.0	10.30	no	no	
W55	20/08/2013	water	A	SRW	21.0	8.16	no	no	
W56	20/08/2013	water	A	SRW	20.0	8.10	no	yes	
W57	20/08/2013	water	A	SRW	22.0	8.92	no	no	
W58	20/08/2013	water	A	SRW	22.0	8.92	no	no	
W59	20/08/2013	water	A	O	21.0	7.61	no	no	
W60	20/08/2013	water	A	SRW	23.0	8.63	no	no	
W61	20/08/2013	water	A	SRW	23.0	8.77	no	no	
S3	20/08/2013	soil	A	SRW	NA	6.74	no	yes	
W62	20/08/2013	water	A	SRW	22.0	4.53	no	no	
W63	20/08/2013	water	A	SRW	23.0	4.97	no	no	
W64	20/08/2013	water	A	SRW	23.5	3.79	no	no	
W65	20/08/2013	water	A	SRW	23.0	9.13	no	yes	
W66	20/08/2013	water	A	SRW	23.0	8.41	no	yes	
W67	20/08/2013	water	A	O	22.0	8.38	no	no	
W68	20/08/2013	water	A	CRS	22.0	8.04	no	no	
W69	20/08/2013	water	A	SRW	22.0	7.12	no	no	
W70	20/08/2013	water	A	SRW	22.0	7.35	no	no	
W71	22/08/2013	water	OV	O	19.0	6.56	no	no	
W72	22/08/2013	water	OV	O	18.0	6.65	no	no	
S4	22/08/2013	soil	OV	O	NA	4.11	no	yes	
W73	22/08/2013	water	OV	SRW	21.0	8.87	no	no	
W74	22/08/2013	water	WV	SRW	21.0	8.89	no	no	
S5	22/08/2013	soil	WV	SRW	NA	8.45	no	yes	
W75	22/08/2013	water	WV	SRW	20.5	8.23	no	no	
W76	22/08/2013	water	WV	SRW	20.0	9.01	no	no	
W77	22/08/2013	water	WV	SRW	21.0	8.72	no	no	
W78	22/08/2013	water	WV	CRS	20.0	7.69	no	no	
W79	22/08/2013	water	WV	CRS	20.0	8.26	no	no	

Continued on next page

Continued from previous page

ID	Date	Type	Region[a]	Class[b]	Temperature	pH	Bcc	Bm	Remarks
W80	22/08/2013	water	WV	SRW	22.0	8.63	no	no	
W81	27/08/2013	water	OV	O	19.0	7.67	no	no	
S6	27/08/2013	soil	OV	O	NA	3.92	NA	NA	PCR inhibitors
W82	27/08/2013	water	OV	O	21.5	7.59	no	no	
S7	27/08/2013	soil	OV	O	NA	8.63	no	yes	
W83	27/08/2013	water	OV	O	20.0	7.37	no	no	
W84	27/08/2013	water	OV	O	21.0	8.70	no	yes	
W85	27/08/2013	water	OV	SRW	22.5	8.80	no	no	
S8	27/08/2013	soil	OV	SRW	NA	8.30	no	yes	
W86	27/08/2013	water	OV	CRS	22.0	7.85	no	no	
W87	27/08/2013	water	OV	O	23.0	8.11	no	no	
S9	27/08/2013	soil	OV	O	NA	3.53	no	yes	
W88	27/08/2013	water	OV	SRW	22.0	8.47	no	no	
W89	27/08/2013	water	OV	O	22.0	8.11	no	no	
W90	27/08/2013	water	OV	O	22.0	7.28	no	no	
S10	27/08/2013	soil	OV	O	NA	8.01	no	yes	
W91	27/08/2013	water	OV	O	24.0	8.60	no	no	
W92	27/08/2013	water	OV	CRS	23.0	8.14	no	no	
W93	27/08/2013	water	OV	CRS	22.0	7.83	no	yes	
W94	29/08/2013	water	OV	CRS	20.0	7.53	no	no	
W95	29/08/2013	water	OV	SRW	20.0	7.87	no	no	
W96	29/08/2013	water	OV	SRW	20.0	7.66	no	no	
W97	29/08/2013	water	OV	CRS	21.0	7.32	no	no	
W98	29/08/2013	water	WV	SRW	21.0	7.96	no	no	
W99	29/08/2013	water	WV	SRW	21.0	8.25	no	no	
W100	29/08/2013	water	WV	SRW	21.0	7.53	no	no	
W101	29/08/2013	water	OV	SRW	21.0	7.75	no	no	
W102	29/08/2013	water	OV	SRW	22.0	8.20	no	no	
W103	2/09/2013	water	A	SRW	20.5	8.80	no	yes	
W104	2/09/2013	water	A	SRW	21.0	8.79	no	no	
W105	2/09/2013	water	A	SRW	19.0	8.70	no	no	
W106	2/09/2013	water	A	CRS	19.0	8.06	no	no	
W107	2/09/2013	water	A	O	21.0	7.48	no	no	
S11	2/09/2013	soil	A	O	NA	3.69	no	yes	
W108	2/09/2013	water	A	O	20.0	7.87	no	no	
W109	2/09/2013	water	A	SRW	21.0	8.63	no	no	
S12	2/09/2013	soil	A	SRW	NA	8.54	no	yes	
W110	2/09/2013	water	VB	SRW	21.0	7.94	no	no	
W111	2/09/2013	water	VB	O	19.0	8.89	no	no	
W112	2/09/2013	water	A	SRW	22.0	8.28	no	no	
W113	2/09/2013	water	A	SRW	21.5	8.05	no	no	
W114	2/09/2013	water	A	SRW	22.0	8.38	no	no	
W115	2/09/2013	water	A	O	21.5	7.73	no	no	
W116	2/09/2013	water	A	O	21.5	7.44	no	no	
W117	2/09/2013	water	A	CRS	21.0	7.74	no	no	
W118	2/09/2013	water	A	SRW	22.0	8.43	no	no	
S13	2/09/2013	soil	A	SRW	NA	6.26	no	yes	
W119	2/09/2013	water	OV	SRW	20.0	8.25	no	no	

Continued on next page

Continued from previous page

ID	Date	Type	Region[a]	Class[b]	Temperature	pH	Bcc	Bm	Remarks
W120	5/09/2013	water	VB	O	17.5	7.25	no	no	
S14	5/09/2013	soil	VB	O	NA	8.44	NA	NA	PCR inhibitors
W121	5/09/2013	water	VB	CRS	18.0	7.73	no	no	
W122	5/09/2013	water	VB	O	21.5	7.16	no	no	
W123	5/09/2013	water	VB	O	18.0	7.54	no	no	
W124	5/09/2013	water	VB	O	21.0	8.22	no	no	
W125	5/09/2013	water	VB	O	24.0	7.83	no	no	
W126	5/09/2013	water	VB	SRW	23.0	8.31	no	no	
W127	5/09/2013	water	VB	SRW	24.0	8.03	no	no	
W128	5/09/2013	water	VB	CRS	23.0	7.65	no	no	
W129	5/09/2013	water	VB	CRS	23.0	7.94	no	no	
W130	11/09/2013	water	L	SRW	17.0	7.24	no	no	
W131	11/09/2013	water	L	SRW	17.5	7.16	no	no	
W132	11/09/2013	water	L	CRS	17.0	7.40	yes	yes	
W133	11/09/2013	water	L	SRW	18.0	7.81	no	no	
W134	11/09/2013	water	L	SRW	17.0	7.98	no	no	
W135	11/09/2013	water	L	SRW	17.0	6.02	no	no	
W136	11/09/2013	water	L	SRW	18.5	6.49	no	no	
W137	11/09/2013	water	L	O	19.0	6.37	no	no	
W138	11/09/2013	water	L	O	16.0	6.40	no	no	
W139	11/09/2013	water	L	CRS	13.5	6.31	no	yes	
W140	11/09/2013	water	L	CRS	20.5	7.94	no	no	
W141	11/09/2013	water	L	O	18.0	6.50	no	no	
S15	11/09/2013	soil	L	O	NA	7.05	no	yes	
W142	11/09/2013	water	L	O	17.0	6.49	no	no	
W143	11/09/2013	water	L	CRS	15.0	7.04	no	no	
S16	11/09/2013	soil	L	CRS	NA	4.06	no	yes	
W144	11/09/2013	water	L	CRS	16.0	7.40	no	no	
W145	11/09/2013	water	L	O	18.5	7.73	no	no	
W146	11/09/2013	water	L	O	19.0	6.00	no	no	
W147	11/09/2013	water	L	O	18.0	7.71	no	no	
W148	11/09/2013	water	L	SRW	19.5	8.33	no	no	
W149	11/09/2013	water	L	SRW	18.5	7.88	no	no	
W150	11/09/2013	water	L	SRW	19.0	7.60	no	no	
S17	17/09/2013	soil	OV	O	NA	7.97	no	yes	
S18	17/09/2013	soil	OV	O	NA	7.94	no	yes	
W151	19/09/2013	water	OV	O	27.0	8.36	no	no	
S19	19/09/2013	soil	OV	O	NA	9.61	no	yes	
S20	19/09/2013	soil	OV	O	NA	5.18	no	yes	
S21	19/09/2013	soil	OV	O	NA	8.90	no	yes	
S22	21/09/2013	soil	OV	O	NA	6.59	no	no	
S23	21/09/2013	soil	OV	O	NA	6.09	no	yes	
S25	22/09/2013	soil	WV	O	NA	6.68	no	no	
S26	12/10/2013	soil	WV	O	NA	7.53	no	yes	
S27	12/10/2013	soil	WV	O	NA	4.72	no	yes	
S28	12/10/2013	soil	WV	O	NA	7.28	no	yes	
S29	12/10/2013	soil	WV	O	NA	7.40	no	yes	
S30	12/10/2013	soil	WV	O	NA	7.77	no	yes	

3.2 Extensive cultivation of soil and water samples yields various pathogens in patients with cystic fibrosis but not *Burkholderia multivorans*

Redrafted from: Charlotte Peeters, Eliza Depoorter, Jessy Praet and Peter Vandamme (2016). Extensive cultivation of soil and water samples yields various pathogens in patients with cystic fibrosis but not *Burkholderia multivorans*. Journal of Cystic Fibrosis, in press.

Author contributions: CP and PV conceived the study and wrote the manuscript. CP, ED and JP performed the experiments. CP analyzed the data. ED and JP proofread the manuscript.

3.2.1 Abstract

Background: While the epidemiology of Bcc bacteria in CF patients suggests that *B. multivorans* is acquired from environmental sources, this species has rarely been isolated from water and soil samples.
Methods: Multiple isolation strategies were applied to water and soil samples that were previously shown to be *B. multivorans* PCR positive. These included direct plating and liquid enrichment procedures and the use of selective media, acclimatizing recovery and co-cultivation with CF sputum. MALDI-TOF MS and sequence analysis of 16S rRNA and housekeeping genes were used to identify all isolates.
Results: None of the approaches yielded *B. multivorans* isolates. Other *Burkholderia* species, several GNNF bacteria (including *Cupriavidus*, *Inquilinus*, *Pandoraea*, *Pseudomonas* and *Stenotrophomonas*) and rapidly growing mycobacteria (including *Mycobacterium chelonae*) were all isolated from water and soil samples.
Conclusions: The use of Bcc isolation media yielded a surprisingly wide array of rare but often clinically relevant CF pathogens, confirming that water and soil are reservoirs of these infectious agents.

3.2.2 Introduction

Bcc bacteria are rare but important pathogens in CF patients which may have a high impact on morbidity, mortality and post-lung transplant survival (De Boeck *et al.*, 2004; LiPuma, 2010). The Bcc comprises 20 validly named species and at least ten additional species await formal description (Vandamme & Peeters, 2014; De Smet *et al.*, 2015a). Bcc bacteria have

a versatile lifestyle and all Bcc species except *Burkholderia metallica* (only clinical) have been isolated from both CF patients and environmental samples (Coenye & Vandamme, 2003; Baldwin *et al.*, 2007) (http://pubmlst.org/bcc/). During the last decade infection control measures reduced patient-to-patient transmission which is mostly associated with *B. cenocepacia*. Subsequently, *B. multivorans* emerged as the most prevalent Bcc pathogen in many countries (Brisse *et al.*, 2004; De Boeck *et al.*, 2004; Govan *et al.*, 2007; LiPuma, 2010; Norskov-Lauritsen *et al.*, 2010; Pope *et al.*, 2010; Medina-Pascual *et al.*, 2012).
The low number of epidemic outbreaks caused by *B. multivorans* (Whiteford *et al.*, 1995; Segonds *et al.*, 1999; Biddick *et al.*, 2003) and the fact that *B. multivorans* isolates from CF patients commonly represent unique strains suggest that there is limited person-to-person transmission and that strains are acquired from non-human sources such as the natural environment (Baldwin *et al.*, 2008). In a previous study, we developed a cultivation-independent PCR assay for the detection of *B. multivorans* in environmental samples (Peeters *et al.*, submitted), demonstrating that *B. multivorans* is indeed widespread in water and soil samples.
Although many studies have described the isolation of *B. cenocepacia* from rhizosphere samples (Balandreau *et al.*, 2001; LiPuma *et al.*, 2002; Pirone *et al.*, 2005; Ramette *et al.*, 2005; Zhang & Xie, 2007), only few reported on the isolation of *B. multivorans* from environmental samples such as maize rhizosphere soil (Ramette *et al.*, 2005), greenhouse soil (Vanlaere *et al.*, 2005) and river (Vermis *et al.*, 2003a) and lake water (Fang *et al.*, 2011). Since *B. multivorans* is the most commonly isolated Bcc species from CF samples in Belgium (De Boeck *et al.*, 2004) we aimed to compare clinical and environmental *B. multivorans* isolates from Belgium in an epidemiological study. Therefore, we applied different isolation strategies including selective and non-selective liquid enrichment (Vanlaere *et al.*, 2005; Ahn *et al.*, 2014), acclimatizing recovery (Hahn *et al.*, 2004; Vartoukian *et al.*, 2010) and co-cultivation with CF sputum (Vartoukian *et al.*, 2010; Epstein, 2013) on *B. multivorans* PCR positive (Peeters *et al.*, submitted) water and soil samples.

3.2.3 Materials and methods

3.2.3.1 Samples

An autoclaved 1 L Duran bottle was opened and filled 10 cm below the water surface to collect water samples. Soil samples were taken 2 cm below the soil surface using a sterile spoon and collected in sterile falcon tubes. Temperature and pH of the water samples were measured on site. The pH of the soil samples was measured after dissolving 10 g of soil in 50 ml distilled water and magnetic stirring for 10 min. Fresh samples were taken for all

experiments (Table 3.4).

Table 3.4: Sample overview

Experiment	Sample	Sampling date	Sampling source[a]	pH	Temperature[b]
A	W2	03/05/2013	Freshwater greenhouse	7.83	23 °C
A	W33	03/05/2013	Pond greenhouse	7.73	19 °C
B	S32	10/01/2014	Soil	6.23	-
B	S35	10/01/2014	Soil	5.62	-
C	W2	19/08/2014	Freshwater greenhouse	8.36	24 °C
D	W2	07/10/2014	Freshwater greenhouse	8.50	25 °C
D	W152	07/10/2014	Pond water	7.33	15 °C
D	S35	07/10/2014	Soil	5.57	-

[a]All samples were taken in the botanical garden of Ghent University, Ghent, Belgium. [b]For soil samples no temperature was recorded.

3.2.3.2 Isolation media

Four different isolation media were used. BCEM (*B. cepacia* complex enrichment medium) (Vanlaere *et al.*, 2005) contained 0.1 % L-arabinose (Sigma), 0.1 % L-threonine (Sigma), 0.1 % D-cellobiose (Sigma), 0.5 % NaCl (Merck), 0.02 % $MgSO_4 \cdot 7\,H_2O$ (Merck), 0.1 % $NH_4H_2PO_4$ (Sigma), 0.1 % K_2HPO_4 (UCB) and 0.002 % yeast extract (Oxoid). MES (4-morpholine-ethanesulfonic acid) buffer (Sigma) adjusted to pH 5.5 was added to a final concentration of 50 mM after autoclaving. Selective agents were added after autoclaving and included 600,000 U/L polymyxin B sulfate (Sigma), 10 mg/L gentamicin (Sigma) and 2 mg/L vancomycin (Sigma). For cycloheximide (Sigma), the final concentration was 10 mg/L and 200 mg/L for enrichment broth and agar medium, respectively. For agar media, 12 g/L (experiment A) or 15 g/L (experiment B) agar was added before autoclaving.
PCAT (*Pseudomonas cepacia* azelaic acid tryptamine) medium (Burbage & Sasser, 1982) contained 0.2 % azelaic acid (Fluka), 0.02 % $MgSO_4 \cdot 7\,H_2O$ (Merck), 0.4 % KH_2PO_4 (Merck), 0.4 % K_2HPO_4 (UCB) and 0.002 % yeast extract (Oxoid). The pH was adjusted to 5.7. Tryptamine (Sigma) was prepared as a 100x stock solution and added to a final concentration of 200 mg/L after autoclaving. For cycloheximide (Sigma), the final concentration was 10 mg/L and 200 mg/L for enrichment broth and agar medium, respectively. For agar media, 12 g/L (experiment A) or 15 g/L (experiment B) agar was added before autoclaving.
The 1/10 PCAT medium contained all ingredients in a tenfold lower concentration compared to PCAT, except cycloheximide and agar.
BCSA (*B. cepacia* selective agar) medium (Henry *et al.*, 1997) contained 1 % sucrose (Merck), 1 % lactose (Merck), 1 % trypticase peptone (Oxoid), 0.5 % NaCl (Merck) and 0.15 % yeast extract (Oxoid). Phenol red and crystal violet were prepared as 100x stock solutions and were

added to a final concentration of 0.008 % and 0.0002 %, respectively. The pH was adjusted to 7.0. Selective agents were added after autoclaving and included 600,000 U/L polymyxin B sulfate (Sigma), 10 mg/L gentamicin (Sigma), 2.5 mg/L vancomycin (Sigma) and 200 mg/L cycloheximide (Sigma). For agar media, 14 g/L agar was added before autoclaving.

3.2.3.3 Isolation experiments

Experiment A. For water samples W2 and W33, three aliquots of 150 ml were filtered using a Nalgene vacuum filter funnel and cellulose nitrate membrane filters with 0.45 µm pore size and 47 mm diameter (Thermo Scientific) to collect bacterial cells. One filter was incubated for 5 days at 28 °C in 25 ml BCEM and PCAT broth each while shaking. Aliquots (50 µl) of tenfold serial dilutions (1E0 to 1E-9) in physiological saline of both enrichment broths were plated on the respective agar media. The third filter was washed with 1 ml physiological saline of which 100 µl was plated directly on BCEM and PCAT agar media. After incubation at 28 °C for 5 and 14 days (direct plating) or 6 days (plating of enrichment broth), colonies were randomly picked and subcultivated on buffered nutrient agar (Oxoid, pH 6.8).

Experiment B. For soil samples S32 and S35, 25 g of soil was homogenized in 225 ml physiological saline using a Stomacher blender for 30 s at 230 rpm. Aliquots (50 µl) of tenfold serial dilutions (1E0 to 1E-9) in physiological saline were plated on BCEM and PCAT agar media. After incubation at 28 °C for 4, 6 and 41 days (S32) or 4 and 6 days (S35), colonies were randomly picked and subcultivated on buffered nutrient agar (Oxoid, pH 6.8).

Experiment C. For water sample W2, three aliquots of 150 ml were filtered as described for experiment A and each filter was incubated in 25 ml enrichment broth: one in 0.1 % peptone for non-selective recovery (NR), one in BCEM broth for selective recovery (SR) and one in a broth containing 50 % BCEM broth and 50 % W2 filtrate for acclimatizing recovery (AR). After 3 days, the AR broth was replaced by pelleting the cell material (10 min at 8000 rpm or 8228 g), removing the supernatants and adding 25 ml of fresh BCEM broth which was further incubated at 28 °C for another 4 days while shaking. Aliquots (50 µl) of tenfold serial dilutions in physiological saline were plated on 1/10 PCAT and BCSA agar media after 2 h (D0) for NR, 1 day (D1) for NR and SR, 3 days (D3) for NR and SR, 7 days (D7) for SR and AR and 14 days (D14) for SR and AR. After incubation at 28 °C for 6 days (D0, D1, D3 and D7) or 8 days (D14), colonies were randomly picked and subcultivated on buffered nutrient agar (Oxoid, pH 6.8).

Experiment D. For water samples W2 and W152, 300 ml was filtered as described for experiment A and each filter was shaken in 25 ml of its own sterile filtrate for 2 h at 28 °C. For soil sample S35, 2.5 g of soil was shaken in 22.5 ml physiological saline for 2 h at 28 °C. Four ml of the concentrated water samples and 2 ml of soil the suspension were added to a single well of a Greiner 6-well plate. Four ml of physiological saline was used as negative control (NC). ThinCert membranes (Greiner 657 640, pore size 0.4 µm) were placed into each well and 2 ml CF sputum was added into the ThinCert compartment. Aliquots (50 µl) of tenfold serial dilutions in physiological saline were plated on 1/10 PCAT and BCSA agar media after 3 (D3), 7 (D7) and 14 (D14) days at 28 °C. After incubation at 28 °C for 6 days (D3 and D14) or 7 days (D7), colonies were randomly picked and subcultivated on buffered nutrient agar (Oxoid, pH 6.8).

3.2.3.4 Identification of isolates

A total of 964 isolates were subcultivated at least twice and checked for purity. MALDI-TOF MS spectra were generated using a 4800 Plus MALDI TOF/TOF Analyzer (AB Sciex) as previously described (Wieme *et al.*, 2014b). Spectra were clustered through curve-based data analysis using the Pearson product-moment correlation coefficient and UPGMA clustering algorithm in BioNumerics 7.5 (Applied Maths). At least one representative isolate was selected per cluster and tentatively identified through 16S rRNA gene sequence analysis. Accurate species level identification of *Burkholderia*, *Stenotrophomonas* and *Mycobacterium* isolates was further obtained through *recA*, *rpoD* and *rpoB* gene sequence analysis, respectively (see supplementary material). Isolates for which only 16S rRNA gene sequences were generated were identified using the EzTaxon database (Kim *et al.*, 2012). An isolate was considered identified at the genus and species level if its 16S rRNA sequence showed at least 97 % and 98.7 % similarity, respectively, to the sequence of the respective taxonomic type strain (Stackebrandt & Ebers, 2006) (Supplementary Table 3.2). The *recA*, *rpoD* and *rpoB* sequences were compared with those of type and reference strains of validly named *Burkholderia*, *Stenotrophomonas* and *Mycobacterium* species, respectively.

3.2.4 Results

3.2.4.1 Experiment A: direct plating and liquid enrichment of water samples

Two water samples (W2 and W33) were plated both directly and after liquid enrichment on BCEM and PCAT medium. MALDI-TOF MS dereplication of 332 isolates resulted in a total of 46 clusters, of which 20 clusters (243 isolates) were identified as various GNNF bacteria and one cluster (comprising spectra of a single isolate) as *Mycobacterium peregrinum* (Table 3.5).

The remaining 25 clusters (88 isolates) were identified as Gram-positive or Gram-negative fermenting bacteria (data not shown).

Table 3.5: Identification and number of isolates for experiment A

	W2				W33			
	DP		EP		DP		EP	
	BCEM	PCAT	BCEM	PCAT	BCEM	PCAT	BCEM	PCAT
Aminobacter aminovorans					1			
Bosea vestrisii	1*							
Burkholderia fungorum			6					
Burkholderia sp.				1		1		
Chromobacterium aquaticum					34		5	
Chryseobacterium sp.					15			
Cupriavidus metallidurans			4					
Flavobacterium arsenitoxidans					5			
Flavobacterium sp.	2							
Pandoraea sputorum			40					
Pseudomonas mosselii						3		
Pseudomonas nitroreducens				53				19
Rhizobium lusitanum		2						
Sphingobacterium multivorum			11					
Sphingobacterium siyangense								32
Stenotrophomonas maltophilia				4				4
Total GNNF	3	2	61	58	55	4	5	55
Mycobacterium peregrinum					1*			
Total isolates	10	4	72	66	63	19	33	65

Each cell represents the number of isolates per MALDI-TOF MS cluster. DP, direct plating; EP, liquid enrichment prior to plating. W2 and W33, water samples W2 and W33 (Table 3.4). *, picked after 14 days of incubation.

3.2.4.2 Experiment B: direct plating of soil samples

Two soil samples (S32 and S35) were plated directly on BCEM and PCAT medium. MALDI-TOF MS dereplication of 294 isolates collected after 4 or 6 days of incubation resulted in a total of 103 clusters, of which 41 clusters (185 isolates) were identified as GNNF bacteria and one cluster (2 isolates) as *Mycobacterium septicum* (Table 3.6). The remaining 61 clusters (107 isolates) were identified as Gram-positive or Gram-negative fermenting bacteria (data not shown). For sample S32, 63 additional isolates were picked after 41 days of incubation, of which two were identified as *M. septicum* and one as *Mycobacterium smegmatis*.

3.2.4.3 Experiment C: liquid recovery of water samples

For water sample W2 SR, NR and AR liquid recovery strategies were applied and the resulting broth was plated on BCSA and 1/10 PCAT medium. MALDI-TOF MS dereplication of 263 isolates resulted in a total of 31 clusters, of which 21 clusters (243 isolates) were identified

Table 3.6: Identification and number of isolates for experiment B

	S32		S35	
	BCEM	**PCAT**	**BCEM**	**PCAT**
Burkholderia caledonica				21(3)
Burkholderia choica		47(6)		1
Burkholderia cordobensis		2		
Burkholderia glathei		9		1
Burkholderia graminis				15
Burkholderia humi				3
Burkholderia jiangsuensis		14(3)		11
Burkholderia phytofirmans			8	2
Burkholderia sp.		13(9)	1	26(5)
Chryseobacterium lactis	1			
Pseudomonas jessenii		5		2
Pseudomonas migulae				2
Rhizobium metallidurans		1		
Total GNNF	1	91	9	84
Mycobacterium smegmatis	1*			
Mycobacterium septicum	2*	2		
Total isolates	16	153	20	105

Each cell represents the number of isolates per MALDI-TOF MS cluster. The number of clusters is given in between brackets if more than one. S32 and S35, soil samples S32 and S35 (Table 3.4). *, picked after 41 days of incubation and not examined by MALDI-TOF MS.

as GNNF bacteria and one cluster (2 isolates) as *Mycobacterium chelonae* (Table 3.7). The remaining 9 clusters (18 isolates) were identified as Gram-positive or Gram-negative fermenting bacteria (data not shown).

3.2.4.4 Experiment D: liquid enrichment in the presence of CF sputum

Two water samples (W2 and W152) and one soil sample (S35) were incubated for up to two weeks with CF sputum in a transmembrane system prior to plating on BCSA and 1/10 PCAT. For S35, no growth was observed on BCSA. MALDI-TOF MS dereplication of 75 isolates resulted in a total of 19 clusters, of which 10 clusters (58 isolates) were identified as GNNF bacteria and one cluster (3 isolates) as *M. chelonae* (Table 3.8). The remaining 8 clusters (14 isolates) were identified as Gram-positive or Gram-negative fermenting bacteria (data not shown).

3.2.5 Discussion

While the epidemiology of Bcc bacteria in CF patients suggests that *B. multivorans* is acquired from non-human sources such as the natural environment (Baldwin *et al.*, 2008) it has rarely been isolated from environmental samples. The comparison of clinical and environmental

Table 3.7: Identification and number of isolates for experiment C

	NR		SR		AR	
	1/10 PCAT	**BCSA**	**1/10 PCAT**	**BCSA**	**1/10 PCAT**	**BCSA**
Acinetobacter calcoaceticus	17					
Chryseobacterium rhizoplanae	2	4	36	6	2	1
Comamonas sp.	2					
Delftia lacustris					2	15
Inquilinus limosus	2					
Pseudomonas sp.	1					
Rhizobium radiobacter	4					
Sphingobacterium detergens			68	2	51	2
Sphingobacterium multivorum			2			
Stenotrophomonas maltophilia	21(2)					
Stenotrophomonas rhizophila			3			
Total GNNF	49	4	109	8	55	18
Mycobacterium chelonae		2				
Total isolates	57	6	111	8	62	19

Each cell represents the number of isolates per MALDI-TOF MS cluster. The number of clusters is given in between brackets if more than one. NR, non-selective recovery; SR, selective recovery; AR, acclimatizing recovery.

isolates may provide new insights in the epidemiology of this CF pathogen. We therefore applied in a stepwise manner multiple isolation strategies to water and soil samples that were previously shown to be PCR positive for *B. multivorans* (Peeters *et al.*, submitted). Since the diversity of the obtained isolates is strongly dependent on the selective medium (Tabacchioni *et al.*, 2000), two selective media were used in parallel in each experiment.

Several isolation media have been developed for the recovery of Bcc bacteria from CF sputum (Gilligan *et al.*, 1985; Welch *et al.*, 1987; Henry *et al.*, 1997). However, because of their high nutrient and antibiotic concentrations and/or low specificity these media are generally not suited for the isolation of Bcc bacteria from environmental samples and specific media for such samples have been developed (Burbage & Sasser, 1982; Hagedorn *et al.*, 1987). BCEM medium was used in our first isolation experiments since its use yielded a few *B. multivorans* isolates from water and soil samples in earlier studies (Vermis *et al.*, 2003a; Vanlaere *et al.*, 2005); the pH was set to pH 5.5 (Burbage & Sasser, 1982; Hagedorn *et al.*, 1987) and L-arabinose, D-cellobiose and L-threonine were simultaneously included as carbon sources (Vermis *et al.*, 2003a).

Direct plating of two soil samples yielded no *B. multivorans* isolates, neither on BCEM nor on PCAT medium which is in agreement with previous isolation studies in which soil samples were plated on PCAT medium and many other Bcc species were isolated, but no *B. multivorans* (Balandreau *et al.*, 2001; Pirone *et al.*, 2005; Zhang & Xie, 2007). Direct plating of water samples yielded no *B. multivorans* isolates either. Vermis *et al.* (2003a)

Table 3.8: Identification and number of isolates for experiment D

	W2		W152		S35
	1/10 PCAT	**BCSA**	**1/10 PCAT**	**BCSA**	**1/10 PCAT**
Delftia lacustris				5	
Elizabethkingia miricola				6	
Pseudomonas kilonensis					1
Pseudomonas protegens			11		5
Pseudomonas soli	3				
Sphingobacterium faecium				3	
Stenotrophomonas maltophilia	6		6		
Stenotrophomonas rhizophila					12
Total GNNF	9	0	17	14	18
Mycobacterium chelonae		3			
Total isolates	9	4	19	14	29

Each cell represents the number of isolates per MALDI-TOF MS cluster. W2, W152 and S35, water sample W2, water sample W152 and soil sample S35 (Table 3.4).

reported that liquid enrichment prior to plating may enhance the isolation of Bcc from water samples. Concordantly, Ahn *et al.* (2014) showed that broth media were more effective than solid media for the recovery of *B. cenocepacia* from distilled water. However, Fang *et al.* (2011) isolated several Bcc bacteria via direct plating of concentrated water samples on PCAT, suggesting that liquid enrichment is not always necessary. In our study, neither strategy yielded *B. multivorans* isolates. Nevertheless, a higher number of other, non-Bcc *Burkholderia* isolates was obtained using liquid enrichment prior to plating compared to direct plating (Table 3.5).

Since conventional growth media are often considered too nutrient-rich (Vartoukian *et al.*, 2010) to allow for optimal recovery of environmental isolates, a tenfold diluted formulation of PCAT was subsequently compared with conventional PCAT for direct plating of soil samples. Preliminary experiments (data not shown) showed that a higher number of distinct colony morphologies were present on the diluted formulation suggesting that 1/10 PCAT indeed yielded a higher diversity. Therefore, 1/10 PCAT instead of PCAT was used in the remaining experiments.

Because direct plating of soil and liquid enrichment of water prior to plating did not yield *B. multivorans* isolates, we applied several alternative strategies to enable the cultivation of fastidious bacteria (Vartoukian *et al.*, 2010; Epstein, 2013). Besides liquid enrichment of a concentrated water sample in selective broth, we also applied non-selective recovery in 0.1 % peptone solution. This way, debilitated cells are given the chance to resuscitate prior to being exposed to the selective medium (Straka & Stokes, 1957; Ahn *et al.*, 2014). Furthermore, acclimatizing recovery is based on the filtration-acclimatization method (Hahn

et al., 2004) and allows the inclusion of essential nutrients and/or signaling molecules from the original environment, together with a gradual adaptation to increasing nutrient and antibiotic concentrations (Hahn *et al.*, 2004; Vartoukian *et al.*, 2010). Finally, we tried to mimic the CF lung environment by incubating a water sample in the presence of CF sputum, separated only by a semipermeable membrane (Epstein, 2013). Neither of these liquid enrichment strategies yielded *B. multivorans* isolates.

We cannot explain why we failed to isolate *B. multivorans* from these PCR positive water and soil samples. Sequence analysis of *B. multivorans*-specific amplicons of each of these samples earlier demonstrated that they were true positives (Peeters *et al.*, submitted). The presence of *B. multivorans* DNA implies that viable target organisms were present at or before the time of sampling (Josephson *et al.*, 1993). Even after exposure to CF sputum for up to two weeks, no *B. multivorans* was isolated, while this species can be readily cultivated from CF sputum samples. Additionally, Lin *et al.* (2011) showed that *B. multivorans* can survive in a broader range of pH, temperatures and salt concentrations than *B. pseudomallei*, again demonstrating that this species should be able to adapt well to environmental changes and varying growth conditions.

Although no *B. multivorans* was isolated, soil samples in particular yielded several non-Bcc *Burkholderia* species (Tables 3.5-3.6). These organisms were primarily isolated on PCAT medium. Additionally, although only two soil and three water samples were examined, a large number of other GNNF bacteria, including known CF pathogens such as *Cupriavidus*, *Inquilinus*, *Pandoraea*, *Pseudomonas* and *Stenotrophomonas* spp., were isolated. The prevalence of *Stenotrophomonas maltophilia* infection in CF patients ranges from 0 to 17.5 % in European countries (Zolin *et al.*, 2014). *Cupriavidus*, *Inquilinus* and *Pandoraea* are rare causes of severe respiratory exacerbations in CF patients (Jorgensen *et al.*, 2003; Schmoldt *et al.*, 2006; Kalka-Moll *et al.*, 2009) but their general prevalence in CF is unknown (LiPuma, 2010). Our findings confirm that water and soil are indeed natural reservoirs of these rare but potentially virulent pathogens (LiPuma, 2010). The same five samples also yielded mycobacterial isolates. Rapidly growing mycobacteria have emerged as significant pathogens in CF and their prevalence is increasing (Bar-On *et al.*, 2015; Preece *et al.*, 2016), ranging from 0 to 4.4 % in European countries (Zolin *et al.*, 2014). Similar to Bcc bacteria, they can be recovered from various environmental sources, both natural and engineered (van Ingen *et al.*, 2009). In the present study, *M. chelonae* was isolated from water sample W2 using BCSA as selective isolation medium. A second water sample (W33) yielded *M. peregrinum*. The latter strain was isolated after direct plating on BCEM and 14 days of incubation, confirming that an extended incubation time of BCSA may enhance the recovery of some mycobacteria (Esther *et al.*, 2011). From soil sample S32, *M. septicum* was isolated on both

PCAT and BCEM, while *M. smegmatis* was isolated on BCEM only.
When comparing the three isolation media used (BCEM, PCAT and BCSA), it was clear that each medium selected for specific genera. *Pandoraea* and *Chromobacterium* isolates were recovered exclusively from BCEM, while *Pseudomonas* was recovered exclusively from PCAT medium. The latter is in accordance with the presence of gentamicin and polymyxin B sulfate in BCEM and BCSA. Also the absence of *S. maltophilia* isolates from BCSA agrees with earlier reports (Henry *et al.*, 1999).

3.2.6 Acknowledgments

C. P. is indebted to the Special Research Council of Ghent University.

3.2.7 Supplementary material

3.2.7.1 Methods gene sequence analysis

DNA was prepared by alkaline lysis. One colony was suspended in 20 µl lysis buffer (0.25 % SDS, 0.05 M NaOH) and heated at 95 °C for 15 min. After cooling on ice, the lysate was diluted with 180 µl of sterile MQ and centrifuged for 5 min at max speed. Sequencing of the *recA* (*Burkholderia* species) and 16S rRNA gene were performed as described previously (Peeters *et al.*, 2013) with the *recA* primers from Spilker *et al.* (2009).
For *Stenotrophomonas* species, the *rpoD* gene was amplified and sequenced with forward primer 5'-YAT GMG NGA RAT GGG NAC NGT-3' and reverse primer 5'-NGC YTC NAC CAT YTC YTT YTT-3' (Svensson-Stadler, L.A., personal communication). Each 25 µl PCR reaction consisted of 1x PCR buffer (Qiagen), 0.75 U of Taq polymerase (Qiagen), 200 µM of each dNTP (Applied Biosytems), 1x Q-solution (Qiagen), 1 µM of each primer and 2 µl of DNA. PCR was performed using a MJ Research PTC-100 thermal cycler. Initial denaturation for 2 min at 95 °C was followed by 35 cycles of 30 s at 95 °C, 1 min at 55 °C and 2 min at 72 °C, and a final elongation for 10 min at 72 °C.
For *Mycobacterium* species, the *rpoB* gene was amplified and sequenced using the primers MycoF and MycoR from Adekambi *et al.* (2003). Each 25 µl PCR reaction consisted of 1x PCR buffer (Qiagen), 0.625 U of Taq polymerase (Qiagen), 200 µM of each dNTP (Applied Biosytems), 1x Q-solution (Qiagen), 0.5 µM of each primer and 2.5 µl of DNA. PCR was performed using a Verity thermal cycler following the PCR protocol of Adekambi *et al.* (2003).
Sequence assembly and analysis were performed using BioNumerics 7.5 (Applied Maths).

3.2.7.2 Isolates from the present study

Supplementary Table 3.2: Isolates from the present study

Isolates from the present study with their identification and the accession numbers of the gene sequences used for identification.

Experiment	Sample	Isolate	MALDI cluster	Identification	Gene	Accession
A	W33	R-52011	W33-16	*Aminobacter aminovorans*	16S rRNA	LN995690
A	W2	R-51039	W2-25	*Bosea vestrisii*	16S rRNA	LN995688
A	W2	R-51028	W2-17	*Burkholderia fungorum*	*recA*	LN995631
A	W2	R-51032	W2-17	*Burkholderia fungorum*	*recA*	LN995632
A	W2	R-51037	W2-23	*Burkholderia* sp.	*recA*	LN995633
A	W33	R-51042	W33-20	*Burkholderia* sp.	*recA*	LN995634
A	W33	R-50671	W33-12	*Chromobacterium aquaticum*	16S rRNA	LN995680
A	W33	R-50668	W33-13	*Chryseobacterium* sp.	16S rRNA	LN995678
A	W2	R-51029	W2-15	*Cupriavidus metallidurans*	16S rRNA	LN995683
A	W33	R-51041	W33-15	*Flavobacterium arsenitoxidans*	16S rRNA	LN995689
A	W2	R-51025	W2-26	*Flavobacterium* sp.	16S rRNA	LN995681
A	W33	R-52013	bad-5-M	*Mycobacterium peregrinum*	16S rRNA	LN995692
A	W33	R-52013	bad-5-M	*Mycobacterium peregrinum*	*rpoB*	LN995670
A	W2	R-51031	W2-18	*Pandoraea sputorum*	16S rRNA	LN995684
A	W2	R-51036	W2-18	*Pandoraea sputorum*	16S rRNA	LN995687
A	W33	R-52012	W33-14	*Pseudomonas mosselii*	16S rRNA	LN995691
A	W33	R-50665	W33-23	*Pseudomonas nitroreducens*	16S rRNA	LN995677
A	W2	R-54719	W2-21	*Pseudomonas nitroreducens*	16S rRNA	LN995712
A	W2	R-51027	W2-11	*Rhizobium lusitanum*	16S rRNA	LN995682
A	W2	R-51033	W2-30	*Sphingobacterium multivorum*	16S rRNA	LN995685
A	W2	R-51034	W2-30	*Sphingobacterium multivorum*	16S rRNA	LN995686
A	W33	R-50669	W33-26	*Sphingobacterium siyangense*	16S rRNA	LN995679
A	W2	R-51038	W2-29	*Stenotrophomonas maltophilia*	*rpoD*	LN995623
A	W33	R-51045	W33-25	*Stenotrophomonas maltophilia*	*rpoD*	LN995624
B	S35	R-52624	S35-2	*Burkholderia caledonica*	*recA*	LN995657
B	S35	R-52625	S35-4	*Burkholderia caledonica*	*recA*	LN995658
B	S35	R-52635	S35-5	*Burkholderia caledonica*	*recA*	LN995667
B	S32	R-52598	S32-10	*Burkholderia choica*	*recA*	LN995635
B	S32	R-52600	S32-9	*Burkholderia choica*	*recA*	LN995637
B	S32	R-52613	S32-12	*Burkholderia choica*	*recA*	LN995648
B	S32	R-52614	S32-8	*Burkholderia choica*	*recA*	LN995649
B	S32	R-52615	S32-11	*Burkholderia choica*	*recA*	LN995650
B	S32	R-52619	bad-6-B1	*Burkholderia choica*	*recA*	LN995653
B	S35	R-52638	bad-6-B3	*Burkholderia choica*	*recA*	LN995668
B	S32	R-52616	S32-13	*Burkholderia cordobensis*	*recA*	LN995651
B	S32	R-52610	S32-7	*Burkholderia glathei*	*recA*	LN995645
B	S35	R-52628	S35-14	*Burkholderia glathei*	*recA*	LN995660
B	S35	R-52632	S35-12	*Burkholderia graminis*	*recA*	LN995664
B	S35	R-52630	S35-15	*Burkholderia humi*	*recA*	LN995662
B	S32	R-52599	S32-6	*Burkholderia jiangsuensis*	*recA*	LN995636
B	S32	R-52608	S32-46	*Burkholderia jiangsuensis*	*recA*	LN995643

Continued on next page

Continued from previous page

Experiment	Sample	Isolate	MALDI cluster	Identification	Gene	Accession
B	S32	R-52611	S32-2	*Burkholderia jiangsuensis*	*recA*	LN995646
B	S35	R-52623	S35-8	*Burkholderia jiangsuensis*	*recA*	LN995656
B	S35	R-52629	S35-1a	*Burkholderia phytofirmans*	*recA*	LN995661
B	S32	R-52602	S32-53	*Burkholderia* sp.	*recA*	LN995638
B	S32	R-52603	S32-24	*Burkholderia* sp.	*recA*	LN995639
B	S32	R-52605	S32-51	*Burkholderia* sp.	*recA*	LN995640
B	S32	R-52606	S32-25	*Burkholderia* sp.	*recA*	LN995641
B	S32	R-52607	S32-15	*Burkholderia* sp.	*recA*	LN995642
B	S32	R-52609	S32-1	*Burkholderia* sp.	*recA*	LN995644
B	S32	R-52612	S32-3	*Burkholderia* sp.	*recA*	LN995647
B	S32	R-52617	S32-52	*Burkholderia* sp.	*recA*	LN995652
B	S32	R-52621	bad-6-B2	*Burkholderia* sp.	*recA*	LN995654
B	S32	R-52622	S35-10	*Burkholderia* sp.	*recA*	LN995655
B	S35	R-52627	S35-16	*Burkholderia* sp.	*recA*	LN995659
B	S35	R-52631	S35-7	*Burkholderia* sp.	*recA*	LN995663
B	S35	R-52633	S35-6	*Burkholderia* sp.	*recA*	LN995665
B	S35	R-52634	S35-13	*Burkholderia* sp.	*recA*	LN995666
B	S35	R-52639	S35-11	*Burkholderia* sp.	*recA*	LN995669
B	S32	R-52618	S32-33	*Chryseobacterium lactis*	16S rRNA	LN995695
B	S32	R-52604	S32-30	*Mycobacterium septicum*	16S rRNA	LN995694
B	S32	R-52750	NA	*Mycobacterium septicum*	16S rRNA	LN995699
B	S32	R-52752	NA	*Mycobacterium septicum*	16S rRNA	LN995701
B	S32	R-52604	S32-30	*Mycobacterium septicum*	*rpoB*	LN995671
B	S32	R-52750	NA	*Mycobacterium septicum*	*rpoB*	LN995672
B	S32	R-52752	NA	*Mycobacterium septicum*	*rpoB*	LN995674
B	S32	R-52751	NA	*Mycobacterium smegmatis*	16S rRNA	LN995700
B	S32	R-52751	NA	*Mycobacterium smegmatis*	*rpoB*	LN995673
B	S32	R-52601	S32-41	*Pseudomonas jessenii*	16S rRNA	LN995693
B	S35	R-52636	S35-32	*Pseudomonas jessenii*	16S rRNA	LN995697
B	S35	R-52626	S35-21	*Pseudomonas migulae*	16S rRNA	LN995696
B	S32	R-52643	S32-31	*Rhizobium metallidurans*	16S rRNA	LN995698
C	W2	R-53748	W2-65	*Acinetobacter calcoaceticus*	16S rRNA	LN995703
C	W2	R-54350	W2-46	*Chryseobacterium rhizoplanae*	16S rRNA	LN995706
C	W2	R-54353	W2-53	*Comamonas* sp.	16S rRNA	LN995708
C	W2	R-54356	W2-60	*Delftia lacustris*	16S rRNA	LN995711
C	W2	R-53747	W2-47	*Inquilinus limosus*	16S rRNA	LN995702
C	W2	R-54347	W2-52	*Mycobacterium chelonae*	16S rRNA	LN995704
C	W2	R-54347	W2-52	*Mycobacterium chelonae*	*rpoB*	LN995675
C	W2	R-54348	W2-66	*Pseudomonas* sp.	16S rRNA	LN995705
C	W2	R-54352	W2-54	*Rhizobium radiobacter*	16S rRNA	LN995707
C	W2	R-54354	W2-57	*Sphingobacterium detergens*	16S rRNA	LN995709
C	W2	R-54355	W2-59	*Sphingobacterium multivorum*	16S rRNA	LN995710
C	W2	R-54349	W2-55	*Stenotrophomonas maltophilia*	*rpoD*	LN995626
C	W2	R-54351	W2-48	*Stenotrophomonas maltophilia*	*rpoD*	LN995627
C	W2	R-53749	W2-49	*Stenotrophomonas rhizophila*	*rpoD*	LN995625
D	W152	R-54734	W152-25	*Delftia lacustris*	16S rRNA	LN995716
D	W152	R-54733	W152-22	*Elizabethkingia miricola*	16S rRNA	LN995715
D	W2	R-54726	W2-71	*Mycobacterium chelonae*	16S rRNA	LN995714

Continued on next page

Continued from previous page

Experiment	Sample	Isolate	MALDI cluster	Identification	Gene	Accession
D	W2	R-54726	W2-71	*Mycobacterium chelonae*	*rpoB*	LN995676
D	S35	R-54757	S35-47	*Pseudomonas kilonensis*	16S rRNA	LN995719
D	W152	R-54746	W152-23	*Pseudomonas protegens*	16S rRNA	LN995718
D	S35	R-54758	S35-46	*Pseudomonas protegens*	16S rRNA	LN995720
D	W2	R-54724	W2-70	*Pseudomonas soli*	16S rRNA	LN995713
D	W152	R-54745	W152-21	*Sphingobacterium faecium*	16S rRNA	LN995717
D	W2	R-54722	W2-73	*Stenotrophomonas maltophilia*	*rpoD*	LN995628
D	W152	R-54741	W152-32	*Stenotrophomonas maltophilia*	*rpoD*	LN995629
D	S35	R-54756	S35-40	*Stenotrophomonas rhizophila*	*rpoD*	LN995630

3.3 The quest for environmental *Burkholderia multivorans* isolates - unpublished isolation experiments

3.3.1 Introduction

This section presents the results of unpublished isolation experiments and serves as a complement to Section 3.2. Experiment 4 presents the optimization of BCEM with respect to C-source composition and pH. Because the traditional approaches, i.e direct plating and liquid enrichment of water samples and direct plating of soil samples (Section 3.2), did not yield *B. multivorans* isolates from PCR positive samples, we applied multiple cultivation strategies in a stepwise manner. Experiment 7 and 8 examine the effect of lowering the pH, extending the incubation time and shaking the enrichment broth. Additionally, we applied several strategies to enable the cultivation of fastidious bacteria (Vartoukian *et al.*, 2010; Epstein, 2013), including acclimatizing recovery in broth medium (experiment 8), resuscitation at room temperature (experiments 11 and 12), use of diluted nutrient media (experiments 11 and 12) and co-cultivation with amoebae (experiment 13) and helper strain panels (experiment 16). Finally, we used germinating seedlings to "fish" *B. multivorans* out of environmental samples. A discussion of the results in this section is presented in Section 6.2 of the general discussion.

3.3.2 Samples

Fresh samples were taken for all experiments (Table 3.10) as described previously (Section 3.2.3.1).

3.3.3 Experiment 4: comparison of different BCEM formulations with PCAT for liquid enrichment and plating of water samples

3.3.3.1 Setup and methods

The present experiment compared different formulations of BCEM (with respect to C-sources, pH and buffer, see Table 3.11) with PCAT and examined the effect of different agar concentrations (1.2 % versus 1.8 %). Media C and D were based on the EB1 and EB2 formulations of Vermis *et al.* (2003a) and contained only two C-sources.
Two fresh water samples W2 and W33 (Table 3.10) that were PCR positive for *B. multivorans* were filtered as described previously (Section 3.2.3.3, experiment A). One filter was incubated in 25 ml of each broth (Table 3.11) while shaking. After 5 days of incubation at 28 °C, 2 ml broth samples were taken for pH measurement, total DNA extraction and plating of

Table 3.10: Sample overview

Experiment	Sample	Sampling date	Sampling source[a]	pH	Temperature[b]
4	W2	20/03/2013	Freshwater greenhouse	7.83	23 °C
4	W33	20/03/2013	Pond greenhouse	7.73	19 °C
7	W2	28/02/2014	Freshwater greenhouse	8.12	22 °C
7	W152	28/02/2014	Pond water	7.95	6 °C
8	W2	16/04/2014	Freshwater greenhouse	8.19	22 °C
9, 11, 12	W2	07/05/2014	Freshwater greenhouse	8.14	22 °C
9, 11, 12	S35	07/05/2014	Soil	6.17	-
13	W2	26/08/2014	Freshwater greenhouse	8.42	24 °C
13	W152	26/08/2014	Pond water	7.75	18 °C
16	W2	07/10/2014	Freshwater greenhouse	8.50	25 °C
16	W152	07/10/2014	Pond water	7.33	15 °C
16	S35	07/10/2014	Soil	5.57	-

[a]All samples were taken in the botanical garden of Ghent University, Ghent, Belgium. [b]For soil samples no temperature was recorded.

aliquots (50 µl) of tenfold serial dilutions in physiological saline (1E-0, 1E-3 and 1E-5) on the respective agar media (Table 3.11) with 1.2 % or 1.8 % agar. After 3 days of incubation at 28 °C, total cell mass was harvested from each agar plate for DNA extraction following the Pitcher protocol (Pitcher *et al.*, 1989). The presence of *B. multivorans* and the relative abundance of *Burkholderia* bacteria in each sample were assessed using a *B. multivorans*-specific *recA*-based PCR assay (Section 3.1.3.4) and a *Burkholderia* qPCR, respectively. In case the *B. multivorans*-specific PCR assay yielded visible amplicons, these were sequenced to exclude false positive results (Section 3.1.3.4).

Table 3.11: Isolation media experiment 4

Medium	C- and N-sources	pH	Reference
A	L-arabinose, D-cellobiose and L-threonine	5.5 - MES buffer	BCEM (Section 3.2.3.2)
B	L-arabinose, D-cellobiose and L-threonine	Not adjusted (±6.5)	-
C	L-arabinose and D-cellobiose	Not adjusted (±6.5)	EB1 (Vermis *et al.*, 2003a)
D	L-arabinose and L-threonine	Not adjusted (±6.5)	EB2 (Vermis *et al.*, 2003a) (Vanlaere *et al.*, 2005)
E	Azelaic acid and tryptamine	Not adjusted (5.8)	PCAT (Section 3.2.3.2)

Medium A, B, C and D contained the same selective agents as BCEM (Section 3.2.3.2).

Quantitative real-time PCR (qPCR) was performed on a LightCycler 480 II (Roche) using FrameStar 480 plates (BioKe 4TI-0951). Each reaction (20 µl) was composed of 1x SensiMix SYBR No-ROX (Bioline), 0.15 µM (universal 16S rRNA assay) or 0.7 µM (*Burkholderia* 16S rRNA assay) of each primer, 2 µl of template DNA and MQ water. Per sample, reactions were run in triplicate and fluorescence was detected at the end of each cycle. Specificity of the PCR products was evaluated by melting curve analysis. Universal primers Eub338

and Eub518 (Fierer *et al.*, 2005) targeted a 200 bp fragment of the 16S rRNA gene and were used to measure total bacterial abundance. PCR cycling conditions for this primer set were 10 min at 95 °C, followed by 45 cycles of 10 s at 95 °C, 30 s at 60 °C and 1 s at 72 °C. *Burkholderia*-specific primers BKH812F and BKH1249R (Bergmark *et al.*, 2012) targeted a 438 bp fragment of the 16S rRNA gene and were used to measure total *Burkholderia* abundance. PCR cycling conditions for this primer set were based on the protocol of Stopnisek *et al.* (2014) and were 10 min at 95 °C, followed by 45 cycles of 30 s at 95 °C, 30 s at 64 °C and 45 s at 72 °C. External standard curves (R^2>0.99) were prepared from tenfold serial dilutions of *B. multivorans* R-20526 with known concentrations. Data were presented as the ratio of *Burkholderia* bacteria relative to the total number of bacteria.

3.3.3.2 Results

Growth was observed for both water samples on all agar media until dilution 1E-5. Overall, similar numbers of colonies were observed for agar media with 1.2 % or 1.8 % agar, except for medium A for which more growth was observed on plates with 1.2 % agar. Colonies were generally larger on agar media with 1.2 % agar. The pH of the enrichment broth after 5 days of incubation at 28 °C and the presence of *B. multivorans* in total cell mass from the enrichment broths and agar plates is presented in Table 3.12. The MES buffer in medium A (50 mM) was effective in keeping the pH of the enrichment broth at 5.5 for both samples. Only for media A and E both the broth and agar medium were PCR positive for *B. multivorans*. Water sample W33 yielded a Bcc-specific amplicon on medium B (detected only in the broth) and D (detected both in the broth and on the agar plates). The relative abundance of *Burkholderia* bacteria in the total cell mass from broth and agar media is presented in Fig. 3.1. Only for water sample W2 in medium A an increase in relative abundance of *Burkholderia* bacteria was recorded.

3.3.3.3 Conclusions

Overall, similar numbers of colonies were observed for agar media with 1.2 % and 1.8 % agar after liquid enrichment and plating of water samples. *B. multivorans* was detected in the total cell mass from both the enrichment broth and agar plates only when BCEM with three C-sources at pH 5.5 or PCAT medium was used. Although inconsistent qPCR results were recorded for the two water samples, BCEM with three C-sources at pH 5.5 was generally better than PCAT in increasing the relative abundance of *Burkholderia* bacteria. In conclusion, the optimized formulation of BCEM included three C-sources (i.e. L-arabinose, D-cellobiose and L-threonine) and MES buffer at pH 5.5.

Table 3.12: Presence of *B. multivorans* in BCEM and PCAT enrichment broths and on BCEM and PCAT agar media (experiment 4)

Medium	W2			W33		
	broth	**1.2 % agar**	**1.8 % agar**	**broth**	**1.2 % agar**	**1.8 % agar**
A	pH 5.45			pH 5.45		
	Bcc: +	Bcc: +	Bcc: −	Bcc: +	Bcc: +	Bcc: +
	Bm: +	Bm: +	Bm: −	Bm: +	Bm: +	Bm: +
B	pH 5.70			pH 6.42		
	Bcc: +	Bcc: −	Bcc: −	Bcc: +[a]	Bcc: +	Bcc: +
	Bm: +	Bm: −	Bm: −	Bm: −	Bm: +	Bm: +
C	pH 5.66			pH 5.65		
	Bcc: +	Bcc: −	Bcc: −	Bcc: +	Bcc: −	Bcc: w
	Bm: +	Bm: −	Bm: −	Bm: +	Bm: −	Bm: −
D	pH 5.95			pH 6.15		
	Bcc: +	Bcc: +	Bcc: +	Bcc: +[b]	Bcc: +[c]	Bcc: +[c]
	Bm: +	Bm: +	Bm: +	Bm: −	Bm: −	Bm: −
E	pH 5.86			pH 5.88		
	Bcc: +	Bcc: +	Bcc: +	Bcc: +	Bcc: +	Bcc: +
	Bm: +	Bm: +	Bm: +	Bm: +	Bm: w	Bm: +

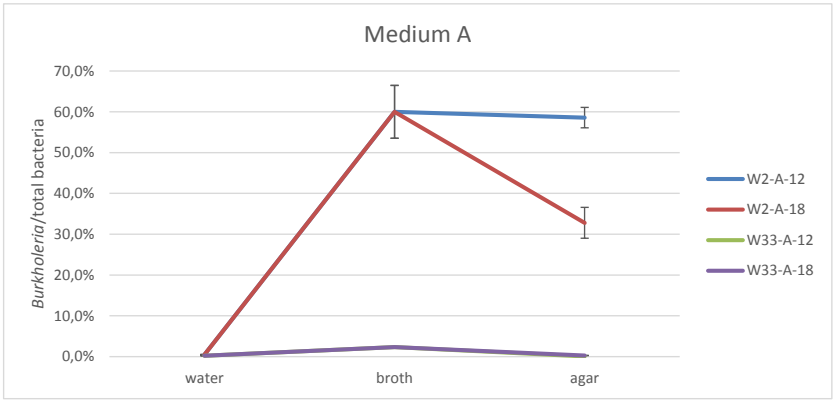
W2 and W33, water samples. Bcc, Bcc-specific amplicon in the first round of the PCR assay; Bm, *B. multivorans*-specific amplicon in the second round of the PCR assay; +, PCR positive; −, PCR negative; w, weak amplicon. [a]96 % similarity with *B. arboris recA* sequences. [b]99 % similarity with *B. arboris* and *B. lata recA* sequences. [c]97 % similarity with Bcc *recA* sequences.

3.3.4 Experiment 7: effect of pH and incubation time on liquid enrichment of water samples

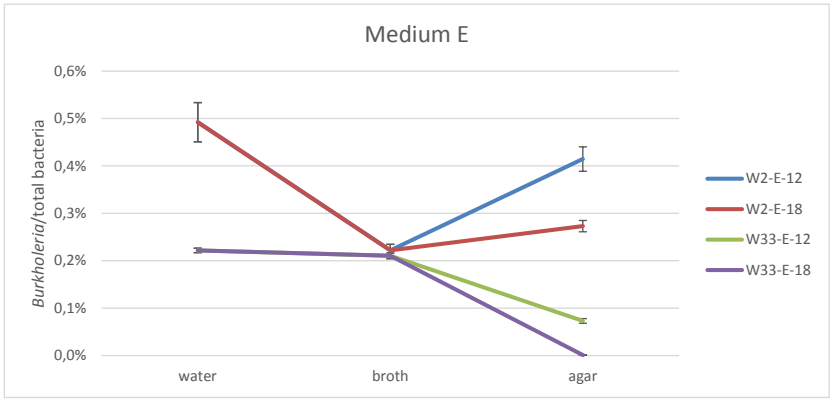
3.3.4.1 Setup and methods

The present experiment examined the effect of the pH and incubation time of the enrichment broth.

Two fresh water samples W2 and W152 (Table 3.10) that were PCR positive for *B. multivorans* were filtered as described previously (Section 3.2.3.3, experiment A). One liter of filtrate from water samples W2 and W152 was inoculated with a loopful (10 µl) of cells of *B. multivorans* R-20825 (grown for 48 h at 28 °C on buffered nutrient agar) and subsequently filtered as a positive control (PC). Filtrate from water samples W2 and W152 was used as negative control (NC). Per condition, four filters were cut into smaller pieces and incubated in 100 ml of liquid BCEM (as described in section 3.2.3.2) at pH 4.5 or pH 5.5 (with 50 mM MES buffer). After 3, 5, 7 and 10 days of incubation at 28 °C, 10 ml broth samples were taken for pH measurement and total DNA extraction following the Pitcher protocol (Pitcher *et al.*, 1989). The presence of *B. multivorans* and the relative abundance of *Burkholderia* bacteria in each sample were assessed using a *B. multivorans*-specific *recA*-based PCR assay (Section



(a) BCEM



(b) PCAT

Figure 3.1: Relative abundance of *Burkholderia* bacteria in enrichment broth and on agar media of **(a)** BCEM and **(b)** PCAT (experiment 4). W2 and W33, water samples. A, medium A; E, medium E. 12, 1.2 % agar; 18; 1.8 % agar.

3.1.3.4) and a *Burkholderia* qPCR, respectively. In case the *B. multivorans*-specific PCR assay yielded visible amplicons, these were sequenced to exclude false positive results (Section 3.1.3.4).

3.3.4.2 Results

The presence of *B. multivorans* and relative abundance of *Burkholderia* bacteria in total cell mass from the BCEM broth with pH 4.5 or 5.5 at different time points is presented in Table 3.13 and Fig. 3.2, respectively. All PC samples were PCR positive for *B. multivorans*, while all NC samples that were taken after 10 days of incubation were PCR negative for *B. multivorans*. After 10 days of incubation at 28 °C, sample W2 in BCEM broth with pH 4.5 showed an increase in relative abundance of *Burkholderia* bacteria. Contrastingly, sample

W152 incubated in BCEM broth with pH 4.5 showed a decrease in relative abundance of *Burkholderia* bacteria at day 3. In BCEM broth with pH 5.5, both samples showed an enrichment after 3 days, again followed by a decline of the relative number of *Burkholderia* bacteria (Fig. 3.2). A similar trend was observed for the presence of *B. multivorans*, in which all broths were initially PCR positive, turned negative at day 5, and turned positive again at days 7 and 10 (only W2) (Table 3.13).

Table 3.13: Presence of *B. multivorans* in BCEM broth at pH 4.5 or 5.5 at different time points (experiment 7)

Sample	pH BCEM	Day 3	Day 5	Day 7	Day 10
W2	4.5	−	−	w	+
	5.5	w	−	w	w
W152	4.5	w	−	w	−
	5.5	+	−	−	−

W2 and W152, water samples. +, PCR positive for *B. multivorans*; −, PCR negative for *B. multivorans*; w, weak amplicon for *B. multivorans*.

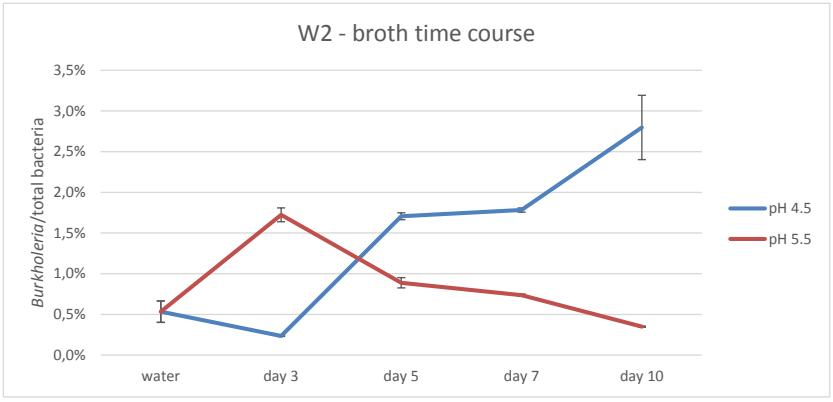
3.3.4.3 Conclusions

Although an increase was observed in the relative abundance of *Burkholderia* bacteria in BCEM broth with pH 5.5 after 3 days of incubation, their abundance again dropped at the later time points. A similar trend was observed for the presence of *B. multivorans*. Lowering the pH of the enrichment broth to 4.5 and increasing the incubation time yielded varying results for different samples but generally did not enhance the liquid enrichment of *Burkholderia* bacteria from water samples.

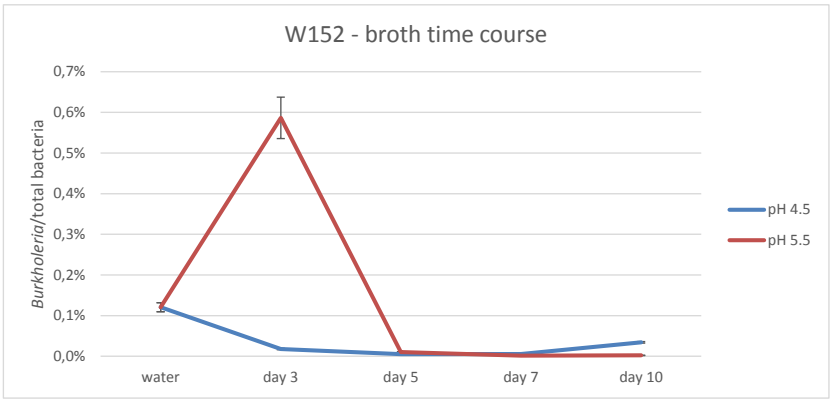
3.3.5 Experiment 8: effect of shaking and acclimatizing recovery on liquid enrichment of water samples

3.3.5.1 Setup and methods

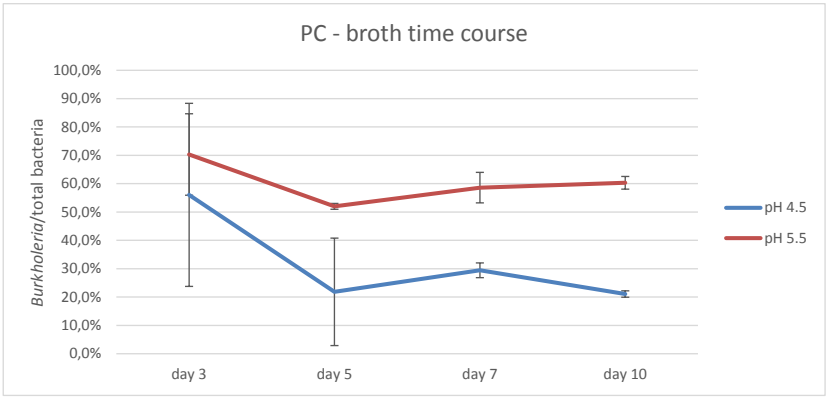
The present experiment examined the effect of the pH and shaking of the enrichment broth and tested acclimatizing recovery. The latter strategy allows a gradual increase in nutrients and selective agents and the inclusion of nutrients and signals from the original environment (Vartoukian *et al.*, 2010; Epstein, 2013).
One fresh water sample W2 (Table 3.10) that was PCR positive for *B. multivorans* was filtered as described previously (Section 3.2.3.3, experiment A). Per condition, two filters were cut into smaller pieces and incubated in 50 ml of liquid BCEM (as described in section 3.2.3.2) at pH 4.5 or pH 5.5 (with 50 mM MES buffer). Per pH condition, one bottle was



(a) W2



(b) W152



(c) *B. multivorans* R-20825

Figure 3.2: Relative abundance of *Burkholderia* bacteria in BCEM broth at pH 4.5 and 5.5 at different time points (experiment 7). W2 and W152, water samples; PC, positive control.

incubated while shaking, while the other bottle was incubated while not shaking. After 1, 2, 3 and 4 weeks of incubation at 28 °C, 10 ml broth samples were taken for pH measurement and total DNA extraction following the Pitcher protocol (Pitcher *et al.*, 1989).
For the acclimatizing recovery experiment, two filters were incubated in 50 ml of broth which comprised a mixture of filtrate from the original water sample and BCEM at pH 4.5 or 5.5. The ratio of BCEM:filtrate was increased weekly by 20 % of the total volume (Table 3.14). The broths were therefore divided into two aliquots of 25 ml that were centrifuged for 10 min at 8000 rpm. The supernatant was removed and used to measure the pH of the enrichment broth. One pellet was stored at −20 °C for total DNA extraction, while the other pellet was used to inoculate the next batch of 50 ml BCEM/filtrate broth. After 1, 2, 3, 4, 5 and 6 weeks of incubation at 28 °C, 10 ml broth samples were taken for pH measurement and total DNA extraction following the Pitcher protocol (Pitcher *et al.*, 1989). The presence of *B. multivorans* and the relative abundance of *Burkholderia* bacteria in each sample were assessed using a *B. multivorans*-specific *recA*-based PCR assay (Section 3.1.3.4) and a *Burkholderia* qPCR, respectively. In case the *B. multivorans*-specific PCR assay yielded visible amplicons, these were sequenced to exclude false positive results (Section 3.1.3.4).

Table 3.14: Increasing ratio of BCEM:filtrate in liquid enrichment broth

	Ratio BCEM:filtrate	BCEM (ml)	Filtrate (ml)
Week 1	0:5	0	50
Week 2	1:4	10	20
Week 3	2:3	20	30
Week 4	3:2	30	20
Week 5	4:1	40	10
Week 6	5:0	50	0

3.3.5.2 Results

The presence of *B. multivorans* and relative abundance of *Burkholderia* bacteria in total cell mass from the different BCEM broths at different time points is presented in Table 3.15 and Fig. 3.3, respectively. Both for the shaking and acclimatizing recovery experiment a similar trend was found in which the broths were PCR positive for *B. multivorans* after 1 week of incubation. The broths of week 2 and 3 are PCR negative, while after 4 weeks of incubation all samples were again PCR positive for *B. multivorans*. Because the overall trend for the presence of *B. multivorans* was the same in the shaking experiment for pH 4.5 and 5.5, the qPCR was performed only for the acclimatizing recovery experiment on the broths with pH 4.5. The qPCR results showed that shaking of the enrichment broth led to a faster increase in

relative abundance of *Burkholderia* bacteria. The acclimatizing recovery experiment yielded different results dependent on the pH of the enrichment broth (Fig. 3.3).

Table 3.15: Presence of *B. multivorans* in BCEM broth to assess the effect of shaking and acclimatizing recovery (experiment 8)

Condition	pH BCEM	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
Shaking	4.5	+	−	−	+	nd	nd
	5.5	+	−	−	+	nd	nd
Non-shaking	4.5	+	−	−	+	nd	nd
	5.5	−	−	−	+	nd	nd
Acclimatizing recovery	4.5	+	−	−	+	+	+
	5.5	+	−	−	+	+	+

+, PCR positive for *B. multivorans*; −, PCR negative for *B. multivorans*; nd, not determined.

3.3.5.3 Conclusions

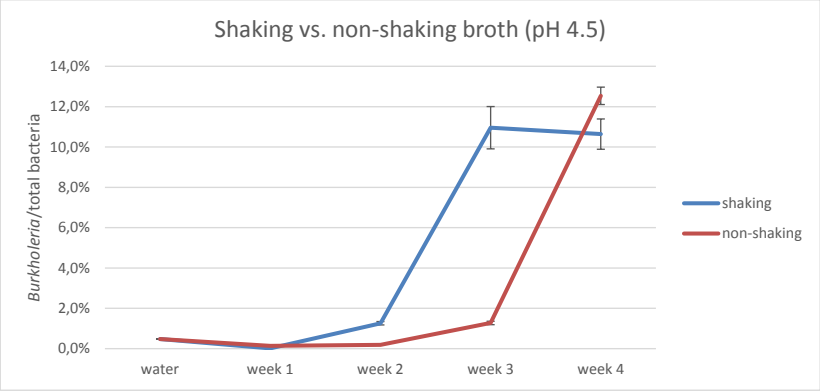
Shaking the enrichment broth resulted in a faster increase of the relative abundance of *Burkholderia* bacteria. The acclimatizing recovery experiment showed the highest relative abundance of *Burkholderia* bacteria after one week (100 % filtrate) and 6 weeks (100 % BCEM broth) of incubation. For both the shaking and acclimatizing recovery experiment, the broths were generally PCR positive for *B. multivorans* after 1 week of incubation, turned PCR negative in week 2 and 3, and PCR positive again after 4 weeks of incubation.

3.3.6 Experiment 9: "fishing" with seedlings

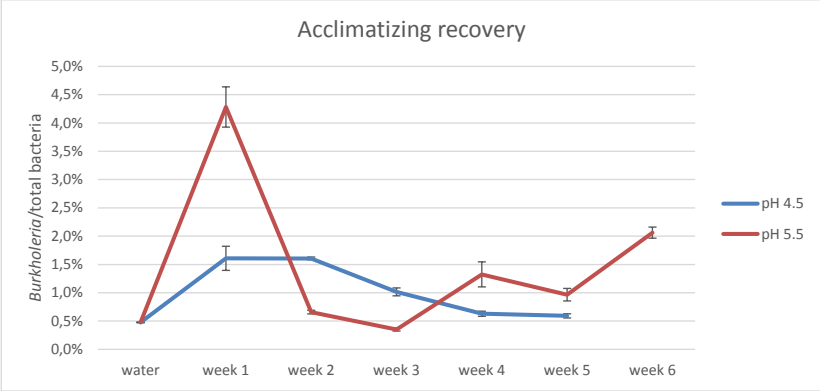
3.3.6.1 Setup and methods

The present experiment examined whether the adhesion of *Burkholderia* bacteria to plant roots could be used as a "fishing" strategy to isolate *B. multivorans* from water and soil samples.

One fresh water sample W2 and one fresh soil sample S35 (Table 3.10) that were PCR positive for *B. multivorans* were obtained as described previously (Section 3.2.3.1). For water sample W2, two times 150 ml of the water sample was transferred to an autoclaved glass petri plate. For soil sample S35, two times 15 g of the soil sample was transferred to an autoclaved glass petri plate and suspended in 135 ml of autoclaved 0.85 % NaCl solution. A half loop of cells of *B. multivorans* R-20526 and R-20825 each (grown for 24 h at 37 °C on buffered nutrient agar) was resuspended in 9 ml of 0.85 % NaCl solution and two ml of this solution was added to one of the two plates with the water and soil sample to serve as a positive control (W2-PC and S35-PC). An autoclaved glass petri plate with 150 ml of 0.85 %



(a) Shaking of the enrichment broth



(b) Acclimatizing recovery

Figure 3.3: Relative abundance of *Burkholderia* bacteria in BCEM broth to assess the effect of **(a)** shaking and **(b)** acclimatizing recovery (experiment 8).

NaCl solution served as negative control (NC). Seeds of pea (Wonder of Kelvedon), maize (Tasty sweet F1 hybrid), lupine, barley and rice were surface sterilized (by soaking them for 2 min in a 1 % NaOCl solution, 2 min in 70 % ethanol, 2 min in sterile MQ and 4 times 1 min in sterile MQ) and transferred to the glass petri plates containing either water or soil sample. After 9 days of incubation at room temperature on a windowsill, one type of seed from each condition was used for further analysis. The roots of the seedlings were aseptically removed and transferred into 500 µl 0.85 % NaCl solution. Using a surface sterilized micropestle, the roots were crushed for 30 sec and another 500 µl 0.85 % NaCl solution was added. Aliquots (50 µl) of tenfold serial dilutions (1E-0 until 1E-2) in physiological saline were plated on PCAT and BCSA agar media (Section 3.2.3.2). After 6 (W2) or 7 (S35) days of incubation at 28 °C for , 10-15 isolates per sample, seed and agar medium were randomly picked and subcultivated on buffered nutrient agar (Oxoid, pH 6.8). Isolates were identified using *recA*

gene sequence analysis as described previously (Section 3.2.7.1).

3.3.6.2 Results

In W2 and W2-PC, seeds of pea and rice germinated, as well as one maize seed in W2-PC. In S35 and S35-PC, only pea seeds germinated. The other seeds in W2 and S35 were overgrown by fungi. In the NC sample none of the seeds germinated and all seeds were overgrown by fungi. For the pea seed in W2, no isolates were obtained from BCSA and 0/10 isolates from PCAT were identified as *Burkholderia*. For the rice seed in W2, 0/10 isolates from BCSA and 0/8 isolates from PCAT were identified as *Burkholderia*. For the pea seed in S35, 0/2 isolates from BCSA and 4/23 isolates from PCAT were identified as *Burkholderia*. Two isolates were identified as *B. caledonica*, one as *B. phytofirmans* and one as *Burkholderia* sp. The identification and number of isolates for the spiked samples W2-PC and S35-PC is presented in Table 3.16.

Table 3.16: Identification and number of isolates from seedlings germinated in spiked samples W2-PC and S35-PC (experiment 9)

	W2-PC						**S35-PC**	
	Pea		**Rice**		**Maize**		**Pea**	
	PCAT	**BCSA**	**PCAT**	**BCSA**	**PCAT**	**BCSA**	**PCAT**	**BCSA**
B. multivorans PC	5	3	5		4		1	
B. phytofirmans							1	
Total isolates	9	4	8	2	20	9	10	5

PC, *B. multivorans* positive control strains R-20526 and R-20825.

3.3.6.3 Conclusions

B. multivorans was retrieved from the spiked PC samples with pea, rice and maize from W2, and with pea for S35. More isolates of the PC strains were retrieved from PCAT than from BCSA. "Fishing" with seedlings did however not yield *B. multivorans* from the unspiked water and soil samples.

3.3.7 Experiment 11 and 12: effect of resuscitation at room temperature and comparison of PCAT to 1/10 PCAT medium

3.3.7.1 Setup and methods

The present experiment examined the effect of a reduced nutrient concentration of the PCAT isolation medium and the resuscitation of environmental bacteria at room temperature.

One fresh soil sample S35 (Table 3.10) that was PCR positive for *B. multivorans* was homogenized as described previously (Section 3.2.3.3, experiment B). Aliquots (50 µl) of tenfold serial dilutions of the soil suspension in physiological saline (1E-0 until 1E-5) were plated on PCAT and 1/10 PCAT (Table 3.2.3.2). The remainder of the soil sample was left at room temperature for two more weeks (with loosened screw cap to allow oxygen flow) and the plating procedure was repeated after 1 and 2 weeks. Colonies were picked from both media from the last plating after 16 days of incubation at 28 °C. One fresh water sample W2 (Table 3.10) that was PCR positive for *B. multivorans* was left for 16 days at room temperature (with loosened screw cap to allow oxygen flow) prior to direct plating (50 µl) on PCAT and 1/10 PCAT. After 5 days of incubation at 28 °C growth on the agar plates was evaluated.

3.3.7.2 Results

For sample S35, growth was observed until dilutions of 1E-4 or 1E-3 onto 1/10 PCAT medium, whereas PCAT medium only yielded growth up until 1E-2 (Table 3.17). A larger diversity of colony morphologies was observed on 1/10 PCAT medium after incubating the soil sample for 1 or 2 weeks at 28 °C. Of the 35 isolates from 1/10 PCAT medium, 23 were identified as *Burkholderia*, but none of them was *B. multivorans*. Of the 23 isolates from PCAT medium, 19 were identified as *Burkholderia*, but none of them was *B. multivorans*.
Sample W2, which was left for 16 days at room temperature prior to plating, yielded no colonies on PCAT but 25-35 colonies on 1/10 PCAT. None of the 32 colonies that were picked from 1/10 PCAT medium was identified as *Burkholderia*.

Table 3.17: Growth of sample S35 on PCAT and 1/10 PCAT after 0, 1 and 2 weeks of incubation at room temperature (experiment 11)

	Week 0		Week 1		Week 2	
Dilution	**PCAT**	**1/10 PCAT**	**PCAT**	**1/10 PCAT**	**PCAT**	**1/10 PCAT**
1E-0	+	+	+	+	+	+
1E-1	+	+	+	+	+	+
1E-2	+	+	+	+	+	+
1E-3	NG	+	NG	+	NG	+
1E-4	NG	+	NG	+	NG	NG
1E-5	NG	NG	NG	NG	NG	NG

NG, no growth observed.

3.3.7.3 Conclusions

The nutrient diluted (1/10) formulation of PCAT yielded more growth for both water and soil samples, and generally yielded a larger number of distinct colony morphologies. Incubating the water and soil samples for up to two weeks at room temperature prior to plating did not yield *B. multivorans* isolates.

3.3.8 Experiment 13: "fishing" with amoebae

3.3.8.1 Setup and methods

The present experiment examined the effect of co-cultivation with amoebae to resuscitate *B. multivorans* in environmental samples (Steinert *et al.*, 1997; Marolda *et al.*, 1999; La Scola *et al.*, 2001; Lamothe *et al.*, 2004; Pagnier *et al.*, 2008). For fresh water samples W2 and W152 (Table 3.10), 150 ml aliquots were filtered as described previously (Section 3.2.3.3, experiment A) and two filters were shaken for 2 h at 28 °C in 25 ml PAS medium (Chavatte *et al.*, 2014). One ml of the concentrated water sample in PAS medium was added to a single well of a 12-well plate with a monolayer of *Acanthamoeba castellanii* in duplicate. PAS medium was used as negative control (NC). After 1 (D1), 3 (D3) and 7 (D7) days of incubation at 25 °C aliquots (100 µl for D1 and 300 µl for D3 and D7) of tenfold serial dilutions were plated on 1/10 PCAT and BCSA agar media (Section 3.2.3.2). After 6 days of incubation at 28 °C colonies were randomly picked and subcultivated on buffered nutrient agar (Oxoid, pH 6.8). Dereplication and identification of the isolates was performed as described previously (Section 3.2.3.4).

3.3.8.2 Results

MALDI-TOF MS dereplication of 176 isolates resulted in a total of 26 clusters, of which none were identified as *Burkholderia*, 16 clusters (146 isolates) were identified as other GNNF bacteria and 3 clusters (4 isolates) as *Mycobacterium* species (Table 3.18). The remaining 7 clusters (26 isolates) were identified as Gram-positive or Gram-negative fermenting bacteria (data not shown).

3.3.8.3 Conclusions

Co-cultivation of water samples with amoebae (*A. castellanii*) prior to plating did not yield *B. multivorans* isolates.

Table 3.18: Identification and number of isolates for co-cultivation with amoebae (experiment 13)

	W2		W152	
	1/10 PCAT	**BCSA**	**1/10 PCAT**	**BCSA**
Acinetobacter calcoaceticus			1	
Chryseobacterium rhizoplanae	26	3		
Flavobacterium oceanosedimentum	3			
Flavobacterium sp.			1	
Neorhizobium alkalisoli	1			
Pseudomonas sp.			49(2)	25
Rhizobium radiobacter	8(2)			
Rhizobium taibaishanense			1	
Sphingobacterium cladoniae			11	12
Sphingomonas sanxanigenens	4			
Variovorax soli			1	
Total GNNF	42	3	64	37
Mycobacterium abscessus		2		
Mycobacterium chelonae		1		
Mycobacterium sp.		1		
Total isolates	46	7	84	39

Each cell represents the number of isolates per MALDI-TOF MS cluster. The number of clusters is given in between brackets if more than one. W2 and W152, water samples W2 and W152 (Table 3.10).

3.3.9 Experiment 16: co-cultivation with helper strains

3.3.9.1 Setup and methods

The present experiment examined the effect of co-cultivation with helper strains that can provide nutrients and/or signals which may resuscitate *B. multivorans* in environmental samples (Vartoukian *et al.*, 2010; Epstein, 2013). For fresh water samples W2 and W152 (Table 3.10), 300 ml was filtered as described previously (Section 3.2.3.3, experiment A) and each filter was shaken in 25 ml of its own filtrate for 2 h at 28 °C. For fresh soil sample S35 (Table 3.10), 2.5 g of soil was shaken in 22.5 ml physiological saline for 2 h at 28 °C. Four ml of the concentrated water samples and 4 ml of the 2x diluted soil suspension in physiological saline were added to a single well of a Greiner 6-well plate. Four ml of physiological saline was used as negative control (NC). ThinCert membranes (Greiner 657 640, pore size 0.4 µm) were placed into each well and 1 ml of helper strain cell suspension was added to the ThinCert compartment. Each cell suspension contained 3-5 strains of *B. multivorans* (Bm), *Pseudomonas* sp. (P) or *Staphylococcus* sp. (S) (Table 3.19). Aliquots (50 µl) of tenfold serial dilutions in physiological saline were plated on 1/10 PCAT and BCSA agar media (Section 3.2.3.2) after 1, (D1), 3 (D3) and 7 (D7) days of incubation at 28 °C. After 6 (D1 and D3) or 7 days (D7) of incubation at 28 °C, colonies were randomly picked and subcultivated on buffered nutrient agar (Oxoid, pH 6.8). Dereplication and identification of

the isolates was performed as described previously (Section 3.2.3.4).

Table 3.19: Helper strain panels

Strain panel	Strain	Source (ST)
Bm	R-20526	ENV (ST-836)
	R-20825	ENV (ST-650)
	R-53445	CF (ST-25)
	R-53446	CF (ST-741)
	R-53627	CF (ST-16)
P	R-31520	CF
	R-52232	CF
	R-52542	CF
	R-53162	CF
S	R-19214	CF
	R-19219	CF
	R-49429 t1	CF

Bm, *B. multivorans*; P, *Pseudomonas* sp.; S, *Staphylococcus* sp.; CF, cystic fibrosis; ENV, environmental; ST, multilocus sequence type.

3.3.9.2 Results

For sample W2, MALDI-TOF MS dereplication of 279 isolates resulted in a total of 46 clusters, of which 35 clusters (259 isolates) were identified as GNNF bacteria and two clusters (3 isolates) as *M. chelonae* (Table 3.20). The remaining 9 clusters (17 isolates) were identified as Gram-positive or Gram-negative fermenting bacteria (data not shown).
For sample W152, MALDI-TOF MS dereplication of 236 isolates resulted in a total of 47 clusters, of which 34 clusters (116 isolates) were identified as GNNF bacteria (Table 3.21). The remaining 13 clusters (120 isolates) were identified as Gram-positive or Gram-negative fermenting bacteria (data not shown).
For sample S35, MALDI-TOF MS dereplication of 314 isolates resulted in a total of 30 clusters, of which 24 clusters (291 isolates) were identified as GNNF bacteria (Table 3.22). The remaining 6 clusters (23 isolates) were identified as Gram-positive or Gram-negative fermenting bacteria (data not shown).
For the NC, growth was observed at all time points on both the 1/10 PCAT and BCSA agar media, with all colonies showing the same colony morphology. Several colonies were picked at each time point and from each agar medium (17 isolates in total) and all of them were identified as *B. multivorans*. MALDI-TOF MS analysis suggested that all *B. multivorans* isolates that were obtained in this experiment (from the wells with NC, W2, W152 and S35 sample) belonged to the Bm strain panel (Table 3.19) and were thus contaminants. These findings were confirmed by RAPD typing and *gyrB* sequence analysis (data not shown).

Table 3.20: Identification and number of isolates for sample W2 (experiment 16)

	Bm		P		S	
	1/10 PCAT	**BCSA**	**1/10 PCAT**	**BCSA**	**1/10 PCAT**	**BCSA**
Aminobacter sp.	1	42		9		6
Bosea lupini						1
Bosea vestrisii						2
Chromobacterium aquaticum		4	1	5		21
Chryseobacterium sp.		2	1	3		
Novosphingobium resinovorum	1		1			
Pandoraea pnomenusa				1		3
Pseudomonas extremaustralis					1	
Pseudomonas nitroreducens	28		6		6	
Pseudomonas soli	1		1		1	
Pseudomonas taiwanensis			2		4	
Rhizobium radiobacter			2		1	
Sphingobium rhizovicinum	29		22		20	
Stenotrophomonas maltophilia			3			
Bm contaminants		15				
Total GNNF	60	63	39	18	33	33
Mycobacterium chelonae				1		2
Total isolates	62	63	43	20	33	35

Each cell represents the number of isolates per MALDI-TOF MS cluster. The number of clusters is given in between brackets if more than one. Bm, *B. multivorans*; P, *Pseudomonas* sp.; S, *Staphylococcus* sp. (Table 3.19).

3.3.9.3 Conclusions

Co-cultivation of water and soil samples with helper strains prior to plating did not yield *B. multivorans* isolates.

Table 3.21: Identification and number of isolates for sample W152 (experiment 16)

	Bm		P		S	
	1/10 PCAT	**BCSA**	**1/10 PCAT**	**BCSA**	**1/10 PCAT**	**BCSA**
Brevundimonas sp.	1					
Chryseobacterium aquaticum		2		1		1
Chryseobacterium scophthalmum				1		1
Chryseobacterium sp.		1				
Chryseobacterium taeanense				1		
Cloacibacterium rupense				1		
Herbaspirillum frisingense						3
Herbaspirillum huttiense				3		13
Herbaspirillum sp.				1		
Novosphingobium barchaimii			1			
Novosphingobium resinovorum	7		5			
Pseudomonas protegens					4	
Pseudomonas sp.			1			
Rhizobium nepotum			1			
Rhizobium sp.	1					
Rhizobium tibeticum			1			
Sphingobacterium faecium						2(2)
Sphingobacterium kitahiroshimense		2				
Sphingobacterium sp.		1		2		3
Sphingobium sp.					2	
Stenotrophomonas maltophilia			1			
Stenotrophomonas rhizophila					1	
Variovorax guangxiensis			1			
Bm contaminants		20				
Total GNNF	9	26	11	10	7	23
Total isolates	47	26	43	10	57	23

Each cell represents the number of isolates per MALDI-TOF MS cluster. The number of clusters is given in between brackets if more than one. Bm, *B. multivorans*; P, *Pseudomonas* sp.; S, *Staphylococcus* sp. (Table 3.19).

Table 3.22: Identification and number of isolates for sample S35 (experiment 16)

	Bm		P		S	
	1/10 PCAT	**BCSA**	**1/10 PCAT**	**BCSA**	**1/10 PCAT**	**BCSA**
Burkholderia choica					2	
Burkholderia jiangsuensis			2		26(5)	
Burkholderia phytofirmans	10		8		24(5)	
Burkholderia sp.			1		3(2)	
Pseudomonas prosekii					1	
Bm contaminants	43	34	45	32	6	8
Total GNNF	53	34	56	32	62	8
Total isolates	74	34	56	32	64	8

Each cell represents the number of isolates per MALDI-TOF MS cluster. The number of clusters is given in between brackets if more than one. Bm, *B. multivorans*; P, *Pseudomonas* sp.; S, *Staphylococcus* sp. (Table 3.19).

3.4 Multilocus sequence types predict phylogeny and gene content but not isolation source for *Burkholderia multivorans*

Redrafted from: Charlotte Peeters, Vaughn S. Cooper, Philip Hatcher, Bart Verheyde and Peter Vandamme. Multilocus sequence types predict phylogeny and gene content but not isolation source for *Burkholderia multivorans*. In preparation.

Author contributions: CP, VC and PV conceived the study and wrote the manuscript. CP performed the DNA preparations. VC directed the genomic sequencing methods. VC and PH identified the orthologous genes. CP and VC performed the genomic data analysis and statistical analysis. CP and BV performed the whole-genome sequence-based phylogenetic analysis.

3.4.1 Abstract

The natural environment may serve as a reservoir of opportunistic pathogens. A well-established method for studying the epidemiology of such opportunists is multilocus sequence typing, which in many cases has defined strains predisposed to causing infection. *Burkholderia multivorans* is an important pathogen in people with CF and its epidemiology suggests that strains are acquired from non-human sources such as the natural environment. This raises the central question of whether the isolation source (CF or environment) or the ST of *B. multivorans* better predicts their genomic content and functionality. We identified four pairs of *B. multivorans* isolates, representing distinct STs and consisting of one CF and one environmental isolate. All genomes were sequenced using the PacBio SMRT sequencing technology, which resulted in eight high-quality *B. multivorans* genome assemblies. The present study demonstrated that the genomic structure of *B. multivorans* is highly conserved and that the *B. multivorans* genomic lineages are defined by their ST. Orthologous protein families were not uniformly distributed among chromosomes, with core orthologs being enriched on the primary chromosome and ST-specific orthologs being enriched on the second and third chromosome. The ST-specific orthologs were enriched in genes involved in defense mechanisms and secondary metabolism, corroborating the strain-specificity of these virulence characteristics. Finally, the same *B. multivorans* genomic lineages were isolated from CF and environmental samples and on different continents many years apart, demonstrating the evolutionary persistence and ubiquity of these strains in different niches and on different continents.

3.4.2 Introduction

Bcc bacteria are rare but potentially virulent pathogens in CF patients (LiPuma, 2010). Epidemiological surveys revealed that *B. multivorans* is the most prevalent Bcc CF pathogen in many countries (Brisse *et al.*, 2004; De Boeck *et al.*, 2004; Govan *et al.*, 2007; LiPuma, 2010; Norskov-Lauritsen *et al.*, 2010; Pope *et al.*, 2010; Medina-Pascual *et al.*, 2012). The continued emergence of unique *B. multivorans* strains in CF patients suggests acquisition from non-human sources, such as the natural environment (Baldwin *et al.*, 2008). Environmental conditions or non-human hosts in which virulence factors might be adaptive can select for traits that confer virulence and natural environments could therefore serve as reservoirs of opportunistic pathogens (Coenye & Vandamme, 2003; Berg *et al.*, 2005).
MLST is a well-established method for studying the epidemiology and population structure of Bcc organisms (Baldwin *et al.*, 2005; Spilker *et al.*, 2009). The Bcc MLST scheme takes into account the allelic variation of seven housekeeping genes and each strain is defined by its unique allelic profile and ST (Jolley & Maiden, 2010). Baldwin *et al.* (2007) demonstrated that roughly one fifth of the clinical isolates in the Bcc PubMLST database had the same ST as environmental isolates, suggesting these isolates represent the same strain. A follow-up study demonstrated that several *B. multivorans* STs were globally distributed and that the natural environment (e.g. water and soil) may be an important reservoir for infection with this species (Baldwin *et al.*, 2008).
Burkholderia genomes vary in size from 2.4 Mb (*Ca.* Burkholderia schumannianae UZHbot8) (Pinto-Carbo *et al.*, 2016) to 11.5 Mb (*Burkholderia terrae* BS001) (Nazir *et al.*, 2012a), are characterized by a high G+C content (62-68 mol%) and consist of multiple replicons (Mahenthiralingam *et al.*, 2005; Ussery *et al.*, 2009). To gain insight into the overall genome biology of *B. multivorans*, we sequenced the genomes of eight isolates representing four distinct STs. For each ST, a CF and an environmental isolate were sequenced using the PacBio Single-Molecule Real-Time (SMRT) sequencing technology. The present study provides the first comprehensive comparative genome analysis of *B. multivorans* and assesses to which extent isolates with the same ST but from different origin (CF versus environmental) differ in genetic potential.

3.4.3 Materials and methods

3.4.3.1 Studied isolates

We searched the Bcc PubMLST database (http://pubmlst.org/bcc/) (Jolley & Maiden, 2010) for *B. multivorans* STs that included both CF and environmental (ENV) isolates, and selected

eight isolates, representing four STs, for whole-genome sequencing (Table 3.23). Strains were grown aerobically on Tryptone Soya Agar (Oxoid) and incubated at 28 °C. Cultures were preserved in MicroBank™ vials at −80 °C.

Table 3.23: *B. multivorans* isolates included in the present study

Isolate	Strain number	ST	Isolation source	Depositor
ST180-ENV	LMG 29305, J2943	180	Rhizosphere soil (United Kingdom, 2000)	J. Govan
ST180-CF	LMG 29313, 8335	180	CF sputum (Czech republic, 2011)	P. Drevinek
ST189-ENV	LMG 29309	189	Soil succulent (Belgium, 2003)	Own isolate
ST189-CF	LMG 29312, BCC0208	189	CF patient (Canada, 1999)	E. Mahenthiralingam
ST287-ENV	LMG 29306, J2947	287	Rhizosphere soil (United Kingdom, 2000)	J. Govan
ST287-CF	LMG 29311, BCC0059	287	CF patient (Canada, 1995)	E. Mahenthiralingam
ST650-ENV	LMG 29308	650	Pond water (Belgium, 2003)	Own isolate
ST650-CF	LMG 29310, Q113	650	CF patient (Germany, 2010)	B. Kahl

LMG, BCCM/LMG Bacteria Collection, Laboratory of Microbiology, Ghent University, Ghent, Belgium. CF, cystic fibrosis; ENV, environmental; ST, multilocus sequence type.

3.4.3.2 Genome sequencing, assembly and annotation

High-quality DNA was prepared using Qiagen Genomic tips (20/G) and genomes were sequenced using the P5-C3 chemistry on the PacBio SMRT II platform of the Department of Genetics and Genomic Sciences of the Icahn School of Medicine at Mount Sinai (New York, USA). One SMRT cell per isolate was sequenced, except for isolates ST189-CF and ST287-ENV for which a second SMRT cell was run to increase the quality of the raw data. PacBio reads were assembled in the sequencing center using the SMRT analysis software (including HGAP3 and Quiver) and contigs were ordered against the complete reference genome of *B. multivorans* strain ATCC 17616 (PRJNA17407) using Mauve (Rissman *et al.*, 2009). We further polished the assemblies in five steps. The first step consisted of removing spurious contigs that were small in size, had a low coverage and resulted in a highest BLAST hit with the primary chromosome of its own genome (Liao *et al.*, 2015). Reads were mapped using pbalign and QC reports were created based on the resulting BAM files using Qualimap (Garcia-Alcalde *et al.*, 2012). Contigs smaller than 20 kb and with less than 20x coverage or a high variation (SD) in coverage were discarded. In step two, read mappings were used to further polish the contigs using Pilon (Walker *et al.*, 2014b) with default parameters. In step three, contigs with overlapping ends were merged using Gap5 (Bonfield & Whitwham, 2010) to exclude artificially duplicated regions, often including many frameshifts and fragmented open reading frames. In step four, the duplicated ends of circular contigs were trimmed using

Gepard (Krumsiek *et al.*, 2007) and Gap5 as these duplications were a consequence of the circular nature of the replicons in combination with the long-read sequencing technology (Scott & Ely, 2015). Importantly, this artificial duplication of contig ends not only resulted in a highly variable rRNA copy number, but also falsely excluded genes from the ortholog dataset because they were artificially duplicated. Since the merging of overlapping ends by Gap5 might be imperfect, we ran Quiver in a final polishing step. The PacBio sequencing reads of one SMRT cell resulted in a coverage ranging from 76x to 119x.
Annotation was performed using Prokka v1.11 (Seemann, 2014) with a genus-specific database based on reference genomes from the *Burkholderia* Genome database (http://beta.burkholderia.com/) (Winsor *et al.*, 2008). The annotated genome assemblies were submitted to the European Nucleotide Archive and are publicly available through the GenBank/EMBL/DDBJ accession numbers FKJT01000000, FKJS01000000, FKJU01000000, FKJP01000000, FKJV01000000, FKJW01000000, FKJX01000000 and FKJY01000000. The genome sequence of *B. multivorans* strain ATCC 17616 (PRJNA17407) was included as a reference in all further analyses. A multiple genome alignment was performed using Mauve (Darling *et al.*, 2010) to assess the basic genome structure.

3.4.3.3 Analysis of protein-coding genes and ortholog identification

We mapped for each protein-coding gene (CDS) on which chromosome it was located and to which cluster of orthologous groups (COG) it belonged (Galperin *et al.*, 2015). COGs were assigned by a reversed position-specific BLAST (RPSBLAST v2.2.29+) with an e-value cut-off of 1E-3 against the NCBI conserved domain database (CDD v3.14). Orthologous genes were identified as described previously (Lerat *et al.*, 2003; Cooper *et al.*, 2010). In short, homologs were identified as reciprocal best BLAST hits with a normalized bit score, providing an empirically determined taxon-specific threshold. The putative panorthologs (i.e. single-copy orthologous genes conserved in all genomes) were computed while varying the bit score threshold from 0.1 to 0.9 in 0.1 increments and the largest set of panorthologs was selected. For each orthologous protein family, the consensus chromosome location and COG were determined. Conflicts in COG mapping were resolved by the majority rule.

3.4.3.4 Phylogenomic analysis

The whole-genome phylogeny was calculated based on the sequences of the panorthologs as described previously (Pinto-Carbo *et al.*, 2016). In short, amino acid sequences were aligned using MUSCLE (Edgar, 2004) and translated back to the respective nucleotide sequences using T-Coffee (Notredame *et al.*, 2000). Nucleotide alignments were trimmed using trimAl

(Capella-Gutierrez *et al.*, 2009) by removing positions with gaps in more than 50 % of the sequences, and were subsequently concatenated to construct a maximum likelihood tree using RaXML v7.4.2 (Stamatakis, 2014) with the GTRGAMMA substitution model and 1000 rapid bootstrap analyses.
In a second approach, the presence/absence matrix of all orthologs was used in a discrete character-state parsimony analysis using pars from the PHYLIP package (Felsenstein, 2005) to assess the relatedness of the genomes in terms of gene content.

3.4.3.5 Comparison of *B. multivorans* and *B. cenocepacia* COG profiles

Complete genome sequences of *B. cenocepacia* strains J2315, H111, K56-2Valvano, AU1054, HI2424 and MC0-3 were downloaded from the *Burkholderia* Genome database (http://beta.burkholderia.com/) (Winsor *et al.*, 2008). COG mapping of *B. cenocepacia* CDS was performed as described above for *B. multivorans*. The number of CDS per COG category for each species (*B. multivorans* versus *B. cenocepacia*) was counted and the distributions were compared using Pearson chi-square analysis.

3.4.3.6 Data visualization and statistical analyses

Data visualization and statistical analyses were performed using RStudio with R v3.2.3. Pearson's chi-square analyses were used to test the association between different sets of categorical variables. When a significant relationship was found between two variables, we further examined the standardized Pearson residuals. Standardized Pearson residuals having high absolute values indicate a lack of fit of the null hypothesis of independence in each cell (Agresti, 2002) and thus indicate observed cell frequencies in the contingency table that are significantly higher or lower than one would expect based on coincidence. In case multiple COG categories were registered for the same COG, each COG category was counted separately for Pearson chi-square analysis on COG categories. For the 198 CDS that were involved in the translocation within the ST650-CF isolate from the primary to the secondary chromosome, the consensus chromosome mapping was set to the primary chromosome for Pearson chi-square analysis on chromosome distribution.

3.4.4 Results

3.4.4.1 The genomic structure of *B. multivorans* is highly conserved

The final assemblies produced closed genomes for five of the eight *B. multivorans* isolates sequenced in the present study. The genomes were 6.2-6.9 Mb in size with a G+C content of

~67 mol% and the number of predicted CDS ranged from 5,415 to 6,155 CDS per genome (Table 3.24). No clustered regularly interspaced short palindromic repeats (CRISPRs) were identified.

The multiple genome alignment revealed a highly conserved genomic structure with three chromosomes (from here on referred to as C1, C2 and C3). C1, C2 and C3 were on average 3.4 Mb, 2.4 Mb and 0.6 Mb in size. Both ST180 isolates harbored one contig that did not map onto the reference genome of ATCC 17616. These contigs were 22,339 and 28,809 bp in size, had a G+C content of 58 mol% and were therefore considered plasmids. Both plasmids contained genes for an initiator repB protein, an AsnC transcriptional regulator, a cobyrinic acid a,c-diamide synthase (*parA* homologue), multiple integrases and several hypothetical proteins. The multiple genome alignment also revealed a fairly large translocation (207 genes, 198 CDS) within the ST650-CF isolate from C1 to C2 that was delimited by rRNA operons at both ends. All isolates except ST650-CF contained 3, 1 and 1 rRNA copies on C1, C2 and C3, respectively. As a result of the translocation, isolate ST650-CF contained 2, 2 and 1 rRNA copies on C1, C2 and C3, respectively.

Table 3.24: *B. multivorans* genome characteristics

Isolate	Contigs	Size (bp)	%GC	C1	C2	C3	Plasmid	Total CDS	Ortho-logous CDS	Non-ortho-logous CDS
ST180-ENV	4	6,464,081	67.1	1	2	3	4	5,794	5,271	523
ST180-CF	4	6,296,736	67.3	1	2	3	4	5,551	5,266	285
ST189-ENV	3	6,223,431	67.3	1	2	3	-	5,467	5,144	323
ST189-CF	13	6,157,395	67.3	1-5	6-9	10-13	-	5,415	5,132	283
ST287-ENV	4	6,559,547	67.2	1-2	3	4	-	5,800	5,494	306
ST287-CF	6	6,857,684	67.0	1-4	5	6	-	6,155	5,505	650
ST650-ENV	3	6,322,929	67.2	1	2	3	-	5,594	5,275	319
ST650-CF	3	6,308,820	67.2	1	2	3	-	5,599	5,250	349
ATCC 17616	4	7,008,622	66.7	1	2	3	4	6,258	4,893	1,365

C1, C2 and C3: contigs mapping to chromosome 1, 2 and 3 of *B. multivorans* strain ATCC 17616, respectively; bp, base pairs.

3.4.4.2 ST predicts genomic lineage

Orthologous genes were identified to determine the conserved genome content of *B. multivorans*. The ortholog analysis identified 6,254 homologous protein families comprising 47,230 CDS in total (Table 3.24). The largest set of panorthologs, i.e. orthologs conserved in all nine *B. multivorans* genomes and present as single copies, was found at a reciprocal best bit score threshold of 0.7 (see Methods section) and comprised 4,503 ortholog families.

The frequency of orthologous versus non-orthologous CDS varied significantly per isolate ($\chi^2(8)=1829.6$, $p<0.001$) and ST ($\chi^2(3)=67.3$, $p<0.001$), but not isolation source ($p>0.05$). The genomes of isolates ST287-CF and ATCC 17616 were significantly enriched with non-orthologous CDS, while those of ST180-CF, ST287-ENV, ST650-ENV and both ST189 isolates were significantly deprived in non-orthologous CDS (Table 3.24) (Supplementary Table 3.3). Analysis of the relationship between orthologous versus non-orthologous CDS and ST showed that the ST287 genomes were significantly enriched with non-orthologous CDS, while the ST189 and ST650 genomes were significantly deprived in non-orthologous CDS (Supplementary Table 3.4).

The ortholog dataset enabled two subsequent analyses of strain phylogeny. In the first approach, a whole-genome phylogeny was obtained based on nucleotide sequence divergence of the panorthologs (Fig. 3.4). In the second approach, the presence/absence matrix of the ortholog families was used to assess the relatedness of the genomes in terms of gene content using parsimony (Supplementary Fig. 3.2). These analyses both demonstrated that the ST, and not the isolation source, of the *B. multivorans* isolates predicted their phylogeny and gene content. This finding demonstrated that isolates with the same ST represent the same genomic lineage, irrespective of their isolation source.

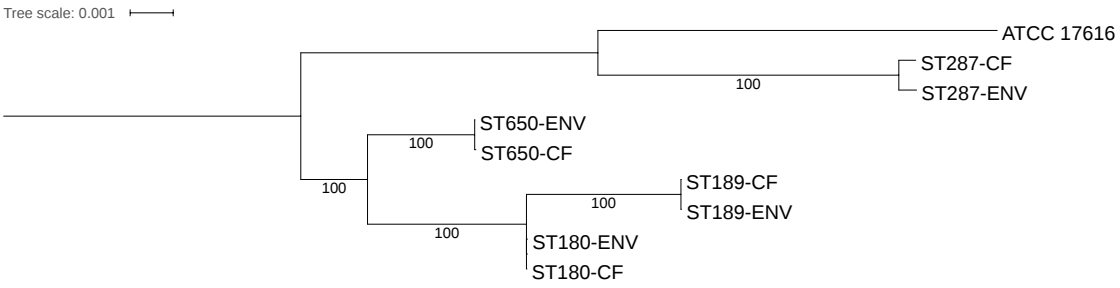


Figure 3.4: Phylogenomic analysis showing the relatedness of the genomes in terms of sequence divergence of the panorthologs. The maximum likelihood tree was inferred using the GTRGAMMA substitution model and is based on a concatenated nucleotide alignment of 4,503 CDS (4,457,847 positions). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap analyses (1,000 replicates) are shown next to the branches. Scale bar represents number of substitutions per site. The tree was rooted on the branch with the largest branch length.

3.4.4.3 Orthologous genes are enriched on C2 and are involved in carbohydrate metabolism and transport

Because the fraction of genes that are involved in housekeeping functions varies among the chromosomes (Mahenthiralingam *et al.*, 2005), we mapped the chromosome location of each CDS. Consistent with the average chromosome size, the total number of CDS was

highest on C1 (27,813 CDS), followed by C2 (18,565 CDS) and C3 (5,047 CDS). The plasmid of the ST180 isolates harbored 208 CDS, of which 206 were non-orthologous CDS. The translocation within the ST650-CF isolate from C1 to C2 comprised 198 CDS, of which 13 were non-orthologous CDS. The frequency of orthologous versus non-orthologous CDS varied significantly among the different chromosomes ($\chi^2(2)=213.4$, $p<0.001$) (Fig. 3.5a). C1 was significantly enriched with non-orthologous CDS, while C2 was significantly enriched with orthologous CDS (Fig. 3.5b).
To identify biological functions that were over- or underrepresented, we assigned each CDS to a COG. Roughly 80 % of the CDS (41,520 CDS in total) could be assigned to a COG and its associated COG functional category. The frequency of orthologous versus non-orthologous CDS varied significantly among the different COG categories ($\chi^2(22)=5101.2$, $p<0.001$) (Fig. 3.5c). The non-orthologous CDS were significantly enriched in the COG categories cell wall/membrane/envelope biogenesis (M), intracellular trafficking, secretion and vesicular transport (U), amino acid transport and metabolism (E) and mobilome (X), while the orthologous CDS were significantly enriched in the COG categories carbohydrate transport and metabolism (G) and general function prediction only (R) (Table 3.25).

3.4.4.4 Each ST harbors unique orthologs

For each ortholog family we examined whether it was present in all eight isolates (i.e. showed core specificity), specific for isolates of one or more STs, specific for isolates of a specific source or randomly present. None of the ortholog families was present in all four isolates of a specific source (CF vs. ENV), but a small number of ortholog families were present in only one, two or three isolates from the same source, thus leaving five relevant specificity groups: core (n=4,684), ST (n=1,362), CF-only (n=38), ENV-only (n=51) and random (n=119). The Venn diagram (Fig. 3.6) visualizes the number of ortholog families in the core and ST specificity groups (n=6,046) and shows that each ST harbors 103-539 orthologs that were not present in any other ST.

3.4.4.5 ST-specific orthologs are enriched on C2 and C3 and are involved in defense mechanisms and secondary metabolism

Based on the chromosome and COG mapping for the individual CDS, we mapped the consensus chromosome location and COG category for each ortholog family. Consistent with the average chromosome size, the number of orthologs was highest on C1 (3,242), followed by C2 (2,264) and C3 (710). For 37 ortholog families there was a conflict in chromosome mapping, and 1 ortholog was located on the plasmid of the ST180 isolates. COGs and

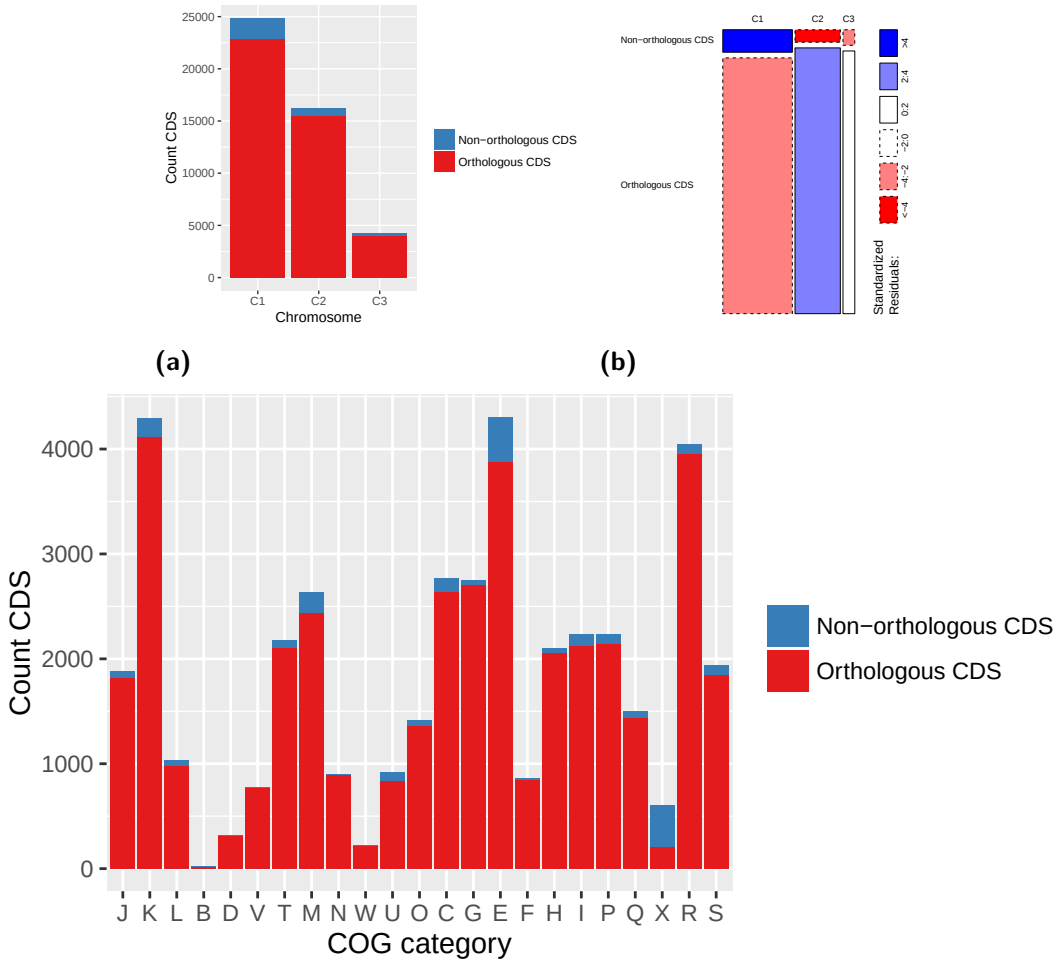


Figure 3.5: The frequency of orthologous versus non-orthologous CDS varies among chromosomes and COG categories. Bar plots show the number of orthologous and non-orthologous CDS per chromosome ($\chi^2(2)=213.4$, $p<0.001$) **(a)** and COG category ($\chi^2(22)=5101.2$, $p<0.001$) **(c)**. Mosaic plots show the standardized residuals of the Pearson chi-square analysis for the number of orthologous and non-orthologous CDS per chromosome **(b)**. Solid and dashed boundaries represent positive and negative residuals, respectively. Rectangles are colored only if the standardized residual is significant at $p<0.05$ (outside ± 1.96). COG categories: J, translation, ribosomal structure and biogenesis; K, transcription; L, replication, recombination and repair; B, chromatin structure and dynamics; D, cell cycle control, cell division, chromosome partitioning; V, defense mechanisms; T, signal transduction mechanisms; M, cell wall/membrane/envelope biogenesis; N, cell motility; W, extracellular structures; U, intracellular trafficking, secretion, and vesicular transport; O, posttranslational modification, protein turnover, chaperones; X, mobilome: prophages, transposons; C, energy production and conversion; G, carbohydrate transport and metabolism; E, amino acid transport and metabolism; F, nucleotide transport and metabolism; H, coenzyme transport and metabolism; I, lipid transport and metabolism; P, inorganic ion transport and metabolism; Q, secondary metabolites biosynthesis, transport and catabolism; R, general function prediction only; S, function unknown.

Table 3.25: The frequency of orthologous versus non-orthologous CDS varies among the COG categories

		Orthologous CDS		Non-orthologous CDS
Information storage and processing				
J	Translation, ribosomal structure and biogenesis	1814 (0.789)	−	65 (-3.359)
K	Transcription	4122 (0.766)	−	176 (-3.258)
L	Replication, recombination and repair	977 (-0.032)		55 (0.137)
B	Chromatin structure and dynamics	24 (0.263)		0 (-1.121)
Cellular processes and signaling				
D	Cell cycle control, cell division, chromosome partitioning	319 (0.848)	−	2 (-3.610)
V	Defense mechanisms	765 (1.199)	−	8 (-5.101)
T	Signal transduction mechanisms	2108 (0.988)	−	69 (-4.207)
M	Cell wall/membrane/envelope biogenesis	2441 (-1.161)	+	196 (4.941)
N	Cell motility	898 (1.510)	−	3 (-6.429)
W	Extracellular structures	225 (0.806)	−	0 (-3.431)
U	Intracellular trafficking, secretion, and vesicular transport	838 (-1.241)	+	85 (5.283)
O	Posttranslational modification, protein turnover, chaperones	1367 (0.711)	−	48 (-3.025)
Metabolism				
C	Energy production and conversion	2641 (0.329)		128 (-1.401)
G	Carbohydrate transport and metabolism	+ 2710 (2.016)	−	41 (-8.579)
E	Amino acid transport and metabolism	− 3880 (-3.142)	+	426 (13.373)
F	Nucleotide transport and metabolism	845 (0.916)	−	19 (-3.897)
H	Coenzyme transport and metabolism	2058 (1.521)	−	42 (-6.475)
I	Lipid transport and metabolism	2127 (0.091)		113 (-0.387)
P	Inorganic ion transport and metabolism	2148 (0.567)	−	91 (-2.415)
Q	Secondary metabolites biosynthesis, transport and catabolism	1442 (0.543)	−	58 (-2.311)
Poorly characterized				
R	General function prediction only	+ 3954 (2.026)	−	86 (-8.623)
S	Function unknown	1851 (0.358)		86 (-1.524)
Mobile elements				
X	Mobilome: prophages, transposons	− 207 (-15.300)	+	398 (65.117)

Pearson chi-square analysis testing the independence of gene conservation (orthologous vs. non-orthologous CDS) and COG category ($\chi^2(22)=5101.2$, $p<0.001$). Each cell in the contingency represents the observed frequency and standardized residual (in between brackets) and is preceded by + or − if the standardized residual is >1.96 or <-1.96, respectively, and significant at $p<0.05$.

their associated COG functional category could be assigned to 4,896 of the 6,254 ortholog families.

The specificity of the ortholog families varied significantly among the chromosomes ($\chi^2(8)=469.8$, $p<0.001$) (Fig. 3.7a) and COG categories ($\chi^2(88)=649.8$, $p<0.001$) (Fig. 3.7c). C2 and C3 were significantly enriched with ST-specific orthologs, while C1 was significantly enriched with orthologs belonging to the specificity groups core, random, CF-only and ENV-only (Fig. 3.7b). The ST-specific orthologs were significantly enriched in the COG categories defense mechanisms (V), secondary metabolites biosynthesis, transport and catabolism (Q), mobilome (X) and general function prediction only (R) (Table 3.26).

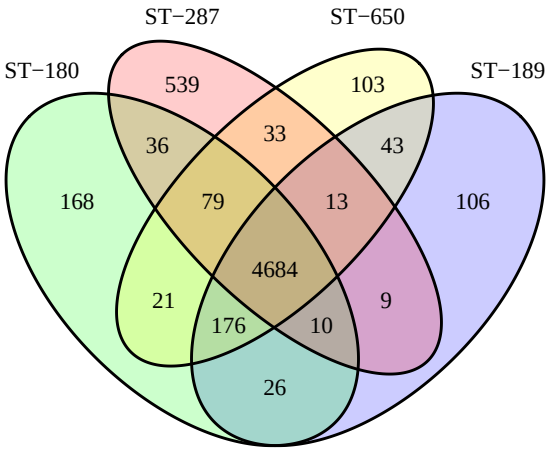


Figure 3.6: Venn diagram showing the number of core and ST-specific ortholog families

3.4.4.6 ST287 harbors extra orthologous and non-orthologous genes

Both ST287 genomes were considerably larger and contained a higher number of CDS (Table 3.24), suggesting that this genomic lineage contains extra genes. ST287 was not only enriched with non-orthologous CDS (Supplementary Table 3.4) but also harbored 539 orthologs that were not present in the other three STs (Fig. 3.6), showing that the extra genes in ST287 are both orthologous and non-orthologous CDS. Conversely, the other three STs (ST180, ST189 and ST650) shared 176 orthologs that were absent in ST287 (Fig. 3.6). We found no significant association between the specificity profile and distribution of COG categories ($p>0.05$), suggesting that the 539 ST287-specific orthologs and 176 orthologs that were specific for the other three STs (ST180, ST189 and ST650) are involved in similar biological processes.

3.4.4.7 C1 is enriched with orthologs showing CF and ENV specificity

C1 was enriched with orthologs that were present in only one, two or three isolates from the same source (CF-only or ENV-only) (Fig. 3.7a and 3.7b). CF-only orthologs were significantly enriched in the COG categories cell cycle control, cell division and chromosome partitioning (D), extracellular structures (W), intracellular trafficking, secretion and vesicular transport (U) and mobilome (X), while ENV-only orthologs were significantly enriched in replication, recombination and repair (L), intracellular trafficking, secretion and vesicular transport (U) and posttranslational modification, protein turnover and chaperones (O) (Fig. 3.7c and Table

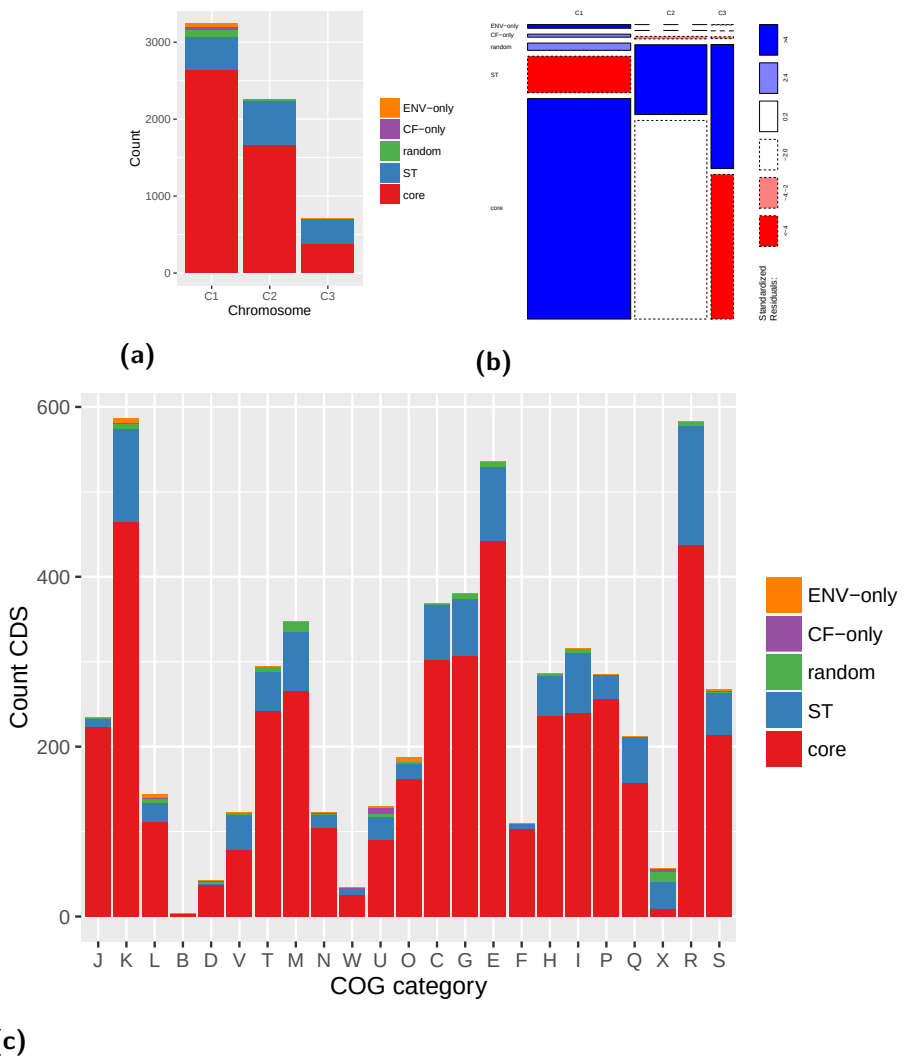


Figure 3.7: Ortholog specificity varies among chromosomes and COG categories. Bar plots show the number of orthologs per specificity group for different chromosomes ($\chi^2(8)=469.8$, $p<0.001$) **(a)** and COG categories ($\chi^2(88)=649.8$, $p<0.001$) **(c)**. Mosaic plots show the standardized residuals of the Pearson chi-square analysis on the number of orthologs per specificity group per chromosome **(b)**. Solid and dashed boundaries represent positive and negative residuals, respectively. Rectangles are colored only if the standardized residual is significant at $p<0.05$ (outside ± 1.96). COG categories: J, translation, ribosomal structure and biogenesis; K, transcription; L, replication, recombination and repair; B, chromatin structure and dynamics; D, cell cycle control, cell division, chromosome partitioning; V, defense mechanisms; T, signal transduction mechanisms; M, cell wall/membrane/envelope biogenesis; N, cell motility; W, extracellular structures; U, intracellular trafficking, secretion, and vesicular transport; O, posttranslational modification, protein turnover, chaperones; X, mobilome: prophages, transposons; C, energy production and conversion; G, carbohydrate transport and metabolism; E, amino acid transport and metabolism; F, nucleotide transport and metabolism; H, coenzyme transport and metabolism; I, lipid transport and metabolism; P, inorganic ion transport and metabolism; Q, secondary metabolites biosynthesis, transport and catabolism; R, general function prediction only; S, function unknown.

3.26).
Additionally, C1 was also enriched with orthologs showing random specificity (Fig. 3.7a and 3.7b) and these orthologs with random specificity were significantly enriched in the COG categories replication, recombination and repair (L), cell cycle control, cell division and chromosome partitioning (D), cell wall/membrane/envelope biogenesis (M) and mobilome (X) (Fig. 3.7c and Table 3.26).

3.4.4.8 Comparison of *B. multivorans* and *B. cenocepacia* average COG profiles

During the past two decades, *B. multivorans* and *B. cenocepacia* have been the most prevalent Bcc pathogens in CF. Historically, *B. cenocepacia* strains have been responsible for large epidemics within the CF community and are often extremely virulent (Drevinek & Mahenthiralingam, 2010). In contrast, only a limited number of *B. multivorans* outbreak strains were described and *B. multivorans* is generally considered a less virulent Bcc pathogen as compared to *B. cenocepacia* (Mahenthiralingam *et al.*, 2005). To examine the species-specific genome content of *B. multivorans*, we compared its average COG profile to that of *B. cenocepacia*. *B. cenocepacia* genomes contained generally more genes (6,477-7,116 CDS per genome) (Supplementary Table 3.5) than *B. multivorans* genomes (5,415-6,155 CDS per genome) (Table 3.24) and more CDS per COG category (Supplementary Fig. 3.3). We compared the average COG profile of the two species by calculating the average number of CDS per genome in each COG category and by comparing these distributions. The distribution of CDS among COG categories varied significantly between the two species ($\chi^2(22)=102.9$, $p<0.001$) (Supplementary Fig. 3.3). *B. cenocepacia* genomes harbor significantly more CDS in the COG categories transcription (K), defense mechanisms (V) and general function prediction only (R) and significantly less in translation, ribosomal structure and biogenesis (J) and replication, recombination and repair (L). Conversely, *B. multivorans* genomes harbor significantly more CDS in the COG categories replication, recombination and repair (L) and less in transcription (K) (Table 3.27).
Finally, we searched for COGs that were exclusively present in either *B. multivorans* or *B. cenocepacia* genomes. In total, 124 COGs were exclusively present in one or more *B. multivorans* genomes, but only 21 COGs were uniquely present in all nine of the *B. multivorans* genomes (Table 3.28). Conversely, 204 COGs were exclusively present in one or more *B. cenocepacia* genomes, but only 72 COGs were uniquely present in all six of the *B. cenocepacia* genomes (Supplementary Table 3.6).

Table 3.26: Ortholog specificity varies among the COG categories

		Core		ST		CF-only		ENV-only		Random
Information storage and processing										
J	Translation, ribosomal structure and biogenesis	+ 223 (2.570)	−	10 (-4.947)		0 (-0.865)		1 (0.044)		1 (-1.306)
K	Transcription	465 (-0.187)		109 (0.375)		2 (0.095)		5 (1.688)		6 (-0.864)
L	Replication, recombination and repair	112 (-0.286)		22 (-0.748)		1 (0.799)	+	4 (4.458)	+	5 (2.013)
B	Chromatin structure and dynamics	3 (0.389)		0 (-0.733)		0 (-0.098)		0 (-0.111)		0 (-0.209)
Cellular processes and signaling										
D	Cell cycle control, cell division, chromosome partitioning	37 (0.594)	−	1 (-2.378)	+	1 (2.368)		0 (-0.414)	+	3 (3.061)
V	Defense mechanisms	79 (-1.872)	+	41 (4.095)		0 (-0.623)		0 (-0.705)		2 (0.172)
T	Signal transduction mechanisms	242 (0.461)		46 (-0.919)		0 (-0.968)		0 (-1.094)		6 (0.839)
M	Cell wall/membrane/envelope biogenesis	266 (-0.677)		69 (0.867)		0 (-1.052)		0 (-1.189)	+	12 (3.102)
N	Cell motility	104 (0.576)		16 (-1.286)		1 (0.971)		1 (0.705)		1 (-0.588)
W	Extracellular structures	25 (-0.416)		8 (0.774)	+	1 (2.709)		0 (-0.372)		0 (-0.703)
U	Intracellular trafficking, secretion, and vesicular transport	90 (-1.362)		27 (0.769)	+	7 (10.233)	+	2 (2.022)		4 (1.538)
O	Posttranslational modification, protein turnover, chaperones	162 (1.028)	−	18 (-2.678)		0 (-0.772)	+	5 (4.858)		2 (-0.434)
Metabolism										
C	Energy production and conversion	302 (0.463)		65 (-0.114)		0 (-1.083)		0 (-1.224)		1 (-1.879)
G	Carbohydrate transport and metabolism	307 (0.192)		67 (-0.130)		0 (-1.100)		0 (-1.244)		6 (0.206)
E	Amino acid transport and metabolism	443 (0.710)		86 (-1.023)		0 (-1.307)		0 (-1.477)		7 (-0.280)
F	Nucleotide transport and metabolism	104 (1.717)	−	5 (-3.313)		0 (-0.592)		0 (-0.669)		1 (-0.472)
H	Coenzyme transport and metabolism	237 (0.560)		46 (-0.731)		0 (-0.955)		0 (-1.079)		3 (-0.565)
I	Lipid transport and metabolism	240 (-0.787)		71 (1.913)		0 (-1.003)		1 (-0.253)		4 (-0.274)
P	Inorganic ion transport and metabolism	256 (1.873)	−	28 (-3.227)		0 (-0.953)		1 (-0.149)	−	0 (-2.034)
Q	Secondary metabolites biosynthesis, transport and catabolism	157 (-0.953)	+	55 (2.762)		0 (-0.822)		0 (-0.929)		0 (-1.754)
Poorly characterized										
R	General function prediction only	437 (-1.337)	+	141 (3.577)		0 (-1.363)		0 (-1.541)		5 (-1.190)
S	Function unknown	214 (-0.010)		49 (0.143)		2 (1.240)		2 (0.870)		1 (-1.465)
Mobile elements										
X	Mobilome: prophages, transposons	− 9 (-5.415)	+	32 (6.818)	+	3 (6.613)		1 (1.594)	+	12 (12.283)

Pearson chi-square analysis testing the independence of ortholog specificity and COG category ($\chi^2(88)=469.8$, $p<0.001$). Each cell in the contingency represents the observed frequency and standardized residual (in between brackets) and is preceded by + or − if the standardized residual is >1.96 or <-1.96, respectively, and significant at $p<0.05$.

Table 3.27: The distribution of *B. multivorans* versus *B. cenocepacia* CDS varies among COG categories

			B. cenocepacia		***B. multivorans***
Information storage and processing					
J	Translation, ribosomal structure and biogenesis	−	1528 (-2.166)		2118 (1.931)
K	Transcription	+	4314 (3.883)	−	4866 (-3.462)
L	Replication, recombination and repair	−	809 (-2.797)	+	1206 (2.494)
B	Chromatin structure and dynamics		21 (-0.057)		27 (0.051)
Cellular processes and signaling					
D	Cell cycle control, cell division, chromosome partitioning		261 (-1.082)		369 (0.964)
V	Defense mechanisms	+	800 (2.046)		880 (-1.825)
T	Signal transduction mechanisms		2032 (0.726)		2482 (-0.647)
M	Cell wall/membrane/envelope biogenesis		2438 (0.658)		2993 (-0.587)
N	Cell motility		715 (-1.808)		1012 (1.612)
W	Extracellular structures		226 (0.949)		253 (-0.846)
U	Intracellular trafficking, secretion, and vesicular transport		750 (-1.798)		1058 (1.603)
O	Posttranslational modification, protein turnover, chaperones		1251 (-0.421)		1607 (0.376)
Metabolism					
C	Energy production and conversion		2357 (-1.677)		3151 (1.496)
G	Carbohydrate transport and metabolism		2519 (0.619)		3098 (-0.552)
E	Amino acid transport and metabolism		3694 (-1.403)		4840 (1.251)
F	Nucleotide transport and metabolism		747 (-0.603)		977 (0.538)
H	Coenzyme transport and metabolism		1854 (-0.344)		2364 (0.297)
I	Lipid transport and metabolism		2090 (1.130)		2513 (-1.008)
P	Inorganic ion transport and metabolism		1883 (-1.820)		2550 (1.623)
Q	Secondary metabolites biosynthesis, transport and catabolism		1367 (0.495)		1678 (-0.441)
Poorly characterized					
R	General function prediction only	+	3843 (1.965)		4562 (-1.752)
S	Function unknown		1685 (-0.887)		2202 (0.791)
Mobile elements					
X	Mobilome: prophages, transposons		630 (0.830)		746 (-0.740)

Pearson chi-square analysis testing the independence of species and COG category ($\chi^2(22)=102.9$, $p<0.001$). Each cell in the contingency represents the observed frequency and standardized residual (in between brackets) and is preceded by + or − if the standardized residual is >1.96 or <-1.96, respectively, and significant at $p<0.05$.

3.4.5 Discussion

While infection control measures reduced patient-to-patient transmission and thereby the prevalence of *B. cenocepacia*, *B. multivorans* is characterized by a limited person-to-person transmission and subsequently emerged as the most prevalent Bcc pathogen in many countries (Brisse *et al.*, 2004; De Boeck *et al.*, 2004; Govan *et al.*, 2007; LiPuma, 2010; Norskov-Lauritsen *et al.*, 2010; Pope *et al.*, 2010; Medina-Pascual *et al.*, 2012). The low number of epidemic outbreaks caused by *B. multivorans* (Whiteford *et al.*, 1995; Segonds *et al.*, 1999; Biddick *et al.*, 2003) and the fact that isolates from CF patients commonly represent unique strains suggest that strains are acquired from non-human sources such as the natural environment (Baldwin *et al.*, 2008). To examine the extent to which the ST of *B. multivorans*

Table 3.28: *B. multivorans*-specific COGs

COG	COG name	COG category
COG0062	NAD(P)H-hydrate repair enzyme Nnr, NAD(P)H-hydrate epimerase domain	F
COG0645	Predicted kinase	R
COG1585	Membrane protein implicated in regulation of membrane protease activity	O
COG2312	Erythromycin esterase homolog	Q
COG4121	tRNA U34 5-methylaminomethyl-2-thiouridine-forming methyltransferase MnmC	J
COG5567	Predicted small periplasmic lipoprotein YifL (function unknown)	S
COG5615	Uncharacterized membrane protein	S
COG2519	tRNA A58 N-methylase Trm61	J
COG2905	Signal-transduction protein containing cAMP-binding, CBS, and nucleotidyl-transferase domains	T
COG3059	Uncharacterized membrane protein YkgB	S
COG3095	Chromosome condensin MukBEF, MukE localization factor	D
COG3220	Uncharacterized conserved protein, UPF0276 family	S
COG4823	Abortive infection bacteriophage resistance protein	V
COG5453	Uncharacterized protein	S
COG1107	Archaea-specific RecJ-like exonuclease, contains DnaJ-type Zn finger domain	L
COG1140	Nitrate reductase beta subunit	CP
COG2180	Nitrate reductase assembly protein NarJ, required for insertion of molybdenum cofactor	CPO
COG2181	Nitrate reductase gamma subunit	CP
COG2202	PAS domain	T
COG2427	Uncharacterized conserved protein YjgD, DUF1641 family	S
COG5013	Nitrate reductase alpha subunit	CP

COG categories: J, translation, ribosomal structure and biogenesis; K, transcription; L, replication, recombination and repair; B, chromatin structure and dynamics; D, cell cycle control, cell division, chromosome partitioning; V, defense mechanisms; T, signal transduction mechanisms; M, cell wall/membrane/envelope biogenesis; N, cell motility; W, extracellular structures; U, intracellular trafficking, secretion, and vesicular transport; O, posttranslational modification, protein turnover, chaperones; X, mobilome: prophages, transposons; C, energy production and conversion; G, carbohydrate transport and metabolism; E, amino acid transport and metabolism; F, nucleotide transport and metabolism; H, coenzyme transport and metabolism; I, lipid transport and metabolism; P, inorganic ion transport and metabolism; Q, secondary metabolites biosynthesis, transport and catabolism; R, general function prediction only; S, function unknown.

isolates from CF versus environmental samples explains their genomic content and functionality, we selected four pairs of *B. multivorans* isolates for whole-genome sequencing, representing distinct STs and consisting of one CF and one environmental isolate each.

MLST is a well-established method for studying the population structure of Bcc organisms (Baldwin *et al.*, 2005; Spilker *et al.*, 2009) and Baldwin *et al.* (2008) previously reported the occurrence of *B. multivorans* STs that were globally distributed. Recently, whole-genome sequencing of *B. pseudomallei* revealed that the unexpected occurrence of two *B. pseudomallei* STs on two continents was due to homoplasy (De Smet *et al.*, 2015b). However, the present study demonstrated that the ST predicted both phylogeny and gene content of *B. multivorans* isolates (Fig. 3.4 and Supplementary Fig. 3.2) and hence corroborated the use of MLST for

epidemiological surveillance of Bcc bacteria.

The clinical isolates of ST189 and ST287 were obtained from samples of Canadian CF patients, but the environmental isolates of these STs were soil isolates from Belgium and the United Kingdom, respectively (Table 3.23). Yet, our analyses showed that, despite this transatlantic barrier, each *B. multivorans* genomic lineage was defined by its ST, harboring a highly conserved set of genes (Supplementary Fig. 3.2). Moreover, isolates belonging to the same ST were isolated up to eleven years apart (Table 3.23). Finally, searching the Bcc PubMLST database (http://pubmlst.org/bcc/) (Jolley & Maiden, 2010) for additional isolates of the studied STs (Table 3.23) revealed yet another ST189 isolate that was isolated in 2000 from an Australian CF patient. Altogether, these findings underscore the ubiquity of *B. multivorans* strains in different niches and on different continents.

To gain insight into the genome biology of *B. multivorans*, we analyzed all protein-coding genes in terms of homology, specificity, chromosome location and predicted function (i.e. COG category). Firstly, we identified orthologous genes because the conservation of genes may hold clues about which genes are essential for the species-specific lifestyle of *B. multivorans*. Secondly, we mapped the chromosome location of each CDS because the different chromosomes are associated with different gene copy numbers, mutation rates and expression levels and because the chromosomal location of a gene has an influence on its evolutionary course (Cooper *et al.*, 2010; Morrow & Cooper, 2012). Finally, we assigned each CDS to a COG to assess which biological functions were over- or underrepresented (Galperin *et al.*, 2015). A large fraction of the orthologs (72 %) was present in all nine *B. multivorans* genomes, showing that the *B. multivorans* isolates possessed a large set of genes regardless of their isolation source. Accordingly, Wolfgang *et al.* (2003) compared clinical and environmental isolates of *Pseudomonas aeruginosa*, which is also a significant CF pathogen, and demonstrated that most strains, regardless of source, possess the basic pathogenic mechanisms necessary to cause a wide variety of human infections.

The highly conserved multireplicon genomic structure found in the present study was in agreement with the general genome architecture of Bcc organisms (Agnoli *et al.*, 2012). Since primary chromosomes contain generally more core genes (Mahenthiralingam *et al.*, 2005) it was not surprising to find that C1 was enriched with core orthologs, while C2 and C3 were enriched with ST-specific orthologs (Fig. 3.7b). These ST-specific orthologs were enriched in genes involved in defense mechanisms and secondary metabolism (Table 3.26), two functional categories that are generally characterized by a large degree of strain-specificity. As shown by Cooper *et al.* (2010), multiple replicons allow for long-term segregation of genes by expression rates and dispensability. This way, secondary chromosomes might serve as evolutionary test beds and the ST-specific orthologs located on C2 and C3 are expected to evolve faster.

C1 was not only enriched with core orthologs but also with orthologs showing random, CF-only and ENV-only specificity (Fig. 3.7b) and non-orthologous CDS (Fig. 3.5b). The enrichment of C1 with random specificity orthologs may be explained by the fact that primarily C1 suffered from unclosed assemblies (Table 3.24) and annotations could be missing at contig ends. Nevertheless, these findings suggest that C1 harbors a rather large number of CDS that are non-orthologous, or CDS that are found only in a smaller subset (CF-only, ENV-only) of the *B. multivorans* genomes in the present study.

Whole-genome phylogenetic analysis (Fig. 3.4) showed that the ST287 genomic lineage was most closely related to ATCC 17616. The larger ST287 genomes (Table 3.24) harbored 539 orthologs that were involved in similar biological processes as the 176 orthologs that were specific for the other three STs (ST180, ST189 and ST650) (Fig. 3.6). These ST287-specific orthologous genes suggest an example of functional redundancy in which different homologous genes encode protein families involved in the same biological process.

Because the absence or presence of specific genes may hold clues about how *B. multivorans* differs in lifestyle and epidemiology from *B. cenocepacia*, we compared the average COG profiles of these two Bcc species. In comparison with *B. cenocepacia*, the genome of *B. multivorans* was enriched in COGs involved in translation (J) and replication (L) and deprived in COGs involved in transcription (K), which may indicate a lower adaptability of *B. multivorans* to varying environments. This different distribution in COGs involved in information storage and processing may also reflect the overall difference in genome size between these two Bcc species (Table 3.24 and Supplementary Table 3.5). Indeed, several studies (Cases *et al.*, 2003; Konstantinidis & Tiedje, 2004) previously demonstrated that the categories translation (J) and replication (L) showed a strong negative correlation with genome size, while transcription (K) showed a strong positive correlation with genome size. Similarly, Carlier *et al.* (2012; 2016) showed that the genomes of the obligate leaf nodule endosymbionts *Candidatus* Burkholderia crenata and *Candidatus* Burkholderia kirkii were smaller, enriched in COG categories J and L and deprived in COG category K when compared to free-living, facultative endophytic *Burkholderia* species. Consequently, we may expect that larger genomes require greater regulatory capacity to control their versatile metabolic capacity, as reflected by the higher number of transcriptional regulators, but that this adaptability also comes with a cost in terms of growth rate.

Next to the differences in average COG profile related to information storage and processing, our comparison also revealed that, as compared to *B. cenocepacia*, *B. multivorans* genomes contained less COGs involved in defense mechanisms (V). This finding correlated with *B. multivorans* generally being less virulent than *B. cenocepacia* (Aris *et al.*, 2001; Jones *et al.*, 2004). Similarly, Bartell *et al.* (2014) recently showed that *B. cenocepacia* produces a wider

array of virulence factors compared to *B. multivorans*. This difference in average COG profile was also reflected by the fact that *B. cenocepacia* genomes harbored several COGs involved in resistance to antimicrobial compounds (Supplementary Table 3.6). *B. multivorans* on the other hand harbored exclusive COGs that suggested an expanded functionality with regards to nitrogen metabolism (Table 3.28). Altogether, the present study did not reveal any difference in the average COG profile between *B. multivorans* and *B. cenocepacia* that could explain their difference in CF epidemiology.

In this high-throughput sequencing era it is relatively straightforward to obtain draft genome sequences to study the molecular epidemiology of bacterial pathogens (McAdam *et al.*, 2014). While short-read sequencing platforms yield draft genome assemblies at a low cost, *Burkholderia* genomes can only be fully resolved using long-read sequencing technologies such as PacBio SMRT sequencing (Shin *et al.*, 2013; Koren & Phillippy, 2015). The present study provides high-quality genome assemblies for eight *B. multivorans* isolates and the final assemblies produced closed genomes for five of the eight isolates (Table 3.24). Although the SMRT analysis software already produced high-quality assemblies there was still a need to further polish the resulting assemblies manually (see Methods section). The circular nature of the replicons in combination with the long-read sequencing technology resulted in artificial duplications, as exemplified by the fact that the *B. multivorans* genomes initially harbored five to seven rRNA operons, while they all harbored five copies after the manual curation. The rRNA copy number is generally quite stable within a species (Stoddard *et al.*, 2015) and is thus an easy quality checkpoint when evaluating the status of PacBio assemblies.

In conclusion, the present study demonstrates that the genomic structure of *B. multivorans* is highly conserved and that the ST predicts the genomic lineage. The high-quality genome assemblies provided in the present study may serve as reference genomes for future studies using transcriptomics and proteomics to try to further elucidate the epidemiology and pathogenicity of this CF pathogen.

3.4.6 Acknowledgments

We thank all strain depositors listed in Table 3.23 to make this study possible. We thank Robert Sebra from the Department of Genetics and Genomic Sciences of the Icahn School of Medicine at Mount Sinai for the PacBio sequencing service and Aurélien Carlier for his advice on the COG data analysis. CP is indebted to the Special Research Council of Ghent University.

3.4.7 Supplementary material

The *B. multivorans* genomes sequenced in this study are publicly available through the GenBank/EMBL/DDBJ accession numbers FKJT01000000, FKJS01000000, FKJU01000000, FKJP01000000, FKJV01000000, FKJW01000000, FKJX01000000 and FKJY01000000.

Supplementary Table 3.3: The frequency of orthologous versus non-orthologous CDS varies among isolates

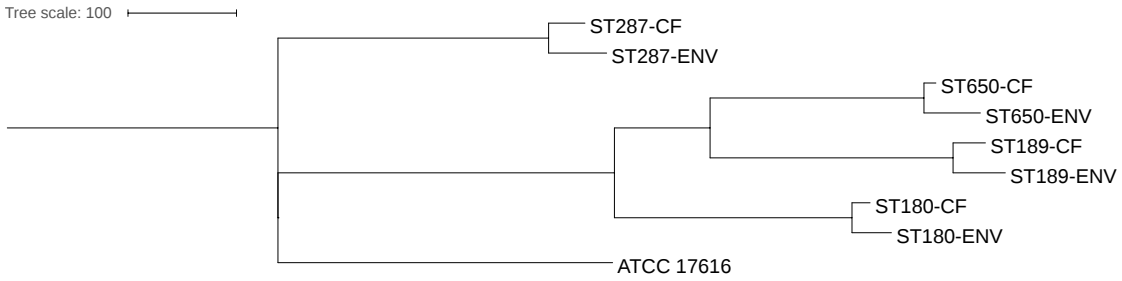
Isolate		Orthologous CDS		Non-orthologous CDS
ST180-ENV		5271 (-0.397)		523 (1.301)
ST180-CF	+	5266 (2.643)	−	285 (-8.658)
ST189-ENV	+	5144 (2.025)	−	323 (-6.632)
ST189-CF	+	5132 (2.540)	−	283 (-8.319)
ST287-ENV	+	5494 (2.589)	−	306 (-8.480)
ST287-CF		5505 (-1.668)	+	650 (5.462)
ST650-ENV	+	5275 (2.209)	−	319 (-7.235)
ST650-CF		5250 (1.795)	−	349 (-5.879)
ATCC 17616	−	4893 (-10.988)	+	1365 (35.988)

Pearson chi-square analysis testing the independence of gene conservation (orthologous vs. non-orthologous CDS) and isolate ($\chi^2(8)=1829.6$, $p<0.001$). Each cell in the contingency represents the observed frequency and standardized residual (in between brackets) and is preceded by + or − if the standardized residual is >1.96 or <-1.96, respectively, and significant at $p<0.05$.

Supplementary Table 3.4: The frequency of orthologous versus non-orthologous CDS varies among STs

ST	Orthologous CDS		Non-orthologous CDS
ST180	10537 (-0.471)		808 (1.757)
ST189	10276 (1.217)	−	606 (-4.541)
ST287	10999 (-1.473)	+	956 (5.499)
ST650	10525 (0.797)	−	668 (-2.974)

Pearson chi-square analysis testing the independence of gene conservation (orthologous vs. non-orthologous CDS) and ST ($\chi^2(3)=67.3$, $p<0.001$). Each cell in the contingency represents the observed frequency and standardized residual (in between brackets) and is preceded by + or − if the standardized residual is >1.96 or <-1.96, respectively, and significant at $p<0.05$.

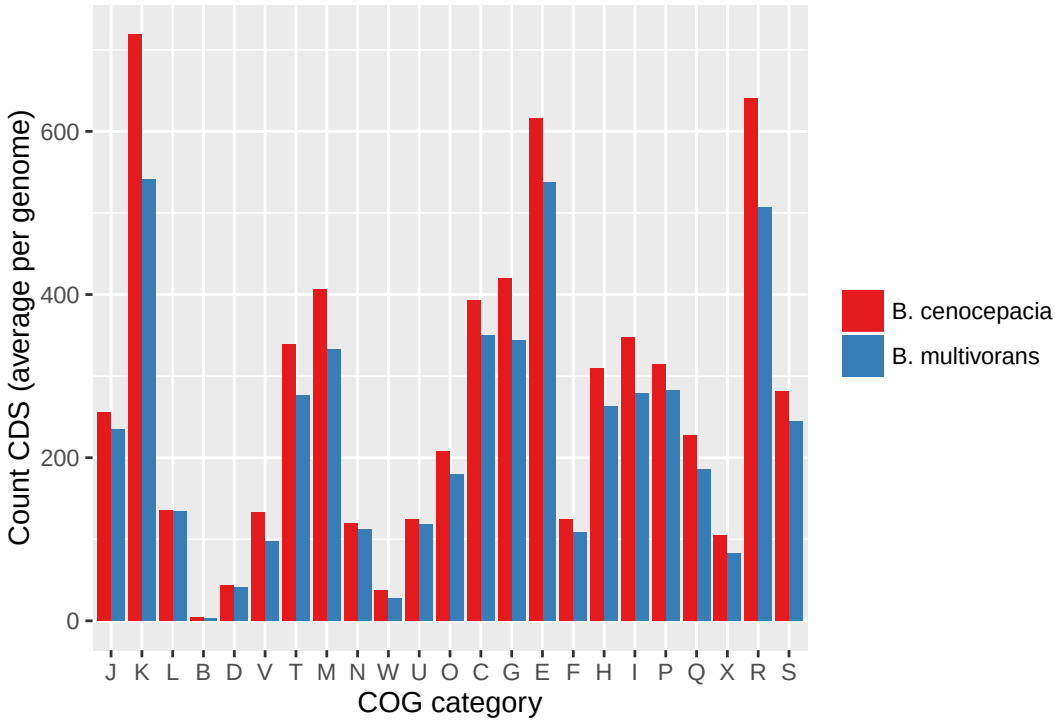


Supplementary Figure 3.2: Parsimony tree showing the relatedness of the genomes in terms of gene content. Scale bar represents number of changes of state required in each character. The tree was rooted on the branch with the largest branch length.

Supplementary Table 3.5: *B. cenocepacia* genomes included in the present study

Isolate	Genomovar	Project	Isolation source	Size (bp)	CDS
J2315	IIIA	PRJNA339	CF	8,055,782	7116
H111	IIIA	PRJNA69823	CF	7,714,893	6932
K56-2Valvano	IIIA	PRJNA62783	CF	7,750,255	6793
AU1054	IIIB	PRJNA13919	CF	7,279,116	6477
HI2424	IIIB	PRJNA13918	ENV	7,702,840	6919
MC0-3	IIIB	PRJNA17929	ENV	7,971,389	7008

CF, cystic fibrosis; ENV, environmental.



Supplementary Figure 3.3: Average COG profiles of *B. multivorans* and *B. cenocepacia*. Bar plot showing the average number of CDS per genome per COG category. COG categories: J, translation, ribosomal structure and biogenesis; K, transcription; L, replication, recombination and repair; B, chromatin structure and dynamics; D, cell cycle control, cell division, chromosome partitioning; V, defense mechanisms; T, signal transduction mechanisms; M, cell wall/membrane/envelope biogenesis; N, cell motility; W, extracellular structures; U, intracellular trafficking, secretion, and vesicular transport; O, posttranslational modification, protein turnover, chaperones; X, mobilome: prophages, transposons; C, energy production and conversion; G, carbohydrate transport and metabolism; E, amino acid transport and metabolism; F, nucleotide transport and metabolism; H, coenzyme transport and metabolism; I, lipid transport and metabolism; P, inorganic ion transport and metabolism; Q, secondary metabolites biosynthesis, transport and catabolism; R, general function prediction only; S, function unknown.

Supplementary Table 3.6: *B. cenocepacia*-specific COGs

COG categories: J, translation, ribosomal structure and biogenesis; K, transcription; L, replication, recombination and repair; B, chromatin structure and dynamics; D, cell cycle control, cell division, chromosome partitioning; V, defense mechanisms; T, signal transduction mechanisms; M, cell wall/membrane/envelope biogenesis; N, cell motility; W, extracellular structures; U, intracellular trafficking, secretion, and vesicular transport; O, posttranslational modification, protein turnover, chaperones; X, mobilome: prophages, transposons; C, energy production and conversion; G, carbohydrate transport and metabolism; E, amino acid transport and metabolism; F, nucleotide transport and metabolism; H, coenzyme transport and metabolism; I, lipid transport and metabolism; P, inorganic ion transport and metabolism; Q, secondary metabolites biosynthesis, transport and catabolism; R, general function prediction only; S, function unknown.

COG	COG name	COG category
COG0063	NAD(P)H-hydrate repair enzyme Nnr, NAD(P)H-hydrate dehydratase domain	F
COG0213	Thymidine phosphorylase	F
COG0274	Deoxyribose-phosphate aldolase	F
COG0295	Cytidine deaminase	F
COG0520	Selenocysteine lyase/Cysteine desulfurase	E
COG0530	Ca2+/Na+ antiporter	P
COG0569	Trk K+ transport system, NAD-binding component	P
COG0639	Diadenosine tetraphosphatase ApaH/serine/threonine protein phosphatase, PP2A family	T
COG0738	Fucose permease	G
COG1027	Aspartate ammonia-lyase	E
COG1056	Nicotinamide mononucleotide adenylyltransferase	H
COG1090	NAD dependent epimerase/dehydratase family enzyme	R
COG1169	Isochorismate synthase EntC	HQ
COG1368	Phosphoglycerol transferase MdoB or a related enzyme of AlkP superfamily	M
COG1501	Alpha-glucosidase, glycosyl hydrolase family GH31	G
COG1510	DNA-binding transcriptional regulator GbsR, MarR family	K
COG1513	Cyanate lyase	P
COG1586	S-adenosylmethionine decarboxylase or arginine decarboxylase	E
COG1621	Sucrose-6-phosphate hydrolase SacC, GH32 family	G
COG1647	Esterase/lipase	Q
COG1708	Predicted nucleotidyltransferase	R
COG1944	Ribosomal protein S12 methylthiotransferase accessory factor YcaO	J
COG1957	Inosine-uridine nucleoside N-ribohydrolase	F
COG1972	Nucleoside permease NupC	F
COG2116	Formate/nitrite transporter FocA, FNT family	P
COG2140	Oxalate decarboxylase/archaeal phosphoglucose isomerase, cupin superfamily	G
COG2206	HD-GYP domain, c-di-GMP phosphodiesterase class II (or its inactivated variant)	T
COG2234	Zn-dependent amino- or carboxypeptidase, M28 family	O
COG2258	Uncharacterized conserved protein YiiM, contains MOSC domain	S
COG2509	Uncharacterized FAD-dependent dehydrogenase	R
COG2602	Beta-lactamase class D	V
COG2936	Predicted acyl esterase	R
COG2974	DNA recombination-dependent growth factor C	L
COG3011	Predicted thiol-disulfide oxidoreductase YuxK, DCC family	R
COG3147	Cell division protein DedD (periplasmic protein involved in septation)	D

Continued on next page

Continued from previous page

COG	COG name	COG category
COG3177	Fic family protein	K
COG3208	Surfactin synthase thioesterase subunit	Q
COG3227	Zn-dependent metalloprotease	O
COG3238	Uncharacterized membrane protein YdcZ, DUF606 family	S
COG3300	MHYT domain, NO-binding membrane sensor	T
COG3393	Predicted acetyltransferase, GNAT family	R
COG3422	Uncharacterized conserved protein YegP, UPF0339 family	S
COG3437	Response regulator c-di-GMP phosphodiesterase, RpfG family, contains REC and HD-GYP domains	T
COG3467	Nitroimidazol reductase NimA or a related FMN-containing flavoprotein, pyridoxamine 5-phosphate oxidase superfamily	V
COG3510	Cephalosporin hydroxylase	V
COG3525	N-acetyl-beta-hexosaminidase	G
COG3530	Uncharacterized conserved protein, DUF3820 family	S
COG3537	Putative alpha-1,2-mannosidase	G
COG3602	Uncharacterized protein	S
COG3637	Opacity protein and related surface antigens	M
COG3670	Carotenoid cleavage dioxygenase or a related enzyme	Q
COG3766	Uncharacterized membrane protein YjfL, UPF0719 family	S
COG3963	Phospholipid N-methyltransferase	I
COG3979	Chitodextrinase	G
COG4117	Thiosulfate reductase cytochrome b subunit	P
COG4260	Membrane protease subunit, stomatin/prohibitin family, contains C-terminal Zn-ribbon domain	O
COG4262	Predicted spermidine synthase with an N-terminal membrane domain	R
COG4285	Uncharacterized conserved protein , conains N-terminal glutamine amidotransferase (GATase1)-like domain	R
COG4448	L-asparaginase II	E
COG4452	Inner membrane protein involved in colicin E2 resistance	V
COG4551	Predicted protein tyrosine phosphatase	R
COG4589	Predicted CDP-diglyceride synthetase/phosphatidate cytidylyltransferase	R
COG4682	Uncharacterized membrane protein YiaA	S
COG4693	Oxidoreductase (NAD-binding), involved in siderophore biosynthesis	P
COG4773	Outer membrane receptor for ferric coprogen and ferric-rhodotorulic acid	P
COG4922	Predicted SnoaL-like aldol condensation-catalyzing enzyme	R
COG5002	Signal transduction histidine kinase	T
COG5460	Uncharacterized conserved protein, DUF2164 family	S
COG5472	Predicted small integral membrane protein	S
COG5500	Uncharacterized membrane protein	S
COG5528	Uncharacterized membrane protein	S
COG5612	Uncharacterized membrane protein	S

4 | New insights into the taxonomy of *Burkholderia cepacia* complex bacteria

During our ongoing survey of the diversity of Bcc bacteria, several isolates of clinical and environmental origin were found to represent unique lineages within the Bcc. These groups of strains could be distinguished both genotypically and phenotypically from the present formally named Bcc species and could therefore be formally classified. Section 4.1 presents a polyphasic taxonomic study that led to the formal classification of *Burkholderia pseudomultivorans* sp. nov. and a re-evaluation of the 3 % concatenated allele sequence divergence threshold (Vanlaere *et al.*, 2009) for species delineation within the Bcc. Section 4.2 presents a polyphasic taxonomic study of several Bcc isolates that were isolated during environmental surveys and the formal classification of *Burkholderia stagnalis* sp. nov. and *Burkholderia territorii* sp. nov.

4.1 *Burkholderia pseudomultivorans* sp. nov., a novel *Burkholderia cepacia* complex species from human respiratory samples and the rhizosphere

Redrafted from: Charlotte Peeters, James E.A. Zlosnik, Theodore Spilker, Trevor J. Hird, John J. LiPuma and Peter Vandamme (2013). *Burkholderia pseudomultivorans* sp. nov., a novel *Burkholderia cepacia* complex species from human respiratory samples and the rhizosphere. Systematic and Applied Microbiology, 36, 483–489.

Author contributions: CP and PV conceived the study and wrote the manuscript. CP performed all experiments except the biochemical characterization and analyzed the data. JZ and TH performed the biochemical characterization. TS and JL isolated the strains. JZ, TS, TH and JL proofread the manuscript.

4.1.1 Abstract

Eleven *B. cepacia*-like isolates of human clinical and environmental origin were examined by a polyphasic approach including *recA* and 16S rRNA sequence analysis, MLSA, DNA base content determination, fatty acid methyl ester analysis, and biochemical characterization. The results of this study demonstrate that these isolates represent a novel species within the Bcc for which we propose the name *Burkholderia pseudomultivorans*. The type strain is strain LMG 26883^T (=CCUG 62895^T). *B. pseudomultivorans* can be differentiated from other Bcc species by *recA* gene sequence analysis, MLSA, and several biochemical tests including growth at 42 °C, acidification of sucrose and adonitol, lysine decarboxylase and β-galactosidase activity, and esculin hydrolysis.

4.1.2 Introduction

The Bcc is a versatile group of closely related bacteria that are ubiquitous in nature (Coenye & Vandamme, 2003). Bcc species have versatile lifestyles and exhibit conflicting biological features: some Bcc strains show biotechnological potential for biocontrol, bioremediation and plant growth promotion (Parke & Gurian-Sherman, 2001), whereas other strains are pathogens for plants and immunocompromised individuals such as persons with CF (Mahenthiralingam *et al.*, 2008). The Bcc currently comprises 17 species (Vanlaere *et al.*, 2009) which exhibit a high degree of 16S rRNA (98-100 %) and *recA* (94-95 %) gene sequence similarity, and moderate levels of DDH (30-50 %) (Coenye *et al.*, 2001d). In contrast to the 16S rRNA

gene, which has only limited taxonomic resolution in the Bcc, the *recA* gene mostly provides the discriminatory power needed for identification of Bcc species (Mahenthiralingam *et al.*, 2000a). The use of multilocus sequence-based approaches as powerful taxonomic tools for both identification and typing (Baldwin *et al.*, 2005; Spilker *et al.*, 2009) was shown valid for the differentiation of Bcc species as well (Vanlaere *et al.*, 2009). Vanlaere *et al.* (2009) compared the average concatenated allele sequence divergence within and between established Bcc species as delineated by DDH studies, and demonstrated that the 3 % concatenated allele sequence divergence level can be used as a threshold value for species delineation within this complex, thus replacing DDH experiments.
In an ongoing survey of the biodiversity of *B. cepacia*-like organisms, 11 isolates from human clinical and environmental origin represented a unique *recA* lineage within the Bcc. The aim of the present study was to examine the taxonomic position of these *B. cepacia*-like isolates using a polyphasic approach.

4.1.3 Materials and methods

4.1.3.1 Bacterial strains and growth conditions

The eleven isolates are listed in Table 4.1. Reference strains of other Bcc species were described previously (Mahenthiralingam *et al.*, 2000b; Coenye *et al.*, 2003; Vanlaere *et al.*, 2008a, 2009). Strains were grown aerobically on Tryptone Soya Agar (Oxoid) and incubated at 28 °C or 37 °C. Cultures were preserved in MicroBank^{TM} vials at −80 °C.

4.1.3.2 DNA preparation

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

4.1.3.3 *recA* gene sequence analysis

The nearly complete sequence of the *recA* gene was amplified by PCR as described previously (Mahenthiralingam *et al.*, 2000a) with forward primer 5' GAT AGC AAG AAG GGC TCC 3' and reverse primer 5' CTC TTC TTC GTC CAT CGC CTC 3' (Baldwin *et al.*, 2005). Sequence analysis was performed with an Applied Biosystems 3130xl Genetic Analyzer and protocols of the manufacturer using the BigDye Terminator Cycle Sequencing Ready kit. The sequencing primers were 5' GAT AGC AAG AAG GGC TCC 3', 5' CTC TTC TTC GTC CAT CGC CTC 3', 5' TGA CCG CCG AGA AGA GCA A 3' and 5' GAC CGA GTC GAT GAC

Table 4.1: Studied isolates, showing their source, ST and allelic profile

Strain	Other strain designations	Source (country, year of isolation)	Depositor
LMG 26883^T	AU3207^T, CCUG 62895^T	CF sputum (USA, 2001)	Own isolate
R-48991	AU4855	CF sputum (USA, 2002)	Own isolate
R-48993	AU6557	CF sputum (USA, 2003)	Own isolate
R-48994	AU12399	CF sputum (USA, 2006)	Own isolate
R-48995	AU19682	CF sputum (USA, 2010)	Own isolate
R-48996	AU20760	Non-CF sputum (USA, 2010)	Own isolate
LMG 16669	J2541, BPC113	Rhizosphere (UK)	J.R.W. Govan
LMG 26914	270t1	CF (Brazil, 1997)	S. Canani
R-9705	41C029	CF (Netherlands)	S. Brisse
LMG 26915	273	CF (Germany)	I. Steinmetz
R-24200	AU6112	CF sputum (USA, 2003)	Own isolate

Strain	ST	Allelic profile						
		atpD	***gltB***	***gyrB***	***recA***	***lepA***	***phaC***	***trpB***
LMG 26883^T	536	150	204	205	250	171	163	304
R-48991	-	-	-	-	-	-	-	-
R-48993	479	123	273	397	176	112	125	270
R-48994	-	-	-	-	-	-	-	-
R-48995	732	310	356	536	210	377	182	363
R-48996	-	-	-	-	-	-	-	-
LMG 16669	755	158	271	305	176	190	125	149
LMG 26914	393	158	238	324	210	237	182	224
R-9705	536	150	204	205	250	171	163	304
LMG 26915	731	158	356	535	210	377	182	364
R-24200	517	158	271	219	176	112	142	149

LMG, BCCM/LMG Bacteria Collection, Laboratory of Microbiology, Ghent University, Ghent, Belgium; CF, isolated from cystic fibrosis patient; ENV, environmental isolate.

GAT 3' (Baldwin *et al.*, 2005). Sequence assembly was performed using BioNumerics v5.10 (Applied Maths). Sequences (929-1050 bp) were aligned based on amino acid sequences using Muscle (Edgar, 2004) in MEGA5 (Tamura *et al.*, 2011). Phylogenetic analysis was conducted in MEGA5 (Tamura *et al.*, 2011). All positions containing gaps and missing data were eliminated, resulting in a total of 860 positions in the final dataset.

4.1.3.4 16S rRNA gene sequence analysis

The nearly complete sequences of the 16S rRNA gene of strains LMG 26883^T and LMG 26915 were amplified by PCR using the conserved primers 5' AGA GTT TGA TCC TGG CTC AG 3' (positions 8-27) and 5' AAG GAG GTG ATC CAG CCG CA 3' (positions 1541-1522). Sequence analysis and assembly were performed as described above for the *recA* gene. The sequencing primers are those described by Coenye *et al.* (1999a), except that BKL1 (position 536-516) was used instead of pD (positions 536-519). The mothur software package (Schloss

et al., 2009) was used to align the 16S rRNA gene sequences of strains LMG 26883^T and LMG 26915 (both 1485 bp) with those of type strains of established Bcc and other *Burkholderia* species (1124-1600 bp) against the Silva reference database (www.arb-silva.de). Uncorrected pairwise distances were calculated using MEGA5 (Tamura *et al.*, 2011). All ambiguous positions were removed for each sequence pair, resulting in a total of 1134 positions in the final dataset.

4.1.3.5 MLSA analysis

MLSA analysis was based on the method described by Spilker *et al.* (2009) with small modifications. The 25 µl PCR reaction was slightly modified and consisted of 1x PCR buffer (Qiagen), 1 U of Taq polymerase (Qiagen), 250 µM of each dNTP (Applied Biosystems), 1x Q-solution (Qiagen) and 0.4 µM (*gltB*, *trpB*, *atpD*, *gyrB*), 1 µM (*recA*, *lepA*) or 2 µM (*phaC*) of each primer. Sequence analysis and assembly were performed as described above for the *recA* gene. A phylogenetic tree of the concatenated sequences (2773 bp) of seven housekeeping gene fragments [*atpD* (443 bp), *gltB* (400 bp), *gyrB* (454 bp), *recA* (393 bp), *lepA* (397 bp), *phaC* (385 bp) and *trpB* (301 bp)] was constructed using MEGA5 (Tamura *et al.*, 2011). The mean number of nucleotide substitutions per site (i.e. the percentage of divergence of concatenated allele sequences) between populations was calculated using the program DnaSP v5.10 (Librado & Rozas, 2009) based on the Jukes-Cantor method (Jukes & Cantor, 1969). Nucleotide sequences of each allele, allelic profiles and sequence types for all isolates from the present study are available on the Bcc PubMLST website (http://pubmlst.org/bcc/) (Jolley *et al.*, 2004).

4.1.3.6 Determination of the DNA base composition

DNA was enzymatically degraded into nucleosides as described by Mesbah *et al.* (1989). The nucleoside mixture obtained was then separated using a Waters Breeze HPLC system and XBridge Shield RP18 column thermostabilized at 37 °C. The solvent was 0.02 M $NH_4H_2PO_4$ (pH 4.0) with 1.5 % (v/v) acetonitrile. Non-methylated lambda phage (Sigma) and *Escherichia coli* LMG 2093 DNA were used as calibration reference and control, respectively.

4.1.3.7 *Burkholderia multivorans*-specific PCR

B. multivorans-specific *recA* and 16S rRNA PCR assays were performed on alkaline lysates as described previously (LiPuma *et al.*, 1999; Mahenthiralingam *et al.*, 2000a).

4.1.3.8 Fatty acid methyl ester analysis

After a 24 h incubation period at 28 °C on Tryptone Soya Agar (BD), a loopful of well-grown cells was harvested and fatty acid methyl esters were prepared, separated and identified using the Microbial Identification System (Microbial ID) as described previously (Vandamme *et al.*, 1992).

4.1.3.9 Biochemical characterization

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

4.1.4 Results

4.1.4.1 *recA* gene sequence analysis

The *recA* sequences of the 11 isolates showed greater than 94 % similarity towards established Bcc species (data not shown). Phylogenetic analysis demonstrated that they constituted one cluster,consisting of two subclusters, designated recA-I and recA-II, and supported by bootstrap values of 69 % and 99 %, respectively (Fig. 4.1).

4.1.4.2 16S rRNA gene sequence analysis

The 16S rRNA sequences of representatives of both *recA* subclusters (LMG 26883^T for recA-I and LMG 26915 for recA-II) showed 99.2 % similarity. Pairwise comparison of these sequences with those of type strains of other Bcc species revealed similarity levels between 98.3 % and 99.8 % (data not shown). Similarity levels towards *B. glumae* and *B. gladioli* were also in the range of 98-99 %, and similarity levels towards other *Burkholderia* species were below 96 % (data not shown).

4.1.4.3 MLSA analysis

Eight isolates (LMG 16669, LMG 26883^T, LMG 26914, LMG 26915, R-9705, R-24200, R-48993, R-48995) were analyzed by MLSA. Phylogenetic analysis of concatenated allele sequences demonstrated that these isolates constituted one cluster within the Bcc, supported by a bootstrap value of 97 % (Fig. 4.2); there was no subdivision in two distinct lineages as observed by *recA* sequence analysis. MLSA data were also used to assign allele types, allelic profiles and STs: the eight isolates were resolved into seven STs, with LMG 26883^T and R-9705 (isolated from CF sputum samples in the USA and the Netherlands, respectively) showing the same ST (Table 4.1).

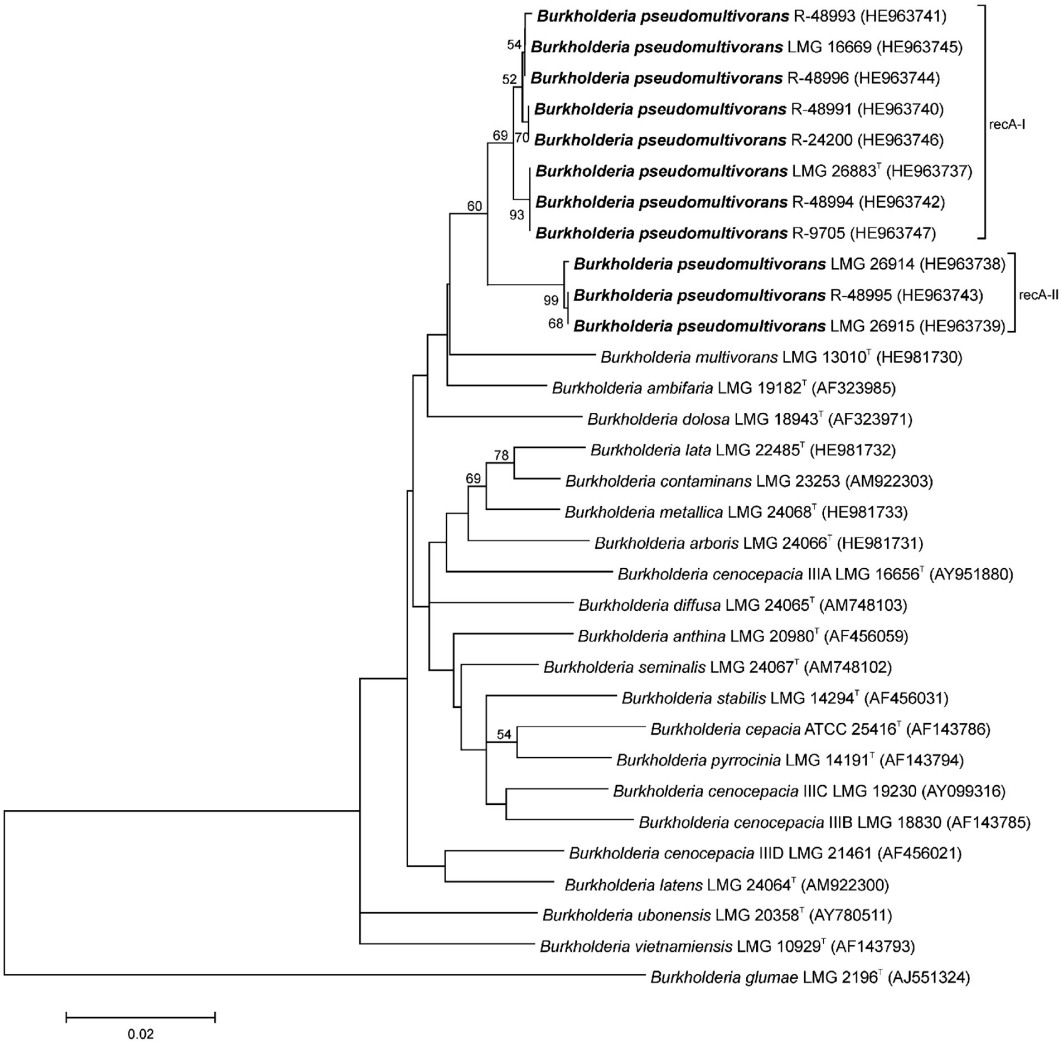


Figure 4.1: Phylogenetic tree based on *recA* gene sequences of established Bcc species and *B. pseudomultivorans* isolates. The bootstrap consensus tree, inferred from 1000 replicates, was constructed using the maximum likelihood method based on the general time reversible model. Branches corresponding to partitions reproduced in less than 50 % bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches if greater than 50 %. A discrete gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.1657)) and allowed for some sites to be evolutionarily invariable ([+I], 57.4297 % sites). The *recA* sequence of *B. glumae* LMG 2196^{T} was used as outgroup. The scale bar indicates the number of substitutions per site.

The intra- and inter-taxon divergence was calculated for the novel taxon and for the other Bcc species (Supplementary Fig. 4.1). For each established Bcc species, all STs were exported from the Bcc MLST website (August 10, 2012). Accuracy of species designations was evaluated by phylogenetic analysis of concatenated MLST loci (data not shown) to verify that the species designation of each ST corresponded with its phylogenetic position. For the novel taxon the concatenated allele sequence divergence was 1.3±0.2 %,and the average divergence towards its nearest neighbor, *B. multivorans*, was 4.6±0.2 %. For all established Bcc species the average within-group divergence remained below the 3 % cut-off value as reported by Vanlaere *et al.* (2009).

4.1.4.4 DNA base composition

The G+C content of both strains LMG 26883^T (recA-I) and LMG 26915 (recA-II) was 67 mol%.

4.1.4.5 *B. multivorans*-specific PCR

All 10 isolates tested (strain LMG 16669 was not examined) yielded an amplicon of the expected size in both the *recA* and 16S rRNA-based *B. multivorans*-specific PCR assays (data not shown), thus yielding false positive results.

4.1.4.6 Cellular fatty acid analysis

The following fatty acid components were present in all 11 isolates: $C_{14:0}$ (4.5±0.5 %), $C_{16:0}$ (20.9±2.8 %), $C_{16:0}$ 2-OH (1.1±0.3 %), $C_{16:0}$ 3-OH (6.7±1.1 %), $C_{16:1}$ 2-OH (1.6±0.7 %), $C_{17:0}$ cyclo (8.0±4.4 %), $C_{18:1}$ 2-OH (1.8±0.9 %), $C_{18:1}$ ω7c (25.2±6.2 %), $C_{19:0}$ cyclo ω8c (4.2±2.2 %), summed feature 2 (comprising $C_{14:0}$ 3-OH, $C_{16:1}$ ISO I, an unidentified fatty acid with equivalent chain length of 10.928 or $C_{12:0}$ ALDE or any combination of these fatty acids) (8.6±1.8 %), summed feature 3 (comprising $C_{16:1}$ ω7c or C_{15} ISO 2-OH or both) (14.1±4.1 %). Some fatty acid components were present in different quantities in recA-I and recA-II isolates: $C_{17:0}$ cyclo (9.2±4.6 % and 4.7±0.5 %, respectively), $C_{19:0}$ cyclo ω8c (5.0±2.0 % and 2.1±0.8 %, respectively) and summed feature 3 (12.8±2.6 % and 17.6±5.9 %, respectively). Summed features 3 and 7 most probably correspond to $C_{16:1}$ ω7c and $C_{18:1}$ ω7c, respectively, as these fatty acids have been reported in *Burkholderia* species (Stead, 1992).

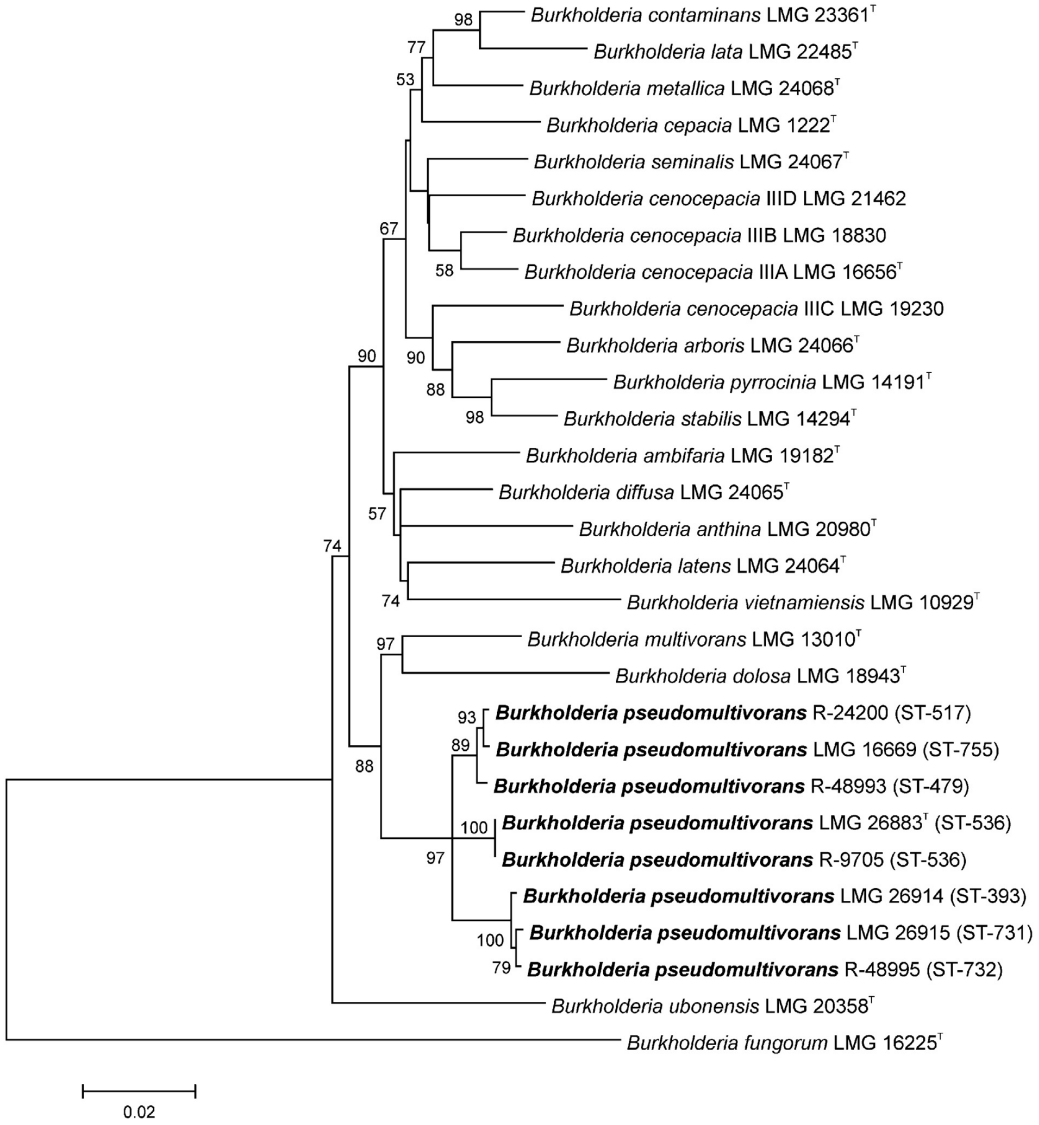


Figure 4.2: Phylogenetic tree based on the concatenated sequences (2773 bp) of seven housekeeping gene fragments [*atpD* (443 bp), *gltB* (400 bp), *gyrB* (454 bp), *recA* (393 bp), *lepA* (397 bp), *phaC* (385 bp) and *trpB* (301 bp)] of established Bcc species and *B. pseudomultivorans* strains. The bootstrap consensus tree, inferred from 1000 replicates, was constructed using the maximum likelihood method based on the general time reversible model. Branches corresponding to partitions reproduced in less than 50 % bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches if greater than 50 %. A discrete gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.2232)) and allowed for some sites to be evolutionarily invariable ([+I], 52.6522 % sites). The sequence of *B. fungorum* LMG 16225^{T} was used as outgroup. The scale bar indicates the number of substitutions per site.

4.1.4.7 Biochemical characterization

Biochemical characteristics were determined for nine isolates representing 6 STs (LMG 16669 and R-9705 were not examined) and are listed below. Characteristics that differed between recA-I and recA-II isolates were acidification of adonitol and esculin hydrolysis (both present in recA-I isolates and absent in recA-II isolates). In contrast to the other isolates, strain R-48995 demonstrated only slow growth on BSA and was non-mucoid (data not shown).

4.1.5 Discussion

In an ongoing survey of the biodiversity of *B. cepacia*-like bacteria, 11 isolates represented a unique *recA* lineage (Fig. 4.1). The *recA* and 16S rRNA gene sequence similarity levels towards other Bcc strains suggested that these isolates represented a distinct Bcc taxon. The DNA G+C content also corresponded with that of other Bcc species (66-69 mol%) (Vandamme & Dawyndt, 2011). We used MLSA analysis as a validated tool (Vanlaere *et al.*, 2009) for replacing DDH experiments to confirm that these strains represent a novel species within the Bcc. In this process, the 3 % concatenated allele sequence divergence threshold value for species delineation was re-evaluated based on a considerably larger set of STs (485 versus 349 STs in the original study). The intra- and inter-taxon divergence of concatenated allele sequences was calculated based on the expanded dataset and the 3 % threshold value for species delineation was further substantiated (Supplementary Fig. 4.1). This analysis also confirmed that the novel isolates constituted a single cluster within the Bcc (Fig. 4.2), of which the intraspecies divergence was below 3 %, while the average divergence to its nearest neighbor *B. multivorans* was above the 3 % threshold value. We therefore conclude that the eleven isolates represent a single novel Bcc species.

Biochemically, this novel species is similar to other Bcc bacteria and *B. gladioli*. However, particularly growth at 42 °C, acidification of sucrose and adonitol, lysine decarboxylase activity, β-galactosidase activity and esculin hydrolysis are helpful to enable its differentiation (Table 4.2). It can also be differentiated from *B. multivorans*, its nearest phylogenetic neighbor (Fig. 4.2 and Supplementary Fig. 4.1): whereas *B. multivorans* is typically sucrose negative and adonitol positive, most of these isolates are either sucrose positive and adonitol negative (n=3) or sucrose positive and adonitol positive (n=4). The most discriminating fatty acid components are $C_{16:0}$, $C_{17:0}$ cyclo, $C_{19:0}$ cyclo ω8c, summed feature 3 and summed feature 7 (Table 4.2), but Bcc species or *B. gladioli* cannot unequivocally be distinguished using their whole-cell fatty acid profiles (Table 4.2).

In conclusion, we demonstrated that eleven human clinical and environmental isolates represent a novel species within the Bcc, with *B. multivorans* as its nearest phylogenetic neighbor. This

species can be identified unambiguously by *recA* gene sequencing and MLSA analysis but would be misidentified as *B. multivorans* using *B. multivorans*-specific *recA* and 16S rRNA PCR assays (LiPuma *et al.*, 1999; Mahenthiralingam *et al.*, 2000a). This novel species consists of two *recA* subgroups which are biochemically different but which cannot be distinguished using MLSA analysis. We therefore propose to formally classify these eleven isolates as the novel species *Burkholderia pseudomultivorans* sp. nov. with strain LMG 26883^T as the type strain.

4.1.5.1 Description of *Burkholderia pseudomultivorans* sp. nov.

B. pseudomultivorans: pseu.do.mul.ti.vo'rans. Gr. adj. *pseudes*, false; N.L. part. adj. *multivorans*, a bacterial specific epithet; N.L. part. adj. *pseudomultivorans*, the false (*Burkholderia*) *multivorans*, referring to the fact that isolates of this species are very similar to *B. multivorans* isolates.

Cells are Gram-negative, non-sporulating rods. All strains grow on BSA, BCSA and MacConkey agar. Growth is observed at 42 °C. No pigment production. Oxidase, β-galactosidase and lysine decarboxylase activity are present. No ornithine decarboxylase or β-haemolysis activity, and no gelatin liquefaction. Strain-dependent reactions for nitrate reduction and esculin hydrolysis. Acidification of glucose, maltose, lactose and xylose, but strain-dependent reactions for sucrose and adonitol. The G+C content is 67 mol%. The following fatty acids are present in all strains: $C_{14:0}$, $C_{16:0}$, $C_{16:0}$ 2-OH, $C_{16:0}$ 3-OH, $C_{16:1}$ 2-OH, $C_{17:0}$ cyclo, $C_{18:1}$ 2OH, $C_{18:1}$ ω7c, $C_{19:0}$ cyclo ω8c, and summed features 2 and 3. Strains have been isolated from human respiratory specimens and from rhizosphere soil.

The type strain is LMG 26883^T (=CCUG 62895^T). Phenotypic characteristics of the type strain are the same as those described above for the species. In addition, the type strain acidifies adonitol but not sucrose, reduces nitrate and hydrolyzes esculin. Its G+C content is 67 %.

4.1.6 Acknowledgments

We thank all strain depositors who contributed to this study.

Table 4.2: Phenotypic characteristics useful for differentiation of *B. pseudomultivorans* from members of the *B. cepacia* complex and *B. gladioli*

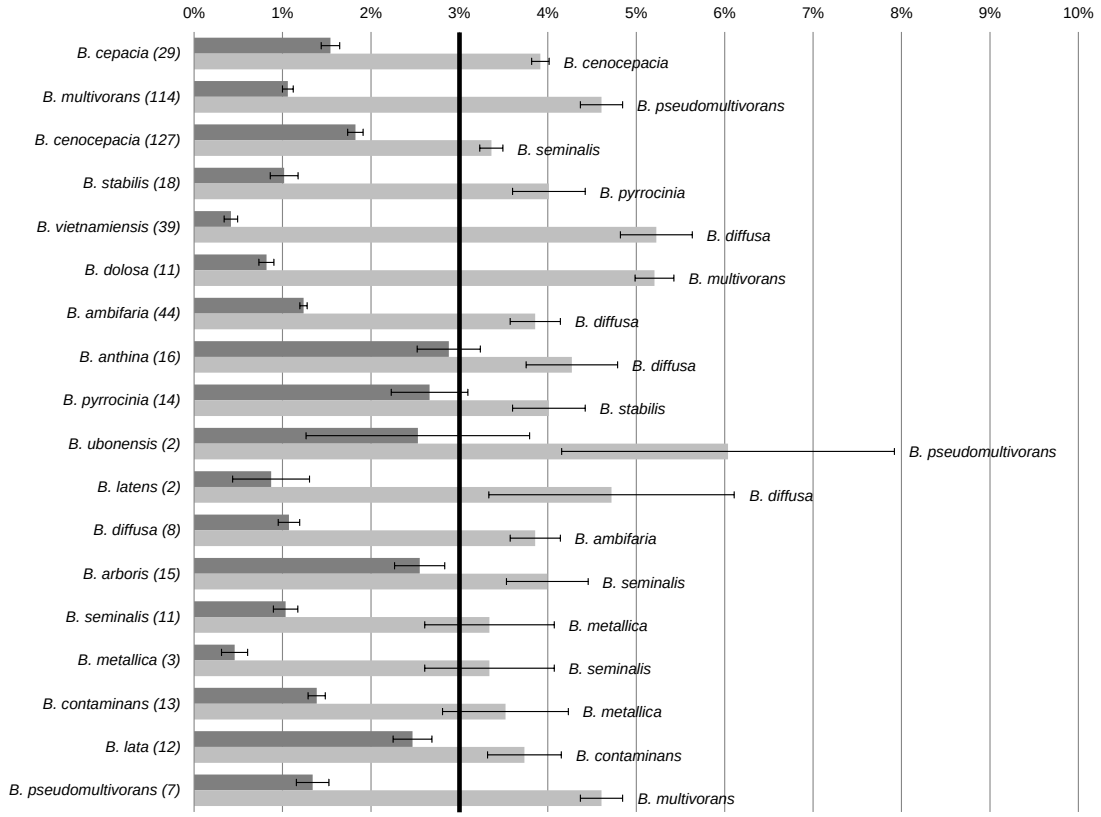
	B. pseudomultivorans	***B. cepacia***	***B. multivorans***	***B. cenocepacia***	***B. stabilis***	***B. vietnamiensis***	***B. dolosa***	***B. ambifaria***	***B. anthina***
Growth at 42 °C	+	v	+	v	−	+	+	v	v
Acidification of									
Sucrose	7(−)	+	−	+	−	+	−	+	v
Adonitol	6(+)	v	+	v	+	−	+	+	v
Lysine decarboxylation	+	+	v	+	+	+	−	+	v
β-galactosidase	+	+	+	+	−	+	+	+	v
Esculin hydrolysis	6(+)	v	−	v	−	−	−	v	−
Fatty acid content									
$C_{16:0}$	20.9±2.8	26.8±3.5	28.9±3.5	26.1±3.5	25.6±4.3	19.5±2.5	26.1±1.4	26.3±2.0	28.4±0.9
$C_{17:0}$ cyclo	8.0±4.4	17.9±3.6	18.2±4.9	21.3±4.2	17.8±3.3	14.0±4.9	16.0±2.2	11.3±4.9	4.6±2.5
$C_{19:0}$ cyclo ω8c	4.2±2.2	12.5±4.7	9.7±3.3	13.4±4.9	15.3±4.2	5.8±2.5	14.8±3.7	4.8±2.8	1.3±0.4
Summed feature 3[a]	14.1±4.1	4.6±2.5	6.5±1.5	3.0±1.0	2.7±0.8	9.8±2.5	4.2±2.0	11.0±3.5	17.1±1.1
Summed feature 7[a]	25.2±6.2	11.8±3.5	11.5±4.8	9.3±4.4	10.7±1.8	19.7±5.5	9.9±4.3	25.6±6.3	27.6±5.2

	B. pyrrocinia	***B. ubonensis***	***B. latens***	***B. diffusa***	***B. arboris***	***B. seminalis***	***B. metallica***	***B. contaminans***	***B. lata***	***B. gladioli***
Growth at 42 °C	v	v	+	v	v	+	+	v	−	−
Acidification of										
Sucrose	+	+	+	+	+	+	+	+	v	−
Adonitol	+	−	+	v	+	+	+	+	v	+
Lysine decarboxylation	+	−	+	+	v	v	+	+	+	−
β-galactosidase	+	−	+	+	+	+	+	+	v	+
Esculin hydrolysis	−	−	−	−	−	v	+	v	v	v
Fatty acid content										
$C_{16:0}$	22.9	16.6	17.8	21.2	17.1	28.1	26.4	21.9	16.0	29.0±1.1
$C_{17:0}$ cyclo	13.5	5.7	12.6	1.59	5.8	11.7	10.4	5.2	4.9	17.2±1.9
$C_{19:0}$ cyclo ω8c	4.7	2.4	9.5	0.4	1.7	1.6	3.0	1.8	1.6	10.0±2.7
Summed feature 3[a]	7.3	16.2	10.3	21.2	15.1	21.1	19.7	16.8	17.7	5.1±2.3
Summed feature 7[a]	22.2	33.4	29.9	35.7	31.4	13.2	19.2	29.3	27.7	11.0±3.7

For characteristics other than fatty acid content data, data were taken from the present study, from Henry *et al.* (2001), Vanlaere *et al.* (2008a, 2009), or from unpublished data from D. Henry and D.P. Speert. +, >90 % of all isolates positive; v, 10-90 % positive; −, <10 % of strains positive. For *B. pseudomultivorans* (n=9), the number of positive reactions is indicated for strain-dependent reactions and the reaction of the type strain is given in parentheses. Data for fatty acid content are percentages of whole-cell fatty acid content and were taken from the present study or from Coenye *et al.* (2001c) and Vandamme *et al.* (1997, 2000, 2002b). [a]Summed feature 3 comprises $C_{16:1}$ ω7c or C_{15} ISO 2-OH or both. Summed feature 7 comprises $C_{18:1}$ ω7c, $C_{18:1}$ ω9t, $C_{18:1}$ ω12t or any combination of these fatty acids.

4.1.7 Supplementary material

The GenBank/EMBL/DDBJ accession numbers for the 16S rRNA gene sequences of strains LMG 26883^T and LMG 26915 are HE962386 and HE962387, respectively. The GenBank/EMBL/DDBJ accession numbers for the *recA* gene sequences of strains LMG 26883^T, LMG 26914, LMG 26915, R-48991, R-48993, R-48994, R-48995, R-48996, LMG 16669, R-24200, R-9705, LMG 13010^T, LMG 24066^T, LMG 22485^T and LMG 24068^T are HE963737, HE963738, HE963739, HE963740, HE963741, HE963742, HE963743, HE963744, HE963745, HE963746, HE963747, HE981730, HE981731, HE981732 and HE981733, respectively.



Supplementary Figure 4.1: Nucleotide divergence of concatenated MLST loci. Dark grey bars represent within group divergence, light grey bars represent the divergence between a Bcc species and its highest similarity neighbor (listed to the right). The number of STs included in the calculations for each species is given in parentheses. For each bar, the SD is indicated by error bars. The 3 % divergence threshold value is marked (Vanlaere *et al.*, 2009).

4.2 *Burkholderia stagnalis* sp. nov. and *Burkholderia territorii* sp. nov., two novel *Burkholderia cepacia* complex species from environmental and human sources

Redrafted from: Birgit De Smet, Mark Mayo, Charlotte Peeters, James E.A. Zlosnik, Theodore Spilker, Trevor J. Hird, John J. LiPuma, Timothy J. Kidd, Mirjam Kaestli, Jennifer L. Ginther, David M. Wagner, Paul Keim, Scott C. Bell, Jan A. Jacobs, Bart J. Currie and Peter Vandamme (2015). *Burkholderia stagnalis* sp. nov. and *Burkholderia territorii* sp. nov., two novel *Burkholderia cepacia* complex species from environmental and human sources. International Journal of Systematic and Evolutionary Microbiology, 65, 2265-2271.

Author contributions: BDS, CP and PV conceived the study. BDS and PV wrote the manuscript. BDS and CP performed all experiments except the biochemical characterization and analyzed the data. JZ and TH performed the biochemical characterization. TS, JL, TK, SB and BC isolated the strains. CP, JZ, TS, TH and JL proofread the manuscript.

4.2.1 Abstract

Nine Bcc bacteria were isolated during environmental surveys for the ecological niche of *B. pseudomallei*, the aetiological agent of melioidosis, in the Northern Territory of Australia. They represented two multilocus sequence analysis-based clusters, referred to as Bcc B and Bcc L. Three additional environmental and clinical Bcc B isolates were identified upon deposition of the sequences in the PubMLST database. Analysis of the concatenated nucleotide sequence divergence levels within both groups (1.4 and 1.9 %, respectively) and towards established Bcc species (4.0 and 3.9 %, respectively) demonstrated that the two taxa represented novel Bcc species. All 12 isolates were further characterized using 16S rRNA and *recA* gene sequence analysis, RAPD analysis, DNA base content determination, fatty acid methyl ester analysis and biochemical profiling. Analysis of *recA* gene sequences revealed a remarkable diversity within each of these taxa, but, together, the results supported the affiliation of the two taxa to the Bcc. Bcc B strains can be differentiated from most other Bcc members by the assimilation of maltose. Bcc L strains can be differentiated from other Bcc members by the absence of assimilation of N-acetylglucosamine. The names *Burkholderia stagnalis* sp. nov. with type strain LMG 28156^T (=CCUG 65686^T) and *Burkholderia territorii* sp. nov. with type strain LMG 28158^T (=CCUG 65687^T) are proposed for Bcc B and Bcc L bacteria, respectively.

4.2.2 Background, methods, results and discussion

The Bcc is a versatile group of closely related bacteria that are ubiquitous in nature (Coenye & Vandamme, 2003; Vanlaere *et al.*, 2009; Peeters *et al.*, 2013; Vandamme & Peeters, 2014). Some Bcc strains show biotechnological potential for biocontrol, bioremediation and plant growth promotion, whereas other strains are pathogens for plants and immunocompromised individuals such as those with CF. Bcc strains are also common nosocomial pathogens (Vonberg & Gastmeier, 2007). The Bcc comprises, at the time of writing, 18 validly named species (Peeters *et al.*, 2013), which exhibit a high degree of 16S rRNA (98-100 %) and *recA* (94-95 %) gene sequence similarity, moderate levels of DDH (30-50 %) (Coenye *et al.*, 2001d) and whole-genome ANI values of between 85.04 and 89.92 % (Vanlaere *et al.*, 2009). Protein encoding genes applied in a MLSA approach have a high discriminatory power for the identification of Bcc isolates, and 3 % concatenated sequence divergence has been validated as a threshold level for species differentiation (Vanlaere *et al.*, 2009; Peeters *et al.*, 2013). The application of this threshold level to analyze publicly available MLST data revealed the presence of at least another 16 novel Bcc species that await formal description (Vandamme & Peeters, 2014).

Nine Bcc bacteria were isolated in ongoing environmental soil and water sampling surveys for the ecological niches of *B. pseudomallei*, the aetiological agent of melioidosis, in the tropical Northern Territory of Australia. As part of the Darwin prospective melioidosis study (Currie *et al.*, 2010) household yards, gardens and domestic water supplies and other locations in Darwin city and its rural surrounds have been opportunistically sampled over the last two decades, with recovered *B. pseudomallei* and other species of the genus *Burkholderia* characterized and stored. Environmental sampling and bacterial isolation methods have been described previously (Mayo *et al.*, 2011; Kaestli *et al.*, 2012).

These nine isolates represented two unique MLSA clusters within the Bcc, referred to earlier as clusters Bcc B and Bcc L (Vandamme & Peeters, 2014). Upon deposition of the sequences in the Bcc PubMLST database (http://pubmlst.org/bcc/), three additional Bcc B isolates were identified: one environmental soil isolate from North Carolina, USA, one clinical isolate from a respiratory specimen (tracheal aspirate) of a native American non-CF adult patient from North Carolina (we were unable to verify if infection could have been acquired abroad) and one isolate from a sputum sample of a patient with CF from Australia (Table 4.3).

All isolates were preserved in MicroBank vials at −80 °C. They were grown aerobically on tryptone soya agar (Oxoid) and incubated at 30 °C for further characterization, except when mentioned otherwise. For PCR experiments, DNA was prepared by alkaline lysis as described by Storms *et al.* (2004).

Table 4.3: Studied isolates, showing their source, ST and allelic profile

Strain	Other strain designations	Source (country, year of isolation)
***B. stagnalis* sp. nov.**		
R-52235	MSMB049, FC1736	Soil (Australia, 2006)
LMG 28156^T	CCUG 65686^T, MSMB050^T, FC1737^T	Soil (Australia, 2006)
R-52237	MSMB085, FC1738	Soil (Australia, 2007)
R-52238	MSMB086, FC1739	Soil (Australia, 2007)
LMG 28157	MSMB097, FC1740	Soil (Australia, 2007)
R-52095	HI3541, FC1742	Soil (USA, 2005)
R-52096	AU7314, FC1741	Tracheal aspirate, non-CF (USA, 2004)
R-52240	MSMB21995, QLD037, FC1812	CF sputum (Australia, 2008)
***B. territorii* sp. nov.**		
LMG 28158^T	CCUG 65687^T, MSMB110^T, FC1743^T	Water (Australia, 2003)
R-52242	MSMB117, FC1744	Water (Australia, 2003)
LMG 28159	MSMB138, FC1745	Water (Australia, 2003)
R-52244	MSMB139, FC1746	Water (Australia, 2003)

Strain	ST	Allelic profile						
		atpD	***gltB***	***gyrB***	***recA***	***lepA***	***phaC***	***trpB***
B. stagnalis sp. nov.								
R-52235	787	170	197	564	345	300	213	381
LMG 28156^T	787	170	197	564	345	300	213	381
R-52237	789	328	379	399	347	300	213	196
R-52238	789	328	379	399	347	300	213	196
LMG 28157	789	328	379	399	347	300	213	196
R-52095	865	205	406	420	366	245	191	247
R-52096	519	170	197	399	190	300	213	196
R-52240	690	300	345	511	319	300	274	351
B. territorii sp. nov.								
LMG 28158^T	791	233	381	567	348	395	304	384
R-52242	792	329	382	568	349	396	305	385
LMG 28159	794	331	384	570	351	398	306	386
R-52244	794	331	384	570	351	398	306	386

LMG, BCCM/LMG Bacteria Collection, Laboratory of Microbiology, Ghent University, Ghent, Belgium; CCUG, Culture collection University of Göteborg, Sweden; MSMB, Menzies school of Health Research, Darwin, Australia; CF, cystic fibrosis patient.

MLSA was performed using standard protocols (Spilker *et al.*, 2009; Peeters *et al.*, 2013). Nucleotide sequences of each allele, allelic profiles and sequence types (STs) for all isolates represented five STs for cluster Bcc B isolates and three STs for cluster Bcc L isolates, and are available on the Bcc PubMLST website (http://pubmlst.org/bcc/) (Jolley & Maiden, 2010). The mean number of nucleotide substitutions per site (i.e. the percentage of divergence of concatenated allele sequences) between established (June 2014) and the newly proposed Bcc species was calculated using the program DnaSP v5.10 (Librado & Rozas, 2009) based on the Jukes-Cantor method (Jukes & Cantor, 1969). A phylogenetic tree of the concatenated sequences (2773 bp) of seven housekeeping gene fragments, namely *atpD* (443 bp), *gltB* (400 bp), *gyrB* (454 bp), *recA* (393 bp), *lepA* (397 bp), *phaC* (385 bp) and *trpB* (301 bp), was constructed using MEGA5 (Tamura *et al.*, 2011).

Within cluster Bcc B (n=5) the concatenated allele sequence divergence was 1.4±0.6 %, and the average divergence towards its nearest neighbor, *B. ubonensis*, was 4.0±0.8 %. Within cluster Bcc L (n=3) the concatenated allele sequence divergence was 1.9±0.7 %, and the average divergence towards its nearest neighbor, *B. diffusa*, was 3.9±0.9 % (data not shown). These concatenated sequence divergence values demonstrate that the two clusters represent novel species within the Bcc (Vanlaere *et al.*, 2009; Peeters *et al.*, 2013).

When a non-Bcc outgroup species (*B. fungorum*) was included in the phylogenetic analysis of the concatenated sequences, cluster Bcc B and cluster Bcc L isolates constituted two clusters supported by high bootstrap values of 93 and 99 %, respectively (Fig. 4.3; see also Vandamme & Peeters, 2014). Yet a bifurcation was apparent in each cluster and the removal of the outgroup taxon reinforced the differences between these subgroups in the resulting phylogenetic trees: Supplementary Fig. 4.3-4.9 show the results of phylogenetic analyses based on single gene sequences. Among Bcc B isolates, R-52095 had *gltB*, *lepA* and *trpB* gene sequences that were very similar to those of the remaining cluster Bcc B strains, whereas its *gyrB*, *phaC* and, to an even greater extent, *atpD* and *recA* gene sequences were divergent from those of the remaining cluster Bcc B strains. Among the cluster Bcc L strains, isolates LMG 28159 and R-52244 differed from LMG 28158^{T} and R-52242 in particular for the *recA*, *trpB* and, to a lesser extent, *lepA* gene sequences, whereas the *atpD*, *gltB*, *gyrB* and *phaC* gene sequences were highly similar (Supplementary Fig. 4.3-4.9).

RAPD analysis of all isolates was performed with primer 270 as described by Mahenthiralingam *et al.* (1996) and profiles were compared visually (data not shown). For the eight Bcc B isolates, five different RAPD profiles where observed, corresponding to the five STs. For the four Bcc L isolates, four different RAPD profiles were observed; isolates LMG 28159 and R-52244, which represented a single ST, had only slightly different RAPD profiles (data not shown).

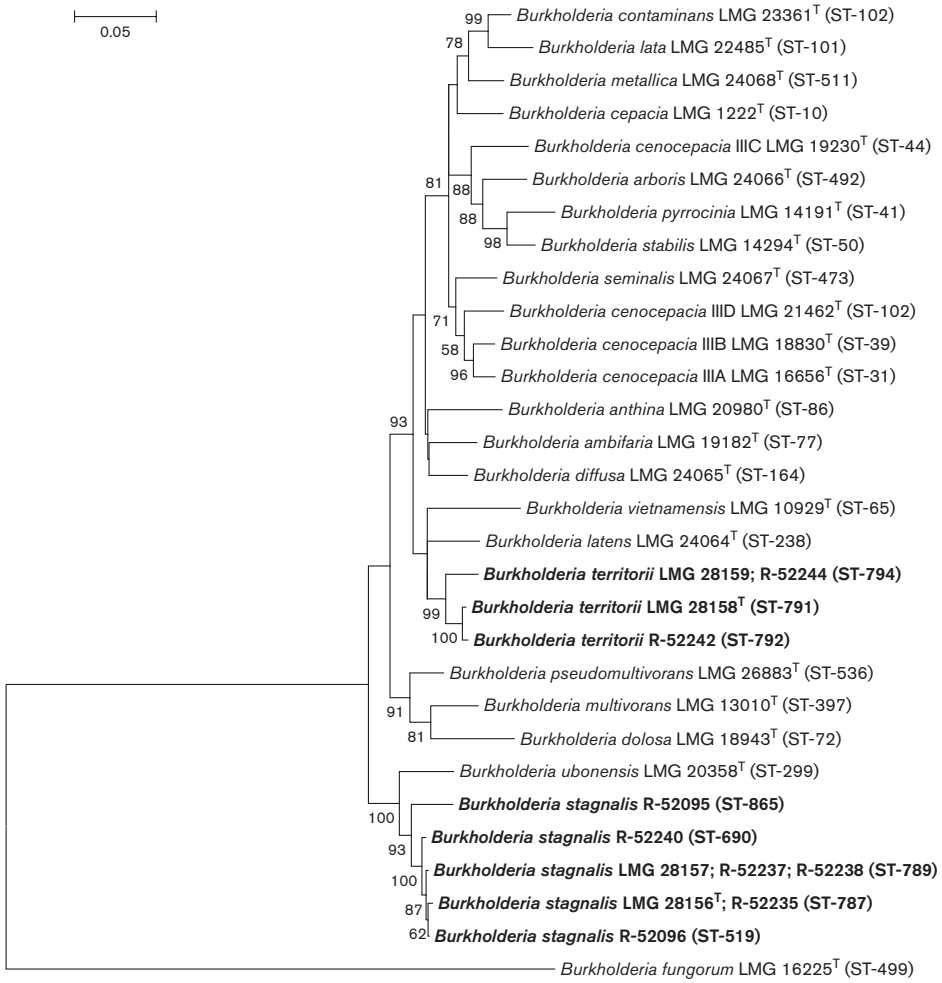


Figure 4.3: Phylogenetic tree based on the concatenated sequences (2773 bp) of seven housekeeping gene fragments [*atpD* (443 bp), *gltB* (400 bp), *gyrB* (454 bp), *recA* (393 bp), *lepA* (397 bp), *phaC* (385 bp) and *trpB* (301 bp)] of established Bcc species and *Burkholderia stagnalis* sp. nov. and *Burkholderia territorii* sp. nov. strains. The bootstrap consensus tree, inferred from 1000 replicates, was reconstructed using the maximum likelihood method based on the general time reversible model. Branches corresponding to partitions reproduced in less than 50 % bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches if greater than 50 %. A discrete gamma distribution was used to model evolutionary rate differences among sites [five categories (+G, parameter=0.3623)] and allowed for some sites to be evolutionarily invariable [(+I), 51.7996 % sites]. The analysis involved 30 nt sequences. Evolutionary analyses were conducted in MEGA5. The sequence of *B. fungorum* LMG 16225^T was used as an outgroup. Bar, 0.05 substitutions per site.

The nearly complete sequences of the 16S rRNA gene of strains LMG 28156^T (1485 bp) and LMG 28158^T (1443 bp) were obtained as described previously (Peeters *et al.*, 2013). Pairwise comparison of these sequences, as determined with the EzTaxon-e server (www.ezbiocloud.net/eztaxon (Kim *et al.*, 2012), with those of type strains of other Bcc species revealed similarity levels of between 98.43 and 99.66 % for the former and between 98.89 and 100 % for the latter (data not shown). Similarity levels towards non-Bcc *Burkholderia* species were in the range 94.28-98.98 %. For both strains the highest similarity values were obtained towards the type strain of *B. glumae*.

Sequences of the *recA* gene were amplified using forward primer 5' AGG ACG ATT CAT GGA AGA WAG C 3' and reverse primer 5' GAC GCA CYG AYG MRT AGA ACT T '3 (Spilker *et al.*, 2009). Sequence assembly was performed using BioNumerics v5.10 (Applied Maths). Sequences (540-660 bp) were aligned based on amino acid sequences using Muscle (Edgar, 2004) in MEGA5 (Tamura *et al.*, 2011). Phylogenetic analysis was conducted in MEGA5 (Tamura *et al.*, 2011). All positions containing gaps and missing data were eliminated, resulting in a total of 473 positions in the final dataset. All Bcc B isolates except isolate R-52095, which occupied a distinct position in the *recA* gene tree, formed a coherent *recA* gene cluster with *B. pseudomultivorans* as its nearest neighbor (Supplementary Fig. 4.2). These results confirmed and extended the considerable differences observed in the 393 bp *recA* gene fragment of the MLSA scheme (Supplementary Fig. 4.6). Similarly, the bifurcation of the four Bcc L strains was confirmed in the 473 bp based analysis (Supplementary Fig. 4.2).

For determination of the DNA base composition, high-molecular-mass DNA was prepared as described by Pitcher *et al.* (1989). DNA was enzymatically 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 thermo-stabilized at 37 °C. The solvent was 0.02 M $NH_4H_2PO_4$ (pH 4.0) with 1.5 % (v/v) acetonitrile. Non-methylated lambda phage (Sigma) and *Escherichia coli* LMG 2093 DNA were used as calibration reference and control, respectively. The DNA G+C content of strains LMG 28156^T and LMG 28158^T was both 67 mol%, which corresponded with that of other Bcc species (66-69 mol%) (Vandamme & Dawyndt, 2011).

For fatty acid methyl ester analysis, all strains were grown on tryptone soya agar (BD) at 28 °C for 24 h. A loopful of well-grown cells was harvested and fatty acid methyl esters were prepared, separated and identified using the Microbial Identification System (Microbial ID) as described previously (Vandamme *et al.*, 1992). The following fatty acid components were detected in Bcc B and Bcc L isolates, respectively: $C_{14:0}$ (4.08±1.3 and 4.3±0.3 %), $C_{16:0}$ (18.7±0.9 and 19.0±0.9 %), $C_{16:0}$ 2-OH (1.5±0.5 and 1.2±0.6 %),$C_{16:0}$ 3-OH (8.4±1.0

and 7.1±0.9 %), $C_{16:1}$ 2-OH (1.9±0.6 and 1.0±0.4 ;), $C_{17:0}$ cyclo (10.2±4.2 and 6.1±3.2 %), $C_{18:1}$ ω7c (24.0±4.5 and 30.1±1.7 %), $C_{19:0}$ cyclo ω8c (4.4±2.4 and 2.1±1.3 %), $C_{18:0}$ 2-OH (2.8±0.7 and 2.5±1.0 %), summed feature 2 (comprising $C_{14:0}$ 3-OH, iso-$C_{16:1}$ I, an unidentified fatty acid with equivalent chain-length of 10.928 or $C_{12:0}$ ALDE or any combination of these) (10.3±1.7 and 10.8±3.3 %) and summed feature 3 (comprising $C_{16:1}$ ω7c and/or iso-$C_{15:0}$ 2-OH) (12.2±4.9 and 15.2±4.0 %). These cellular fatty acid profiles are highly similar to those of other Bcc species and *B. gladioli* (Stead, 1992).
Biochemical characterization of all isolates was performed as described by Henry *et al.* (2001) and a summary of the results can be found in Table 4.4 and in the species descriptions below. Biochemically, all Bcc B and Bcc L strains are similar to other Bcc bacteria. Bcc B strains can be differentiated from other Bcc species (except for *B. ubonensis*) by the assimilation of maltose. The only difference noted between isolate R-52095 and the remaining Bcc B isolates is the presence of β-haemolysis. Bcc L strains can be differentiated from other Bcc species by the absence of assimilation of N-acetylglucosamine. The two Bcc L *recA* gene clusters differed in nitrate reduction capacity.
In conclusion, MLSA demonstrated that 12 Bcc isolates represented two novel species within the Bcc. DNA base content determination, 16S rRNA gene sequence analysis and fatty acid methyl ester analysis confirmed that these isolates conformed to the general characteristics of the genus *Burkholderia*. The two novel species consisted of two *recA* gene lineages that were phenotypically different; yet assimilation of maltose and of N-acetylglucosamine could be used to distinguish these species from other Bcc species. We therefore propose to formally classify Bcc B and Bcc L strains as representing the novel species *Burkholderia stagnalis* sp. nov. with strain LMG 28156^T (=CCUG 65686^T) as the type strain, and *Burkholderia territorii* sp. nov. with strain LMG 28158^T (=CCUG 65687^T) as the type strain, respectively.

4.2.2.1 Description of *Burkholderia stagnalis* sp. nov.

Burkholderia stagnalis (stag.na'lis. L. n. *stagnum* pond, lagoon, lake; -*alis* adjective forming suffix; N.L. fem. adj. *stagnalis* pertaining to a lagoon).
Cells are Gram-stain-negative, aerobic, motile, non-sporeforming rods, about 0.1-0.5 µm wide and 0.6-1.7 µm long. All isolates grow at 37 °C on BSA, BCSA and MacConkey agar and do not produce any pigment. The majority (six of eight) of the isolates show medium-sized grey entire colonies after 3 days of incubation on BSA, and the remaining two isolates show distinctive colony centers Only one isolate (R-52095) shows β-haemolysis on blood agar. The majority of isolates (five of eight) are able to grow at 42 °C (R-52235, LMG 28156^T, R-52237, R-52238, R-52239). All isolates show slow oxidase-positive reactions and lysine

decarboxylase activity, but no nitrate reduction, ornithine decarboxylase or aesculin hydrolysis activity. Furthermore, glucose, maltose, lactose and xylose are acidified, but not adonitol. Assimilation of maltose is observed for all strains. Strain-dependent reactions are observed for acidification of sucrose (negative for R-52240), gelatinase (negative for R-52096) and β-galactosidase (weakly positive for R-52096) activity. API 20NE profiles are 0057777 (six of eight isolates), 0056577 (R-52240) and 0067777 (R-52096). The following fatty acids are present in all isolates: $C_{14:0}$, $C_{16:1}$ 2-OH, $C_{16:0}$ 2-OH, $C_{18:0}$, $C_{19:0}$ cyclo ω8c and $C_{18:1}$ 2-OH all at <5 % (mean value of all isolates); and $C_{16:0}$, $C_{17:0}$ cyclo, $C_{16:0}$ 3-OH, $C_{18:1}$ ω7c and summed features 2 and 3 all at between 5 and 25 % (mean value of all isolates). Isolates have been obtained from soil and human respiratory samples in Australia and the USA.
The type strain is LMG 28156^T (=CCUG 65686^T, originally collected as MSMB050^T). Phenotypic characteristics of the type strain are the same as those described above for all strains of the species. In addition, the type strain grows at 42 °C, does not show β-haemolysis on sheep blood agar, acidifies sucrose, assimilates arabinose and shows gelatinase activity. The DNA G+C content of the type strain is 67 mol%.

4.2.2.2 Description of *Burkholderia territorii* sp. nov.

Burkholderia territorii [ter.ri.to'ri.i. L. gen. n. *territorii* of the territory; referring to the (Northern) Territory of Australia, where this organism has been recovered from environmental groundwater samples].
Cells are Gram-stain-negative, aerobic, motile, non-sporeforming rods, about 0.2-0.5 µm wide and 0.5-1.7 µm long. All isolates grow at 37 °C on BSA, BCSA and MacConkey agar. They show medium-sized grey entire colonies on BSA without haemolysis, no pigment production is observed and all strains grow at 42 °C. Two isolates (R-52243 and R-52244) display mixed colonies types with small and large colonies. However, RAPD on both colony types shows identical patterns. All isolates show slow oxidase-positive reactions, lysine decarboxylase, β-galactosidase and gelatinase activity but no ornithine decarboxylase or aesculin hydrolysis activity. Acidification of glucose, maltose, lactose, xylose and adonitol is positive, with strain-dependent reaction for sucrose acidification (negative for R-52242), as well as nitrate reduction (negative for LMG 28158^T and R-52242). No assimilation of N-acetylglucosamine. API 20NE profiles are 0077567 (LMG 28158^T), 0077577 (R-52242) and 1077577 (R-52243 and R-52244). The following fatty acids are present in all isolates: $C_{14:0}$, $C_{16:1}$ 2-OH, $C_{16:0}$ 2-OH, $C_{18:0}$, $C_{19:0}$ cyclo ω8c and $C_{18:1}$ 2-OH all at <5 %; and $C_{16:0}$, $C_{17:0}$ cyclo, $C_{16:0}$ 3-OH, $C_{18:1}$ ω7c, and summed features 2 and 3 all at between 5 and 30 %. Strains have been isolated from environmental groundwater samples.

The type strain is LMG 28158^T (=CCUG 65687^T, originally collected as MSMB110^T). Phenotypic characteristics of the type strain are the same as those described above for all strains of the species. In addition, the type strain reduces nitrate and acidifies sucrose. The DNA G+C content of the type strain is 67 mol%.

Table 4.4: Phenotypic characteristics useful for differentiation of *B. stagnalis* sp. nov. and *B. territorii* sp. nov. from members of the Bcc

Phenotypic characteristic	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
Growth on MacConkey agar	+	+	+	+	+	+	+	+	+	+	v	+	v	+	v	+	+	+	+	+
Growth at 42 °C	5(+)	+	+	v	−	+	v	v	+	+	v	+	v	−	+	+	v	v	v	v
Pigment	−	−	−	v	v	−	−	v	v	+(y)	v	−	v	−	−	−	v	−	v	−
Haemolysis (sheep blood)	1 β(−)	−	−	v	−	−	−	v	−	−	−	−	−	−	v	−	v	−	v	−
Assimilation of:																				
L-Arabinose	5(+)	+	+	+	v	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+
D-Mannitol	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	v	+
N-Acetylglucosamine	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	v	+
Maltose	+	3(−)	v	−	v	−	−	v	−	−	−	−	−	−	v	−	−	v	v	+
Adipate	+	+	+	+	+	+	+	+	+	+	+	+	+	+	v	+	+	+	+	+
Phenylacetate	+	+	+	v	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Acidification of:																				
Maltose	+	+	+	v	+	+	+	+	+	+	v	+	v	+	+	+	+	+	+	+
Lactose	+	+	+	+	+	+	+	+	+	+	+	+	v	+	+	+	+	+	+	+
D-Xylose	+	+	+	+	+	+	+	w	+	+	+	+	+	v	v	+	+	+	+	+
Sucrose	7(+)	3(+)	v	+	v	+	+	+	+	+	+	−	+	−	+	−	+	v	+	+
Adonitol	−	+	v	+	v	+	v	+	+	+	v	+	v	+	−	+	+	v	+	−
Nitrate reduction	−	2(−)	v	v	v	−	+	v	−	−	−	+	v	−	v	+	v	v	v	v
Activity of:																				
Lysine decarboxylase	+	+	+	+	+	+	+	v	v	+	+	v	+	+	+	−	+	v	+	−
Ornithine decarboxylase	−	−	−	−	v	−	−	+	v	−	v	−	v	+	−	−	−	−	+	−
Aesculin hydrolysis	−	−	v	v	v	−	−	−	v	+	v	−	v	−	−	−	v	−	−	−
Arginine dihydrolase	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+
Gelatinase	7(+)	+	−	+	v	v	v	+	+	+	v	−	v	v	−	−	+	−	+	+
β-galactosidase	1(−)	+	+	+	v	+	+	+	+	+	+	+	+	−	+	+	+	v	+	−

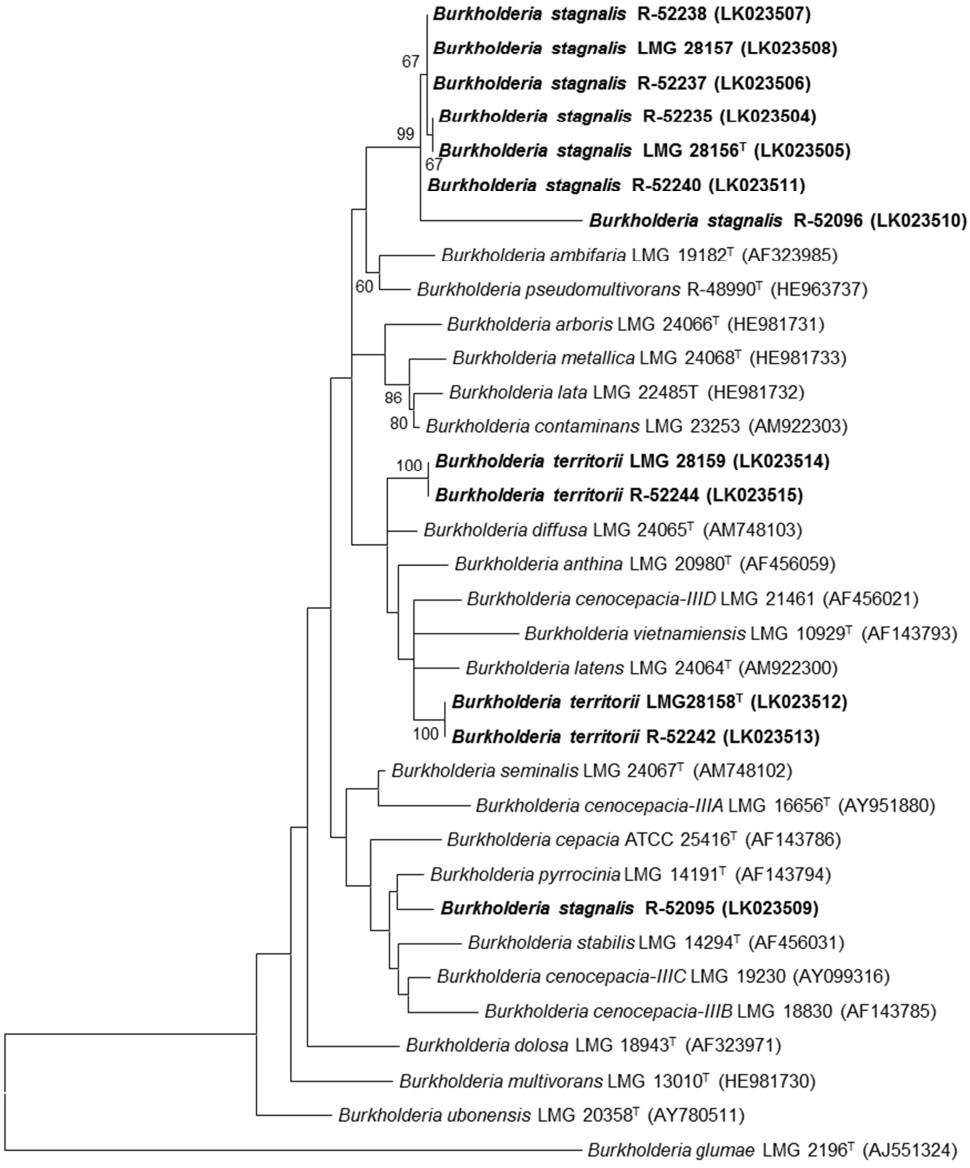
Species: 1, *B. stagnalis* sp. nov.; 2, *B. territorii* sp. nov.; 3, *B. pseudomultivorans*; 4, *B. contaminans*; 5, *B. lata*; 6, *B. latens*; 7, *B. diffusa*; 8, *B. arboris*; 9, *B. seminalis*; 10, *B. metallica*; 11, *B. cepacia*; 12, *B. multivorans*; 13, *B. cenocepacia*; 14, *B. stabilis*; 15, *B. vietnamiensis*; 16, *B. dolosa*; 17, *B. ambifaria*; 18, *B. anthina*; 19, *B. pyrrocinia*; 20, *B. ubonensis*. +, <90 % of all isolates positive; v, 10-90 % positive; −, <10 % of strains positive; w, weak reaction; β, β-haemolysis; y, yellow. For *B. stagnalis* (n=8) and *B. territorii* (n=4), the number of positive reactions is indicated for strain-dependent reactions and the reaction of the type strain is given in parentheses. For characteristics of established Bcc species data were taken from the present study and Peeters *et al.* (2013).

4.2.3 Acknowledgments

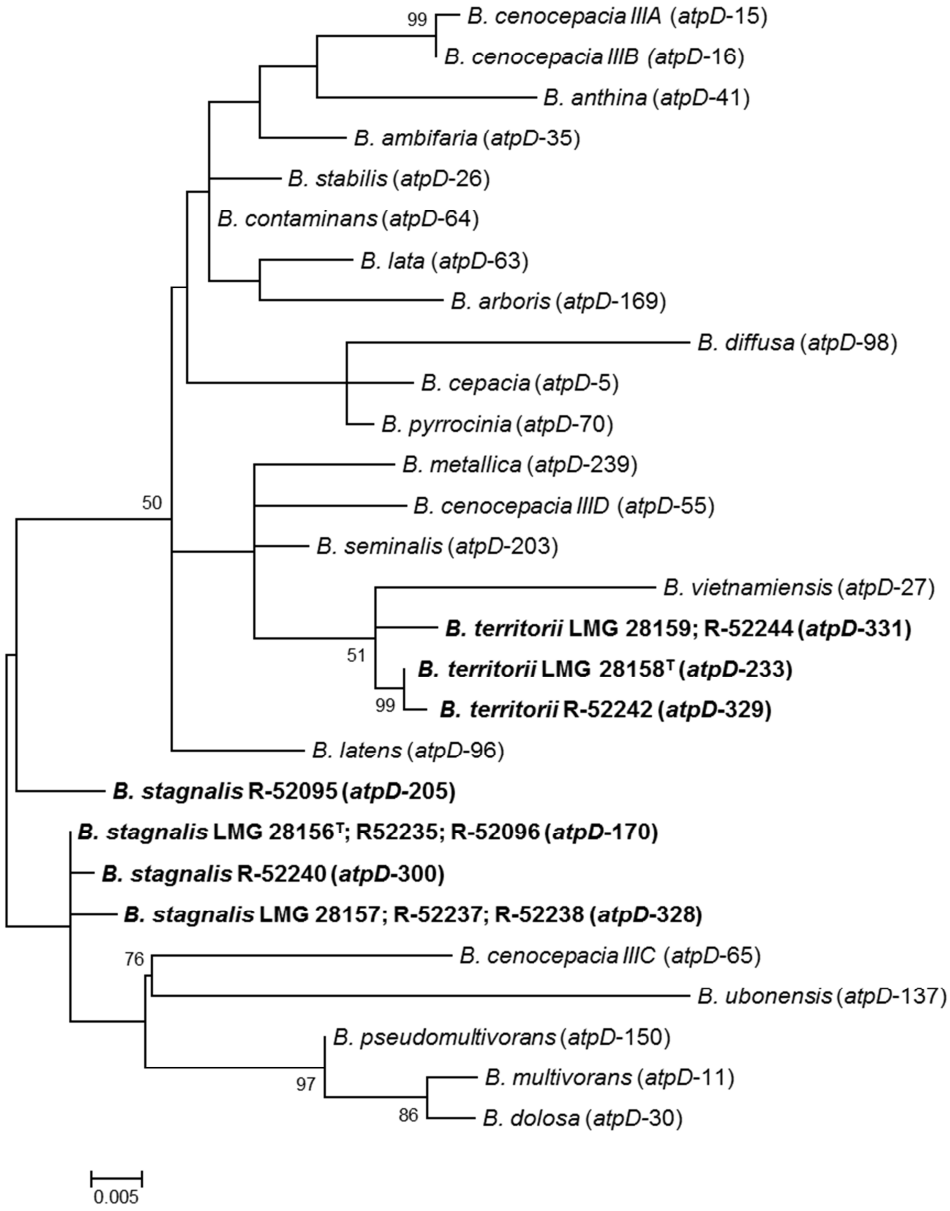
We thank all strain depositors who contributed to this study, Cindy Snauwaert (Laboratory of Microbiology, Faculty of Sciences, Ghent University, Ghent, Belgium), Evie De Brandt (Laboratory of Microbiology, Faculty of Sciences, Ghent University, Ghent, Belgium), Rebecca Hickman (University of British Columbia) and Kay Ramsay (Queensland Childrens Medical Research Institute, The University of Queensland, Brisbane, Queensland, Australia) for technical assistance, and Glenda Harrington (Menzies School of Health Research, Darwin, Northern Territory, Australia) for assistance with environmental sampling. Collection of the environmental samples was supported by grants from the Australian National Health and Medical Research Council (605820 and 1046812). J. E.A. Z. and T. J. H. are very grateful for the support of David P. Speert and financial support provided to D. P. S. by Cystic Fibrosis Canada through an operating grant (20R32976) and support for the Canadian Bcc research and referral repository (20R40868). Initial MLSA was supported by grant funding from the Childrens Health Foundation Queensland (50005). This publication made use of the Bcc PubMLST website (http://pubmlst.org/bcc) developed by Keith Jolley and sited at the University of Oxford (Jolley & Maiden, 2010). The development of this site has been funded by the Wellcome Trust.

4.2.4 Supplementary material

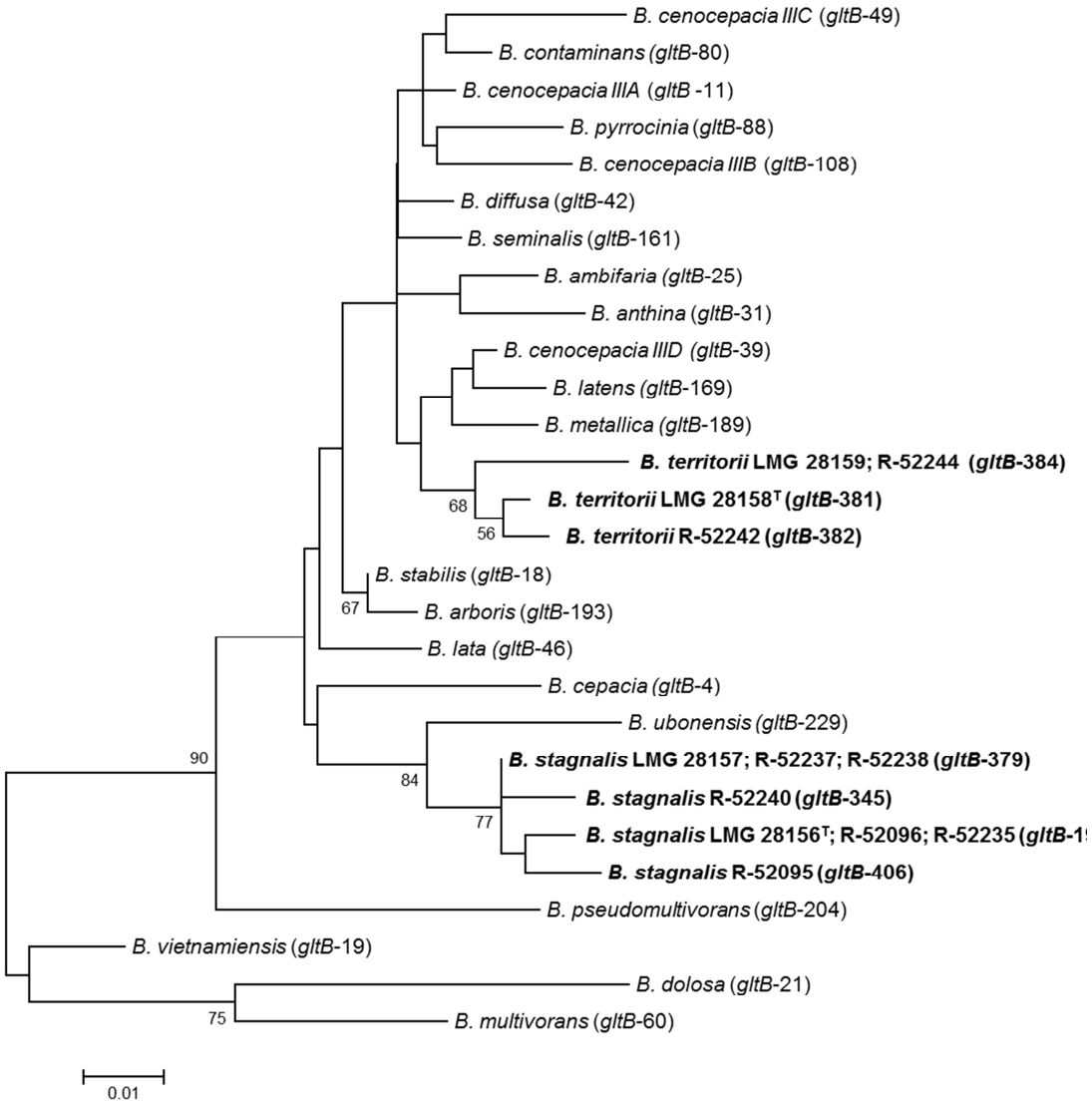
The GenBank/EMBL/DDBJ accession numbers for the 16S rRNA gene sequences of strains LMG 28156^{T} and LMG 28158^{T} are LK023502 and LK023503, respectively. Those for the *recA* gene sequences of strains R-52235, LMG 28156^{T}, R-52237, R-52238, LMG 28157, R-52095, R-52096, R-52240, LMG 28158^{T}, R-52242, LMG 28159 and R-52244 are LK023504-LK023515, respectively.



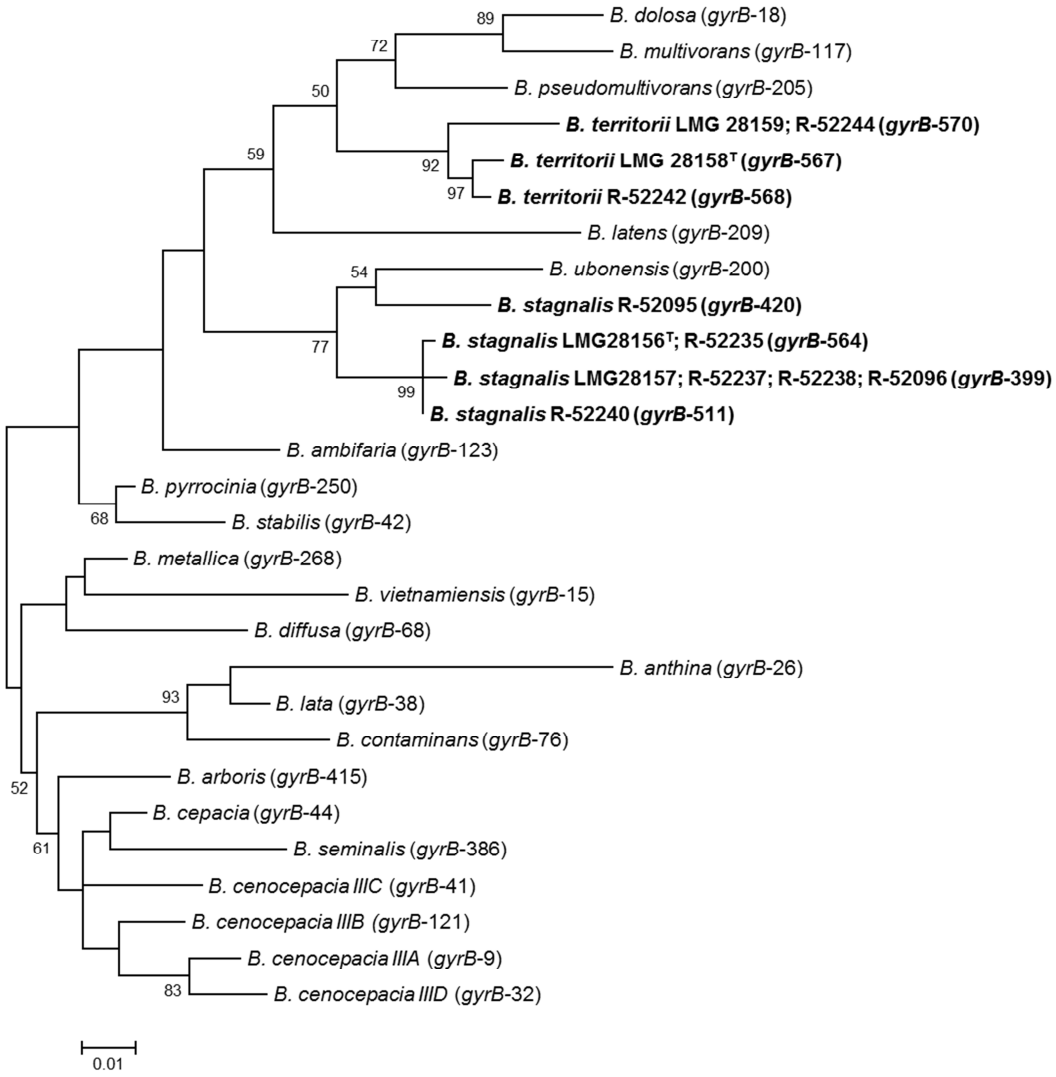
Supplementary Figure 4.2: Phylogenetic tree based on *recA* gene sequences of established Bcc species and *B. stagnalis* and *B. territorii* isolates. The bootstrap consensus tree, inferred from 1000 replicates, was constructed using the maximum likelihood method based on the Tamura 3-parameter model (Tamura, 1992). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test is shown next to the branches if greater than 50 %. A discrete gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.3168)) and allowed for some sites to be evolutionarily invariable ([+I], 39.4400 % sites). The *recA* sequence of *B. glumae* LMG 2196^{T} was used as outgroup. The scale bar indicates the number of substitutions per site. The analysis involved 34 nucleotide sequences. There were a total of 473 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Tamura *et al.*, 2011).



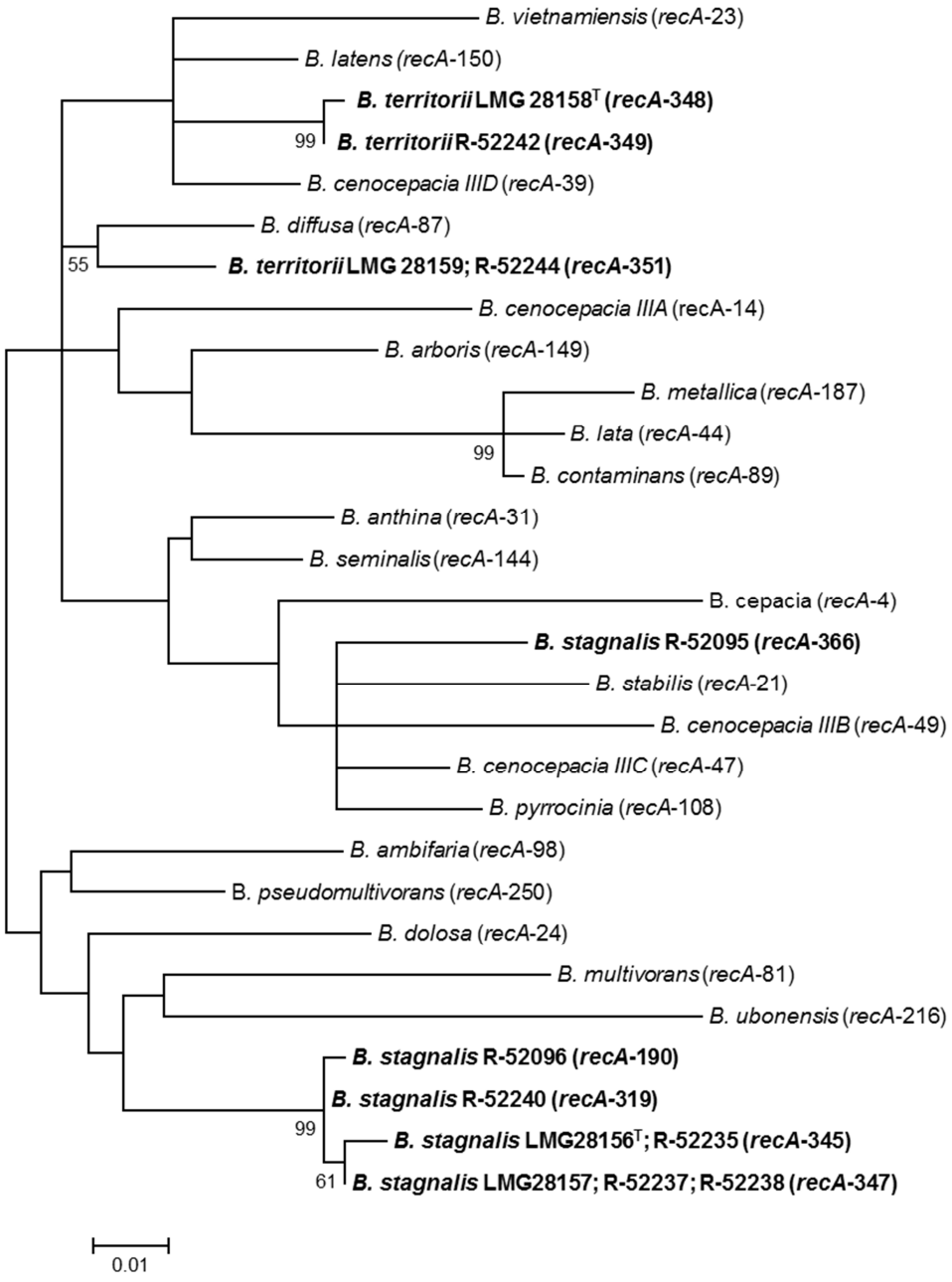
Supplementary Figure 4.3: Phylogenetic tree based on *atpD* locus sequences of established Bcc species and *B.stagnalis* and *B. territorii* isolates. The bootstrap consensus tree, inferred from 1000 replicates was constructed using the maximum likelihood method based on the Tamura 3-parameter model (Tamura, 1992). The percentage of trees in which the associated taxa clustered together is shown next to the branches if greater than 50 %. A discrete gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.2570)) and allowed for some sites to be evolutionarily invariable ([+I], 74.3388 % sites). The scale bar indicates the number of substitutions per site. The analysis involved 28 nucleotide sequences. There were a total of 443 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Tamura *et al.*, 2011).



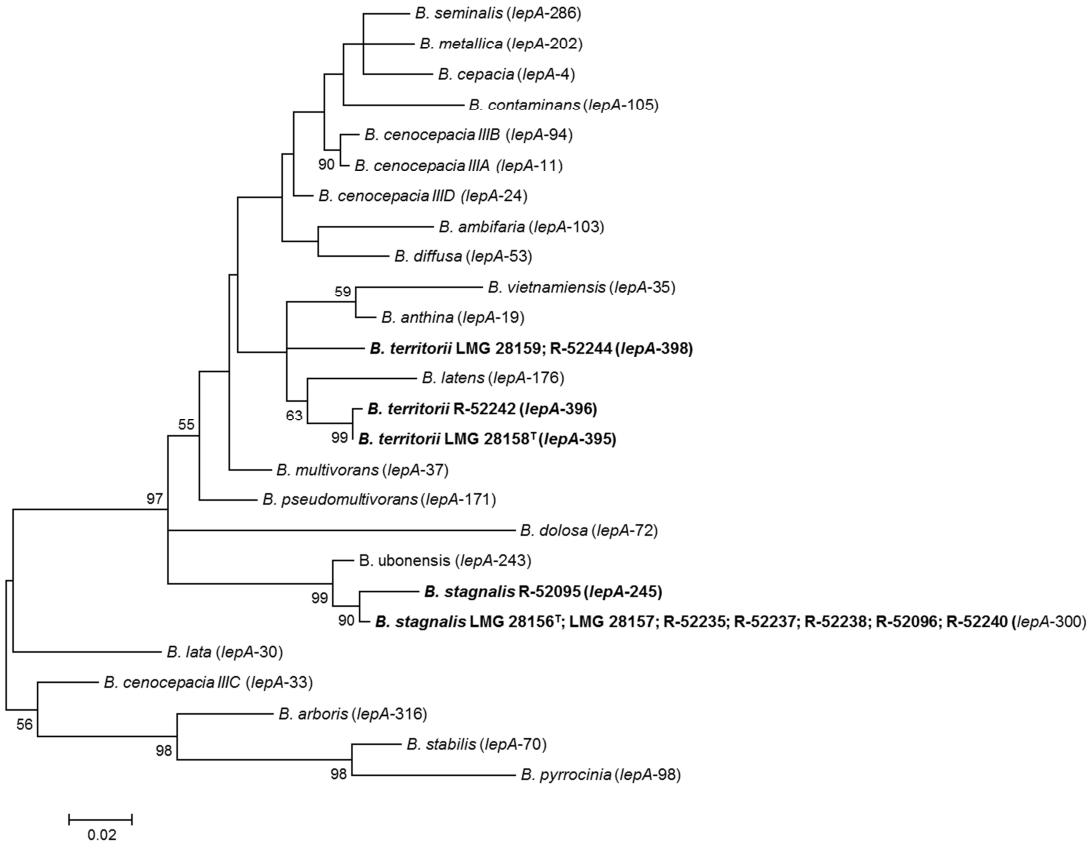
Supplementary Figure 4.4: Phylogenetic tree based on *gltB* locus sequences of established Bcc species and *B. stagnalis* and *B. territorii* isolates. The bootstrap consensus tree, inferred from 1000 replicates was constructed using the maximum likelihood method based on the Tamura 3-parameter model (Tamura, 1992). The percentage of trees in which the associated taxa clustered together is shown next to the branches if greater than 50 %. A discrete gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.4758)) and allowed for some sites to be evolutionarily invariable ([+I], 57.9292 % sites). The scale bar indicates the number of substitutions per site. The analysis involved 28 nucleotide sequences. There were a total of 400 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Tamura *et al.*, 2011).



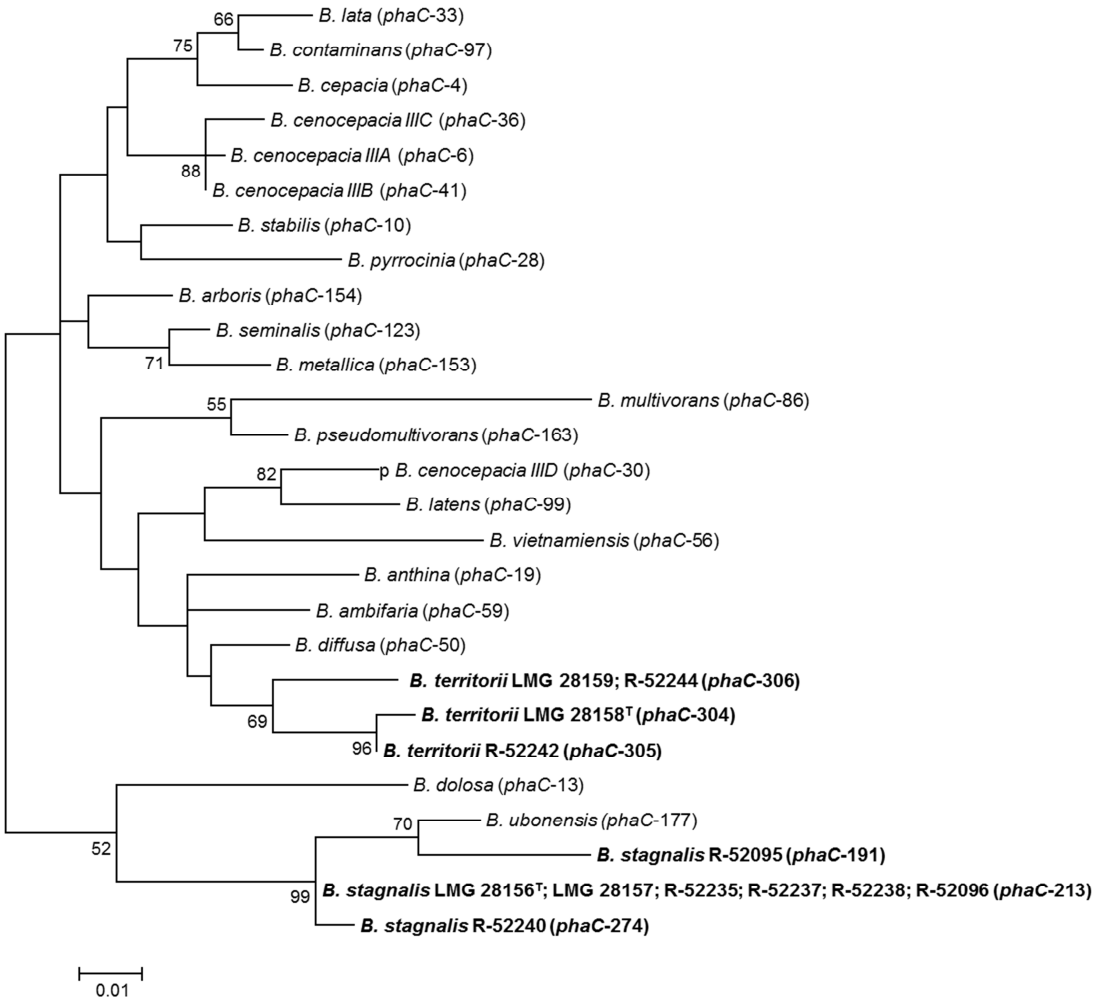
Supplementary Figure 4.5: Phylogenetic tree based on *gyrB* locus sequences of established Bcc species and *B. stagnalis* and *B. territorii* isolates. The bootstrap consensus tree, inferred from 1000 replicates was constructed using the maximum likelihood method based on the general time reversible model. The percentage of trees in which the associated taxa clustered together is shown next to the branches if greater than 50 %. A discrete gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.4373)) and allowed for some sites to be evolutionarily invariable ([+I], 49.3662 % sites). The scale bar indicates the number of substitutions per site. The analysis involved 28 nucleotide sequences. There were a total of 454 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Tamura *et al.*, 2011).



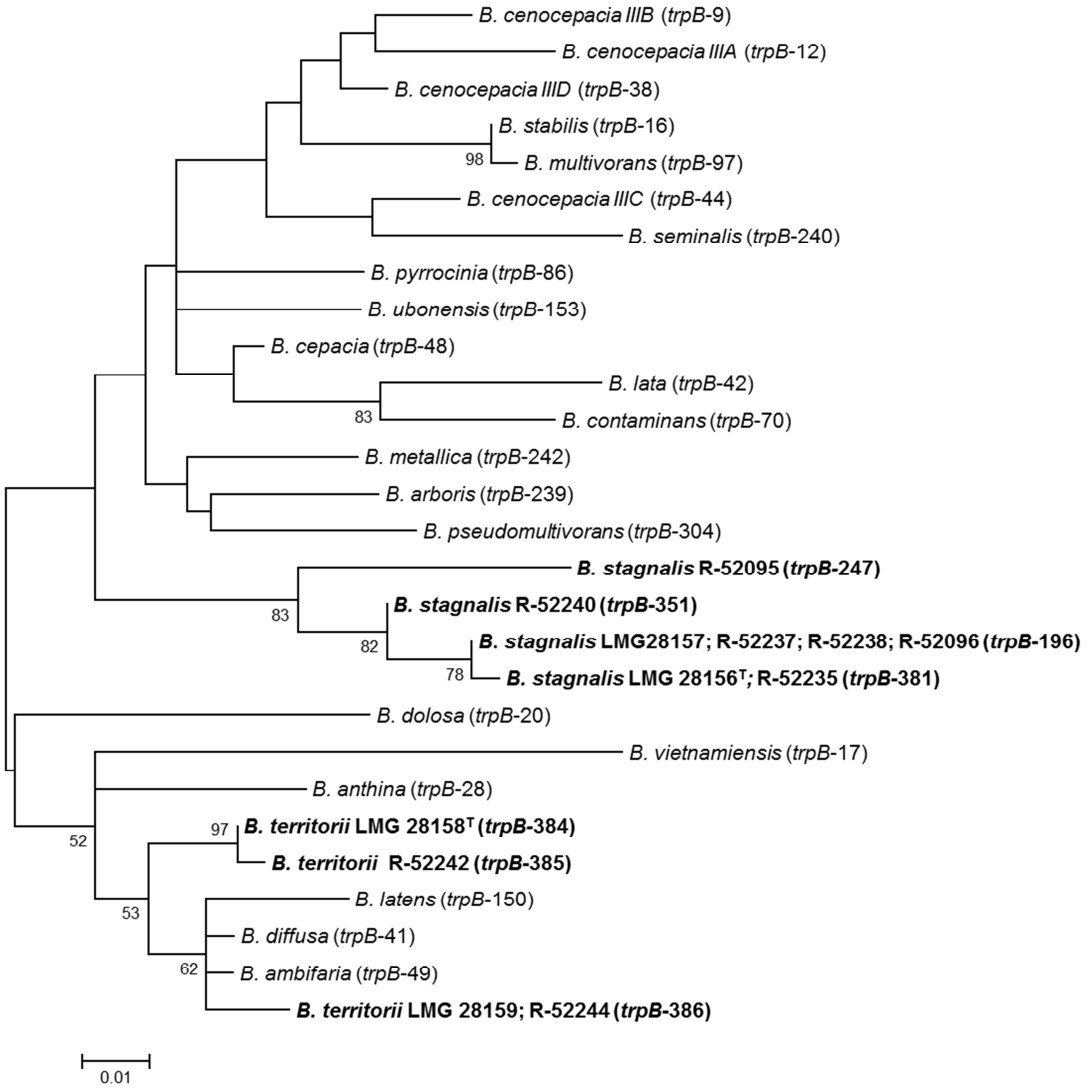
Supplementary Figure 4.6: Phylogenetic tree based on *recA* locus sequences of established Bcc species and *B. stagnalis* and *B. territorii* isolates. The bootstrap consensus tree, inferred from 1000 replicates was constructed using the maximum likelihood method based on the Tamura 3-parameter model (Tamura, 1992). The percentage of trees in which the associated taxa clustered together is shown next to the branches if greater than 50 %. A discrete gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.3316)) and allowed for some sites to be evolutionarily invariable ([+I], 54.4859 % sites). The scale bar indicates the number of substitutions per site. The analysis involved 29 nucleotide sequences. There were a total of 393 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Tamura *et al.*, 2011).



Supplementary Figure 4.7: Phylogenetic tree based on *lepA* locus sequences of established Bcc species and *B. stagnalis* and *B. territorii* isolates. The bootstrap consensus tree, inferred from 1000 replicates was constructed using the maximum likelihood method based on the Tamura 3-parameter model (Tamura, 1992). The percentage of trees in which the associated taxa clustered together is shown next to the branches if greater than 50 %. A discrete gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.6858)) and allowed for some sites to be evolutionarily invariable ([+I], 52.7765 % sites). The scale bar indicates the number of substitutions per site. The analysis involved 26 nucleotide sequences. There were a total of 397 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Tamura *et al.*, 2011).



Supplementary Figure 4.8: Phylogenetic tree based on *phaC* locus sequences of established Bcc species and *B. stagnalis* and *B. territorii* isolates. The bootstrap consensus tree, inferred from 1000 replicates was constructed using the maximum likelihood method based on the Tamura 3-parameter model (Tamura, 1992). The percentage of trees in which the associated taxa clustered together is shown next to the branches if greater than 50 %. A discrete gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.6710)) and allowed for some sites to be evolutionarily invariable ([+I], 67.9788 % sites). The scale bar indicates the number of substitutions per site. The analysis involved 27 nucleotide sequences. There were a total of 385 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Tamura *et al.*, 2011).



Supplementary Figure 4.9: Phylogenetic tree based on *trpB* locus sequences of established Bcc species and *B. stagnalis* and *B. territorii* isolates. The bootstrap consensus tree, inferred from 1000 replicates was constructed using the maximum likelihood method based on the Tamura 3-parameter model (Tamura, 1992). The percentage of trees in which the associated taxa clustered together is shown next to the branches if greater than 50 %. A discrete gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.3701)) and allowed for some sites to be evolutionarily invariable ([+I], 51.1343 % sites). The scale bar indicates the number of substitutions per site. The analysis involved 28 nucleotide sequences. There were a total of 301 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Tamura *et al.*, 2011).

5 | New insights into the taxonomy of *Burkholderia glathei*-like bacteria

The *B. glathei* clade (BGC) represents a phylogenetically divergent clade in the genus *Burkholderia* (Sawana *et al.*, 2014; Vandamme *et al.*, 2014). Most of the 12 validly named BGC species have been isolated from soil but publicly available sequence data indicate that the taxonomic diversity in this clade is severely underestimated (Nogales *et al.*, 2001; Salles *et al.*, 2006b; Pumphrey & Madsen, 2008; Xu *et al.*, 2016). Section 5.1 presents a traditional polyphasic taxonomic study of isolates from agricultural soil in Argentina and the formal classification of *Burkholderia cordobensis* sp. nov. Section 5.2 presents a modern polyphasic taxonomic approach in which whole-genome sequences are used for the assessment of the genomic relatedness among strains and their G+C content. This genomic taxonomic study was performed on 13 clusters of *B. glathei*-like isolates from human clinical and environmental origin and led to the formal taxonomic proposal of 13 novel BGC species.

5.1 *Burkholderia cordobensis* sp. nov., from agricultural soils

Redrafted from: Walter O. Draghi, Charlotte Peeters, Margo Cnockaert, Cindy Snauwaert, Luis G. Wall, Angeles Zorreguieta and Peter Vandamme (2014). *Burkholderia cordobensis* sp. nov., from agricultural soils. International Journal of Systematic and Evolutionary Microbiology, 64, 2003–2008.

Author contributions: WD, CP and PV conceived the study and wrote the manuscript. WD and CP performed all experiments and analyzed the sequence data. WD and MC performed the biochemical characterization. WD and CS performed the fatty acid methyl ester analysis. WD, LW and AZ isolated the strains.

5.1.1 Abstract

Two Gram-negative, rod-shaped bacteria were isolated from agricultural soils in Cordoba province in central Argentina. Their 16S rRNA gene sequences demonstrated that they belong to the genus *Burkholderia*, with *Burkholderia zhejiangensis* as most closely related formally named species; this relationship was confirmed through comparative *gyrB* sequence analysis. Whole-cell fatty acid analysis supported their assignment to the genus *Burkholderia*. *Burkholderia* sp. strain YI23, for which a whole-genome sequence is available, represents the same taxon, as demonstrated by its highly similar 16S rRNA (100 % similarity) and *gyrB* (99.1-99.7 %) gene sequences. The results of DDH experiments and physiological and biochemical characterization further substantiated the genotypic and phenotypic distinctiveness of the Argentinian soil isolates, for which the name *Burkholderia cordobensis* sp. nov. is proposed, with strain MMP81^T (=LMG 27620^T =CCUG 64368^T) as the type strain.

5.1.2 Background, methods, results and discussion

Members of the genus *Burkholderia* are β-Proteobacteria and are aerobic, chemo-organotrophic and motile (except for *B. mallei*), Gram-negative rods that display a wide physiological versatility, allowing them to inhabit extremely different ecological niches (Coenye & Vandamme, 2003; Compant *et al.*, 2008). Several species of the genus *Burkholderia* are human or animal pathogens; these include members of the Bcc, a group of closely related species commonly isolated from respiratory specimens of CF patients (Vandamme & Dawyndt, 2011), and species belonging to the *B. pseudomallei* group, which includes *B. pseudomallei*, the aetiological agent of melioidosis in humans, and *B. mallei*, the causative agent of glanders in animals

(Galyov *et al.*, 2010). However, strains of *Burkholderia*, including Bcc bacteria (Parke & Gurian-Sherman, 2001), may also have beneficial characteristics that include biological nitrogen fixation, phytohormone synthesis, plant defense induction and xenobiotic activity (Vial *et al.*, 2011; Suarez-Moreno *et al.*, 2012). Also, a growing number of species of the genus *Burkholderia* lives associated with other organisms: these include endophytes from plant tissues including root nodules, and inhabitants of insect guts or fungal mycelium and spores (Levy *et al.*, 2003; Kikuchi *et al.*, 2005; Mahenthiralingam *et al.*, 2008; Gyaneshwar *et al.*, 2011; Verstraete *et al.*, 2013).

Strains of *Burkholderia* occur commonly in pristine, agricultural or polluted soils, in which their population density can reach levels of 10^3-10^5 c.f.u. (per g soil) (Pallud *et al.*, 2001; Salles *et al.*, 2006b), but antagonistic species of the genus *Burkholderia* can be lost when soils are used continuously under arable management schemes (Salles *et al.*, 2006a). As part of the BIOSPAS Consortium, we analyzed the composition of populations of *Burkholderia* across Argentinian agricultural soils that were subjected to different agricultural management regimes. A thorough description of site characteristics and treatment definitions was published elsewhere (Figuerola *et al.*, 2012). From these samples, 1 g soil was suspended in 10 ml saline solution (0.85 % w/v NaCl), shaken at 240 rpm for 30 min, vortexed for 1 min, sonicated in a water bath for 30 s and centrifuged for 2 min at 500 rpm. An aliquot (200 µl) of the soil suspension was plated onto modified semi-selective medium PCAT (but using citrulline instead of tryptamine as the main nitrogen source; PCAT refers to *Pseudomonas cepacia*, azelaic acid and tryptamine) containing (per L) 0.1 g $MgSO_4$, 0.2 g citrulline, 2 g azelaic acid, 4 g K_2HPO_4,4 g KH_2PO_4, 0.02 g yeast extract and 15 g agar. The pH was adjusted to 5.7. The medium was supplemented with cycloheximide (0.2 g/L) and crystal violet (0.002 g/L) to inhibit growth of fungi and Gram-positive bacteria, respectively (Burbage & Sasser, 1982). After 5 days of incubation at 28 °C, colonies were selected randomly, subcultured on LB medium (per L: 10 g tryptone, 5 g yeast extract, 10 g NaCl, 15 g agar) to check purity and maintained at −80 °C as a suspension with 20 % (v/v) glycerol. Two isolates from soils from Monte Buey, in Cordoba province (32°58′14″ S 62°27′6″ W), MMP81^T (=LMG 27620^T, isolated in 2010) and MAN52 (=LMG 27621, isolated in 2011), were characterized further in the present study.

When analyzing our data, we found highly similar 16S rRNA gene and *gyrB* sequences in the whole-genome sequence of strain YI23, which was isolated from golf course soil in South Korea and which showed the ability to degrade the pesticide fenitrothion (GenBank accession numbers CP003087-CP003092) (Lim *et al.*, 2012).

For PCR, genomic DNA was prepared using the method described by Niemann *et al.* (1997). RAPD patterns were generated using the primers 5' TGC GCG CGG G 3' and 5' AGC GGG

CCA A 3', as described previously (Williams *et al.*, 1990), and demonstrated that isolates LMG 27620^T and LMG 27621 represent genetically distinct strains (Supplementary Fig. 5.1).
Nearly complete sequences of the 16S rRNA genes of strains LMG 27620^T and LMG 27621 were obtained as described previously (Peeters *et al.*, 2013). The MOTHUR software package (Schloss *et al.*, 2009) was used to align the 16S rRNA gene sequences of strains LMG 27620^T, LMG 27621 and YI23 (1484 bp) with those of type strains of phylogenetically related species of the genus *Burkholderia* (1388-1525 bp) against the SILVA reference database (www.arb-silva.de). Phylogenetic analysis was conducted in MEGA5 (Tamura *et al.*, 2011) (Fig. 5.1). Uncorrected pairwise distances were calculated using MEGA5, and all ambiguous positions were removed for each sequence pair. Strains LMG 27620^T, LMG 27621 and YI23 had identical 16S rRNA gene sequences, which were highly similar to those of *B. zhejiangensis* OP-1^T (99.7 % similarity) and *B. grimmiae* R27^T (99.3 %) (Fig. 5.1).

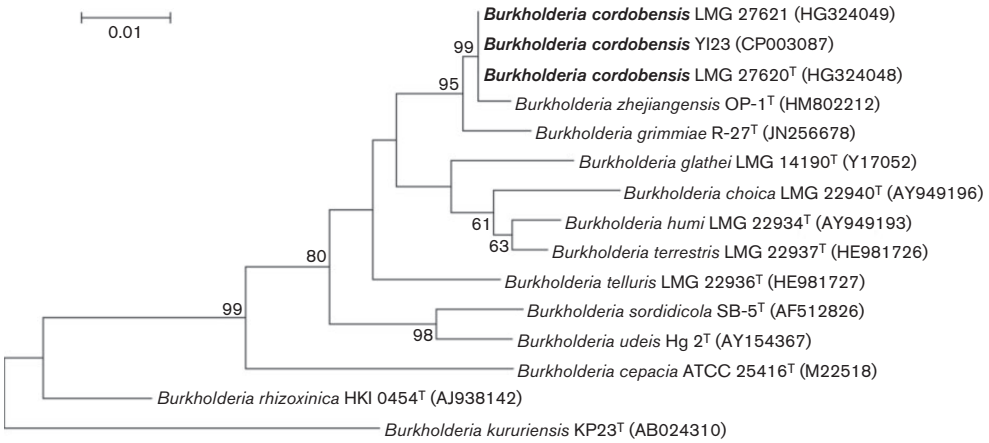


Figure 5.1: Phylogenetic tree based on 16S rRNA gene sequences of strains LMG 27620^T, LMG 27621 and YI23 and phylogenetically related members of the genus *Burkholderia*. The tree was reconstructed using the maximum likelihood method based on the Hasegawa–Kishino–Yano model (Hasegawa *et al.*, 1985). The tree with the highest log-likelihood (-3205.5147) is shown. The percentage of trees in which the associated taxa clustered together (from 1000 bootstrap replicates) is shown next to branches if greater than 50 %. A discrete gamma distribution was used to model evolutionary rate differences among sites [5 categories (+G, parameter=50.0500)] and allowed for some sites to be evolutionarily invariable ([+I], 10.0430 % sites). All positions containing gaps and missing data were eliminated. There were a total of 1335 positions in the final dataset. The 16S rRNA gene sequence of *Burkholderia kururiensis* KP23^T was used as an outgroup. Bar, 0.01 substitutions per site.

Partial sequences of the *gyrB* genes of strains LMG 27620^T and LMG 27621 and their nearest neighbors as determined by 16S rRNA gene sequence proximity were obtained using the method described by Spilker *et al.* (2009) as modified by De Meyer *et al.* (2013). Sequence

assembly was performed using BioNumerics v5.10 (Applied Maths). Sequences (589-704 bp) were aligned based on amino acid sequences using MUSCLE (Edgar, 2004) in MEGA5 (Tamura *et al.*, 2011). Phylogenetic analysis of nucleotide sequences was conducted in MEGA5 (Fig. 5.2) and uncorrected pairwise distances were calculated using MEGA5. Strains LMG 27620^T, LMG 27621 and YI23 formed a homogeneous cluster (sequences were >99 % similar), which was supported by a bootstrap value of 75 % (Fig. 5.2) and grouped with *B. zhejiangensis* LMG 27258^T (95.3-95.9 %) and *B. grimmiae* LMG 27580^T (90.8-91.6 %) as nearest neighbors.

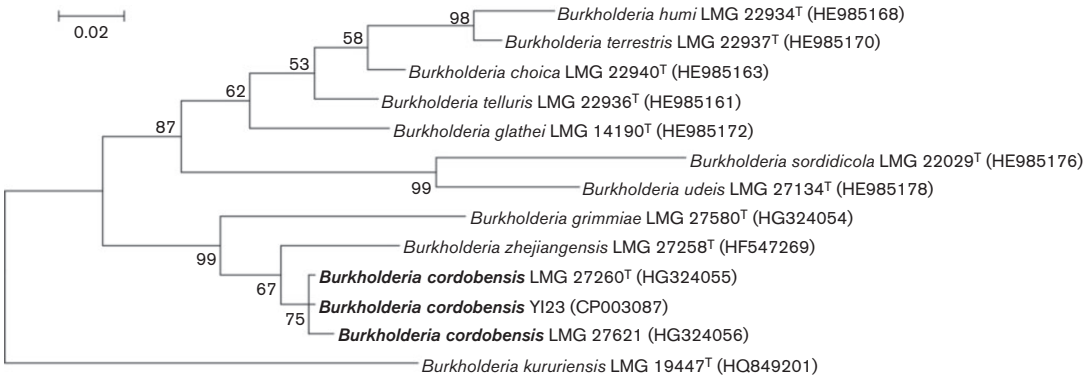


Figure 5.2: Phylogenetic tree based on *gyrB* gene sequences of strains LMG 27620^T, LMG 27621 and YI23 and phylogenetically related members of the genus *Burkholderia*. The tree was reconstructed using the maximum likelihood method based on Tamura-three-parameter model (Tamura, 1992). The tree with the highest log-likelihood (-2409.0868) is shown. The percentage of trees in which the associated taxa clustered together (from 1000 bootstrap replicates) is shown next to branches. A discrete gamma distribution was used to model evolutionary rate differences among sites [5 categories (+G, parameter=50.2933)]. All positions containing gaps and missing data were eliminated. There were a total of 570 positions in the final dataset. The *gyrB* sequence of *Burkholderia kururiensis* LMG 19447^T was used as an outgroup. Bar, 0.02 substitutions per site.

For DDH and the determination of the DNA G+C content, high-molecular-mass DNA was prepared as described by Pitcher *et al.* (1989). DDHs were performed using a microplate method and biotinylated probe DNA (Ezaki *et al.*, 1989). The hybridization temperature was 45±1 °C. Reciprocal reactions (A x B and B x A) were performed and the variation was within the limits of this method (Goris *et al.*, 1998). DDH experiments were performed between strain LMG 27620^T and *B. zhejiangensis* LMG 27258^T, its nearest named phylogenetic neighbor (Fig. 5.1 and 5.2). DDH between the two strains was 44 %. The DNA G+C content was determined by HPLC according to the method of Mesbah & Whitman (1989) using a Waters Breeze HPLC system and XBridge Shield RP18 column thermostabilized at 37 °C.

The solvent was 0.02 M $NH_4H_2PO_4$ (pH 4.0) with 1.5 % (v/v) acetonitrile. Non-methylated lambda phage (Sigma) and *Escherichia coli* DNA were used as calibration reference and control, respectively. The DNA G+C content of strain LMG 27620^T was 63.6 mol%, which is within the range reported for members of the genus *Burkholderia* (59.0-69.9 mol%) (Yabuuchi *et al.*, 1992; Gillis *et al.*, 1995) and which corresponds well to the DNA G+C content of 63.5 mol% reported for the whole-genome sequence of strain YI23 (Lim *et al.*, 2012).
Phenotypic analysis of strains LMG 27620^T and LMG 27621 and *B. zhejiangensis* LMG 27258^T was performed on tryptone soya agar (TSA) at 28 °C unless indicated otherwise (BCCM/LMG medium 14). Cell morphology and motility were observed by phase-contrast microscopy. Oxidase activity was detected by immersion of cells in 1 % N,N,N',N'-tetramethyl p-phenylenediamine solution and catalase activity was determined by bubble formation after flooding a colony with 10 % H_2O_2. Lipase activity was determined according to the method described by Sierra (1957). Growth on MacConkey medium was observed after 48 h of incubation at 28 °C. Starch hydrolysis was observed after 48 h of incubation at 28 °C in LMG 14 medium amended with 2 % starch. DNase activity was observed after 48 h of incubation at 28 °C on BD Difco DNase test agar, according to the method of Jeffries *et al.* (1957). Casein hydrolysis was observed after 48 h of incubation at 28 °C on TSA plates amended with 1.3 % skimmed milk, through the observation of clear haloes around colonies. Other biochemical tests were performed by inoculating API 20NE and API ZYM strips (bioMerieux) according to the manufacturer's instructions and incubating for 48 h at 28 °C or for 4 h at 28 °C, respectively. Growth was tested at 28 °C in nutrient broth (NB; BD Difco) at pH 4-9 using appropriate biological buffers (acetate, citrate/Na_2HPO_4, phosphate buffer and Tris/HCl). Growth on LMG 14 medium was tested at 4, 15, 28, 30, 37, 40 and 45 °C. The results of phenotypic and biochemical tests are given in the species description and in Table 5.1. They allow clear differentiation of the taxon represented by strains LMG 27620^T and LMG 27621 and its nearest phylogenetic neighbor, *B. zhejiangensis*: the former exhibits oxidase activity but not arginine dihydrolase or urease activity, reduces nitrate, hydrolyzes Tween 80 and does not assimilate caprate; opposite test results were obtained for the latter. In addition, on the basis of phenotypic data published by Tian *et al.* (2013) using the same test gallery, strains LMG 27620^T and LMG 27621 can be differentiated from *B. grimmiae* R27^T: the former exhibits no arginine dihydrolase or urease activity and assimilates phenylacetate. In addition, it does not hydrolyze starch and grows on MacConkey agar; opposite test results were obtained for *B. grimmiae* R27^T. Differential characteristics towards other species of the genus *Burkholderia* in this same lineage are presented in Table 5.1.
Whole-cell fatty acid methyl esters were extracted according to the MIDI protocol (www.microbialid.com/PDF/TechNote_101.pdf). All characteristics such as temperature, medium

Table 5.1: Phenotypic characteristics that distinguish *B. cordobensis* sp. nov. from its nearest phylogenetic neighbor

Characteristic	1	2	3	4	5	6	7	8	9	10
Growth at:										
37 °C	+−	+++	+	w	+	−	−	w	−	−
40 °C	−−	+++	+	−	−	−	−	−	−	−
pH 8.0	ww	+++	+	−	−	+	w	−	−	+
Hydrolysis of Tween 60	++	+++	ND	+	−	−	−	+	−	−
Nitrate reduction (API 20NE)	w+	+++	+	−	−	−	+	+	−	+
Beta-galactosidase (API 20NE)	−−	−−−	−	−	−	−	+	−	−	w
Assimilation of (API 20NE)										
Arabinose	+−	+++	+	w	w	−	+	+	−	+
Mannose	++	+++	+	−	+	+	+	+	+	+
Mannitol	++	−−−	+	w	w	+	+	+	+	−
N-Acetylglucosamine	++	+++	+	w	+	+	+	+	+	+
Gluconate	++	+++	+	w	+	+	+	+	+	+
Caprate	−−	+++	−	−	+	w	−	+	w	−
Malate	++	+++	+	w	+	+	v	+	+	+
Citrate	−−	−−−	−	−	+	+	−	+	+	+
Phenylacetate	++	++w	−	−	+	+	−	+	+	−
Enzyme activity (API ZYM)										
C4 lipase	−−	−−−	+	w	+	+	+	+	+	+
C8 lipase	−−	−−−	+	+	w	+	+	w	w	+
Valine arylamidase	−−	−−−	−	w	w	−	−	−	w	+
Cystine arylamidase	−−	−−−	−	−	−	−	−	−	−	−
β-galactosidase	−−	−−−	−	−	−	−	+	−	−	+

Strains: 1. *B. cordobensis* sp. nov. LMG 26720^T and LMG 27621; 2, *B. zhejiangensis* LMG 27258^T, LMG 26180 and LMG 26181; 3, *B. grimmiae* R27^T; 4, *B. choica* LMG 22940^T; 5, *B. glathei* LMG 14190^T; 6, *B. humi* LMG 22934^T; 7, *B. sordidicola* LMG 22029^T; 8, *B. telluris* LMG 22936^T; 9, *B. terrestris* LMG 22937^T; 10, *B. udeis* LMG 27134^T. Results for *B. cordobensis* sp. nov. and its nearest neighbor *B. zhejiangensis* are from the present study; test results of the type strains are given first, followed by the remaining strains in the order given above. Data for *B. grimmiae* R27^T were extracted from Tian *et al.* (2013); data for the remaining strains were taken from Vandamme *et al.* (2013a). +, present; −, absent; w, weak reaction; v, variable; ND, not determined.

and physiological age (overlap area of the second and third quadrant from a quadrant streak for late-exponential-phase growing cells) were as in the MIDI protocol. The profiles were generated using an Agilent Technologies 6890N gas chromatograph and identified and clustered using the Microbial Identification System software and MIDI TSBA database version 5.0. The most abundant fatty acids of strains LMG 27620^T and LMG 27621 were $C_{18:1}$ ω7c, summed feature 3 (probably $C_{16:1}$ ω7c) and $C_{16:0}$. Moderate to small amounts of summed feature 2 (probably $C_{14:0}$ 3-OH), $C_{17:0}$ cyclo, $C_{16:0}$ 3-OH, $C_{14:0}$, $C_{19:0}$ cyclo ω8c, $C_{16:1}$ 2-OH and $C_{16:0}$ 2-OH were also detected. The presence of $C_{16:0}$ 3-OH supports the placement of these strains in the genus *Burkholderia* (Yabuuchi *et al.*, 1992), but the overall profile of the novel taxon is very similar to those of its nearest neighbors (Table 5.2).

The phenotypic, chemotaxonomic and genotypic data from the present study demonstrate that

strains LMG 27620^T, LMG 27621 and YI23 represent a novel species of the genus *Burkholderia* that can be distinguished from its nearest phylogenetic neighbors both phenotypically as well as genotypically. We therefore propose to classify these bacteria in a novel species, for which the name *Burkholderia cordobensis* sp. nov. is proposed.

Table 5.2: Mean fatty acid compositions of *B. cordobensis* sp. nov. and its nearest phylogenetic neighbors

Fatty acid	1	2	3	4	5
$C_{12:0}$	ND	ND	TR	ND	ND
$C_{14:0}$	4.16±0.04	4.08±0.13	4.02	3.54±0.13	4.29±0.13
$C_{16:0}$	17.00±0.80	16.24±0.64	12.14	20.37±4.40	19.11±1.93
$C_{16:0}$ 2-OH	2.49±0.79	3.37±0.82	4.41	3.26±1.54	2.06±0.86
$C_{16:0}$ 3-OH	5.92±0.04	5.91±0.13	5.02	6.93±2.32	6.13±0.49
$C_{16:1}$ 2-OH	TR	TR	4.37	ND	TR
$C_{17:0}$ cyclo	5.31±1.86	4.96±0.65	TR	14.30±4.67	13.05±2.41
$C_{18:1}$ ω7c	36.82±1.2	36.88±1.05	38.78	30.97±1.41	31.10±1.83
$C_{19:0}$ cyclo ω8c	1.82±0.72	2.14±0.37	ND	ND	2.64±1.29
Summed features*					
2	7.21±0.15	6.82±0.35	5.32	7.80±2.28	7.08±0.53
3	17.55±2.11	17.43±0.60	23.16	12.84±4.21	12.79±2.79

Fatty acid	6	7	8	9	10
$C_{12:0}$	ND	3.70	ND	ND	4.63±0.07
$C_{14:0}$	3.71±0.27	0.79	3.00±0.01	3.80	TR
$C_{16:0}$	12.00±0.76	18.29	13.31±0.15	16.02	15.72±0.96
$C_{16:0}$ 2-OH	2.43±0.55	3.32	1.61±0.14	2.29	2.05±0.10
$C_{16:0}$ 3-OH	5.17±0.38	4.90	5.05±0.04	5.36	6.75±0.91
$C_{16:1}$ 2-OH	1.71±0.32	4.06	TR	1.29	2.02±0.22
$C_{17:0}$ cyclo	7.64±2.05	13.94	4.54±0.13	6.58	6.57±0.99
$C_{18:1}$ ω7c	42.19±1.77	27.61	43.66±0.55	38.00	32.56±1.27
$C_{19:0}$ cyclo ω8c	TR	7.01	TR	ND	3.61±0.68
Summed features*					
2	6.24±0.42	4.86	5.71±0.08	6.18	7.53±0.60
3	17.14±1.55	11.53	20.55±0.57	18.98	17.20±1.46

Species/strains: 1, *B. cordobensis* sp. nov. LMG 27620^T and LMG 27621; 2, *B. zhejiangensis* LMG 27258^T, LMG 26180 and LMG 26181; 3, *B. grimmiae* LMG 27258^T; 4, *B. choica*; 5, *B. glathei*; 6, *B. humi*; 7, *B. sordidicola*; 8, *B. telluris*; 9, *B. terrestris*; 10, *B. udeis*. Data for *B. cordobensis* sp. nov. and *B. grimmiae* LMG 27258^T are from the present study. The remaining data were extracted from Vandamme *et al.* (2013a) and were generated under the same cultivation and analysis conditions. Values are mean±SD percentages of total fatty acids. Those fatty acids for which the mean amount for all taxa was <1 % are not included; therefore, the percentages may not add up to 100 %. TR, trace amount (<1 %); ND, not detected. *Summed feature 2 comprised iso-$C_{16:1}$ and/or $C_{14:0}$ 3-OH); summed feature 3 comprised iso-$C_{15:0}$ 2-OH and/or $C_{16:1}$ ω7c.

5.1.2.1 Description of *Burkholderia cordobensis* sp. nov.

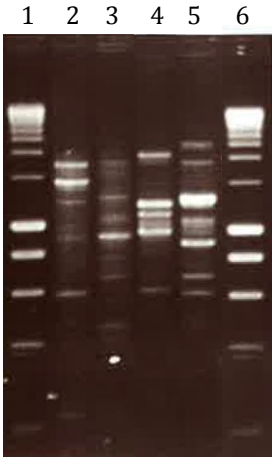
Burkholderia cordobensis (cor.do.ben'sis. N.L. fem. adj. *cordobensis* pertaining to the Argentinian province of Cordoba, where the first strains were isolated).
Gram-negative, aerobic, motile, non-sporeforming rods, about 0.4-0.7 μm wide and 1.2-1.8 μm long. Colonies are round, with entire margins, a convex elevation, a white-creamy color and a moist appearance, and are 1-2 mm in diameter after 48 h of growth on TSA at 28 °C. Grows on MacConkey agar. Growth occurs at 28-37 °C and at pH 6-8 in NB at 28 °C. Catalase and oxidase activities are present, but not arginine dihydrolase, urease, β-glucosidase, gelatinase or β-galactosidase. When tested by using API ZYM strips, activities of the following enzymes are present: alkaline and acid phosphatases, leucyl arylamidase and phosphoamidase; activities are absent for C4 lipase, C8 lipase, C14 lipase, valine and cystine arylamidases, trypsin, chymotrypsin, α-galactosidase, β-galactosidase, β-glucuronidase, α-glucosidase, β-glucosidase, N-acetyl-β-glucosaminidase, α-mannosidase and α-fucosidase. Tweens 20, 40, 60 and 80 are degraded, but casein and starch are not hydrolyzed. When tested by using API 20NE strips, nitrate is reduced and glucose, arabinose (variable; type strain positive), mannose, mannitol, N-acetylglucosamine, gluconate, malate and phenylacetate are assimilated, but not maltose, caprate, adipate or citrate. Negative for fermentation of glucose, activities of tryptophanase, arginine dihydrolase, urease and β-galactosidase (PNPG), hydrolysis of aesculin and gelatin liquefaction. The most abundant fatty acids are $C_{18:1}$ ω7c, summed feature 3 (probably $C_{16:1}$ ω7c) and $C_{16:0}$.
The type strain, MMP81^T (=LMG 27620^T =CCUG 64368^T), was isolated from agricultural soils in Cordoba province, Argentina. The DNA G+C content of the type strain is 63.6 mol%. The whole-genome sequence of strain YI23 is 8.89 Mb and consists of three chromosomes and three plasmids (Lim *et al.*, 2012).

5.1.3 Acknowledgments

This work was partially supported by the Agencia Nacional de Promocion Cientifica y Tecnologica, Argentina (grants PAE 36976-PID 52 and PICT 2010/2691). W.O.D., A.Z. and L.G.W. are members of Research Career at CONICET, Argentina.

5.1.4 Supplementary material

The GenBank/EMBL/DDBJ accession numbers for the 16S rRNA and *gyrB* gene sequences of strains LMG 27620^T and LMG 27621 are HG324048 and HG324055 (LMG 27620^T) and HG324049 and HG324056 (LMG 27621), respectively. The accession number for the *gyrB* gene sequence of *B. grimmiae* LMG 27580^T is HG324054.



Supplementary Figure 5.1: RAPD patterns of strains LMG 27620^{T} (lanes 2 and 4) and LMG 27621 (lanes 3 and 5) obtained using the primers TGC GCG CGG G (lanes 2 and 3) and AGC GGG CCA A (lanes 4 and 5). Lanes 1 and 6 are marker lanes.

5.2 Phylogenomic study of *Burkholderia glathei*-like organisms, proposal of 13 novel *Burkholderia* species and emended descriptions of *Burkholderia sordidicola*, *Burkholderia zhejiangensis* and *Burkholderia grimmiae*

Redrafted from: Charlotte Peeters, Jan P. Meier-Kolthoff, Bart Verheyde, Evie De Brandt, Vaughn S. Cooper and Peter Vandamme (2016). Phylogenomic study of *Burkholderia glathei*-like organisms, proposal of 13 novel *Burkholderia* species and emended descriptions of *Burkholderia sordidicola*, *Burkholderia zhejiangensis* and *Burkholderia grimmiae*. Frontiers in Microbiology, 7, 1-19.

Author contributions: CP and PV conceived the study and wrote the manuscript. CP performed all experiments except the phenotypical characterization and analyzed the data. CP and EDB performed the DNA extractions. EDB performed the phenotypical characterization. VC directed the genomic sequencing methods and initial analysis. JMK performed all GBDP-related analyses. CP and BV performed the ortholog analysis and whole-genome sequence-based phylogenetic analysis. JMK and VC proofread the manuscript.

5.2.1 Abstract

Partial *gyrB* gene sequence analysis of 17 isolates from human and environmental sources revealed 13 clusters of strains and identified them as BGC bacteria. The taxonomic status of these clusters was examined by whole-genome sequence analysis, determination of the G+C content, whole-cell fatty acid analysis and biochemical characterization. The whole-genome sequence-based phylogeny was assessed using the GBDP method and an extended MLSA approach. The results demonstrated that these 17 BGC isolates represented 13 novel *Burkholderia* species that could be distinguished by both genotypic and phenotypic characteristics. BGC strains exhibited a broad metabolic versatility and developed beneficial, symbiotic and pathogenic interactions with different hosts. Our data also confirmed that there is no phylogenetic subdivision in the genus *Burkholderia* that distinguishes beneficial from pathogenic strains. We therefore propose to formally classify the 13 novel BGC *Burkholderia* species as *Burkholderia arvi* sp. nov. (type strain LMG 29317^T =CCUG 68412^T), *Burkholderia hypogeia* sp. nov. (type strain LMG 29322^T =CCUG 68407^T), *Burkholderia ptereochthonis* sp. nov. (type strain LMG 29326^T =CCUG 68403^T), *Burkholderia glebae* sp. nov. (type strain LMG 29325^T =CCUG 68404^T), *Burkholderia pedi* sp. nov. (type strain LMG 29323^T

=CCUG 68406^T), *Burkholderia arationis* sp. nov. (type strain LMG 29324^T =CCUG 68405^T), *Burkholderia fortuita* sp. nov. (type strain LMG 29320^T =CCUG 68409^T), *Burkholderia temeraria* sp. nov. (type strain LMG 29319^T =CCUG 68410^T), *Burkholderia calidae* sp. nov. (type strain LMG 29321^T =CCUG 68408^T), *Burkholderia concitans* sp. nov. (type strain LMG 29315^T =CCUG 68414^T), *Burkholderia turbans* sp. nov. (type strain LMG 29316^T =CCUG 68413^T), *Burkholderia catudaia* sp. nov. (type strain LMG 29318^T =CCUG 68411^T) and *Burkholderia peredens* sp. nov. (type strain LMG 29314^T =CCUG 68415^T). Furthermore, we present emended descriptions of the species *Burkholderia sordidicola*, *Burkholderia zhejiangensis* and *Burkholderia grimmiae*.

5.2.2 Introduction

The genus *Burkholderia* currently comprises 90 validly named species (Euzeby, 1997) and several uncultured *Candidatus* species (Van Oevelen *et al.*, 2004; Verstraete *et al.*, 2011; Lemaire *et al.*, 2012) which occupy very diverse niches (Coenye & Vandamme, 2003). Many *Burkholderia* species have thus far only been isolated as free-living organisms but a growing body of literature reveals that they live in close interaction with numerous plant, animal, fungal or even amoebozoan hosts (Marolda *et al.*, 1999; Van Borm *et al.*, 2002; Kikuchi *et al.*, 2011; Verstraete *et al.*, 2013; Stopnisek *et al.*, 2016; Xu *et al.*, 2016). *Burkholderia* species may be beneficial to their hosts because some strains can fix nitrogen, produce plant hormones or siderophores, or lower pathogen-related ethylene levels; hence they have been exploited for plant growth promotion and biocontrol of plant diseases (Compant *et al.*, 2008; Vial *et al.*, 2011). Yet, other *Burkholderia* species are notorious pathogens in plants, animals and humans (Mahenthiralingam *et al.*, 2008). This ecological diversity is likely attributed to their large, multireplicon genomes (typically between 6 and 9 Mb) which also confer a metabolic versatility allowing them to degrade a wide range of recalcitrant xenobiotics (Parke & Gurian-Sherman, 2001; Coenye & Vandamme, 2003).

Phylogenetic analyses based on the 16S rRNA and protein-coding genes showed that BGC species are phylogenetically divergent from other *Burkholderia* species and form a separate clade (Sawana *et al.*, 2014; Vandamme *et al.*, 2014). Although this clade thus far includes only 12 formally named species, its functional diversity is impressive. In this clade too, most species have been isolated from bulk and rhizosphere soil (Zolg & Ottow, 1975; Viallard *et al.*, 1998; Vandamme *et al.*, 2013a; Draghi *et al.*, 2014; Baek *et al.*, 2015), but also from contaminated soil and sludge from a wastewater treatment system (Lu *et al.*, 2012; Vandamme *et al.*, 2013a; Liu *et al.*, 2014). Two BGC species were associated with less studied hosts like fungi (*B. sordidicola*) and mosses (*B. grimmiae*) (Lim *et al.*, 2003; Tian

et al., 2013) but numerous, mostly uncultivated BGC species adopted endosymbiotic lifestyles in insect guts (Kikuchi *et al.*, 2011; Tago *et al.*, 2015; Xu *et al.*, 2016) or plant leaf tissue (Verstraete *et al.*, 2013; Carlier *et al.*, 2016) and many additional unclassified *B. glathei*-like bacteria have been reported (Nogales *et al.*, 2001; Salles *et al.*, 2006b; Pumphrey & Madsen, 2008; Draghi *et al.*, 2014; Verstraete *et al.*, 2014; Peeters *et al.*, 2016).
The present study aimed to perform a phylogenomic study of established and novel species in the *B. glathei* clade, to formally name the latter and to make reference cultures and whole-genome sequences of each of these versatile bacteria publicly available. The genome sequence-based phylogeny was assessed using the GBDP method (Meier-Kolthoff *et al.*, 2013) and an extended MLSA approach. For phenotypic characterization, whole-cell fatty acid profiling and biochemical analyses were performed.

5.2.3 Materials and methods

5.2.3.1 Bacterial strains and growth conditions

Table 5.3 lists the sources of the 17 studied isolates. Details of type strains of each of the present BGC species were described previously (Zolg & Ottow, 1975; Lim *et al.*, 2003; Lu *et al.*, 2012; Tian *et al.*, 2013; Vandamme *et al.*, 2013a; Draghi *et al.*, 2014; Liu *et al.*, 2014; Baek *et al.*, 2015). Strains were grown aerobically on buffered nutrient agar (Oxoid, pH 6.8) and incubated at 28 °C. Cultures were preserved in MicroBank™ vials at −80 °C.

5.2.3.2 16S rRNA gene sequence analysis

Nearly complete sequences were obtained as described previously (Peeters *et al.*, 2013).

5.2.3.3 *gyrB* gene sequence analysis

Partial *gyrB* gene sequences were obtained as described previously (Spilker *et al.*, 2009; Peeters *et al.*, 2013). Sequence assembly was performed using BioNumerics v7.5 (Applied Maths). Sequences (589-1182 bp) were aligned based on amino acid sequences using Muscle (Edgar, 2004) in MEGA6 (Tamura *et al.*, 2013). All positions with less than 95 % site coverage were eliminated, resulting in a total of 570 positions in the final dataset. Phylogenetic analysis was conducted in MEGA6 (Tamura *et al.*, 2013).

5.2.3.4 Whole-genome sequencing

Genomic DNA of 20 strains (Table 5.4) was prepared as described by Pitcher *et al.* (1989). Genomic libraries were prepared using the Nextera kit following the methods introduced by

Table 5.3: Strains included in the present study

Strain	Other strains designations	Source	Depositor	Reference[a]
***Burkholderia arvi* sp. nov.**				
LMG 29317^T	CCUG 68412^T, MAN34^T	Soil (Argentina, 2010)	Walter Draghi	(1)
***Burkholderia hypogeia* sp. nov.**				
LMG 29322^T	CCUG 68407^T	Soil (Belgium, 2014)	Own isolate	(2)
***Burkholderia ptereochthonis* sp. nov.**				
LMG 29326^T	CCUG 68403^T	Soil (Belgium, 2014)	Own isolate	(2)
***Burkholderia glebae* sp. nov.**				
LMG 29325^T	CCUG 68404^T	Soil (Belgium, 2014)	Own isolate	(2)
LMG 22938	RA57-7	Soil (Netherlands)	Joana Salles	(3)
***Burkholderia pedi* sp. nov.**				
LMG 29323^T	CCUG 68406^T	Soil (Belgium, 2014)	Own isolate	(2)
R-52605		Soil (Belgium, 2014)	Own isolate	(2)
***Burkholderia arationis* sp. nov.**				
LMG 29324^T	CCUG 68405^T	Soil (Belgium, 2014)	Own isolate	(2)
R-23361	RG47-6	Soil (Netherlands)	Joana Salles	(3)
***Burkholderia fortuita* sp. nov.**				
LMG 29320^T	CCUG 68409^T	Soil (South Africa, 2013)	Brecht Verstraete	(4)
***Burkholderia temeraria* sp. nov.**				
LMG 29319^T	CCUG 68410^T	Soil (South Africa, 2013)	Brecht Verstraete	(4)
***Burkholderia calidae* sp. nov.**				
LMG 29321^T	CCUG 68408^T	Water (Belgium, 2013)	Own isolate	(2)
***Burkholderia concitans* sp. nov.**				
LMG 29315^T	CCUG 68414^T, AU12121^T	Lung tissue (USA, 2006)	John J. LiPuma	
R-46586	AU21394	Blood (USA, 2010)	John J. LiPuma	
***Burkholderia turbans* sp. nov.**				
LMG 29316^T	CCUG 68413^T, HI4065^T	Pleural fluid (USA, 2006)	John J. LiPuma	
***Burkholderia catudaia* sp. nov.**				
LMG 29318^T	CCUG 68411^T	Soil (South Africa, 2013)	Brecht Verstraete	(4)
***Burkholderia peredens* sp. nov.**				
LMG 29314^T	CCUG 68415^T, NF100^T	Soil (Japan)	M. Hayatsu	(5)

LMG, BCCM/LMG Bacteria Collection, Laboratory of Microbiology, Ghent University, Ghent, Belgium. [a](1) Draghi *et al.* (2014), (2) sections 3.2 and 3.3, Peeters *et al.* (2016), (3) Salles *et al.* (2006b), (4) Verstraete *et al.* (2014), (5) Hayatsu *et al.* (2000).

Baym *et al.* (2015) and the 151 bp paired-end libraries were sequenced on the Illumina HiSeq platform of the University of New Hampshire Hubbard Center for Genomics Studies with an average insert size of 386 bp. Quality reports were created by FastQC. Adaptors and low-quality reads were trimmed using Trimmomatic (Bolger *et al.*, 2014) with the following options: ILLUMINACLIP:NexteraPE-PE.fa:2:30:10 MAXINFO:60:0.4 MINLEN:60. Assembly was performed using SPAdes (Bankevich *et al.*, 2012) with default k-mer sizes (21, 33, 55, 77) and mismatch correction (option –careful). Contigs with length $<$500 bp and coverage $<$2 were discarded from the resulting assemblies. Raw reads were mapped against the assemblies using bwa-mem (Li, 2013) and contigs were polished using Pilon (Walker *et al.*, 2014b) with default parameters. Quast (Gurevich *et al.*, 2013) was used to create quality reports of the resulting assemblies. Annotation was performed using Prokka 1.11 (Seemann, 2014) with a genus-specific database based on reference genomes from the *Burkholderia* Genome Database (Winsor *et al.*, 2008).

Table 5.4: Genomes included in the present study

Strain	Project	Contigs[a]	Size (bp)	%GC	Reference[b]
B. glathei LMG 14190^T	PRJEB6934	139	8,049,485	64.7	(1)
B. sordidicola LMG 22029^T	PRJEB12475	72	6,874,511	60.2	This study
B. zhejiangensis OP-1^T	PRJNA238427	116	7,767,215	62.7	(2)
B. grimmiae R27^T	PRJNA238424	160	6,704,301	63.0	(2)
B. choica LMG 22940^T	PRJEB12479	657	9,776,207	62.7	This study
B. humi LMG 22934^T	PRJEB12476	272	7,619,203	62.8	This study
B. telluris LMG 22936^T	PRJEB12477	163	7,056,109	64.0	This study
B. terrestris LMG 22937^T	PRJEB12478	645	8,201,357	62.6	This study
B. udeis LMG 27134^T	PRJEB12480	242	10,051,569	60.0	This study
B. cordobensis LMG 27620^T	PRJEB12481	74	8,208,096	63.7	This study
B. jiangsuensis MP-1^T	PRJNA238425	168	8,611,053	62.6	(2)
B. megalochromosomata JC2949^T	PRJNA241423[c]	285	9,506,519	62.7	(3)
B. arvi sp. nov. LMG 29317^T	PRJEB12485	351	9,665,767	62.4	This study
B. hypogeia sp. nov. LMG 29322^T	PRJEB12491	94	8,333,271	63.2	This study
B. ptereochthonis sp. nov. LMG 29326^T	PRJEB12495	117	7,714,803	64.2	This study
B. glebae sp. nov. LMG 29325^T	PRJEB12494	194	7,842,312	62.7	This study
B. pedi sp. nov. LMG 29323^T	PRJEB12492	142	9,141,307	63.0	This study
B arationis sp. nov. LMG 29324^T	PRJEB12493	629	9,377,494	62.8	This study
B. fortuita sp. nov. LMG 29320^T	PRJEB12489	50	7,360,810	62.9	This study
B. temeraria sp. nov. LMG 29319^T	PRJEB12488	129	8,325,519	62.7	This study
B. calidae sp. nov. LMG 29321^T	PRJEB12490	379	9,609,693	62.5	This study
B. concitans sp. nov. LMG 29315^T	PRJEB12483	47	6,166,171	63.2	This study
B. turbans sp. nov. LMG 29316^T	PRJEB12484	120	7,352,555	63.1	This study
B. catudaia sp. nov. LMG 29318^T	PRJEB12486	156	7,726,733	62.8	This study
B. peredens sp. nov. LMG 29314^T	PRJEB12482	78	6,726,081	63.1	This study
B. cordobensis YI23	PRJNA74517	6	8,896,411	63.3	(4)
Burkholderia sp. PML1(12)	PRJNA53985	377	9,368,249	60.1	(5)
Burkholderia sp. S170	PRJNA248610	216	10,261,891	59.6	(6)
B. zhejiangensis CEIB S4-3	PRJNA264584	154	7,666,841	62.8	(7)
B. zhejiangensis SJ98	PRJNA81431	14	7,878,727	62.7	(8)
Burkholderia sp. Leaf177	PRJNA297956	27	6,804,288	59.2	(9)
B. concitans sp. nov. MR1	PRJNA269162	58	6,019,671	63.3	(10)
Burkholderia sp. RPE64	PRJDB1103	5	6,964,487	63.2	(11)
B. cordobensis RPE67	PRJDB1660	6	8,685,756	63.4	(12)
Ca. B. kirkii UZHbot1	PRJNA69825	305	3,990,738	62.9	(13)
Ca. B. kirkii UZHbot2	PRJNA253356	48	3,914,712	64.0	(14)
Ca. B. pumila UZHbot3	PRJNA253357	519	3,681,223	59.3	(14)
Ca. B. verschuerenii UZHbot4	PRJNA253359	446	6,188,480	61.9	(14)
Ca. B. humilis UZHbot5	PRJNA253360	354	5,148,994	60.1	(14)
Ca. B. calva UZHbot6	PRJNA253361	307	4,208,605	61.4	(14)
Ca. B. brachyanthoides UZHbot7	PRJNA253362	684	3,545,532	61.2	(14)
Ca. B. schumannianae UZHbot8	PRJNA253363	283	2,362,726	63.1	(14)
Ca. B. crenata UZHbot9	PRJNA253365	643	2,843,741	59.0	(15)

[a]Status complete: RPE64, RPE67, YI23; status draft assembly: all other genomes. [b](1) Stopnisek *et al.* (2016), (2) Liu *et al.* (2014), (3) (Baek *et al.*, 2015), (4) (Lim *et al.*, 2012), (5) Uroz & Oger (2015), (6) Llado *et al.* (2014), (7) Hernandez-Mendoza *et al.* (2014), (8) Kumar *et al.* (2012), (9) Bai *et al.* (2015), (10) Pawitwar *et al.* (2015), (11) Shibata *et al.* (2013), (12) Takeshita *et al.* (2014), (13) Carlier & Eberl (2012), (14) Pinto-Carbo *et al.* (2016), (15) Carlier *et al.* (2016). [c]Genome sequence not publicly available, contig sequences were provided by J. Chun (Baek *et al.*, 2015).

5.2.3.5 Publicly available genomes

Twenty three publicly available whole-genome sequences of BGC bacteria were downloaded from the NCBI database (Table 5.4). *B. gladioli* BSR3 (Seo *et al.*, 2011) was used as an outgroup in all phylogenomic analyses. For *B. megalochromosomata* JC2949^T the whole-genome sequence was not publicly available (February 1st, 2016) and the contig sequences were provided by J. Chun (Baek *et al.*, 2015). For *B. sordidicola* S170, *B. zhejiangensis* CEIB S4-3 and *B. megalochromosomata* JC2949^T no annotation was available and annotation was performed using Prokka as described above.

5.2.3.6 Phylogenomic analysis

The latest version of the GBDP approach was applied (Meier-Kolthoff *et al.*, 2013) to calculate the intergenomic distance between each pair of genomes (based on the nucleotide data) and included the calculation of 100 replicate distances to assess pseudo-bootstrap support (Meier-Kolthoff *et al.*, 2014a). Distance calculations were conducted under the recommended settings of the Genome-to-Genome Distance Calculator (GGDC 2.1; http://ggdc.dsmz.de), as described earlier (Meier-Kolthoff *et al.*, 2013). The GBDP trimming algorithm and formula d_5 were chosen because of their advantages for phylogenetic inference (Meier-Kolthoff *et al.*, 2014a) and according distance matrices were prepared (a single matrix for the original distances plus 100 matrices containing the replicates). A phylogenomic tree with branch support (Meier-Kolthoff *et al.*, 2014a) was inferred using FastME v2.07 with tree bisection and reconnection post-processing (Lefort *et al.*, 2015). Moreover, pairwise digital DNA-DNA hybridization (dDDH) values and their confidence intervals were also determined using GGDC 2.1 under recommended settings (Meier-Kolthoff *et al.*, 2013). The potential affiliation of the novel strains to existing species was determined by clustering using a 70 % dDDH radius around each of the 12 BGC type strains as previously applied (Liu *et al.*, 2015). Visualization and annotation of the phylogenetic tree was performed using iTOL (Letunic & Bork, 2011).

As an alternative for the GBDP method, an extended MLSA analysis was performed in which a whole-genome phylogeny was calculated based on single-copy orthologous genes as described previously (Pinto-Carbo *et al.*, 2016). In short, single-copy orthologs were identified using blastp and OrthoMCL v2.0.9 (with e-value cut-off 1E-6 and 50 % match cut-off) (Fischer *et al.*, 2011) and aligned based on their amino acid sequences using MUSCLE. The alignments were trimmed using TrimAl (removing positions with gaps in more than 50 % of the sequences) and concatenated to construct a maximum likelihood tree using RaXML v7.4.2 (Stamatakis, 2014) with the WAG amino acid substitution model and 100 rapid bootstrap analyses.

5.2.3.7 Phenotypic characterization

Phenotypic and cellular fatty acid analyses were performed as described previously (Draghi *et al.*, 2014).

5.2.4 Results

5.2.4.1 16S rRNA gene sequence analysis

The 16S rRNA gene sequences determined in the present study are publicly available through the GenBank/EMBL/DDBJ accession numbers LT158612-LT158624.

5.2.4.2 *gyrB* gene sequence analysis

Partial *gyrB* gene sequences were compared to those of the type strains of the 12 validly named BGC species (Fig. 5.3). The 17 unclassified isolates represented 13 taxa which showed 83.4-96.2 % pairwise identity with the *gyrB* sequences of the type strains of other BGC species. The *gyrB* gene sequences determined in the present study are publicly available through the GenBank/EMBL/DDBJ accession numbers LT158625-LT158641.

5.2.4.3 Whole-genome sequencing

To further characterize the taxonomic status of these 13 taxa, we determined the whole-genome sequence of one strain per *gyrB* cluster and of *B. sordidicola* LMG 22029^T, *B. choica* LMG 22940^T, *B. humi* LMG 22934^T, *B. telluris* LMG 22936^T, *B. terrestris* LMG 22937^T, *B. udeis* LMG 27134^T and *B. cordobensis* LMG 27620^T. The assembly of the Illumina HiSeq 150 bp paired end reads resulted in assemblies with 47 to 657 contigs and a total of 6,166,171 to 10,051,569 bp (Table 5.4). The annotated assemblies of these 20 genomes were submitted to the European Nucleotide Archive and are publicly available through the GenBank/EMBL/DDBJ accession numbers listed in Table 5.4 and the species descriptions. The genome sequences of the remaining five BGC type strains and of 18 additional strains were publicly available (Table 5.4).

5.2.4.4 DNA base composition

The G+C content of all type strains was calculated from their genome sequences and ranged from 62.4 to 64.2 mol% (Table 5.4).

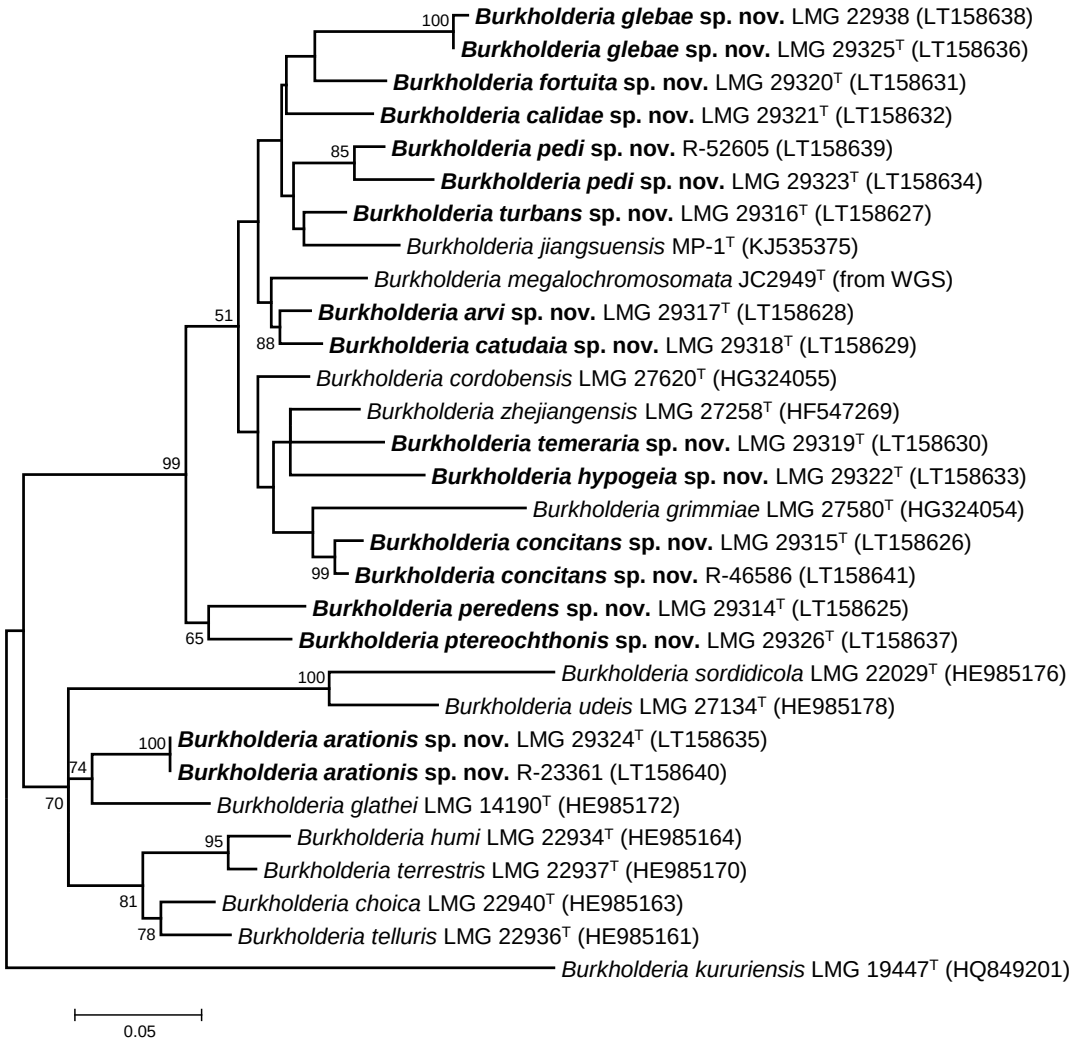


Figure 5.3: Phylogenetic tree based on partial *gyrB* sequences of the 17 isolates in this study and type strains of phylogenetically related *Burkholderia* species. The optimal tree (highest log likelihood) was constructed using the maximum likelihood method and general time reversible model in MEGA6 (Tamura *et al.*, 2013). A discrete gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.5462)) and allowed for some sites to be evolutionarily invariable ([+I], 37.9331 % sites). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches if greater than 50 %. For *B. megalochromosomata* JC2949^T the *gyrB* gene sequence was extracted from the genome sequence. The *gyrB* sequence of *B. kururiensis* LMG 19447^T was used as outgroup. The scale bar indicates the number of substitutions per site.

5.2.4.5 Phylogenomic analysis

The pairwise intergenomic distances and dDDH estimates of the 44 genome sequences are listed in Supplementary Table 5.1. The phylogenetic tree inferred from the intergenomic distances (Fig. 5.4) was well resolved and most branches showed a very high bootstrap support (average support: 94.8 %). Species delineation based on the pairwise dDDH values and a 70 % dDDH radius around each type strain yielded 39 species which included the present 12 validly named species as well as the 13 novel species delineated by means of partial *gyrB* gene sequences (Fig. 5.3).
For the extended MLSA approach, we identified 332 single-copy orthologs that were present in all 44 genomes. The maximum likelihood phylogenetic tree based on the concatenated amino acid alignment (Fig. 5.5) was well resolved and showed a high bootstrap support on almost all branches.
The topologies of the two phylogenomic trees (Fig. 5.4 and 5.5) were very similar and both revealed six clusters of species (A-F). The main difference in tree topology related to the phylogenetic position of the *Candidatus* species in cluster C. This cluster was supported by a 100 % bootstrap value in both analyses but its relative position to cluster D species differed in the two trees (Fig. 5.4 and 5.5). Additionally, the internal branching order of cluster C, E and F species differed minimally between both analyses. Both phylogenomic analyses showed that strain MR1 clustered with *B. concitans* sp. nov. and that strain RPE67 clustered with *B. cordobensis*. Finally, the large distances between strains PML1(12) and S170, and the type strains of *B. glathei* and *B. sordidicola*, respectively, indicated that both strains were misidentified and wrongly annotated in the NCBI database as *B. glathei* and *B. sordidicola*, respectively (Fig. 5.4 and 5.5). Both strains occupy unique positions in the phylogenomic trees and represent additional novel BGC species.

5.2.4.6 Cellular fatty acid analysis

The fatty acid profiles of all strains are shown in Table 5.5. Branched chain fatty acids have not been reported in members of the genus *Burkholderia* and therefore summed features 2 and 3 very likely represent $C_{14:0}$ 3-OH and $C_{16:1}$ ω7c, respectively (Yabuuchi *et al.*, 1992). The main fatty acid components are $C_{16:0}$, $C_{18:1}$ ω7c and summed feature 3 (most probably representing $C_{16:1}$ ω7c).

5.2.4.7 Biochemical characterization

An overview of biochemical characteristics useful for distinguishing the BGC species is shown in Table 5.6.

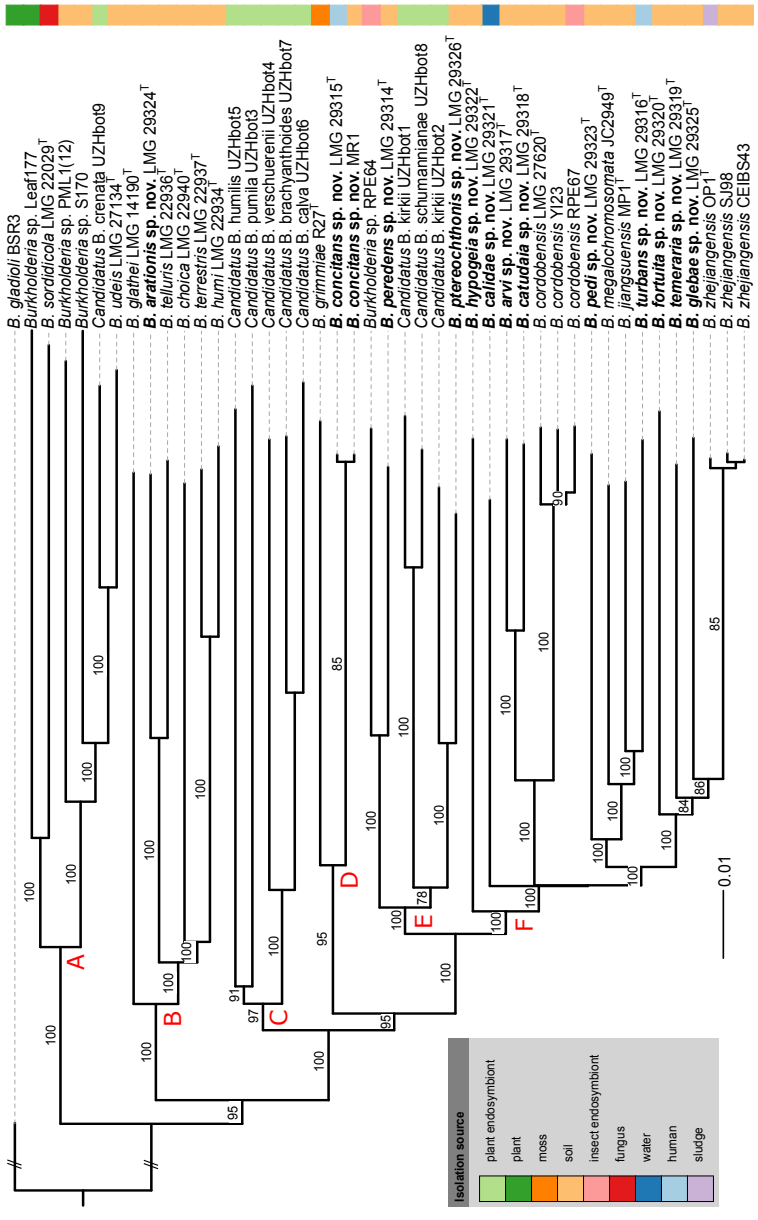


Figure 5.4: Whole-genome sequence-based phylogenomic tree of all BGC genomes inferred by GBDP. The outer column shows the isolation source of the strains. Pseudo-bootstrap support values above 60 % are shown. The tree reveals a high average support of 94.8 %. Long terminal branches are due to the distinct scaling used by GBDP's formula d_5. *B. gladioli* BSR3 was used as outgroup. Red capital letters define subtrees that also occur in the tree depicted in Fig. 5.5.

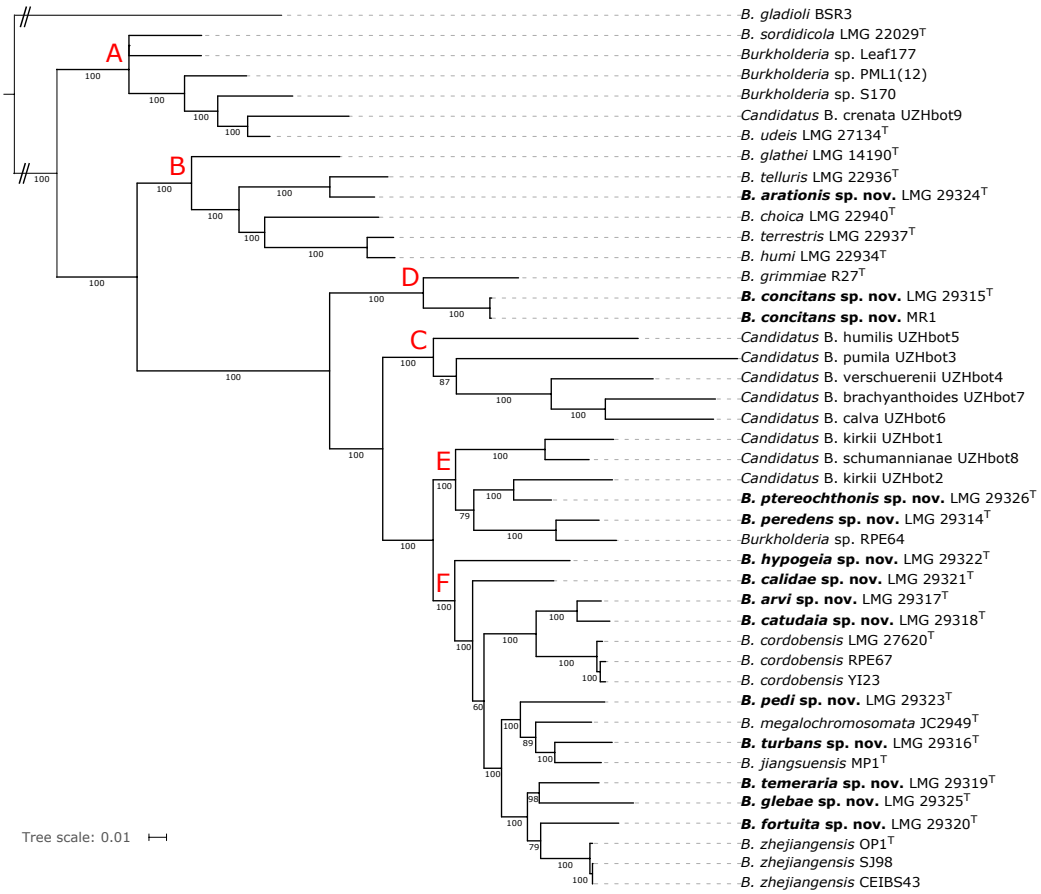


Figure 5.5: Whole-genome phylogeny based on single-copy orthologs of all BGC genomes. The phylogenetic tree was constructed using the WAG protein substitution model and RAxML and is based on an amino acid alignment with 105,141 positions from 332 single-copy orthologous genes. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) are shown next to the branches if greater than 60 %. *B. gladioli* BSR3 was used as outgroup. Red capital letters define subtrees that also occur in the tree depicted in Fig. 5.4.

Table 5.5: Mean fatty acid composition of all examined strains of BGC species

Fatty acid	1	2	3	4	5	6	7	8	9	10	11	12
$C_{12:0}$	ND	3.70	ND	TR	ND	ND	ND	ND	4.63±0.07	ND	ND	1.01
$C_{14:0}$	4.29±0.13	0.79	4.08±0.13	4.02	3.54±0.13	3.71±0.27	3.00±0.01	3.80	TR	4.16±0.04	4.36	3.97
$C_{16:0}$	19.11±1.93	18.29	16.24±0.64	12.14	20.37±4.40	12.00±0.76	13.31±0.15	16.02	15.72±0.96	17.00±0.80	15.58	15.23
$C_{16:0}$ 2-OH	2.06±0.86	3.32	3.37±0.82	4.41	3.26±1.54	2.43±0.55	1.61±0.14	2.29	2.05±0.10	2.49±0.79	4.10	1.53
$C_{16:0}$ 3-OH	6.13±0.49	4.90	5.91±0.13	5.02	6.93±2.32	5.17±0.38	5.05±0.04	5.36	6.75±0.91	5.92±0.04	6.44	4.89
$C_{16:1}$ 2-OH	TR	4.06	TR	4.37	ND	1.71±0.32	TR	1.29	2.02±0.22	TR	1.77	TR
$C_{17:0}$ cyclo	13.05±2.41	13.94	4.96±0.65	TR	14.30±4.67	7.64±2.05	4.54±0.13	6.58	6.57±0.99	5.31±1.86	3.96	TR
$C_{18:1}$ ω7c	31.10±1.83	27.61	36.88±1.05	38.78	30.97±1.41	42.19±1.77	43.66±0.55	38.00	32.56±1.27	36.62±1.20	35.48	42.25
$C_{19:0}$ cyclo ω8c	2.64±1.29	7.01	2.14±0.37	ND	ND	TR	TR	ND	3.61±0.68	1.82±0.72	1.73	TR
Summed feature 2	7.08±0.53	4.86	6.82±0.35	5.32	7.80±2.28	6.24±0.42	5.71±0.08	6.18	7.53±0.60	7.21±0.15	8.13	7.04
Summed feature 3	12.79±2.79	11.53	17.43±0.60	23.16	12.84±4.21	17.14±1.55	20.55±0.57	18.98	17.20±1.46	17.55±2.11	18.44	20.64

Fatty acid	13	14	15	16	17	18	19	20	21	22	23	24	25
$C_{12:0}$	ND	ND	ND	TR	TR	ND	ND	ND	ND	TR	ND	ND	ND
$C_{14:0}$	4.13	4.40	4.18	3.91±0.13	4.11±0.28	4.08±0.78	4.36	4.21	2.15	4.40±0.05	4.86	4.43	4.31
$C_{16:0}$	15.87	18.25	16.71	15.38±2.61	14.88±0.02	17.09±0.73	14.58	14.51	17.84	16.34±1.82	26.27	16.97	15.11
$C_{16:0}$ 2-OH	1.45	2.35	4.08	4.33±0.53	3.31±1.51	1.71±1.71	3.45	3.68	2.48	7.31±1.43	4.20	2.26	2.46
$C_{16:0}$ 3-OH	5.28	5.64	6.94	5.63±0.21	6.49±0.16	10.29±2.71	5.83	5.67	4.89	6.27±1.50	4.68	5.16	6.27
$C_{16:1}$ 2-OH	ND	TR	1.22	1.91±0.05	1.27±0.39	TR	1.22	TR	ND	1.99±0.31	2.21	TR	1.52
$C_{17:0}$ cyclo	4.01	6.77	9.10	5.97±3.01	7.56±4.61	7.65±7.65	5.06	7.85	4.56	13.95±3.77	13.76	3.64	2.02
$C_{18:1}$ ω7c	41.10	31.35	27.67	32.48±4.27	32.00±3.63	32.23±4.96	34.89	35.05	41.75	23.16±4.03	20.42	34.53	35.09
$C_{19:0}$ cyclo ω8c	1.73	2.48	3.21	3.80±1.45	1.96±1.22	ND	1.64	4.40	TR	8.65±4.56	3.60	1.20	ND
Summed feature 2	6.16	8.14	8.81	7.45±0.86	8.73±0.59	10.12±2.55	7.13	6.84	6.12	6.45±1.41	5.83	7.90	8.18
Summed feature 3	19.32	19.45	15.97	18.32±0.18	18.44±5.56	15.70±5.44	20.31	15.52	18.83	8.82±4.11	12.69	23.12	25.05

Species: 1, *B. glathei* (5 strains); 2, *B. sordidicola* (1); 3, *B. zhejiangensis* (3); 4, *B. grimmiae* (1); 5, *B. choica* (2); 6, *B. humi* (6); 7, *B. telluris* (2); 8, *B. terrestris* (1); 9, *B. udeis* (3); 10, *B. cordobensis* (2); 11, *B. jiangsuensis* (1); 12, *B. megalochromosomata* (1); 13, *Burkholderia arvi* sp. nov. (1); 14, *Burkholderia hypogeia* sp. nov. (1); 15, *Burkholderia ptereochthonis* sp. nov. (1); 16, *Burkholderia glebae* sp. nov. (2); 17, *Burkholderia pedi* sp. nov. (2); 18, *Burkholderia arationis* sp. nov. (2); 19, *Burkholderia fortuita* sp. nov. (1); 20, *Burkholderia temeraria* sp. nov. (1); 21, *Burkholderia calidae* sp. nov. (1); 22, *Burkholderia concitans* sp. nov. (2); 23, *Burkholderia turbans* sp. nov. (1); 24, *Burkholderia catudaia* sp. nov. (1); 25, *Burkholderia peredens* sp. nov. (1). Data for *B. glathei*, *B. sordidicola*, *B. zhejiangensis*, *B. choica*, *B. humi*, *B. telluris*, *B. terrestris* and *B. udeis* were extracted from Vandamme *et al.* (2013a). Data for *B. cordobensis* and *B. grimmiae* were extracted from Draghi *et al.* (2014). All other data are from the present study. Values are mean±SD percentages of total fatty acids. Those fatty acids for which the mean amount for all taxa was <1 % are not included, therefore, the percentages may not add up to 100 %. TR, trace amount (<1 %); ND, not detected. Summed feature 2 comprises iso-$C_{16:1}$I and/or $C_{14:0}$ 3-OH; summed feature 3 comprises iso-$C_{15:0}$ 2-OH and/or $C_{16:1}$ ω7c.

Table 5.6: Differential biochemical characteristics of all examined strains of BGC species

Characteristic	1	2	3	4	6	5	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Growth at																									
15 °C	+	+	w	+	+	+	+	+	+	−	+	w	+	+	+	+ +	+ +	+ +	+	+	+	+ +	+	+	+
20 °C	+	+	ND	+	+	+	+	+	+	ND	ND	+	+	+	+	+ +	+ +	+ +	+	+	+	+ +	+	+	+
37 °C	+	−	+	+	w	−	w	−	−	+	+	−	w	w	+	− −	− −	− −	+	w	w	+ −	w	w	w
pH 5	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	− −	− −	− −	−	−	−	− −	−	−	−
pH 6	+	+	+	+	+	+	+	+	+	+	+	+	+	w	−	− −	+ +	− +	w	w	w	+ +	w	w	−
pH 7	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	w +	w +	− −	w	w	w	+ −	+	w	+
pH 8	−	w	+	+	−	+	−	−	+	w	−	+	−	−	−	− +	− +	− −	−	−	−	− −	−	−	−
Hydrolysis of																									
Tween 60	+	−	+	ND	+	−	+	NG	NG	+	+	+	+	+	+	+ +	+ +	+ +	+	−	−	+ +	+	+	+
Tween 80	−	−	+	+	−	−	−	−	−	+	+	+	−	−	−	− −	− −	− −	−	−	−	− −	−	−	−
API 20NE																									
Nitrate reduction	−	+	+	+	−	−	+	−	+	w	+	−	+	+	−	+ +	+ +	− −	−	−	+	− −	−	+	−
Urease	−	−	+	+	−	−	−	−	−	−	−	−	−	−	−	− −	− −	− −	−	−	−	− −	−	−	−
β-Galactosidase	−	+	−	−	−	−	−	−	w	−	−	+	w	−	−	− −	w +	− −	−	−	−	− −	−	−	−
Assimilation of																									
Arabinose	w	+	+	+	w	−	+	−	+	+	w	−	+	+	−	+ w	+ +	+ +	+	+	+	w +	+	+	w
Mannose	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+ +	+ +	+ +	+	+	+	+ +	+	+	+
Mannitol	w	+	−	+	w	+	+	+	−	+	+	+	+	+	+	+ +	+ +	+ +	+	+	+	+ +	+	+	+
N-Acetylglucosamine	+	+	+	+	w	+	+	+	+	+	+	+	+	+	+	+ +	+ +	+ +	+	+	+	+ +	+	+	+
Gluconate	+	+	+	+	w	+	+	+	+	+	+	+	+	+	+	+ +	+ +	+ +	+	+	+	+ +	+	+	+
Caprate	+	−	+	−	−	w	+	w	−	−	−	−	−	−	−	− −	− w	w w	−	−	+	− −	+	−	−
Malate	+	w	+	+	w	+	+	+	+	+	+	+	+	+	+	+ w	+ w	+ +	+	+	+	+ +	+	+	+
Citrate	+	−	−	−	−	+	+	+	+	−	w	−	w	−	−	+ w	− −	+ +	−	w	w	w −	−	−	−
Phenylacetate	+	−	+	−	−	+	+	+	−	+	+	+	+	+	+	+ +	+ +	w +	+	+	+	w w	+	+	+
Enzyme activity (API ZYM)																									
C4 lipase	+	+	−	+	w	+	w	+	+	−	+	−	−	+	+	w −	− +	+ +	−	+	−	+ +	w	−	w
C8 lipase	w	+	−	+	+	+	w	w	+	−	w	−	−	w	−	− −	− w	− w	−	−	w	w w	−	−	w
Valine arylamidase	w	−	−	−	w	−	−	w	−	−	−	−	−	−	−	− −	− +	− w	−	−	−	+ +	−	−	−
Cystine arylamidase	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	− −	− w	− −	−	−	−	− w	−	−	−
β-Galactosidase	−	+	−	−	−	−	−	−	+	−	−	−	−	−	−	− −	−w	− −	−	−	−	− −	−	−	−

Species: 1, *B. glathei* LMG 14190^T; 2, *B. sordidicola* LMG 22029^T; 3, *B. zhejiangensis* LMG 27258^T; 4, *B. grimmiae* R27^T; 5, *B. choica* LMG 22940^T; 6, *B. humi* LMG 22934^T; 7, *B. telluris* LMG 22936^T; 8, *B. terrestris* LMG 22937^T; 9, *B. udeis* LMG 27134^T; 10, *B. cordobensis* LMG 27620^T; 11, *B. jiangsuensis* LMG 27927^T; 12, *B. megalochromosomata* LMG 29263^T; 13, *Burkholderia arvi* sp. nov. LMG 29317^T; 14, *Burkholderia hypogeia* sp. nov. LMG 29322^T; 15, *Burkholderia ptereochthonis* sp. nov. LMG 29326^T; 16, *Burkholderia glebae* sp. nov. LMG 29325^T and LMG 22938; 17, *Burkholderia pedi* sp. nov. LMG 29323^T and R-52605; 18, *Burkholderia arationis* sp. nov. LMG 29324^T and R-23361; 19, *Burkholderia fortuita* sp. nov. LMG 29320^T; 20, *Burkholderia temeraria* sp. nov. LMG 29319^T; 21, *Burkholderia calidae* sp. nov. LMG 29321^T; 22, *Burkholderia concitans* sp. nov. LMG 29315^T and R-46586; 23, *Burkholderia turbans* sp. nov. LMG 29316^T; 24, *Burkholderia catudaia* sp. nov. LMG 29318^T; 25, *Burkholderia peredens* sp. nov. LMG 29314^T. Data for *B. glathei*, *B. sordidicola*, *B. choica*, *B. humi*, *B. telluris*, *B. terrestris* and *B. udeis* were extracted from Vandamme *et al.* (2013a). Data for *B. grimmiae* were extracted from Tian *et al.* (2013). Data for *B. cordobensis* and *B. zhejiangensis* were extracted from Draghi *et al.* (2014). All other data are from the present study. Test results of the type strains are given first, followed by the remaining strains in the order given above. +, present; −, absent; w, weak reaction; v, variable; ND, not determined; NG, no growth.

5.2.5 Discussion

While soil is a well-known source of free-living *Burkholderia* species, these organisms often live in close interaction with plants, animals, fungi or amoebae (Marolda *et al.*, 1999; Van Borm *et al.*, 2002; Kikuchi *et al.*, 2011; Verstraete *et al.*, 2013; Stopnisek *et al.*, 2016; Xu *et al.*, 2016). The BGC represents a poorly known line of descent within the genus *Burkholderia* and most of the 12 validly named BGC species have been isolated from soil. Yet, publicly available sequence data indicate that the taxonomic diversity in this clade is severely underestimated (Nogales *et al.*, 2001; Salles *et al.*, 2006b; Pumphrey & Madsen, 2008; Draghi *et al.*, 2014; Verstraete *et al.*, 2014; Peeters *et al.*, 2016). In the present study, *gyrB* gene sequence analysis was used to screen our strain collection and 17 isolates from human and environmental samples were identified as *B. glathei*-like bacteria. The *gyrB* sequence similarity levels towards other BGC species suggested that the 17 isolates in this study represented 13 novel species (Fig. 5.3). To further characterize the taxonomic status of these isolates, we analyzed the genome sequence of 13 isolates representative for the 13 *gyrB* sequence clusters and of 7 BGC type strains and compared those to 23 whole-genome sequences of BGC strains that were publicly available. Additionally, we also studied their chemotaxonomic and biochemical properties to comply with the polyphasic taxonomic consensus approach to bacterial systematics (Vandamme *et al.*, 1996).

In this genomics era, state-of-the-art sequencing technologies enable direct access to the information contained in whole-genome sequences and it is no longer adequate to deduce genome relatedness through traditional DDH experiments (Vandamme & Peeters, 2014; Whitman, 2015). Genomic taxonomy can be studied through various parameters including ANI, GBDP, MUMi and core gene identity (Konstantinidis & Tiedje, 2005a; Goris *et al.*, 2007; Deloger *et al.*, 2009; Vanlaere *et al.*, 2009; Meier-Kolthoff *et al.*, 2013). Although there is a general consensus that genome sequencing could revolutionize prokaryotic systematics (Sutcliffe *et al.*, 2013; Meier-Kolthoff *et al.*, 2014c; Rossello-Mora & Amann, 2015; Thompson *et al.*, 2015), traditional DDH experiments are still being performed and new genome-based methods are evaluated in terms of their correspondence to the existing classifications which are based on DDH data (Wayne *et al.*, 1987; Stackebrandt *et al.*, 2002). The GGDC implementation of the GBDP method provides a quick and reliable alternative to the wet-lab DDH technique and its dDDH prediction capability (including confidence intervals) produces classifications which correlate better with the traditional DDH values than do any of the ANI implementations (Meier-Kolthoff *et al.*, 2013). Among several advantages, GBDP is independent from genome annotation, is applicable to both nucleotide and amino acid data and is immune against problems caused by incompletely sequenced or low-quality draft

genomes. Finally, GBDP provides branch support values for the resulting phylogenetic trees (Meier-Kolthoff *et al.*, 2013, 2014a).

We complemented the results of the GBDP analysis with a whole-genome sequence-based phylogeny based on the sequence analysis of 332 single-copy orthologous genes in all BGC genomes. This extended MLSA approach takes only the coding part of the genomes into account and is therefore not influenced by non-coding sequences or pseudogenes that might have a different evolutionary history than the rest of the genome. It depends however on genome annotation, is unable to cope with problems caused by incompletely sequenced or low-quality draft genomes, and its calculations are more compute-intensive and cannot be carried out incrementally. Although the GBDP and extended MLSA methods used different algorithms, the conclusions drawn from their phylogenies were consistent thus illustrating the robustness of whole-genome sequence-based taxonomic methods (Colston *et al.*, 2014).

The GGDC dDDH values and the application of the 70 % dDDH cut-off for species delineation (Supplementary Table 5.1) demonstrated that the 13 clusters delineated through *gyrB* sequence analysis (Fig. 5.3) represented 13 novel BGC species and thus confirmed that *gyrB* gene sequence analysis is a reliable tool for the identification of *Burkholderia* species (Tayeb *et al.*, 2008; Vandamme *et al.*, 2013a). Both phylogenomic analyses identified strain MR1, which was isolated from Florida golf course soil and which was shown to reduce the herbicide methylarsenate, as *B. concitans* sp. nov. Next to strain YI23, which was previously identified as *B. cordobensis* by Draghi *et al.* (2014), the present study also identified strain RPE67, which was isolated from the gut of a stink bug, as *B. cordobensis*. Finally, both phylogenomic analyses also showed that strain PML1(12), an ectomycorrhizosphere-inhabiting bacterium with mineral-weathering ability (Uroz & Oger, 2015), strain S170, a potential plant growth promoter isolated from coniferous forest soil (Llado *et al.*, 2014), strain RPE64, a bacterial symbiont of the bean bug *Riptortus pedestris* (Shibata *et al.*, 2013) and strain Leaf177, an *Arabidopsis* leaf isolate (Bai *et al.*, 2015) all represent novel BGC species.

Burkholderia genomes vary in size from 3.75 Mb (*B. rhizoxinica* HKI 454) to 11.3 Mb (*B. terrae* BS001), are characterized by a high G+C content (60-68 %) and consist of multiple replicons (Winsor *et al.*, 2008; Ussery *et al.*, 2009). The DNA G+C content of the 13 novel species was calculated from their genome sequences and was in the range of that reported for other BGC species (60-65 mol%). For 10 of the 12 established BGC species, the G+C content was previously calculated by traditional wet-lab methods and the reported values differed by 0.1 to 3.3 mol% from the values calculated from their genome sequences (Table 5.7). As reported by Meier-Kolthoff *et al.* (2014c) the G+C content calculations based on genome sequences show a higher precision than calculations based on traditional wet-lab methods because the latter methods do not count nucleotides but estimate the genomic

G+C content based on the physical properties of the extracted and/or digested genomic DNA. The difference between literature data (Lim *et al.*, 2003; Lu *et al.*, 2012; Tian *et al.*, 2013) and the genome sequence-based G+C content values of *B. sordidicola* LMG 22029^T, *B. zhejiangensis* OP-1^T and *B. grimmiae* R27^T is larger than 1 % and we therefore present emended descriptions of these species. The genome sizes of the type strains of the 13 novel species ranged from 6.2 Mb (*B. concitans* sp. nov. LMG 29315^T) to 9.7 Mb (*B. arvi* sp. nov. LMG 29317^T) and corresponded with the genome sizes of other free-living BGC species (Table 5.4). Consistent with reductive genome evolution in obligatory symbionts, the smallest BGC genomes belong to the obligatory leaf endosymbionts (2.4 to 6.2 Mb) (Carlier & Eberl, 2012; Carlier *et al.*, 2016; Pinto-Carbo *et al.*, 2016).

Table 5.7: G+C content (mol%) of validly named BGC species

Strain	Wet-lab calculation	Calculation from WGS
B. glathei LMG 14190^T	64.8 (Zolg & Ottow, 1975)	64.7
B. sordidicola LMG 22029^T	61.3 (Lim *et al.*, 2003)	60.2
B. zhejiangensis OP-1^T	59.4 (Lu *et al.*, 2012)	62.7
B. grimmiae R27^T	64.6 (Tian *et al.*, 2013)	63.0
B. choica LMG 22940^T	63 (Vandamme *et al.*, 2013a)	62.7
B. humi LMG 22934^T	63 (Vandamme *et al.*, 2013a)	62.8
B. telluris LMG 22936^T	64 (Vandamme *et al.*, 2013a)	64.0
B. terrestris LMG 22937^T	62 (Vandamme *et al.*, 2013a)	62.6
B. udeis LMG 27134^T	60 (Vandamme *et al.*, 2013a)	60.0
B. cordobensis LMG 27620^T	63.6 (Draghi *et al.*, 2014)	63.7
B. jiangsuensis MP-1^T	-	62.6
B. megalochromosomata JC2949^T	-	62.7

WGS, whole-genome sequence

Biochemically, these novel species are similar to their nearest neighbors. However, tests particularly useful for distinguishing BGC species are growth at 37 °C and at pH 8, hydrolysis of tween 60 and 80, nitrate reduction, assimilation of arabinose, caprate and citrate, beta-galactosidase activity and C4 lipase (Table 5.6). The most discriminating fatty acids are $C_{16:0}$ 3-OH, $C_{17:0}$ cyclo, $C_{19:0}$ cyclo ω8c and summed features 2 and 3 (Table 5.5). The overall fatty acid profiles of the novel taxa are similar to those of their nearest neighbors and support their placement in the genus *Burkholderia* (Yabuuchi *et al.*, 1992).

The present study again underscores the multifaceted nature of *Burkholderia* bacteria (Coenye & Vandamme, 2003; Mahenthiralingam *et al.*, 2005) and highlights that also BGC species have evolved a broad range of interactions with different hosts. *B. cordobensis* is a striking example of phenotypic and geographic breadth: it was recovered from agricultural soil in Argentina (strain LMG 27620^T) (Draghi *et al.*, 2014), from golf course soil in South Korea (strain YI23) (Lim *et al.*, 2012) and from the gut of the bean bug *Riptortus pedestris* in

Japan (strain RPE67) (Takeshita *et al.*, 2014). The two latter strains (YI23 and RPE67) have fenitrothion degrading properties. The former two strains (LMG 27620^T and YI23) were free-living but the latter (RPE67) is an endosymbiont of stink bugs that is not vertically transmitted but acquired from soil by the nymphal insect (Kikuchi *et al.*, 2007). The insecticide resistance to fenitrothion in the pest insects was shown to be established by the endosymbiotic *Burkholderia* strain in the insect gut (Kikuchi *et al.*, 2012) and was shown to emerge as a consequence of repeated insecticide use (Tago *et al.*, 2015). The *Riptortus pedestris-B. cordobensis* association thus appears to be a rather young endosymbiosis and contrasts with the symbiosis observed between plant species of the *Rubiaceae* and *Primulaceae* families and several *Candidatus Burkholderia* species. The latter *Candidatus* designation is a provisional taxonomic status for organisms that have been characterized but that cannot be cultivated at present (Schleifer, 2009). These obligate leaf endosymbionts are vertically transmitted and represent an obligatory symbiosis which was estimated to originate millions of years ago (Lemaire *et al.*, 2011b).

BGC species harbor both beneficial and pathogenic strains. Strains PML1(12) and S170 show biotechnological potential for mineral-weathering and plant growth promotion, respectively, and are exemplary for the metabolic versatility of *Burkholderia* organisms (Llado *et al.*, 2014; Uroz & Oger, 2015). Mineral-weathering bacteria dissolute key nutrients from minerals and thereby increase the bioavailability of chemical nutrients in the environment (Uroz *et al.*, 2009). On the other hand, three strains analyzed in the present study were isolated from human clinical samples, i.e. blood, pleural fluid and lung tissue (Table 5.3) and were classified as two novel species (*Burkholderia concitans* sp. nov. and *Burkholderia turbans* sp. nov.). They represent, to our knowledge, the first examples of human clinical isolates in the *B. glathei* clade. Strikingly, strain MR1, which was isolated from Florida golf course soil and shown to reduce the herbicide methylarsenate, was also identified as *Burkholderia concitans* sp. nov., and this species thus represents yet another human clinical *Burkholderia* species with interesting biotechnological properties (Coenye *et al.*, 2001b; Coenye & Vandamme, 2003; Goris *et al.*, 2004; Mahenthiralingam *et al.*, 2005). This study therefore further underscores that there is no phylogenetic subdivision in the genus *Burkholderia* that distinguishes beneficial from pathogenic strains (Angus *et al.*, 2014; Sawana *et al.*, 2014; Estrada-de los Santos *et al.*, 2015; Dobritsa & Samadpour, 2016).

In summary, the present study provides genotypic, chemotaxonomic and phenotypic data which enable the differentiation of 13 novel species in the genus *Burkholderia* and we propose the names *Burkholderia arvi* sp. nov., *Burkholderia hypogeia* sp. nov., *Burkholderia ptereochthonis* sp. nov., *Burkholderia glebae* sp. nov., *Burkholderia pedi* sp. nov., *Burkholderia arationis* sp. nov., *Burkholderia fortuita* sp. nov., *Burkholderia temeraria* sp. nov., *Burkholderia calidae*

sp. nov., *Burkholderia concitans* sp. nov., *Burkholderia turbans* sp. nov., *Burkholderia catudaia* sp. nov. and *Burkholderia peredens* sp. nov., with strains LMG 29317^T, LMG 29322^T, LMG 29326^T, LMG 29325^T, LMG 29323^T, LMG 29324^T, LMG 29320^T, LMG 29319^T, LMG 29321^T, LMG 29315^T, LMG 29316^T, LMG 29318^T and LMG 29314^T as type strains, respectively. By making reference cultures and whole-genome sequences of each of these versatile bacteria publicly available, we aim to contribute to future knowledge about the metabolic versatility and pathogenicity of *Burkholderia* organisms.

5.2.5.1 Description of *Burkholderia arvi* sp. nov.

Burkholderia arvi (ar'vi. L. gen. n. *arvi* of a field).
Cells are Gram-negative, non-motile rods (less than 1 µm wide and about 1 µm long) with rounded ends that occur as single units or in pairs. After 48 h of incubation on trypticase soy agar at 28 °C, colonies are round (typically less than 1 mm in diameter), smooth, shiny, non-translucent, with entire margins and a white-creamy color. Grows on MacConkey agar. Growth occurs at 15-37 °C and at pH 6-7 in NB at 28 °C. Catalase and oxidase activities are present. Hydrolyzes tween 60, but not tween 80, starch and casein. When tested using API 20NE strips, positive for nitrate reduction, beta-galactosidase (PNPG) (weak) and assimilation of glucose, arabinose, mannose, mannitol, N-acetyl-glucosamine, gluconate, malate, citrate (weak) and phenylacetate; negative for production of indol, fermentation of glucose, arginine dihydrolase, urease, esculin hydrolysis, gelatin liquefaction and assimilation of maltose, caprate and adipate. When tested using API ZYM strips, positive for alkaline phosphatase, leucyl arylamidase, acid phosphatase and phosphoamidase (weak); negative for C4 lipase, C8 lipase, C14 lipase, valine arylamidase, cystine arylamidase, trypsin, chymotrypsin, alpha-galactosidase, beta-galactosidase, beta-glucuronidase, alpha-glucosidase, beta-glucosidase, N-acetyl-beta-glucosaminidase, alpha-mannosidase and alpha-fucosidase. The following fatty acids are present: $C_{16:0}$, $C_{16:0}$ 3-OH, $C_{18:1}$ ω7c, summed feature 2 (most likely $C_{14:0}$ 3-OH) and summed feature 3 (most likely $C_{16:1}$ ω7c) in moderate amounts (>5 %), and $C_{14:0}$, $C_{16:0}$ 2-OH, $C_{17:0}$ cyclo and $C_{19:0}$ cyclo ω8c in minor amounts (1-5 %).
The type strain is LMG 29317^T (=CCUG 68412^T) and was isolated from agricultural soil in Argentina in 2010 (Draghi *et al.*, 2014). Its G+C content is 62.4 mol% (calculated based on its genome sequence). The 16S rRNA, *gyrB* and whole-genome sequence of LMG 29317^T are publicly available through the accession numbers LT158615, LT158628 and FCOM01000000, respectively.

5.2.5.2 Description of *Burkholderia hypogeia* sp. nov.

Burkholderia hypogeia (hy.po.ge'ia. Gr. adj. *hypogeios* subterraneous; N. L. fem. adj. *hypogeia*, subterraneous, earth-born).
Cells are Gram-negative, non-motile rods (about 1 µm wide and 1-2 µm long) with rounded ends that occur as single units or in pairs. After 48 h of incubation on trypticase soy agar at 28 °C, colonies are round (typically less than 1 mm in diameter), smooth, shiny, non-translucent, with entire margins and a white-creamy color. Grows on MacConkey agar. Growth occurs at 15-37 °C and at pH 6 in NB at 28 °C. Catalase and oxidase activities are present. Hydrolyzes tween 60, but not tween 80, starch and casein. When tested using API 20NE strips, positive for nitrate reduction and assimilation of glucose, arabinose, mannose, mannitol, N-acetyl-glucosamine, gluconate, malate and phenylacetate; negative for production of indol, fermentation of glucose, arginine dihydrolase, urease, esculin hydrolysis, gelatin liquefaction, beta-galactosidase (PNPG) and assimilation of maltose, caprate, adipate and citrate. When tested using API ZYM strips, positive for alkaline phosphatase (weak), C4 lipase, C8 lipase (weak), leucyl arylamidase, acid phosphatase and phosphoamidase (weak); negative for C14 lipase, valine arylamidase, cystine arylamidase, trypsin, chymotrypsin, alpha-galactosidase, beta-galactosidase, beta-glucuronidase, alpha-glucosidase, beta-glucosidase, N-acetyl-beta-glucosaminidase, alpha-mannosidase and alpha-fucosidase. The following fatty acids are present: $C_{16:0}$, $C_{16:0}$ 3-OH, $C_{17:0}$ cyclo, $C_{18:1}$ ω7c, summed feature 2 (most likely $C_{14:0}$ 3-OH) and summed feature 3 (most likely $C_{16:1}$ ω7c) in moderate amounts (>5 %), and $C_{14:0}$, C_2-OH and $C_{19:0}$ cyclo ω8c in minor amounts (1-5 %).
The type strain is LMG 29322^T (=CCUG 68407^T) and was isolated from greenhouse soil in Belgium in 2014 (Peeters *et al.*, 2016). Its G+C content is 63.2 mol% (calculated based on its genome sequence). The 16S rRNA, *gyrB* and whole-genome sequence of LMG 29322^T are publicly available through the accession numbers LT158620, LT158633 and FCOA01000000, respectively.

5.2.5.3 Description of *Burkholderia ptereochthonis* sp. nov.

Burkholderia ptereochthonis (pte.re.o.chtho'nis Gr. n. *pteris* fern; Gr. n. *chthon* soil; N. L. gen. n. *ptereochthonis*, from fern soil).
Cells are Gram-negative, non-motile rods (less than 1 µm wide and about 1 µm long) with rounded ends that occur as single units or in pairs. After 48 h of incubation on trypticase soy agar at 28 °C, colonies are round (typically less than 1 mm in diameter), smooth, shiny, non-translucent, with entire margins and a white-creamy color. Grows on MacConkey agar. Growth occurs at 15-37 °C and at pH 7 in NB at 28 °C. Catalase and oxidase activities are

present. Hydrolyzes tween 60, but not tween 80, starch and casein. When tested using API 20NE strips, positive for the assimilation of glucose, mannose, mannitol, N-acetyl-glucosamine, gluconate, malate and phenylacetate; negative for nitrate reduction, production of indol, fermentation of glucose, arginine dihydrolase, urease, esculin hydrolysis, gelatin liquefaction, beta-galactosidase (PNPG) and assimilation of arabinose, maltose, caprate, adipate and citrate. When tested using API ZYM strips, positive for alkaline phosphatase, C4 lipase, leucyl arylamidase, acid phosphatase and phosphoamidase (weak); negative for C8 lipase, C14 lipase, valine arylamidase, cystine arylamidase, trypsin, chymotrypsin, alpha-galactosidase, beta-galactosidase, beta-glucuronidase, alpha-glucosidase, beta-glucosidase, N-acetyl-beta-glucosaminidase, alpha-mannosidase and alpha-fucosidase. The following fatty acids are present: $C_{16:0}$, $C_{16:0}$ 3-OH, $C_{17:0}$ cyclo, $C_{18:1}$ ω7c, summed feature 2 (most likely $C_{14:0}$ 3-OH) and summed feature 3 (most likely $C_{16:1}$ ω7c) in moderate amounts (>5 %), and $C_{14:0}$, $C_{16:0}$ 2-OH, $C_{16:1}$ 2-OH and $C_{19:0}$ cyclo ω8c in minor amounts (1-5 %).
The type strain is LMG 29326^T (=CCUG 68403^T) and was isolated from botanical garden soil in Belgium in 2014 (Peeters *et al.*, 2016). Its G+C content is 64.2 mol% (calculated based on its genome sequence). The 16S rRNA, *gyrB* and whole-genome sequence of LMG 29326^T are publicly available through the accession numbers LT158624, LT158637 and FCOB01000000, respectively.

5.2.5.4 Description of *Burkholderia glebae* sp. nov.

Burkholderia glebae (gle'bae. L. gen. n. *glebae* from a lump or clod of earth, soil).
Cells are Gram-negative, non-motile rods (less than 1 µm wide and about 1 µm long) with rounded ends that occur as single units or in pairs. After 48 h of incubation on trypticase soy agar at 28 °C, colonies are round, tiny (typically less than 0.5 mm in diameter), non-translucent, with a white-creamy color. Grows on MacConkey agar. Growth occurs at 15-28 °C and at pH 7-8 in NB at 28 °C (for the type strain only at pH 7). Catalase and oxidase activities are present. Hydrolyzes tween 60, but not tween 80, starch and casein. When tested using API 20NE strips, positive for nitrate reduction and assimilation of glucose, arabinose, mannose, mannitol, N-acetyl-glucosamine, gluconate, malate, citrate and phenylacetate; negative for production of indol, fermentation of glucose, arginine dihydrolase, urease, esculin hydrolysis, gelatin liquefaction, beta-galactosidase (PNPG) and assimilation of maltose, caprate and adipate. When tested using API ZYM strips, positive for leucyl arylamidase, acid phosphatase and phosphoamidase; negative for C8 lipase, C14 lipase, valine arylamidase, cystine arylamidase, trypsin, chymotrypsin, alpha-galactosidase, beta-galactosidase, beta-glucuronidase, alpha-glucosidase, beta-glucosidase, N-acetyl-beta-glucosaminidase, alpha-

mannosidase and alpha-fucosidase; strain-dependent reactions for alkaline phosphatase (type strain negative) and C4 lipase (type strain weak). The following fatty acids are present in all isolates: $C_{16:0}$, $C_{16:0}$ 3-OH, $C_{17:0}$ cyclo, $C_{18:1}$ ω7c, summed feature 2 (most likely $C_{14:0}$ 3-OH) and summed feature 3 (most likely $C_{16:1}$ ω7c) in moderate amounts (>5 %), and $C_{14:0}$, $C_{16:0}$ 2-OH, $C_{16:1}$ 2-OH and $C_{19:0}$ cyclo ω8c in minor amounts (1-5 %) (mean value of all isolates).
The type strain is LMG 29325^T (=CCUG 68404^T) and was isolated from botanical garden soil in Belgium in 2014 (Peeters *et al.*, 2016). Its G+C content is 62.7 mol% (calculated based on its genome sequence). The 16S rRNA, *gyrB* and whole-genome sequence of LMG 29325^T are publicly available through the accession numbers LT158623, LT158636 and FCOJ01000000, respectively. An additional strain has been isolated from soil in the Netherlands (Table 5.3).

5.2.5.5 Description of *Burkholderia pedi* sp. nov.

Burkholderia pedi (pe'di. Gr. n. *pedon* soil, earth; N. L. gen. n. *pedi*, from soil).
Cells are Gram-negative, non-motile rods (less than 1 µm wide and 1-2 µm long) with rounded ends that occur as single units or in pairs. After 48 h of incubation on trypticase soy agar at 28 °C, colonies are round (typically less than 1 mm in diameter), smooth, shiny, non-translucent, with entire margins and a beige color. Grows on MacConkey agar. Growth occurs at 15-28 °C and at pH 6-8 in NB at 28 °C (type strain only in pH 6-7). Catalase and oxidase activities are present. Hydrolyzes tween 60, but not tween 80, starch and casein. When tested using API 20NE strips, positive for nitrate reduction, beta-galactosidase (PNPG) and assimilation of glucose, arabinose, mannose, mannitol, N-acetyl-glucosamine, gluconate, adipate, malate and phenylacetate; negative for production of indol, fermentation of glucose, urease, esculin hydrolysis, gelatin liquefaction and assimilation of maltose and citrate; strain-dependent reactions for arginine dihydrolase (type strain negative) and the assimilation of caprate (type strain negative). When tested using API ZYM strips, positive for alkaline phosphatase, leucyl arylamidase, acid phosphatase and phosphoamidase; negative for C14 lipase, trypsin, chymotrypsin, alpha-galactosidase, beta-glucuronidase, alpha-glucosidase, beta-glucosidase, alpha-mannosidase and alpha-fucosidase; strain-dependent reactions for C4 lipase (type strain negative), C8 lipase (type strain negative), valine arylamidase (type strain negative), cystine arylamidase (type strain negative), beta-galactosidase (type strain negative) and N-acetyl-beta-glucosaminidase (type strain negative). The following fatty acids are present in all isolates: $C_{16:0}$, $C_{16:0}$ 3-OH, $C_{17:0}$ cyclo, $C_{18:1}$ ω7c, summed feature 2 (most likely $C_{14:0}$ 3-OH) and summed feature 3 (most likely $C_{16:1}$ ω7c) in moderate amounts

(>5 %), and $C_{14:0}$, $C_{16:0}$ 2-OH, $C_{16:1}$ 2-OH and $C_{19:0}$ cyclo ω8c in minor amounts (1-5 %) (mean value of all isolates).
The type strain is LMG 29323^T (=CCUG 68406^T) and was isolated from greenhouse soil in Belgium in 2014 (Peeters *et al.*, 2016). Its G+C content is 63.0 mol% (calculated based on its genome sequence). The 16S rRNA, *gyrB* and whole-genome sequence of LMG 29323^T are publicly available through the accession numbers LT158621, LT158634 and FCOE01000000, respectively. An additional strain has been isolated from the same sample (Table 5.3).

5.2.5.6 Description of *Burkholderia arationis* sp. nov.

Burkholderia arationis (a.ra.ti.o'nis. L. gen. n. *arationis* from a field).
Cells are Gram-negative, non-motile rods (less than 1 µm wide and about 1 µm long) with rounded ends that occur as single units or in pairs. After 48 h of incubation on trypticase soy agar at 28 °C, colonies are round (typically less than 1 mm in diameter), smooth, shiny, translucent, with entire margins and a white-creamy color. Grows on MacConkey agar. Growth occurs at 15-28 °C and at pH 6 in NB at 28 °C (the type strain did not grow in liquid NB medium). Catalase and oxidase activities are present. Hydrolyzes tween 60, but not tween 80, starch and casein. When tested using API 20NE strips, positive for assimilation of glucose, arabinose, mannose, mannitol, N-acetyl-glucosamine, gluconate, caprate (weak), adipate (weak), malate, citrate and phenylacetate; negative for nitrate reduction, production of indol, fermentation of glucose, arginine dihydrolase, urease, esculin hydrolysis, gelatin liquefaction, beta-galactosidase (PNPG) and assimilation of maltose. When tested using API ZYM strips, positive for alkaline phosphatase, C4 lipase, leucyl arylamidase, acid phosphatase and phosphoamidase; negative for C14 lipase, cystine arylamidase, trypsin, alpha-galactosidase, beta-galactosidase, beta-glucuronidase, alpha-glucosidase, beta-glucosidase, N-acetyl-beta-glucosaminidase, alpha-mannosidase and alpha-fucosidase; strain-dependent reactions for C8 lipase (type strain negative), valine arylamidase (type strain negative) and chymotrypsin (type strain negative). The following fatty acids are present in all isolates: $C_{16:0}$, $C_{16:0}$ 3-OH, $C_{18:1}$ ω7c, summed feature 2 (most likely $C_{14:0}$ 3-OH) and summed feature 3 (most likely $C_{16:1}$ ω7c) in moderate amounts (>5 %), and $C_{14:0}$ in minor amounts (1-5 %) (mean value of all isolates).
The type strain is LMG 29324^T (=CCUG 68405^T) and was isolated from botanical garden soil in Belgium in 2014 (Peeters *et al.*, 2016). Its G+C content is 62.8 mol% (calculated based on its genome sequence). The 16S rRNA, *gyrB* and whole-genome sequence of LMG 29324^T are publicly available through the accession numbers LT158622, LT158635 and FCOG01000000, respectively. An additional strain has been isolated from soil in the Netherlands (Table

5.3).

5.2.5.7 Description of *Burkholderia fortuita* sp. nov.

Burkholderia fortuita (for.tu.i'ta. L. fem. adj. *fortuita* accidental, unpremeditated; referring to its fortuitous isolation when searching for *Burkholderia caledonica* endophytes).
Cells are Gram-negative, non-motile rods (less than 1 µm wide and about 1 µm long) with rounded ends that occur as single units or in pairs. After 48 h of incubation on trypticase soy agar at 28 °C, colonies are round (typically less than 1 mm in diameter), smooth, shiny, non-translucent, with entire margins and a beige color. Grows on MacConkey agar. Growth occurs at 15-37 °C and at pH 6-7 in NB at 28 °C. Catalase and oxidase activities are present. Hydrolyzes tween 60, but not tween 80, starch and casein. When tested using API 20NE strips, positive for the assimilation of glucose, arabinose, mannose, mannitol, N-acetyl-glucosamine, gluconate, malate and phenylacetate; negative for nitrate reduction, production of indol, fermentation of glucose, arginine dihydrolase, urease, esculin hydrolysis, gelatin liquefaction, beta-galactosidase (PNPG) and assimilation of maltose, caprate, adipate and citrate. When tested using API ZYM strips, positive for alkaline phosphatase (weak), leucyl arylamidase, acid phosphatase and phosphoamidase (weak); negative for C4 lipase, C8 lipase, C14 lipase, valine arylamidase, cystine arylamidase, trypsin, chymotrypsin, alpha-galactosidase, beta-galactosidase, beta-glucuronidase, alpha-glucosidase, beta-glucosidase, N-acetyl-beta-glucosaminidase, alpha-mannosidase and alpha-fucosidase. The following fatty acids are present: $C_{16:0}$, $C_{16:0}$ 3-OH, $C_{17:0}$ cyclo, $C_{18:1}$ ω7c, summed feature 2 (most likely $C_{14:0}$ 3-OH) and summed feature 3 (most likely $C_{16:1}$ ω7c) in moderate amounts (>5 %), and $C_{14:0}$, $C1_{6:0}$ 2-OH, $C_{16:1}$ 2-OH and $C_{19:0}$ cyclo ω8c in minor amounts (1-5 %).
The type strain is LMG 29320^T (=CCUG 68409^T) and was isolated from *Fadogia homblei* rhizosphere soil in South Africa in 2013 (Verstraete *et al.*, 2014). Its G+C content is 62.9 mol% (calculated based on its genome sequence). The 16S rRNA, *gyrB* and whole-genome sequence of LMG 29320^T are publicly available through the accession numbers LT158618, LT158631 and FCNX01000000, respectively.

5.2.5.8 Description of *Burkholderia temeraria* sp. nov.

Burkholderia temeraria (te.me.ra'ri.a. L. fem. adj. *temeraria* accidental, inconsiderate; referring to its accidental isolation when searching for *Burkholderia caledonica* endophytes).
Cells are Gram-negative, non-motile rods (less than 1 µm wide and about 1 µm long) with rounded ends that occur as single units or in pairs. After 48 h of incubation on trypticase soy agar at 28 °C, colonies are round (typically less than 1 mm in diameter), smooth, shiny,

non-translucent, with entire margins and a white-creamy color. Grows on MacConkey agar. Growth occurs at 15-37 °C and at pH 6-7 in NB at 28 °C. Catalase and oxidase activities are present. Does not hydrolyze tween 60, tween 80, starch and casein. When tested using API 20NE strips, positive for the assimilation of glucose, arabinose, mannose, mannitol, N-acetyl-glucosamine, gluconate, malate, citrate (weak) and phenylacetate; negative for nitrate reduction, production of indol, fermentation of glucose, arginine dihydrolase, urease, esculin hydrolysis, gelatin liquefaction, beta-galactosidase (PNPG) and assimilation of maltose, caprate and adipate. When tested using API ZYM strips, positive for alkaline phosphatase, C4 lipase, leucyl arylamidase, acid phosphatase and phosphoamidase (weak); negative for C8 lipase, C14 lipase, valine arylamidase, cystine arylamidase, trypsin, chymotrypsin, alpha-galactosidase, beta-galactosidase, beta-glucuronidase, alpha-glucosidase, beta-glucosidase, N-acetyl-beta-glucosaminidase, alpha-mannosidase and alpha-fucosidase. The following fatty acids are present: $C_{16:0}$, $C_{16:0}$ 3-OH, $C_{17:}0$ cyclo, $C_{18:1}$ ω7c, summed feature 2 (most likely $C_{14:0}$ 3-OH) and summed feature 3 (most likely $C_{16:1}$ ω7c) in moderate amounts (>5 %), and $C_{14:0}$, $C_{16:0}$ 2-OH and $C_{19:0}$ cyclo ω8c in minor amounts (1-55 %).
The type strain is LMG 29319^T (=CCUG 68410^T) and was isolated from *Fadogia homblei* rhizosphere soil in South Africa in 2013 (Verstraete *et al.*, 2014). Its G+C content is 62.7 mol% (calculated based on its genome sequence). The 16S rRNA, *gyrB* and whole-genome sequence of LMG 29319^T are publicly available through the accession numbers LT158617, LT158630 and FCOI01000000, respectively.

5.2.5.9 Description of *Burkholderia calidae* sp. nov.

Burkholderia calidae (ca'li.dae. L. gen. n. *calidae* from warm water, because this strain was isolated from pond water in a tropical garden).
Cells are Gram-negative, non-motile rods (about 1 µm wide and 1 µm long) with rounded ends that occur as single units or in pairs. After 48 h of incubation on trypticase soy agar at 28 °C, colonies are round (typically about 1 mm in diameter), smooth, shiny, non-translucent, with entire margins and a white-creamy color. Grows on MacConkey agar. Growth occurs at 15-37 °C and at pH 6-7 in NB at 28 °C. Catalase and oxidase activities are present. Does not hydrolyze tween 60, tween 80, starch and casein. When tested using API 20NE strips, positive for nitrate reduction and assimilation of glucose, arabinose, mannose, mannitol, N-acetyl-glucosamine, gluconate, caprate, malate, citrate (weak) and phenylacetate; negative for production of indol, fermentation of glucose, arginine dihydrolase, urease, esculin hydrolysis, gelatin liquefaction, beta-galactosidase (PNPG) and assimilation of maltose and adipate. When tested using API ZYM strips, positive for alkaline phosphatase (weak), C8 lipase

(weak), leucyl arylamidase (weak), acid phosphatase and phosphoamidase (weak); negative for C4 lipase, C14 lipase, valine arylamidase, cystine arylamidase, trypsin, chymotrypsin, alpha-galactosidase, beta-galactosidase, beta-glucuronidase, alpha-glucosidase, beta-glucosidase, N-acetyl-beta-glucosaminidase, alpha-mannosidase and alpha-fucosidase. The following fatty acids are present: $C_{16:0}$, $C_{18:1}$ ω7c, summed feature 2 (most likely $C_{14:0}$ 3-OH) and summed feature 3 (most likely $C_{16:1}$ ω7c) in moderate amounts (>5 %), and $C_{14:0}$, $C_{16:0}$ 2-OH, $C_{16:0}$ 3-OH and $C_{17:0}$ cyclo in minor amounts (1-5 %).
The type strain is LMG 29321^T (=CCUG 68408^T) and was isolated from greenhouse pond water in Belgium in 2013 (Peeters *et al.*, 2016). Its G+C content is 62.5 mol% (calculated based on its genome sequence). The 16S rRNA, *gyrB* and whole-genome sequence of LMG 29321^T are publicly available through the accession numbers LT158619, LT158632 and FCOX01000000, respectively.

5.2.5.10 Description of *Burkholderia concitans* sp. nov.

Burkholderia concitans (con.ci'tans. L. fem. part. pres. *concitans* disturbing, upsetting; because the isolation of this bacterium from human sources, including blood, further disturbs the image of this lineage of *Burkholderia* species as benign bacteria).
Cells are Gram-negative, non-motile rods (less than 1 µm wide and about 1 µm long) with rounded ends that occur as single units or in pairs. After 48 h of incubation on trypticase soy agar at 28 °C, colonies are round (typically less than 1 mm in diameter), smooth, shiny, non-translucent, with entire margins and a white-creamy color. Grows on MacConkey agar. Growth occurs at 15-28 °C (additionally, the type strains grows at 37 °C) and at pH 6-7 in NB at 28 °C. Catalase and oxidase activities are present. Hydrolyzes tween 60, but not tween 80, starch and casein. When tested using API 20NE strips, positive for the assimilation of glucose, arabinose, mannose, mannitol, N-acetyl-glucosamine, gluconate, malate and phenylacetate; negative for nitrate reduction, production of indol, fermentation of glucose, arginine dihydrolase, urease, esculin hydrolysis, gelatin liquefaction, beta-galactosidase (PNPG) and assimilation of maltose, caprate and adipate; strain-dependent reactions for the assimilation of citrate (type strain weak). When tested using API ZYM strips, positive for alkaline phosphatase, C4 lipase, C8 lipase (weak), leucyl arylamidase, valine arylamidase, acid phosphatase and phosphoamidase; negative for C14 lipase, trypsin, chymotrypsin, alpha-galactosidase, beta-galactosidase, beta-glucuronidase, alpha-glucosidase, beta-glucosidase, N-acetyl-beta-glucosaminidase, alpha-mannosidase and alpha-fucosidase; strain-dependent reactions for cystine arylamidase (type strain negative). The following fatty acids are present in all isolates: $C_{16:0}$, $C_{16:0}$ 3-OH, $C_{17:0}$ cyclo, $C_{18:1}$ ω7c, $C_{19:0}$ cyclo ω8c, summed feature 2

(most likely $C_{14:0}$ 3-OH) and summed feature 3 (most likely $C_{16:1}$ ω7c) in moderate amounts (>5 %), and $C_{14:0}$, $C_{16:0}$ 2-OH and $C_{16:1}$ 2-OH in minor amounts (1-5 %) (mean value of all isolates).
The type strain is LMG 29315^T (=CCUG 68414^T) and was isolated from human lung tissue in the USA in 2006. Its G+C content is 63.2 mol%. The 16S rRNA, *gyrB* and whole-genome sequence of LMG 29315^T are publicly available through the accession numbers LT158613, LT158626 and FCNV01000000, respectively. An additional strain has been isolated from human blood in the USA in 2010 (Table 5.3).

5.2.5.11 Description of *Burkholderia turbans* sp. nov.

Burkholderia turbans (tur'bans. L. fem. part. pres. *turbans* disturbing, agitating, because the isolation of this bacterium from human pleural fluid further disturbs the image of this lineage of *Burkholderia* species as benign bacteria).
Cells are Gram-negative, non-motile rods (about 1 µm wide and 1-1.5 µm long) with rounded ends that occur as single units or in pairs. After 48 h of incubation on trypticase soy agar at 28 °C, colonies are round (typically less than 1 mm in diameter), smooth, shiny, non-translucent, with entire margins and a white-creamy color. Grows on MacConkey agar. Growth occurs at 15-37 °C and at pH 6-7 in NB at 28 °C. Catalase and oxidase activities are present. Hydrolyzes tween 60, but not tween 80, starch and casein. When tested using API 20NE strips, positive for the assimilation of glucose, arabinose, mannose, mannitol, N-acetyl-glucosamine, gluconate, caprate, malate and phenylacetate; negative for nitrate reduction, production of indol, fermentation of glucose, arginine dihydrolase, urease, esculin hydrolysis, gelatin liquefaction, beta-galactosidase (PNPG) and assimilation of maltose, adipate and citrate. When tested using API ZYM strips, positive for alkaline phosphatase, C4 lipase (weak), leucyl arylamidase, acid phosphatase and phosphoamidase (weak); negative for C8 lipase, C14 lipase, valine arylamidase, cystine arylamidase, trypsin, chymotrypsin, alpha-galactosidase, beta-galactosidase, beta-glucuronidase, alpha-glucosidase, beta-glucosidase, N-acetyl-beta-glucosaminidase, alpha-mannosidase and alpha-fucosidase. The following fatty acids are present: $C_{16:0}$, $C_{17:0}$ cyclo, $C_{18:1}$ ω7c, summed feature 2 (most likely $C_{14:0}$ 3-OH) and summed feature 3 (most likely $C_{16:1}$ ω7c) in moderate amounts (>5 %), and $C_{14:0}$, $C_{16:0}$ 2-OH, $C_{16:0}$ 3-OH, $C_{16:1}$ 2-OH and $C_{19:0}$ cyclo ω8c in minor amounts (1-5 %).
The type strain is LMG 29316^T (=CCUG 68413^T) and was isolated from human pleural fluid in the USA in 2006. Its G+C content is 63.1 mol% (calculated based on its genome sequence). The 16S rRNA, *gyrB* and whole-genome sequence of LMG 29316T are publicly available through the accession numbers LT158614, LT158627 and FCOD01000000, respectively.

5.2.5.12 Description of *Burkholderia catudaia* sp. nov.

Burkholderia catudaia (ca.tu.da'ia. Gr. adj. *catudaios* subterraneous; N. L. fem. adj. *catudaia*, earth-born).
Cells are Gram-negative, non-motile rods (about 1 µm wide and 1-2 µm long) with rounded ends that occur as single units or in pairs. After 48 h of incubation on trypticase soy agar at 28 °C, colonies are round (typically less than 1 mm in diameter), smooth, shiny, non-translucent, with entire margins and a white-creamy color. Grows on MacConkey agar. Growth occurs at 15-37 °C and at pH 6-7 in NB at 28 °C. Catalase and oxidase activities are present. Hydrolyzes tween 60, but not tween 80, starch and casein. When tested using API 20NE strips, positive for nitrate reduction and assimilation of glucose, arabinose, mannose, mannitol, N-acetyl-glucosamine, gluconate, malate and phenylacetate; negative for production of indol, fermentation of glucose, arginine dihydrolase, urease, esculin hydrolysis, gelatin liquefaction, beta-galactosidase (PNPG) and assimilation of maltose, caprate, adipate and citrate. When tested using API ZYM strips, positive for alkaline phosphatase (weak), leucyl arylamidase, acid phosphatase and phosphoamidase (weak); negative for C4 lipase, C8 lipase, C14 lipase, valine arylamidase, cystine arylamidase, trypsin, chymotrypsin, alpha-galactosidase, beta-galactosidase, beta-glucuronidase, alpha-glucosidase, beta-glucosidase, N-acetyl-beta-glucosaminidase, alpha-mannosidase and alpha-fucosidase. The following fatty acids are present: $C_{16:0}$, $C_{16:0}$ 3-OH, $C_{18:1}$ ω7c, summed feature 2 (most likely $C_{14:0}$ 3-OH) and summed feature 3 (most likely $C_{16:1}$ ω7c) in moderate amounts (>5 %), and $C_{14:0}$, $C_{16:0}$ 2-OH, $C_{17:0}$ cyclo and $C_{19:0}$ cyclo ω8c in minor amounts (1-5 %).
The type strain is LMG 29318^T (=CCUG 68411^T) and was isolated from *Fadogia homblei* rhizosphere soil in South Africa in 2013 (Verstraete *et al.*, 2014). Its G+C content is 62.8 mol% (calculated based on its genome sequence). The 16S rRNA, *gyrB* and whole-genome sequence of LMG 29318^T are publicly available through the accession numbers LT158616, LT158629 and FCOF01000000, respectively.

5.2.5.13 Description of *Burkholderia peredens* sp. nov.

Burkholderia peredens (per.e'dens. L. fem. part. pres. *peredens* consuming, devouring; referring to the capacity of this bacterium to degrade fenitrothion).
Cells are Gram-negative, non-motile rods (about 1 µm wide and 1-2 µm long) with rounded ends that occur as single units or in pairs. After 48 h of incubation on trypticase soy agar at 28 °C, colonies are round (typically less than 1 mm in diameter), smooth, shiny, non-translucent, with entire margins and a white-creamy color. Grows on MacConkey agar. Growth occurs at 15-37 °C and at pH 7 in NB at 28 °C. Catalase and oxidase activities are

present. Hydrolyzes tween 60, but not tween 80, starch and casein. When tested using API 20NE strips, positive for the assimilation of glucose, arabinose (weak), mannose, mannitol, N-acetyl-glucosamine, gluconate, malate and phenylacetate; negative for nitrate reduction, production of indol, fermentation of glucose, arginine dihydrolase, urease, esculin hydrolysis, gelatin liquefaction, beta-galactosidase (PNPG) and assimilation of maltose, caprate, adipate and citrate. When tested using API ZYM strips, positive for alkaline phosphatase, C4 lipase (weak), C8 lipase (weak), leucyl arylamidase, acid phosphatase and phosphoamidase (weak); negative for C14 lipase, valine arylamidase, cystine arylamidase, trypsin, chymotrypsin, alpha-galactosidase, beta-galactosidase, beta-glucuronidase, alpha-glucosidase, beta-glucosidase, N-acetyl-beta-glucosaminidase, alpha-mannosidase and alpha-fucosidase. The following fatty acids are present: $C_{16:0}$, $C_{16:0}$ 3-OH, $C_{18:1}$ ω7c, summed feature 2 (most likely $C_{14:0}$ 3-OH) and summed feature 3 (most likely $C_{16:1}$ ω7c) in moderate amounts (>5 %), and $C_{14:0}$, $C_{16:0}$ 2-OH, $C_{16:1}$ 2-OH and $C_{17:0}$ cyclo in minor amounts (1-5 %).
The type strain is LMG 29314^T (=CCUG 68415^T) and was isolated from soil in Japan (Hayatsu *et al.*, 2000). Its G+C content is 63.1 mol% (calculated based on its genome sequence). The 16S rRNA, *gyrB* and whole-genome sequence of LMG 29314^T are publicly available through the accession numbers LT158612, LT158625 and FCOH01000000, respectively.

5.2.5.14 Emended description of the species *Burkholderia sordidicola* Lim *et al.* 2003

The description of the species *Burkholderia sordidicola* is the one given by Lim *et al.* (2003) with the following modification. The G+C content of the type strain is 60.2 %.

5.2.5.15 Emended description of the species *Burkholderia zhejiangensis* Lu *et al.* 2012

The description of the species *Burkholderia zhejiangensis* is the one given by Lu *et al.* (2012) with the following modification. The G+C content of the type strain is 62.7 %.

5.2.5.16 Emended description of the species *Burkholderia grimmiae* Tian *et al.* 2013

The description of the species *Burkholderia grimmiae* is the one given by Tian *et al.* (2013) with the following modification. The G+C content of the type strain is 63.0 %.

5.2.6 Acknowledgments

CP is indebted to the Special Research Council of Ghent University. We thank Marcus Dillon for constructing sequencing libraries, Aurélien Carlier for his advice on genome analysis and all strain depositors listed in Table 5.3 to make this study possible.

5.2.7 Supplementary material

The GenBank/EMBL/DDBJ accession numbers for the 16S rRNA and *gyrB* gene sequences determined in this study are LT158612-LT158624 and LT158625-LT158641, respectively.

Supplementary Table 5.1: Pairwise GGD and dDDH values

Supplementary Table 5.1 is available online at http://journal.frontiersin.org/article/10.3389/fmicb.2016.00877.

Part IV

General Discussion and Future Perspectives

6 | Reflections on the epidemiology of *Burkholderia multivorans*

The first goal of the present thesis was to gain better insight into the epidemiology of *B. multivorans* by examining its environmental niche and by comparing the genomes of clinical and environmental *B. multivorans* isolates. **Chapter 3** presented the experimental work which demonstrated **(i)** that *B. multivorans* DNA is present in water and – to a greater extent – soil samples (Section 3.1), **(ii)** that *B. multivorans* could not be isolated from PCR positive water and soil samples (Sections 3.2 and 3.3), and **(iii)** that the genomic lineages of *B. multivorans* are defined by their multilocus sequence type (Section 3.4). The present chapter provides a general discussion of these results and some future perspectives.

6.1 Cultivation-independent detection of *Burkholderia multivorans* in environmental samples

Although semi-selective isolation media have been developed for the isolation of Bcc bacteria from environmental samples (Burbage & Sasser, 1982; Hagedorn *et al.*, 1987; Vermis *et al.*, 2003a; Vanlaere *et al.*, 2005), the procedure is still very time-consuming and labor-intensive and thus far only few environmental *B. multivorans* isolates have been obtained using this strategy. The aim of our first study (Section 3.1) was to gain insight into the environmental niche of *B. multivorans* without the need for labor-intensive isolation studies. We therefore followed a cultivation-independent approach that was based on a total DNA extraction from water and soil samples and a *B. multivorans*-specific *recA* PCR assay.

PCR-based diagnostic tests were developed in the 1990s to distinguish Bcc species and *B. multivorans*-specific primers based on both the 16S rRNA and *recA* gene were available from previous studies (LiPuma *et al.*, 1999; Mahenthiralingam *et al.*, 2000a). In later studies, these species-specific primers were used for the early detection and genomovar determination

of Bcc organisms in CF sputum samples (Drevinek *et al.*, 2002). In addition, Miller *et al.* (2002) used the 16S rRNA-based PCR assays for the detection of Bcc bacteria in soil samples. Because our preliminary experiments showed that the 16S rRNA-based PCR assay yielded many false positive results when applied to water samples (Section 3.1.4.1), we used *recA*-based primers to further develop our PCR assay. The goal of this first study was to gain insight into the environmental occurrence of *B. multivorans* in a time and cost-effective way and we therefore preferred to apply a conventional PCR method with gel-based detection of the resulting amplicons, rather than a qPCR. The used species-specific *recA* primers (Mahenthiralingam *et al.*, 2000a) were previously evaluated for their sensitivity and specificity by Vermis *et al.* (2002a) and were found to be 100 % specific (as tested on 508 Bcc isolates). Additionally, the combination of two different primer pairs in our nested PCR design (with primers specific for Bcc and *B. multivorans*) helped to further decrease the number of false positives. However, preliminary experiments (data not shown) showed that false positive amplicons were occasionally obtained when applying our PCR assay to water isolates belonging to the genera *Ochrobactrum*, *Pseudochrobactrum*, *Sphingomonas* and *Chryseobacterium*. Therefore, all amplicons were sequenced to ensure that only true positives were recorded. Whereas this final check excluded a considerable number of false positives from the water samples, this was not the case for the soil samples (Section 3.1.5), demonstrating that this *recA*-based PCR assay in its current form is not suited for routine analysis of water samples and that one should re-evaluate the specificity of PCR-based methods when applying them to environmental samples. Accordingly, Bergmark *et al.* (2012) showed that a high specificity as calculated from *in silico* analysis does not guarantee a high specificity of the primers when applied to environmental samples.

Because the few environmental *B. multivorans* isolates that have been reported mainly originated from water samples (Vermis *et al.*, 2003a; Fang *et al.*, 2011), we hypothesized that water rather than soil was the most likely environmental niche of *B. multivorans* and our study therefore focused primarily on water samples. However, only 12 out of 112 water samples (11 %) examined were *B. multivorans* PCR positive in contrast to 23 out of 25 soil samples (92 %) examined (Section 3.1.4.2). Our finding that *B. multivorans* is widely distributed in soil samples contrasted with results of previous isolation campaigns, but nevertheless agreed with the notion that the soil environment typically harbors large genome-sized organisms (Konstantinidis & Tiedje, 2004; Raes *et al.*, 2007).

Because Stopnisek *et al.* (2014) previously demonstrated that the biogeographical distribution of *Burkholderia* bacteria in soil could be attributed to their acid tolerance, we hypothesized to detect *B. multivorans* more commonly in acidic versus alkaline waters. However, binary logistic regression showed that the probability of detecting *B. multivorans* was higher in water

samples with a higher pH and in streams compared to swimming and recreational waters (Section 3.1.4.3). This unexpected finding may again indicate that water is not the natural reservoir of *B. multivorans*. However, the number of PCR positive water samples was rather small and one should therefore be cautious when extrapolating the results of our binary logistic regression model. Staley *et al.* (2016) recently demonstrated that physico-chemical parameters poorly explained the variation in relative abundances of bacterial families as determined through 16S rRNA amplicon sequencing. Furthermore, the same study showed that sediments and soils may act as a reservoir for the biodiversity detected in Mississippi river water. With the ever-decreasing cost of high-throughput sequencing, a shotgun metagenomic analysis of a large number of environmental samples may shed more light on the specific determinants of the environmental occurrence of *B. multivorans*.
In conclusion, *B. multivorans* can be detected in water and - to a greater extent - soil samples, demonstrating that this Bcc species is widespread in the natural environment. Therefore, we subsequently aimed to isolate *B. multivorans* from PCR positive environmental samples in order to genotype the environmental isolates and to compare them with isolates from Belgian CF patients.

6.2 Cultivating the obvious: the failure to isolate *Burkholderia multivorans* from environmental samples

While the epidemiology of Bcc bacteria in CF patients suggests that *B. multivorans* is acquired from non-human sources such as the natural environment (Baldwin *et al.*, 2008), it has rarely been isolated from environmental samples (Table 6.1). The comparison of clinical and environmental isolates may provide new insights into the epidemiology of this CF pathogen and we therefore applied multiple isolation strategies in a stepwise manner to water and soil samples that were previously shown to be PCR positive for *B. multivorans* (Section 3.1 and 6.1).

Table 6.1: Environmental *B. multivorans* isolates

Source	Number of isolates / samples	Liquid enrichment prior to plating	C- and N-sources	Selective agents	pH	ST	References
Schelde and Leie river (June-October 2002, Belgium)	8 / 4	yes	L-arabinose and D-cellobiose (EB1) or L-arabinose and L-threonine (EB2)	Polymyxin B sulfate, vancomycin, C-390	Not adjusted	374 418 782 783 784 836	(Vermis *et al.*, 2003a)
Succulent soil in veranda of CF patient (October-November 2003, Belgium)	2 / 1	yes	L-arabinose and L-threonine	Polymyxin B sulfate, vancomycin, C-390, gentamicin	Not adjusted	189 781	(Vanlaere *et al.*, 2005)
Maize rhizosphere soil (July 2003, USA)	3 / 1	no	Glucose and L-asparagine (TB-T)	Trypan blue, crystal violet, tetracycline	5.5	NA	(Ramette *et al.*, 2005)
West Lake (October 2006, China)	48 / 1	no	Azelaic acid and tryptamine (PCAT)	-	5.8	274	(Fang *et al.*, 2011)

Since the 1980s, several semi-selective isolation media have been developed for the recovery of Bcc bacteria from CF sputum. These include *Pseudomonas cepacia* agar (Gilligan *et al.*, 1985), OFPBL (oxidation-fermentation base supplemented with agar, polymyxin B, bacitracin and lactose) (Welch *et al.*, 1987) and BCSA (Henry *et al.*, 1997). Several studies (Henry *et al.*, 1999; Vermis *et al.*, 2003b) compared the recovery of Bcc bacteria on different growth media and concluded that BCSA is best for Bcc recovery from CF samples. Because of their high nutrient and antibiotic concentrations and/or low specificity these media are generally not suited for the isolation of Bcc bacteria from environmental samples (Hagedorn *et al.*, 1987; Vartoukian *et al.*, 2010). Therefore, other media were developed for the isolation of Bcc bacteria from complex environmental samples such as PCAT (Burbage & Sasser, 1982), TB-T (Hagedorn *et al.*, 1987) and BCEM (Vermis *et al.*, 2003a; Vanlaere *et al.*, 2005). When comparing PCAT with TB-T, the latter was shown to be more selective but a lot less sensitive, i.e. less Bcc isolates could be recovered on TB-T (Tabacchioni *et al.*, 2000; Pallud *et al.*, 2001; Vermis *et al.*, 2003b). Since the diversity of the obtained isolates is strongly dependent on the selective medium (Tabacchioni *et al.*, 2000; Vermis *et al.*, 2003a), two selective media were used in each isolation experiment performed in the present study.

BCEM was designed based on a re-assessment of C-source utilization and antibiotic susceptibility of Bcc bacteria (Vermis *et al.*, 2003b) and its application yielded a few *B. multivorans* isolates from water and soil samples (Table 6.1). We therefore used this medium as a starting point for further optimization with respect to the C-source composition and pH (Experiment 4, section 3.3.3). Because an increased pH results in a loss of selectivity (Hagedorn *et al.*, 1987) and because *Burkholderia* bacteria are known to be acid tolerant (Stopnisek *et al.*, 2014) we lowered the pH of BCEM to pH 5.5, analogous to that of PCAT and TB-T (Burbage & Sasser, 1982; Hagedorn *et al.*, 1987). Measuring the pH of BCEM enrichment broth after several days of incubation showed that MES buffer (Angle *et al.*, 1991) was effective in maintaining the pH of the enrichment broth at pH 5.5 (Experiment 4, section 3.3.3). Vermis *et al.* (2003a) used two combinations of C-sources (i.e. L-arabinose and D-cellobiose or L-arabinose and L-threonine), while Vanlaere *et al.* (2005) previously used only the latter formulation (Table 6.1). Because experiment 4 (Section 3.3.3) showed that *B. multivorans* DNA was only detected in an enrichment broth in which the three carbon sources were combined, our BCEM formulation subsequently contained all three C-sources, i.e. L-arabinose, D-cellobiose and L-threonine.

Because *B. multivorans* DNA was detected through PCR experiments in the total cell mass that was harvested from BCEM agar plates (Experiment 4, section 3.3.3), we applied previously described isolation strategies, i.e. direct plating and liquid enrichment of water samples (Experiment A, section 3.2.4.1) and direct plating of soil samples (Experiment B,

section 3.2.4.2), to *B. multivorans* PCR positive samples. Direct plating of two soil samples yielded no *B. multivorans* isolates, neither on BCEM nor on PCAT medium (Experiment B, section 3.2.4.2), which is in agreement with previous isolation studies in which soil samples were plated on PCAT medium, but no *B. multivorans* was isolated (Balandreau *et al.*, 2001; Pirone *et al.*, 2005; Zhang & Xie, 2007). Direct plating of water samples yielded no *B. multivorans* isolates either (Experiment A, section 3.2.4.1). Vermis *et al.* (2003a) reported that liquid enrichment prior to plating may enhance the isolation of Bcc from water samples. Concordantly, Ahn *et al.* (2014) showed that broth media were more effective than solid media for the recovery of *B. cenocepacia* from distilled water. However, Fang *et al.* (2011) isolated several Bcc bacteria via direct plating of concentrated water samples on PCAT, suggesting that liquid enrichment is not always necessary. In our study, neither strategy yielded *B. multivorans* isolates. Nevertheless, a higher number of other, non-Bcc, *Burkholderia* isolates was obtained when using liquid enrichment prior to plating (Experiment A, section 3.2.4.1). A first difference with the study of Vanlaere *et al.* (2005) was that in the present study BCEM did not contain the selective component C-390 because it was not available for purchase. However, Vanlaere *et al.* (2005) also reported the isolation other GNNF bacteria such as *Comamonas*, *Pandoraea*, *Chryseobacterium*, *Ralstonia*, *Herbaspirillum* and *Pseudomonas*, suggesting that C-390 was not fully effective to inhibit the growth of these genera. A second difference with the study of Vanlaere *et al.* (2005) was our use of MALDI-TOF MS (Ghyselinck *et al.*, 2011) coupled to sequence analysis of various genes (Mahenthiralingam *et al.*, 2000a) (rather than RAPD typing, growth in Stewart's medium and *recA* RFLP) which allowed to dereplicate and identify a much larger number of isolates (e.g. 626 isolates in experiments A and B, sections 3.2.4.1 and 3.2.4.2).

Because direct plating and liquid enrichment of water samples and direct plating of soil samples did not yield *B. multivorans* isolates, we attempted to improve the liquid enrichment of *Burkholderia* bacteria and examined the effect of lowering the pH, extending the incubation time and shaking the enrichment broth (Experiments 7 and 8, sections 3.3.4 and 3.3.5). In contrast with what was expected based on the genus-wide acid tolerance of *Burkholderia* (Stopnisek *et al.*, 2014), lowering the pH of the BCEM enrichment broth to pH 4.5 did not result in a general increase of the relative abundance of *Burkholderia* bacteria as determined through qPCR. Although varying and inconsistent results were obtained with different water samples, shaking the BCEM enrichment broth at pH 5.5 generally increased the relative abundance of *Burkholderia* bacteria and the detection of *B. multivorans* (as determined through qPCR and the *B. multivorans*-specific PCR assay, respectively), confirming the validity of the approach that was used in previous experiments (Experiment A, section 3.2.4.2).

As an alternative approach, we applied several strategies which had been reported to enable

the cultivation of fastidious bacteria (Vartoukian *et al.*, 2010; Epstein, 2013). Besides liquid enrichment of a concentrated water sample in a semi-selective broth, a non-selective recovery strategy in a 0.1 % peptone solution was applied to give debilitated cells the opportunity to resuscitate prior to being exposed to the selective agar medium (Straka & Stokes, 1957; Ahn *et al.*, 2014). This approach (Experiment C, section 3.2.4.3) yielded several GNNF opportunistic pathogens in CF patients including *Acinetobacter calcoaceticus*, *Inquilinus limosus* and *S. maltophilia*, yet no *B. multivorans*. Acclimatizing recovery is an approach based on the filtration-acclimatization method and allows a gradual transition from the original environment, i.e. the water sample, to semi-selective conditions with high nutrient and antibiotic concentrations (Hahn *et al.*, 2004). Also the use of this strategy (Experiment C, section 3.2.4.3) did not yield *B. multivorans* isolates from water sample W2.

The use of a transmembrane system allows the diffusion of essential nutrients and/or signaling molecules between two compartments and previously led to the successful cultivation of several fastidious bacteria (Vartoukian *et al.*, 2010; Epstein, 2013). Because Bcc bacteria thrive in the CF lung as opportunistic pathogens and can be readily isolated from CF sputum samples we tried to mimic the CF lung environment by incubating a water sample in the presence of CF sputum with its accompanying microbiota in a transmembrane system (Experiment D, section 3.2.4.4). Simultaneously, the same samples were co-incubated with helper strain panels comprising other bacteria that also occur in the CF lung (i.e. *B. multivorans*, *P. aeruginosa* and *S. aureus*) (LiPuma, 2010) and that may influence the growth or cultivability of *B. multivorans* (Experiment 16, section 3.3.9). While these experiments again yielded several opportunistic GNNF pathogens relevant in CF, none were *B. multivorans*.

Both Vermis *et al.* (2003a) and Fang *et al.* (2011) reported an increased isolation of *B. multivorans* in (early) fall, suggesting a seasonal effect. Accordingly, Purdy-Gibson *et al.* (2015) found that *P. aeruginosa* could readily be isolated from drains in the home environment, but that the recovery was significantly higher in fall relative to other seasons. Because bacteria may be dormant in cold environments we examined whether the temperature of the samples had an effect on the cultivability of *B. multivorans*. Experiments 11 and 12 (Section 3.3.7) tested the resuscitation of water and soil bacteria at room temperature and showed that although many *Burkholderia* isolates were obtained from the soil sample, none of them were *B. multivorans*. Moreover, several fresh, *B. multivorans* PCR positive samples were taken in August and October (Table 3.4 and 3.10) and water sample W2 originated from a pond in a greenhouse at constant temperature of about 25 °C, eliminating the seasonal temperature effect.

Because the use of diluted nutrient media has led to the successful cultivation of previously uncultivable bacteria from various aquatic and terrestrial habitats (Vartoukian *et al.*, 2010)

and because Bcc bacteria are known to survive in nutrient-poor conditions (Ahn *et al.*, 2014), we also examined the effect of a reduced nutrient concentration in the isolation medium (Experiments 11 and 12, section 3.3.7). For both tested water and soil samples, a larger number of colonies and a larger number of distinct colony morphologies were observed on 1/10 PCAT compared to standard PCAT, but none of these experiments yielded *B. multivorans* isolates. Nevertheless, 1/10 PCAT was used instead of PCAT in the remaining experiments.

Almost two decades ago Steinert *et al.* (1997) demonstrated that viable but non-cultivable *Legionella pneumophila* could be resuscitated by co-cultivation with *Acanthamoeba*. A more recent study used *Acanthamoeba* co-cultivation to isolate amoeba-resisting bacteria from environmental water samples (Pagnier *et al.*, 2008). Since amoebae can be found in freshwater environments and several Bcc strains were shown to survive in *Acanthamoeba* cells, amoebae have been suggested to be a natural reservoir for Bcc strains (Marolda *et al.*, 1999; Inglis & Sagripanti, 2006). More recently, a cultivation-independent study showed that *Burkholderia* was one of the genera that was found in free-living amoebae isolated from drinking water (Delafont *et al.*, 2013). We therefore examined the effect of co-cultivation of two water samples with *Acanthamoeba castellanii*, but no *B. multivorans* isolates were obtained (Experiment 13, section 3.3.8).

Finally, *Burkholderia* bacteria are frequently isolated from the plant rhizosphere and many species form intimate interactions with the roots of plants such as maize, lupine, rice and pea (Compant *et al.*, 2008). Vidal-Quist *et al.* (2014) recently showed that plant root colonization is an intrinsic property of Bcc bacteria and reported an active migration of Bcc bacteria toward plant roots. We therefore used germinating seedlings to "fish" Bcc bacteria out of an environmental sample (Experiment 9, section 3.3.6). Seeds of maize, pea, rice, barley and lupine were surface sterilized, germinated in either a water sample or a soil suspension and the roots were plated onto PCAT and BCSA medium. Pea seedling roots yielded *B. caledonica*, *B. phytofirmans* and one *Burkholderia* sp. isolate, but no *B. multivorans*. In line with these findings, both *B. caledonica* and *B. phytofirmans* were originally isolated from rhizosphere soil and were reported to form close associations with plants (Coenye *et al.*, 2001b; Sessitsch *et al.*, 2005). We also spiked control experiments with two *B. multivorans* strains which were recovered in high numbers from the pea, rice and maize seedling roots.

The present series of experiments demonstrated that despite numerous enrichment and cultivation efforts and the analysis of more than 2,400 isolates, none of the *B. multivorans* PCR positive environmental samples yielded *B. multivorans* isolates. Although different samples yielded inconsistent results, at least some samples showed the presence of *B. multivorans* in enrichment broth or in the total cell mass that was harvested from agar plates when BCEM

at pH 5.5 was used. The subsequent experiments yielded a wide diversity of *Burkholderia* and other GNNF bacteria, but none of the picked isolated was *B. multivorans*.

6.3 The discrepancy between PCR and cultivation-based detection of *Burkholderia multivorans* in environmental samples

The notion that only a small fraction of a diverse bacterial population can by cultivated using conventional cultivation-based techniques is known as "the great plate count anomaly" (Staley & Konopka, 1985) and led to a huge interest in cultivation-independent methods to study the bacterial diversity in complex environments such as the human gut and terrestrial and aquatic habitats (Venter *et al.*, 2004; Rousk *et al.*, 2010; Arumugam *et al.*, 2011). Metagenomic approaches have revolutionized microbiology by greatly improving our knowledge about the rare biosphere and the huge diversity of microbial life in a wide range of environments. However, one of the major drawbacks of the current metagenomic approaches is known as the "depth bias", referring to the fact that sequence-based techniques may miss clinically relevant minority populations (Lagier *et al.*, 2012). Another important drawback, in particular of the 16S rRNA amplicon-based approaches, is the limited discriminatory power of the short 16S rRNA gene fragments generated. Although this approach is very powerful to gain insight into the general diversity of a sample, a short fragment of the 16S rRNA gene does not allow to classify the obtained sequences at the species level. Finally, major discrepancies among studies reflect the biases of the employed techniques (Lagier *et al.*, 2012; Clooney *et al.*, 2016) and sequence-based microbiome analyses may be prone to major contaminations if the inclusion of proper control samples is ignored (Salter *et al.*, 2014). Recently, these drawbacks of metagenomic approaches led to a renewed interest in cultivation-based methods to study bacterial diversity (Lagier *et al.*, 2012; Shade *et al.*, 2012; Walker *et al.*, 2014a; Lagier *et al.*, 2015b).

In the present study, we wanted to gain insight into to environmental occurrence of *B. multivorans* by using a fast, cultivation-independent PCR assay, and we subsequently aimed to obtain environmental *B. multivorans* isolates from the PCR positive samples to be able to compare them with clinical *B. multivorans* isolates in an epidemiological study. After all, pure bacterial cultures remain essential for studying their virulence, their antibiotic susceptibility and their genome sequence in order to facilitate the understanding and treatment of the caused disease (Lagier *et al.*, 2015a).

Altogether, our series of experiments demonstrated a large discrepancy between PCR and

cultivation-based detection of *B. multivorans* in environmental samples. Although our cultivation-independent approach demonstrated that *B. multivorans* DNA is present in the natural environment, it should be noted that the detection of *B. multivorans* DNA does not necessarily imply that viable target organisms were present at the time of sampling (Josephson *et al.*, 1993). Nevertheless, the fact that DNA of this organism was present at the time of sampling and the fact that DNA degrades rather fast in soil (Bogas *et al.*, 2009) together imply that the target organism was present alive in these environmental samples at some point (Miller *et al.*, 2002). Additionally, the finding that at least some samples showed an enrichment (Experiments 7 and 8, sections 3.3.4 and 3.3.5) and/or the presence of *B. multivorans* among the colonies growing onto the selective agar media (Experiment 4, section 3.3.3) indicates that live cells must have been present in the environmental samples at the time of sampling.

More than a decade ago, Miller *et al.* (2002) reported that although 93 % of the examined (sub)urban soil samples were PCR positive for Bcc, only 15 % of these samples yielded Bcc isolates. Because of this discrepancy the authors concluded that the cultivation-based recovery of Bcc may underestimate their environmental presence. Payne *et al.* (2006) used a *recA*-based PCR assay to examine the *Burkholderia* diversity associated with the maize rhizosphere and found *recA* phylotypes that matched previously isolated Bcc species. However, 90 % of the *recA* sequences in the latter study were suggestive of novel taxa belonging to yet uncultivated *Burkholderia* species. Also for CF sputum samples this discrepancy between cultivation-based and DNA-based findings has been reported (Whitby *et al.*, 1998; Flight *et al.*, 2015).

Our failure to isolate *B. multivorans* from environmental samples is in stark contrast with the ever increasing number of unique *B. multivorans* strains that are isolated from CF patients. This finding may be explained by the fact that *B. multivorans* may not have been enriched sufficiently to be present on the agar plates with the highest dilutions from which isolates were picked. Another explanation for these observations is that *B. multivorans* cells in environmental samples are alive but acquired a viable but non-cultivable state. Because cells in a viable but non-cultivable state were shown not to lose their virulence potential (McDougald *et al.*, 1998; Oliver, 2010), it is currently unclear if and how these mere observations should be implemented in infection control guidelines. Future research will need to address why *B. multivorans* cannot readily be isolated from water and soil samples which are PCR positive. Live/dead staining in combination with FISH could be used to determine the number of viable and dormant *B. multivorans* cells in environmental samples and enrichment broths (Savichtcheva *et al.*, 2005). Alternatively, the survival and recovery of *B. multivorans* in a soil or water environment could be monitored via microcosm experiments (Delmont *et al.*,

2015).

The finding that *B. multivorans* DNA was detected far more often in soil than in water samples contrasted with the relatively large number of water isolates that are currently available. This may suggest that, if present in the natural environment, *B. multivorans* can be isolated more easily from water than from soil samples. *B. multivorans* can be readily cultivated from CF sputum samples and was shown to survive in a broader range of pH, temperatures and salt concentrations than *B. pseudomallei* (Lin *et al.*, 2011), suggesting that *B. multivorans* should be able to adapt well to environmental changes and varying growth conditions. Future cultivation efforts could try to improve the liquid enrichment conditions and capture the genetic diversity of *B. multivorans* in an enrichment broth using a shotgun metagenomic approach or single-cell genomics. An alternative enrichment strategy could exploit the capability of *B. multivorans* to form biofilms by providing solid surfaces for adhesion of *B. multivorans* in the enrichment broth (Gich *et al.*, 2012). Finally, an alternative "fishing" strategy could exploit the close *Burkholderia*-fungi interactions and the capability of *Burkholderia* bacteria to adhere to fungi to "fish" *B. multivorans* out of environmental samples or enrichment broths (Stopnisek *et al.*, 2016).

6.4 Genomic insights into the epidemiology of *Burkholderia multivorans*

The comparative genomics study of four pairs of clinical and environmental isolates representing four STs demonstrated that the genomic lineages of *B. multivorans* are well predicted by their ST (Section 3.4). The finding that the ST of the *B. multivorans* isolates predicted their phylogeny and gene content corroborated the use of MLST for epidemiological surveillance of Bcc bacteria in CF (Baldwin *et al.*, 2005). This confirms that the previous finding of the same ST in both clinical and environmental *B. multivorans* isolates thus indicates that the natural environment is a potential reservoir for *B. multivorans* (Baldwin *et al.*, 2008). Finally, this also demonstrated that the previous record of the same ST on different continents (Baldwin *et al.*, 2008) is not due to homoplasy as was previously shown for two STs of *B. pseudomallei* (De Smet *et al.*, 2015b).

The advent of long-read sequencing technologies such as PacBio SMRT sequencing has made it possible to resolve sequence repeats and assemble raw sequencing reads into high-quality, finished assemblies (Shin *et al.*, 2013; Koren & Phillippy, 2015). The availability of complete genomes enables the study of genome structure and the evolutionary consequences of genes being localized on different chromosomes. The highly conserved multireplicon genomic

structure that we found for *B. multivorans* agrees with the general genome architecture of Bcc organisms (Agnoli *et al.*, 2012). In line with the general genome biology of *Burkholderia* (Ussery *et al.*, 2009) the first and largest *B. multivorans* chromosome was enriched in core orthologs, while the other two chromosomes were enriched in ST-specific orthologs. As shown by Cooper *et al.* (2010), secondary chromosomes might serve as evolutionary test beds and the ST-specific orthologs located on the second and third chromosome are therefore expected to evolve faster. Although the three large replicons are generally considered chromosomes because they contain essential genes (Holden *et al.*, 2009) and at least one rRNA operon (Bentley & Parkhill, 2004), Agnoli *et al.* (2012, 2014) showed that the third chromosome of Bcc organisms is rather a megaplasmid than a chromosome that encodes accessory functions such as virulence and secondary metabolism and is involved in stress tolerance.

The comparative genomics study showed that a large set of orthologous genes was present in strains of the same ST regardless of their isolation source. However, differences between clinical and environmental isolates may occur at the level of gene regulation, rather than presence or absence of genes (Parke & Gurian-Sherman, 2001). Zhu *et al.* (2016) recently compared whole-genome sequences of four *Burkholderia seminalis* strains and concluded that the examined *B. seminalis* strains differed in metabolic capacities and in survival capacity in different niches. The strains showed an increased fitness in medium that simulated their original niche and this strain-specific niche survival was linked with the differential expression of genes. The availability of eight high-quality genome assemblies for *B. multivorans* isolates from both CF patients and the natural environment enables future research focusing on gene expression by means of transcriptome, proteome or phenome analysis.

Bacterial respiratory infections in CF patients present an important model for studying microbial adaptation (Silva *et al.*, 2016). Future research could examine the adaptation of *B. multivorans* to the CF lung using the four pairs of clinical and environmental isolates (Section 3.4) in competition assays or animal infection models. Examples of animal infections models that are frequently used to assess virulence are *Galleria mellonella* and *Caenorhabditis elegans* (Valvano *et al.*, 2005; Loutet & Valvano, 2010). However, the relevance of these infection models to chronic CF lung infection is unclear because virulence in these infection models is generally not a reliable predictor for the clinical outcome of Bcc infection in CF patients (Eberl & Vandamme, 2016). For *B. multivorans* there is a complete lack of a suitable infection model, as exemplified by the fact that *B. multivorans* is mostly avirulent in *G. mellonella* and *C. elegans* infection models, while most putative virulence factors that were suggested to be indicative for pathogenic *Burkholderia* species were identified in this Bcc species (Angus *et al.*, 2014; Eberl & Vandamme, 2016).

As *B. multivorans* harbors a ~6.5 Mb genome, it is equipped with the metabolic versatility

needed to thrive in a complex, variable environment such as soil. The fact that the soil environment typically harbors large genome-sized organisms (Konstantinidis & Tiedje, 2004; Raes *et al.*, 2007) corroborates the finding of our cultivation-independent study that *B. multivorans* DNA is widespread in soil (Section 3.1). This metabolic versatility found in *B. multivorans* genomes nevertheless contrasts with the failure to isolate *B. multivorans* from environmental samples (Sections 3.2 and 3.3). Future research could use the *B. multivorans* genomes to try and explain why *B. multivorans* cannot readily be isolated from water and soil samples and subsequently use this genomic information to design new isolation media as previously done for *Tropheryma whipplei* (Renesto *et al.*, 2003). Alternatively, one could search for unique catabolic activities which may help in designing novel semi-selective isolation media. Finally, future studies could focus on the genome biology of this organism and try to infer the lifestyle of this organism based on genome data (Barberan *et al.*, 2014; Livermore *et al.*, 2014).

7 | Epidemiology and taxonomy of cystic fibrosis pathogens

The second goal of the present thesis was to contribute to the general knowledge of the prevalence and epidemiology of GNNF bacteria in CF. **Chapter 4** presented **(i)** the re-evaluation and application of the 3 % threshold value of concatenated allele sequence divergence for species delineation within the Bcc (Section 4.1) and **(ii)** the use of MLSA in two polyphasic studies that resulted in the formal classification of three novel Bcc species (Sections 4.1 and 4.2). **Chapter 5** presented **(i)** a traditional polyphasic taxonomic study in which species delineation was based on DDH (Section 5.1) and **(ii)** a modern polyphasic taxonomic study in which species delineation was based on whole-genome sequence-based parameters (Section 5.2). The present chapter provides a general discussion of these results, reflections on the NRC Bcc and the use of MLST for surveillance of Bcc infections in CF and some future perspectives.

7.1 Epidemiology of Gram-negative non-fermenting cystic fibrosis pathogens

The LM-UGent research group has been studying the diversity of *Burkholderia* bacteria and other GNNF bacteria since the early 1990s. For two decades, strains of CF patients were identified sporadically when requested by CF care and there was no systematic surveillance of Bcc infections. Since 2011, NRCs for human microbiology are being financed by the Health Insurance System of the Belgian Ministry of Social Affairs and the NRC Bcc is charged with the surveillance of respiratory infections caused by GNNF bacilli (except *P. aeruginosa* and *Acinetobacter* spp.) in Belgian CF patients. While this NRC primarily aims to identify and type CF isolates, also non-CF human clinical and non-clinical isolates can be analyzed if this provides clinically relevant information related to the epidemiology of GNNF pathogens in CF.

Altogether, these long-term studies contributed to the general knowledge of the diversity of GNNF bacteria in CF patients and in the environment.

7.1.1 The NRC Bcc and its role in the surveillance of GNNF pathogens in Belgian CF patients

The NRC Bcc exists as a cooperation between UZ Brussel and LM-UGent (https://nrchm.wiv-isp.be/nl/ref_centra_labo/burkholderia_cepacia_complex). UZ Brussel is responsible for the preliminary identification and characterization of isolates using MALDI-TOF MS, RAPD typing and antibiotic sensitivity testing, and aims to analyze two isolates per species per patient per year. LM-UGent is responsible for the molecular identification and typing of the isolates by gene sequence analysis and MLST, respectively, and aims to analyze one isolate per species per patient per year.
In the five-year period 2011-2015, 565 isolates were identified, including 72 non-CF isolates. Of the 493 CF isolates, 170 were identified as *Burkholderia* (Table 7.1). The 169 Bcc CF isolates (originating from 77 patients) were identified using *recA* gene sequence analysis and 60 % of the isolates belonged to *B. multivorans* (Fig. 7.1), confirming that *B. multivorans* is still the most prevalent Bcc species in Belgian CF patients (De Boeck *et al.*, 2004). MLST analysis of these Bcc CF isolates revealed many unique *B. multivorans* strains but also three *B. multivorans* strains that were present in multiple Belgian CF patients and centers (Table 7.2). Future research is needed to examine these cases of implied patient-to-patient spread in order to improve infection control guidelines.

Table 7.1: Isolate overview NRC Bcc 2011-2015

	CF					Non-CF
	Total	***Burkholderia***	***Achromobacter***	***Stenotrophomonas***	**Other GNNF**	**Total**
2011	38	30	3	0	5	7
2012	67	40	5	17	5	14
2013	98	35	26	23	14	22
2014	152	32	56	23	41	15
2015	138	33	47	24	34	14
Total	493	170	137	87	99	72

The usefulness of analyzing non-CF Bcc isolates is illustrated by the epidemiological investigation of a *B. cepacia* outbreak among hospitalized non-CF patients at the end of 2013. *B. cepacia* was found in one batch of liquid soap and the same strain (ST-848) was retrieved from respiratory, blood and urine samples of hospitalized patients (Jans *et al.*, 2014). Bcc organisms have been reported as contaminating agents of many pharmaceutical products

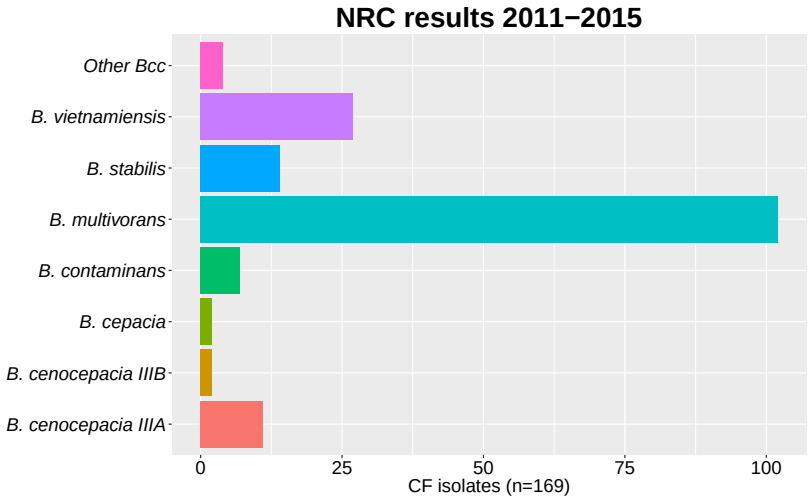


Figure 7.1: Bcc species distribution in Belgian CF patients (NRC Bcc 2011-2015)

Table 7.2: *B. multivorans* strains present in multiple Belgian CF patients and centers (NRC Bcc 2011-2015)

Strain	Isolates	Patients	Centers
ST-739	8	6	4
ST-741	19	10	6
ST-742	13	3	3

including cosmetics and disinfectants, posing a significant health risk for CF and other immunocompromised patients (Weber *et al.*, 2007; Torbeck *et al.*, 2011; De Smet *et al.*, 2013).

7.1.2 The Bcc PubMLST database and its role in epidemiological studies worldwide

One of the major advantages of the MLST approach over previous typing and identification methods is its reproducibility and portability (Jolley & Maiden, 2010). The public database (http://pubmlst.org/bcc/) allows researchers worldwide to analyze and deposit data, enabling the study of the prevalence and epidemiology of Bcc bacteria. As curator of the Bcc PubMLST database, we experienced a geographical expansion of interest in Bcc and *Burkholderia* organisms. While the first few hundred STs were primarily originating from European, American and Canadian isolates, there has been an increase in submissions from Asian and South-American countries. The large number of recent publications that used MLST and the accompanying PubMLST database as a tool for epidemiological studies shows that MLST

has become a well-established method for studying the population structure of Bcc organisms (Zurita *et al.*, 2014; Segonds *et al.*, 2015; Wang *et al.*, 2015; Vasiljevic *et al.*, 2016).
Although whole-genome sequencing is revolutionizing (clinical) microbiology practices (Didelot *et al.*, 2012; McAdam *et al.*, 2014), a systematic and standardized description of genotypic variation as provided by MLST will remain important (Maiden *et al.*, 2013; Jolley & Maiden, 2014). Whole-genome sequencing will therefore likely serve as a complement for existing MLST schemes rather than a replacement. Depending on the taxonomic resolution required, different sets of genes can be extracted from genomic sequences and the application of multiple MLST schemes can provide resolutions from the domain to the strain level (Jolley & Maiden, 2010; Jolley *et al.*, 2012).

7.1.3 The natural environment as a niche for opportunistic CF pathogens

A multitude of cultivation strategies were applied to environmental water and soil samples during our search for environmental *B. multivorans* isolates (Sections 3.2 and 3.3). Using Bcc isolation media for the study of only a few environmental samples, a wide diversity of *Burkholderia* bacteria and many other GNNF bacteria relevant in CF were isolated from both water and soil samples. Among the GNNF bacteria isolated were *Cupriavidus*, *Pandoraea*, *Inquilinus*, *Pseudomonas* and *Stenotrophomonas* species, confirming that water and soil are indeed potential reservoirs of these rare but potentially virulent pathogens (LiPuma, 2010). Rather unexpectedly, also several *Mycobacterium* species were isolated (Sections 3.2 and 3.3). Rapidly growing mycobacteria have emerged as significant pathogens in CF (Bar-On *et al.*, 2015) and, similar to Bcc bacteria, they can be recovered from environmental sources (van Ingen *et al.*, 2009).
As argued by Berg *et al.* (2005, 2014), several mechanisms that are important for colonization of the rhizosphere and antagonistic activity against plant pathogens (i.e. survival in the soil environment) are similar to those responsible for colonization of human tissues and pathogenicity. Similar to the natural environment, the human body imposes harsh conditions upon any organism attempting to colonize it. Opportunistic bacteria may overcome these by a broad metabolic versatility, production of siderophores to scavenge iron and the formation of biofilms which protect them from the host immune system.

7.2 Taxonomy of Gram-negative non-fermenting cystic fibrosis pathogens

7.2.1 Novel insights into the taxonomy of Gram-negative non-fermenting CF pathogens

The devastating effect of pulmonary infections in CF patients triggered numerous polyphasic taxonomic studies of clinical and environmental *B. cepacia*-like GNNF bacteria that contributed to the general knowledge of the prevalence and epidemiology of GNNF bacteria in CF. The taxonomic work performed in the context of the present thesis led to the formal classification of several novel species in the genera *Burkholderia*. Concomitantly with the analyses of novel *Burkholderia* isolates, we also contributed to the description of several novel *Achromobacter* and *Bordetella* species.

A first taxonomic effort focused on the diversity of Bcc bacteria and resulted in the formal classification of three novel Bcc species (Chapter 4), extending the list of validly named Bcc species from 17 to 20.

A second effort focused on the diversity of *B. glathei*-like bacteria and resulted in the formal classification of 14 novel *B. glathei*-like species from both human clinical and environmental sources (Chapter 5). Most of the novel species in this study originated from soil (Section 5.2) (Draghi *et al.*, 2014; Verstraete *et al.*, 2014), illustrating the huge diversity of *Burkholderia* organisms in the natural environment. However, *B. glathei*-like bacteria developed a wide range of interactions with different types of hosts including plants, insects and fungi (Section 5.2).

Finally, although not part of the present PhD thesis, a third effort focused on the diversity of *Achromobacter* and *Bordetella* bacteria in CF and resulted in the formal classification of four novel *Achromobacter* species (Vandamme *et al.*, 2013b) and three novel *Bordetella* species (Vandamme *et al.*, 2015).

7.2.2 Taxonomic methodology in a changing landscape

The CF lung can harbor a wide range of bacteria (LiPuma, 2010) and accurate identification of CF pathogens at the species level is important to assess the clinical impact of these generally rare opportunistic pathogens. The polyphasic taxonomic approach originated more than two decades ago by the complementation of early classifications based on morphology and biochemical data with chemotaxonomic and genotypic data (Vandamme *et al.*, 1996). This consensus approach integrated genotypic, phenotypic and phylogenetic information to

provide a multifaceted view on the biology of a bacterium. Using this consensus approach to prokaryotic systematics, a bacterial species is defined as a group of strains that are characterized by a certain degree of phenotypic consistency, more than 70 % DDH and more than 97 % of 16S rRNA gene sequence similarity (Stackebrandt & Goebel, 1994; Vandamme *et al.*, 1996). The ever-decreasing cost of high-throughput sequencing technologies has been revolutionizing taxonomic methodologies however (Klenk & Goker, 2010; Sutcliffe *et al.*, 2013; Whitman, 2015) and throughout the present thesis, an evolution in taxonomic methods can be observed with regards to phylogeny and species delineation.

7.2.2.1 Evolution in methods to infer phylogeny

Long before the advent of polyphasic taxonomy, there was a general agreement among taxonomists that the genomic DNA sequence should be the reference standard to determine phylogeny and that phylogeny should determine taxonomy (Wayne *et al.*, 1987). Sequence analysis of the universal 16S rRNA gene was recognized as a good tool to determine the phylogenetic position of any prokaryotic organism, but lacked the resolving power to discriminate closely related species (Fox *et al.*, 1992; Stackebrandt & Goebel, 1994). Also within the genera *Burkholderia*, *Achromobacter* and *Bordetella*, the 16S rRNA gene only has limited taxonomic resolution for species identification. As shown in chapters 4 and 5, sequence analysis of the *recA* and *gyrB* housekeeping genes is a powerful taxonomic tool for *Burkholderia* species identification (Mahenthiralingam *et al.*, 2000a; Tabacchioni *et al.*, 2008; Tayeb *et al.*, 2008; Vandamme *et al.*, 2013a). Similarly, sequence analysis of an internal fragment of *nrdA* gene was successfully applied for species differentiation within the genera *Achromobacter* and *Bordetella* (Vandamme *et al.*, 2013b, 2015).

Because the analysis of multiple genes provides a buffer against the distorting effects of recombination at a single locus, MLSA emerged as a most useful taxonomic tool for grouping isolates into robust, major genetic lineages (Gevers *et al.*, 2005). Especially for depicting relationships within and between closely related species such as Bcc (Chapter 4) and *Achromobacter* species (Vandamme *et al.*, 2013b, 2016), MLSA has a taxonomic resolution superior to that of the traditional 16S rRNA gene sequence analysis.

Finally, the number of finished or draft bacterial genome sequences has risen exponentially and phylogenetic analysis is entering the new era of phylogenomics (Eisen & Fraser, 2003; Delsuc *et al.*, 2005). As an extension of the MLSA principle, phylogenetic trees with superior resolution can be inferred based on the multiple sequence alignment of all conserved protein-coding genes in the genomes under study (see section 2.1.3.2 for a detailed overview of tools for studying genomic taxonomy). Section 5.2 presented a phylogenomic study of 13

clusters of *B. glathei*-like bacteria and the resulting phylogenetic tree (Fig. 5.5) showed a superior resolution to the *gyrB* phylogenetic tree, together with very high bootstrap support values.

7.2.2.2 Evolution in methods for species delineation

DDH was introduced as a tool to approach whole-genome sequence derived information as close as possible in an era in which microbiologists had no access to DNA sequencing (Wayne *et al.*, 1987). Following the advent of the first sequencing technologies, MLSA was proposed to replace DDH as a tool for species delineation (Stackebrandt *et al.*, 2002; Gevers *et al.*, 2005). Vanlaere *et al.* (2009) compared the average concatenated allele sequence divergence within and between established Bcc species as delineated by DDH and demonstrated that a 3 % concatenated allele sequence divergence level could be used as a threshold value for species delineation within the Bcc, thus replacing the need to perform DDH experiments in this complex. This 3 % threshold value was evaluated and further substantiated in section 4.1 and this method for Bcc species delineation was subsequently applied for the formal classification of three novel Bcc species (Chapter 4). Similarly, a 2.1 % threshold value of concatenated allele sequence divergence was applied for species delineation within the genus *Achromobacter* (Vandamme *et al.*, 2013b, 2016). Although DDH was replaced successfully by MLSA approaches in specific genera, the latter method has the disadvantage that the cut-off for concatenated allele sequence divergence must be validated for each MLSA scheme separately and is thus not universally applicable.
In the present genomics era, several universal whole-genome sequence-based parameters have been proposed as a replacement for DDH and a detailed overview was presented in section 2.1.3.2 and Table 2.1. The GBDP method was introduced as an *in silico* replacement for DDH (dDDH) and calculates pairwise inter-genomic distances based on the results of a genome-wide homology search (Henz *et al.*, 2005; Meier-Kolthoff *et al.*, 2013). Importantly, GBDP is the only method that provides confidence intervals for species delineation and support values for the groupings in the resulting phylogenetic trees (Meier-Kolthoff *et al.*, 2013). While *B. cordobensis* sp. nov. (Section 5.1) was still delineated based on traditional DDH studies, thirteen novel *B. glathei*-like species (Section 5.2) were delineated based on the genomic relatedness among strains and the 70 % dDDH threshold value.

7.2.3 Time to revisit polyphasic taxonomy

In the present thesis we described three novel Bcc species and more than a dozen novel *B. glathei*-like species, but there is still an overwhelming number of putative novel *Burkholderia*

species awaiting formal classification. A phylogenetic analysis of the concatenated sequences of the seven loci of all STs in the PubMLST database (Fig. 2.2) revealed the presence of at least 12 putative novel species within the Bcc alone. Furthermore, when the huge taxonomic diversity of *Burkholderia* bacteria in the natural environment (Sections 3.2 and 3.3) will be complemented with culturomics approaches (Editorial Nature Reviews Microbiology, 2012; Lagier *et al.*, 2015b), the rate at which novel species may be discovered can only be expected to increase. Therefore, the genus *Burkholderia* represents a good example of the current failure of polyphasic taxonomy to truly report and name bacterial diversity (Vandamme & Peeters, 2014).

Although an evolution in taxonomic methods can be observed in which whole-genome sequence-based tools replace traditional tools for both species delineation and inferring phylogeny (Section 7.2.2), the current implementation of whole-genome sequence-based parameters into a "modern" polyphasic approach remains problematic for two reasons.

The first issue relates to the fact that although DDH is gradually being replaced by whole-genome sequence-based parameters, these new tools are being validated and calibrated by comparing them to the classifications that were built on the 70 % DDH cut-off for species delineation. However, this 70 % DDH cut-off was on its turn calibrated empirically to yield many of the phenotype-based species already recognized at the time of its inception (Cohan, 2002; Gevers *et al.*, 2005). As argued by Cohan (2002), the practice of calibrating each new molecular technique to yield the clusters that were previously determined by phenotypic criteria is caused by the lack of a theory-based species concept. However, in an era in which microbiologists have direct access to whole-genome sequence information, it makes no sense to try and mimic the results that were obtained through DDH, a method that was introduced in a pre-genomics era to approach whole-genome sequence derived information as close as possible (Wayne *et al.*, 1987; Vandamme & Peeters, 2014).

The second issue relates to the fact that taxonomists are being forced to perform an increasing number of, in particular biochemical and phenotypical, characterizations (Tindall *et al.*, 2010; Ramasamy *et al.*, 2014) that are of questionable taxonomic and practical relevance (Sutcliffe *et al.*, 2012). The polyphasic taxonomic approach was introduced as a consensus approach to provide a multifaceted view on the biology of a bacterium in an era in which microbiologists had no direct access to the genetic source code responsible for these characteristics. This search for differentiating biochemical reactions is a practice that originated from the era of traditional diagnostic microbiology. However, the biochemical diagnostic methods that were previously used in routine clinical microbiology have nowadays been replaced by MALDI-TOF MS (Cherkaoui *et al.*, 2010; Van Veen *et al.*, 2010; Carbonnelle *et al.*, 2011). The extended biochemical and phenotypical characterizations that were performed as part of the polyphasic

taxonomic studies in chapters 4 and 5 were thus performed to comply with the traditional requirements for the formal classification of bacteria that are imposed by the leading journals in the field of prokaryotic taxonomy. This cost- and time-inefficient practice of performing extended phenotypical characterizations presents a significant barrier to progress in the field of taxonomy (Sutcliffe *et al.*, 2012).

In conclusion, it is time to revisit the currently applied polyphasic taxonomic approach and revitalize the practice of prokaryotic systematics (Sutcliffe *et al.*, 2012, 2013; Vandamme & Peeters, 2014). Although it can be argued that this compliance of whole-genome sequence-based parameters with the existing classifications is a convenient way to replace DDH without the need to revise the classification schemes, we should not hold on to existing classification schemes if this hampers the introduction of new, more accurate genome-based methods in prokaryotic taxonomy. In this genomics era, a whole-genome sequence and a minimal description of phenotypic characteristics would provide a basic biological identity card that could be considered sufficient, cost-effective and appropriate for a species description (Vandamme & Peeters, 2014). In the future, a robust, whole-genome sequence-based species definition may help to classify the backlog of diversity awaiting formal description and the huge diversity that is yet to be discovered.

Bibliography

Adekambi, T., Colson, P., & Drancourt, M. (2003). rpoB-based identification of nonpigmented and late-pigmenting rapidly growing mycobacteria. *Journal of Clinical Microbiology*, *41*, 5699–5708. doi:10.1128/JCM.41.12.5699-5708.2003.

Adiba, S., Nizak, C., van Baalen, M., Denamur, E., & Depaulis, F. (2010). From grazing resistance to pathogenesis: the coincidental evolution of virulence factors. *PLoS ONE*, *5*, e11882. doi:10.1371/journal.pone.0011882.

Agnoli, K., Frauenknecht, C., Freitag, R., Schwager, S., Jenul, C., Vergunst, A., Carlier, A., & Eberl, L. (2014). The third replicon of members of the Burkholderia cepacia complex, plasmid pC3, plays a role in stress tolerance. *Applied and Environmental Microbiology*, *80*, 1340–1348. doi:10.1128/AEM.03330-13.

Agnoli, K., Schwager, S., Uehlinger, S., Vergunst, A., Viteri, D. F., Nguyen, D. T., Sokol, P. A., Carlier, A., & Eberl, L. (2012). Exposing the third chromosome of Burkholderia cepacia complex strains as a virulence plasmid. *Molecular Microbiology*, *83*, 362–378. doi:10.1111/j.1365-2958.2011.07937.x.

Agresti, A. (2002). Inference for contingency tables. In *Categorical Data Analysis* chapter 3. (pp. 70–114). John Wiley & Sons, Inc. (2nd ed.). doi:10.1002/0471249688.ch3.

Ahn, Y., Kim, J. M., Ahn, H., Lee, Y.-J., LiPuma, J. J., Hussong, D., & Cerniglia, C. E. (2014). Evaluation of liquid and solid culture media for the recovery and enrichment of Burkholderia cenocepacia from distilled water. *Journal of Industrial Microbiology and Biotechnology*, *41*, 1109–1118. doi:10.1007/s10295-014-1442-3.

Alby, K., Gilligan, P. H., & Miller, M. B. (2013). Comparison of matrix-assisted laser desorption ionization–time of flight (MALDI-TOF) mass spectrometry platforms for the identification of Gram-negative rods from patients with cystic fibrosis. *Journal of Clinical Microbiology*, *51*, 3852–3854. doi:10.1128/JCM.01618-13.

Andrade, G., Mihara, K. L., Linderman, R. G., & Bethlenfalvay, G. J. (1997). Bacteria from rhizosphere and hyphosphere soils of different arbuscular-mycorrhizal fungi. *Plant and Soil*, *192*, 71–79. doi:10.1023/A:1004249629643.

Angle, J. S., McGrath, S. P., & Chaney, R. L. (1991). New culture medium containing ionic concentrations of nutrients similar to concentrations found in the soil solution. *Applied and Environmental Microbiology*, *57*, 3674–3676.

Angus, A. A., Agapakis, C. M., Fong, S., Yerrapragada, S., Estrada-de los Santos, P., Yang, P., Song, N., Kano, S., Caballero-Mellado, J., de Faria, S. M., Dakora, F. D., Weinstock, G., & Hirsch, A. M. (2014). Plant-associated symbiotic Burkholderia species lack hallmark strategies required in mammalian pathogenesis. *PLoS ONE*, *9*, e83779. doi:10.1371/journal.pone.0083779.

Arbeli, Z., & Fuentes, C. L. (2007). Improved purification and PCR amplification of DNA from environmental samples. *FEMS Microbiology Letters*, *272*, 269–275. doi:10.1111/j.1574-6968.2007.00764.x.

Aris, R. M., Routh, J. C., LiPuma, J. J., Heath, D. G., & Gilligan, P. H. (2001). Lung transplantation for cystic fibrosis patients with Burkholderia cepacia complex. *American Journal of Respiratory and Critical Care Medicine*, *164*, 2102–2106. doi:10.1164/ajrccm.164.11.2107022.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., Fernandes, G. R., Tap, J., Bruls, T., Batto, J.-M., Bertalan, M., Borruel, N., Casellas, F., Fernandez, L., Gautier, L., Hansen, T., Hattori, M., Hayashi, T., Kleerebezem, M., Kurokawa, K., Leclerc, M., Levenez, F., Manichanh, C., Nielsen, H. B., Nielsen, T., Pons, N., Poulain, J., Qin, J., Sicheritz-Ponten, T., Tims, S., Torrents, D., Ugarte, E., Zoetendal, E. G., Wang, J., Guarner, F., Pedersen, O., de Vos, W. M., Brunak, S., Doré, J., Antolín, M., Artiguenave, F., Blottiere, H. M., Almeida, M., Brechot, C., Cara, C., Chervaux, C., Cultrone, A., Delorme, C., Denariaz, G., Dervyn, R., Foerstner, K. U., Friss, C., van de Guchte, M., Guedon, E., Haimet, F., Huber, W., van Hylckama-Vlieg, J., Jamet, A., Juste, C., Kaci, G., Knol, J., Lakhdari, O., Layec, S.,

Le Roux, K., Maguin, E., Mérieux, A., Melo Minardi, R., M'rini, C., Muller, J., Oozeer, R., Parkhill, J., Renault, P., Rescigno, M., Sanchez, N., Sunagawa, S., Torrejon, A., Turner, K., Vandemeulebrouck, G., Varela, E., Winogradsky, Y., Zeller, G., Weissenbach, J., Ehrlich, S. D., & Bork, P. (2011). Enterotypes of the human gut microbiome. *Nature*, *473*, 174–180. doi:10.1038/nature09944.

Auch, A. F., von Jan, M., Klenk, H.-P., & Goker, M. (2010). Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Standards in Genomic Sciences*, *2*, 117–134. doi:10.4056/sigs.531120.

Baek, I., Seo, B., Lee, I., Lee, K., Park, S.-C., Yi, H., & Chun, J. (2015). Burkholderia megalochromosomata sp. nov., isolated from grassland soil. *International Journal of Systematic and Evolutionary Microbiology*, *65*, 959–964. doi:10.1099/ijs.0.000046.

Bai, Y., Muller, D. B., Srinivas, G., Garrido-Oter, R., Potthoff, E., Rott, M., Dombrowski, N., Munch, P. C., Spaepen, S., Remus-Emsermann, M., Huttel, B., McHardy, A. C., Vorholt, J. A., & Schulze-Lefert, P. (2015). Functional overlap of the Arabidopsis leaf and root microbiota. *Nature*, *528*, 364–369. doi:10.1038/nature16192.

Balandreau, J., Viallard, V., Cournoyer, B., Coenye, T., Laevens, S., & Vandamme, P. (2001). Burkholderia cepacia genomovar III is a common plant-associated bacterium. *Applied and Environmental Microbiology*, *67*, 982–985. doi:10.1128/AEM.67.2.982-985.2001.

Baldwin, A., Mahenthiralingam, E., Drevinek, P., Pope, C., Waine, D. J., Henry, D. A., Speert, D. P., Carter, P., Vandamme, P., LiPuma, J. J., & Dowson, C. G. (2008). Elucidating global epidemiology of Burkholderia multivorans in cases of cystic fibrosis by multilocus sequence typing. *Journal of Clinical Microbiology*, *46*, 290–295. doi:10.1128/JCM.01818-07.

Baldwin, A., Mahenthiralingam, E., Drevinek, P., Vandamme, P., Govan, J. R., Waine, D. J., Chiarini, L., Dalmastri, C., Henry, D. A., Speert, D. P., Honeybourne, D., Maiden, M. C. J., & Dowson, C. G. (2007). Environmental Burkholderia cepacia complex isolates in human infections. *Emerging Infectious Diseases*, *13*, 458–461. doi:10.3201/eid1303.060403.

Baldwin, A., Mahenthiralingam, E., Thickett, K. M., Honeybourne, D., Maiden, M. C. J., 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. *Journal of Clinical Microbiology*, *43*, 4665–4673. doi:10.1128/JCM.43.9.4665-4673.2005.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, *19*, 455–477. doi:10.1089/cmb.2012.0021.

Bar-On, O., Mussaffi, H., Mei-Zahav, M., Prais, D., Steuer, G., Stafler, P., Hananya, S., & Blau, H. (2015). Increasing nontuberculous mycobacteria infection in cystic fibrosis. *Journal of Cystic Fibrosis*, *14*, 53–62. doi:10.1016/j.jcf.2014.05.008.

Barberan, A., Ramirez, K. S., Leff, J. W., Bradford, M. A., Wall, D. H., & Fierer, N. (2014). Why are some microbes more ubiquitous than others? Predicting the habitat breadth of soil bacteria. *Ecology Letters*, *17*, 794–802. doi:10.1111/ele.12282.

Bartell, J. A., Yen, P., Varga, J. J., Goldberg, J. B., & Papin, J. A. (2014). Comparative metabolic systems analysis of pathogenic Burkholderia. *Journal of Bacteriology*, *196*, 210–226. doi:10.1128/JB.00997-13.

Baym, M., Kryazhimskiy, S., Lieberman, T. D., Chung, H., Desai, M. M., & Kishony, R. (2015). Inexpensive multiplexed library preparation for megabase-sized genomes. *PLoS ONE*, *10*. doi:10.1371/journal.pone.0128036.

Beckman, W., & Lessie, T. G. (1979). Response of Pseudomonas cepacia to beta-lactam antibiotics: utilization of penicillin G as the carbon source. *Journal of Bacteriology*, *140*, 1126–1128.

Bentley, S. D., & Parkhill, J. (2004). Comparative genomic structure of prokaryotes. *Annual Review of Genetics*, *38*, 771–791. doi:10.1146/annurev.genet.38.072902.094318.

Berg, G., Eberl, L., & Hartmann, A. (2005). The rhizosphere as a reservoir for opportunistic human pathogenic bacteria. *Environmental Microbiology*, *7*, 1673–1685. doi:10.1111/j.1462-2920.2005.00891.x.

Berg, G., Erlacher, A., Smalla, K., & Krause, R. (2014). Vegetable microbiomes: is there a connection among opportunistic infections, human health and our 'gut feeling'? *Microbial Biotechnology*, *7*, 487–495. doi:10.1111/1751-7915.12159.

Bergmark, L., Poulsen, P. H. B., Al-Soud, W. A., Norman, A., Hansen, L. H., & Sorensen, S. J. (2012). Assessment of the specificity of Burkholderia and Pseudomonas qPCR assays for detection of these genera in soil using 454 pyrosequencing. *FEMS Microbiology Letters*, *333*, 77–84. doi:10.1111/j.1574-6968.2012.02601.x.

Bevivino, A., Costa, B., Cantale, C., Cesarini, S., Chiarini, L., Tabacchioni, S., Caballero-Mellado, J., & Dalmastri, C. (2011). Genetic relationships among Italian and Mexican maize-rhizosphere Burkholderia cepacia complex (BCC) populations belonging to Burkholderia cenocepacia IIIB and BCC6 group. *BMC Microbiology*, *11*. doi:10.1186/1471-2180-11-228.

Bevivino, A., Dalmastri, C., Tabacchioni, S., Chiarini, L., Belli, M. L., Piana, S., Materazzo, A., Vandamme, P., & Manno, G. (2002). Burkholderia cepacia complex bacteria from clinical and environmental sources in Italy: genomovar status and distribution of traits related to virulence and transmissibility. *Journal of Clinical Microbiology*, *40*, 846–851. doi:10.1128/JCM.40.3.846-851.2002.

Bevivino, A., Tabacchioni, S., Chiarini, L., Carusi, M. V., Del Gallo, M., & Visca, P. (1994). Phenotypic comparison between rhizosphere and clinical isolates of Burkholderia cepacia. *Microbiology*, *140*, 1069–1077.

Biddick, R., Spilker, T., Martin, A., & LiPuma, J. J. (2003). Evidence of transmission of Burkholderia cepacia, Burkholderia multivorans and Burkholderia dolosa among persons with cystic fibrosis. *FEMS Microbiology Letters*, *228*, 57–62. doi:10.1016/S0378-1097(03)00724-9.

Blackburn, L., Brownlee, K., Conway, S., & Denton, M. (2004). 'Cepacia syndrome' with Burkholderia multivorans, 9 years after initial colonization. *Journal of Cystic Fibrosis*, *3*, 133–134. doi:10.1016/j.jcf.2004.03.007.

Bogas, V., Carvalho, M., Anjos, M., Pinheiro, M., & Corte-Real, F. (2009). Genetic identification of degraded DNA samples buried in different types of soil. *Forensic Science International: Genetics Supplement Series 2*, *2*, 169–171. doi:10.1016/j.fsigss.2009.08.031.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*, 2114–2120. doi:10.1093/bioinformatics/btu170.

Bonfield, J. K., & Whitwham, A. (2010). Gap5 - editing the billion fragment sequence assembly. *Bioinformatics*, *26*, 1699–1703. doi:10.1093/bioinformatics/btq268.

Bopp, L. H. (1986). Degradation of highly chlorinated PCBs by Pseudomonas strain LB400. *Journal of Industrial Microbiology*, *1*, 23–29. doi:10.1007/BF01569413.

Boucias, D. G., Garcia-Maruniak, A., Cherry, R., Lu, H., Maruniak, J. E., & Lietze, V.-U. (2012). Detection and characterization of bacterial symbionts in the Heteropteran, Blissus insularis. *FEMS Microbiology Ecology*, *82*, 629–641. doi:10.1111/j.1574-6941.2012.01433.x.

Brisse, S., Cordevant, C., Vandamme, P., Bidet, P., Loukil, C., Chabanon, G., Lange, M., & Bingen, E. (2004). Species distribution and ribotype diversity of Burkholderia cepacia complex isolates from French patients with cystic fibrosis. *Journal of Clinical Microbiology*, *42*, 4824–4827. doi:10.1128/JCM.42.10.4824-4827.2004.

Brisse, S., Verduin, C. M., Milatovic, D., Fluit, A., Verhoef, J., Laevens, S., Vandamme, P., Tummler, B., Verbrugh, H. A., & van Belkum, A. (2000). Distinguishing species of the Burkholderia cepacia complex and Burkholderia gladioli by automated ribotyping. *Journal of Clinical Microbiology*, *38*, 1876–1884.

Burbage, D. A., & Sasser, M. (1982). A medium selective for Pseudomonas cepacia. *Phytopathology*, *76*, 706.

Burge, C., Campbell, A. M., & Karlin, S. (1992). Over- and under-representation of short oligonucleotides in DNA sequences. *Proceedings of the National Academy of Sciences*, *89*, 1358–1362. doi:10.1073/pnas.89.4.1358.

Burkholder, W. (1950). Sour skin, a bacterial rot of onion bulbs. *Phytopathology*, *40*, 115–117.

Burns, J. L., Jonas, M., Chi, E. Y., Clark, D. K., Berger, A., & Griffith, A. (1996). Invasion of respiratory epithelial cells by (Pseudomonas) cepacia. *Microbiology*, *64*, 4054–4059.

Butler, S. L., Doherty, C. J., Hughes, J. E., Nelson, J. W., & Govan, J. R. W. (1995). Burkholderia cepacia and cystic fibrosis - do natural environments present a potential hazard? *Journal of Clinical Microbiology*, *33*, 1001–1004.

Caballero-Mellado, J., Martinez-Aguilar, L., Paredes-Valdez, G., & Estrada-de los Santos, P. (2004). Burkholderia unamae sp. nov., an N2-fixing rhizospheric and endophytic species. *International Journal of Systematic and Evolutionary Microbiology*, *54*, 1165–1172. doi:10.1099/ijs.0.02951-0.

Caballero-Mellado, J., Onofre-Lemus, J., Estrada-de los Santos, P., & Martinez-Aguilar, L. (2007). The tomato rhizosphere, an environment rich in nitrogen-fixing Burkholderia species with capabilities of interest for agriculture and bioremediation. *Applied and Environmental Microbiology*, *73*, 5308–5319. doi:10.1128/AEM.00324-07.

Capella-Gutierrez, S., Silla-Martinez, J. M., & Gabaldon, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*, 1972–1973. doi:10.1093/bioinformatics/btp348.

Carbonnelle, E., Mesquita, C., Bille, E., Day, N., Dauphin, B., Beretti, J.-L., Ferroni, A., Gutmann, L., & Nassif, X. (2011). MALDI-TOF mass spectrometry tools for bacterial identification in clinical microbiology laboratory. *Clinical Biochemistry*, *44*, 104–109. doi:10.1016/j.clinbiochem.2010.06.017.

Carlier, A., Fehr, L., Pinto-Carbo, M., Schaberle, T., Reher, R., Dessein, S., Konig, G., & Eberl, L. (2016). The genome analysis of Candidatus Burkholderia crenata reveals that secondary metabolism may be a key function of the Ardisia crenata leaf nodule symbiosis. *Environmental Microbiology*, *18*, 2507–2522. doi:10.1111/1462-2920.13184.

Carlier, A. L., & Eberl, L. (2012). The eroded genome of a Psychotria leaf symbiont: hypotheses about lifestyle and interactions with its plant host. *Environmental Microbiology*, *14*, 2757–2769. doi:10.1111/j.1462-2920.2012.02763.x.

Cases, I., de Lorenzo, V., & Ouzounis, C. A. (2003). Transcription regulation and environmental adaptation in bacteria. *Trends in Microbiology*, *11*, 248–253. doi:10.1016/S0966-842X(03)00103-3.

Castanheira, N., Dourado, A. C., Kruz, S., Alves, P. I. L., Delgado-Rodriguez, A. I., Pais, I., Semedo, J., Scotti-Campos, P., Sanchez, C., Borges, N., Carvalho, G., Barreto Crespo, M. T., & Fareleira, P. (2016). Plant growth-promoting Burkholderia species isolated from annual ryegrass in Portuguese soils. *Journal of Applied Microbiology*, *120*, 724–739. doi:10.1111/jam.13025.

Castorena, G., Mugica, V., Le Borgne, S., Acuna, M. E., Bustos-Jaimes, I., & Aburto, J. (2006). Carbazole biodegradation in gas oil/water biphasic media by a new isolated bacterium Burkholderia sp. strain IMP5GC. *Journal of Applied Microbiology*, *100*, 739–745. doi:10.1111/j.1365-2672.2005.02799.x.

Cesarini, S., Bevivino, A., Tabacchioni, S., Chiarini, L., & Dalmastri, C. (2009). RecA gene sequence and

multilocus sequence typing for species-level resolution of Burkholderia cepacia complex isolates. *Letters in Applied Microbiology*, *49*, 580–588. doi:10.1111/j.1472-765X.2009.02709.x.

Chain, P. S. G., Denef, V. J., Konstantinidis, K. T., Vergez, L. M., Agullo, L., Reyes, V. L., Hauser, L., Cordova, M., Gomez, L., Gonzalez, M., Land, M., Lao, V., Larimer, F., LiPuma, J. J., Mahenthiralingam, E., Malfatti, S. A., Marx, C. J., Parnell, J. J., Ramette, A., Richardson, P., Seeger, M., Smith, D., Spilker, T., Sul, W. J., Tsoi, T. V., Ulrich, L. E., Zhulin, I. B., & Tiedje, J. M. (2006). Burkholderia xenovorans LB400 harbors a multi-replicon, 9.73-Mbp genome shaped for versatility. *Proceedings of the National Academy of Sciences*, *103*, 15280–15287. doi:10.1073/pnas.0606924103.

Chavatte, N., Bare, J., Lambrecht, E., Van Damme, I., Vaerewijck, M., Sabbe, K., & Houf, K. (2014). Co-occurrence of free-living protozoa and foodborne pathogens on dishcloths: implications for food safety. *International Journal of Food Microbiology*, *191*, 89–96. doi:10.1016/j.ijfoodmicro.2014.08.030.

Chen, J. S., Witzmann, K. A., Spilker, T., Fink, R. J., & LiPuma, J. J. (2001). Endemicity and inter-city spread of Burkholderia cepacia genomovar III in cystic fibrosis. *The Journal of Pediatrics*, *139*, 643–649. doi:10.1067/mpd.2001.118430.

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. *International Journal of Systematic and Evolutionary Microbiology*, *57*, 1055–1059. doi:10.1099/ijs.0.64873-0.

Chen, W.-M., de Faria, S. M., Straliotto, R., Pitard, R. M., Simoes-Araujo, 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. *Applied and Environmental Microbiology*, *71*, 7461–7471. doi:10.1128/AEM.71.11.7461-7471.2005.

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. *International Journal of Systematic and Evolutionary Microbiology*, *56*, 1847–1851. doi:10.1099/ijs.0.64325-0.

Cherkaoui, A., Hibbs, J., Emonet, S., Tangomo, M., Girard, M., Francois, P., & Schrenzel, J. (2010). Comparison of two matrix-assisted laser desorption ionization-time of flight mass spectrometry methods with conventional phenotypic identification for routine identification of bacteria to the species level. *Journal of Clinical Microbiology*, *48*, 1169–1175. doi:10.1128/JCM.01881-09.

Ciccarelli, F. D., Doerks, T., von Mering, C., Creevey, C. J., Snel, B., & Bork, P. (2006). Toward automatic reconstruction of a highly resolved tree of life. *Science*, *311*, 1283–1287. doi:10.1126/science.1123061.

Clark, A. E., Kaleta, E. J., Arora, A., & Wolk, D. M. (2013). Matrix-assisted laser desorption ionization-time of flight mass spectrometry: a fundamental shift in the routine practice of clinical microbiology. *Clinical Microbiology Reviews*, *26*, 547–603. doi:10.1128/CMR.00072-12.

Clooney, A. G., Fouhy, F., Sleator, R. D., O' Driscoll, A., Stanton, C., Cotter, P. D., & Claesson, M. J. (2016). Comparing apples and oranges?: next generation sequencing and its impact on microbiome analysis. *PLoS ONE*, *11*, e0148028. doi:10.1371/journal.pone.0148028.

Coenye, T., Falsen, E., Vancanneyt, M., Hoste, B., Govan, J. R. W., Kersters, K., & Vandamme, P. (1999a). Classification of Alcaligenes faecalis-like isolates from the environment and human clinical samples as Ralstonia gilardii sp. nov. *International Journal of Systematic Bacteriology*, *49*, 405–413.

Coenye, T., Goris, J., Spilker, T., Vandamme, P., & LiPuma, J. J. (2002a). Characterization of unusual bacteria isolated from respiratory secretions of cystic fibrosis patients and description of Inquilinus limosus gen. nov.,

sp. nov. *Journal of Clinical Microbiology*, *40*, 2062–2069. doi:10.1128/JCM.40.6.2062-2069.2002.

Coenye, T., Henry, D., Speert, D. P., & Vandamme, P. (2004). Burkholderia phenoliruptrix sp. nov., to accommodate the 2,4,5-trichlorophenoxyacetic acid and halophenol-degrading strain AC1100. *Systematic and Applied Microbiology*, *27*, 623–627.

Coenye, T., Laevens, S., Gillis, M., & Vandamme, P. (2001a). Genotypic and chemotaxonomic evidence for the reclassification of Pseudomonas woodsii (Smith 1911) Stevens 1925 as Burkholderia andropogonis (Smith 1911) Gillis et al. 1995. *International Journal of Systematic and Evolutionary Microbiology*, *51*, 183–185.

Coenye, T., Laevens, S., Willems, A., Ohlen, M., Hannant, W., Govan, J. R. W., Gillis, M., Falsen, E., & Vandamme, P. (2001b). Burkholderia fungorum sp. nov. and Burkholderia caledonica sp. nov., two new species isolated from the environment, animals and human clinical samples. *International Journal of Systematic and Evolutionary Microbiology*, *51*, 1099–1107.

Coenye, T., & LiPuma, J. J. (2002). Multilocus restriction typing: a novel tool for studying global epidemiology of Burkholderia cepacia complex infection in cystic fibrosis. *Journal of Infectious Diseases*, *185*, 1454–1462. doi:10.1086/340279.

Coenye, T., LiPuma, J. J., Henry, D., Hoste, B., Vandemeulebroecke, K., Gillis, M., Speert, D. P., & Vandamme, P. (2001c). Burkholderia cepacia genomovar VI, a new member of the Burkholderia cepacia complex isolated from cystic fibrosis patients. *International Journal of Systematic and Evolutionary Microbiology*, *51*, 271–279.

Coenye, T., Schouls, L. M., Govan, J. R. W., Kersters, K., & Vandamme, P. (1999b). Identification of Burkholderia species and genomovars from cystic fibrosis patients by AFLP fingerprinting. *International Journal of Systematic Bacteriology*, *49*, 1657–1666.

Coenye, T., Spilker, T., Martin, A., & LiPuma, J. J. (2002b). Comparative assessment of genotyping methods for epidemiologic study of Burkholderia cepacia genomovar III. *Journal of Clinical Microbiology*, *40*, 3300–3307. doi:10.1128/JCM.40.9.3300-3307.2002.

Coenye, T., & Vandamme, P. (2003). Diversity and significance of Burkholderia species occupying diverse ecological niches. *Environmental Microbiology*, *5*, 719–729. doi:10.1046/j.1462-2920.2003.00471.x.

Coenye, T., & Vandamme, P. (2004). Use of the genomic signature in bacterial classification and identification. *Systematic and Applied Microbiology*, *27*, 175–185. doi:10.1078/072320204322881790.

Coenye, T., Vandamme, P., Govan, J. R. W., & LiPuma, J. J. (2001d). Taxonomy and identification of the Burkholderia cepacia complex. *Journal of Clinical Microbiology*, *39*, 3427–3436. doi:10.1128/JCM.39.10.3427-3436.2001.

Coenye, T., Vandamme, P., LiPuma, J. J., Govan, J. R. W., & Mahenthiralingam, E. (2003). Updated version of the Burkholderia cepacia complex experimental strain panel. *Journal of Clinical Microbiology*, *41*, 2797–2798. doi:10.1128/JCM.41.6.2797-2798.2003.

Cohan, F. M. (2002). What are bacterial species? *Annual Review of Microbiology*, *56*, 457–487. doi:10.1146/annurev.micro.56.012302.160634.

Colston, S. M., Fullmer, M. S., Beka, L., Lamy, B., Gogarten, J. P., & Graf, J. (2014). Bioinformatic genome comparisons for taxonomic and phylogenetic assignments using Aeromonas as a test case. *mBio*, *5*, e02136–14. doi:10.1128/mBio.02136-14.

Compant, S., Nowak, J., Coenye, T., Clement, C., & Ait Barka, E. (2008). Diversity and occurrence of Burkholderia spp. in the natural environment. *FEMS Microbiology Reviews*, *32*, 607–626. doi:10.1111/j.1574-6976.2008.00113.x.

Cooper, V. S., Vohr, S. H., Wrocklage, S. C., & Hatcher, P. J. (2010). Why genes evolve faster on secondary chromosomes in bacteria. *PLoS Computational Biology*, *6*, e1000732. doi:10.1371/journal.

pcbi.1000732.

Cordova-Kreylos, A. L., Fernandez, L. E., Koivunen, M., Yang, A., Flor-Weiler, L., & Marrone, P. G. (2013). Isolation and characterization of Burkholderia rinojensis sp. nov., a non-Burkholderia cepacia complex soil bacterium with insecticidal and miticidal activities. *Applied and Environmental Microbiology*, *79*, 7669–7678. doi:10.1128/AEM.02365-13.

Coutinho, C. P., Barreto, C., Pereira, L., Lito, L., Cristino, J. M., & Sa-Correia, I. (2015). Incidence of Burkholderia contaminans at a cystic fibrosis centre with an unusually high representation of Burkholderia cepacia during 15 years of epidemiological surveillance. *Journal of Medical Microbiology*, *64*, 927–935. doi:10.1099/jmm.0.000094.

Currie, B. J., Ward, L., & Cheng, A. C. (2010). The epidemiology and clinical spectrum of melioidosis: 540 cases from the 20 year Darwin prospective study. *PLoS Neglected Tropical Diseases*, *4*. doi:10.1371/journal.pntd.0000900.

Darling, A. E., Mau, B., & Perna, N. T. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE*, *5*, e11147. doi:10.1371/journal.pone.0011147.

Daubin, V., Gouy, M., & Perriere, G. (2001). Bacterial molecular phylogeny using supertree approach. *Genome Informatics*, *164*, 155–164.

De Boeck, K., Malfroot, A., Van Schil, L., Lebecque, P., Knoop, C., Govan, J. R. W., Doherty, C., Laevens, S., & Vandamme, P. (2004). Epidemiology of Burkholderia cepacia complex colonisation in cystic fibrosis patients. *European Respiratory Journal*, *23*, 851–856. doi:10.1183/09031936.04.00118804.

De Meyer, S. E., Cnockaert, M., Ardley, J. K., Trengove, R. D., Garau, G., Howieson, J. G., & Vandamme, P. (2013). Burkholderia rhynchosiae sp. nov., isolated from Rhynchosia ferulifolia root nodules. *International Journal of Systematic and Evolutionary Microbiology*, *63*, 3944–3949. doi:10.1099/ijs.0.048751-0.

De Smet, B., Mayo, M., Peeters, C., Zlosnik, J. E. A., Spilker, T., Hird, T. J., LiPuma, J. J., Kidd, T. J., Kaestli, M., Ginther, J. L., Wagner, D. M., Keim, P., Bell, S. C., Jacobs, J. A., Currie, B. J., & Vandamme, P. (2015a). Burkholderia stagnalis sp. nov. and Burkholderia territorii sp. nov., two novel Burkholderia cepacia complex species from environmental and human sources. *International Journal of Systematic and Evolutionary Microbiology*, *65*, 2265–2271. doi:10.1099/ijs.0.000251.

De Smet, B., Sarovich, D. S., Price, E. P., Mayo, M., Theobald, V., Kham, C., Heng, S., Thong, P., Holden, M. T. G., Parkhill, J., Peacock, S. J., Spratt, B. G., Jacobs, J. A., Vandamme, P., & Currie, B. J. (2015b). Whole-genome sequencing confirms that Burkholderia pseudomallei multilocus sequence types common to both Cambodia and Australia are due to homoplasy. *Journal of Clinical Microbiology*, *53*, 323–326. doi:10.1128/JCM.02574-14.

De Smet, B., Veng, C., Kruy, L., Kham, C., van Griensven, J., Peeters, C., Ieng, S., Phe, T., Vlieghe, E., Vandamme, P., & Jacobs, J. (2013). Outbreak of Burkholderia cepacia bloodstream infections traced to the use of Ringer lactate solution as multiple-dose vial for catheter flushing, Phnom Penh, Cambodia. *Clinical Microbiology and Infection*, *19*, 832–837. doi:10.1111/1469-0691.12047.

De Soyza, A., McDowell, A., Archer, L., Dark, J. H., Elborn, S. J., Mahenthiralingam, E., Gould, K., & Corris, P. A. (2001). Burkholderia cepacia complex genomovars and pulmonary transplantation outcomes in patients with cystic fibrosis. *The Lancet*, *358*, 1780–1781.

Degand, N., Carbonnelle, E., Dauphin, B., Beretti, J.-L., Le Bourgeois, M., Sermet-Gaudelus, I., Segonds, C., Berche, P., Nassif, X., & Ferroni, A. (2008). Matrix-assisted laser desorption ionization-time of flight mass spectrometry for identification of nonfermenting Gram-negative bacilli isolated from cystic fibrosis patients. *Journal of Clinical Microbiology*, *46*, 3361–3367. doi:10.1128/JCM.00569-08.

Delafont, V., Brouke, A., Bouchon, D., Moulin, L., & Hechard, Y. (2013). Microbiome of free-living amoebae

isolated from drinking water. *Water Research*, *47*, 6958–6965. doi:10.1016/j.watres.2013.07.047.

Delmont, T. O., Eren, A. M., Maccario, L., Prestat, E., Esen, z. C., Pelletier, E., Le Paslier, D., Simonet, P., & Vogel, T. M. (2015). Reconstructing rare soil microbial genomes using in situ enrichments and metagenomics. *Frontiers in Microbiology*, *6*. doi:10.3389/fmicb.2015.00358.

Deloger, M., El Karoui, M., & Petit, M.-A. (2009). A genomic distance based on MUM indicates discontinuity between most bacterial species and genera. *Journal of Bacteriology*, *191*, 91–99. doi:10.1128/JB.01202-08.

Delsuc, F., Brinkmann, H., & Philippe, H. (2005). Phylogenomics and the reconstruction of the tree of life. *Nature Reviews Genetics*, *6*, 361–375. doi:10.1038/nrg1603.

Depoorter, E., Bull, M. J., Peeters, C., Coenye, T., Vandamme, P., & Mahenthiralingam, E. (2016). Burkholderia: an update on taxonomy and biotechnological potential as antibiotic producers. *Applied Microbiology and Biotechnology*, *100*, 5215–5229. doi:10.1007/s00253-016-7520-x.

Deris, Z. Z., Van Rostenberghe, H., Habsah, H., Noraida, R., Tan, G. C., Chan, Y. Y., Rosliza, A. R., & Ravichandran, M. (2010). First isolation of Burkholderia tropica from a neonatal patient successfully treated with imipenem. *International Journal of Infectious Diseases*, *14*, e73–e74. doi:10.1016/j.ijid.2009.03.005.

Desai, A. P., Stanley, T., Atuan, M., McKey, J., Lipuma, J. J., Rogers, B., & Jerris, R. (2012). Use of matrix assisted laser desorption ionisation-time of flight mass spectrometry in a paediatric clinical laboratory for identification of bacteria commonly isolated from cystic fibrosis patients. *Journal of Clinical Pathology*, *65*, 835–838. doi:10.1136/jclinpath-2012-200772.

Didelot, X., Bowden, R., Wilson, D. J., Peto, T. E. a., & Crook, D. W. (2012). Transforming clinical microbiology with bacterial genome sequencing. *Nature Reviews Genetics*, *13*, 601–612. doi:10.1038/nrg3226.

DiSalvo, S., Haselkorn, T. S., Bashir, U., Jimenez, D., Brock, D. A., Queller, D. C., & Strassmann, J. E. (2015). Burkholderia bacteria infectiously induce the proto-farming symbiosis of Dictyostelium amoebae and food bacteria. *Proceedings of the National Academy of Sciences*, *112*, E5029–E5037. doi:10.1073/pnas.1511878112.

Dobritsa, A. P., & Samadpour, M. (2016). Transfer of eleven Burkholderia species to the genus Paraburkholderia and proposal of Caballeronia gen. nov., a new genus to accommodate twelve species of Burkholderia and Paraburkholderia. *International Journal of Systematic and Evolutionary Microbiology*, *in press*. doi:10.1099/ijsem.0.001065.

Dolezel, J., Bartos, J., Voglmayr, H., & Greilhuber, J. (2003). Nuclear DNA content and genome size of trout and human. *Cytometry*, *51A*, 127–128. doi:10.1002/cyto.a.10013.

Draghi, W. O., Peeters, C., Cnockaert, M., Snauwaert, C., Wall, L. G., Zorreguieta, A., & Vandamme, P. (2014). Burkholderia cordobensis sp. nov., from agricultural soils. *International Journal of Systematic and Evolutionary Microbiology*, *64*, 2003–2008. doi:10.1099/ijs.0.059667-0.

Drevinek, P., Hrbackova, H., Cinek, O., Bartosova, J., Nyc, O., Nemec, A., & Pohunek, P. (2002). Direct PCR detection of Burkholderia cepacia complex and identification of its genomovars by using sputum as source of DNA. *Journal of Clinical Microbiology*, *40*, 3485–3488. doi:10.1128/JCM.40.9.3485-3488.2002.

Drevinek, P., & Mahenthiralingam, E. (2010). Burkholderia cenocepacia in cystic fibrosis: epidemiology and molecular mechanisms of virulence. *Clinical Microbiology and Infection*, *16*, 821–830. doi:10.1111/j.1469-0691.2010.03237.x.

Drevinek, P., Vosahlikova, S., Dedeckova, K., Cinek, O., & Mahenthiralingam, E. (2010). Direct culture-independent strain typing of Burkholderia cepacia complex in sputum samples from patients with cystic fibrosis. *Journal of Clinical Microbiology*, *48*, 1888–1891. doi:10.1128/JCM.02359-09.

Eberl, L., & Vandamme, P. (2016). Members of the genus Burkholderia: good and bad guys. *F1000Research*, *5*, 1007. doi:10.12688/f1000research.8221.1.

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, *5*, 113. doi:10.1186/1471-2105-5-113.

Editorial Nature Reviews Microbiology (2012). The cultural revolution. *Nature Reviews Microbiology*, *11*, 1–1. doi:10.1038/nrmicro2948.

Eisen, J. A., & Fraser, C. M. (2003). Phylogenomics: intersection of evolution and genomics. *Science*, *300*, 1706–1707. doi:10.1126/science.1086292.

Environmental Protection Agency (2003). Burkholderia cepacia complex: significant new use rule. *Federal Register*, *68*.

Epstein, S. (2013). The phenomenon of microbial uncultivability. *Current Opinion in Microbiology*, *16*, 636–642. doi:10.1016/j.mib.2013.08.003.

Esther, C. R., Hoberman, S., Fine, J., Allen, S., Culbreath, K., Rodino, K., Kerr, A., & Gilligan, P. (2011). Detection of rapidly growing mycobacteria in routine cultures of samples from patients with cystic fibrosis. *Journal of Clinical Microbiology*, *49*, 1421–1425. doi:10.1128/JCM.02379-10.

Estrada-de los Santos, P., Rojas-Rojas, F. U., Tapia-Garcia, E. Y., Vasquez-Murrieta, M. S., & Hirsch, A. M. (2015). To split or not to split: an opinion on dividing the genus Burkholderia. *Annals of Microbiology*, (pp. 1–12). doi:10.1007/s13213-015-1183-1.

Estrada-de los Santos, P., Vinuesa, P., Martinez-Aguilar, L., Hirsch, A. M., & Caballero-Mellado, J. (2013). Phylogenetic analysis of Burkholderia species by multilocus sequence analysis. *Current Microbiology*, *67*, 51–60. doi:10.1007/s00284-013-0330-9.

Euzeby, J. P. (1997). List of bacterial names with standing in nomenclature: a folder available on the internet. *International Journal of Systematic and Evolutionary Microbiology*, *47*, 590–592. doi:10.1099/00207713-47-2-590.

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. *International Journal of Systematic Bacteriology*, *39*, 224–229. doi:10.1099/00207713-39-3-224.

Fang, Y., Xie, G.-L., Lou, M.-M., Li, B., & Muhammad, I. (2011). Diversity analysis of Burkholderia cepacia complex in the water bodies of West Lake, Hangzhou, China. *Journal of Microbiology*, *49*, 309–314. doi:10.1007/s12275-011-0267-2.

Farrell, P. M. (2008). The prevalence of cystic fibrosis in the European Union. *Journal of Cystic Fibrosis*, *7*, 450–453. doi:10.1016/j.jcf.2008.03.007.

Felsenstein, J. (2005). PHYLIP (Phylogeny Inference Package) version 3.6.

Fernandez-Olmos, A., Garcia-Castillo, M., Morosini, M.-I., Lamas, A., Maiz, L., & Canton, R. (2012). MALDI-TOF MS improves routine identification of non-fermenting Gram negative isolates from cystic fibrosis patients. *Journal of Cystic Fibrosis*, *11*, 59–62. doi:10.1016/j.jcf.2011.09.001.

Fierer, N., Jackson, J. A., Vilgalys, R., & Jackson, R. B. (2005). Assessment of soil microbial community structure by use of taxon-specific quantitative PCR assays. *Applied and Environmental Microbiology*, *71*, 4117–4120. doi:10.1128/AEM.71.7.4117-4120.2005.

Fierer, N., & Jackson, R. B. (2006). The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences*, *103*, 626–631. doi:10.1073/pnas.0507535103.

Figuerola, E. L. M., Guerrero, L. D., Rosa, S. M., Simonetti, L., Duval, M. E., Galantini, J. A., Bedano, J. C., Wall, L. G., & Erijman, L. (2012). Bacterial indicator of agricultural management for soil under no-till crop

production. *PLoS ONE*, *7*. doi:10.1371/journal.pone.0051075.

Fiore, A., Laevens, S., Bevivino, A., Dalmastri, C., Tabacchioni, S., Vandamme, P., & Chiarini, L. (2001). Burkholderia cepacia complex: distribution of genomovars among isolates from the maize rhizosphere in Italy. *Environmental Microbiology*, *3*, 137–143. doi:10.1046/j.1462-2920.2001.00175.x.

Fischer, S., Brunk, B. P., Chen, F., Gao, X., Harb, O. S., Iodice, J. B., Shanmugam, D., Roos, D. S., & Stoeckert, C. J. (2011). Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. In *Current Protocols in Bioinformatics* (pp. 6.12.1–6.12.19). Hoboken, NJ, USA: John Wiley & Sons, Inc. volume 35. doi:10.1002/0471250953.bi0612s35.

Flight, W. G., Smith, A., Paisey, C., Marchesi, J. R., Bull, M. J., Norville, P. J., Mutton, K. J., Webb, A. K., Bright-Thomas, R. J., Jones, A. M., & Mahenthiralingam, E. (2015). Rapid detection of emerging pathogens and loss of microbial diversity associated with severe lung disease in cystic fibrosis. *Journal of Clinical Microbiology*, *53*, 2022–2029. doi:10.1128/JCM.00432-15.

Fox, G. E., Wisotzkey, J. D., & Jurtshuk, P. (1992). How close is close - 16S rRNA sequence identity may not be sufficient to guarantee species identity. *International Journal of Systematic Bacteriology*, *42*, 166–170.

France, M. W., Dodd, M. E., Govan, J. R., Doherty, C. J., Webb, A., & Jones, A. M. (2008). The changing epidemiology of Burkholderia species infection at an adult cystic fibrosis centre. *Journal of Cystic Fibrosis*, *7*, 368–372. doi:10.1016/j.jcf.2008.01.002.

Furuya, N., Masunaga, T., Khan, A. A., Iiyama, K., Matsumoto, M., & Matsuyama, N. (2000). Bacterial wilt of Russell prairie gentian caused by Burkholderia caryophylli. *Journal of General Plant Pathology*, *66*, 316–322.

Galperin, M. Y., Makarova, K. S., Wolf, Y. I., & Koonin, E. V. (2015). Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Research*, *43*, D261–D269. doi:10.1093/nar/gku1223.

Galyov, E. E., Brett, P. J., & DeShazer, D. (2010). Molecular insights into Burkholderia pseudomallei and Burkholderia mallei pathogenesis. *Annual Review of Microbiology*, *64*, 495–517. doi:10.1146/annurev.micro.112408.134030.

Garcia-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L. M., Gotz, S., Tarazona, S., Dopazo, J., Meyer, T. F., & Conesa, A. (2012). Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics*, *28*, 2678–2679. doi:10.1093/bioinformatics/bts503.

Gerrits, G. P., Klaassen, C., Coenye, T., Vandamme, P., & Meis, J. F. (2005). Burkholderia fungorum septicemia. *Emerging Infectious Diseases*, *11*, 1115–1117.

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). Re-evaluating prokaryotic species. *Nature Reviews Microbiology*, *3*, 733–739. doi:10.1038/nrmicro1236.

Ghyselinck, J., Van Hoorde, K., Hoste, B., Heylen, K., & De Vos, P. (2011). Evaluation of MALDI-TOF MS as a tool for high-throughput dereplication. *Journal of Microbiological Methods*, *86*, 327–336. doi:10.1016/j.mimet.2011.06.004.

Gich, F., Janys, M. A., Konig, M., & Overmann, J. (2012). Enrichment of previously uncultured bacteria from natural complex communities by adhesion to solid surfaces. *Environmental Microbiology*, *14*, 2984–2997. doi:10.1111/j.1462-2920.2012.02868.x.

Gilligan, P. H., Gage, P. A., Bradshaw, L. M., Schidlow, D. V., & DeCicco, B. T. (1985). Isolation medium for the recovery of Pseudomonas cepacia from respiratory secretions of patients with cystic fibrosis. *Journal of Clinical Microbiology*, *22*, 5–8.

Gillis, M., Van Van, T., Bardin, R., Goor, M., Hebbar, P., Willems, A., Segers, P., Kersters, K., Heulin, T., &

Fernandez, M. P. (1995). Polyphasic taxonomy in the genus Burkholderia leading to an emended description of the genus and proposition of Burkholderia vietnamiensis sp. nov. for N2-fixing isolates from rice in Vietnam. *International Journal of Systematic Bacteriology*, *45*, 274–289. doi:10.1099/00207713-45-2-274.

Goris, J., De Vos, P., Caballero-Mellado, J., Park, J., Falsen, E., Quensen, J. F., Tiedje, J. M., & Vandamme, P. (2004). Classification of the biphenyl- and polychlorinated biphenyl-degrading strain LB400T and relatives as Burkholderia xenovorans sp. nov. *International Journal of Systematic and Evolutionary Microbiology*, *54*, 1677–1681. doi:10.1099/ijs.0.63101-0.

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. *International Journal of Systematic and Evolutionary Microbiology*, *57*, 81–91. doi:10.1099/ijs.0.64483-0.

Goris, J., Suzuki, K.-i., De Vos, P., Nakase, T., & Kersters, K. (1998). Evaluation of a microplate DNA-DNA hybridization method compared with the initial renaturation method. *Canadian Journal of Microbiology*, *44*, 1148–1153.

Govan, J. R. W., Brown, A. R., & Jones, A. M. (2007). Evolving epidemiology of Pseudomonas aeruginosa and the Burkholderia cepacia complex in cystic fibrosis lung infection. *Future Microbiology*, *2*, 153–164. doi:10.2217/17460913.2.2.153.

Govan, J. R. W., Brown, P. H., Maddison, J., Doherty, C. J., Nelson, J. W., Dodd, M., Greening, A. P., & Webb, A. K. (1993). Evidence for transmission of Pseudomonas cepacia by social contact in cystic fibrosis. *The Lancet*, *342*, 15–19.

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, *29*, 1072–1075. doi:10.1093/bioinformatics/btt086.

Gyaneshwar, P., Hirsch, A. M., Moulin, L., Chen, W.-M., Elliott, G. N., Bontemps, C., Estrada-de los Santos, P., Gross, E., dos Reis, F. B., Sprent, J. I., Young, J. P. W., & James, E. K. (2011). Legume-nodulating Betaproteobacteria: diversity, host range, and future prospects. *Molecular Plant-Microbe Interactions*, *24*, 1276–1288. doi:10.1094/MPMI-06-11-0172.

Hagedorn, C., Gould, W. D., Bardinelli, T. R., & Gustavson, D. R. (1987). A selective medium for enumeration and recovery of Pseudomonas cepacia biotypes from soil. *Applied and Environmental Microbiology*, *53*, 2265–2268.

Hahn, M. W., Stadler, P., Wu, Q. L., & Pockl, M. (2004). The filtration-acclimatization method for isolation of an important fraction of the not readily cultivable bacteria. *Journal of Microbiological Methods*, *57*, 379–390. doi:10.1016/j.mimet.2004.02.004.

Hall, B. G. (2011). *Phylogenetic trees made easy: a how-to manual*. (4th ed.). Sinauer Associates.

Hall, C. M., Busch, J. D., Shippy, K., Allender, C. J., Kaestli, M., Mayo, M., Sahl, J. W., Schupp, J. M., Colman, R. E., Keim, P., Currie, B. J., & Wagner, D. M. (2015). Diverse Burkholderia species isolated from soils in the southern United States with no evidence of B. pseudomallei. *PLOS ONE*, *10*, e0143254. doi:10.1371/journal.pone.0143254.

Harrison, F. (2007). Microbial ecology of the cystic fibrosis lung. *Microbiology*, *153*, 917–923. doi:10.1099/mic.0.2006/004077-0.

Hasegawa, M., Kishino, H., & Yano, T.-a. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, *22*, 160–174.

Hayatsu, M., Hirano, M., & Tokuda, S. (2000). Involvement of two plasmids in fenitrothion degradation by Burkholderia sp. strain NF100. *Applied and Environmental Microbiology*, *66*, 1737–1740. doi:10.1128/AEM.66.4.1737-1740.2000.

Henry, D., Campbell, M., McGimpsey, C., Clarke, A., Louden, L., Burns, J. L., Roe, M. H., Vandamme, P.,

& Speert, D. (1999). Comparison of isolation media for recovery of Burkholderia cepacia complex from respiratory secretions of patients with cystic fibrosis. *Journal of Clinical Microbiology*, *37*, 1004–1007.

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. *Journal of Clinical Microbiology*, *35*, 614–619.

Henry, D. A., Mahenthiralingam, E., Vandamme, P., Coenye, T., & Speert, D. P. (2001). Phenotypic methods for determining genomovar status of the Burkholderia cepacia complex. *Journal of Clinical Microbiology*, *39*, 1073–1078. doi:10.1128/JCM.39.3.1073-1078.2001.

Henz, S. R., Huson, D. H., Auch, A. F., Nieselt-Struwe, K., & Schuster, S. C. (2005). Whole-genome prokaryotic phylogeny. *Bioinformatics*, *21*, 2329–2335. doi:10.1093/bioinformatics/bth324.

Heo, S. T., Kim, S. J., Jeong, Y. G., Bae, I. G., Jin, J. S., & Lee, J. C. (2008). Hospital outbreak of Burkholderia stabilis bacteraemia related to contaminated chlorhexidine in haematological malignancy patients with indwelling catheters. *Journal of Hospital Infection*, *70*, 241–245. doi:10.1016/j.jhin.2008.07.019.

Hernandez-Mendoza, A., Martinez-Ocampo, F., Lozano-Aguirre Beltran, L. F., Popoca-Ursino, E. C., Ortiz-Hernandez, L., Sanchez-Salinas, E., & Dantan-Gonzalez, E. (2014). Draft genome sequence of the organophosphorus compound-degrading Burkholderia zhejiangensis strain CEIB S4-3. *Genome Announcements*, *2*, e01323–14–e01323–14. doi:10.1128/genomeA.01323-14.

Heuer, H., & Smalla, K. (2012). Plasmids foster diversification and adaptation of bacterial populations in soil. *FEMS Microbiology Reviews*, *36*, 1083–1104. doi:10.1111/j.1574-6976.2012.00337.x.

Ho, Y.-N., Chiang, H.-M., Chao, C.-P., Su, C.-C., Hsu, H.-F., Guo, C.-t., Hsieh, J.-L., & Huang, C.-C. (2015). In planta biocontrol of soilborne Fusarium wilt of banana through a plant endophytic bacterium, Burkholderia cenocepacia 869T2. *Plant and Soil*, *387*, 295–306. doi:10.1007/s11104-014-2297-0.

Holden, M. T. G., Seth-Smith, H. M. B., Crossman, L. C., Sebaihia, M., Bentley, S. D., Cerdeno-Tarraga, A. M., Thomson, N. R., Bason, N., Quail, M. A., Sharp, S., Cherevach, I., Churcher, C., Goodhead, I., Hauser, H., Holroyd, N., Mungall, K., Scott, P., Walker, D., White, B., Rose, H., Iversen, P., Mil-Homens, D., Rocha, E. P. C., Fialho, A. M., Baldwin, A., Dowson, C., Barrell, B. G., Govan, J. R., Vandamme, P., Hart, C. A., Mahenthiralingam, E., & Parkhill, J. (2009). The genome of Burkholderia cenocepacia J2315, an epidemic pathogen of cystic fibrosis patients. *Journal of Bacteriology*, *191*, 261–277. doi:10.1128/JB.01230-08.

Holmes, A., Govan, J., & Goldstein, R. (1998). Agricultural use of Burkholderia (Pseudomonas) cepacia: a threat to human health? *Emerging Infectious Diseases*, *4*, 221–227.

Hong, Q., Zhang, Z., Hong, Y., & Li, S. (2007). A microcosm study on bioremediation of fenitrothion-contaminated soil using Burkholderia sp. FDS-1. *International Biodeterioration & Biodegradation*, *59*, 55–61. doi:10.1016/j.ibiod.2006.07.013.

Hsueh, P.-T., Liu, J.-K., Chen, Y.-L., Liu, P.-J., Ni, W.-F., Chen, Y.-S., Wu, K.-M., & Lin, H.-H. (2015). Genomic sequence of Burkholderia multivorans NKI379, a soil bacterium that inhibits the growth of Burkholderia pseudomallei. *Genome Announcements*, *3*, e01294–15. doi:10.1128/genomeA.01294-15.

van Ingen, J., Boeree, M. J., Dekhuijzen, P. N. R., & van Soolingen, D. (2009). Environmental sources of rapid growing nontuberculous mycobacteria causing disease in humans. *Clinical Microbiology and Infection*, *15*, 888–893. doi:10.1111/j.1469-0691.2009.03013.x.

Inglis, T. J. J., & Sagripanti, J.-L. (2006). Environmental factors that affect the survival and persistence of Burkholderia pseudomallei. *Applied and Environmental Microbiology*, *72*, 6865–6875. doi:10.1128/AEM.01036-06.

Isles, A., Maclusky, I., Corey, M., Gold, R., Prober, C., Fleming, P., & Levison, H. (1984). Pseudomonas cepacia infection in cystic fibrosis: an emerging problem. *Journal of Pediatrics*, *104*, 206–210.

Izumi, H., Moore, E. R. B., Killham, K., Alexander, I. J., & Anderson, I. C. (2007). Characterisation of endobacterial communities in ectomycorrhizas by DNA- and RNA-based molecular methods. *Soil Biology and Biochemistry*, *39*, 891–899. doi:10.1016/j.soilbio.2006.10.011.

Jacobsen, C. S. (1997). Plant protection and rhizosphere colonization of barley by seed inoculated herbicide degrading Burkholderia (Pseudomonas) cepacia DBO1(pRO101) in 2,4-D contaminated soil. *Plant and Soil*, *189*, 139–144. doi:10.1023/A:1004296615446.

Jans, B., Pierard, D., Vrindts, E., & Vandamme, P. (2014). *Impact van het gebruik van een vloeibare zeep besmet met Burkholderia cepacia in Belgische ziekenhuizen*. Technical Report WIV-ISP.

Jeffries, C. D., Holtman, D. F., & Guse, D. G. (1957). Rapid method for determining the activity of microorganisms on nucleic acids. *Journal Of Bacteriology*, *73*, 590–591.

Jeong, Y., Kim, J., Kim, S., Kang, Y., Nagamatsu, T., & Hwang, I. (2003). Toxoflavin produced by Burkholderia glumae causing rice grain rot is responsible for inducing bacterial wilt in many field crops. *Plant Disease*, *87*, 890–895. doi:10.1094/PDIS.2003.87.8.890.

Jolley, K. A., Bliss, C. M., Bennett, J. S., Bratcher, H. B., Brehony, C., Colles, F. M., Wimalarathna, H., Harrison, O. B., Sheppard, S. K., Cody, A. J., & Maiden, M. C. J. (2012). Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. *Microbiology*, *158*, 1005–1015. doi:10.1099/mic.0.055459-0.

Jolley, K. A., Chan, M.-S., & Maiden, M. C. J. (2004). mlstdbNet – distributed multi-locus sequence typing (MLST) databases. *BMC Bioinformatics*, *5*, 86. doi:10.1186/1471-2105-5-86.

Jolley, K. A., & Maiden, M. C. (2010). BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics*, *11*. doi:10.1186/1471-2105-11-595.

Jolley, K. A., & Maiden, M. C. (2014). Using MLST to study bacterial variation: prospects in the genomic era. *Future Microbiology*, *9*, 623–630. doi:10.2217/fmb.14.24.

Jones, A. M., Dodd, M. E., Govan, J. R. W., Barcus, V., Doherty, C. J., Morris, J., & Webb, A. K. (2004). Burkholderia cenocepacia and Burkholderia multivorans: influence on survival in cystic fibrosis. *Thorax*, *59*, 948–951. doi:10.1136/thx.2003.017210.

Jorgensen, I. M., Johansen, H. K., Frederiksen, B., Pressler, T., Hansen, A., Vandamme, P., Hoiby, N., & Koch, C. (2003). Epidemic spread of Pandoraea apista, a new pathogen causing severe lung disease in cystic fibrosis patients. *Pediatric Pulmonology*, *36*, 439–446. doi:10.1002/ppul.10383.

Josephson, K. L., Gerba, C. P., & Pepper, I. L. (1993). Polmerase chain detection of nonviable bacterial pathogens. *Applied and Environmental Microbiology*, *59*, 3513–3515.

Juhas, M., van der Meer, J. R., Gaillard, M., Harding, R. M., Hood, D. W., & Crook, D. W. (2009). Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiology Reviews*, *33*, 376–393. doi:10.1111/j.1574-6976.2008.00136.x.

Jukes, T. H., & Cantor, C. R. (1969). Evolution of protein molecules. *Mammalian Protein Metabolism*, *3*, 21–132.

Kaestli, M., Schmid, M., Mayo, M., Rothballer, M., Harrington, G., Richardson, L., Hill, A., Hill, J., Tuanyok, A., Keim, P., Hartmann, A., & Currie, B. J. (2012). Out of the ground: aerial and exotic habitats of the melioidosis bacterium Burkholderia pseudomallei in grasses in Australia. *Environmental Microbiology*, *14*, 2058–2070. doi:10.1111/j.1462-2920.2011.02671.x.

Kalka-Moll, W. M., LiPuma, J. J., Accurso, F. J., Plum, G., Van Koningsbruggen, S., & Vandamme, P. (2009). Airway infection with a novel Cupriavidus species in persons with cystic fibrosis. *Journal of Clinical Microbiology*, *47*, 3026–3028. doi:10.1128/JCM.00846-09.

Karlin, S. (1998). Global dinucleotide signatures and analysis of genomic heterogeneity. *Current Opinion in*

Microbiology, *1*, 598–610. doi:10.1016/S1369-5274(98)80095-7.

Karlin, S., Campbell, A. M., & Mrazek, J. (1998). Comparative DNA analysis across diverse genomes. *Annual Review of Genetics*, *32*, 185–225. doi:10.1146/annurev.genet.32.1.185.

Karlin, S., Mrazek, J., & Campbell, A. M. (1997). Compositional biases of bacterial genomes and evolutionary implications. *Journal of Bacteriology*, *179*, 3899–3913.

Kikuchi, Y., Hayatsu, M., Hosokawa, T., Nagayama, A., Tago, K., & Fukatsu, T. (2012). Symbiont-mediated insecticide resistance. *Proceedings of the National Academy of Sciences*, *109*, 8618–8622. doi:10.1073/pnas.1200231109.

Kikuchi, Y., Hosokawa, T., & Fukatsu, T. (2007). Insect-microbe mutualism without vertical transmission: a stinkbug acquires a beneficial gut symbiont from the environment every generation. *Applied and Environmental Microbiology*, *73*, 4308–4316. doi:10.1128/AEM.00067-07.

Kikuchi, Y., Hosokawa, T., & Fukatsu, T. (2011). An ancient but promiscuous host–symbiont association between Burkholderia gut symbionts and their heteropteran hosts. *The ISME Journal*, *5*, 446–460. doi:10.1038/ismej.2010.150.

Kikuchi, Y., Meng, X.-Y., & Fukatsu, T. (2005). Gut symbiotic bacteria of the genus Burkholderia in the broad-headed bugs Riptortus clavatus and Leptocorisa chinensis (Heteroptera: Alydidae). *Applied and Environmental Microbiology*, *71*, 4035–4043. doi:10.1128/AEM.71.7.4035-4043.2005.

Kim, D. Y., Park, D. S., Kwon, S. B., Chung, M. G., Bae, K. S., Park, H. Y., & Rhee, Y. H. (2009). Biosynthesis of poly(3-hydroxybutyrate-co-3-hydroxyvalerate) copolyesters with a high molar fraction of 3-hydroxyvalerate by an insect-symbiotic burkholderia sp. IS-01. *Journal of Microbiology*, *47*, 651–656. doi:10.1007/s12275-009-0109-7.

Kim, O.-S., Cho, Y.-J., Lee, K., Yoon, S.-H., Kim, M., Na, H., Park, S.-C., Jeon, Y. S., Lee, J.-H., Yi, H., Won, S., & Chun, J. (2012). Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. *International Journal of Systematic and Evolutionary Microbiology*, *62*, 716–721. doi:10.1099/ijs.0.038075-0.

King, E. B., & Parke, J. L. (1993). Biocontrol of Aphanomyces root rot and Pythium damping-off by Pseudomonas cepacia AMMD on four pea cultivars. *Plant Disease*, *77*, 1185–1188.

Klenk, H. P., & Goker, M. (2010). En route to a genome-based classification of Archaea and Bacteria? *Systematic and Applied Microbiology*, *33*, 175–182. doi:10.1016/j.syapm.2010.03.003.

Konstantinidis, K. T., & Tiedje, J. M. (2004). Trends between gene content and genome size in prokaryotic species with larger genomes. *Proceedings of the National Academy of Sciences*, *101*, 3160–3165. doi:10.1073/pnas.0308653100.

Konstantinidis, K. T., & Tiedje, J. M. (2005a). Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences*, *102*, 2567–2572. doi:10.1073/pnas.0409727102.

Konstantinidis, K. T., & Tiedje, J. M. (2005b). Towards a genome-based taxonomy for prokaryotes. *Journal of Bacteriology*, *187*, 6258–6264. doi:10.1128/JB.187.18.6258-6264.2005.

Koren, S., & Phillippy, A. M. (2015). One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly. *Current Opinion in Microbiology*, *23*, 110–120. doi:10.1016/j.mib.2014.11.014.

Krumsiek, J., Arnold, R., & Rattei, T. (2007). Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics*, *23*, 1026–1028. doi:10.1093/bioinformatics/btm039.

Kumar, S., Vikram, S., & Raghava, G. P. S. (2012). Genome sequence of the nitroaromatic compound-degrading bacterium Burkholderia sp. strain SJ98. *Journal of Bacteriology*, *194*, 3286–3286. doi:10.1128/

JB.00497-12.

Kwak, Y., & Shin, J.-H. (2015). Complete genome sequence of Burkholderia pyrrocinia 2327T, the first industrial bacterium which produced antifungal antibiotic pyrrolnitrin. *Journal of Biotechnology*, *211*, 3–4. doi:10.1016/j.jbiotec.2015.06.420.

La Scola, B., Mezi, L., Weiller, P. J., & Raoult, D. (2001). Isolation of Legionella anisa using an amoebic coculture procedure. *Journal of Clinical Microbiology*, *39*, 365–366. doi:10.1128/JCM.39.1.365-366.2001.

Lagier, J.-C., Armougom, F., Million, M., Hugon, P., Pagnier, I., Robert, C., Bittar, F., Fournous, G., Gimenez, G., Maraninchi, M., Trape, J.-F., Koonin, E., La Scola, B., & Raoult, D. (2012). Microbial culturomics: paradigm shift in the human gut microbiome study. *Clinical Microbiology and Infection*, *18*, 1185–1193. doi:10.1111/1469-0691.12023.

Lagier, J.-C., Edouard, S., Pagnier, I., Mediannikov, O., Drancourt, M., & Raoult, D. (2015a). Current and past strategies for bacterial culture in clinical microbiology. *Clinical Microbiology Reviews*, *28*, 208–236. doi:10.1128/CMR.00110-14.

Lagier, J.-C., Hugon, P., Khelaifia, S., Fournier, P.-E., La Scola, B., & Raoult, D. (2015b). The rebirth of culture in microbiology through the example of culturomics to study human gut microbiota. *Clinical Microbiology Reviews*, *28*, 237–264. doi:10.1128/CMR.00014-14.

Lambiase, A., Del Pezzo, M., Cerbone, D., Raia, V., Rossano, F., & Catania, M. R. (2013). Rapid identification of Burkholderia cepacia complex species recovered from cystic fibrosis patients using matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Journal of Microbiological Methods*, *92*, 145–149. doi:10.1016/j.mimet.2012.11.010.

Lamothe, J., Thyssen, S., & Valvano, M. A. (2004). Burkholderia cepacia complex isolates survive intracellularly without replication within acidic vacuoles of Acanthamoeba polyphaga. *Cellular Microbiology*, *6*, 1127–1138. doi:10.1111/j.1462-5822.2004.00424.x.

Landers, P., Kerr, K. G., Rowbotham, T. J., Tipper, J. L., Keig, P. M., Ingham, E., & Denton, M. (2000). Survival and growth of Burkholderia cepacia within the free-living amoeba Acanthamoeba polyphaga. *European Journal of Clinical Microbiology and Infectious Diseases*, *19*, 121–123. doi:10.1007/s100960050442.

Lee, C. J., Lee, J. T., Kwon, J. H., Kim, B. C., & Park, W. (2005). Occurrence of bacterial soft rot of onion plants caused by Burkholderia gladioli pv. alliicola in Korea. *Australasian Plant Pathology*, *34*, 287. doi:10.1071/AP05024.

Lefort, V., Desper, R., & Gascuel, O. (2015). FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program. *Molecular Biology and Evolution*, *32*, 2798–2800. doi:10.1093/molbev/msv150.

Lemaire, B., Robbrecht, E., van Wyk, B., Van Oevelen, S., Verstraete, B., Prinsen, E., Smets, E., & Dessein, S. (2011a). Identification, origin, and evolution of leaf nodulating symbionts of Sericanthe (Rubiaceae). *The Journal of Microbiology*, *49*, 935–941. doi:10.1007/s12275-011-1163-5.

Lemaire, B., Van Oevelen, S., De Block, P., Verstraete, B., Smets, E., Prinsen, E., & Dessein, S. (2012). Identification of the bacterial endosymbionts in leaf nodules of Pavetta (Rubiaceae). *International Journal of Systematic and Evolutionary Microbiology*, *62*, 202–209. doi:10.1099/ijs.0.028019-0.

Lemaire, B., Vandamme, P., Merckx, V., Smets, E., & Dessein, S. (2011b). Bacterial leaf symbiosis in angiosperms: host specificity without co-speciation. *PLoS ONE*, *6*, e24430. doi:10.1371/journal.pone.0024430.

Lerat, E., Daubin, V., & Moran, N. A. (2003). From gene trees to organismal phylogeny in prokaryotes: the case of the γ-Proteobacteria. *PLoS Biology*, *1*, e19. doi:10.1371/journal.pbio.0000019.

Lessie, T. G., Hendrickson, W., Manning, B., & Devereux, R. (1996). Genomic complexity and plasticity of

Burkholderia cepacia. *FEMS Microbiology Letters*, *144*, 117–128.

Letunic, I., & Bork, P. (2011). Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Research*, *39*, W475–W478. doi:10.1093/nar/gkr201.

Levy, A., Chang, B. J., Abbott, L. K., Kuo, J., Harnett, G., & Inglis, T. J. J. (2003). Invasion of spores of the arbuscular mycorrhizal fungus Gigaspora decipiens by Burkholderia spp. *Applied and Environmental Microbiology*, *69*, 6250–6256. doi:10.1128/AEM.69.10.6250-6256.2003.

Li, G.-X., Wu, X.-Q., & Ye, J.-R. (2013). Biosafety and colonization of Burkholderia multivorans WS-FJ9 and its growth-promoting effects on poplars. *Applied Microbiology and Biotechnology*, *97*, 10489–10498. doi:10.1007/s00253-013-5276-0.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*, .

Liao, Y.-C., Lin, S.-H., & Lin, H.-H. (2015). Completing bacterial genome assemblies: strategy and performance comparisons. *Scientific Reports*, *5*, 8747. doi:10.1038/srep08747.

Librado, P., & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, *25*, 1451–1452. doi:10.1093/bioinformatics/btp187.

Lieberman, T. D., Michel, J.-B., Aingaran, M., Potter-Bynoe, G., Roux, D., Davis, M. R., Skurnik, D., Leiby, N., LiPuma, J. J., Goldberg, J. B., McAdam, A. J., Priebe, G. P., & Kishony, R. (2011). Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. *Nature Genetics*, *43*, 1275–1280. doi:10.1038/ng.997.

Lim, J. S., Choi, B. S., Choi, A. Y., Kim, K. D., Kim, D. I., Choi, I. Y., & Ka, J.-O. (2012). Complete genome sequence of the fenitrothion-degrading Burkholderia sp. strain YI23. *Journal of Bacteriology*, *194*, 896–896. doi:10.1128/JB.06479-11.

Lim, Y. W., Baik, K. S., Han, S. K., Kim, S. B., & Bae, K. S. (2003). Burkholderia sordidicola sp. nov., isolated from the white-rot fungus Phanerochaete sordida. *International Journal of Systematic and Evolutionary Microbiology*, *53*, 1631–1636. doi:10.1099/ijs.0.02456-0.

Lin, H.-H., Chen, Y.-S., Li, Y.-C., Tseng, I.-L., Hsieh, T.-H., Buu, L.-M., & Chen, Y.-L. (2011). Burkholderia multivorans acts as an antagonist against the growth of Burkholderia pseudomallei in soil. *Microbiology and Immunology*, *55*, 616–624. doi:10.1111/j.1348-0421.2011.00365.x.

LiPuma, J., & Mahenthiralingam, E. (1999). Commercial use of Burkholderia cepacia. *Emerging Infectious Diseases*, *5*, 305–306.

LiPuma, J. J. (2010). The changing microbial epidemiology in cystic fibrosis. *Clinical Microbiology Reviews*, *23*, 299–323. doi:10.1128/cmr.00068-09.

LiPuma, J. J., Dasen, S. E., Nielson, D., Stern, R. C., & Stull, T. L. (1990). Person-to-person transmission of Pseudomonas cepacia between patients with cystic fibrosis. *The Lancet*, *336*, 1094–1096. doi:10.1016/0140-6736(90)92571-X.

LiPuma, J. J., Dulaney, B. J., McMenamin, J. D., Whitby, P. W., Stull, T. L., Coenye, T., & Vandamme, P. (1999). Development of rRNA-based PCR assays for identification of Burkholderia cepacia complex isolates recovered from cystic fibrosis patients. *Journal of Clinical Microbiology*, *37*, 3167–3170.

LiPuma, J. J., Spilker, T., Coenye, T., & Gonzalez, C. F. (2002). An epidemic Burkholderia cepacia complex strain identified in soil. *The Lancet*, *359*, 2002–2003. doi:10.1016/S0140-6736(02)08836-0.

Liu, X.-Y., Li, C. L., Luo, X.-J., Lai, Q.-L., & Xu, J.-H. (2014). Burkholderia jiangsuensis sp. nov., a methyl parathion degrading bacterium, isolated from methyl parathion contaminated soil. *International Journal of Systematic and Evolutionary Microbiology*, *64*, 3247–3253. doi:10.1099/ijs.0.064444-0.

Liu, Y., Lai, Q., Goker, M., Meier-Kolthoff, J. P., Wang, M., Sun, Y., Wang, L., & Shao, Z. (2015). Genomic insights into the taxonomic status of the Bacillus cereus group. *Scientific Reports*, *5*, 14082.

doi:10.1038/srep14082.

Livermore, J. A., Emrich, S. J., Tan, J., & Jones, S. E. (2014). Freshwater bacterial lifestyles inferred from comparative genomics. *Environmental Microbiology*, *16*, 746–758. doi:10.1111/1462-2920.12199.

Llado, S., Xu, Z., Sorensen, S. J., & Baldrian, P. (2014). Draft genome sequence of Burkholderia sordidicola S170, a potential plant growth promoter isolated from coniferous forest soil in the Czech Republic. *Genome Announcements*, *2*, e00810–14–e00810–14. doi:10.1128/genomeA.00810-14.

Loutet, S. A., & Valvano, M. A. (2010). A decade of Burkholderia cenocepacia virulence determinant research. *Infection and Immunity*, *78*, 4088–4100. doi:10.1128/IAI.00212-10.

Lu, P., Zheng, L.-Q., Sun, J.-J., Liu, H.-M., Li, S.-P., Hong, Q., & Li, W.-J. (2012). Burkholderia zhejiangensis sp. nov., a methyl-parathion-degrading bacterium isolated from a wastewater-treatment system. *International Journal of Systematic and Evolutionary Microbiology*, *62*, 1337–1341. doi:10.1099/ijs.0.035428-0.

Lugtenberg, B., & Kamilova, F. (2009). Plant-growth-promoting rhizobacteria. *Annual Review of Microbiology*, *63*, 541–556. doi:10.1146/annurev.micro.62.081307.162918.

Lundgren, J. G., Lehman, R. M., & Chee-sanford, J. (2007). Bacterial communities within digestive tracts of ground beetles (Coleoptera: Carabidae). *Annals of the Entomological Society of America*, *100*, 275–282.

Mahenthiralingam, E., Baldwin, A., & Dowson, C. G. (2008). Burkholderia cepacia complex bacteria: opportunistic pathogens with important natural biology. *Journal of Applied Microbiology*, *104*, 1539–1551. doi:10.1111/j.1365-2672.2007.03706.x.

Mahenthiralingam, E., Baldwin, A., & Vandamme, P. (2002). Burkholderia cepacia complex infection in patients with cystic fibrosis. *Journal of Medical Microbiology*, *51*, 533–538.

Mahenthiralingam, E., Bischof, J., Byrne, S. K., Radomski, C., Davies, J. E., Av-gay, Y., & Vandamme, P. (2000a). DNA-based diagnostic approaches for identification of Burkholderia cepacia complex, Burkholderia vietnamiensis, Burkholderia multivorans, Burkholderia stabilis, and Burkholderia cepacia genomovars I and III. *Journal of Clinical Microbiology*, *38*, 3165–3173.

Mahenthiralingam, E., Campbell, M. E., Henry, D. A., & Speert, D. P. (1996). Epidemiology of Burkholderia cepacia infection in patients with cystic fibrosis: analysis by randomly amplified polymorphic DNA fingerprinting. *Journal of Clinical Microbiology*, *34*, 2914–2920.

Mahenthiralingam, E., Coenye, T., Chung, J. W., Speert, D. P., Govan, J. R. W., Taylor, P., & Vandamme, P. (2000b). Diagnostically and experimentally useful panel of strains from the Burkholderia cepacia complex. *Journal of Clinical Microbiology*, *38*, 910–913.

Mahenthiralingam, E., Urban, T. A., & Goldberg, J. B. (2005). The multifarious, multireplicon Burkholderia cepacia complex. *Nature Reviews Microbiology*, *3*, 144–156. doi:10.1038/nrmicro1085.

Mahenthiralingam, E., Vandamme, P., Campbell, M. E., Henry, D. A., Gravelle, A. M., Wong, L. T. K., Davidson, A. G. F., Wilcox, P. G., Nakielna, B., & Speert, D. P. (2001). Infection with Burkholderia cepacia complex genomovars in patients with cystic fibrosis: virulent transmissible strains of genomovar III can replace Burkholderia multivorans. *Clinical Infectious Diseases*, *33*, 1469–1475.

Maiden, M. C. J., van Rensburg, M. J. J., Bray, J. E., Earle, S. G., Ford, S. A., Jolley, K. A., & McCarthy, N. D. (2013). MLST revisited: the gene-by-gene approach to bacterial genomics. *Nature Reviews Microbiology*, *11*, 728–736. doi:10.1038/nrmicro3093.

Marks, L. R., Dodd, H., Russo, T. A., & Berenson, C. S. (2016). Burkholderia ginsengisoli bacteraemia: emergence of a novel pathogen. *BMJ case reports*, *online*. doi:10.1136/bcr-2015-213584.

Marolda, C. L., Hauroder, B., John, M. A., Michel, R., & Valvano, M. A. (1999). Intracellular survival and saprophytic growth of isolates from the Burkholderia cepacia complex in free-living amoebae. *Microbiology*, *145*, 1509–1517.

Mars, A. E., Houwing, J., Dolfing, J., & Janssen, D. B. (1996). Degradation of toluene and trichloroethylene by Burkholderia cepacia G4 in growth-limited fed-batch culture. *Applied and Environmental Microbiology*, *62*, 886–91.

Martina, P., Bettiol, M., Vescina, C., Montanaro, P., Mannino, M. C., Prieto, C. I., Vay, C., Naumann, D., Schmitt, J., Yantorno, O., Lagares, A., & Bosch, A. (2013). Genetic diversity of Burkholderia contaminans isolates from cystic fibrosis patients in Argentina. *Journal of Clinical Microbiology*, *51*, 339–344. doi:10.1128/JCM.02500-12.

Mayo, M., Kaestli, M., Harrington, G., Cheng, A. C., Ward, L., Karp, D., Jolly, P., Godoy, D., Spratt, B. G., & Currie, B. J. (2011). Burkholderia pseudomallei in unchlorinated domestic bore water, tropical northern Australia. *Emerging Infectious Diseases*, *17*, 1283–1285. doi:10.3201/eid1707.100614.

McAdam, P. R., Richardson, E. J., & Fitzgerald, J. R. (2014). High-throughput sequencing for the study of bacterial pathogen biology. *Current Opinion in Microbiology*, *19*, 106–113. doi:10.1016/j.mib.2014.06.002.

McDougald, D., Rice, S. A., Weichart, D., & Kjelleberg, S. (1998). Nonculturability: adaption or debilitation? *FEMS Microbiology Ecology*, *25*.

McDowell, A., Mahenthiralingam, E., Moore, J. E., Dunbar, K. E. A., Webb, A. K., Dodd, M. E., Martin, S. L., Millar, B. C., Scott, C. J., Crowe, M., & Elborn, J. S. (2001). PCR-based detection and identification of Burkholderia cepacia complex pathogens in sputum from cystic fibrosis patients. *Journal of Clinical Microbiology*, *39*, 4247–4255. doi:10.1128/JCM.39.12.4247-4255.2001.

McMenamin, J. D., Zaccone, T. M., Coenye, T., Vandamme, P., & LiPuma, J. J. (2000). Misidentification of Burkholderia cepacia in US cystic fibrosis treatment centers. *Chest*, *117*, 1661–1665. doi:10.1378/chest.117.6.1661.

Medina-Pascual, M. J., Valdezate, S., Carrasco, G., Villalon, P., Garrido, N., & Saez-Nieto, J. A. (2015). Increase in isolation of Burkholderia contaminans from Spanish patients with cystic fibrosis. *Clinical Microbiology and Infection*, *21*, 150–156. doi:10.1016/j.cmi.2014.07.014.

Medina-Pascual, M. J., Valdezate, S., Villalon, P., Garrido, N., Rubio, V., & Saez-Nieto, J. A. (2012). Identification, molecular characterisation and antimicrobial susceptibility of genomovars of the Burkholderia cepacia complex in Spain. *European Journal of Clinical Microbiology and Infectious Diseases*, *31*, 3385–3396. doi:10.1007/s10096-012-1707-6.

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H.-P., & Goker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics*, *14*, 60. doi:10.1186/1471-2105-14-60.

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H.-P., & Goker, M. (2014a). Highly parallelized inference of large genome-based phylogenies. *Concurrency and Computation: Practice and Experience*, *26*, 1715–1729. doi:10.1002/cpe.3112.

Meier-Kolthoff, J. P., Hahnke, R. L., Petersen, J., Scheuner, C., Michael, V., Fiebig, A., Rohde, C., Rohde, M., Fartmann, B., Goodwin, L. A., Chertkov, O., Reddy, T., Pati, A., Ivanova, N. N., Markowitz, V., Kyrpides, N. C., Woyke, T., Goker, M., & Klenk, H.-P. (2014b). Complete genome sequence of DSM 30083T, the type strain (U5/41T) of Escherichia coli, and a proposal for delineating subspecies in microbial taxonomy. *Standards in Genomic Sciences*, *9*, 2. doi:10.1186/1944-3277-9-2.

Meier-Kolthoff, J. P., Klenk, H.-P., & Goker, M. (2014c). Taxonomic use of DNA G+C content and DNA–DNA hybridization in the genomic age. *International Journal of Systematic and Evolutionary Microbiology*, *64*, 352–356. doi:10.1099/ijs.0.056994-0.

Mesbah, M., Premachandran, U., & Whitman, W. B. (1989). Precise measurement of the G+C content of

deoxyribonucleic acid by high-performance liquid-chromatography. *International Journal of Systematic Bacteriology*, *39*, 159–167.

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. *Journal of Chromatography*, *479*, 297–306.

Millar-Jones, L., Ryley, H. C., Paull, A., & Goodchild, M. C. (1998). Transmission and prevalence of Burkholderia cepacia in Welsh cystic fibrosis patients. *Respiratory Medicine*, *92*, 178–183.

Miller, D. N. (2001). Evaluation of gel filtration resins for the removal of PCR-inhibitory substances from soils and sediments. *Journal of Microbiological Methods*, *44*, 49–58. doi:10.1016/S0167-7012(00)00228-1.

Miller, S. C. M., LiPuma, J. J., & Parke, J. L. (2002). Culture-based and non-growth-dependent detection of the Burkholderia cepacia complex in soil environments. *Applied and Environmental Microbiology*, *68*, 3750–3758. doi:10.1128/aem.68.8.3750-3758.2002.

Moore, J. E., Millar, B. C., Xu, J., Crowe, M., Redmond, A. O. B., & Elborn, J. S. (2002). Misidentification of a genomovar of Burkholderia cepacia by recA restriction fragment length polymorphism. *Journal of Clinical Pathology*, *55*, 309–311.

Moreira, D. (1998). Efficient removal of PCR inhibitors using agarose-embedded DNA preparations. *Nucleic Acids Research*, *26*, 3309–3310.

Morrow, J. D., & Cooper, V. S. (2012). Evolutionary effects of translocations in bacterial genomes. *Genome Biology and Evolution*, *4*, 1256–1262. doi:10.1093/gbe/evs099.

Murray, S., Charbeneau, J., Marshall, B. C., & LiPuma, J. J. (2008). Impact of Burkholderia infection on lung transplantation in cystic fibrosis. *American Journal of Respiratory and Critical Care Medicine*, *178*, 363–371. doi:10.1164/rccm.200712-1834OC.

Nazir, R., Hansen, M. A., Sorensen, S., & van Elsas, J. D. (2012a). Draft genome sequence of the soil bacterium Burkholderia terrae strain BS001, which interacts with fungal surface structures. *Journal of Bacteriology*, *194*, 4480–4481. doi:10.1128/JB.00725-12.

Nazir, R., Zhang, M., de Boer, W., & van Elsas, J. D. (2012b). The capacity to comigrate with Lyophyllum sp. strain Karsten through different soils is spread among several phylogenetic groups within the genus Burkholderia. *Soil Biology and Biochemistry*, *50*, 221–233. doi:10.1016/j.soilbio.2012.03.015.

Niemann, S., Puhler, A., Tichy, H. V., Simon, R., & Selbitschka, W. (1997). Evaluation of the resolving power of three different DNA fingerprinting methods to discriminate among isolates of a natural Rhizobium meliloti population. *Journal of Applied Microbiology*, *82*, 477–484.

Nogales, B., Moore, E. R. B., Llobet-Brossa, E., Rossello-Mora, R., Amann, R., & Timmis, K. N. (2001). Combined use of 16S ribosomal DNA and 16S rRNA to study the bacterial community of polychlorinated biphenyl-polluted soil. *Applied and Environmental Microbiology*, *67*, 1874–1884. doi:10.1128/AEM.67.4.1874-1884.2001.

Norskov-Lauritsen, N., Johansen, H. K., Fenger, M. G., Nielsen, X. C., Pressler, T., Olesen, H. V., & Hoiby, N. (2010). Unusual distribution of Burkholderia cepacia complex species in Danish cystic fibrosis clinics may stem from restricted transmission between patients. *Journal of Clinical Microbiology*, *48*, 2981–2983. doi:10.1128/JCM.00383-10.

Notredame, C., Higgins, D. G., & Heringa, J. (2000). T-coffee: a novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology*, *302*, 205–217. doi:10.1006/jmbi.2000.4042.

Oliver, J. D. (2010). Recent findings on the viable but nonculturable state in pathogenic bacteria. *FEMS Microbiology Reviews*, *34*, 415–425. doi:10.1111/j.1574-6976.2009.00200.x.

Oren, A., & Garrity, G. M. (2015). List of new names and new combinations previously effectively, but not

validly, published. *International Journal of Systematic and Evolutionary Microbiology*, *65*, 2017–2025. doi:10.1099/ijs.0.000317.

O'Sullivan, B. P., & Freedman, S. D. (2009). Cystic fibrosis. *The Lancet*, *373*, 1891–1904. doi:10.1016/S0140-6736(09)60327-5.

Pagnier, I., Raoult, D., & La Scola, B. (2008). Isolation and identification of amoeba-resisting bacteria from water in human environment by using an Acanthamoeba polyphaga co-culture procedure. *Environmental Microbiology*, *10*, 1135–1144. doi:10.1111/j.1462-2920.2007.01530.x.

Pallen, M. J., Loman, N. J., & Penn, C. W. (2010). High-throughput sequencing and clinical microbiology: progress, opportunities and challenges. *Current Opinion in Microbiology*, *13*, 625–631. doi:10.1016/j.mib.2010.08.003.

Palleroni, N. J., Kunisawa, R., Contopoulou, R., & Doudoroff, M. (1973). Nucleic acid homologies in the genus Pseudomonas. *International Journal of Systematic Bacteriology*, *23*, 333 – 339. doi:10.1099/00207713-23-4-333.

Pallud, C., Viallard, V., Balandreau, J., Normand, P., & Grundmann, G. (2001). Combined use of a specific probe and PCAT medium to study Burkholderia in soil. *Journal of Microbiological Methods*, *47*, 25–34. doi:10.1016/S0167-7012(01)00287-1.

Parke, J. L., & Gurian-Sherman, D. (2001). Diversity of the Burkholderia cepacia complex and implications for risk assessment of biological control strains. *Annual Review of Phytopathology*, *39*, 225–258. doi:10.1146/annurev.phyto.39.1.225.

Partida-Martinez, L. P., Groth, I., Schmitt, I., Richter, W., Roth, M., & Hertweck, C. (2007). Burkholderia rhizoxinica sp. nov. and Burkholderia endofungorum sp. nov., bacterial endosymbionts of the plant-pathogenic fungus Rhizopus microsporus. *International Journal of Systematic and Evolutionary Microbiology*, *57*, 2583–2590. doi:10.1099/ijs.0.64660-0.

Partida-Martinez, L. P., & Hertweck, C. (2005). Pathogenic fungus harbours endosymbiotic bacteria for toxin production. *Nature*, *437*, 884–888. doi:10.1038/nature03997.

Pawitwar, S. S., Utturkar, S. M., Brown, S. D., Yoshinaga, M., & Rosen, B. P. (2015). Draft genome sequence of Burkholderia sp. MR1, a methylarsenate-reducing bacterial isolate from Florida golf course soil. *Genome Announcements*, *3*, e00608–15. doi:10.1128/genomeA.00608-15.

Payne, G. W., Ramette, A., Rose, H. L., Weightman, A. J., Jones, T. H., Tiedje, J. M., & Mahenthiralingam, E. (2006). Application of a recA gene-based identification approach to the maize rhizosphere reveals novel diversity in Burkholderia species. *FEMS Microbiology Letters*, *259*, 126–132. doi:10.1111/j.1574-6968.2006.00257.x.

Peeters, C., Depoorter, E., Praet, J., & Vandamme, P. (2016). Extensive cultivation of soil and water samples yields various pathogens in patients with cystic fibrosis but not Burkholderia multivorans. *Journal of Cystic Fibrosis*, *in press*. doi:10.1016/j.jcf.2016.02.014.

Peeters, C., Zlosnik, J. E. A., Spilker, T., Hird, T. J., LiPuma, J. J., & Vandamme, P. (2013). Burkholderia pseudomultivorans sp. nov., a novel Burkholderia cepacia complex species from human respiratory samples and the rhizosphere. *Systematic and Applied Microbiology*, *36*, 483–489. doi:10.1016/j.syapm.2013.06.003.

Perin, L., Martinez-Aguilar, L., Paredes-Valdez, G., Baldani, J. I., Estrada-de los Santos, P., Reis, V. M., & Caballero-Mellado, J. (2006). Burkholderia silvatlantica sp. nov., a diazotrophic bacterium associated with sugar cane and maize. *International Journal of Systematic and Evolutionary Microbiology*, *56*, 1931–1937. doi:10.1099/ijs.0.64362-0.

Pinto-Carbo, M., Sieber, S., Dessein, S., Wicker, T., Verstraete, B., Gademann, K., Eberl, L., & Carlier, A.

(2016). Evidence of horizontal gene transfer between obligate leaf nodule symbionts. *The ISME Journal*, *10*, 2092–2105. doi:10.1038/ismej.2016.27.

Pirone, L., Chiarini, L., Dalmastri, C., Bevivino, A., & Tabacchioni, S. (2005). Detection of cultured and uncultured Burkholderia cepacia complex bacteria naturally occurring in the maize rhizosphere. *Environmental Microbiology*, *7*, 1734–1742. doi:10.1111/j.1462-2920.2005.00897.x.

Pitcher, D. G., Saunders, N. A., & Owen, R. J. (1989). Rapid extraction of bacterial genomic DNA with guanidium thiocyanate. *Letters in Applied Microbiology*, *8*, 151–156.

Pope, C. E., Short, P., & Carter, P. E. (2010). Species distribution of Burkholderia cepacia complex isolates in cystic fibrosis and non-cystic fibrosis patients in New Zealand. *Journal of Cystic Fibrosis*, *9*, 442–446. doi:10.1016/j.jcf.2010.08.011.

Preece, C. L., Perry, A., Gray, B., Kenna, D. T., Jones, A. L., Cummings, S. P., Robb, A., Thomas, M. F., Brodlie, M., O'Brien, C. J., Bourke, S. J., & Perry, J. D. (2016). A novel culture medium for isolation of rapidly-growing mycobacteria from the sputum of patients with cystic fibrosis. *Journal of Cystic Fibrosis*, *15*, 186–191. doi:10.1016/j.jcf.2015.05.002.

Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS ONE*, *5*, e9490. doi:10.1371/journal.pone.0009490.

Pruesse, E., Peplies, J., & Glockner, F. O. (2012). SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics*, *28*, 1823–1829. doi:10.1093/bioinformatics/bts252.

Pumphrey, G. M., & Madsen, E. L. (2008). Field-based stable isotope probing reveals the identities of benzoic acid-metabolizing microorganisms and their in situ growth in agricultural soil. *Applied and Environmental Microbiology*, *74*, 4111–4118. doi:10.1128/AEM.00464-08.

Purdy-Gibson, M. E., France, M., Hundley, T. C., Eid, N., & Remold, S. K. (2015). Pseudomonas aeruginosa in CF and non-CF homes is found predominantly in drains. *Journal of Cystic Fibrosis*, *14*, 341–346. doi:10.1016/j.jcf.2014.10.008.

Raes, J., Korbel, J. O., Lercher, M. J., von Mering, C., & Bork, P. (2007). Prediction of effective genome size in metagenomic samples. *Genome Biology*, *8*, R10. doi:10.1186/gb-2007-8-1-r10.

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. *International Journal of Systematic and Evolutionary Microbiology*, *64*, 384–391. doi:10.1099/ijs.0.057091-0.

Ramette, A., LiPuma, J. J., & Tiedje, J. M. (2005). Species abundance and diversity of Burkholderia cepacia complex in the environment. *Applied and Environmental Microbiology*, *71*, 1193–1201. doi:10.1128/AEM.71.3.1193-1201.2005.

Renesto, P., Crapoulet, N., Ogata, H., La Scola, B., Vestris, G., Claverie, J.-M., & Raoult, D. (2003). Genome-based design of a cell-free culture medium for Tropheryma whipplei. *The Lancet*, *362*, 447–449. doi:10.1016/S0140-6736(03)14071-8.

Renwick, J., McNally, P., John, B., DeSantis, T., Linnane, B., & Murphy, P. (2014). The microbial community of the cystic fibrosis airway is disrupted in early life. *PLoS ONE*, *9*, e109798. doi:10.1371/journal.pone.0109798.

Richter, M., & Rossello-Mora, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences*, *106*, 19126–19131. doi:10.1073/pnas.0906412106.

Rissman, A. I., Mau, B., Biehl, B. S., Darling, A. E., Glasner, J. D., & Perna, N. T. (2009). Reordering contigs of draft genomes using the Mauve Aligner. *Bioinformatics*, *25*, 2071–2073. doi:10.1093/bioinformatics/

btp356.

Rossello-Mora, R., & Amann, R. (2015). Past and future species definitions for Bacteria and Archaea. *Systematic and Applied Microbiology*, *38*, 209–216. doi:10.1016/j.syapm.2015.02.001.

Rousk, J., Baath, E., Brookes, P. C., Lauber, C. L., Lozupone, C., Caporaso, J. G., Knight, R., & Fierer, N. (2010). Soil bacterial and fungal communities across a pH gradient in an arable soil. *The ISME Journal*, *4*, 1340–1351. doi:10.1038/ismej.2010.58.

Rowe, S. M., Miller, S., & Sorscher, E. J. (2005). Cystic fibrosis. *The New England Journal of Medicine*, *352*, 1992–2001.

Saenz, V. L., Maggi, R. G., Breitschwerdt, E. B., Kim, J., Vargo, E. L., & Schal, C. (2013). Survey of Bartonella spp. in U.S. bed bugs detects Burkholderia multivorans but not Bartonella. *PLoS ONE*, *8*, e73661. doi:10.1371/journal.pone.0073661.

Saini, L. S., Galsworthy, S. B., John, M. A., & Valvano, M. A. (1999). Intracellular survival of Burkholderia cepacia complex isolates in the presence of macrophage cell activation. *Microbiology*, *145*, 3465–3475. doi:10.1099/00221287-145-12-3465.

Salles, J. F., van Elsas, J. D., & van Veen, J. A. (2006a). Effect of agricultural management regime on Burkholderia community structure in soil. *Microbial Ecology*, *52*, 267–279. doi:10.1007/s00248-006-9048-6.

Salles, J. F., Samyn, E., Vandamme, P., van Veen, J. A., & van Elsas, J. D. (2006b). Changes in agricultural management drive the diversity of Burkholderia species isolated from soil on PCAT medium. *Soil Biology and Biochemistry*, *38*, 661–673. doi:10.1016/j.soilbio.2005.06.018.

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., Turner, P., Parkhill, J., Loman, N. J., & Walker, A. W. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology*, *12*, 87. doi:10.1186/s12915-014-0087-z.

Santos, A. V., Dillon, R. J., Dillon, V. M., Reynolds, S. E., & Samuels, R. I. (2004). Ocurrence of the antibiotic producing bacterium Burkholderia sp. in colonies of the leaf-cutting ant Atta sexdens rubropilosa. *FEMS Microbiology Letters*, *239*, 319–323. doi:10.1016/j.femsle.2004.09.005.

Savichtcheva, O., Okayama, N., Ito, T., & Okabe, S. (2005). Application of a direct fluorescence-based live/dead staining combined with fluorescence in situ hybridization for assessment of survival rate of Bacteroides spp. in drinking water. *Biotechnology and Bioengineering*, *92*, 356–363. doi:10.1002/bit.20608.

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. *Frontiers in Genetics*, *5*. doi:10.3389/fgene.2014.00429.

Scherlach, K., Graupner, K., & Hertweck, C. (2013). Molecular bacteria-fungi interactions: effects on environment, food, and medicine. *Annual Review of Microbiology*, *67*, 375–397. doi:10.1146/annurev-micro-092412-155702.

Schleifer, K. H. (2009). Classification of Bacteria and Archaea: past, present and future. *Systematic and Applied Microbiology*, *32*, 533–542. doi:10.1016/j.syapm.2009.09.002.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J., & Weber, C. F. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, *75*, 7537–7541. doi:10.1128/AEM.01541-09.

Schmoldt, S., Latzin, P., Heesemann, J., Griese, M., Imhof, A., & Hogardt, M. (2006). Clonal analysis of

Inquilinus limosus isolates from six cystic fibrosis patients and specific serum antibody response. *Journal of Medical Microbiology*, *55*, 1425–1433. doi:10.1099/jmm.0.46466-0.

Scott, D., & Ely, B. (2015). Comparison of genome sequencing technology and assembly methods for the analysis of a GC-rich bacterial genome. *Current Microbiology*, *70*, 338–344. doi:10.1007/s00284-014-0721-6.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, *30*, 2068–2069. doi:10.1093/bioinformatics/btu153.

Segonds, C., Heulin, T., Marty, N., & Chabanon, G. (1999). Differentiation of Burkholderia species by PCR-restriction fragment length polymorphism analysis of the 16S rRNA gene and application to cystic fibrosis isolates. *Journal of Clinical Microbiology*, *37*, 2201–2208.

Segonds, C., Thouverez, M., Barthe, A., Bossuet-Greif, N., Tisseyre, L., Plesiat, P., Vergnaud, G., Chabanon, G., & Pourcel, C. (2015). Development of a multiple-locus variable-number tandem-repeat typing scheme for genetic fingerprinting of Burkholderia cenocepacia and application to nationwide epidemiological analysis. *Journal of Clinical Microbiology*, *53*, 398–409. doi:10.1128/JCM.02473-14.

Seo, Y.-S., Lim, J., Choi, B.-S., Kim, H., Goo, E., Lee, B., Lim, J.-S., Choi, I.-Y., Moon, J. S., Kim, J., & Hwang, I. (2011). Complete genome sequence of Burkholderia gladioli BSR3. *Journal of Bacteriology*, *193*, 3149–3149. doi:10.1128/JB.00420-11.

Seo, Y.-S., Lim, J. Y., Park, J., Kim, S., Lee, H.-H., Cheong, H., Kim, S.-M., Moon, J. S., & Hwang, I. (2015). Comparative genome analysis of rice-pathogenic Burkholderia provides insight into capacity to adapt to different environments and hosts. *BMC Genomics*, *16*, 349. doi:10.1186/s12864-015-1558-5.

Sessitsch, A., Coenye, T., Sturz, A. V., Vandamme, P., Ait Barka, E., Salles, J. F., Van Elsas, J. D., Faure, D., Reiter, B., Glick, B. R., Wang-Pruski, G., & Nowak, J. (2005). Burkholderia phytofirmans sp. nov., a novel plant-associated bacterium with plant-beneficial properties. *International Journal of Systematic and Evolutionary Microbiology*, *55*, 1187–1192. doi:10.1099/ijs.0.63149-0.

Shade, A., Hogan, C. S., Klimowicz, A. K., Linske, M., McManus, P. S., & Handelsman, J. (2012). Culturing captures members of the soil rare biosphere. *Environmental Microbiology*, *14*, 2247–2252. doi:10.1111/j.1462-2920.2012.02817.x.

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. *International Journal of Systematic and Evolutionary Microbiology*, *62*, 1–35. doi:10.1099/ijs.0.037408-0.

Sheu, S.-Y., Chou, J.-H., Bontemps, C., Elliott, G. N., Gross, E., dos Reis, 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. *International Journal of Systematic and Evolutionary Microbiology*, *63*, 435–441. doi:10.1099/ijs.0.039859-0.

Shibata, T. F., Maeda, T., Nikoh, N., Yamaguchi, K., Oshima, K., Hattori, M., Nishiyama, T., Hasebe, M., Fukatsu, T., Kikuchi, Y., & Shigenobu, S. (2013). Complete genome sequence of Burkholderia sp. strain RPE64, bacterial symbiont of the bean bug Riptortus pedestris. *Genome Announcements*, *1*, e00441–13–e00441–13. doi:10.1128/genomeA.00441-13.

Shin, S. C., Ahn, D. H., Kim, S. J., Lee, H., Oh, T.-J., Lee, J. E., & Park, H. (2013). Advantages of single-molecule real-time sequencing in high-GC content genomes. *PLoS ONE*, *8*, e68824. doi:10.1371/journal.pone.0068824.

Sierra, G. (1957). A simple method for the detection of lipolytic activity of micro-organisms and some observations on the influence of the contact between cells and fatty substrates. *Antonie van Leeuwenhoek*, *23*, 15–22.

Silva, I. N., Santos, P. M., Santos, M. R., Zlosnik, J. E. A., Speert, D. P., Buskirk, S. W., Bruger, E. L., Waters, C. M., Cooper, V. S., & Moreira, L. M. (2016). Long-term evolution of Burkholderia multivorans during a chronic cystic fibrosis infection reveals shifting forces of selection. *mSystems*, *1*, e00029–16. doi:10.1128/mSystems.00029-16.

Sleator, R. D. (2011). Phylogenetics. *Archives of Microbiology*, *193*, 235–239. doi:10.1007/s00203-011-0677-x.

Sleator, R. D. (2013). A beginner's guide to phylogenetics. *Microbial Ecology*, *66*, 1–4. doi:10.1007/s00248-013-0236-x.

Souza Dias, M. B., Cavassin, L. G., Stempliuk, V., Xavier, L. S., Lobo, R. D., Sampaio, J. L. M., Pignatari, A. C., Borrasca, V. L., Bierrenbach, A. L., & Toscano, C. M. (2013). Multi-institutional outbreak of Burkholderia cepacia complex associated with contaminated mannitol solution prepared in compounding pharmacy. *American Journal of Infection Control*, *41*, 1038–1042. doi:10.1016/j.ajic.2013.01.033.

Speert, D. P., Henry, D., Vandamme, P., Corey, M., & Mahenthiralingam, E. (2002). Epidemiology of Burkholderia cepacia complex in patients with cystic fibrosis, Canada. *Emerging Infectious Diseases*, *8*, 181–187.

Spilker, T., Baldwin, A., Bumford, A., Dowson, C. G., Mahenthiralingam, E., & LiPuma, J. J. (2009). Expanded multilocus sequence typing for Burkholderia species. *Journal of Clinical Microbiology*, *47*, 2607–2610. doi:10.1128/JCM.00770-09.

Stackebrandt, E., & Ebers, J. (2006). Taxonomic parameters revisited: tarnished gold standards. *Microbiology Today*, *33*, 152–155.

Stackebrandt, E., Frederiksen, W., Garrity, G. M., Grimont, P. A. D., Kampfer, P., Maiden, M. C. J., Nesme, X., Rossello-Mora, R., Swings, J., Truper, H. G., Vauterin, L., Ward, A. C., & Whitman, W. B. (2002). Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *International Journal of Systematic and Evolutionary Microbiology*, *52*, 1043–1047. doi:10.1099/00207713-52-3-1043.

Stackebrandt, E., & Goebel, B. M. (1994). A place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *International Journal of Systematic Bacteriology*, *44*, 846–849.

Staley, C., Gould, T. J., Wang, P., Phillips, J., Cotner, J. B., & Sadowsky, M. J. (2016). Sediments and soils act as reservoirs for taxonomic and functional bacterial diversity in the upper Mississippi River. *Microbial Ecology*, *71*, 814–824. doi:10.1007/s00248-016-0729-5.

Staley, J. T., & Konopka, A. (1985). Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annual Review of Microbiology*, *39*, 321–346. doi:10.1146/annurev.mi.39.100185.001541.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*, 1312–1313. doi:10.1093/bioinformatics/btu033.

Stead, D. E. (1992). Grouping of plant-pathogenic and some other Pseudomonas spp by using cellular fatty-acid profiles. *International Journal of Systematic Bacteriology*, *42*, 281–295.

Steinert, M., Emody, L., Amann, R., & Hacker, J. (1997). Resuscitation of viable but nonculturable Legionella pneumophila Philadelphia JR32 by Acanthamoeba castellanii. *Applied and Environmental Microbiology*, *63*, 2047–2053.

Stoddard, S. F., Smith, B. J., Hein, R., Roller, B. R. K., & Schmidt, T. M. (2015). rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Research*, *43*, D593–D598. doi:10.1093/nar/gku1201.

Stokell, J. R., Gharaibeh, R. Z., Hamp, T. J., Zapata, M. J., Fodor, A. A., & Steck, T. R. (2015). Analysis of

changes in diversity and abundance of the microbial community in a cystic fibrosis patient over a multiyear period. *Journal of Clinical Microbiology*, *53*, 237–247. doi:10.1128/JCM.02555-14.

Stopnisek, N., Bodenhausen, N., Frey, B., Fierer, N., Eberl, L., & Weisskopf, L. (2014). Genus-wide acid tolerance accounts for the biogeographical distribution of soil Burkholderia populations. *Environmental Microbiology*, *16*, 1503–1512. doi:10.1111/1462-2920.12211.

Stopnisek, N., Zuhlke, D., Carlier, A., Barberan, A., Fierer, N., Becher, D., Riedel, K., Eberl, L., & Weisskopf, L. (2016). Molecular mechanisms underlying the close association between soil Burkholderia and fungi. *The ISME Journal*, *10*, 253–264. doi:10.1038/ismej.2015.73.

Storms, V., Van Den Vreken, N., Coenye, T., Mahenthiralingam, E., LiPuma, J. J., Gillis, M., & Vandamme, P. (2004). Polyphasic characterisation of Burkholderia cepacia-like isolates leading to the emended description of Burkholderia pyrrocinia. *Systematic and Applied Microbiology*, *27*, 517–526.

Storms, V., Van Den Vreken, N., Gillis, M., & Vandamme, P. (2002). Evaluation of tRNA intergenic length polymorphism (tDNA-PCR) for the differentiation of the members of the Burkholderia cepacia complex. *Systematic and Applied Microbiology*, *25*, 376–385. doi:10.1078/0723-2020-00126.

Straka, R. P., & Stokes, J. L. (1957). Rapid destruction of bacteria in commonly used diluents and its elimination. *Applied and Environmental Microbiology*, *5*, 21–24.

Suarez-Moreno, Z. R., Caballero-Mellado, J., Coutinho, B. G., Mendonca-Previato, L., James, E. K., & Venturi, V. (2012). Common features of environmental and potentially beneficial plant-associated Burkholderia. *Microbial Ecology*, *63*, 249–266. doi:10.1007/s00248-011-9929-1.

Sutcliffe, I. C., Trujillo, M. E., & Goodfellow, M. (2012). A call to arms for systematists: revitalising the purpose and practises underpinning the description of novel microbial taxa. *Antonie van Leeuwenhoek*, *101*, 13–20. doi:10.1007/s10482-011-9664-0.

Sutcliffe, I. C., Trujillo, M. E., Whitman, W. B., & Goodfellow, M. (2013). A call to action for the International Committee on Systematics of Prokaryotes. *Trends in Microbiology*, *21*, 51–52. doi:10.1016/j.tim.2012.11.004.

Tabacchioni, S., Chiarini, L., Bevivino, A., Cantale, C., & Dalmastri, C. (2000). Bias caused by using different isolation media for assessing the genetic diversity of a natural microbial population. *Microbial Ecology*, *40*, 169–176.

Tabacchioni, S., Ferri, L., Manno, G., Mentasti, M., Cocchi, P., Campana, S., Ravenni, N., Taccetti, G., Dalmastri, C., Chiarini, L., Bevivino, A., & Fani, R. (2008). Use of the gyrB gene to discriminate among species of the Burkholderia cepacia complex. *FEMS Microbiology Letters*, *281*, 175–182. doi:10.1111/j.1574-6968.2008.01105.x.

Tago, K., Kikuchi, Y., Nakaoka, S., Katsuyama, C., & Hayatsu, M. (2015). Insecticide applications to soil contribute to the development of Burkholderia mediating insecticide resistance in stinkbugs. *Molecular Ecology*, *24*, 3766–3778. doi:10.1111/mec.13265.

Takeshita, K., Shibata, T. F., Nikoh, N., Nishiyama, T., Hasebe, M., Fukatsu, T., Shigenobu, S., & Kikuchi, Y. (2014). Whole-genome sequence of Burkholderia sp. strain RPE67, a bacterial gut symbiont of the bean bug Riptortus pedestris. *Genome Announcements*, *2*, e00556–14–e00556–14. doi:10.1128/genomeA.00556-14.

Talavera, G., & Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology*, *56*, 564–577. doi:10.1080/10635150701472164.

Tamura, K. (1992). Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G + C-content biases. *Molecular Biology and Evolution*, *9*, 678–687.

Tamura, K., & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of

mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, *10*, 512–526.

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. *Molecular Biology and Evolution*, *28*, 2731–2739. doi:10.1093/molbev/msr121.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, *30*, 2725–2729. doi:10.1093/molbev/mst197.

Tayeb, L. A., Lefevre, M., Passet, V., Diancourt, L., Brisse, S., & Grimont, P. A. D. (2008). Comparative phylogenies of Burkholderia, Ralstonia, Comamonas, Brevundimonas and related organisms derived from rpoB, gyrB and rrs gene sequences. *Research in Microbiology*, *159*, 169–177. doi:10.1016/j.resmic.2007.12.005.

Thompson, C. C., Amaral, G. R., Campeao, M., Edwards, R. A., Polz, M. F., Dutilh, B. E., Ussery, D. W., Sawabe, T., Swings, J., & Thompson, F. L. (2015). Microbial taxonomy in the post-genomic era: rebuilding from scratch? *Archives of Microbiology*, *197*, 359–370. doi:10.1007/s00203-014-1071-2.

Tian, Y., Kong, B. H., Liu, S. L., Li, C. L., Yu, R., Liu, L., & Li, Y. H. (2013). Burkholderia grimmiae sp. nov., isolated from a xerophilous moss (Grimmia montana). *International Journal of Systematic and Evolutionary Microbiology*, *63*, 2108–2113. doi:10.1099/ijs.0.045492-0.

Tindall, B. J., Rossello-Mora, R., Busse, H.-J., Ludwig, W., & Kampfer, P. (2010). Notes on the characterization of prokaryote strains for taxonomic purposes. *International Journal of Systematic and Evolutionary Microbiology*, *60*, 249–266. doi:10.1099/ijs.0.016949-0.

Torbeck, L., Raccasi, D., Guilfoyle, D. E., Friedman, R. L., & Hussong, D. (2011). Burkholderia cepacia: this decision is overdue. *PDA Journal of Pharmaceutical Science and Technology*, *65*, 535–543. doi:10.5731/pdajpst.2011.00793.

Turton, J. F., Kaufmann, M. E., Mustafa, N., Kawa, S., Clode, F. E., & Pitt, T. L. (2003). Molecular comparison of isolates of Burkholderia multivorans from patients with cystic fibrosis in the United Kingdom. *Journal of Clinical Microbiology*, *41*, 5750–5754. doi:10.1128/JCM.41.12.5750-5754.2003.

Uroz, S., Calvaruso, C., Turpault, M.-P., & Frey-Klett, P. (2009). Mineral weathering by bacteria: ecology, actors and mechanisms. *Trends in Microbiology*, *17*, 378–387. doi:10.1016/j.tim.2009.05.004.

Uroz, S., & Oger, P. (2015). Draft genome sequence of Burkholderia sp. strain PML1(12), an ectomycorrhizosphere-inhabiting bacterium with effective mineral-weathering ability. *Genome Announcements*, *3*, e00798–15–e00798–15. doi:10.1128/genomeA.00798-15.

Uroz, S., Oger, P., Morin, E., & Frey-Klett, P. (2012). Distinct ectomycorrhizospheres share similar bacterial communities as revealed by pyrosequencing-based analysis of 16S rRNA Genes. *Applied and Environmental Microbiology*, *78*, 3020–3024. doi:10.1128/AEM.06742-11.

Ursing, J. B., Rossello-Mora, R. A., Garcia-Valdes, E., & Lalucat, J. (1995). Taxonomic note - a pragmatic approach to the nomenclature of phenotypically similar genomic groups. *International Journal of Systematic Bacteriology*, *45*, 604.

Ussery, D. W., Kiil, K., Lagesen, K., Sicheritz-Ponten, T., Bohlin, J., & Wassenaar, T. M. (2009). The genus Burkholderia: analysis of 56 genomic sequences. *Microbial Pathogenomics*, *6*, 140–157.

Valvano, M. A., Keith, K. E., & Cardona, S. T. (2005). Survival and persistence of opportunistic Burkholderia species in host cells. *Current Opinion in Microbiology*, *8*, 99–105. doi:10.1016/j.mib.2004.12.002.

Van, V. T., Berge, O., Ngo Ke, S., Balandreau, J., & Heulin, T. (2000). Repeated beneficial effects of rice inoculation with a strain of Burkholderia vietnamiensis on early and late yield components in low fertility sulphate acid soils of Vietnam. *Plant and Soil*, *218/2*, 273–284. doi:10.1023/A:1014986916913.

Van Borm, S., Buschinger, A., Boomsma, J. J., & Billen, J. (2002). Tetraponera ants have gut symbionts related to nitrogen-fixing root-nodule bacteria. *Proceedings of the Royal Society B*, *269*, 2023–2027. doi:10.1098/rspb.2002.2101.

Van Oevelen, S., De Wachter, R., Vandamme, P., Robbrecht, E., & Prinsen, E. (2002). Identification of the bacterial endosymbionts in leaf galls of Psychotria (Rubiaceae, angiosperms) and proposal of 'Candidatus Burkholderia kirkii' sp. nov. *International Journal of Systematic and Evolutionary Microbiology*, *52*, 2023–2027. doi:10.1099/00207713-52-6-2023.

Van Oevelen, S., De Wachter, R., Vandamme, P., Robbrecht, E., & Prinsen, E. (2004). 'Candidatus Burkholderia calva' and 'Candidatus Burkholderia nigropunctata' as leaf gall endosymbionts of African Psychotria. *International Journal of Systematic and Evolutionary Microbiology*, *54*, 2237–2239. doi:10.1099/ijs.0.63188-0.

Van Veen, S. Q., Claas, E. C. J., & Kuijper, E. J. (2010). High-throughput identification of bacteria and yeast by matrix-assisted laser desorption ionization-time of flight mass spectrometry in conventional medical microbiology laboratories. *Journal of Clinical Microbiology*, *48*, 900–907. doi:10.1128/JCM.02071-09.

Vandamme, P., & Dawyndt, P. (2011). Classification and identification of the Burkholderia cepacia complex: past, present and future. *Systematic and Applied Microbiology*, *34*, 87–95. doi:10.1016/j.syapm.2010.10.002.

Vandamme, P., De Brandt, E., Houf, K., Salles, J. F., van Elsas, J. D., Spilker, T., & LiPuma, J. J. (2013a). Burkholderia humi sp. nov., Burkholderia choica sp. nov., Burkholderia telluris sp. nov., Burkholderia terrestris sp. nov. and Burkholderia udeis sp. nov.: Burkholderia glathei-like bacteria from soil and rhizosphere soil. *International Journal of Systematic and Evolutionary Microbiology*, *63*, 4707–4718. doi:10.1099/ijs.0.048900-0.

Vandamme, P., Goris, J., Chen, W.-M., de Vos, P., & Willems, A. (2002a). Burkholderia tuberum sp. nov. and Burkholderia phymatum sp. nov., nodulate the roots of tropical legumes. *Systematic and Applied Microbiology*, *25*, 507–12. doi:10.1078/07232020260517634.

Vandamme, P., Henry, D., Coenye, T., Nzula, S., Vancanneyt, M., LiPuma, J. J., Speert, D. P., Govan, J. R., & Mahenthiralingam, E. (2002b). Burkholderia anthina sp. nov. and Burkholderia pyrrocinia , two additional Burkholderia cepacia complex bacteria, may confound results of new molecular diagnostic tools. *FEMS Immunology and Medical Microbiology*, *33*, 143–149. doi:10.1111/j.1574-695X.2002.tb00584.x.

Vandamme, P., Holmes, B., Coenye, T., Goris, J., Mahenthiralingam, E., LiPuma, J. J., & Govan, J. R. W. (2003). Burkholderia cenocepacia sp. nov.—a new twist to an old story. *Research in Microbiology*, *154*, 91–96. doi:10.1016/S0923-2508(03)00026-3.

Vandamme, P., Holmes, B., Vancanneyt, M., Coenye, T., Hoste, B., Coopman, R., Revets, H., Lauwers, S., Gillis, M., Kersters, K., & Govan, J. R. W. (1997). Occurrence of multiple genomovars of Burkholderia cepacia in cystic fibrosis patients and proposal of Burkholderia multivorans sp. nov. *International Journal of Systematic Bacteriology*, *47*, 1188–1200.

Vandamme, P., Mahenthiralingam, E., Holmes, B., Coenye, T., Hoste, B., De Vos, P., Henry, D., & Speert, D. P. (2000). Identification and population structure of Burkholderia stabilis sp. nov. (formerly Burkholderia cepacia genomovar IV). *Journal of Clinical Microbiology*, *38*, 1042–1047.

Vandamme, P., Moore, E. R. B., Cnockaert, M., Peeters, C., Svensson-Stadler, L., Houf, K., Spilker, T., & LiPuma, J. J. (2013b). Classification of Achromobacter genogroups 2, 5, 7 and 14 as Achromobacter insuavis sp. nov., Achromobacter aegrifaciens sp. nov., Achromobacter anxifer sp. nov. and Achromobacter dolens sp. nov., respectively. *Systematic and Applied Microbiology*, *36*, 474–482. doi:10.1016/j.syapm.2013.06.005.

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

Vandamme, P., Pot, B., Gillis, M., De Vos, P., Kersters, K., & Swings, J. (1996). Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiological Reviews*, *60*, 407–438.

Vandamme, P., Vancanneyt, M., Pot, B., Mels, L., Hoste, B., Dewettinck, D., Vlaes, L., Van Den Borre, C., Higgins, R., Hommez, J., Kersters, K., Butzler, J. P., & Goossens, H. (1992). Polyphasic taxonomic study of the emended genus Arcobacter with Arcobacter butzleri comb. nov. and Arcobacter skirrowii sp. nov., an aerotolerant bacterium isolated from veterinary specimens. *International Journal of Systematic Bacteriology*, *42*, 344–356. doi:10.1099/00207713-42-3-344.

Vandamme, P., Verheyde, B., Peeters, C., & Dawyndt, P. (2014). Genomic taxonomy and biodiversity of the Burkholderia cepacia complex. In T. Coenye, & E. Mahenthiralingam (Eds.), *Burkholderia: from genomes to function* chapter 1. (pp. 15–29). Norfolk, UK: Caister Academic Press. (1st ed.).

Vandamme, P. A., Peeters, C., Cnockaert, M., Gomila, M., Moore, E. R. B., Spilker, T., & LiPuma, J. J. (2016). Reclassification of Achromobacter spiritinus Vandamme et al. 2013 as a later heterotypic synonym of Achromobacter marplatensis Gomila et al. 2011. *International Journal of Systematic and Evolutionary Microbiology*, *66*, 1641–1644. doi:10.1099/ijsem.0.000872.

Vandamme, P. A., Peeters, C., Cnockaert, M., Inganas, E., Falsen, E., Moore, E. R. B., Nunes, O. C., Manaia, C. M., Spilker, T., & LiPuma, J. J. (2015). Bordetella bronchialis sp. nov., Bordetella flabilis sp. nov. and Bordetella sputigena sp. nov., isolated from human respiratory specimens, and reclassification of Achromobacter sediminum Zhang et al. 2014 as Verticia sediminum gen. nov., comb. nov. *International Journal of Systematic and Evolutionary Microbiology*, *65*, 3674–3682. doi:10.1099/ijsem.0.000473.

Vanlaere, E., Baldwin, A., Gevers, D., Henry, D., De Brandt, E., LiPuma, J. J., Mahenthiralingam, E., Speert, D. P., Dowson, C., & Vandamme, P. (2009). Taxon K, a complex within the Burkholderia cepacia complex, comprises at least two novel species, Burkholderia contaminans sp. nov. and Burkholderia lata sp. nov. *International Journal of Systematic and Evolutionary Microbiology*, *59*, 102–111. doi:10.1099/ijs.0.001123-0.

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 Microbiology Letters*, *249*, 303–307. doi:10.1016/j.femsle.2005.06.026.

Vanlaere, E., LiPuma, J. J., Baldwin, A., Henry, D., De Brandt, E., Mahenthiralingam, E., Speert, D., Dowson, C., & Vandamme, P. (2008a). 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. *International Journal of Systematic and Evolutionary Microbiology*, *58*, 1580–1590. doi:10.1099/ijs.0.65634-0.

Vanlaere, E., van der Meer, J. R., Falsen, E., Salles, J. F., de Brandt, E., & Vandamme, P. (2008b). Burkholderia sartisoli sp. nov., isolated from a polycyclic aromatic hydrocarbon-contaminated soil. *International Journal of Systematic and Evolutionary Microbiology*, *58*, 420–423. doi:10.1099/ijs.0.65451-0.

Varghese, N. J., Mukherjee, S., Ivanova, N., Konstantinidis, K. T., Mavrommatis, K., Kyrpides, N. C., & Pati, A. (2015). Microbial species delineation using whole genome sequences. *Nucleic Acids Research*, *43*, 6761–6771. doi:10.1093/nar/gkv657.

Vartoukian, S. R., Palmer, R. M., & Wade, W. G. (2010). Strategies for culture of 'unculturable' bacteria. *FEMS Microbiology Letters*, *309*. doi:10.1111/j.1574-6968.2010.02000.x.

Vasiljevic, Z. V., Novovic, K., Kojic, M., Minic, P., Sovtic, A., Djukic, S., & Jovcic, B. (2016). Burkholderia

cepacia complex in Serbian patients with cystic fibrosis: prevalence and molecular epidemiology. *European Journal of Clinical Microbiology and Infectious Diseases*, *35*, 1277–1284. doi:10.1007/s10096-016-2662-4.

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., Wu, D., Paulsen, I., Nelson, K. E., Nelson, W., Fouts, D. E., Levy, S., Knap, A. H., Lomas, M. W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y.-H., & Smith, H. O. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science*, *304*, 66–74. doi:10.1126/science.1093857.

Vermis, K., Brachkova, M., Vandamme, P., & Nelis, H. (2003a). Isolation of Burkholderia cepacia complex genomovars from waters. *Systematic and Applied Microbiology*, *26*, 595–600. doi:10.1078/072320203770865909.

Vermis, K., Coenye, T., Mahenthiralingam, E., Nelis, H. J., & Vandamme, P. (2002a). Evaluation of species-specific recA-based PCR tests for genomovar level identification within the Burkholderia cepacia complex. *Journal of Medical Microbiology*, *51*, 937–940.

Vermis, K., Vandamme, P. A. R., & Nelis, H. J. (2003b). Burkholderia cepacia complex genomovars: utilization of carbon sources, susceptibility to antimicrobial agents and growth on selective media. *Journal of Applied Microbiology*, *95*, 1191–1199. doi:10.1046/j.1365-2672.2003.02054.x.

Vermis, K., Vandekerckhove, C., Nelis, H. J., & Vandamme, P. A. R. (2002b). Evaluation of restriction fragment length polymorphism analysis of 16S rDNA as a tool for genomovar characterisation within the Burkholderia cepacia complex. *FEMS Microbiology Letters*, *214*, 1–5. doi:10.1111/j.1574-6968.2002.tb11316.x.

Verstraete, B., Janssens, S., Smets, E., & Dessein, S. (2013). Symbiotic ß-Proteobacteria beyond Legumes: Burkholderia in Rubiaceae. *PLoS ONE*, *8*. doi:10.1371/journal.pone.0055260.

Verstraete, B., Peeters, C., van Wyk, B., Smets, E., Dessein, S., & Vandamme, P. (2014). Intraspecific variation in Burkholderia caledonica: Europe vs. Africa and soil vs. endophytic isolates. *Systematic and Applied Microbiology*, *37*, 194–199. doi:10.1016/j.syapm.2013.12.001.

Verstraete, B., Van Elst, D., Steyn, H., Van Wyk, B., Lemaire, B., Smets, E., & Dessein, S. (2011). Endophytic bacteria in toxic South African plants: identification, phylogeny and possible involvement in Gousiekte. *PLoS ONE*, *6*, e19265. doi:10.1371/journal.pone.0019265.g001.

Vial, L., Chapalain, A., Groleau, M.-C., & Deziel, E. (2011). The various lifestyles of the Burkholderia cepacia complex species: a tribute to adaptation. *Environmental Microbiology*, *13*, 1–12. doi:10.1111/j.1462-2920.2010.02343.x.

Viallard, V., Poirier, I., Cournoyer, B., Haurat, J., Wiebkin, S., Ophel-Keller, K., & Balandreau, J. (1998). Burkholderia graminis sp. nov., a rhizospheric Burkholderia species, and reassessment of [Pseudomonas] phenazinium, [Pseudomonas] pyrrocinia and [Pseudomonas] glathei as Burkholderia. *International Journal of Systematic Bacteriology*, *48*, 549–563.

Vidal-Quist, J. C., O'Sullivan, L. A., Desert, A., Fivian-Hughes, A. S., Millet, C., Jones, T. H., Weightman, A. J., Rogers, H. J., Berry, C., & Mahenthiralingam, E. (2014). Arabidopsis thaliana and Pisum sativum models demonstrate that root colonization is an intrinsic trait of Burkholderia cepacia complex bacteria. *Microbiology*, *160*, 373–384. doi:10.1099/mic.0.074351-0.

Vonberg, R. P., & Gastmeier, P. (2007). Hospital-acquired infections related to contaminated substances. *Journal of Hospital Infection*, *65*, 15–23. doi:10.1016/j.jhin.2006.09.018.

Walker, A. W., Duncan, S. H., Louis, P., & Flint, H. J. (2014a). Phylogeny, culturing, and metagenomics of the human gut microbiota. *Trends in Microbiology*, *22*, 267–274. doi:10.1016/j.tim.2014.03.001.

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. a., Zeng, Q., Wortman, J., Young, S. K., & Earl, A. M. (2014b). Pilon: an integrated tool for comprehensive microbial

variant detection and genome assembly improvement. *PLoS ONE*, *9*, e112963. doi:10.1371/journal.pone.0112963.

Wang, L., Wang, M., Zhang, J., Wu, W., Lu, Y., & Fan, Y. (2015). An outbreak of Burkholderia stabilis colonization in a nasal ward. *International Journal of Infectious Diseases*, *33*, 71–74. doi:10.1016/j.ijid.2014.12.046.

Wang, X. Q., Liu, A. X., Guerrero, A., Liu, J., Yu, X. Q., Deng, P., Ma, L., Baird, S. M., Smith, L., Li, X. D., & Lu, S. E. (2016). Occidiofungin is an important component responsible for the antifungal activity of Burkholderia pyrrocinia strain Lyc2. *Journal of Applied Microbiology*, *120*, 607–618. doi:10.1111/jam.13036.

Warmink, J. A., Nazir, R., Corten, B., & van Elsas, J. D. (2011). Hitchhikers on the fungal highway: The helper effect for bacterial migration via fungal hyphae. *Soil Biology and Biochemistry*, *43*, 760–765. doi:10.1016/j.soilbio.2010.12.009.

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., Stackebrandt, E., Starr, M. P., & Truper, H. G. (1987). Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *International Journal of Systematic Bacteriology*, *37*, 463–464.

Weber, D. J., Rutala, W. A., & Sickbert-Bennett, E. E. (2007). Outbreaks associated with contaminated antiseptics and disinfectants. *Antimicrobial Agents and Chemotherapy*, *51*, 4217–4224. doi:10.1128/AAC.00138-07.

Welch, D. F., Muszynski, M. J., Pai, C. H., Marcon, M. J., Hribar, M. M., Gilligan, P. H., Matsen, J. M., Ahlin, P. A., Hilman, B. C., & Chartrand, S. A. (1987). Selective and differential medium for recovery of Pseudomonas cepacia from the respiratory tracts of patients with cystic fibrosis. *Journal of Clinical Microbiology*, *25*, 1730–1734.

Whitby, P. W., Carter, K. B., Hatter, K. L., LiPuma, J. J., & Stull, T. L. (2000). Identification of members of the Burkholderia cepacia complex by species-specific PCR. *Journal of Clinical Microbiology*, *38*, 2962–2965.

Whitby, P. W., Dick, H. L. N., Campbell, P. W., Tullis, D. E., Matlow, A., & Stull, T. L. (1998). Comparison of culture and PCR for detection of Burkholderia cepacia in sputum samples of patients with cystic fibrosis. *Journal of Clinical Microbiology*, *36*, 1642–1645.

Whiteford, M. L., Wilkinson, J. D., McColl, J. H., Conlon, F. M., Michie, J. R., Evans, T. J., & Paton, J. Y. (1995). Outcome of Burkholderia (Pseudomonas) cepacia colonisation in children with cystic fibrosis following a hospital outbreak. *Thorax*, *50*, 1194–1198. doi:10.1136/thx.50.11.1194.

Whitman, W. B. (2015). Genome sequences as the type material for taxonomic descriptions of prokaryotes. *Systematic and Applied Microbiology*, *38*, 217–222. doi:10.1016/j.syapm.2015.02.003.

Wieme, A. D., Spitaels, F., Aerts, M., De Bruyne, K., Van Landschoot, A., & Vandamme, P. (2014a). Effects of growth medium on matrix-assisted laser desorption-ionization time of flight mass spectra: a case study of acetic acid bacteria. *Applied and Environmental Microbiology*, *80*, 1528–1538. doi:10.1128/AEM.03708-13.

Wieme, A. D., Spitaels, F., Aerts, M., De Bruyne, K., Van Landschoot, A., & Vandamme, P. (2014b). Identification of beer-spoilage bacteria using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *International Journal of Food Microbiology*, *185*, 41–50. doi:10.1016/j.ijfoodmicro.2014.05.003.

Williams, J. G. K., Kubelik, A. R., Livak, K. J., Rafalski, J. A., & Tingey, S. V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research*, *18*, 6531–6535.

Winsor, G. L., Khaira, B., Van Rossum, T., Lo, R., Whiteside, M. D., & Brinkman, F. S. L. (2008). The Burkholderia Genome Database: facilitating flexible queries and comparative analyses. *Bioinformatics*, *24*,

2803–2804. doi:10.1093/bioinformatics/btn524.

Wolfgang, M. C., Kulasekara, B. R., Liang, X., Boyd, D., Wu, K., Yang, Q., Miyada, C. G., & Lory, S. (2003). Conservation of genome content and virulence determinants among clinical and environmental isolates of Pseudomonas aeruginosa. *Proceedings of the National Academy of Sciences*, *100*, 8484–8489. doi:10.1073/pnas.0832438100.

Xu, Y., Buss, E. A., & Boucias, D. G. (2016). Culturing and characterization of gut symbiont Burkholderia spp. from the Southern chinch bug, Blissus insularis (Hemiptera: Blissidae). *Applied and Environmental Microbiology*, *82*, 3319–3330. doi:10.1128/AEM.00367-16.

Yabuuchi, E., Kosako, Y., Oyaizu, H., Yano, I., Hotta, H., Hashimoto, Y., Ezaki, T., & Arakawa, M. (1992). Proposal of Burkholderia gen. nov. and transfer of seven species of the genus Pseudomonas homology group II to the new genus, with the type species Burkholderia cepacia (Palleroni and Holmes 1981) comb. nov. *Microbiology and Immunology*, *36*, 1251–1275.

Yoo, S.-H., Kim, B.-Y., Weon, H.-Y., Kwon, S.-W., Go, S.-J., & Stackebrandt, E. (2007). Burkholderia soli sp. nov., isolated from soil cultivated with Korean ginseng. *International Journal of Systematic and Evolutionary Microbiology*, *57*, 122–125. doi:10.1099/ijs.0.64471-0.

Zhang, H., Hanada, S., Shigematsu, T., Shibuya, K., Kamagata, Y., Kanagawa, T., & Kurane, R. (2000). Burkholderia kururiensis sp. nov., a trichloroethylene (TCE)-degrading bacterium isolated from an aquifer polluted with TCE. *International Journal of Systematic and Evolutionary Microbiology*, *50*, 743–749. doi:10.1099/00207713-50-2-743.

Zhang, L., & Xie, G. (2007). Diversity and distribution of Burkholderia cepacia complex in the rhizosphere of rice and maize. *FEMS Microbiology Letters*, *266*, 231–235. doi:10.1111/j.1574-6968.2006.00530.x.

Zhao, J., Schloss, P. D., Kalikin, L. M., Carmody, L. A., Foster, B. K., Petrosino, J. F., Cavalcoli, J. D., VanDevanter, D. R., Murray, S., Li, J. Z., Young, V. B., & LiPuma, J. J. (2012). Decade-long bacterial community dynamics in cystic fibrosis airways. *Proceedings of the National Academy of Sciences*, *109*, 5809–5814. doi:10.1073/pnas.1120577109.

Zhu, B., Ibrahim, M., Cui, Z., Xie, G., Jin, G., Kube, M., Li, B., & Zhou, X. (2016). Multi-omics analysis of niche specificity provides new insights into ecological adaptation in bacteria. *The ISME Journal*, (pp. 1–4). doi:10.1038/ismej.2015.251.

Zolg, W., & Ottow, J. C. G. (1975). Pseudomonas glathei sp. nov., a new nitrogen-scavenging rod isolated from acid lateritic relicts in Germany. *Zeitschrift für allgemeine Mikrobiologie*, *15*, 287–299. doi:10.1002/jobm.19750150410.

Zolin, A., McKone, E., & van Rens, J. (2014). *ECFSPR Annual Report 2010*. Technical Report ECFS.

Zuleta, L. F. G., Cunha, C. d. O., de Carvalho, F. M., Ciapina, L. P., Souza, R. C., Mercante, F. M., de Faria, S. M., Baldani, J. I., Straliotto, R., Hungria, M., & de Vasconcelos, A. T. R. (2014). The complete genome of Burkholderia phenoliruptrix strain BR3459a, a symbiont of Mimosa flocculosa: highlighting the coexistence of symbiotic and pathogenic genes. *BMC Genomics*, *15*, 535. doi:10.1186/1471-2164-15-535.

Zurita, J., Mejia, L., Zapata, S., Trueba, G., Vargas, A. C., Aguirre, S., & Falconi, G. (2014). Healthcare-associated respiratory tract infection and colonization in an intensive care unit caused by Burkholderia cepacia isolated in mouthwash. *International Journal of Infectious Diseases*, *29*, 96–99. doi:10.1016/j.ijid.2014.07.016.

Curriculum Vitae

PERSONAL INFORMATION

Name	Charlotte Peeters
Address	Spinmolenplein 53, 9000 Gent
Phone	+32 (0) 473 91 16 57
Email	charlotte.peeters@ugent.be
Nationality	Belgian
Date of birth	February 25, 1988

WORK EXPERIENCE

2011-now	PhD student Biochemistry and Biotechnology, LM-UGent, Ghent University Dissertation: '*Burkholderia multivorans* and beyond: epidemiology, genomics and taxonomy of opportunistic pathogens in cystic fibrosis patients' Promoter: Prof. Dr. Peter Vandamme

EDUCATION

2009-2011	Master of Science in Biochemistry and Biotechnology Specialization: Microbial Biotechnology Ghent University - Summa cum laude
	Dissertation: 'Polyphasic identification of *Staphylococcus* species isolated from raw milk cheeses: MALDI-TOF MS as an alternative identification tool' Promoters: Prof. Dr. Peter Vandamme and Dr. Koenraad Van Hoorde
2006-2009	Bachelor in Biochemistry and Biotechnology Ghent University - Magna cum laude

TRAINING

2015	Visit Cooper Lab, University of New Hampshire, NH, USA for comparative genome analysis (January-February 2015)
2014	Crash course in R for microbial ecologists, Patrick Schloss, Detroit, USA (July 8-10, 2014)
	COST training school ES1103: Linking microbial ecology & evolution, Berlin, Germany (June 24-27, 2014)
2013	COST training school ES1103: Bioinformatics for Microbial Community Analysis, Liverpool, United Kingdom (December 11-14, 2013)
	Getting Started with High-performance Computing part I: Unix/Linux, Ghent University, Belgium (May 29-31, 2013)
2012	Basic training workshop on BioNumerics and GelCompar II, Applied Maths, Sint-Martens-Latem, Belgium (May 21-22, 2012)
2011-2016	Ghent University Doctoral Training Program, organized by Doctoral Schools of Life Sciences and Medicine: Advanced academic English writing skills, Personal effectiveness

SKILLS & EXPERIENCE

Social and organizational skills

- Active participation in outreach activities and science popularization ('Dag van de Wetenschap' 2012, 2013 and 2015)
- Coordination of practical course Microbiology of 3rd Bachelor in Biochemistry and Biotechnology

Computer skills

- Microsoft Windows and Office (Word, Excel, Powerpoint, Access)
- Linux and command-line interface
- R software for data analysis and visualization
- Basic programming skills in Microsoft Office Visual Basic

Educational skills

- Supervision of master thesis students Biochemistry and Biotechnology (Jessy Praet and Eliza Depoorter) and Pharmaceutical sciences (Charlotte Schelfhout)
- Supervision of students during practical courses in Microbiology of 2nd and 3rd Bachelor in Biochemistry and Biotechnology
- Supervision of lab technicians in the context of the National Reference Center for *Burkholderia cepacia* complex bacteria from Belgian cystic fibrosis patients
- Guidance and participation in isolation study of *Burkholderia* bacteria from South-African soil samples (Brecht Verstraete, KU Leuven)
- Guidance and participation in polyphasic taxonomic study of *Burkholderia* bacteria isolated from Argentinian soil samples (Walter Draghi, University of La Plata, Argentina)
- Guidance and participation in study of clinical *Burkholderia* isolates from Cambodia (Birgit De Smet, ITG Antwerpen)
- Supervision of study of *Burkholderia* and *Achromobacter* isolates from Brazilian cystic fibrosis patients (Carolina Capizzani, University of São Paulo, Brazil)

Services

- Identification and typing of clinical isolates for the National Reference Center for *Burkholderia cepacia* complex bacteria from Belgian cystic fibrosis patients
- Curator of the Bcc PubMLST database (pubmlst.org/bcc)
- Design and implementation of a central database for all data related to isolates, DNA extracts, whole-genome sequences, etc. in LM-UGent
- Implementation of a pipeline for whole-genome sequencing for the National Reference Center for *Burkholderia cepacia* complex bacteria from Belgian cystic fibrosis patients

SCIENTIFIC OUTPUT

2016 **Peeters, C.**, Daenekindt, S., & Vandamme, P. (2016a). PCR detection of *Burkholderia multivorans* in water and soil samples. BMC Microbiology, 16, 1–7.

Peeters, C., Meier-Kolthoff, J. P., Verheyde, B., De Brandt, E., Cooper, V. S. & Vandamme, P. (2016b). Phylogenomic study of *Burkholderia glathei*-like organisms, proposal of 13 novel *Burkholderia* species and emended descriptions of *Burkholderia sordidicola*, *Burkholderia zhejiangensis* and *Burkholderia grimmiae*. Frontiers in Microbiology, 7, 1-19.

Peeters, C., Depoorter, E., Praet, J. & Vandamme, P. (2016). Extensive cultivation of soil and water samples yields various pathogens in patients with cystic fibrosis but not *Burkholderia multivorans*. Journal of Cystic Fibrosis, in press.

Vandamme, P. A., **Peeters, C.**, Inganas, E., Cnockaert, M., Houf, K., Spilker, T., Moore, E. R. B. & LiPuma, J. J. (2016b). Taxonomic dissection of *Achromobacter denitrificans* Coenye et al. 2003 and proposal of *Achromobacter agilis* sp. nov., nom. rev., *Achromobacter pestifer* sp. nov., nom. rev., *Achromobacter kerstersii* sp. nov. and *Achromobacter deleyi* sp. nov. International Journal of Systematic and Evolutionary Microbiology, in press.

Depoorter, E., Bull, M. J., **Peeters, C.**, Coenye, T., Vandamme, P. & Mahenthiralingam, E. (2016). *Burkholderia*: an update on taxonomy and biotechnological potential as antibiotic producers. Applied Microbiology and Biotechnology, 100, 5215–5229.

Vandamme, P. A., **Peeters, C.**, Cnockaert, M., Gomila, M., Moore, E. R. B., Spilker, T. & LiPuma, J. J. (2016). Reclassification of *Achromobacter spiritinus* Vandamme et al. 2013 as a later heterotypic synonym of *Achromobacter marplatensis* Gomila et al. 2011. International Journal of Systematic and Evolutionary Microbiology, 66, 1641–1644.

2015 Vandamme, P. A., **Peeters, C.**, Cnockaert, M., Inganas, E., Falsen, E., Moore, E. R. B., Nunes, O. C., Manaia, C. M., Spilker, T. & LiPuma, J. J. (2015). *Bordetella bronchialis* sp. nov., *Bordetella flabilis* sp. nov. and *Bordetella sputigena* sp. nov., isolated from human respiratory specimens, and reclassification of *Achromobacter sediminum* Zhang et al. 2014 as *Verticia sediminum* gen. nov., comb. nov. International Journal of Systematic and Evolutionary Microbiology, 65, 3674–3682.

De Smet, B., Mayo, M., **Peeters, C.**, Zlosnik, J. E. A., Spilker, T., Hird, T. J., LiPuma, J. J., Kidd, T. J., Kaestli, M. & other authors. (2015). *Burkholderia stagnalis* sp. nov. and *Burkholderia territorii* sp. nov., two novel *Burkholderia cepacia* complex species from environmental and human sources. International Journal of Systematic and Evolutionary Microbiology, 65, 2265–2271.

2014 Vandamme, P. & **Peeters, C.** (2014). Time to revisit polyphasic taxonomy. Antonie van Leeuwenhoek, 106, 57–65.

Verstraete, B., **Peeters, C.**, van Wyk, B., Smets, E., Dessein, S. & Vandamme, P. (2014). Intraspecific variation in *Burkholderia caledonica*: Europe vs. Africa and soil vs. endophytic isolates. Systematic and Applied Microbiology, 37, 194–199.

Draghi, W. O., **Peeters, C.**, Cnockaert, M., Snauwaert, C., Wall, L. G., Zorreguieta, A. & Vandamme, P. (2014). *Burkholderia cordobensis* sp. nov., from agricultural soils. International Journal of Systematic and Evolutionary Microbiology, 64, 2003–2008.

Vandamme, P., Verheyde, B., **Peeters, C.** & Dawyndt, P. (2014). Genomic taxonomy and biodiversity of the *Burkholderia cepacia* complex. In *Burkholderia*: from

genomes to function, pp. 15–29. Edited by T. Coenye & E. Mahenthiralingam. Caister Academic Press.

2013 **Peeters, C.**, Zlosnik, J. E. A., Spilker, T., Hird, T. J., LiPuma, J. J. & Vandamme, P. (2013). *Burkholderia pseudomultivorans* sp. nov., a novel *Burkholderia cepacia* complex species from human respiratory samples and the rhizosphere. Systematic and Applied Microbiology, 36, 483–489.

De Smet, B., Veng, C., Kruy, L., Kham, C., van Griensven, J., **Peeters, C.**, Ieng, S., Phe, T., Vlieghe, E. & other authors. (2013). Outbreak of *Burkholderia cepacia* bloodstream infections traced to the use of Ringer lactate solution as multiple-dose vial for catheter flushing, Phnom Penh, Cambodia. Clinical Microbiology and Infection, 19, 832–837.

Vandamme, P., Moore, E. R. B., Cnockaert, M., **Peeters, C.**, Svensson-Stadler, L., Houf, K., Spilker, T. & LiPuma, J. J. (2013). Classification of *Achromobacter* genogroups 2, 5, 7 and 14 as *Achromobacter insuavis* sp. nov., *Achromobacter aegrifaciens* sp. nov., *Achromobacter anxifer* sp. nov. and *Achromobacter dolens* sp. nov., respectively. Systematic and Applied Microbiology, 36, 474–482.

CONFERENCE PRESENTATIONS

2015 Oral presentation at the International *Burkholderia cepacia* Working Group Meeting, Vancouver, Canada (April 2015).

2014 Oral presentation at the International *Burkholderia cepacia* Working Group Meeting, Nimes, France (April 2014)

Poster and oral presentation at the 8th European Young Investigator Meeting, Paris, France (February 2014)

2013 Oral presentation at the International *Burkholderia cepacia* Working Group Meeting, Ann Arbor, USA (April 2013)

ADDITIONAL INFORMATION

Personality: eager to learn, motivated team-player, good social skills

Languages

Dutch: mother tongue

English: proficient user

French: basic user

Other: Driver's license B

Acknowledgments

Dankwoord

Het is goed geweest. Het waren vijf leuke, mooie, intense jaren.

Dit hele doctoraatstraject had voor mij iets weg van een 400m op de piste. Eén volle baanronde de benen vanonder je lijf spurten, verzuren maar blijven gaan om toch te proberen eervol te finishen. De eerste bocht (lezen en experimenteren): zo snel mogelijk op snelheid komen zonder onnodig krachten te verspillen. De eerste rechte lijn (op volle toeren): in een goed ritme, net niet voluit toch zo snel mogelijk vooruit. De laatste bocht (de mid-PhD crisis): je begint je af te vragen waarom je hieraan begonnen bent want het begint pijn te doen en de meet is nog ver. De laatste rechte lijn (de laatste loodjes): je lichaam schreeuwt stop maar je draait de gas volledig open want dit is de snelste weg naar de meet. Eens aan de meet, de voldoening dat je doorgezet hebt.
Net zoals vroeger tijdens een loopwedstrijd waren ook tijdens dit doctoraat de supporters onmisbaar. Bedankt aan iedereen die ik tegenkwam! Ik wil hierbij graag enkele mensen in het bijzonder bedanken. En omdat woorden hierbij veelal te kort schieten heb ik ze aangevuld met wat muziek.
Peter, een welgemeende dankuwel! Je hebt me als promotor begeleid maar ook losgelaten. Op eigen benen kon ik zelf dingen uitzoeken, op poten zetten, met mijn kop tegen de muur lopen en terug rechtkrabbelen. Bedankt voor de open discussies waardoor ik bij alles kon meedenken. Het was een mooi, afwisselend parcours en het heeft van mij niet alleen een betere wetenschapper maar ook een wijzere mens gemaakt. De IBCWG congressen waren stuk voor stuk fantastische belevenissen en een grote inspiratie om door te gaan, bedankt dat ik je hiernaartoe mocht vergezellen!

`No Ceiling - Eddie Vedder`

Tom, Aurélien, Kurt and Leo, thank you for being part of my examination committee. Your critical reading and inspiring comments made it possible to finish with satisfaction.

`Divenire - Ludovico Einaudi`

ISABEL, je was lang mijn partner in crime voor zowat alles en je was o zo veel een schitterend voorbeeld van hoe je bovenal mens kan blijven onder de nooit aflatende druk om dit doctoraat tot een goed einde te brengen. Onze avonturen in Liverpool ("do you know picnic?"), Berlijn (na look moet je appels eten) en Detroit (literatuurstudie aan het zwembad en fitness om 4u 's ochtends) zullen me altijd bijblijven! HELEN, we zijn samen begonnen en we deelden de gewoonte om zo efficiënt mogelijk (onmogelijke?) to do lijstjes af te werken. Onze vriendschap is er het zuiverste bewijs van dat samen in de shit zitten een sterke band schept! ANNELEEN, ik heb je pas echt goed leren kennen toen we samen al eens laat achterbleven op onze gang en het was de overuren meer dan waard! Samen mijten meppen en boeken invriezen, leve de chaos! ANNE, je was een fantastisch voorbeeld van hoe je ook een tomeloze inzet en energie aan de dag kunt leggen voor niet-doctoraat dingen en dit was voor mij een belangrijke inspiratie! Jullie zijn karaktervrouwtjes!

Human - The Killers
Up To You Now - Ben Harper & Relentless7
Pompeii - Bastille

JESSY en ELIZA, het doet me deugd te zien dat jullie elk jullie eigen weg gevonden hebben in dit wereldje. Jullie waren schatten van thesisstudenten! De laatste maanden waren jullie telkens weer een rots in de branding en ik hoop dat die spontane brainstormmomenten altijd zullen blijven komen!

It's Oh So Quiet - Björk
Bad Day - REM

BART, ik vind het nog steeds fantastisch wat er op het raakvlak tussen IT en biologie allemaal is ontstaan. Tijdens de vele uren miserie oplossen met BioNumerics en het op poten zetten van de masterDB is een vriendschap ontstaan die me de laatste maanden door veel miserie geholpen heeft. Ik voel me vereerd om meter te mogen zijn van Katinka!

Rigoureus - The Scene
Losing My Religion - REM

MAARTEN, je eindeloze mopjes, droge humor en relativeringsvermogen ("chill!") waren goud waard! Het is me een eer en genoegen jou te morgen opvolgen ik zal mijn uiterste best doen om je wijsheid door te geven! (Ik beloof hierbij ook plechtig om de kasten in de practicumzaal op orde te houden.)

Giorgio by Moroder - Daft Punk

Bram, ook jij merci voor de eeuwig aanhoudende stroom mopjes en onnozeliteiten. Het was heerlijk om tijdens een zotte brainstorm de zoveelste theorie te verzinnen omtrent de niet-cultivatie van onze beestjes. Guillaume, merci voor je oneindig enthousiasme tijdens de practica! Charles en alle anderen die een bijgedragen hebben aan de fantastische collegiale sfeer, merci! Dat er nog veel Friday After Works mogen plaatsvinden!

`Another Brick In The Wall - Pink Floyd`

Evie, we hebben samen de kar van het NRC getrokken en onze heerlijk directe no-nonsense communicatie heeft ervoor gezorgd dat we tussendoor ook nog eens goed konden lachen of een menselijk gesprek voeren. Je was vaak een reddingsboei als er weer eens chaos heerste... Evelien, ook jij bent nu deel van een bijster goeie tandem om de nooit aflatende stroom van NRC stammen het hoofd te bieden. Liesbeth, dat er nog veel smoelen mogen getrokken worden! Margo, Cindy, Bart (H.), Jeannine, Mario en alle andere collega's, ook jullie waren onmisbaar in dit geheel.

`Iedereen is van de wereld - The Scene`

Elien & Sarah, ik heb jullie als student leren kennen maar het is ondertussen duidelijk geworden dat onze vriendschap veel verder draagt dan onze studententijd. Ik kan me geen betere vriendinnen bedenken! Het was heerlijk om met jullie af en toe eens even dat doctoraat volledig te kunnen vergeten en te genieten van muziek, drank, shoppen en eten.

`You've Got The Love - Florence + The Machine`
`Nothing Really Ends - dEUS`
`Beat It - Michael Jackson`
`Freed from Desire - Gala`

Mama, papa, Nico en Kaatje. Ik kan jullie niet genoeg bedanken voor jullie onvoorwaardelijke steun en jullie onvermurwbare geloof in een goeie afloop. Jullie hebben me altijd geleerd mijn best te doen, los van punten of beloningen, en dit was dé sleutel om dit doctoraat tot een goed einde te brengen! Oma, jij was er het levende bewijs van dat je niet groot hoeft te zijn om sterk te zijn. Jullie hebben ons geleerd mooi en goed te leven, ook bij tegenslag. Laat die sproeten maar komen, ik ben er trots op! Maarten, het is ondertussen tien jaar geleden dat ik je terugvond in Gent en het is fantastisch om een grote broer te hebben. En Isabelle, ik had nooit gedacht er zo'n toffe schoonzus bij te krijgen, laat staan schatten van een neefje en nichtje! Robin, Thomas, Sofie en Aline, merci om deel uit te maken van onze complexe gezinnen, ik zou het voor geen geld hebben willen missen.

`Rattle That Lock - David Gilmour`

Godspeed - Ozark Henry
Aan de meet - Raymond van het Groenewoud
Corner Of The Earth - Jamiroquai
Maanlicht - Raymond van het Groenewoud

Tot slot, STIJN. Talloze keren heb je me aangemoedigd om te stoppen met werken, naar huis te komen, te slapen, op te staan, even niets te doen, te blijven liggen, even dat doctoraat te vergeten, of toch maar nog een beetje te proberen schrijven... Het was heerlijk om thuis te komen bij jou, en op slag alles te kunnen relativeren. Woorden schieten tekort om je voor alles te bedanken!

Personal Jesus - Depeche Mode
Private Investigations - Dire Straits
High Hopes - Pink Floyd
Better Together - Jack Johnson

Charlotte, Lot, Lottie, Lotte, Lotje, Charlie
Gent, 26 augustus 2016