Contents

List of Abbreviations

AAI	-	average amino acid identity
adj.	-	adjective
AFLP	-	Amplified fragment length polymorphism
ANI	-	average nucleotide identity
ATCC	-	American Type Culture Collection
atpA	-	gene encoding ATP synthase α subunit
atpD	-	gene encoding ATP synthase β subunit
BCC	-	Bacterial Culture Collection, Forestry and Agricultural Biotechnology Institute (FABI)
BCCM/LMG	-	Belgian Coordinated Collection of Microorganisms/Laboratory of Microbiology, Ghent University
BD	-	Bacterial Disease, Plant Pathogenic and Plant Protecting Bacteria (PPPPB) Culture Collection
Bp	-	base pair
° C	-	degrees Celsius
carA	-	gene encoding carbamoyl phosphate synthase
CCUG	-	Culture Collection, University of Göteborg
CDC	-	Centres for Disease Control
cm	-	centimetre
comb. nov.	-	combination nova
ΔT_m	-	thermal denaturation midpoint
dim.	-	diminutive
DKGA	-	2,5-diketo-D-gluconic acid
DNA	-	deoxyribonucleic acid
dNTP's	-	deoxynucleotide triphosphate
EMBL	-	European Molecular Biology Laboratory
ERIC	-	enterobacterial repetitive intergenic consensus
FABI	-	Forestry and Agricultural Biotechnology Institute
fem.	-	feminine

Fig.	-	figure
gen.	-	genitive
Gr.	-	Greek
groEL	-	gene encoding heat shock protein
GTR	-	general time reversible
gyrB	-	gene encoding DNA gyrase
HPLC	-	high performance liquid chromatography
HR	-	hypersensitivity reaction
hsp60	-	gene encoding heat shock protein 60
H_2S	-	Hydrogen sulphide
IJSEM	-	International Journal of Systematic and Evolutionary Microbiology
infB	-	gene encoding initiation translation factor 2
K3P	-	Kimura
KCN	-	Potassium cyanide
L.	-	Latin
LMG	-	Laboratory of Microbiology, Ghent University
µl	-	microlitre
µm	-	micrometre
µM	-	micromolar
M.L.	-	medieval Latin
MLSA	-	multilocus sequence analysis
MLST	-	multilocus sequence typing
mol %	-	moles percent guanosine plus cytosine
n.	-	noun
NCPPB	-	National Collection of Plant Pathogenic Bacteria
Neut.	-	neuter
N.L.	-	new Latin
NRF	-	National Research Foundation
ONPG	-	*o*-Nitrophenyl-β-D-galactopyranoside
PCR	-	polymerase chain reaction
RBR	-	relative binding ratio
recA	-	gene encoding recombinase A
recN	-	gene encoding recombinase N

rep-PCR	-	repetitive extragenic palindromic-PCR
rRNA	-	ribosomal ribonucleic acid
rpoB	-	gene encoding RNA polymerase β subunit
sp. nov.	-	species nova
ssp.	-	subspecies
ST	-	sequence type
subsp.	-	subspecies
THRIP	-	Technology and Human Resources for Industry Programme
TN93	-	Tamura-Nei
TPCP	-	Tree Protection Co-operative Programme
tRNA	-	transfer ribonucleic acid
TSI	-	triple sugar iron
tuf	-	gene encoding elongation factor
UPGMA	-	unweighted pair groups method using arithmetic average
U.S.A.	-	United States of America
v.	-	verb
V	-	volt

Prepared Manuscripts

C. L. Brady, I. Cleenwerck, S. N. Venter, M. Vancanneyt, J. Swings and T. A. Coutinho. (2008). Phylogeny and identification of *Pantoea* species associated with the environment, humans and plants based on multilocus sequence analysis (MLSA). Submitted to *Syst Appl Microbiol*

C. L. Brady, S. N. Venter, I. Cleenwerck, K. Engelbeen, M. Vancanneyt, J. Swings and T. A. Coutinho. (2008). *Pantoea vagens* sp. nov., *Pantoea eucalypti* sp. nov., *Pantoea deleyii* sp. nov. and *Pantoea anthophila* sp. nov., four novel species belonging to the Genus *Pantoea*. Prepared for *Int J Syst Evol Microbiol*

C. L. Brady, I. Cleenwerck, S. N. Venter, K. Engelbeen, P. de Vos and T. A. Coutinho. (2008). Description of four novel *Pantoea* species from human clinical samples, *Pantoea septica* sp. nov., *Pantoea eucrina* sp. nov., *Pantoea brenneri* sp. nov. and *Pantoea conspicua* sp. nov. Prepared for *Int J Syst Evol Microbiol*

C. L. Brady, S. N. Venter, I. Cleenwerck, K. Vandemeulebroecke, P. de Vos and T. A. Coutinho. (2008). Transfer of *Pantoea citrea*, *Pantoea punctata* and *Pantoea terrea* to the genus *Tatumella* emend. as *Tatumella citrea* comb. nov., *Tatumella punctata* comb. nov., and *Tatumella terrea* comb. nov. (Kageyama *et al*., 1992) and description of *Tatumella morbirosei* sp. nov. Prepared for *Int J Syst Evol Microbiol*

C. L. Brady, S. N. Venter, I. Cleenwerck, K. Engelbeen, P. de Vos, M. J. Wingfield, N. Telechea and T. A. Coutinho. (2008). Isolation of *Enterobacter cowanii* from *Eucalyptus* showing symptoms of bacterial blight and dieback in Uruguay. Prepared for *Lett Appl Microbiol*

CHAPTER 1

Introduction

The *Enterobacteriaceae* represents a diverse group of genera and species. Microorganisms placed in this family are typically associated with humans and disease. However, the majority of plant-pathogenic bacteria are also contained within the *Enterobacteriaceae*. Species belonging to this family are phenotypically and phylogenetically closely-related, which has lead to the incorrect identification of numerous strains and the creation of many taxonomic problems over the years. The *Erwinia herbicola-Enterobacter agglomerans* complex is a prime example of this predicament.

In 1972, the names *Enterobacter agglomerans* and *Erwinia herbicola* were synonomized and the epithet *agglomerans* employed for all members of the "*herbicola-lathyri*" bacteria (Ewing and Fife, 1972). According to Ewing and Fife, the "*herbicola-lathyri*" bacteria included *Erwinia herbicola*, *Erwinia lathyri*, *Erwinia ananas*, *Erwinia milletiae*, *Erwinia cassavae* and *Erwinia uredovora*. However, plant pathologists continued to use the name *Erwinia herbicola* whilst clinical microbiologists preferred *Enterobacter agglomerans*. Both names appeared on the Approved Lists of Bacterial Names in 1980 (Skerman *et al.*, 1980), resulting in many mis-identified strains and taxonomic problems. Species and strains belonging to this group of bacteria became known as the *Erwinia herbicola-Enterobacter agglomerans* complex.

The genus *Erwinia* has long been acknowledged as a depository for plant-associated and plant-pathogenic members of the family *Enterobacteriaceae* (Grimont and Grimont, 2006). Over the years many *Erwinia* species have been transferred to the genera *Enterobacter*, *Pectobacterium*, *Brenneria* and *Pantoea*, creating much confusion regarding the correct taxonomy of numerous species. The taxonomy of *Erwinia herbicola*, *Erwinia milletiae* and *Enterobacter agglomerans* was finally resolved in the late 1980's and these three species were transferred to the newly-created genus *Pantoea* as *Pantoea agglomerans* (Beji *et al*., 1988; Gavini *et al*., 1989). Several years later, *Erwinia ananas* and *Erwinia uredovora* were synonomized and transferred to *Pantoea* as *Pantoea ananas*, along with *Erwinia stewartii* which became

Pantoea stewartii (Mergaert *et al*., 1993). Subsequently the epithet *ananas* was changed to *ananatis* in agreement with the International Code of Nomenclature of Bacteria (Trüper and De' Clari, 1997). Despite the inclusion of the new combinations in Bergey's Manual of Systematic Bacteriology (Grimont and Grimont, 2005), incorrect nomenclature is still used repeatedly in literature maintaining the confusion created in the early 1970's.

A recent review of *Pantoea* indicated the existence of several hybridisation groups (DNA hybridization groups I, II, IV and V) from a study by Brenner *et al*. (1984) which, based on 16S rRNA and *rpoB* sequencing data, should be included in the genus. The review also stated that the genus *Pantoea* could be divided into two groups of species: the core *Pantoea* species including *P. agglomerans*, *P. dispersa*, *P. ananatis* and *P. stewartii* and the "Japanese" species with *P. citrea*, *P. punctata* and *P. terrea* (Grimont and Grimont, 2005). The taxonomic position of the "Japanese" species within the genus *Pantoea* was also questioned.

Species of the genus *Pantoea* are primarily known as plant pathogens or plant-associated bacteria. *P. agglomerans* causes disease on the plant hosts gypsophila, beet, onion and cotton (Cooksey, 1986; Burr *et al*., 1991; Medrano and Bell, 2007) and is also considered an opportunistic human pathogen. *P. ananatis* causes disease on a wide range of plant hosts including maize, rice, onion and melon (Goszczynska *et al*., 2007; Cother *et al*., 2004; Gitaitis and Gay, 1997; Wells *et al.*, 1987; Bruton *et al.*, 1991). Probably, the most well known disease caused by a *Pantoea* species (*P. stewartii* ssp. *stewartii*) is Stewart's vascular wilt of corn. This species is also the only quarantine pathogen in the genus *Pantoea* (Coplin *et al.*, 2002). *P. citrea* causes the discolouration of pineapple, known as pink disease, following the heating process of canning (Cha *et al*., 1997).

Species belonging to the genus *Pantoea* are ubiquitous and frequently isolated from the environment. Like other members of the *Enterobacteriaceae* such as *Enterobacter*, *Erwinia* and *Rahnella*, *Pantoea* species are often found as epiphytes or endophytes on a range of plant hosts. *P. ananatis* was recovered from 25 asymptomatic weed species and crop plants, including crabgrass, Texas millet and tall verbena and Bermuda grass, cowpea and soybean (Gitaitis *et al.*, 2002). This species was also the most frequently isolated endophyte from maize kernels (Rijavec *et al*., 2007) and papaya shoot tips (Thomas *et al*., 2007). *P. agglomerans* has been found to exist as an endophyte in grapevines (Bell *et al*., 1995), tangerine and sweet-orange plants (Elvira-Recuenco and van Vuurde, 2000), sweet potato stems (Asis and Adachi, 2003), carrots (Surette *et al*., 2003) and soybean root nodules (Li *et al*., 2008). *Pantoea* species have also been isolated as endophytes from sweet corn and cotton (McInroy and Kloepper, 1995) and wild strawberries (Kukkurainen *et al*., 2005).

Identification of species, previously belonging to the *Erwinia herbicola-Enterobacter agglomerans* complex, has long been based on phenotypic and biochemical characteristics. However, as there is a high degree of phenotypic similarity between genera of the *Enterobacteriaceae,* this has led to the misidentification of many strains. It has even been suggested that caution be exercised when identifying strains belonging to the former *Erwinia herbicola - Enterobacter agglomerans* complex, as well as *Pantoea* species, based solely on commercialized phenotypic identification systems (Gavini *et al*., 1989). In recent years identification of *Pantoea* strains has been based on PCR assays with species-specific primers, as is the case with *P. ananatis* (Gitaitis *et al*., 2002, Walcott *et al*., 2002), 16S rRNA sequencing (Cother *et al*., 2004, Coutinho *et al*., 2002, Schmid *et al*., 2002, Medrano and Bell, 2007), DNA-DNA hybridisation (Coutinho *et al*., 2002, Gavini *et al*., 1989, Kageyama *et al*., 1992, Mergaert *et al*., 1993) and Amplified Fragment Length Polymorphism (AFLP) analysis (Brady *et al*., 2007). However, species-specific PCR assays do not exist for all *Pantoea* species and 16S rRNA sequences cannot be relied upon to clearly delineate all *Pantoea* species. It has been observed, based on 16S rRNA sequencing, that strains can generally be assigned to the genus *Pantoea*, but often not to a specific species.

The first report of a *Pantoea* species causing disease on plant hosts in South Africa occurred in the late 1970's, when *P. agglomerans* was found to cause stalk and leaf necrosis of onion (Hattingh and Walters, 1981). Almost twenty years later, *P. ananatis* was identified as the causal agent of bacterial blight and dieback of *Eucalyptus* in South Africa (Coutinho *et al*., 2002). A similar disease was later noted on young *Eucalyptus* trees in Uganda, Argentina and Uruguay. The bacterial strains isolated from the diseased material were morphologically and phenotypically similar to

Pantoea. Subsequently, *P. ananatis* was detected in onion seed (Goszczynska *et al*., 2006) and found to cause brown stalk rot of maize in South Africa (Goszczynska *et al*., 2007). Isolated simultaneously with *P. ananatis* from maize, were additional *Pantoea* strains which also caused brown stalk rot. *Pantoea* strains were also isolated from diseased onion in South Africa and the U.S.A. but could not be identified based on partial 16S rRNA sequencing (Goszczynska *et al*., 2006).

The increasing appearance of *Pantoea* species and unidentified *Pantoea* pathogens, causing either new diseases or outbreaks in countries where they had not previously been recorded, highlights the need for a discriminatory technique to conclusively resolve the identity of such isolates worldwide. The technique should be rapid, readily available, inter-laboratory reproducible and should be able to categorically identify *Pantoea* strains to the species level, and also resolve the taxonomic framework of the genus *Pantoea* and groups from the former *Erwinia herbicola-Enterobacter agglomerans* complex. The confusing taxonomy of the genus *Pantoea* has made it difficult to correctly identify strains to the species level. Resolving the taxonomy of the genus would further assist the conclusive identification of environmental *Pantoea* strains.

Aim:

To examine the taxonomy of species within the genus *Pantoea* using a multigene approach

Objectives:

- To develop a rapid, molecular-based technique for the identification of all *Pantoea* strains
- To conclusively identity *Pantoea* strains from *Eucalyptus* and maize from South America and South Africa
- To determine if protein profile group VII (Beji *et al*., 1988) and DNA hybridization groups I, II, IV and V (Brenner *et al*., 1984) should be included in the genus *Pantoea*
- To resolve the taxonomic position of the "Japanese" *Pantoea* species, namely *P. citrea*, *P. punctata* and *P. terrea*

- To identify several non-pigmented, slime-producing endophytic strains isolated with the *Pantoea* strains from *Eucalyptus* in Uruguay

References

Asis, C. A. & Adachi, K. (2003). Isolation of endophytic diazotroph *Pantoea agglomerans* and nondiazotroph *Enterobacter asburiae* from sweetpotato stem in Japan. *Lett Appl Microbiol* **38,** 19-23.

Beji, A., Mergaert, J., Gavini, F., Izard, D., Kersters, K., Leclerc, H. & De Ley, J. (1988). Subjective synonymy of *Erwinia herbicola*, *Erwinia milletiae*, and *Enterobacter agglomerans* and redefinition of the taxon by genotypic and phenotypic data. *Int J Syst Bacteriol* **38,** 77-88.

Bell, C. R., Dickie, G. A., Harvey, W. L. G. & Chan, J. W. Y. F. (1995). Endophytic bacteria in grapevine. *Can J Microbiol* **41,** 46-53.

Brady, C., Venter, S., Cleenwerck, I., Vancanneyt, M., Swings, J. & Coutinho, T. (2007). A FAFLP system for the improved identification of plant-pathogenic and plant-associated species of the genus *Pantoea*. *Syst Appl Microbiol* **30,** 413-417.

Brenner, D. J., Fanning, G. R., Knutson, J. K. L., Steigerwalt, A. G. & Krichevsky, M. I. (1984). Attempts to classify Herbicola group-*Enterobacter agglomerans* strains by deoxyribonucleic acid hydridization and phenotypic tests. *Int J Syst Bacteriol* **34,** 45-55.

Bruton, B. D., Wells, J. M., Lester, G. E. & Patterson, C. L. (1991). Pathogenicity and characterization of *Erwinia ananas* causing a postharvest disease of cantaloup fruit. *Plant Dis* **75,** 180-183.

Burr, T. J., Katz, B. H., Abawi, G. S. & Crosier, D. C. (1991). Comparison of tumorigenic strains of *Erwinia herbicola* isolated from table beet with *E. h. gypsophilae*. *Plant Dis* **75,** 855-858.

Cha, J. -S., Pujol, C., Ducusin, A. R., Macion, E. A., Hubbard, C. H. & Kado, C. I. (1997). Studies on *Pantoea citrea*, the causal agent of pink disease of pineapple. *J Phytopathol* **145,** 313-319.

Cooksey, D. A. (1986). Galls of *Gypsophila paniculata* caused by *Erwinia herbicola*. *Plant Dis* **70,** 464-468.

Coplin, D. L., Majerczak, D. R., Zhang, Y., Kim, W. S., Jock, S. & Geider, K. (2002). Identification of *Pantoea stewartii* subsp. *stewartii* by PCR and strain differentiation by PFGE. *Plant Dis* **86,** 304-311.

Cother, E. J., Reinke, R., McKenzie, C., Lanoiselet, V. M. & Noble, D. H. (2004). An unusual stem necrosis of rice caused by *Pantoea ananas* and the first record of this pathogen on rice in Australia. *Austral Plant Pathol* **33,** 495-503.

Coutinho, T. A., Preisig, O., Mergaert, J., Cnockaert, M. C., Riedel, K. -H., Swings, J. & Wingfield, M. J. (2002). Bacterial blight and eieback of *Eucalyptus* species, hybrids, and clones in South Africa. *Plant Dis* **86,** 20-25.

Elvira-Recuenco, M. & van Vuurde, J. W. (2000). Natural incidence of endophytic bacteria in pea cultivars under field conditions. *Can J Microbiol* **46,** 1036-1041.

Ewing, W. H. & Fife, M. A. (1972). *Enterobacter agglomerans* (Beijerinck) comb. nov. (the Herbicola-Lathyri bacteria). *Int J Syst Bacteriol* **22,** 4-11.

Gavini, F., Mergaert, J., Beji, A., Mielcarek, C., Izard, D., Kersters, K. & de Ley, J. (1989). Transfer of *Enterobacter agglomerans* (Beijerinck 1888) Ewing and Fife 1972 to *Pantoea* gen. nov. as *Pantoea agglomerans* comb. nov. and description of *Pantoea dispersa* sp. nov. *Int J Syst Bacteriol* **39,** 337-345.

Gitaitis, R. D. & Gay, J. D. (1997). First report of leaf blight, seed stalk rot, and bulb decay of onion by *Pantoea ananatis* in Georgia. *Plant Dis* **81,** 1096.

Gitaitis, R., Walcott, R., Culpepper, S., Sanders, H., Zolobowska, L. & Langston, D. (2002). Recovery of *Pantoea ananatis*, causal agent of center rot of onion, from weeds and crops in Georgia, USA. *Crop Protection* **21,** 983-989.

Goszczynska, T., Moloto, V. M., Venter, S. N. & Coutinho, T. A. (2006). Isolation and identification of *Pantoea ananatis* from onion seed in South Africa. *Seed Sci Tech* **34,** 655-668.

Goszczynska, T., Botha, W. J., Venter, S. N. & Coutinho, T. A. (2007). Isolation and identification of the causal agent of brown stalk rot, a new disease of maize in South Africa. *Plant Dis* **91,** 711-718.

Grimont, P. A. D. & Grimont, F. (2005). Genus: *Pantoea* In Volume Two: The *Proteobacteria*, Part B: The *Gammaproteobacteria*. *In Bergey's Manual of Systematic Bacteriology*, pp. 713-720. Edited by D. J. Brenner, N. R. Krieg & J. T. Staley. New York: Springer.

Grimont, F. & Grimont, P. A. D. (2006). The Genus *Enterobacter*. *In The Prokaryotes: Proteobacteria: Gamma Subclass,* pp. 197-214. Edited by M. Dworkin, S. Falkow, E. Rosenberg, K. H. Schleifer & E. Stackebrandt. New York: Springer.

Hattingh, M. J. & Walters, D. F. (1981). Stalk and leaf necrosis of onion caused by *Erwinia herbicola*. *Plant Dis* **65,** 615-618.

Kageyama, B., Nakae, M., Yagi, S. & Sonoyama, T. (1992). *Pantoea punctata* sp. nov., *Pantoea citrea* sp. nov., and *Pantoea terrea* sp. nov. isolated from fruit and soil samples. *Int J Syst Bacteriol* **42,** 203-210.

Kukkurainen, S., Leino, A., Vahamiko, S., Karkkainen, H. R. & Ahanen, K. (2005). Occurrence and location of endophytic bacteria in garden and wild strawberry. *HortScience* **40,** 348-352.

Li, J. H., Wang, E. T., Chen, W. F. & Chen, W. X. (2008). Genetic diversity and potential for promotion of plant growth detected in nodule endophytic bacteria of soybean grown in Heilongjiang province of China. *Soil Biol Biochem* **40,** 238-246.

McInroy, J. A. & Kloepper, J. W. (1995). Survey of indigenous bacterial endophytes from cotton and sweet corn. *Plant Soil* **173,** 337-342.

Medrano, E. G. & Bell, A. A. (2007). Role of *Pantoea agglomerans* in opportunistic bacterial seed and boll rot of cotton (*Gossypium hirsutum*) grown in the field. *J Appl Microbiol* **102,** 134-143.

Mergaert, J., Verdonck, L. & Kersters, K. (1993). Transfer of *Erwinia ananas* (synonym, *Erwinia uredovora*) and *Erwinia stewartii* to the genus *Pantoea* emend. as *Pantoea ananas* (Serrano 1928) comb. nov. and *Pantoea stewartii* (Smith 1898) comb. nov., respectively, and description of *Pantoea stewartii* subsp. *indologenes* subsp. nov. *Int J Syst Bacteriol* **43,** 162-173.

Rijavec, T., Lapanje, A., Dermastia, M. & Rupnik, M. (2007). Isolation of bacterial endophytes from germinated maize kernels. *Can J Microbiol* **53,** 802-808.

Schmid, H., Schubert, S., Weber, C. & Bogner, J. R. (2003). Isolation of a *Pantoea dispersa*-like strain fron a 71-year-old woman with acute myeloid leukemia and multiple myeloma. *Infection* **31,** 66-67.

Skerman, V. B. D., McGowan, V. & Sneath, P. H. A. (1980). Approved lists of bacterial names. *Int J Syst Bacteriol* **30,** 225-420.

Surette, M. A., Sturz, A. V., Lada, R. R. & Nowak, J. (2003). Bacterial endophytes in processing carrots (*Daucus carota* L. var. sativus): their localization, population density, biodiversity and their effects on plant growth. *Plant Soil* **253,** 381-390.

Thomas, P., Kumari, S., Swarna, G. K. & Gowda, T. K. S. (2007). Papaya shoot tip associated endophytic bacteria isolated from in vitro cultures and host-endophyte interaction in vitro and in vivo. *Can J Microbiol* **53,** 380-390.

Truper, H. G. & De' Clari, L. (1997). Taxonomic note: necessary correction of specific epithets fromed as substantives (nouns) "in apposition". *Int J Syst Bacteriol* **47,** 908-909.

Walcott, R. R., Gitaitis, R. D., Castro, A. C., Sanders Jr., F. H. & Diaz-Perez, J. C. (2002). Natural infestation of onion seed by *Pantoea ananatis*, causal agent of center rot. *Plant Dis* **86,** 106-111.

Wells, J. M., Sheug, W. S., Ceponis, M. J. & Chen, T. A. (1987). Isolation and characterization of strains of *Erwinia ananas* from honeydew melons. *Phtypathology* **77,** 511-514.

CHAPTER 2

Classification and taxonomy of the *Enterobacteriaceae*, with a focus on the genus *Pantoea*

2.1 Introduction

The organisms described in the family *Enterobacteriaceae* are typically facultatively anaerobic, Gram-negative rods and most are motile by means of peritrichous flagella. They grow well at 37 ºC and are oxidase negative and catalase positive, with few exceptions (Brenner and Farmer, 2005). The majority of organisms placed in the *Enterobacteriaceae* are associated with the digestive tract and human disease, for example, *Escherichia*, *Salmonella* and *Shigella*. However, phytopathogenic bacteria, including *Erwinia*, *Brenneria*, *Pectobacterium*, *Dickeya* and *Pantoea*, are also found in this family. In the past 25 years, the number of genera and species within the family *Enterobacteriaceae* has increased exponentially. When the eighth edition of Bergey's Manual of Determinative Bacteriology was published in 1974, the *Enterobacteriaceae* consisted of 12 genera and 36 species (Buchanan and Gibbons, 1974). The latest edition of Bergey's Manual of Systematic Bacteriology describes 44 genera and 176 species (Brenner and Farmer, 2005). The members of this family can generally be separated into four categories, depending on where they are isolated from: 1) human pathogens, 2) phytopathogens, 3) insect pathogens, symbionts and endosymbionts and 4) environmental, industrial and animals. There are, however, some genera which can overlap several categories (Janda, 2006). One such genus is *Pantoea*, where the species are primarily known as phytopathogens, but are also regularly isolated from human and clinical samples and from the environment.

The more common enteric species of the family *Enterobacteriaceae* can be differentiated by phenotypic and biochemical tests, usually with commercialized identification systems. However, infrequently isolated species or environmental strains are more difficult to identify especially if they have an atypical biochemical profile or belong to a rare or novel species. Species belonging to the genus *Pantoea* are particularly difficult to identify, owing to high phenotypic similarity, a lack of distinguishing characteristics and a somewhat confusing taxonomy. Methods based on genotypic information such as 16S rRNA and protein-encoding gene sequencing, have been employed for differentiation of numerous members of the *Enterobacteriaceae*. By combining phenotypic and genotypic information in a polyphasic approach, the identification of closely related, or infrequently isolated, enterobacterial species has improved in recent years.

The aim of this review is to examine the current species definition, the techniques used for species delineation and their application in the taxonomic framework of the genus *Pantoea*.

2.2 The Genus *Pantoea*

The genus *Pantoea* was formed to accommodate two hydridization groups from the *Erwinia herbicola-Enterobacter agglomerans* complex that did not correlate with either *Erwinia* or *Enterobacter* (Gavini *et al*., 1989). The earliest reports of bacteria that were later included in the complex were isolated from plants, seeds and fruit and assigned the names *Bacterium herbicola aureum* (Düggeli, 1904 cited by Graham & Hodgkiss, 1967) and *Erwinia lathyri* (Manns & Daubenhaus, 1913 cited by Graham & Hodgkiss, 1967). The first recorded isolation of these bacteria from humans occurred in 1928 when strains were isolated from stool samples of patients suffering from typhoid fever and named *Bacterium typhi flavum* (Dresel & Stickl, 1928 cited by Graham & Hodgkiss, 1967). Another species which later joined the *Erwinia herbicola-Enterobacter agglomerans* complex was *Pseudomonas trifolii*, which became *Xanthomonas trifolii* or *Xanthomonas herbicola* (Hüss, 1907; James, 1955 cited by Graham & Hodgkiss, 1967). In 1964, it was suggested by Dye that *X. trifolii* and *E. lathyri* had similar morphological and biochemical characteristics and should be re-classified as *Erwinia herbicola* (Dye, 1964). Graham and Hodgkiss (1967) noted the similarities between *B. typhi flavum* and the chromagenic bacteria *E. herbicola*, *E. lathyri*, *E. ananas*, *E. cassavae*, *E. milletiae* and *E. uredovora*. In 1972, Ewing and Fife compared the "*herbicola-lathyri* bacteria" with isolates implicated in a nosocomial septicaemia outbreak in the U.S.A. in 1971 and proposed that all of those strains should be incorporated into the genus *Enterobacter* as *Enterobacter agglomerans* (Ewing & Fife, 1972). The epithet *agglomerans* (Beijerinck, 1888) having priority over *herbicola* and *trifolii*. The names *Erwinia herbicola* and *Enterobacter agglomerans* were both included in the Approved Lists of Bacterial

Names, resulting in general confusion regarding the correct taxonomy of these bacteria (Skerman *et al*., 1980). In the next 16 years, several studies were performed on the *Erwinia herbicola-Enterobacter agglomerans* complex in attempts to resolve the nomenclature of these strains (Gavini *et al*., 1983, Mergaert *et al*., 1983, Brenner *et al*., 1984, Verdonck *et al*., 1987, Beji *et al*., 1988). The most successful of these studies was by Brenner *et al*. (1984), who performed DNA-DNA hybridization on 124 strains belonging to the *Erwinia herbicola-Enterobacter agglomerans* complex. Ninety strains were divided into 13 hybridization groups (DNA hybridization groups I to XIII) and the remaining 34 strains did not fall into any group. This study paved the way for the description of several new species.

DNA hybridization group XIII contained strains received as *Erwinia herbicola* ssp. *herbicola*, *Erwinia lathyri*, *Erwinia milletiae* and *Xanthomonas trifolii* (Brenner *et al*. 1984). Type strains and reference strains of these species were later hybridized to the type strain of *Enterobacter agglomerans* (ATCC 27155^{T}) and demonstrated more than 90 % DNA homology. Based on this DNA hybridization data, as well as protein electropherograms and phenotypic data, the synonymy of *Erwinia herbicola*, *Erwinia milletiae* and *Enterobacter agglomerans* was proposed (Beji *et al*., 1988). In agreement with Ewing and Fife (1972), the epithet *agglomerans* had priority, but the placement of the species in a genus was undecided. A year later, a new genus *Pantoea* was proposed to contain the species *agglomerans* which included the synonyms *Erwinia herbicola* and *Erwinia milletiae* (Gavini *et al*., 1989). Also described was a new species, *Pantoea dispersa* containing strains belonging to DNA hybridization group III from Brenner *et al.* (1984).

In Japan in 1988, bacterial strains that produce 2,5-diketo-D-gluconic acid (DKGA) were isolated from fruit and soils samples. As these strains shared the general characteristics of the genus *Erwinia*, they were tentatively named "*Erwinia citreus*", "*Erwinia punctata*" and "*Erwinia terreus*" (Sonoyama *et al*., 1988). After further testing, it was concluded that these DKGA-producing strains belonged to the *Erwinia herbicola-Enterobacter agglomerans* complex, as they were phenotypically related to DNA hybridization groups II, II and IV of Brenner *et al*. (1984). Following DNA hybridization and further phenotypic tests, the DKGA-producing strains were described and classified in the genus *Pantoea* as *P. citrea, P. punctata* and *P. terrea* (Kageyama *et al.*, 1992).

A year later it was proposed to transfer *Erwinia ananas*, *Erwinia uredovora* and *Erwinia stewartii* to the genus *Pantoea* following DNA hybridization and protein profiling (Mergaert *et al.*, 1993). *E. ananas* and *E. uredovora* were shown to be subjective synonyms and united as a single species which was classified as *Pantoea ananas.* Several strains from Brenner's DNA hybridization group VI were found in the same protein profile groups as *Pantoea ananas*, resolving another group from the *Erwinia herbicola-Enterobacter agglomerans* complex. The epithet *ananas* was later corrected to *ananatis* in accordance with the International Code of Nomenclature of Bacteria (Trüper and De' Clari, 1997). Two separate subspecies were created within the species *Pantoea stewartii* (formerly *Erwinia stewartii*), *P. stewartii* subsp. *stewartii* and *P. stewartii* subsp. *indologenes* (Mergaert *et al.*, 1993). These two subspecies shared 60 – 83 % DNA homology but were considerably different in biochemical characteristics and fatty acid composition. Out of the 13 DNA hybridization groups of Brenner *et al*. (1984), three groups (DNA hybridization groups III, VI and XIII) have been conclusively classified as *Pantoea* species and four groups (DNA hybridization groups I, II, IV and V) provisionally assigned to the genus *Pantoea* (Grimont and Grimont, 2005). The remaining six DNA hybridization groups have been assigned to other genera within the *Enterobacteriaceae*.

Species of *Pantoea* are generally acknowledged as plant-associated bacteria and are widely distributed in the environment. The type species of the genus, *P. agglomerans*, has been found to cause crown and root gall disease of beet and gypsophila, leaf blight and bulb rot of onions, seed and boll rot of cotton and leaf blight and vascular wilt of maize and sorghum (Cooksey, 1986; Burr *et al*., 1991; Edens *et al*., 2006; Medrano and Bell, 2007; Morales-Valenzuela *et al*., 2007). *P. agglomerans* is also associated with human and clinical samples, and is regarded as a rare opportunistic pathogen (Bicudo *et al*., 2007; De Champs *et al*., 2000; Fulleron *et al*., 2007; Kratz *et al*., 2003; Lim *et al*., 2006). *P. dispersa* has been isolated from soil, plant surfaces, seed and humans (Gavini *et al*., 1989; Schmid *et al*., 2003). *P. citrea* and *P. punctata* have both been isolated from mandarin oranges, and *P. citrea* is the causal agent of pink disease of pineapple whilst *P. terrea* is found in soil (Kageyama *et al*., 1992; Cha *et al*., 1997).

P. ananatis is the most varied species in the genus, causing a variety of diseases on a diverse range of hosts. This bacterium has been identified as the causal agent of brown rot of pineapple fruitlets and soft rot of sugarcane, brown spot of honeydew melon, postharvest disease of cantaloupe fruit, leaf blight, seed stalk rot and bulb decay of onion, necrotic leaf blotch disease of sudangrass, leaf spot of maize, bacterial blight of *Eucalyptus*, stem necrosis of rice and brown stalk rot of maize (Serrano, 1928; Wells *et al*., 1987; Bruton *et al*., 1991; Gitaitis and Gay, 1997; Azad *et al.*, 2000; Paccola-Meirelles *et al*., 2001; Coutinho *et al*., 2002; Cother *et al*., 2004; Goszczynska *et al*., 2007). The causal agent of Stewart's vascular wilt of sweet corn is *P. stewartii* subsp. *stewartii* (Stewart, 1897) and *P. stewartii* subsp. *indologenes* has been linked with leaf spot of millet (Mergaert *et al*., 1993) and leaf blotch of sudangrass (Azad *et al*., 2000).

2.3 Species Definitions and Concepts

The past 20 years have seen many taxonomical rearrangements within the family *Enterobacteriaceae*, as well as an exponential increase in the number of genera and species described. This can be largely attributed to the advances in molecular microbiology, including PCR and sequencing. Species which were previously indiscernible in their phenotype have been shown to be phylogenetically unrelated. As the techniques used for species differentiation and description have improved over time, so the species concept for prokaryotes has developed. The original species definitions based on morphological characteristics have been proven to be inadequate, but improved definitions have been developed based on new information units such as chemotaxonomic markers, DNA properties and rRNA sequences. A prokaryotic species is presently defined as "a category that circumscribes a (preferably) genomically coherent group of individual isolates/strains sharing a high degree of similarity in (many) independent features, comparatively tested under highly standardized conditions" (Rosselló-Mora and Amann, 2001). Practically, a species can currently be defined as "a group of strains, including the type strain, sharing greater than 70 % DNA-DNA relatedness and with 5 °C or less ΔT_m" (Wayne *et al*., 1987).

There have been several concerns raised regarding the validity of a species definition based on DNA relatedness. These concerns include: a) DNA relatedness studies are subject to sampling bias, b) the species delineation cut-off of 70 % was calibrated empirically and does not correspond to a theory-based concept, c) results are not comparable due to different methods, and d) DNA relatedness tests are difficult and tedious to perform. The above issues were addressed by Brenner *et al*. (2005) in the latest edition of Bergey's Manual of Systematic Bacteriology, and it was concluded that the advantages of DNA-DNA relatedness far outweigh the limitations and also that the practical definition of a species may not be faultless but it is both reliable and stable. Despite this, there is still a call for a more rapid technique for species circumscription. At a meeting of the *ad hoc* committee for the re-evaluation of species definition in bacteriology, it was proposed that alternative genomic methods should be used to describe new species, as long as there is a sufficient degree of similarity between the technique used and DNA-DNA hybridization data (Stackebrandt *et al*., 2002). It was also stated that sequencing of protein-encoding genes shows promise in identification and definition of bacterial species, and that the sequencing data from a minimum of five genes would provide an informative level of phylogenetic data.

Recently, there has been much controversy and debate regarding which species concept is most applicable to bacteria. One of the most persistent arguments is that of the ecotype concept of bacterial species (Cohan, 2001). An ecotype is defined as a set of strains that are ecologically similar to each other, such that an adaptive mutant from one ecotype can out-compete to extinction all other individuals from the same ecotype. This concept is supported by Gevers *et al*. (2005), although it is suggested that defining species as ecotypes might be reserved for pathogenic bacteria, where the ecotype is obvious, and for undescribed species. For the present, the most widely applicable species concept for prokaryotes is the phylo-phenetic species concept and is circumscribed by three different approaches: 1) delineation of genomic boundaries following whole genome hybridization, 2) description of the phenotype of the taxon and 3) position of the taxon within a reconstructed genealogy (Rosselló-Mora and Amann, 2001). This current species concept is in agreement with a polyphasic approach to taxonomy, where phenotypic- and genomic-information are thoroughly investigated (Vandamme *et al*., 1996).

2.4 Phenotypic Information

The phenotype of a bacterium is described as the observable or measurable physical and biochemical characteristics, as a result of genotype and the environment (Sneath, 1989). Before molecular techniques became readily available, classification of bacteria was based solely on morphology, physiology and biochemistry. The morphology of a bacterium includes both cellular (shape, endospore, flagella, inclusion bodies and Gram staining) and colonial (colour, dimensions and form) characteristics. The physiological and biochemical features include growth at different temperatures and on different media, pH values, salt concentrations, atmospheric conditions, antimicrobial activity, expression of various enzymes and metabolism of compounds (Vandamme *et al.*, 1996). Determination of a prokaryotic phenotype cannot be based simply on the morphology of an organism as the majority of bacteria, especially the *Enterobacteriaceae*, lack complex morphological characteristics and do not enter life cycles with different morphological stages. Although time-consuming and tedious, the phenotypic properties of an organism are required to generate useful classification systems, if the procedures used are highly standardized. Phenotypic data is the basis for formal descriptions of species, subspecies and genera. When analyzing the phenotype of a prokaryotic species, it is important to select strains representative of the known diversity and environmental niches of the group studied. Therefore, it is necessary to include recent isolates, as well as type strains and reference strains from accredited culture collections (Rosselló-Mora and Amann, 2001).

Commercial identification systems, including API (bioMérieux), Microlog (Biolog) and Biotype (bioMérieux), are used frequently by most research groups to obtain phenotypic data. Because of the ease of application and interpretation of results, commercial identification systems have become increasingly popular. However, these commercial assays have several disadvantages. The majority of these systems have a reduced number of tests, which decreases the amount of phenotypic information available for species description. Identification of a species utilizing commercial systems is based on a computer-generated identification database, which means that the identification result is dependent on the quality of the database (Rosselló-Mora and Amann, 2001). In the case of the genus *Pantoea*, it has been suggested that caution be exercised when basing identification solely on commercialised phenotypic identification systems (Gavini *et al.*, 1989).

Phenotypic data can be subjected to numerical analysis, whereby the data is encoded as binary. Positive reactions are numbered 1 and negative reactions as 0, and a similarity coefficient is applied to the data, generating a dendrogram. This approach can prove useful for phenotypically distinct species. Several numerical analysis studies were performed on strains belonging to the *Erwinia herbicola-Enterobacter agglomerans* complex (Mergaert *et al.*, 1984; Verdonck *et al.*, 1987; Beji *et al.*, 1988) and while strains could be assigned to profile groups or phena, it was not possible to resolve any major taxonomical or phylogenetic issues at that time. These studies highlight a short-coming when analyzing the phenotype of a prokaryote: that the whole information potential of the genome is not expressed. Therefore, it is not possible to infer phylogeny through phenotype as the genes responsible for those characteristics may be exclusive to certain taxa. This reinforces the concept of polyphasic taxonomy, both phenotypic- and genomic-information are necessary for reliable differentiation and classification of bacteria.

2.5 Genomic Information

The genotype refers to the genetic information of an organism which acts together with environmental factors to determine phenotype. However, when referring to large amounts of genome information it is preferable to use the term genomic instead of genotype (Sneath, 1989). With the advances in molecular microbiology, many techniques have been developed aimed at retrieving genomic information. Amplified Fragment Length Polymorphism (AFLP), rep-PCR and sequencing being among the more commonly applied techniques. However, there is still one technique which is considered the “gold standard” for species delineation and description and to which all new techniques are compared, DNA-DNA hybridization.

2.5.1 DNA-DNA Hybridization

DNA-DNA hybridization is based on measuring the reassociation of a denatured mixture of DNAs incubated under stringent conditions. The DNAs from different organisms will form hybrid molecules depending of the similarity of their nucleotide sequences. The higher the genetic similarity between two organisms is, the higher their degree of reassociation. There have been several techniques developed for whole genome hybridization studies since the first attempt by Schildkraut *et al*. (1961), based on either free solution methods or fixed DNA methods. The technique currently most widely used is the microtitre plate method described by Ezaki *et al*. (1989). The target DNA is bound to a microtitre plate to which the test DNA, labelled with biotin, is added. The level of DNA-DNA binding is measured using fluorogenic or chemiluminescent substrates.

The degree of DNA relatedness can either be measured by the relative binding ratio (RBR) or the difference in thermal denaturation midpoint (ΔT_m). The RBR measures the relative amount of heterologous, double-stranded hybrid DNA compared to that of the homologous hybrid DNA of the reference strain. RBR is expressed as percentage similarity, considering that the reference genome hybridizes 100 % with itself. ΔT_m is a more reliable parameter for measuring the degree of DNA relatedness, as it is independent from the quantity and quality of the DNAs used in the hybridization. ΔT_m reflects the thermal stability of the DNA hybrids formed during hybridization, and is the difference between the melting temperature of a given homologous DNA and that of a hybrid DNA (Rosselló-Mora, 2006). It is advised that prokaryotic species should share more than 70 % DNA-DNA relatedness and 5 °C or less ΔT_m (Wayne *et al*., 1987).

Even though ΔT_m may be the more reliable parameter for determining DNA relatedness, the RBR is a frequently used parameter for species delineation and new species descriptions. It has been consistently stressed that the species delineation cut-off of 70 % is only a recommendation, and should not be used as a strict boundary (Rosselló-Mora, 2003). It was suggested that the definition of a species could be made more robust by stating that strains showing more than 80 % DNA-DNA similarity belong to one genomic species, whilst strains sharing less than 60 % DNA-DNA similarity do not belong to the same genomic species; and for strains showing any value between 60 and 80 %, the distribution of values within and between species should be studied carefully (Grimont, 1988). The importance of phenotypic data has also been emphasized and can essentially override the phylogenetic concept of species in exceptional cases (Wayne *et al*., 1987; Stackebrandt *et al*., 2002). Therefore, a nomenspecies can contain more than one genomic group, referred to as a genomovar or genomospecies, if strains within these groups are phenotypically similar (Ursing *et al*., 1995). This concept has been applied in several studies where strains are suspected as belonging to new species, but cannot be phenotypically differentiated. In which case they are either included in a single species and designated as genomovars, or described as genomospecies within the genus (Brenner *et al*., 1993; Brenner *et al*., 1998; Vandamme *et al*., 1998).

2.5.2 DNA Base Ratio

A second parameter recommended for species delineation and description is the determination of the DNA base ratio, or G + C content which is calculated as a percentage of G + C. The G + C contents for prokaryotes range from 20 to 80 mol %. The higher the difference in G + C content for two organisms, the less related they are. Organisms that differ by more than 10 mol % do not belong to the same genus, whilst a difference of 5 mol % indicates organisms which belong to different species (Rosselló-Mora and Amann, 2001).

2.5.3 Amplified Fragment Length Polymorphism

Amplified fragment length polymorphism (AFLP) is a genomic fingerprinting method based on the selective PCR amplification of restriction fragments from a total digestion of genomic DNA with two restriction endonucleases and can be applied to DNA from any origin (Vos *et al*., 1995). AFLP can be used for both identification and typing as it can discriminate to below the species level (Savelkoul *et al*., 1999). The technique consists of three steps: digestion of genomic DNA with two restriction endonucleases and ligation of double-stranded adaptors to the resulting restriction fragments; preamplification and selective amplification of restriction fragments with two sets of primers, complimentary to the ligated adaptors, and lastly, electrophoretic separation of the amplified products. AFLP analysis has been used extensively in the identification, classification and epidemiology of bacteria belonging to the *Enterobacteriaceae*, including *Escherichia coli*, *Salmonella*, *Erwinia*, *Pantoea* and *Klebsiella* (Arnold *et al*., 1999; Aarts *et al*., 1998; Avrova *et al*., 2002; Brady *et al*., 2007; Jonas *et al*., 2004).

Shortly after the development of the AFLP technique, a study was performed comparing genomic fingerprinting with DNA-DNA hybridization data using *Xanthomonas* as a model system (Rademaker *et al*., 2000). A high correlation was observed between AFLP clusters and DNA-DNA hybridization results, suggesting that genomic fingerprinting techniques reflect the genotypic, phylogenetic and taxonomic relationships of organisms. These conclusions were supported by several studies on a wide range of bacteria such as *Aeromonas*, *Acinetobacter*, *Agrobacterium* and *Bradyrhizobium* (Huys *et al*., 1996; Janssen *et al*., 1997; Mougel *et al*., 2002; Portier *et al*., 2006; Willems *et al*., 2001).

2.5.4 Repetitive Extragenic Palindromic-PCR

Repetitive extragenic palindromic-PCR (Rep-PCR) is based on amplification from the sites of repetitive extragenic palindromic (REP) elements, enterobacterial repetitive intergenic consensus (ERIC) sequences and BOX elements found at different positions on the bacterial genome. The term "rep-PCR" includes amplification of any of the repetitive elements. REP elements are 38 bp long sequences consisting of six degenerate positions and a five bp variable loop between each side of a conserved palindromic stem (Stern *et al*., 1984). ERIC sequences are 126 bp long elements which contain a highly conserved central inverted repeat and are located in extragenic regions of genomes of the *Enterobacteriaceae* (Hulton *et al*., 1991). The third set of repetitive elements, are the BOX elements which are less commonly used for genomic fingerprinting than REP and ERIC sequences. BOX elements are located within intergenic regions and are mosaic repetitive elements composed of various combinations of three subunit sequences referred to as boxA, boxB and boxC (Martin *et al*., 1992). Amplification of rep elements can be performed with a single primer, a single set of primers or multiple sets of primers.

Rep-PCR has also been widely applied to members of the *Enterobacteriaceae* such as *E.coli*, *Brenneria*, *Erwinia*, *Salmonella* and *Serratia* (dos Anjos Borges *et al*., 2003; Moretti *et al*., 2004; Norman *et al*., 2003; Weigel *et al*., 2004; Zhang *et al*., 2003) for genetic diversity, identification and characterization studies. Additionally, high congruence has been observed between rep-PCR and DNA-DNA hybridization data for rhizobia and *Xanthomonas* (Nick *et al*., 1999; Rademaker *et al*., 2000) and between rep-PCR and pulsed-field gel electrophoresis (PFGE) for *Salmonella* (Weigel *et al*., 2004).

2.5.5 16S rRNA Sequence Analysis

At the beginning of the 1970's, 16S rRNA sequence data was first used for comparative analyses of prokaryote phylogeny (Fox *et al*., 1980). Eventually all living organisms were divided into three primary domains, the Archeae, the Bacteria and the Eucarya, in a universal phylogenetic tree (Woese, 1987; Woese *et al*., 1990). In the years following this breakthrough, rRNA became the gene of choice for phylogeny and identification of prokaryotic species for a number of reasons. rRNA molecules are universally present and have a conserved function; they are easy to sequence and their many conserved regions allow rapid alignment; rRNA sequences are readily available from a number of web-based databases and it is often easier to identify bacteria by rRNA sequencing than by biochemical or physiological tests (Cilia *et al*., 1996).

In prokaryotes, the rRNA operon consists of three conserved genes: the small subunit 16S rRNA gene, the large subunit 23S rRNA genes and the 5S rRNA gene. Sequencing of the 5S rRNA gene was initially used for phylogenetic studies of the prokaryotes. However, it was soon found that the 16S rRNA gene contained more highly conserved regions interspersed with variable and hypervariable sequences making it easier to design universal primers, and the 5S rRNA gene contained relatively little information due to its short length. Although the 23S rRNA gene is phylogenetically more discriminatory than the 16S rRNA gene, it is twice the length and therefore less popular (Ludwig and Schleifer, 1999). Consequently, 16S rRNA sequencing has become one of the most widely used standard techniques in microbial taxonomy and is used frequently in new species descriptions.

It was noted that strains which share 70 % or more DNA-DNA relatedness, will typically have 16S rRNA sequence similarity greater than 97 % (Stackebrandt and Goebel, 1994). However, this is not to say that strains which have more than 97 % 16S rRNA sequence similarity will have 70 % or more DNA-DNA similarity. There are several groups of organisms which have almost identical 16S rRNA sequences, but have DNA hybridization values of less than 70 %, for example, *Bacillus globisporus* and *B. psychrophilus* (Fox *et al*., 1992). Because the 16S rRNA gene is highly conserved, there is no linear correlation between DNA-DNA hybridization values and 16S rRNA sequence similarity for closely related organisms (Grimont, 1988; Stackebrandt and Goebel, 1994). Therefore, 16S rRNA sequence data should never be used unaided for new species descriptions but rather as an integral component of a polyphasic approach.

16S rRNA sequencing has resolved the phylogenetic positions of a number of genera and species within the family *Enterobacteriaceae*. In 1998, Hauben *et al*. examined the phylogenetic position of phytopathogens within the *Enterobacteriaceae* using 16S rRNA sequence data. This study resulted in the division of the genus *Erwinia* into three genera: the true *Erwinia* genus, *Pectobacterium* and *Brenneria* (Hauben *et al*., 1998). A fourth genus, *Dickeya*, was later established to include two species previously belonging to *Pectobacterium* and *Brenneria* and four novel species (Samson *et al*., 2005). Several studies based on 16S rRNA sequencing have revealed the polyphyletic nature of the genera *Serratia* (Spröer *et al*., 1999; Dauga, 2002), *Klebsiella* and *Enterobacter* (Mollet *et al*., 1997; Hauben *et al*., 1998; Dauga, 2002). Despite the many advantages of 16S rRNA sequencing in both phylogeny and identification, it will never be the technique to replace DNA-DNA hybridization for species circumscription because of its lack of resolving power at the species level.

2.5.6 Protein-encoding Genes

Protein-encoding genes have two main advantages over rRNA genes for phylogenetic analyses: 1) protein-encoding genes evolve much faster than rRNA genes, making it easier to analyze closely related bacteria and 2) the alignment of protein-encoding genes can be made easier by using the translated amino acid sequence (Harayama and Kasai, 2006). As previously mentioned, it was recommended by the committee for the re-evaluation of the species definition that sequencing of protein-encoding genes could play an important role in the circumscription of species (Stackebrandt *et al*., 2002). The committee suggested that when evaluating this approach, organisms should be selected for which there is extensive DNA-DNA hybridization data and intraspecific diversity DNA profiles available. Such a study was undertaken by Zeigler (2003), who compared sequence data from 32 protein-encoding genes with the genomes of 44 bacteria that could be grouped into 16 genera. It was observed that DNA identity scores for eight of the 32 genes correlated strongly with the overall sequence identity scores for the genomes, making them outstanding candidates for a species prediction sequence set. *recN* was the candidate gene with the highest potential for predicting genome relatedness, whilst 16S rDNA was the poorest. The 16S rDNA gene was included in the analysis due to its extensive use in taxonomic studies, despite not encoding a protein. This study proved that it is possible to predict genome relatedness based on the sequences of protein-encoding genes and supported the recommendation of Stackebrandt *et al.* (2002). However, the requirement of a minimum of five genes may be excessive as Zeigler (2003) found that even single gene alignments could predict the overall genome similarity, although this could be improved by using two or three genes.

Zeigler (2003) identified candidate genes for a species prediction set by applying four criteria: 1) genes should be widely distributed among genomes, 2) each of the candidate genes should be single copy within a given genome, 3) individual gene sequences should be long enough to contain important phylogenetic information, but short enough to be sequenced economically and with ease and 4) gene sequences must predict whole genome relationships with precision and accuracy. However, it was noted by Santos and Ochman (2004) that there was no consensus among genes used for species prediction and no means to rapidly sequence uncharacterized bacterial species. They identified 143 genes which were present in single copy in 95 % of the bacterial genomes examined. Of the 143 genes, a set of only 39 included at least two highly conserved regions for primer design. Based on this set of genes, conserved primers with G + C rich clamps were designed for ten functionally diverse genes. Of the taxonomically diverse range of bacteria tested, 60 % were amplified with each primer pair, creating a starting point for the development of multigene schemes for numerous bacterial species.

An additional requirement for the selection of candidate genes for phylogenetic analysis is that the genes should not be subject to horizontal gene transfer (Yamamoto and Harayama, 1996). However, only a small percentage of genes are unlikely to have undergone horizontal gene transfer (Brown *et al*., 2001). It has been noted that the horizontal transfer of genes correlates strongly with gene function. Genes which are involved in transcription, translation and related processes (informational genes) are less likely to undergo horizontal gene transfer than genes involved in metabolic functions (operational genes) (Rivera *et al*., 1998). This may be due to informational genes being part of large complex systems, and interacting with other proteins, while operational genes are not (Jain *et al*., 1999).

Gevers *et al*. (2004) analyzed 106 bacterial genomes and found that many contained a significant number of paralogs (homologous genes within a genome belonging to the same gene family, created by duplication). Gene duplication is considered an important evolutionary step towards diversity in the metabolic function of an organism. The data revealed that the largest group of paralogs within genomes encodes ABC-type transporters, transcriptional regulators or dehydrogenases. This study also indicated a subset of gene families that is found in all 106 of the genomes analyzed which can be considered as core housekeeping genes. These genes encode ribosome proteins, translation elongation factors, tRNA-synthetases, ABC-type transporters, topoisomerases, polymerases and ATP/GTPases (Gevers *et al*., 2004). It is evident that many of the housekeeping genes in the bacterial core may have undergone a duplication event, meaning that they can be paralogs. Therefore, protein-encoding (or housekeeping) genes should be selected with care if they are to be used in phylogenetic analyses (Harayama and Kasai, 2006).

The theory that each sequence cluster containing a bacterial species should correspond to an ecologically distinct population has been strongly advocated by Cohan (1994). Palys *et al*. (1997) verified this theory by demonstrating that protein-encoding genes can be used successfully in classifying the ecological diversity of bacteria, and are more effective in doing so than DNA-DNA hybridization. This lead to a second study which proved that two ecologically distinct species of *Bacillus*, that are basically identical in their 16S rRNA sequences, could be easily separated on the basis of protein-encoding genes (Palys *et al*., 2000). It was also confirmed that the inability of 16S rRNA sequencing to distinguish the two species was as a result of the extremely slow rate of evolution of the 16S rRNA gene. In theory, ecologically distinct organisms will eventually diverge into separate clusters for any gene. However, in practice only protein-encoding genes evolve rapidly enough to differentiate between closely related organisms (Palys *et al*., 2000).

a) Single Gene Phylogeny

Single gene phylogeny is based on the sequence comparison of one highly variable gene. Several genes have shown a higher resolving power for species delineation compared to the 16S rRNA gene, one such gene being *rpoB*. The *rpoB* gene encodes for RNA polymerase β-subunit and has been used in several identification and phylogenetic studies in recent years. One of the first studies to examine the usefulness of *rpoB* sequence analysis for bacterial identification was by Mollet *et al*. (1997). It was found that partial *rpoB* sequence data was more discriminative than 16S rRNA sequence data for the majority of *Enterobacteriaceae* examined. Also, *rpoB* sequences could be used for phylogenetic analyses as it was demonstrated that the genus *Klebsiella* is polyphyletic (Mollet *et al*., 1997). Subsequently, a phylogenetic study was performed on the genus *Klebsiella* and with the help of *rpoB* sequence data, three species were transferred from *Klebsiella* to a new genus *Raoultella* (Drancourt *et al*., 2001). Following these studies, a new *Klebsiella* species, *K. singaporensis*, and three new *Enterobacter* species, *E. radicincitans*, *E. turicensis* and *E. helveticus*, were described using *rpoB* sequence data as a supporting phylogenetic parameter (Li *et al*., 2004; Kämpfer *et al*., 2005; Stephan *et al*., 2007).

gyrB, encoding DNA gyrase β-subunit, has also been used in numerous phylogenetic studies and has resolved the taxonomic positions of bacterial strains. The phylogenetic relationships of 49 *Acinetobacter* strains were determined based on *gyrB* sequence data, which showed a strong congruence with DNA-DNA hybridization data (Yamamoto *et al*., 1999). *gyrB* gene sequences were shown to differentiate the closely related bacteria *E.coli*, *Shigella* and *Salmonella*, and may also be an alternative to 16S rRNA for species classification (Fukushima *et al*., 2002). A study of the phylogeny of the *Enterobacteriaceae* utilizing the *gyrB* gene demonstrated the reliability of this gene for inferring intra- and intergenic relationships, especially within the genus *Serratia* (Dauga, 2002). Similar conclusions were drawn from a study of the genus *Aeromonas*, where *gyrB* sequence data could successfully identify and differentiate all known species within the genus (Yáñez *et al*., 2003). The *gyrB* gene has also proven more discriminatory than the 16S rRNA gene for differentiation of the *Bacillus anthracis-cereus-thuringiensis* group, although *gyrB* could not discriminate virulent strains from avirulent strains of *B. anthracis* (La Duc *et al*., 2004). Additional popular protein-encoding genes which have been used for phylogenetic analyses include *groEL* (Harada and Ishikawa, 1997; McGhee *et al*., 2002), *infB* (Hedegaard *et al*., 1999), *hsp60* (Iversen *et al*., 2004) and *atpA* (Naser *et al*., 2005a).

b) <u>Multigene Phylogeny and Multilocus Sequence Analysis</u>

Multigene phylogeny is based on the assumption that a higher resolving power for relationships between species can be achieved by examining more than one protein-encoding gene. By concatenating the individual protein-encoding gene sequences, an overall consensus phylogenetic tree can support or reject phylogenies observed using single genes. Concatenated alignments have been shown to construct highly robust universal trees (Brown *et al*., 2001). The phylogeny of species belonging to the family *Enterobacteriaceae* was examined using two protein-encoding genes, *tuf* and *atpD*. The concatenation of the two genes improved the bootstrap values and resolved some inconsistencies seen in the single gene trees (Paradis *et al*., 2005). A similar study determined the relationships of plant pathogenic enterobacteria using three protein-encoding genes, *atpD*, *carA* and *recA* (Young and Park, 2007). A consensus tree based on the concatenation of the three genes supported many of the phylogenetic arrangements seen in the single gene tree.

A new technique, multilocus sequence analysis (MLSA), encompasses the concept of multigene phylogeny (Nasser *et al.*, 2005b). The term MLSA was proposed to emphasize the distinction of this technique from multilocus sequence typing (MLST) (Gevers *et al.*, 2005). MLSA is based on the sequence comparison of several conserved protein-coding genes and can determine the diversity and phylogenetic relatedness between related taxa using dendrograms constructed from the sequence data. There are three major approaches for dendrogram, or tree, construction: distance matrix, maximum parsimony and maximum likelihood. These are based on evolutionary models which define probabilities for the transition from one nucleotide to another (Harayama and Kasai, 2006). Because the approaches are based on different models, it is unlikely that the tree topologies will be identical. To construct a more robust tree, it is recommended that the different methods be used together with calculations based on several data sets (Rosselló-Mora and Amann, 2001).

The first study to utilize MLSA was aimed at members of the *Vibrionaceae* and was found to be more discriminatory among species than 16S rRNA sequences (Thompson *et al.*, 2005). MLSA was also applied to species of *Enterococcus* and showed great promise (Naser *et al.*, 2005b). The same MLSA scheme was applied to the genus *Lactobacillus* (Naser *et al.*, 2007), and led to the descriptions of several new species and reclassifications in both *Lactobacillus* and *Enterococcus*. Since the emergence of MLSA as a technique to rival 16S rRNA sequencing, it has been used in an increasing number of diversity, taxonomic and phylogenetic studies.

MLSA was used to examine the diversity between dairy and non-dairy *Lactococcus lactis* isolates using six protein-encoding genes. The study revealed two major genomic lineages within *Lactococcus lactis* (Rademaker *et al.*, 2007). Numerous taxonomic studies and novel species descriptions on a wide range of bacteria have been based on MLSA in the last three years. Included among them are species belonging to the genera *Borrelia* (Richter *et al.*, 2006; Postic *et al.*, 2007), *Enterococcus* (Naser *et al.*, 2005c) and *Leuconostoc* (De Bruyne *et al.*, 2007). In addition MLSA has proven useful for phylogenetic studies, which in some cases has lead to the revision of taxa. A MLSA scheme for the rhizobia provided support for the merging of *Ensifer* and *Sinorhizobium* into a single genus (Martens *et al.*, 2007; Martens *et al.*, 2008). Two genomovars could be distinguished within *Flavobacterium* *columnare* by MLSA, which also demonstrated the host-specific nature of the strains (Olivares-Fuster *et al*., 2007). MLSA of a serogroup of *E. coli* revealed that the isolates could be separated into four separate evolutionary clusters within the *E.coli* phylogeny and were related to enteropathogenic *E. coli* and enterohemorrhagic *E.coli* (Tarr *et al*., 2008).

c) Multilocus Sequence Typing

Multilocus sequence typing (MLST) goes one step further than MLSA and can group strains into the major genetic lineages within a species (Cooper and Feil, 2004). MLST was developed to characterize isolates of *Neisseria meningitidis* into the major meningococcal lineages (Maiden *et al*., 1998). MLST is based on multilocus enzyme electrophoresis (MLEE), where the alleles at each of seven house-keeping loci are assigned directly by nucleotide sequencing instead of comparing the variation in the electrophoresis of the enzymes they encode (Maiden *et al*., 1998). For each gene fragment, every unique sequence is designated a different allele, even those sequences that differ by a single nucleotide. The alleles present at each of the seven house-keeping genes are combined into an allelic profile and are assigned a sequence type (ST) designation (Maiden *et al*., 1998). By comparing the allelic profiles of isolates, epidemiological relationships can be determined. Closely related strains will have identical ST's (or ST's that differ at one or two loci), whereas unrelated isolates will have unrelated ST's (Urwin and Maiden, 2003).

In the design of a new MLST system, several factors must be taken into account. A diverse collection of 100 isolates should be used in the initial evaluation. House-keeping gene sequences of ± 450 bp are good targets for MLST, as these size fragments can be sequenced accurately on both strands using a single pair of primers. Although seven genes are generally used in MLST, it is recommended to examine additional candidate loci as the genes may experience levels of recombination or selection. In the design of oligonucleotide primers, a nested strategy should be used where the DNA fragments amplified are longer than required for the final sequences (Urwin and Maiden, 2003). Also, the house-keeping genes selected for the MLST system should be conserved, widely separated on the bacterial chromosome and should not be adjacent to genes which may be under selective pressure.

MLST has typically been used over the years for long term epidemiological and intraspecies diversity studies. Schemes have been developed for numerous medically important bacteria, including *N. meningitidis*, *Streptococcus pneumoniae*, *Staphylococcus aureus*, *Streptococcus pyogenes* and *Haemophilus influenzae* (Enright and Spratt, 1999) and for two plant pathogens, *Pseudomonas syringae* (Sarkar and Guttman, 2004) and *Xylella fastidiosa* (Scally *et al.*, 2005). A major benefit of MLST is that allelic profiles can be compared to those available on the extensive databases accessible via the internet (Aanensen and Spratt, 2005). Additionally, software is available for determining the genetic and evolutionary relationships between bacteria using MLST data (Feil and Enright, 2004; Spratt *et al*., 2004).

2.5.7 Genome-based Phylogeny

The entire bacterial genome is obviously the most complete source of genomic information. Large-scale sequencing of genomes is becoming increasingly routine resulting in a deluge of genetic information now available for making phylogenetic inferences. Ideally, whole-organism phylogenies could be constructed by comparing the overall similarity between entire genomes, where the degree of similarity depends on the fraction of genes shared (Francino *et al*., 2006). The majority of genomic studies to date have focussed on the comparison of genome-based phylogenies with those obtained by 16S rRNA sequencing, rather than the differences between taxa (Konstantinidis and Tiedje, 2005a). Genomic signatures, obtained from bacterial whole genome sequences, were compared with 16S rRNA sequence similarity and DNA-DNA hybridization data. A high correlation was observed between genomic signatures and DNA-DNA hybridization values, but the overall correlation between genome signatures and 16S rRNA sequence similarity was low, except between closely related organisms (Coenye and Vandamme, 2004).

Even with the advances made in large-scale sequencing, it is still not feasible to sequence the entire genomes of all bacterial species for phylogenetic studies. The best alternative is to identify genes which are representative of the whole genome. This was achieved by Ziegler (2003), who demonstrated that certain gene sequences diverge at a rate that reflects the overall rate of genome divergence and identified genes that could predict whole genome relatedness. Two novel parameters, average nucleotide identity (ANI) and average amino acid identity (AAI) have been developed by Konstantinidis and Tiedje (2005a; 2005b) to measure the whole genome relatedness between strains. Pairwise, whole genome comparisons were performed on closely related bacterial strains to determine the conserved protein-encoding genes and strain-specific genes. The ANI of the shared genes between two strains was found to be a robust parameter for comparing genetic relatedness and showed a strong correlation to DNA-DNA hybridization values. ANI values of 94 % corresponded to the suggested 70 % DNA-DNA hybridization standard (Konstantinidis and Tiedje, 2005b). AAI also offers high resolving power within species, as values between 95-96 % correspond to the 70 % DNA-DNA hybridization standard (Konstantinidis and Tiedje, 2005a). Additionally, phylogenetic trees constructed using AAI were congruent with those based on concatenated sequences of conserved genes within the genomes examined.

2.6 Conclusions:

For the time being, the current species definition of a prokaryotic species is practical and reliable (Brenner *et al.*, 2005). There are possible alternatives to DNA-DNA hybridization for the circumscription of species, the most promising being MLSA. However, for MLSA to replace DNA-DNA hybridization a universal set of genes must be identified which could be applied to the classification all of all prokaryotes (Gevers *et al.*, 2005). Until an alternative standardized method for bacterial classification is available, it will still be considered standard practice to confirm new species descriptions with DNA-DNA hybridization data.

It is evident that the taxonomy of the genus *Pantoea* must be updated, especially as there is still reference to "*Erwinia herbicola*" and "*Enterobacter agglomerans*" in recent literature. Due to the short-comings of 16S rRNA sequencing, this gene is only suitable for assigning isolates to the genus level, and is therefore not an option for a taxonomic study of the genus *Pantoea*. In contrast MLSA appears to be a robust, reliable method for taxonomic and phylogenetic studies. The availability of high throughput sequencing and the ease with which primers can be designed, make this technique particularly attractive compared to the more laborious fingerprinting methods of rep-PCR and AFLP.

Once the taxonomy of the genus *Pantoea* has been resolved, the MLSA scheme could be expanded to examine the epidemiological issues of species within the genus. Little is known concerning the specificity and genetic relatedness of *Pantoea* strains, especially those from different hosts and varied geographical regions like *P. agglomerans* and *P. ananatis*. Additionally the epidemiology of the diseases caused by these pathogens is unclear. MLST has been proven to be an excellent tool for studying the global, or long term, epidemiology of both medical and plant-pathogenic bacteria. An improved understanding of the taxonomy and epidemiology of species belonging to the genus *Pantoea* could hopefully lead to control measures for the diseases caused by them.

2.7 References

Aanensen, D. M. & Spratt, B. G. (2005). The multilocus sequence typing network: mlst.net. *Nucleic Acids Res* **33,** W728-33.

Aarts, H. J. M., Van Lith, L. A. J. T. & Keijer, J. (1998). High-resolution genotyping of *Salmonella* strains by AFLP-fingerprinting. *Lett Appl Microbiol* **26,** 131-135.

Arnold, C., Metherell, L., Clewley, J. P. & Stanley, J. (1999). Predictive modelling of fluorescent AFLP: a new approach to the molecular epidemiology of *E. coli*. *Res Microbiol* **150,** 33-44.

Avrova, A. O., Hyman, L. J., Toth, R. L. & Toth, I. K. (2002). Application of amplified fragment length polymorphism fingerprinting for taxonomy and identification of the soft eot bacteria *Erwinia carotovora* and *Erwinia chrysanthemi*. *Appl Environ Microbiol* **68,** 1499-1508.

Azad, H. R., Holmes, G. J. & Cooksey, D. A. (2000). A new leaf blotch disease of sudangrass caused by *Pantoea ananas* and *Pantoea stewartii*. *Plant Dis* **84,** 973-979.

Beji, A., Mergaert, J., Gavini, F., Izard, D., Kersters, K., Leclerc, H. & De Ley, J. (1988). Subjective synonymy of *Erwinia herbicola*, *Erwinia milletiae*, and *Enterobacter agglomerans* and redefinition of the taxon by genotypic and phenotypic data. *Int J Syst Bacteriol* **38,** 77-88.

Bicudo, E. L., Macedo, V. O. & Carrara, M. A. (2007). Nosocomial outbreak of *Pantoea agglomerans* in a pediatric urgent care center. *Braz J Infect Dis* **11,** 281-284.

Brady, C., Venter, S., Cleenwerck, I., Vancanneyt, M., Swings, J. & Coutinho, T. (2007). A FAFLP system for the improved identification of plant-pathogenic and plant-associated species of the genus *Pantoea*. *Systematic and Applied Microbiology*, **30,** 413-417.

Brenner, D. J. & Farmer, J. J. I. (2005). Order XIII: "Enterobacteriales" *In Bergey's Manual of Systematic Bacteriology, Volume Two, The Proteobacteria. Part B, The Gammaproteobacteria*. pp. 587-607. Edited by D. J. Brenner, N. R. Krieg & J. T. Staley. Springer, New York.

Brenner, D. J., Staley, J. T. & Krieg, N. R. (2005). Classification of Procaryotic organisms and the concept of bacterial speciation. *In Bergey's Manual of Systematic Bacteriology, The Archaea and the Deeply Branching and Phototrophic Bacteria*, pp. 27-31. Edited by D. R. Boone, R. W. Castenholz & G. M. Garrity. New York: Springer.

Brenner, D. J., Fanning, G. R., Knutson, J. K. L., Steigerwalt, A. G. & Krichevsky, M. I. (1984). Attempts to classify Herbicola group-*Enterobacter agglomerans* strains by deoxyribonucleic acid hydridization and phenotypic tests. *Int J Syst Bacteriol* **34,** 45-55.

Brenner, D. J., Grimont, P. A. D., Steigerwalt, A. G., Fanning, G. R., Ageron, E. & Riddle, C. F. (1993). Classification of citrobacteria by DNA hybridization: Designation of *Citrobacter farmeri* sp. nov., *Citrobacter youngae* sp. nov., *Citrobacter braakii* sp. nov., *Citrobacter werkmanii* sp. nov., *Citrobacter sedlakii* sp. nov., and three unnamed *Citrobacter* genomospecies. *Int J Syst Bacteriol* **43,** 645-658.

Brenner, D. J., Muller, H. E., Steigerwalt, A. G., Whitney, A. M., O'Hara, C. M. & Kampfer, P. (1998). Two new *Rahnella* genomospecies that cannot be phenotypically differentiated from *Rahnella aquatilis*. *Int J Syst Bacteriol* **48,** 141-149.

Brown, J. R., Douady, C. J., Italia, M. J., Marshall, W. E. & Stanhope, M. J. (2001). Universal trees based on large combined protein sequence data sets. *Nat Genet* **28,** 281.

Bruton, B. D., Wells, J. M., Lester, G. E. & Patterson, C. L. (1991). Pathogenicity and characterization of *Erwinia ananas* causing a postharvest disease of cantaloup fruit. *Plant Dis* **75,** 180-183.

Buchanan, R. E. & Gibbons, W. E. (1974). *Bergey's Manual of Determinative Bacteriology*, 8th edn. Baltimore: The Williams & Wilkins Co.

Burr, T. J., Katz, B. H., Abawi, G. S. & Crosier, D. C. (1991). Comparison of tumorigenic strains of *Erwinia herbicola* isolated from table beet with *E. h. gypsophilae*. *Plant Dis* **75,** 855-858.

Cha, J. -S., Pujol, C., Ducusin, A. R., Macion, E. A., Hubbard, C. H. & Kado, C. I. (1997). Studies on *Pantoea citrea*, the causal agent of pink disease of pineapple. *J Phytopathol* **145,** 313-319.

Cilia, V., Lafay, B. & Christen, R. (1996). Sequence heterogeneities among 16S ribosomal RNA sequences, and their effect on phylogenetic analyses at the species level. *Mol Biol Evol* **13,** 451-461.

Coenye, T. & Vandamme, P. (2004). Use of the genomic signature in bacterial classification and identification. *Syst Appl Microbiol* **27,** 175-185.

Cohan, F. M. (1994). The effects of rare but promiscuous genetic exchange on evolutionary divergence in prokaryotes. *Am Nat* **143,** 965-986.

Cohan, F. M. (2001). Bacterial Species and Speciation. *Syst Biol* **50,** 513-524.

Cooksey, D. A. (1986). Galls of *Gypsophila paniculata* caused by *Erwinia herbicola*. *Plant Dis* **70,** 464-468.

Cooper, J. E. & Feil, E. J. (2004). Multilocus sequence typing-what is resolved? *Trends Microbiol* **12,** 373-377.

Cother, E. J., Reinke, R., McKenzie, C., Lanoiselet, V. M. & Noble, D. H. (2004). An unusual stem necrosis of rice caused by *Pantoea ananas* and the first record of this pathogen on rice in Australia. *Austral Plant Pathol* **33,** 495-503.

Coutinho, T. A., Preisig, O., Mergaert, J., Cnockaert, M. C., Riedel, K. -H., Swings, J. & Wingfield, M. J. (2002). Bacterial blight and dieback of *Eucalyptus* species, hybrids, and clones in South Africa. *Plant Dis* **86,** 20-25.

Dauga, C. (2002). Evolution of the *gyrB* gene and the molecular phylogeny of *Enterobacteriaceae*: a model molecule for molecular systematic studies. *Int J Syst Evol Microbiol* **52,** 531-547.

De Bruyne K., Vandamme, P., Schillinger, U., Caroline, L., Boehringer, B., Franz, C. M. A. P., Cleenwerck, I., Vancanneyt, M. & De, V. L. (2007). *Leuconostoc holzapfelii* sp. nov., isolated from Ethiopian coffee fermentation and assessment of sequence analysis of housekeeping genes from delineation of *Leuconostoc* species. *Int J Syst Evol Microbiol* **57,** 2952-2959.

De Champs, C., Le Seaux, S., Dubost, J. J., Boisgard, S., Sauvezie, B. & Sirot, J. (2000). Isolation of *Pantoea agglomerans* in two cases of septic monoarthritis after plant thorn and wood sliver injuries. *J Clin Microbiol* **38,** 460-461.

De Vuyst, L. & Vancanneyt, M. (2007). Biodiversity and identification of sourdough lactic acid bacteria. *Food Microbiol* **24,** 120-127.

dos Anjos Borges, L.G., Vechia, V. D. & Corcao, G. (2003). Characterization and genetic diversity via REP-PCR of *Escherichia coli* isolates from polluted waters in southern Brazil. *FEMS Microbiol Ecol* **45,** 173-180.

Drancourt, M., Bollet, C., Carta, A. & Rousselier, P. (2001). Phylogenetic analyses of *Klebsiella* species delineate *Klebsiella* and *Raoultella* gen. nov., with description of *Raoultella ornithinolytica* comb. nov., *Raoultella terrigena* comb. nov. and *Raoultella planticola* comb. nov. *Int J Syst Evol Microbiol* **51,** 925-932.

Dye, D. W. (1964). The taxonomic position of *Xanthomonas trifolii* (Huss, 1907) James, 1955. *N Z J Sci* **7,** 261-269.

Edens, D. G., Gitaitis, R. D., Sanders, F. H. & Nischwitz, C. (2006). First report of *Pantoea agglomerans* causing a leaf blight and bulb rot of onions in Georgia. *Plant Dis* **90,** 1551.

Elvira-Recuenco, M. & van Vuurde, J. W. (2000). Natural incidence of endophytic bacteria in pea cultivars under field conditions. *Can J Microbioly* **46,** 1036-1041.

Enright, M. C. & Spratt, B. G. (1999). Multilocus sequence typing. *Trends Microbiol* **7,** 482-487.

Ewing, W. H. & Fife, M. A. (1972). *Enterobacter agglomerans* (Beijerinck) comb. nov. (the Herbicola-Lathyri bacteria). *Int J Syst Bacteriol* **22,** 4-11.

Ezaki, T., Hashimoto, Y. & Yabuuchi, E. (1989). Fluorometric deoxyribonucleic acid-deoxyribonucleic acid hybridization in micro-dilution wells as an alternative to membrane filter hybridization in which radioisotopes are used to determine genetic relatedness among bacterial strains. *Int J Syst Bacteriol* **39,** 224-229.

Feil, E. J. & Enright, M. C. (2004). Analyses of clonality and the evolution of bacterial pathogens. *Curr Opin Microbiol* **7,** 308-313.

Fox, G. E., Stackebrandt, E., Hespell, R. B., Gibson, J., Maniloff, J., Dyer, T. A., Wolfe, R. S., Balch, W. E., Tanner, R. S., Magrum, L. J., Zablen, L. B., Blakemore, R., Gupta, R., Bonen, L., Lewis, B. J., Stahl, D. A., Luehrsen, K. R., Chen, K. N. & Woese, C. R. (1980). The phylogeny of prokaryotes. *Science* **209,** 457-463.

Fox, G. E., Wisotzkey, J. D. & Jurtshuk, P.,Jr (1992). How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. *Int J Syst Bacteriol* **42,** 166-170.

Francino, M. P., Santos, S. R. & Ochman, H. (2006). Phylogenetic relationships of bacteria with special reference to endosymbionts and enteric species. *In The*

Prokaryotes: Proteobacteria: Gamma Subclass, pp. 41-59. Edited by M. Dworkin, S. Falkow, E. Rosenberg, K. -H. Schleifer & E. Stackebrandt. New York: Springer.

Fukushima, M., Kakinuma, K. & Kawaguchi, R. (2002). Phylogenetic analysis of *Salmonella*, *Shigella*, and *Escherichia coli* strains on the basis of the *gyrB* gene sequence. *J Clin Microbiol* **40,** 2779-2785.

Fullerton, D. G., Lwin, A. A. & Lal, S. (2007). *Pantoea agglomerans* liver abscess presenting with a painful thigh. *Eur J Gastroenterol Hepatol* **19,** 433-435.

Gavini, F., Lefebvre, B. & Leclerc, H. (1983). Taxonomic study of strains belonging or related to the genus *Erwinia*, *herbicola* group, and to the species *Enterobacter agglomerans*. *Syst Appl Microbiol* **4,** 218-235.

Gavini, F., Mergaert, J., Beji, A., Mielcarek, C., Izard, D., Kersters, K. & de Ley, J. (1989). Transfer of *Enterobacter agglomerans* (Beijerinck 1888) Ewing and Fife 1972 to *Pantoea* gen. nov. as *Pantoea agglomerans* comb. nov. and description of *Pantoea dispersa* sp. nov. *Int J Syst Bacteriol* **39,** 337-345.

Gevers, D., Vandepoele, K., Simillion, C. & Van de Peer, Y. (2004). Gene duplication and biased functional retention of paralogs in bacterial genomes. *Trends Microbiol* **12,** 148-154.

Gevers, D., Cohan, F. M., Lawrence, J. G. & other authors (2005). Opinion: Re-evaluating prokaryotic species. *Nat Rev Microbiol* **3,** 733-739.

Gitaitis, R. D. & Gay, J. D. (1997). First report of leaf blight, seed stalk rot, and bulb decay of onion by *Pantoea ananatis* in Georgia. *Plant Dis* **81,** 1096.

Goszczynska, T., Botha, W. J., Venter, S. N. & Coutinho, T. A. (2007). Isolation and identification of the causal agent of brown stalk rot, a new disease of maize in South Africa. *Plant Dis* **91,** 711-718.

Graham, D. C. & Hodgkiss, W. (1967). Identity of Gram negative, yellow pigmented, fermentative bacteria isolated from plants and animals. *J Appl Bacteriol* **30,** 175-189.

Grimont, P. A. D. (1988). Use of DNA reassociation in bacterial classification. *Can J Microbiol* **34,** 541-546.

Grimont, P. A. D. & Grimont, F. (2005). Genus: *Pantoea* In Volume Two: The *Proteobacteria*, Part B: The *Gammaproteobacteria*. *In Bergey's Manual of Systematic Bacteriology*, pp. 713-720. Edited by D. J. Brenner, N. R. Krieg & J. T. Staley. New York: Springer.

Harada, H. & Ishikawa, H. (1997). Phylogenetical relationship based on *groE* genes among phenotypically related *Enterobacter*, *Pantoea*, *Klebsiella*, *Serratia* and *Erwinia* species. *J Gen Appl Microbiol* **43,** 355-361.

Harayama, S. & Kasai, H. (2006). Bacterial phylogeny reconstruction from molecular sequences. *In Molecular Identification, Systematics, and Population Structure of Prokaryotes*, pp. 105-139. Edited by E. Stackebrandt. New York: Springer.

Hauben, L., Moore, E. R. B., Vauterin, L., Steenackers, M., Mergaert, J., Verdonck, L. & Swings, J. (1998). Phylogenetic position of phytopathogens within the *Enterobacteriaceae*. *Syst Appl Microbiol* **21,** 384-397.

Hedegaard, J., Steffensen, S. A., Norskov-Lauritsen, N., Mortensen, K. K. & Sperling-Petersen, H. U. (1999). Identification of *Enterobacteriaceae* by partial sequencing of the gene encoding translation initiation factor 2. *Int J Syst Bacteriol* **49,** 1531-1538.

Hulton, C. S., Higgins, C. F. & Sharp, P. M. (1991). ERIC sequences: a novel family of repetitive elements in the genomes of *Escherichia coli*, *Salmonella typhimurium* and other enterobacteria. *Mol Microbiol* **5,** 825-834.

Huys, G., Coopman, R., Janssen, P. & Kersters, K. (1996). High-resolution genotypic analysis of the genus *Aeromonas* by AFLP fingerprinting. *Int J Syst Bacteriol* **46,** 572-580.

Iversen, C., Waddington, M., On, S. L. W. & Forsythe, S. (2004). Identification and phylogeny of *Enterobacter sakazakii* relative to *Enterobacter* and *Citrobacter* species. *J Clin Microbiol* **42,** 5368-5370.

Jain, R., Rivera, M. C. & Lake, J. A. (1999). Horizontal gene transfer among genomes: the complexity hypothesis. *Proc Natl Acad Sci USA* **96,** 3801-3806.

Janda, J. M. (2006). New members of the family *Enterobacteriaceae*. *In The Prokaryotes: Proteobacteria: Gamma Subclass*, pp. 4-40. Edited by M. Dworkin, S. Falkow, E. Rosenberg, K. -H. Schleifer & E. Stackebrandt. New York: Springer.

Janssen, P., Maquelin, K., Coopman, R., Tjernberg, I., Bouvet, P., Kersters, K. & Dijkshoorn, L. (1997). Discrimination of *Acinetobacter* genomic species by AFLP fingerprinting. *Int J Syst Bacteriol* **47,** 1179-1187.

Jonas, D., Spitzmueller, B., Daschner, F. D., Verhoef, J. & Brisse, S. (2004). Discrimination of *Klebsiella pneumoniae* and *Klebsiella oxytoca* phylogenetic groups and other *Klebsiella* species by use of amplified fragment length polymorphism. *Res Microbiol* **155,** 17-23.

Kageyama, B., Nakae, M., Yagi, S. & Sonoyama, T. (1992). *Pantoea punctata* sp. nov., *Pantoea citrea* sp. nov., and *Pantoea terrea* sp. nov. isolated from fruit and soil samples. *Int J Syst Bacteriol* **42,** 203-210.

Kämpfer, P., Ruppel, S. & Remus, R. (2005). *Enterobacter radicincitans* sp. nov., a plant growth promoting species of the family *Enterobacteriaceae*. *Syst Appl Microbiol* **28,** 213-221.

Konstantinidis, K. T. & Tiedje, J. M. (2005a). Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci USA* **102,** 2567-2572.

Konstantinidis, K. T. & Tiedje, J. M. (2005b). Towards a genome-based taxonomy for prokaryotes. *J Bacteriol* **187,** 6258-6264.

Kratz, A., Greenberg, D., Barki, Y., Cohen, E. & Lifshitz, M. (2003). *Pantoea agglomerans* as a cause of septic arthritis after palm tree thorn injury; case report and literature review. *Arch Dis Child* **88,** 542-544.

Kuhnert, P. & Korczak, B. M. (2006). Prediction of whole-genome DNA-DNA similarity, determination of G+C content and phylogenetic analysis within the family *Pasteurellaceae* by multilocus sequence analysis (MLSA). *Microbiology* **152,** 2537-2548.

La Duc, M. T., Satomi, M., Agata, N. & Venkateswaran, K. (2004). *gyrB* as a phylogenetic discriminator for members of the *Bacillus anthracis–cereus–thuringiensis* group. *J Microbiol Methods,* **56,** 383-394.

Li, X., Zhang, D., Chen, F., Ma, J., Dong, Y. & Zhang, L. (2004). *Klebsiella singaporensis* sp. nov., a novel isomaltulose-producing bacterium. *Int J Syst Evol Microbiol* **54,** 2131-2136.

Lim, P., Chen, S., Tsai, C. & Pai, M. (2006). *Pantoea* peritonitis in a patient receiving chronic ambulatory peritoneal dialysis. *Nephrology* **11,** 97-99.

Ludwig, W. & Schleifer, K. –H. (1999). Phylogeny of bacteria beyond the 16S rRNA standard. *ASM News* **65,** 752-757.

Maiden, M., Bygraves, J., Feil, E. & other authors (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci USA* **95,** 3140-3145.

Martens, M., Delaere, M., Coopman, R., De Vos, P., Gillis, M. & Willems, A. (2007). Multilocus sequence analysis of *Ensifer* and related taxa. *Int J Syst Evol Microbiol* **57,** 489-503.

Martens, M., Dawyndt, P., Coopman, R., Gillis, M., De Vos, P. & Willems, A. (2008). Advantages of multilocus sequence analysis for taxonomic studies: a case study using 10 housekeeping genes in the genus *Ensifer* (including former *Sinorhizobium*). *Int J Syst Evol Microbiol* **58,** 200-214.

Martin, B., Humbert, O., Camara, M. & other authors (1992). A highly conserved repeated DNA element located in the chromosome of *Streptococcus pneumoniae*. *Nucl Acids Res* **20,** 3479-3483.

McCrea, K. W., Xie, J., LaCross, N., Patel, M., Mukundan, D., Murphy, T. F., Marrs, C. F. & Gilsdorf, J. R. (2008). Relationships of nontypeable *Haemophilus influenzae* strains to hemolytic and nonhemolytic *Haemophilus haemolyticus* strains. *J Clin Microbiol* **46,** 406-416.

McGhee, G. C., Schnabel, E. L., Maxson-Stein, K., Jones, B., Stromberg, V. K., Lacy, G. H. & Jones, A. L. (2002). Relatedness of chromosomal and plasmid DNAs of *Erwinia pyrifoliae* and *Erwinia amylovora*. *Appl Environ Microbiol* **68,** 6182-6192.

Medrano, E. G. & Bell, A. A. (2007). Role of *Pantoea agglomerans* in opportunistic bacterial seed and boll rot of cotton (*Gossypium hirsutum)* grown in the field. *J Appl Microbiol* **102,** 134-143.

Mergaert, J., Gavini, F., Kersters, K., Leclerc, H. & De Ley, J. (1983). Phenotypic and protein electrophoretic similarities between strains of *Enterobacter agglomerans*, *Erwinia herbicola*, and *Erwinia milletiae* from clinical or plant origin. *Curr Microbiol* **8,** 327-331.

Mergaert, J., Verdonck, L., Kersters, K., Swings, J., Boeufgras, J. M. & De Ley, J. (1984). Numerical taxonomy of *Erwinia* species using API systems. *J Gen Microbiol* **130,** 1893-1910.

Mergaert, J., Verdonck, L. & Kersters, K. (1993). Transfer of *Erwinia ananas* (synonym, *Erwinia uredovora*) and *Erwinia stewartii* to the genus *Pantoea* emend. as *Pantoea ananas* (Serrano 1928) comb. nov. and *Pantoea stewartii* (Smith 1898) comb.

nov., respectively, and description of *Pantoea stewartii* subsp. *indologenes* subsp. nov. *Int J Syst Bacteriol* **43,** 162-173.

Mollet, C., Drancourt, M. & Raoult, D. (1997). *rpoB* Sequence analysis as a novel basis for bacterial identification. *Mol Microbiol* **26,** 1005-1011.

Morales-Valenzuela, G., Silva-Rojas, H. V., Ochoa-Martinez, D. & other authors (2007). First report of *Pantoea agglomerans* causing leaf blight and vascular wilt in maize and sorghum in Mexico. *Plant Dis* **91,** 1365.

Moretti, C., Silvestri, F. M., Rossini, E., Natalini, G. & Buonaurio, R. (2004). Diagnostic tools for the identification of *Brenneria nigrifluens*, the causal agent of Persian walnut bark canker. *J Plant Pathol* **86,** 300-301.

Mougel, C., Thioulouse, J., Perriere, G. & Nesme, X. (2002). A mathematical method for determining genome divergence and species delineation using AFLP. *Int J Syst Evol Microbiol* **52,** 573-586.

Naser, S., Thompson, F. L., Hoste, B., Gevers, D., Vandemeulebroecke, K., Cleenwerck, I., Thompson, C. C., Vancanneyt, M. & Swings, J. (2005a). Phylogeny and identification of enterococci using *atpA* gene sequence analysis. *J Clin Microbiol* **43,** 2224-2230.

Naser, S. M., Thompson, F. L., Hoste, B., Gevers, D., Dawyndt, P., Vancanneyt, M. & Swings, J. (2005b). Application of multilocus sequence analysis (MLSA) for rapid identification of *Enterococcus* species based on *rpoA* and *pheS* genes. *Microbiology* **151,** 2141-2150.

Naser, S. M., Vancanneyt, M., De Graef, E. & other authors (2005c). *Enterococcus canintestini* sp. nov., from faecal samples of healthy dogs. *Int J Syst Evol Microbiol* **55,** 2177-2182.

Naser, S. M., Dawyndt, P., Hoste, B., Gevers, D., Vandemeulebroecke, K., Cleenwerck, I., Vancanneyt, M. & Swings, J. (2007). Identification of lactobacilli by *pheS* and *rpoA* gene sequence analyses. *Int J Syst Evol Microbiol* **57,** 2777-2789.

Nick, G., Jusssila, M., Hoste, B., Maarit, R., Kaijalainen, S., De Lajudie, P., Gillis, M., de Bruijn, E. J. & Lindstrom, K. (1999). *Rhizobia* isolated from root nodules of tropical leguminous trees characterized using DNA-DNA dot-blot hybridization and rep-PCR genomic fingerprinting. *Syst Appl Microbiol* **22,** 287-299.

Norman, D. J., Yuen, J. M. F., Resendiz, R. & Boswell, J. (2003). Characterization of *Erwinia* populations from nursery retention ponds and lakes infecting ornamental plants in Florida. *Plant Dis* **87,** 193-196.

Olivares-Fuster, O., Baker, J. L., Terhune, J. S., Shoemaker, C. A., Klesius, P. H. & Arias, C. R. (2007). Host-specific association between *Flavobacterium columnare* genomovars and fish species. *Syst Appl Microbiol* **30,** 624-633.

Paccola-Meirelles, L. D., Ferreira, A. S., Meirelles, W. F., Marriel, I. E. & Casela, C. R. (2001). Detection of a bacterium associated with a leaf spot disease of maize in Brazil. *J Phytopathol* **149,** 275-279.

Palys, T., Nakamura, L. K. & Cohan, F. M. (1997). Discovery and classification of ecological diversity in the bacterial world: the role of DNA sequence data. *Int J Syst Bacteriol* **47,** 1145-1156.

Palys, T., Berger, E., Mitrica, I., Nakamura, L. K. & Cohan, F. M. (2000). Protein-coding genes as molecular markers for ecologically distinct populations: the case of two *Bacillus* species. *Int J Syst Evol Microbiol* **50,** 1021-1028.

Paradis, S., Boissinot, M., Paquette, N. & other authors (2005). Phylogeny of the *Enterobacteriaceae* based on genes encoding elongation factor Tu and F-ATPase beta-subunit. *Int J Syst Evol Microbiol* **55,** 2013-2025.

Portier, P., Fischer-Le Saux, M., Mougel, C., Lerondelle, C., Chapulliot, D., Thioulouse, J. & Nesme, X. (2006). Identification of genomic species in *Agrobacterium* Biovar 1 by AFLP genomic markers. *Appl Environ Microbiol* **72,** 7123-7131.

Postic, D., Garnier, M. & Baranton, G. (2007). Multilocus sequence analysis of atypical *Borrelia burgdorferi sensu lato* isolates - description of *Borrelia californiensis* sp. nov., and genomospecies 1 and 2. *Int J Med Microbiol* **297,** 263-271.

Rademaker, J. L. W., Hoste, B., Louws, F. J., Kersters, K., Swings, J., Vauterin, L., Vauterin, P. & De Bruijn, F. J. (2000). Comparison of AFLP and rep-PCR genomic fingerprinting with DNA-DNA homology studies: *Xanthomonas* as a model system. *Int J Syst Evol Microbiol* **50,** 665-677.

Rademaker, J. L. W., Herbet, H., Starrenburg, M. J. C., Naser, S. M., Gevers, D., Kelly, W. J., Hugenholtz, J., Swings, J. & van Hylckama Vlieg, J.E.T. (2007). Diversity analysis of dairy and nondairy *Lactococcus lactis* isolates, ssing a novel multilocus sequence analysis scheme and (GTG) sub(5)-PCR fingerprinting. *Appl Environ Microbiol* **73,** 7128-7137.

Richter, D., Postic, D., Sertour, N., Livey, I., Matuschka, F. & Baranton, G. (2006). Delineation of *Borrelia burgdorferi sensu lato* species by multilocus sequence analysis and confirmation of the delineation of *Borrelia spielmanii* sp. nov. *Int J Syst Evol Microbiol* **56,** 873-881.

Rivera, M. C., Jain, R., Moore, J. E. & Lake, J. A. (1998). Genomic evidence for two functionally distinct gene classes. *Proc Natl Acad Sci USA* **95,** 6239-6244.

Rossello-Mora, R. (2006). DNA-DNA reassociation methods applied to microbial taxonomy and their critical evaluation. *In Molecular Identification, Systematics, and Population Structure of Prokaryotes,* pp. 23-50. Edited by E. Stackebrandt. New York: Springer.

Rossello-Mora, R. (2003). Opinion: the species problem, can we achieve a universal concept? *Syst Appl Microbiol* **26,** 323-326.

Rossello-Mora, R. & Amann, R. (2001). The species concept for prokaryotes. *FEMS Microbiol Rev* **25,** 39-67.

Samson, R., Legendre, J. B., Christen, R., Achouak, W. & Gardan, L. (2005). Transfer of *Pectobacterium chrysanthemi* (Burkholder *et al.* 1953) Brenner *et al.* 1973 and *Brenneria paradisiaca* to the genus *Dickeya* gen. nov. as *Dickeya chrysanthemi* comb. nov. and *Dickeya paradisiaca* comb. nov. and delineation of four novel species, *Dickeya dadantii* sp. nov., *Dickeya dianthicola* sp. nov., *Dickeya dieffenbachiae* sp. nov. and *Dickeya zeae* sp. nov. *Int J Syst Evol Microbiol* **55,** 1415-1427.

Santos, S. R. & Ochman, H. (2004). Identification and phylogenetic sorting of bacterial lineages with universally conserved genes and proteins. *Environ Microbiol* **6,** 754-759.

Sarkar, S. F. & Guttman, D. S. (2004). Evolution of the core genome of *Pseudomonas syringae*, a highly clonal, endemic plant pathogen. *Appl Environ Microbiol* **70,** 1999-2012.

Savelkoul, P. H. M., Aarts, H. J. M., De Haas, J., Dijkshoorn, L., Duim, B., Otsen, M., Rademaker, J. L. W., Schouls, L. & Lenstra, J. A. (1999). Amplified-fragment length polymorphism analysis: the state of an art. *J Clin Microbiol* **37,** 3083-3091.

Scally, M., Schuenzel, E. L., Stouthamer, R. & Nunney, L. (2005). Multilocus sequence type system for the plant pathogen *Xylella fastidiosa* and relative contributions of recombination and point mutation to clonal diversity. *Appl Environ Microbiol* **71,** 8491-8499.

Schildkraut, C. L., Marmur, J. & Doty, P. (1961). The formation of hybrid DNA molecules and their use in studies of DNA homologies. *J Mol Biol* **3,** 595-617.

Schmid, H., Schubert, S., Weber, C. & Bogner, J. R. (2003). Isolation of a *Pantoea dispersa*-like strain fron a 71-year-old woman with acute myeloid leukemia and multiple myeloma. *Infection* **31,** 66-67.

Serrano, F. B. (1928). Bacterial fruitlet brown rot of pineapples in the Philippines. *Philipp J Sci* **36,** 271-305.

Skerman, V. B. D., McGowan, V. & Sneath, P. H. A. (1980). Approved lists of bacterial names. *Int J Syst Bacteriol* **30,** 225-420.

Sneath, P. H. A. (1989). Analysis and interpretation of sequence data for bacterial systematics: the view of a numerical taxonomist. *Syst Appl Microbiol* **12,** 15-31.

Sonoyama, T., Yagi, S. & Kageyama, B. (1988). Facultatively anaerobic bacteria showing high productivities of 2,5-diketo-D-gluconate from D-glucose. *Agric Biol Chem* **52,** 667-674.

Spratt, B. G., Hanage, W. P., Li, B., Aanensen, D. M. & Feil, E. J. (2004). Displaying the relatedness among isolates of bacterial species - the eBURST approach. *FEMS Microbiol Lett* **241,** 129-134.

Spröer, C., Mendrock, U., Swiderski, J., Lang, E. & Stackebrandt, E. (1999). The phylogenetic position of *Serratia*, *Buttiauxella* and some other genera of the family *Enterobacteriaceae*. *Int J Syst Bacteriol* **49,** 1433-1438.

Stackebrandt, E. (2002). From species definition to species concept: population genetics if going to influence the systematics of prokaryotes. *WFCC Newsletter* **35,** 1-4.

Stackebrandt, E., Frederiksen, W., Garrity, G. M. & other authors (2002). Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int J Syst Evol Microbiol* **52,** 1043-1047.

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

Stephan, R., Van Trappen, S., Cleenwerck, I., Vancanneyt, M., De Vos, P. & Lehner, A. (2007). *Enterobacter turicensis* sp. nov. and *Enterobacter helveticus* sp. nov., isolated from fruit powder. *Int J Syst Evol Microbiol* **57,** 820-826.

Stern, M. J., Ames, G. F., Smith, N. H., Robinson, E. C. & Higgins, C. F. (1984). Repetitive extragenic palindromic sequences: a major component of the bacterial genome. *Cell* **37,** 1015-1026.

Stewart, F. C. (1897). A bacterial disease of sweet corn. *New York State Agric Exp Stn Bull* **130,** 422-439.

Tarr, C. L., Nelson, A. M., Beutin, L., Olsen, K. E. P. & Whittam, T. S. (2008). Molecular characterization reveals similar virulence gene content in unrelated clonal groups of *Escherichia coli* of serogroup O174 (OX3). *J Bacteriol* **190,** 1344-1349.

Thompson, F. L., Gevers, D., Thompson, C. C., Dawyndt, P., Naser, S., Hoste, B., Munn, C. B. & Swings, J. (2005). Phylogeny and molecular identification of Vibrios on the basis of multilocus sequence analysis. *Appl Environ Microbiol* **71,** 5107-5115.

Truper, H. G. & De' Clari, L. (1997). Taxonomic note: necessary correction of specific epithets fromed as substantives (nouns) "in apposition". *Int J Syst Bacteriol* **47,** 908-909.

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. *Int J Syst Bacteriol* **45,** 604.

Urwin, R. & Maiden, M. C. J. (2003). Multi-locus sequence typing: a tool for global epidemiology. *Trends Microbiol* **11,** 479-487.

Vandamme, P., Pot, B., Gillis, M., Vos, P. d., Kersters, K. & Swings, J. (1996). Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol Rev* **60,** 407-438.

Vandamme, P., Segers, P., Ryll, M.& other authors (1998). *Pelistega europaea* gen. nov., sp. nov., a bacterium associated with respiratory disease in pigeons: taxonomic structure and phylogenetic allocation. *Int J Syst Bacteriol* **48,** 431-440.

Verdonck, L., Mergaert, J., Rijckaert, C., Swings, J., Kersters, K. & De Ley, J. (1987). Genus *Erwinia* : Numerical analysis of phenotypic features. *Int J Syst Bacteriol* **37,** 4-18.

Vos, P., Hogers, R., Bleeker, M.& other authors (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res* **23,** 4407-4414.

Wayne, L. G., Brenner, D. J., Colwell, R. R. & other authors (1987). Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int J Syst Bacteriol* **37,** 463-464.

Weigel, R. M., Qiao, B., Teferedegne, B., Suh, D. K., Barber, D. A., Isaacson, R. E. & White, B. A. (2004). Comparison of pulsed field gel electrophoresis and repetitive sequence polymerase chain reaction as genotyping methods for detection of genetic diversity and inferring transmission of *Salmonella*. *Vet Microbiol* **100,** 205-217.

Wells, J. M., Sheug, W. S., Ceponis, M. J. & Chen, T. A. (1987). Isolation and characterization of strains of *Erwinia ananas* from honeydew melons. *Phtypathology* **77,** 511-514.

Willems, A., Coopman, R. & Gillis, M. (2001). Comparison of sequence analysis of 16S-23S rDNA spacer regions, AFLP analysis and DNA - DNA hybridizations in *Bradyrhizobium*. *Int J Syst Evol Microbiol* **51,** 623-632.

Woese, C.R. (1987). Bacterial evolution. *Microbiol Rev* **51,** 221-271

Woese, C. R., Kandler, O. & Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria and Eucarya. *Proc Natl Acad Sci USA* **87,** 4576-4579.

Yamamoto, S., Bouvet, P. J. & Harayama, S. (1999). Phylogenetic structures of the genus *Acinetobacter* based on *gyrB* sequences: comparison with the grouping by DNA-DNA hybridization. *Int J Syst Bacteriol* **49,** 87-95.

Yamamoto, S. & Harayama, S. (1996). Phylogenetic analysis of *Acinetobacter* strains based on the nucleotide sequences of *gyrB* genes and on the amino acid sequences of their products. *Int J Syst Evol Microbiol* **46,** 506-511.

Yanez, M. A., Catalan, V., Apraiz, D., Figueras, M.J. & Martinez-Murcia, A.J. (2003). Phylogenetic analysis of members of the genus *Aeromonas* based on *gyrB* gene sequences. *Int J Syst Evol Microbiol* **53,** 875-883.

Young, J. M. & Park, D. -C. (2007). Relationships of plant pathogenic enterobacteria based on partial *atpD*, *carA*, and *recA* as individual and concatenated nucleotide and peptide sequences. *Syst Appl Microbiol* **30,** 343-354.

Zeigler, D. R. (2003). Gene sequences useful for predicting relatedness of whole genomes in bacteria. *Int J Syst Evol Microbiol* **53,** 1893-1900.

Zhang, Q., Weyant, R., Steigerwalt, A. G., White, L. A., Melcher, U., Bruton, B. D., Pair, S. D., Mitchell, F. L. & Fletcher, J. (2003). Genotyping of *Serratia marcescens* strains associated with cucurbit yellow vine disease by repetitive elements-based polymerase chain reaction and DNA-DNA hybridization. *Phytopathology* **93,** 1240-1246.

CHAPTER 3

Phylogeny and identification of *Pantoea* species associated with the environment, humans and plants based on multilocus sequence analysis (MLSA)

As submitted to: Systematic and Applied Microbiology

ABSTRACT

Species belonging to the genus of *Pantoea* are commonly isolated from plants, humans and the environment. The species of the genus are phenotypically closely related, making rapid identification of *Pantoea* strains to the species level difficult. Multilocus sequence analysis (MLSA) was evaluated as means for rapid classification and identification of *Pantoea* strains. Four housekeeping genes, *rpoB*, *atpD*, *gyrB* and *infB*, were sequenced for a total of 102 strains assigned to the genus. Included in the study were (1) reference strains from the seven validly described species of *Pantoea*, (2) strains belonging to Brenner DNA groups II, IV and V, previously isolated from clinical samples and difficult to identify because of high phenotypic homogeneity to *P. agglomerans* or *P. ananatis*, and (3) isolates from diseased *Eucalyptus*, maize and onion, assigned to the genus on the basis of phenotypic tests. Phylogenetic trees were constructed from the sequences of the four housekeeping genes. The *Pantoea* strains grouped into a monophyletic cluster when the tree was based on concatenated sequences of the four genes, although two sublineages could be observed with high bootstrap support. The MLSA data further suggested the existence of ten potential novel species, phylogenetically related to the currently recognized *Pantoea* species. When compared with DNA-DNA hybridization data a good congruence was observed between both methods, with *gyrB* sequence data being the most consistent. In conclusion, MLSA of partial nucleotide sequences of the genes *rpoB*, *atpD*, *gyrB* and *infB* can be used for classification and identification of *Pantoea* strains.

INTRODUCTION

The genus *Pantoea* belongs within the family *Enterobacteriaceae* and was proposed by Gavini *et al*. (14) for two groups of strains that were, at that time, assigned to the *Erwinia herbicola-Enterobacter agglomerans* complex. This complex covered many phena (13, 40) and genomic groups (3), some of which were later designated as new genera (16). The genus *Pantoea* comprises at present seven validly-described species, namely *Pantoea agglomerans* and *P. dispersa* (14), *P. citrea*, *P. punctata* and *P. terrea* (21) and *P. ananatis* and *P. stewartii* (28). However, Grimont and Grimont (16) stated that the genus *Pantoea* can be envisioned to include DNA groups I, II, IV and V as determined by Brenner *et al*. (3). It was further observed that the species *P. citrea*, *P. punctata* and *P. terrea*, isolated in Japan and described by Kageyama *et al*. (21) differed from the "core" *Pantoea* species in several biochemical or nutritional characteristics. Grimont and Grimont (16) determined the phylogenetic position of all validly-described *Pantoea* species and DNA groups of Brenner *et al*. (3) using 16S rRNA- and *rpoB*-sequence comparisons and found that the "Japanese" species constituted a cluster that joined the *Pantoea* cluster at a lower level. They concluded that more taxonomic work was needed to justify the assignment of these species to the genus *Pantoea*.

Several species belonging to the genus *Pantoea* are known as plant pathogens. Stewart's vascular wilt is a disease of sweet corn and maize caused by *Pantoea stewartii* subsp. *stewartii* (36), *Pantoea agglomerans* causes crown and root gall disease of gypsophila and beet (4, 5) and *P. ananatis* causes a variety of diseases on a wide range of hosts including bacterial blight and dieback of *Eucalyptus* (7), stem necrosis of rice (6) and brown stalk rot of maize (15). Recently, *Pantoea* strains were isolated from young *Eucalyptus* trees in Uganda, Argentina and Uruguay, showing a disease similar to bacterial blight and dieback of *Eucalyptus* in South Africa. *Pantoea* strains, not belonging to *P. ananatis* and causing brown stalk rot of maize were also isolated. Similarly, *Pantoea* strains that could not clearly be identified to the species level were isolated from diseased onion in South Africa and the U.S.A. All of these isolates were assigned to the genus *Pantoea* on the basis of phenotypic tests.

All strains of DNA groups I, II, IV and V determined by Brenner *et al*. (3) were isolated from human sources. In recent years *Pantoea* strains have been consistently linked with human infections (10, 12, 24, 26, 34, 39). The increasing isolations of *Pantoea* strains from the environment, from human infections and diseased plant material highlights the need for a technique that enables fast and reliable classification and identification of *Pantoea* strains worldwide.

Partial sequences of protein-encoding genes have been proven useful for species identification and as phylogenetic markers in the family *Enterobacteriaceae*. The following genes have been evaluated for these purposes: *rpoB* (11, 22, 25, 29, 35), *infB* (19), *groEL* (27); *gyrB* (9), *tuf* and *atpD* (31, 42), *carA* and *recA* (42). Results have shown that they are more reliable than 16S rRNA gene sequences for species identification and for determining intra- and some inter-generic relationships (9, 19, 29, 31). For determination of phylogenetic relationships, it has been advised to use sequence data from more than one gene (9, 23), to reduce the possibility of ambiguities caused by genetic recombination or specific selection (19). 16S rRNA gene sequences, on the other hand, appeared to be useful for determination of phylogenies between distantly related *Enterobacteriaceae* (9).

In the present study, the phylogeny of all validly described species of *Pantoea*, DNA groups II, IV and V of Brenner *et al*. (3), which belong in the genus *Pantoea* according to Grimont and Grimont (16), as well as isolates assigned to *Pantoea* on the basis of phenotypic tests, were investigated using multilocus sequence analysis (MLSA) of the protein-encoding genes *rpoB*, *atpD*, *gyrB* and *infB*. These genes encode RNA polymerase ß subunit, ATP synthase ß subunit, DNA gyrase and initiation translation factor 2, respectively and have been used successfully in previous phylogenetic studies of the *Enterobacteriaceae* (9, 19, 29, 31).

MATERIAL AND METHODS

Strains investigated and DNA extraction

A detailed summary of the 102 *Pantoea* strains used in this study is listed in Table 1. All strains used in this study are maintained in the Bacterial Culture Collection (BCC) of the Forestry and Agricultural Biotechnology Institute (FABI) and representative isolates have been deposited in the BCCM/LMG Collection, University of Ghent, Belgium.

Genomic DNA was extracted from each of the bacterial strains using an alkalic extraction method (30) and stored at -20 °C.

Primer design

External primers for amplification and internal primers for sequencing were designed based on sequence alignments of strains representative of multiple species belonging to the family *Enterobacteriaceae*. The primers used in previous studies of the *Enterobacteriaceae* (9, 19, 29, 31), were the basis for the design of the primers used in this study. The sequences for both the amplification and sequencing primers for all four genes are listed in Table 2.

PCR and sequencing

PCR was performed on each of the strains listed in Table 1 using each of the four sets of amplification primers. Each 50 µl PCR reaction consisted of 5 µl 10 x PCR buffer, 5 µl dNTP's (200 µM each), 0.5 µl forward primer (50 µM), 0.5 µl reverse primer (50 µM), 1 µL AmpliTaq DNA polymerase (1U/µl), 5 µl template DNA and 33 µl sterile MilliQ water. The amplification conditions included denaturation at 95 °C for 5 minutes, 3 cycles of denaturation at 95 °C for 1 minute, annealing at 55 °C for 2 minutes 15 seconds and elongation at 72 °C for 1 minute 15 seconds, followed by 30 cycles of denaturation at 95 °C for 35 seconds, annealing at 55 °C for 1 minute 15 seconds and elongation at 72 °C for 1 minute 15 seconds and a further 7 minutes of elongation at 72 °C. An annealing temperature of 50 °C was used for several strains which would not amplify at 55 °C. PCR products were separated on 1 % agarose gels at 75 V for 45 minutes. Those reactions resulting in positive PCR products of the expected size were purified using NucleoFast 96 PCR plates (Machery-Nagel). Sequencing reactions were performed using 3 µl purified PCR product, 2 µl 5 x sequencing buffer, 0.2 µl Big Dye sequencing reaction mix, 3 µl primer (4 µM) and 1.8 µl sterile MilliQ water. The sequencing conditions included denaturation at 96 °C for 5 seconds, 25 cycles of denaturation at 96 °C for 10 seconds, annealing at 55 °C for 10 seconds and elongation at 60 °C for 4 minutes. Sequencing reactions were purified using the Montage Seq 96 Sequencing Reaction Cleanup Kit (Millipore). All PCR set-up and purification steps were carried out on a Genesis Workstation 200 (Tecan).

Sequence analysis

The GenBank/EMBL accession numbers for the sequences presented in this study are: EF988667-EF988752, EU145244-EU145259, EU344753-EU344756 (*atpD* gene), EF988753-EF988838, EU145260-EU145275, EU344757-EU344760 (*gyrB* gene), EF988839-EF988924, EU145276-EU145291, EU344761-EU344764 (*infB* gene) and EF988925-EF989010, EU145292-EU145307, EU344765-EU344768 (*rpoB* gene).

Consensus sequences for each strain were assembled by manual alignment of the internal sequences using BioEdit Sequence Alignment Editor v 5.0.9 (18). The consensus sequences were then aligned for each gene using ClustalX (38) and the overhangs were trimmed. All four data sets were tested for substitution saturation at the first, second and third codons using the DAMBE software package by plotting the transitions (s) and transversions (v) against the genetic distance calculated with the Jukes-Cantor (JC69) model (41). A partition-homogeneity test was performed in PAUP 4.0b10 (37) to establish if the four genes could be combined to form a single concatenated data set. The Modeltest 3.7 programme (32) was then applied to all four data sets, as well as the concatenated data set, to determine the best-fit evolutionary model to apply to each gene. Maximum likelihood and neighbour joining trees were drawn using Phyml (17) and PAUP 4.0b10 (37), respectively by applying the models and parameters determined by Modeltest. Bootstrap analysis with 1 000 replicates was performed on all five trees to assess the reliability of the clusters generated. *Escherichia coli*, *Shigella dysenteriae* and *Citrobacter rodentium* were chosen as outgroups and *Erwinia* and *Tatumella*, the closest phylogenetically related neighbours of *Pantoea* were also included in the trees (42). The sequences for the four housekeeping genes of *E. coli*, *S. dysenteriae*, *C. rodentium* and *Er. amylovora* were obtained from the genome sequencing databases of the Sanger Institute (http://www.sanger.ac.uk) and the University of Wisconsin (https://asap.ahabs.wisc.edu/asap). The genes for *Erwinia billingiae* (LMG 2613^T), *Erwinia rhapontici* (LMG 2688^T), *Erwinia toletana* (LMG 24162) and *Tatumella ptyseos* (LMG 7888^T) were sequenced along with the *Pantoea* strains. The MLSA data from the four housekeeping genes were compared amongst each other, and to DNA-DNA hybridization values (data not shown) in Bionumerics (Applied Maths), by calculating the correlation between the experiment types, to determine the congruence. A scatter plot was constructed depicting the correlation between *gyrB* sequence similarity (the most congruent), and corresponding DNA-DNA hybridization values.

RESULTS

Following sequence alignment and trimming of the overhangs, the lengths of the four genes were as follows: *rpoB* = 637 bp, *atpD* = 657 bp, *gyrB* = 742 bp and *infB* = 615 bp. The results from the substitution saturation tests in DAMBE revealed that there was no saturation at the first or second codon positions for all four genes as neither the transitions nor the transversions reached a plateau (graphs not shown). Therefore the sequences for these four genes are informative at the first and second codon and the phylogenetic signal is intact. Furthermore, there is no substitution saturation at the third codon for *rpoB* or *atpD* (Fig. 1a & 1b), indicating that these housekeeping genes are stable and not under selective pressure. For *gyrB* the transitions outnumber the transversions and are slowly reaching a plateau at a genetic distance greater than 0.92, indicating there is possible substitution saturation at the third codon position (Fig. 1c). There is definite saturation at the third codon position for the *infB* gene. The transitions reach a plateau at a genetic distance greater than 0.53 while the transversions continue to increase linearly (Fig. 1d). In this case, the phylogenetic signal is lost as substitution saturation was reached and these sequences fail to be informative at the third codon.

The partition-homogeneity test revealed that all four data sets were combinable with each other for the construction of a concatenated tree. Only the *P* values for concatenation of *infB* and *atpD*, and *atpD* and *gyrB* may be considered borderline at 0.01. It has been suggested that *P* values between 0.01 and 0.001 indicate data sets that cannot be successfully combined (8). The *P* values for the remaining data set combinations were all well above 0.01, allowing the concatenation of all four data sets. In order to correct the effect of substitution saturation on the sequence data analysis, the data sets were subjected to model tests. The models selected by Modeltest were the general time reversible (GTR) model for *gyrB*, *atpD* and the concatenated data set, the Tamura-Nei (TN93) model for *infB* and the Kimura (K3P) model for *rpoB*.

On the basis of the housekeeping gene sequence data, all seven validly described species of *Pantoea* were clearly differentiated in each of the maximum likelihood trees constructed (neighbour joining trees not shown), but the phylogenetic position of some species varied between trees. Additionally, ten separate MLSA groups (A-J) containing *Pantoea* strains were visible in each of the five trees (Fig. 2a-d & 3). The majority of the validly described *Pantoea* species and MLSA groups A-J were supported by strong bootstrap values. The only exceptions included marginally lower intra- and inter-species bootstraps in the *rpoB* tree (Fig. 2a) and several weaker interspecies bootstrap values in the *infB* (Fig. 2d) and *atpD* (Fig. 2b) trees. The bootstrap values for the concatenated data set tree were by far the highest and most stable, not only at the intraspecies level, but also between species of the genus (Fig. 3). Strains of the same *Pantoea* species had at least 98.9 % *rpoB*, 98.3 % *atpD*, 96.4 % *gyrB* and 97.2 % *infB* gene sequence similarity, whereas at the interspecies level the sequence similarity was at a maximum of 98.4 % for *rpoB*, 97.9 % for *atpD*, 94.8 % for *gyrB* and 96.9 % for *infB*.

The congruence between the four genes used in this study varied from 86.3 % to 92.6 % and the congruence between MLSA and DNA-DNA hybridization data (data not shown) ranged from 84.5 % to 95.2 %, with *gyrB* being most congruent (95.2 %) and *atpD* the least (84.5 %). A scatter plot comparing the DNA-DNA hybridization values and *gyrB* sequence data is presented in Fig. 4.

DISCUSSION

Pantoea strains are isolated from the environment on a regular basis. The isolates can be human and clinical strains, the causal agents of diseases on plants, epi- and endophytes or merely present in water and soil samples. Due to the increasing number of isolations, which rarely result in conclusive identification, a rapid technique is required to classify and identify these phenotypically-related strains to the species level. The difficulties experienced in identifying *Pantoea* species is exacerbated by the uncertain phylogeny of the genus. The concatenated data from the MLSA scheme was used to verify the phylogenetic position of the genus *Pantoea* within the family *Enterobacteriaceae*. Together and separately all four housekeeping genes, *rpoB*, *atpD*, *gyrB* and *infB*, can delineate the seven validly described species and revealed ten potential novel species.

The concatenated tree appeared to be most reliable for determining phylogenetic relationships amongst *Pantoea* strains (Fig. 3). In this tree, *Pantoea* strains form a monophyletic cluster that contains two subclusters, supported by high bootstrap values. The first subcluster contains *P. agglomerans*, *P. ananatis*, *P. stewartii*, *P. dispersa* (the "core" species) and nine MLSA groups of potential novel *Pantoea* species. The second subcluster contains the "Japanese species" (*P. punctata*, *P. citrea* and *P. terrea*), *Tatumella ptyseos* and another potential novel species. The largest MLSA group includes isolates from *Eucalyptus* leaves showing symptoms of bacterial blight in Uganda, Argentina and Uruguay as well as strains isolated from maize infected with brown stalk rot in South Africa. In all of the trees, this MLSA group A clusters closely to *P. agglomerans* strains, but distinctly diverges into a separate cluster. MLSA group B also clusters in close proximity to *P. agglomerans* and contains strains from *Eucalyptus* infected with bacterial blight, but only from trees in Uruguay. Two strains from a study by Beji *et al.* (1), LMG 2558 and LMG 2560, were assigned to *P. agglomerans* based on their protein profiles but later excluded from the species by Gavini *et al*., (14). However, based on the MLSA data presented in this study these two strains, forming MLSA group C, constitute a potential novel species. A single strain, LMG 24200, forms MLSA group D. This strain was isolated from infected *Eucalyptus* in Uganda and was expected to cluster in MLSA group A along with LMG 24199 and BCC 107, also isolated from Uganda. However, as this strain retains its position in all five trees, LMG 24200 represents a potential new species of *Pantoea*. MLSA group E contains human strains belonging to the group referred to as Brenner DNA group V, whilst a single strain from the same group (R-35496) constitutes MLSA group F. The isolates from onion form MLSA group G, which consistently branches off from the *P. ananatis* cluster with strong bootstrap support in the majority of the trees. MLSA groups H and I are comprised of human strains from Brenner DNA groups II and IV, respectively. MLSA groups A – I all group within the first subcluster, which is referred to as the "core" *Pantoea* group with *P. agglomerans*, *P. ananatis*, *P. stewartii* and *P. dispersa* (16). MLSA group J, falls within the second subcluster, the "Japanese" *Pantoea* clade, which joins the "core" *Pantoea* clade at a lower level in all of the maximum likelihood trees. MLSA group J contains two strains thought to be *P. citrea* as they were identified as the causal agent of pink disease of pineapple (33). However, MLSA group J forms a cluster distinct from the type strain of *P. citrea* (LMG 22049^T) with an extended branch length that is observed in all five trees. Two strains classified as *P. terrea* (LMG 23565 and CCUG 30163) cluster with the type strain of *Tatumella ptyseos* (LMG 7888^T) instead of with the type strain of *P. terrea* (LMG 22051^T), suggesting that they were wrongly classified. The clusters observed in this MLSA study were confirmed by DNA-DNA hybridization data (data not shown).

The concatenated data set tree (Fig. 3) revealed that there are four major groupings of *Pantoea* species within the first subcluster, all supported by a bootstrap value of 100 %: (1) MLSA group A, *P. agglomerans*, MLSA groups B, C, D, E and F, (2) *P. ananatis*, MLSA group G and both subspecies of *P. stewartii*, (3) MLSA group H, and (4) MLSA group I and *P. dispersa* on the border of the *Pantoea* "core" species. A similar pattern is seen throughout the maximum likelihood trees of *rpoB* and *infB*. In the concatenated, *rpoB*, *gyrB* and *infB* trees, the "Japanese" *Pantoea* species consistently form a distinct clade with an extended branch length, casting doubt on the inclusion of these species within the genus supporting the statement of Grimont and Grimont (16). In the *atpD* tree (Fig. 2b) the "Japanese" species cluster between groups 1 and 2, separating the "core" *Pantoea* species. This unlikely topology is possibly the result of horizontal gene transfer in the *atpD* gene. A study of the phylogeny of the *Enterobacteriaceae* based on the *atpD* gene by Paradis *et al.* (31), revealed an indel in the *atpD* genes of *P. agglomerans* and *P. dispersa* which could explain the different topology in the *atpD* tree. A second branch having another position in the concatenated tree and the *atpD* and *infB* trees, contains the single strain LMG 24194. In the concatenated tree, *atpD* and *infB* trees, this strain groups on the border of the "core" *Pantoea* clade, but in the *gyrB* and *rpoB* trees LMG 24194 groups with *Erwinia* species (Fig. 2a-d). LMG 24194 was thought to belong to a potential new *Pantoea* species but as this strain does not retain its position in all five trees, its taxonomic position cannot be clearly concluded at present.

The results from this MLSA study further support the statement of Grimont and Grimont (16), that "more taxonomic work is needed to justify the assignment of *P. citrea*, *P. terrea* and *P. punctata* to the genus *Pantoea*" as the "Japanese" species cluster together at a level distant to the *Pantoea* "core" species. The study by Paradis *et al.* (31) brought attention to a clear phylogenetic affiliation between the genera *Pantoea* and *Tatumella*. Based on that observation the type strain of *Tatumella ptyseos* was included in our MLSA study. In all five maximum likelihood trees, *T. ptyseos* groups within the "Japanese" species cluster with bootstrap support of 100 %, prompting closer examination of these species and the genus *Tatumella*. *T. ptyseos* is the only species belonging to the genus *Tatumella* which was proposed for a group of organisms isolated from clinical sources (20).

A novel study examining the relationships of plant pathogenic enterobacteria based on the housekeeping genes *atpD*, *carA* and *recA*, suggests that there is no justification for the separation of *Erwinia* and *Pantoea* into two separate genera (42). In the present study, *Erwinia* species cluster at a lower level to the "Japanese" *Pantoea* species in the concatenated tree (Fig. 3) and usually on the border of the *Pantoea* "core" species in the single gene trees (Fig. 2a-d). Since concatenated trees are preferred over single housekeeping gene trees for making phylogenetic inferences (9, 23), this current study clearly indicates that *Erwinia* and *Pantoea* should not be united into a single genus.

A high level of congruence was observed between *gyrB* sequence data and DNA-DNA hybridization values (Fig. 4). It was generally noted that strains sharing more than 70 % DNA similarity have high *gyrB* sequence similarity. The only notable exceptions are strains belonging to the subspecies of *P. stewartii* which share lower DNA relatedness (60-65 %) but high *gyrB* sequence similarity (± 99%), indicated by an arrow in Fig. 4. The lower DNA-DNA hybridization values could suggest that the two subspecies of *P. stewartii*, subspecies *stewartii* and subspecies *indologenes*, ought to be divided into separate species. However, the high *gyrB* sequence similarity supports their status as subspecies of *P. stewartii*.

The MLSA technique was examined for usefulness in the classification and identification of *Pantoea* strains to the species level and was found to be successful. The high bootstrap values at the species level indicate that the four housekeeping genes used in this study, *rpoB*, *atpD*, *gyrB* and *infB*, are reliable genetic markers for differentiation of *Pantoea* species. Not only could the MLSA scheme distinguish between the seven validly published species of *Pantoea*, it also revealed ten potential new species. The potential species observed in each of the MLSA trees in this study are supported by both AFLP analysis (2) and DNA-DNA hybridization data. The potential new *Pantoea* species are now in the process of being described using the MLSA data as a supporting technique. In conclusion, MLSA provides a rapid technique for reliable classification and identification of *Pantoea* strains to the species level and is clearly more discriminatory than 16S rRNA sequencing. Furthermore, this study has improved our understanding of the phylogeny of the genus *Pantoea*.

ACKNOWLEDGEMENTS

This study was funded by the South African-Flemish Bilateral Agreement, the National Research Foundation (NRF), the Tree Protection Co-operative Programme (TPCP) and the THRIP support programme of the Department of Trade and Industry, South Africa. The BCCM/LMG Bacteria collection is supported by the Federal Public Planning Service-Science Policy, Belgium. The authors wish to acknowledge Bart Hoste, Katrien Vandemeulebroecke, Dirk Gevers and Peter Dawyndt for their help with the primer design and technical assistance. We thank Mike Wingfield for collecting the diseased *Eucalyptus* material, Teresa Goszczynska for allowing us to include the strains isolated from onion and maize, Mrs. Mohr for providing the Brenner strains from the CDC, and Emma Steenkamp for helping with the phylogenetic analysis.

REFERENCES

1. **Beji, A., J. Mergaert, F. Gavini, D. Izard, K. Kersters, H. Leclerc and J. De Ley.** 1988. Subjective synonymy of *Erwinia herbicola*, *Erwinia milletiae*, and *Enterobacter agglomerans* and redefinition of the taxon by genotypic and phenotypic data. Int. J. Syst. Bacteriol. **38:** 77-88.

2. **Brady, C.L., S.N. Venter, I. Cleenwerck, M. Vancanneyt, J. Swings and T.A. Coutinho.** 2007. A FAFLP system for the improved identification of plant-pathogenic and plant-associated species of the genus *Pantoea*. Syst. Appl. Microbiol. **30:** 413-417.

3. **Brenner, D.J., G.R. Fanning, J.K. Leete Knutson, A.G. Steigerwalt and M.I. Krichevsky.** 1984. Attempts to classify Herbicola Group-*Enterobacter agglomerans* strains by deoxyribonucleic acid hybridization and phenotypic tests. Int. J. Syst. Bacteriol. **34:** 45-55.

4. **Burr, T.J., B.H. Katz, G.S. Abawi and D.C. Crosier.** 1991. Comparison of tumorigenic strains of *Erwinia herbicola* isolated from table beet with *E. h. gypsophilae*. Plant Dis. **75:** 855-858.

5. **Cooksey, D.A.** 1986. Galls of *Gypsophila paniculata* caused by *Erwinia herbicola*. Plant Dis. **70:** 464-468.

6. **Cother, E.J., R. Reinke, C. McKenzie, V.M. Lanoiselet and D.H. Noble.** 2004. An unusual stem necrosis of rice caused by *Pantoea ananas* and the first record of this pathogen on rice in Australia. Austral. Plant Pathol. **33:** 495-503.

7. **Coutinho, T.A., O. Preisig, J. Mergaert, M.C. Cnockaert, K.H. Riedel, J. Swings and M.J. Wingfield.** 2002. Bacterial blight and dieback of *Eucalyptus* species, hybrids and clones in South Africa. Plant Dis. **86:** 20-25.

8. **Cunningham, C.W.** 1997. Can three incongruence tests predict when data should be combined? Mol. Biol. Evol. **14:** 733-740.

9. **Dauga, C.** 2002. Evolution of the *gyrB* gene and the molecular phylogeny of *Enterobacteriaceae*: a model molecule for molecular systematic studies. Int. J. Syst. Evol. Microbiol. **52**: 531–547.

10. **De Baere, T., R. Verhelst, C. Labit, G. Verschraegen, G. Wauters, G. Claeys and M. Vaneechoutte.** 2004. Bacteremic infection with *Pantoea ananatis*. J. Clin. Microbiol. **42:** 4393-4395.

11. **Drancourt, M., C. Bollet, A. Carta and P. Rousselier.** 2001. Phylogenetic analyses of *Klebsiella* species delineate *Klebsiella* and *Raoultella* gen. nov., with description of *Raoultella ornithinolytica* comb. nov., *Raoultella terrigena* comb. nov. and *Raoultella planticola* comb. nov. Int. J. Syst. Evol. Microbiol. **51:** 925-932

12. **Fullerton, D.G., A.A. Lwina and S. Lalb.** 2007. *Pantoea agglomerans* liver abscess presenting with a painful thigh. Eur. J. Gastroenterol. Hepatol. **19:** 433–435.

13. **Gavini, F., B. Lefebvre and H. Leclerc.** 1983. Taxonomic study of strains belonging or related to the genus *Erwinia*, herbicola group, and to the species *Enterobacter agglomerans*. Syst. Appl. Microbiol. **4:** 218-235.

14. **Gavini, F., J. Mergaert, A. Beji, C. Mielcarek, D. Izard, K. Kersters and J. De Ley.** 1989. Transfer of *Enterobacter agglomerans* (Beijerinck 1988) Ewing and Fife 1972 to *Pantoea* gen. nov. as *Pantoea agglomerans* comb. nov. and description of *Pantoea dispersa* sp. nov. Int. J. Syst. Bacteriol. **39:** 337-345.

15. **Goszczynska, T., W.J. Botha, S.N. Venter and T.A. Coutinho.** 2007. Isolation and identification of the causal agent of brown stalk rot, a new disease of maize in South Africa. Plant Dis. **91:** 711-718.

16. **Grimont, P.A.D. and F. Grimont.** 2005. Genus: *Pantoea*. In: D.J. Brenner, N.R. Krieg and J.T. Staley, eds. Bergey's Manual of Systematic Bacteriology. 2nd Edition. Volume Two, The *Proteobacteria*. Part B, The *Gammaproteobacteria*.

17. **Guindon, S. and O. Gascuel.** 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. **52:** 696-704.

18. **Hall, T.A.** 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids Symp. Ser. **41:** 95-98.

19. **Hedegaard, J., S.A. de A. Steffensen, N. Nørskov-Lauritsen, K.K. Mortensen and H.U. Sperling-Petersen.** 1999. Identification of *Enterobacteriaceae* by partial sequencing of the gene encoding translation initiation factor 2. Int. J. Syst. Bacteriol. **49**: 1531–1538.

20. **Hollis, D.G., F.W. Hickman, G.R. Fanning, JJ. Farmer III, R.E. Weaver and D.J. Brenner.** 1981. *Tatumella ptyseos* gen. nov., sp. nov., a member of the family *Enterobacteriaceae* found in clinical specimens. J. Clin. Microbio. 14: 79-88

21. **Kageyama, B., M. Nakae, S. Yagi and T. Sonoyama.** 1992. *Pantoea punctata* sp. nov., *Pantoea citrea* sp. nov., and *Pantoea terrea* sp. nov. isolated from fruit and soil samples. Int. J. Syst. Bacteriol. **42:** 203-210.

22. **Kämpfer, P., S. Ruppel and R. Remus.** 2005. *Enterobacter radicincitans* sp. nov., a plant growth promoting species of the family *Enterobacteriaceae*. Syst Appl Microbiol **28:** 213–221.

23. **Konstantinidis, K.T., A. Ramette and J.M. Tiedje.** 2006. Towards a more robust assessment of intraspecies diversity, using fewer genetic markers. Appl. Environ. Microbiol. **72:** 7286-7293.

24. **Kratz, A., D. Greenberg, Y. Barki, E. Cohen and M. Lifshitz.** 2003. *Pantoea agglomerans* as a cause of septic arthritis after palm tree thorn injury; case report and literature review. Arch. Dis. Child. **88:** 542-544.

25. **Li, X., D. Zhang, F. Chen, J. Ma, Y. Dong and L. Zhang.** 2004. *Klebsiella singaporensis* sp. nov., a novel isomaltulose-producing bacterium. Int J Syst Evol Microbiol **54:** 2131–2136.

26. **Lim, P., S. Chen, C. Tsai and M. Pai.** 2006. *Pantoea* peritonitis in a patient receiving chronic ambulatory peritoneal dialysis. Nephrology **11:** 97-99.

27. **McGhee, G.C., E.L. Schnabel, K. Maxson-Stein, B. Jones, V.K. Stromberg, G.H. Lacy, and A.L. Jones.** 2002. Relatedness of chromosomal and plasmid DNAs of *Erwinia pyrifoliae* and *Erwinia amylovora*. Appl. Environ. Microbiol. **68:** 6182-6192.

28. **Mergaert, J., L. Verdonck, and K. Kersters.** 1993. Transfer of *Erwinia ananas* (synonym, *Erwinia uredovora*) and *Erwinia stewartii* to the Genus *Pantoea* emend. as *Pantoea ananas* (Serrano 1928) comb. nov. and *Pantoea stewartii* (Smith 1898) comb. nov., Respectively, and Description of *Pantoea stewartii* subsp. *indologenes* subsp.nov. Int. J. Syst. Bacteriol. **43:** 162-173.

29. **Mollet, C., M. Drancourt and D. Raoult.** 1997. *rpoB* Sequence analysis as a novel basis for bacterial identification. Mol. Microbiol. **26:** 1005-1011.

30. **Niemann, S., A. Pühler, H.-V. Tichy, R. Simon and W. Selbitschka.** 1997. Evaluation of the resolving power of three different DNA fingerprinting methods to discriminate among isolates of a natural *Rhizobium meliloti* population. J. Appl. Microbiol. **82:** 477-484

31. **Paradis, S., M. Biossinot, N. Paquette, S.D. Bélanger, E.A. Martel, D.K. Boudreau, F.J. Picard, M. Ouellette, P.H. Roy and M.G. Bergeron.** 2005. Phylogeny of the *Enterobacteriaceae* based on genes encoding elongation factor Tu and F-ATPase β-subunit. Int. J. Syst. Evol. Microbiol. **55:** 2013-2025.

32. **Posada, D. and K.A. Crandall.** 1998. Modeltest: testing the model of DNA substitution. Bioinformatics **14:** 817-818.

33. **Pujol, C.J. and C.I. Kado.** 2000. Genetic and biochemical characterization of the pathway in *Pantoea citrea* leading to pink disease of pineapple. J. Bacteriol. **182:** 2230-2237.

34. **Schmid, H., C. Weber and J.R. Bogner.** 2003. Isolation of a *Pantoea dispersa*-like strain from a 71-year-old woman with acute myeloid leukemia and multiple myeloma. Infection **31:** 66-67.

35. **Stephan R., S. van Trappen, I. Cleenwerck, M. Vancanneyt, P. de Vos and A. Lehner.** 2007. *Enterobacter turicensis* sp.nov. and *Enterobacter helveticus* sp.nov., isolated from fruit powder. Int. J. Syst. Evol. Microbiol. **57:** 820-826.

36. **Stewart, F.C.** 1897. A bacterial disease of sweet corn. New York State Agric. Exp. Stn Bull. **130:** 422-439.

37. **Swofford, D.L.** 2000. PAUP*: Phylogenetic Analysis Using Parsimony and other methods (software). Sinauer Associates, Sunderland, MA.

38. **Thompson, J.D., T.J. Gibson, F. Plewniak, F. Jeanmougin and D.G. Higgins.** 1997. The ClustalX-Windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucl. Acids Res. **25:** 4876-4882.

39. **Van Rostenberghe, H., R. Noraida, W.I. Wan Pauzi, H. Habsah, M. Zeehaida, A.R. Rosliza, I. Fatimah, N.Y. Nik Sharimah and H. Maimunah.** 2006. The clinical picture of neonatal infection with *Pantoea* species. Jpn. J. Infect. Dis. **59:** 120-121.

40. **Verdonck, L., J. Mergaert, C. Rijckaert, J. Swings, K. Kersters and J. De Ley.** 1987. The genus *Erwinia*: A numerical analysis of phenotypic features. Int. J. Syst. Bacteriol. **37:** 4-18

41. **Xia, X., and Z. Xie.** 2001. DAMBE: Data analysis in molecular biology and evolution. J. Heredity **92:** 371-373.

42. **Young, J.M. and D.C. Park.** 2007. Relationships of plant pathogenic enterobacteria based on partial *atpD*, *carA* and *recA* as individual and concatenated nucleotide and peptide sequences. Syst. Appl. Microbiol. **30:** 343-354

Table 1: Strains used in this study

Species name	Strain no.	Source	Place of isolation
Pantoea agglomerans	LMG 1286^T	Human	Zimbabwe
	LMG 2554	Scarlet runner bean	UK
	LMG 2565	Cereal	Canada
	LMG 2572	Wheat	Canada
	LMG 2596	Onion	South Africa
	LMG 2660	Wisteria	Japan
	SUH 2 (syn. LMG 2596)	Onion	South Africa
Pantoea agglomerans pv. *gypsophilae*	LMG 2553	Gypsophila	Unknown
Pantoea agglomerans pv. *betae*	BCC 734	Beet	Unknown
Pantoea ananatis	LMG 2665^T	Pineapple	Brazil
	LMG 2668	Pineapple	Hawaii
	LMG 2676	*Puccinia graminis*, uredia	USA
	LMG 2678	*Puccinia graminis*, uredia	Zimbabwe
	LMG 20103	*Eucalyptus*	South Africa
	LMG 20104	*Eucalyptus*	South Africa
	LMG 20106	*Eucalyptus*	South Africa
	BCC 114	*Eucalyptus* & Colletogloeopsis canker	South Africa
	BCC 150 = ATCC 35400	Honeydew melon	USA
	LMG 24190 = R-27854	Onion	USA
	LMG 24193 = R-27860	Onion seed	South Africa
	BD 333	Onion seed	South Africa
	BD 336	Onion seed	South Africa
	LMG 24191 = R-27858	Maize	South Africa
	LMG 24192 = R-27859	Maize	South Africa
	BD 561	Maize	South Africa
	BD 577	Maize	South Africa
	BD 588	Maize	South Africa
	BD 602	Maize	South Africa
	BD 622	Maize	South Africa
	BD 640	Maize	South Africa
	BD 647	Maize	South Africa
Pantoea stewartii ssp. *stewartii*	LMG 2715^T	Corn	USA
	LMG 2713	Corn	USA
	LMG 2718	Corn	USA

Pantoea stewartii ssp. *indologenes*	LMG 2632^T	Fox millet	India
	LMG 2630	Guar gum powder	Unknown
	LMG 2631	Millet	India
	LMG 2671	Pineapple	Hawaii
	LMG 2673	Pineapple	Hawaii
	BCC 099	Sudangrass	USA
	BCC 118	*Eucalyptus* & Colletogloeopsis canker	South Africa
Pantoea dispersa	LMG 2603^T	Soil	Japan
	LMG 2602	Sorghum	India
	LMG 2604	Wild rose	Netherlands
	LMG 2749	Human	Unknown
Pantoea citrea	LMG 22049^T = SHS 2003	Mandarin orange	Japan
	LMG 23359	Pineapple	Philippines
	LMG 23360	Pineapple	Philippines
Pantoea punctata	LMG 22050^T = SHS 2006	Mandarin orange	Japan
	LMG 22097 = SHS 2004	Mandarin orange	Japan
	LMG 22098 = SHS 2005	Persimmon	Japan
	LMG 23562 = SHS 2004	Mandarin orange	Japan
	LMG 23563 = SHS 2007	Mandarin orange	Japan
	CCUG 30157 = SHS 2004	Mandarin orange	Japan
	CCUG 30160 = SHS 2007	Mandarin orange	Japan
Pantoea terrea	LMG 22051^T = SHS 2008	Soil	Japan
	LMG 23564 = SHS 2009	Soil	Japan
	LMG 23565 = SHS 2010	Soil	Japan
	CCUG 30162 = SHS 2009	Soil	Japan
	CCUG 30163 = SHS 2010	Soil	Japan
Pantoea sp.	LMG 24194 = R-25665	*Eucalyptus*	Argentina
	LMG 24195 = R-24584	*Eucalyptus*	Uruguay
	LMG 24196 = R-25674	*Eucalyptus*	Argentina
	LMG 24197 = R-25678	*Eucalyptus*	Uruguay
	LMG 24198 = R-25679	*Eucalyptus*	Uruguay
	LMG 24199 = R-21566	*Eucalyptus*	Uganda
	LMG 24200 = R-31523	*Eucalyptus*	Uganda
	BCC 002	*Eucalyptus*	Argentina
	BCC 004	*Eucalyptus*	Argentina
	BCC 006	*Eucalyptus*	Argentina
	BCC 067	*Eucalyptus*	Colombia
	BCC 072	*Eucalyptus*	Uruguay

Pantoea sp.	BCC 075	*Eucalyptus*	Uruguay
	BCC 079	*Eucalyptus*	Uruguay
	BCC 081	*Eucalyptus*	Uruguay
	BCC 082	*Eucalyptus*	Uruguay
	BCC 107	*Eucalyptus*	Uganda
	BCC 208	*Eucalyptus*	Uganda
	BCC 427	*Eucalyptus*	Uganda
	BCC 756	*Eucalyptus*	Uruguay
	BCC 757	*Eucalyptus*	Uruguay
	BCC 760	*Eucalyptus*	Uruguay
	LMG 24201 = R-30991	Maize	South Africa
	BD 502	Maize	South Africa
	LMG 24202 = R-27853	Onion	USA
	LMG 24203 = R-21588	Onion	South Africa
	LMG 24248 = R-27856	Onion	South Africa
	LMG 2558 = NCPPB 1682	Balsam	India
	LMG 2560 = NCPPB 1941	Marigold	Unknown
Pantoea sp. (Brenner HG II)	LMG 5345 = CDC 3123-70	Human	USA
	R-35488 = CDC 238-70 = LMG 24526	Human	USA
	R-35489 = CDC 1778-70 = LMG 24527	Human	USA
	R-35490 = CDC 217-71 = LMG 24528	Human	USA
Pantoea sp. (Brenner HG IV)	LMG 2781 = CDC 1741-71	Human	USA
	LMG 5346 (syn. LMG 2781)	Human	USA
	R-35491 = CDC 3638-70 = LMG 24529	Human	USA
	R-35492 = CDC 5795-70 = LMG 24530	Human	USA
	R-35493 = CDC 6148-70 = LMG 24531	Human	USA
Pantoea sp. (Brenner HG V)	LMG 5343 = CDC 3482-71	Human	USA
	R-35494 = CDC 2928-68 = LMG 24532	Human	USA
	R-35495 = CDC 2525-70 = LMG 24533	Human	USA
	R-35496 = CDC 3527-71 = LMG 24534	Human	USA
Erwinia billingiae	LMG 2613^T	Pear	UK
Erwinia rhapontici	LMG 2688^T	Rhubarb	UK
Erwinia toletana	LMG 24162	Olive tree	Spain
Tatumella ptyseos	LMG 7888^T	Human	USA

Footnote: LMG = BCCM/LMG Bacteria Collection, Ghent University, Belgium. BCC = Bacterial Culture Collection, Forestry and Agricultural Biotechnology Institute, Pretoria, South Africa. ATCC = American Type Culture Collection, Rockville, Maryland, U.S.A. CCUG = Culture Collection, University of Göteborg, Sweden. CDC = Centres for Disease Control, Atlanta, Georgia, U.S.A. BD = Plant Pathogenic and Plant Protecting Bacteria (PPPPB) Culture Collection, ARC-PPRI, Pretoria, South Africa. NCPPB = National Collection of Plant Pathogenic Bacteria, York, United Kingdom

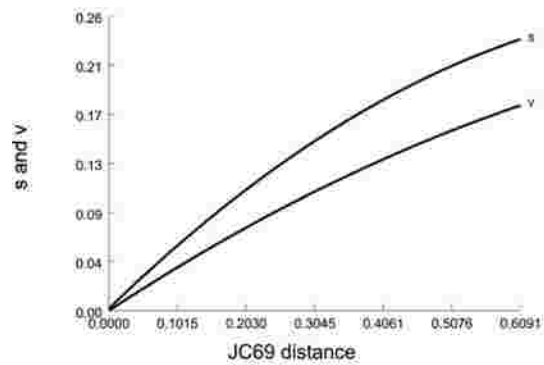
^{T} = type strain

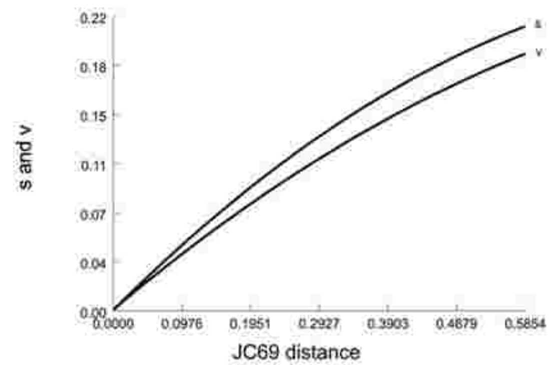
Table 2: Amplification and sequencing primers for *rpoB*, *atpD*, *gyrB* and *infB*

Amplification primers	**Sequence (5’ ? 3’)**
rpoB CM7-F	AAC CAG TTC CGC GTT GGC CTG
rpoB CM31b-R	CCT GAA CAA CAC GCT CGG A
atpD 01-F	RTA ATY GGM GCS GTR GTN GAY GT
atpD 02-R	TCA TCC GCM GGW ACR TAW AYN GCC TG
gyrB 01-F	TAA RTT YGA YGA YAA CTC YTA YAA AGT
gyrB 02-R	CMC CYT CCA CCA RGT AMA GTT
infB 01-F	ATY ATG GGH CAY GTH GAY CA
infB 02-R	ACK GAG TAR TAA CGC AGA TCC A

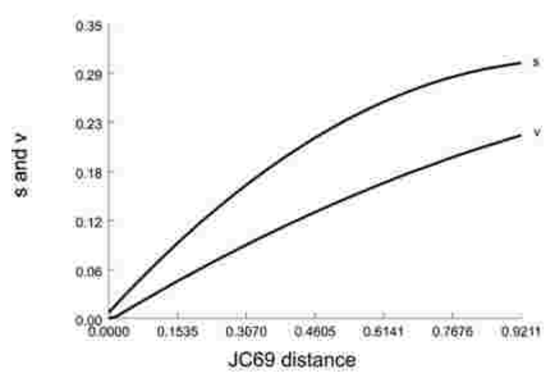
Sequencing primers	**Sequence (5’ ? 3’)**
rpoB CM81-F	CAG TTC CGC GTT GGC CTG
rpoB CM81b-F	TGA TCA ACG CCA AGC C
rpoB CM32b-R	CGG ACC GGC CTG ACG TTG CAT
atpD 03-F	TGC TGG AAG TKC AGC ARC AG
atpD 04-R	CCM AGY ART GCG GAT ACT TC
gyrB 07-F	GTV CGT TTC TGG CCV AG
gyrB 08-R	CTT TAC GRC GKG TCA TWT CAC
infB 03-F	ACG GBA TGA TYA CST TCC TGG
infB 04-R	AGY TTA GAT TTC TGC TGA CG

Figure 1: Substitution saturation of MLSA housekeeping genes **a)** *rpoB* **b)** *atpD* **c)** *gyrB* **d)** *infB*. Transitions (s) and tranversions (v) at the third codon are plotted against Jukes-Cantor's genetic distance (JC69). No substitution saturation is visible at the third codon for *rpoB* and *atpD*. There is possible substitution saturation for *gyrB*, and definite saturation for *infB* at the third codons.





[redacted]



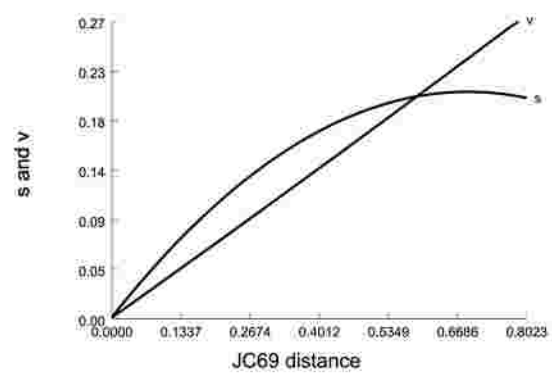


Fig. 1c

Fig. 1d

Figure 2a: Maximum likelihood tree based on partial *rpoB* sequences of 103 *Pantoea* strains. The tree was generated by the Phyml software using the Kimura (K3P) model as selected by Modeltest. Bootstrap values after 1 000 replicates are shown. *Citrobacter rodentium* was included as an outgroup.

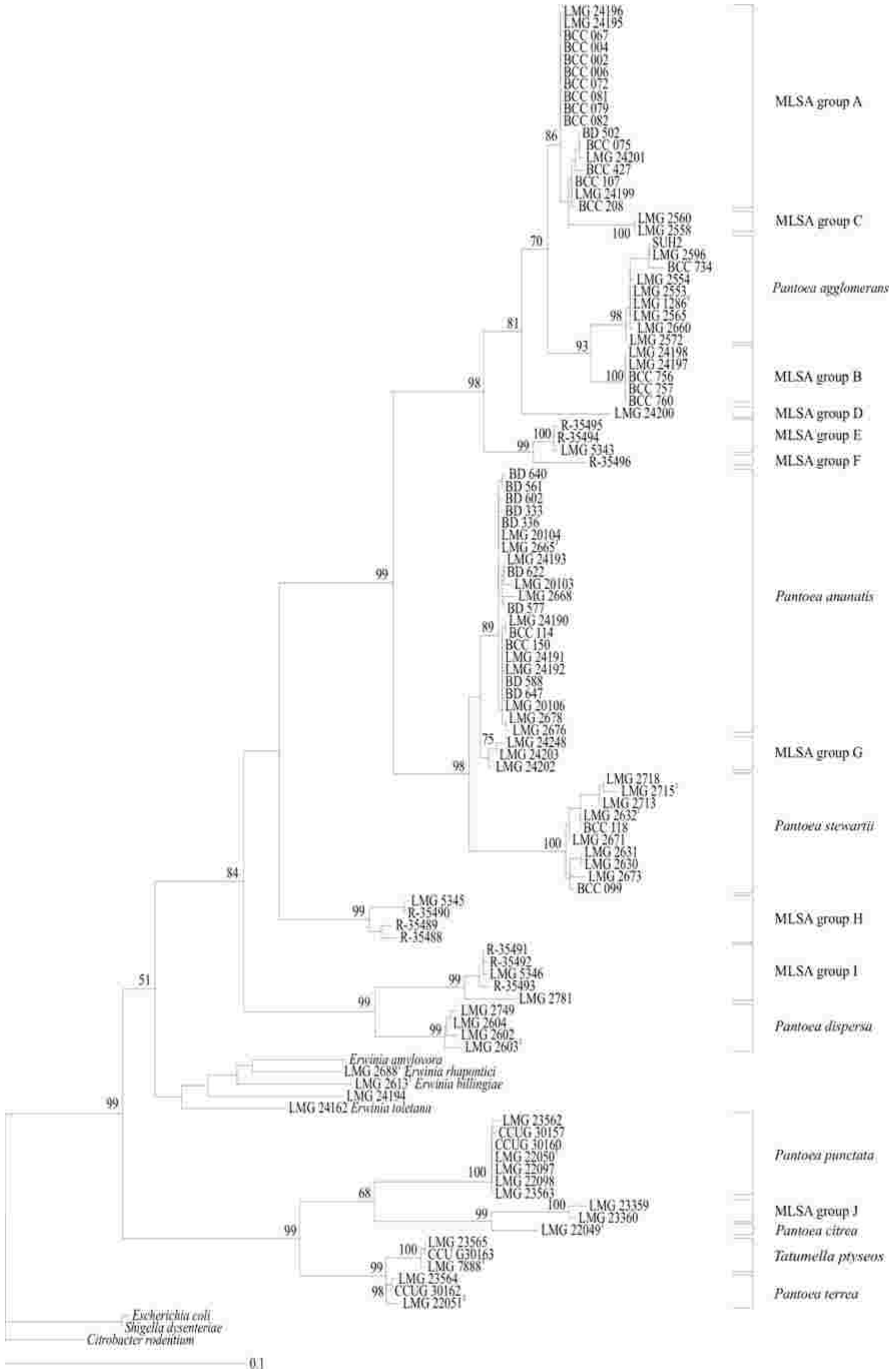

LMG 24196
LMG 24195
BCC 067
BCC 004
BCC 002
BCC 006
BCC 072
BCC 081
BCC 079
BCC 082
BD 502
BCC 075
LMG 24201
BCC 427
BCC 107
LMG 24199
BCC 208
LMG 2560
LMG 2558
SUH2
LMG 2596
BCC 734
LMG 2554
LMG 2553
LMG 1286
LMG 2565
LMG 2660
LMG 2572
LMG 24198
LMG 24197
BCC 756
BCC 757
BCC 760
LMG 24200
R-35495
R-35494
LMG 5343
R-35496
BD 640
BD 561
BD 602
BD 333
BD 336
LMG 20104
LMG 2665
LMG 24193
BD 622
LMG 20103
LMG 2668
BD 577
LMG 24190
BCC 114
BCC 150
LMG 24191
LMG 24192
BD 588
BD 647
LMG 20106
LMG 2678
LMG 2676
LMG 24248
LMG 24203
LMG 24202
LMG 2718
LMG 2715
LMG 2713
LMG 2632
BCC 118
LMG 2671
LMG 2631
LMG 2630
LMG 2673
BCC 099
LMG 5345
R-35490
R-35489
R-35488
R-35491
R-35492
LMG 5346
R-35493
LMG 2781
LMG 2749
LMG 2604
LMG 2602
LMG 2603
Erwinia amylovora
LMG 2688 Erwinia rhapontici
LMG 2613 Erwinia billingiae
LMG 24194
LMG 24162 Erwinia toletana
LMG 23562
CCUG 30157
CCUG 30160
LMG 22050
LMG 22097
LMG 22098
LMG 23563
LMG 23359
LMG 23360
LMG 22049
LMG 23565
CCU G30163
LMG 7888
LMG 23564
CCUG 30162
LMG 22051
Escherichia coli
Shigella dysenteriae
Citrobacter rodentium
0.1
86
70
100
81
98
93
100
98
99
100
99
89
75
98
100
84
99
51
99
99
99
99
100
68
99
100
99
100
99
98
MLSA group A
MLSA group C
Pantoea agglomerans
MLSA group B
MLSA group D
MLSA group E
MLSA group F
Pantoea ananatis
MLSA group G
Pantoea stewartii
MLSA group H
MLSA group I
Pantoea dispersa
Pantoea punctata
MLSA group J
Pantoea citrea
Tatumella ptyseos
Pantoea terrea

Figure 2b: Maximum likelihood tree based on partial *atpD* sequences of 103 *Pantoea* strains. The tree was generated by the Phyml software using the general time reversible (GTR) model as selected by Modeltest. Bootstrap values after 1 000 replicates are shown. *Citrobacter rodentium* was included as an outgroup.

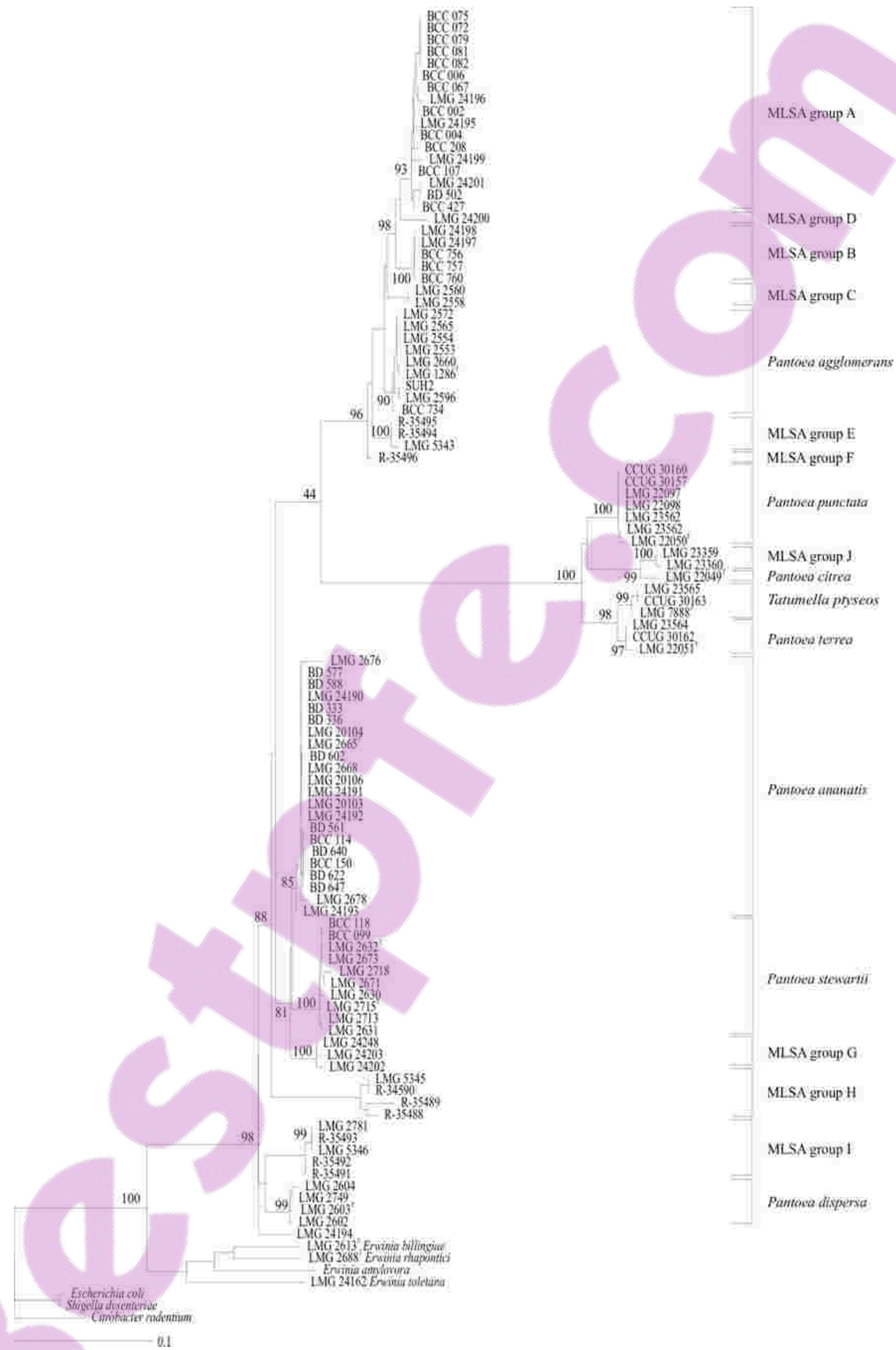

BCC 075
BCC 072
BCC 079
BCC 081
BCC 082
BCC 006
BCC 067
LMG 24196
BCC 002
LMG 24195
BCC 004
BCC 208
LMG 24199
93
BCC 107
LMG 24201
BD 502
BCC 427
LMG 24200
98
LMG 24198
LMG 24197
BCC 756
BCC 757
100
BCC 760
LMG 2560
LMG 2558
LMG 2572
LMG 2565
LMG 2554
LMG 2553
LMG 2660
LMG 1286
SUH2
90
LMG 2596
BCC 734
96
R-35495
100
R-35494
LMG 5343
R-35496
44
CCUG 30160
CCUG 30157
LMG 22097
100
LMG 22098
LMG 23562
LMG 23562
LMG 22050
100
LMG 23359
LMG 23360
100
99
LMG 22049
LMG 23565
99
CCUG 30163
LMG 7888
98
LMG 23564
CCUG 30162
97
LMG 22051
LMG 2676
BD 577
BD 588
LMG 24190
BD 333
BD 336
LMG 20104
LMG 2665
BD 602
LMG 2668
LMG 20106
LMG 24191
LMG 20103
LMG 24192
BD 561
BCC 114
BD 640
BCC 150
BD 622
85
BD 647
LMG 2678
LMG 24193
88
BCC 118
BCC 099
LMG 2632
LMG 2673
LMG 2718
LMG 2671
LMG 2630
81
100
LMG 2715
LMG 2713
LMG 2631
LMG 24248
100
LMG 24203
LMG 24202
LMG 5345
R-34590
R-35489
R-35488
98
99
LMG 2781
R-35493
LMG 5346
R-35492
R-35491
LMG 2604
100
99
LMG 2749
LMG 2603
LMG 2602
LMG 24194
LMG 2613 Erwinia billingiae
LMG 2688 Erwinia rhapontici
Erwinia amylovora
LMG 24162 Erwinia toletana
Escherichia coli
Shigella dysenteriae
Citrobacter rodentium
0.1
MLSA group A
MLSA group D
MLSA group B
MLSA group C
Pantoea agglomerans
MLSA group E
MLSA group F
Pantoea punctata
MLSA group J
Pantoea citrea
Tatumella ptyseos
Pantoea terrea
Pantoea ananatis
Pantoea stewartii
MLSA group G
MLSA group H
MLSA group I
Pantoea dispersa

Figure 2c: Maximum likelihood tree based on partial *gyrB* sequences of 103 *Pantoea* strains. The tree was generated by the Phyml software using the general time reversible (GTR) model as selected by Modeltest. Bootstrap values after 1 000 replicates are shown. *Citrobacter rodentium* was included as an outgroup.

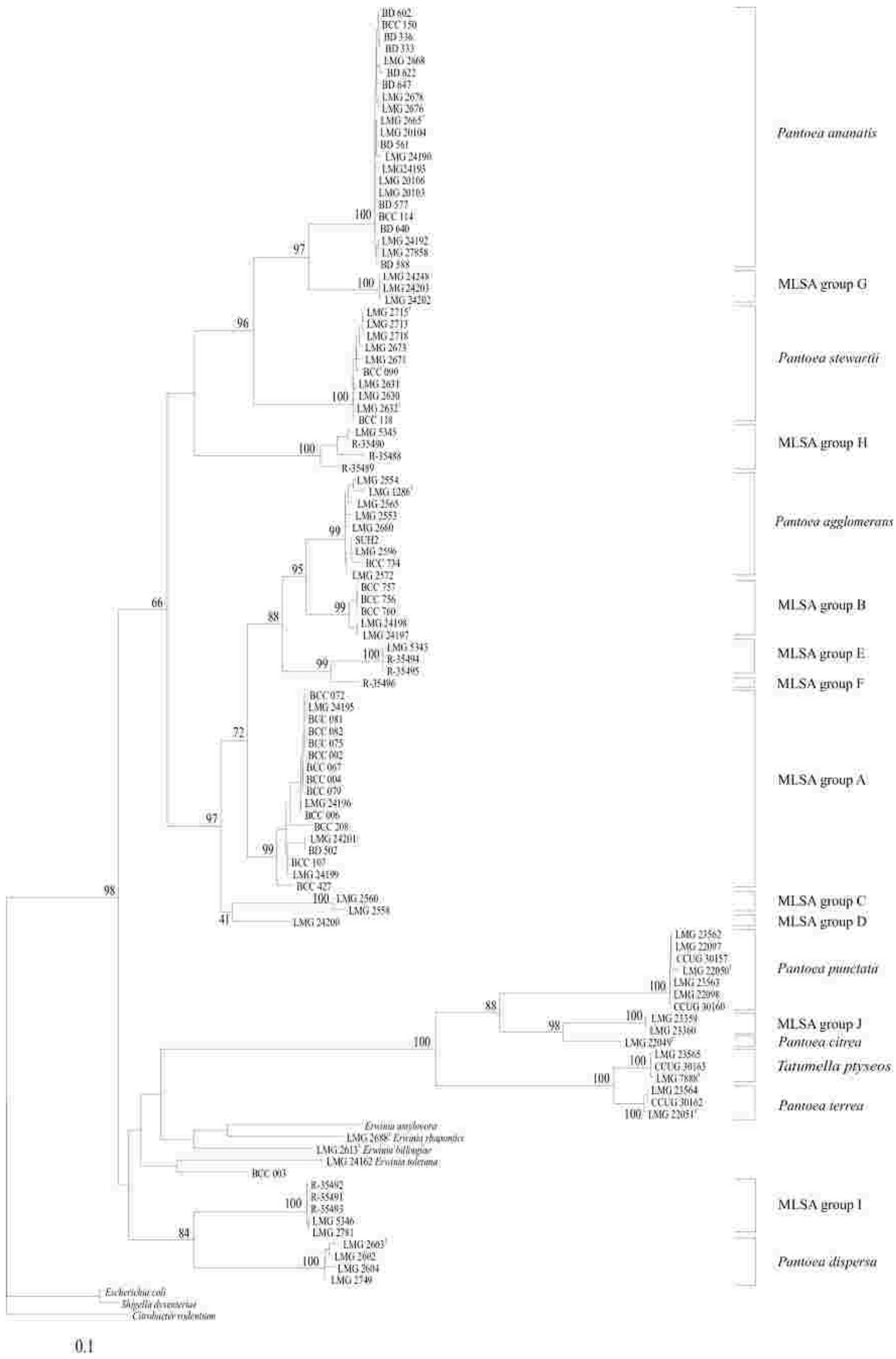

Pantoea ananatis
MLSA group G
Pantoea stewartii
MLSA group H
Pantoea agglomerans
MLSA group B
MLSA group E
MLSA group F
MLSA group A
MLSA group C
MLSA group D
Pantoea punctata
MLSA group J
Pantoea citrea
Tatumella ptyseos
Pantoea terrea
MLSA group I
Pantoea dispersa
Erwinia amylovora
Erwinia rhapontici
Erwinia billingiae
Erwinia toletana
Escherichia coli
Shigella dysenteriae
Citrobacter rodentium
0.1

Figure 2d: Maximum likelihood tree based on partial *infB* sequences of 103 *Pantoea* strains. The tree was generated by the Phyml software using the Tamura-Nei (TN93) model as selected by Modeltest. Bootstrap values after 1 000 replicates are shown. *Citrobacter rodentium* was included as an outgroup.

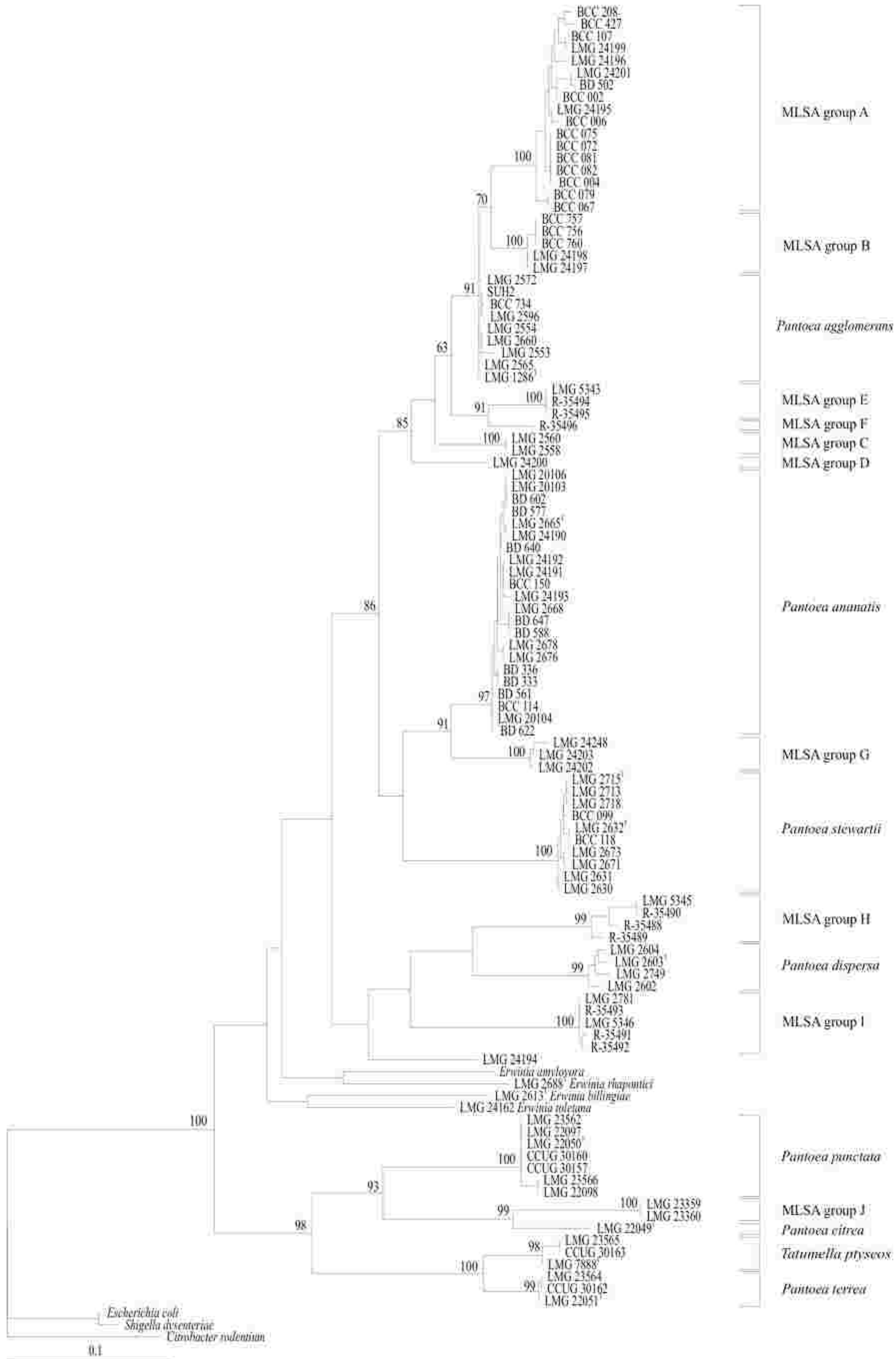

BCC 208
BCC 427
BCC 107
LMG 24199
LMG 24196
LMG 24201
BD 502
BCC 002
LMG 24195
BCC 006
BCC 075
BCC 072
BCC 081
BCC 082
BCC 004
BCC 079
BCC 067
BCC 757
BCC 756
BCC 760
LMG 24198
LMG 24197
LMG 2572
SUH2
BCC 734
LMG 2596
LMG 2554
LMG 2660
LMG 2553
LMG 2565
LMG 1286
LMG 5343
R-35494
R-35495
R-35496
LMG 2560
LMG 2558
LMG 24200
LMG 20106
LMG 20103
BD 602
BD 577
LMG 2665
LMG 24190
BD 640
LMG 24192
LMG 24191
BCC 150
LMG 24193
LMG 2668
BD 647
BD 588
LMG 2678
LMG 2676
BD 336
BD 333
BD 561
BCC 114
LMG 20104
BD 622
LMG 24248
LMG 24203
LMG 24202
LMG 2715
LMG 2713
LMG 2718
BCC 099
LMG 2632
BCC 118
LMG 2673
LMG 2671
LMG 2631
LMG 2630
LMG 5345
R-35490
R-35488
R-35489
LMG 2604
LMG 2603
LMG 2749
LMG 2602
LMG 2781
R-35493
LMG 5346
R-35491
R-35492
LMG 24194
Erwinia amylovora
LMG 2688 Erwinia rhapontici
LMG 2613 Erwinia billingiae
LMG 24162 Erwinia toletana
LMG 23562
LMG 22097
LMG 22050
CCUG 30160
CCUG 30157
LMG 23566
LMG 22098
LMG 23359
LMG 23360
LMG 22049
LMG 23565
CCUG 30163
LMG 7888
LMG 23564
CCUG 30162
LMG 22051
Escherichia coli
Shigella dysenteriae
Citrobacter rodentium
0.1
100
70
100
91
63
91
100
85
100
86
97
91
100
100
99
99
100
100
100
93
99
100
98
98
100
99
MLSA group A
MLSA group B
Pantoea agglomerans
MLSA group E
MLSA group F
MLSA group C
MLSA group D
Pantoea ananatis
MLSA group G
Pantoea stewartii
MLSA group H
Pantoea dispersa
MLSA group I
Pantoea punctata
MLSA group J
Pantoea citrea
Tatumella ptyseos
Pantoea terrea

Figure 3: Maximum likelihood tree based on the concatenated partial sequences of *rpoB*, *atpD*, *gyrB* and *infB* of 103 *Pantoea* strains. The tree was generated by the Phyml software using the general time reversible (GTR) model as selected by Modeltest. Bootstrap values after 1 000 replicates are shown. *Citrobacter rodentium* was included as an outgroup.

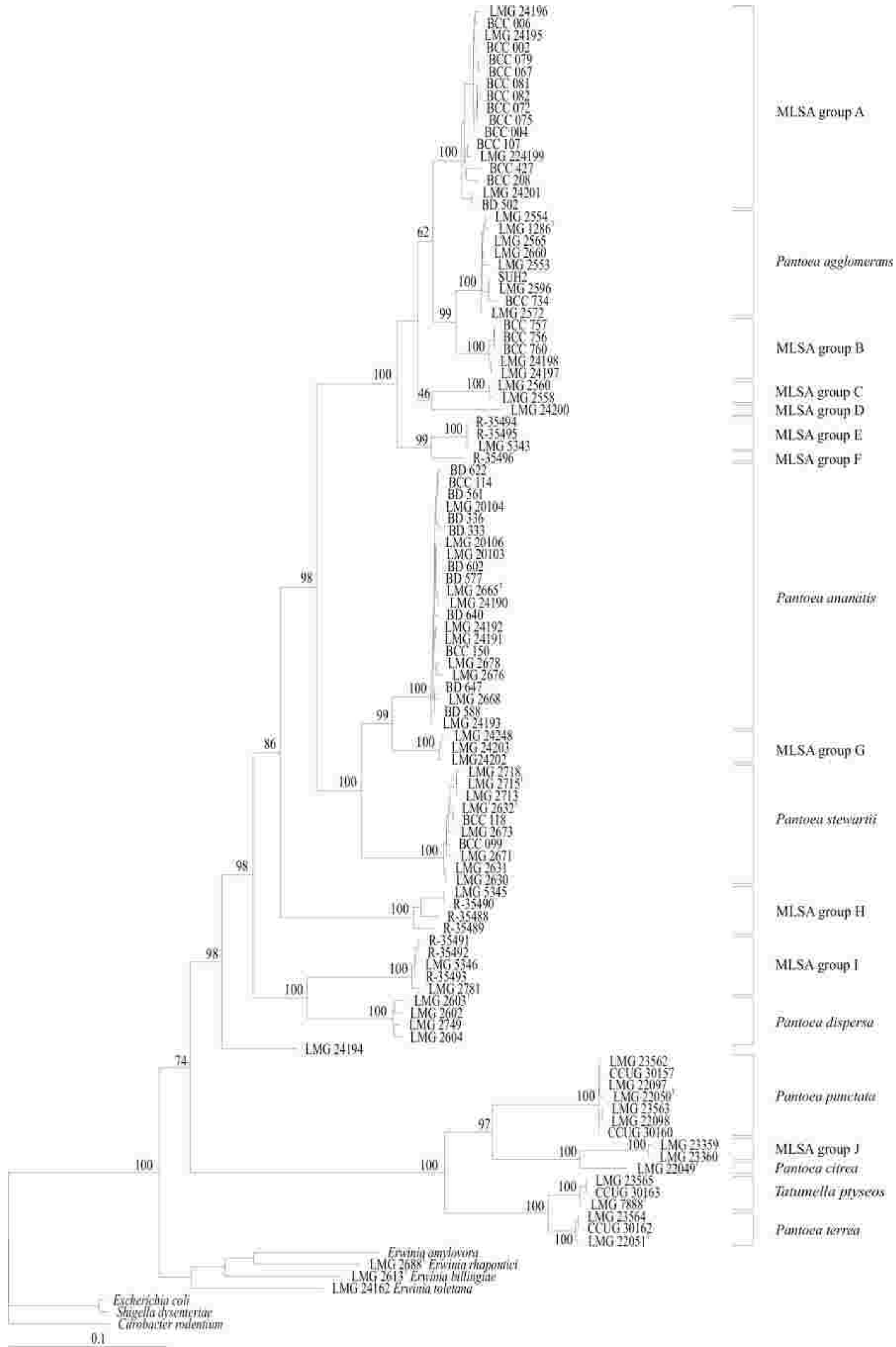
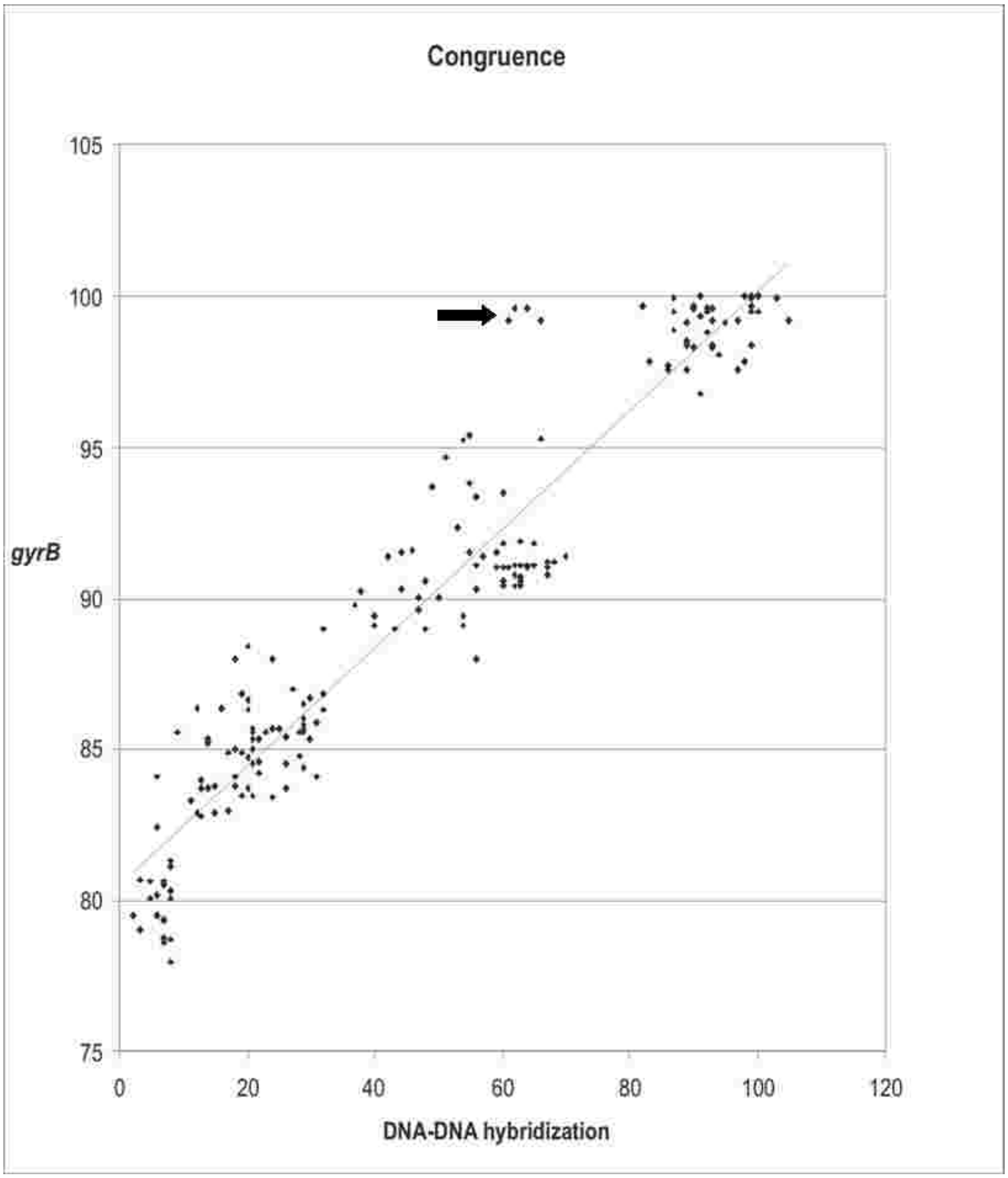

LMG 24196
BCC 006
LMG 24195
BCC 002
BCC 079
BCC 067
BCC 081
BCC 082
BCC 072
BCC 075
BCC 004
BCC 107
LMG 224199
BCC 427
BCC 208
LMG 24201
BD 502
LMG 2554
LMG 1286T
LMG 2565
LMG 2660
LMG 2553
SUH2
LMG 2596
BCC 734
LMG 2572
BCC 757
BCC 756
BCC 760
LMG 24198
LMG 24197
LMG 2560
LMG 2558
LMG 24200
R-35494
R-35495
LMG 5343
R-35496
BD 622
BCC 114
BD 561
LMG 20104
BD 336
BD 333
LMG 20106
LMG 20103
BD 602
BD 577
LMG 2665T
LMG 24190
BD 640
LMG 24192
LMG 24191
BCC 150
LMG 2678
LMG 2676
BD 647
LMG 2668
BD 588
LMG 24193
LMG 24248
LMG 24203
LMG24202
LMG 2718
LMG 2715T
LMG 2713
LMG 2632
BCC 118
LMG 2673
BCC 099
LMG 2671
LMG 2631
LMG 2630
LMG 5345
R-35490
R-35488
R-35489
R-35491
R-35492
LMG 5346
R-35493
LMG 2781
LMG 2603T
LMG 2602
LMG 2749
LMG 2604
LMG 24194
LMG 23562
CCUG 30157
LMG 22097
LMG 22050T
LMG 23563
LMG 22098
CCUG 30160
LMG 23359
LMG 23360
LMG 22049T
LMG 23565
CCUG 30163
LMG 7888
LMG 23564
CCUG 30162
LMG 22051T
Erwinia amylovora
LMG 2688 Erwinia rhapontici
LMG 2613 Erwinia billingiae
LMG 24162 Erwinia toletana
Escherichia coli
Shigella dysenteriae
Citrobacter rodentium
0.1
MLSA group A
Pantoea agglomerans
MLSA group B
MLSA group C
MLSA group D
MLSA group E
MLSA group F
Pantoea ananatis
MLSA group G
Pantoea stewartii
MLSA group H
MLSA group I
Pantoea dispersa
Pantoea punctata
MLSA group J
Pantoea citrea
Tatumella ptyseos
Pantoea terrea

Figure 4: Scatter plot comparing the congruence of partial *gyrB* sequence data and DNA-DNA hybridization values for the genus *Pantoea*. The arrow indicates strains of *P. stewartii* which have lower DNA-DNA hybridization values, but high *gyrB* sequence similarity.


Congruence
gyrB
DNA-DNA hybridization
105
100
95
90
85
80
75
0
20
40
60
80
100
120

CHAPTER 4

Pantoea vagens sp. nov., *Pantoea eucalypti* sp. nov., *Pantoea deleyii* sp. nov. and *Pantoea anthophila* sp. nov., four novel species belonging to the genus *Pantoea*

As submitted to: International Journal of Systematic and Evolutionary Microbiology

Summary

Bacteria, isolated from *Eucalyptus* leaves and shoots showing symptoms of blight and die-back collected in Uganda, Uruguay and Argentina, and from maize displaying brown stalk rot symptoms in South Africa, were tentatively placed in the genus *Pantoea* on the basis of phenotypic and biochemical tests. These isolates, together with two strains (LMG 2558 and LMG 2560) previously assigned to *Pantoea agglomerans* based on protein electropherograms (Beji *et al*., 1988) but later excluded from this species by Gavini *et al*. (1989), were further investigated using molecular techniques. 16S rRNA sequencing and multilocus sequence analyses (MLSA) revealed that the strains were phylogenetically closely related to *Pantoea agglomerans*, *Pantoea stewartii* and *Pantoea ananatis*. MLSA and AFLP analysis placed the strains into four separate clusters, not containing any of the type strains of species of the genus *Pantoea*. DNA-DNA hybridization confirmed the classification of the isolates into four novel species, for which the names *Pantoea vagens* sp. nov. (LMG 24199^T = BD 765^T), *Pantoea eucalypti* sp. nov. (LMG 24198^T = BD 766^T), *Pantoea deleyii* sp. nov. (LMG 24200^T = BD 767^T) and *Pantoea anthophila* sp. nov. (LMG 2558^T = NCPPB 1682^T) are proposed.

The genus *Pantoea* was described to include several species belonging to the "*Erwinia herbicola-Enterobacter agglomerans*" complex (Gavini *et al*., 1989). Presently, the genus comprises seven validly described species, namely, *Pantoea agglomerans* and *Pantoea dispersa* (Gavini *et al*., 1989), *Pantoea citrea, Pantoea punctata* and *Pantoea terrea* (Kageyama *et al*., 1992) and *Pantoea ananatis* and *Pantoea stewartii* (Mergaert *et al*., 1993). Species of *Pantoea* are diverse in their origin and geographical spread, and have been isolated from plants as well as from clinical samples. *Pantoea ananatis* is the causal agent of bacterial blight and die-back of *Eucalyptus* in South Africa (Coutinho *et al*., 2002). Recently, *Pantoea*-like strains were isolated from *Eucalyptus* trees exhibiting a similar disease in Uganda, Argentina and Uruguay in 2001 and were thought to belong to *P. ananatis*. *P. ananatis* was also found to cause brown stalk rot of maize in South Africa (Goszczynska *et al*., 2007). Isolated simultaneously with *P. ananatis* from maize, were additional *Pantoea*-like strains which also caused brown stalk rot. Multilocus sequence analysis (MLSA) has been shown to separate the isolates from *Eucalyptus* and maize into three clusters, indicating that these strains probably constitute three novel species of *Pantoea* (Brady *et al*., submitted). Also included in the MLSA study were two strains, LMG 2558 (= NCPPB 1682) and LMG 2560 (= NCPPB 1941), mentioned in the publication by Beji *et al*. (1988) belonging to protein profile group VII which were assigned to *Pantoea agglomerans* based on their electropherograms despite never being hybridized to the type strain of *Pantoea agglomerans*. It was suggested by Gavini *et al*. (1989) to exclude strains from protein profile group VII from *Pantoea agglomerans*, but to provisionally include these strains in *Pantoea* until their correct classification was determined. The MLSA results (Brady *et al*., submitted) indicated that strains LMG 2558 and LMG 2560 from Beji protein profile group VII constitute another novel species belonging to *Pantoea* as recommended by Gavini *et al*. (1989).

Isolates were obtained from *Eucalyptus* leaves showing typical bacterial blight symptoms including leaf spots and water-soaked lesions. The leaves were surface-sterilized, crushed with sterile water and the resulting suspension was streaked on nutrient agar and incubated at 30 °C for three days. Single colonies were obtained by re-streaking and incubation under the same conditions. Isolates from diseased maize plants were received from Dr T. Goszczynska (Plant Protection Research Institute, South Africa) (Goszczynska *et al*., 2007). Additional strains used in this study were obtained from the BCCM/LMG Bacteria Collection (http://www.belspo.be/bccm) and the Centers for Disease Control, Atlanta, Georgia, U.S.A. The strains used in this study are listed in Table 1.

Genomic DNA was extracted from all strains using the DNeasy Tissue Kit (Qiagen). Amplified fragment polymorphism analysis (AFLP) was performed according to the method previously published (Brady *et al*., 2007) using the selective primer combination Eco-C/Mse-GC. Band patterns were analysed with BioNumerics 4.0 (Applied Maths) and compared with a database containing profiles of reference strains of all validly described *Pantoea* species. A UPGMA dendrogram was constructed using the Pearson correlation. The isolates from *Eucalyptus* and maize were divided into three clusters by AFLP analysis and the two strains from the study by Beji *et al*. (1988) were contained in a separate cluster (see Supplementary Fig. A in IJSEM Online). These AFLP clusters did not contain any reference strains, suggesting that the strains belonged to novel species.

Complete 16S rRNA sequences were determined for selected strains from each AFLP cluster using the primers and conditions determined by Coenye *et al*. (1999). MLSA based on *rpoB*, *atpD*, *gyrB* and *infB* gene sequences was performed on each strain (Brady *et al*, submitted). The GenBank/EMBL accession numbers for the 16S rRNA gene sequences for *P. vagens* R-21566 (= LMG 24199^T), *P. eucalypti* R-25679 (= LMG 24198^T), *P. deleyii* R-31523 (=LMG 24200^T) and *P. anthophila* LMG 2558^T are EF688012, EF688009, EF688011 and EF688010, respectively and EU216734-EU216737 for Brenner's hybridization groups II, IV and V.

The sequences were aligned using ClustalX (Thompson *et al*., 1997) and the overhangs trimmed. The Modeltest 3.7 programme (Posada & Crandall, 1998) was then applied to the data sets to determine the best-fit evolutionary model to apply to each gene. Maximum likelihood and neighbour joining analyses were performed using Phyml (Guindon & Gascuel, 2003) and PAUP 4.0b10 (Swofford, 2000) respectively, by applying the models and parameters determined by Modeltest, (only Maximum likelihood phylogenetic trees are shown). Bootstrap analysis with 1000 replicates was performed on the trees to assess the reliability of the clusters. The 16S rRNA sequence similarity of all four novel species was greater than 98 % to *P. agglomerans*, *P.*

ananatis, *P. stewartii*, and *P. dispersa*. In the 16S rRNA phylogenetic tree, the four novel species cluster within the *Pantoea* "core" group along with *P. agglomerans*, *P. stewartii* and *P. ananatis* (Fig. 1). MLSA revealed that the isolates from *Eucalyptus* and maize, and the strains from Beji protein profile group VII form four well-supported clusters in the concatenated tree (Fig. 2) which were referred to as MLSA groups A, B,C and D (Brady *et al*., submitted). The four MLSA groups could also be clearly differentiated from Brenner's hybridization groups II, IV and V (Brenner *et al*., 1984), referred to as MLSA groups E, F, H and I in the *Pantoea* MLSA study (Brady *et al*., submitted) (Figs. 1 & 2), and belonging to *Pantoea* according to Grimont & Grimont (2005).

High quality DNA for DNA-DNA hybridization of strains was prepared by the method of Wilson (1987), with minor modifications (Cleenwerck *et al*., 2002). DNA-DNA hybridizations were performed using the microplate method (Ezaki *et al*., 1989) with some modifications (Cleenwerck *et al*., 2002). The hybridization temperature was 45 °C ± 1 °C and reciprocal reactions were performed with DNA from all strains. Representative strains from MLSA groups A, B, C and D were selected and hybridized to the type strains of *Pantoea agglomerans* (LMG 1286^T), *Pantoea ananatis* (LMG 2665^T), *Pantoea stewartii* (LMG 2715^T) and *Pantoea dispersa* (LMG 2603^T) and among each other. DNA-DNA hybridization was also performed with strains from other validly described species of the genus *Pantoea* as well as with strains of Brenner's hybridization groups II, IV and V. These results are available in Supplementary Table A on IJSEM Online. The level of DNA-DNA binding between the representative strains of MLSA groups A-D and the type strain of *P. agglomerans* was less than 68 %, less than 30 % to the type strains of *P. ananatis* and *P. dispersa* and less than 15 % to the type strain of *P. stewartii*. The strains of MLSA groups A-D exhibited less than 44 % DNA similarity when hybridized to strains from Brenner's hybridization groups II, IV and V. When seven strains from MLSA group A, isolated from *Eucalyptus* and maize, were hybridized among each other (LMG 24199^T, LMG 24195, LMG 24196, LMG 24201, BCC 072, BCC 081 and BCC 427), they exhibited levels of DNA similarity ranging from 83 % to 103 % (data partially presented in supplementary Table A). The hybridization values between strains LMG 24198^T and LMG 24197 (MLSA group B), also isolated from *Eucalyptus*, and LMG

2558^{T} and LMG 2560 (MLSA group C) were even higher at 99 % and 105 %, respectively.

We propose the names *Pantoea vagens* sp. nov. (MLSA group A) for the strains isolated from *Eucalyptus* and maize, *Pantoea eucalypti* sp. nov. (MLSA group B) for the strains isolated from *Eucalyptus* in Uruguay, *Pantoea deleyii* sp. nov. (MLSA group D) for the strain isolated from *Eucalyptus* in Uganda and *Pantoea anthophila* sp. nov. (MLSA group C), for the two strains belonging to Beji protein profile group VII (Brady *et al*., submitted).

The G + C content range of the four novel species, determined by HPLC as published by Mesbah *et al*. (1989), are as follows: *Pantoea vagens* sp. nov. (LMG 24199^{T}, LMG 24195, LMG 24196, LMG 24201, BCC 072, BCC 081 and BCC 427) 55.2-55.8 mol %; *Pantoea eucalypti* sp. nov. (LMG 24198^{T}, LMG 24197) 54.3-54.5 mol %; *Pantoea deleyii* sp. nov. (LMG 24200^{T}) 58.6 mol % and *Pantoea anthophila* sp. nov. (LMG 2558^{T}, LMG 2560) 57.4-57.5 mol %.

Physiological and biochemical tests were performed on selected isolates using API 20E, API 50E and Biotype-100 strips (bioMérieux) as well as Biolog GN plates (Biolog). Results are given in the species descriptions below. The four novel species can be distinguished from their closest phylogenetic neighbours, *P. agglomerans* and *P. ananatis*, using the characteristics listed in Table 2.

Description of *Pantoea vagens* sp. nov.

Pantoea vagens (vá gens. L. present participle of *vagen* meaning to roam, referring to the wide distribution of the species).

Cells are Gram-negative, short rods (0.9 x 1.5-3.0 µm) occurring singly or in pairs, motile and non-sporeforming. Colonies are beige to yellow, round, convex and smooth with entire margins. Facultatively-anaerobic, oxidase negative, acetoin and ß-galactosidase are produced. Indole, H_2S and urease are not produced. The following carbon sources are utilized at 28 °C within three to six days: L-arabinose, D-ribose, D-xylose, D-galactose, D-glucose, D-fructose, D-mannose, maltotriose, L-rhamnose, inositol, esculin, D-mannitol, N-acetylglucosamine, D-maltose, D-saccharose, D-trehalose, D-cellobiose, sucrose, glycerol, L-tartrate, succinate, fumarate, L-aspartate, L-glutamate, cis-aconitate, trans-aconitate, L-proline, D-alanine, L-alanine, L-serine, malonic acid, tween 40, tween 80, D-lyxose (weak), D-fucose (weak) and citric acid (weak). The following carbon sources are not utilized at 28 °C within three to six days: erythritol, D-arabinose, L-xylose, D-adonitol, L-sorbose, dulcitol, D-sorbitol, amygdalin, inulin, D-raffinose, lactose, lactulose, glycogen, xylitol, D-turanose, D-tagatose, L-arabitol, gluconate, 2-ketogluconate, 5-ketogluconate, L-fucose, D-glycopyranose, D-tartrate, L-tryptophan, L-histidine, glutarate, malonate, propionate and L-tyrosine.

The G + C content of the type strain is 55.4 mol %. Strains belonging to this species were isolated from *Eucalyptus* showing symptoms of bacterial blight and die-back in Uganda, Uruguay and Argentina and from maize causing brown stalk rot in South Africa. The type strain is R-21566^T (= LMG 24199^T = BD 765^T) and was isolated from *Eucalyptus* in Uganda.

Description of *Pantoea eucalypti* sp. nov.

Pantoea eucalypti (eu.ca.lýp.ti. L. genitive of *Eucalyptus*, referring to the host from which the strains where isolated).

Cells are Gram-negative, short rods (0.9 x 1.5-3.0 µm) occurring singly or in pairs, motile and non-sporeforming. Colonies are beige to yellow with a darker centre, round, convex and smooth with entire margins. Facultatively-anaerobic, oxidase negative, acetoin and ß-galactosidase are produced. Indole, H_2S and urease are not produced. The following carbon sources are utilized at 28 °C within three to six days: L-arabinose, D-ribose, D-xylose, D-galactose, D-glucose, D-fructose, D-mannose, maltotriose, L-rhamnose, inositol, esculin, D-mannitol, N-acetylglucosamine, D-maltose, D-saccharose, D-trehalose, D-cellobiose, lactose, sucrose, glycerol, L-pyroglutamic acid, L-tartrate, succinate, fumarate, L-aspartate, L-glutamate, cis-aconitate, trans-aconitate, L-proline, D-alanine, L-alanine, L-serine, tween 40, tween 80, D-lyxose (weak), D-fucose and citric acid. The following carbon sources are not utilized at 28 °C within three to six days: erythritol, D-arabinose, L-xylose, D-adonitol, L-sorbose, dulcitol, D-sorbitol, amygdalin, inulin, D-raffinose, lactulose, glycogen, xylitol, D-turanose, D-tagatose, L-arabitol, gluconate, 2-ketogluconate, 5-ketogluconate, L-fucose, D-glycopyranose, D-tartrate, L-tryptophan, L-histidine, glutarate, malonate, propionate, L-tyrosine and malonic acid.

The G + C content of the type strain is 54.5 mol %. Strains belonging to this species were isolated from *Eucalyptus* showing symptoms of bacterial blight and die-back in Uruguay. The type strain is R-25679^T (= LMG 24198^T = BD 766^T) and was isolated from *Eucalyptus* in Uruguay.

Description of *Pantoea deleyii* sp. nov.

Pantoea deleyii (de.leý.ii. L. genitive of deley, named for Jozef De Ley who contributed to the formation of the genus *Pantoea*.)

Cells are Gram-negative, short rods (0.9 x 1.5-3.0 μm) occurring singly or in pairs, motile and non-sporeforming. Colonies are beige to yellow, round, convex and smooth with entire margins. Facultatively-anaerobic, oxidase negative, acetoin and ß-galactosidase are produced. Indole, H_2S and urease are not produced. The following carbon sources are utilized at 28 °C within three to six days: D-arabinose, L-arabinose, D-ribose, D-xylose, D-galactose, D-glucose, D-fructose, D-mannose, maltotriose, L-rhamnose, esculin, D-mannitol, D-turanose, N-acetylglucosamine, D-maltose, D-saccharose, D-trehalose, sucrose, glycerol, L-pyroglutamic acid, succinate, fumarate, L-aspartate, L-glutamate, cis-aconitate, trans-aconitate, L-proline, D-alanine, L-alanine, L-serine, malonic acid, tween 40, tween 80, D-lyxose, D-fucose, D-arabitol and citric acid. The following carbon sources are not utilized at 28 °C within three to six days: erythritol, L-xylose, D-adonitol, D-cellobiose, lactose, lactulose, L-sorbose, dulcitol, inositol, D-sorbitol, amygdalin, inulin, D-raffinose, glycogen, xylitol, D-tagatose, L-arabitol, L-tartrate, gluconate, 2-ketogluconate, 5-ketogluconate, L-fucose, D-glycopyranose, D-tartrate, L-tryptophan, L-histidine, glutarate, malonate, propionate and L-tyrosine.

The G + C content of the type strain is 58.6 mol %. Isolated from *Eucalyptus* showing symptoms of bacterial blight and die-back in Uganda. The type strain is R-31523^T (= LMG 24200^T = BD 767^T).

Description of *Pantoea anthophila* sp. nov.

Pantoea anthophila (an.thó.phi.la. Gr.N. *anthos* meaning flower and Gr.V. *philos* meaning loving as in flower-loving, pertaining to the habitat of the species.)

Cells are Gram-negative, short rods (0.9 x 1.2-2.5 µm) occurring singly or in pairs, motile and non-sporeforming. Colonies are beige to yellow, round, convex and smooth with entire margins. Cells are Gram-negative, short rods (0.9 x 1.5-3.0 µm) occurring singly or in pairs, motile and non-sporeforming. Colonies are beige to yellow, round, convex and smooth with entire margins. Facultatively-anaerobic, oxidase negative, acetoin and ß-galactosidase are produced. Indole, H_2S and urease are not produced. The following carbon sources are utilized at 28 °C within three to six days: L-arabinose, D-ribose, D-xylose, D-galactose, D-glucose, D-fructose, D-mannose, maltotriose, L-rhamnose, inositol, esculin, D-mannitol, N-acetylglucosamine, D-maltose, D-saccharose, D-trehalose, D-cellobiose, sucrose, glycerol, succinate, fumarate, L-aspartate, L-glutamate, cis-aconitate, trans-aconitate, L-proline, D-alanine, L-alanine, L-serine, malonic acid, tween 40, tween 80 and citric acid. The following carbon sources are not utilized at 28 °C within three to six days: erythritol, D-arabinose, L-xylose, D-lyxose, D-adonitol, L-sorbose, dulcitol, D-sorbitol, amygdalin, inulin, D-raffinose, lactose, lactulose, glycogen, xylitol, D-turanose, D-tagatose, L-arabitol, L-tartrate, gluconate, 2-ketogluconate, 5-ketogluconate, L-fucose, D-fucose, D-glycopyranose, D-tartrate, L-tryptophan, L-histidine, glutarate, malonate, propionate and L-tyrosine.

The G + C content of the type strain is 57.5 mol %. Strains belonging to this species have been isolated from flowering shrubs. The type strain is LMG 2558^T (= NCPPB 1682^T) and was isolated from *Impatiens balsamina* in India.

Acknowledgements

This study was partially supported by the South African-Flemish Bilateral Agreement, the National Research Foundation (NRF), the Tree Protection Co-operative Programme (TPCP) and the THRIP support programme of the Department of Trade and Industry, South Africa. The BCCM/LMG Bacteria collection is supported by the Federal Public Planning Service-Science Policy, Belgium. The authors wish to acknowledge Katrien Vandemeulebroecke for technical assistance, Mike Wingfield, Teresa Goszczynska, Jolanda Roux, Grace Nakabonge and Izette Greyling for allowing us to include several strains isolated from *Eucalyptus*, onion and maize in the study and Mrs. Mohr for providing us with the CDC strains.

References

Beji, A., Mergaert, J., Gavini, F., Izard, D., Kersters, K., Leclerc, H. & De Ley, J. (1988). Subjective synonymy of *Erwinia herbicola*, *Erwinia milletiae*, and *Enterobacter agglomerans* and redefinition of the taxon by genotypic and phenotypic data. *Int J Syst Bacteriol* **38,** 77-88

Brady, C.L., Venter, S.N., Cleenwerck, I., Vancanneyt, M., Swings, J. & Coutinho, T.A. (2007). A FAFLP system for the improved identification of plant-pathogenic and plant-associated species of the genus *Pantoea*. *Syst Appl Microbiol* **30,** 413-417

Brady, C.L., Cleenwerck, I., Venter, S.N., Vancanneyt, M., Swings, J. & Coutinho, T.A. (2008). Phylogeny and identification of *Pantoea* species associated with the environment, humans and plants based on mulitlocus sequence analysis (MLSA). Submitted to *Appl Environ Microbiol*

Brenner, D.J., Fanning, G.R., Leete Knutson, J.K., Steigerwalt, A.G. & Krichevsky, M.I. (1984). Attempts to classify Herbicola Group-*Enterobacter agglomerans* strains by deoxyribonucleic acid hybridization and phenotypic tests. *Int J Syst Bacteriol* **34,** 45-55

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

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

Coutinho, T.A., Preisig, O., Mergaert, J., Cnockaert, M.C., Riedel, K.H., Swings, J. & Wingfield, M.J. (2002). Bacterial blight and dieback of *Eucalyptus* species, hybrids and clones in South Africa. *Plant Dis* **86,** 20-25

Ezaki, T., Hashimoto, Y. & Yabuuchi, E. (1989). Fluorometric deoxyribonucleic acid-deoxyribonucleic acid hybridization in micro-dilution wells as an alternative to membrane filter hybridization in which radioisotopes are used to determine genetic relatedness among bacterial strains. *Int J Syst Bacteriol* **39,** 224-229

Gavini, F., Mergaert, J., Beji, A., Mielcarek, C., Izard, D., Kersters, K. & De Ley, J. (1989). Transfer of *Enterobacter agglomerans* (Beijerinck 1988) Ewing and Fife 1972 to *Pantoea* gen. nov. as *Pantoea agglomerans* comb. nov. and description of *Pantoea dispersa* sp. nov. *Int J Syst Bacteriol* **39,** 337-345

Goszczynska, T., Botha, W.J., Venter, S.N. & Coutinho, T.A. (2007). Isolation and identification of the causal agent of brown stalk rot, a new disease of maize in South Africa. *Plant Dis* **91,** 711-718

Grimont, P.A.D. & Grimont, F. (2005). Genus: *Pantoea*. In *Bergey's Manual of Systematic Bacteriology*, Volume Two, The *Proteobacteria*, Part B, The *Gammaproteobacteria*, pp. 713-720, Edited by D.J. Brenner, N.R. Krieg & J.T. Staley. 2nd Edition. New York: Springer

Guindon, S. & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52,** 696-704

Kageyama, B., Nakae, M., Yagi, S. & Sonoyama, T. (1992). *Pantoea punctata* sp. nov., *Pantoea citrea* sp. nov., and *Pantoea terrea* sp. nov. isolated from fruit and soil samples. *Int J Syst Bacteriol* **42,** 203-210

Mergaert, J., Verdonck, L. & Kersters, K. (1993). Transfer of *Erwinia ananas* (synonym, *Erwinia uredovora*) and *Erwinia stewartii* to the Genus *Pantoea* emend. as *Pantoea ananas* (Serrano 1928) comb. nov. and *Pantoea stewartii* (Smith 1898) comb. nov., Respectively, and Description of *Pantoea stewartii* subsp. *indologenes* subsp.nov. *Int J Syst Bacteriol* **43,** 162-173

Mesbah, M., Premachandran, U. & Whitman, W.B. (1989). Precise measurement of the G + C content of deoxyribonucleic acid by high-performance liquid chromatography. *Int J Syst Bacteriol* **39,** 159-167

Posada, D. & Crandall, K.A. (1998). Modeltest: testing the model of DNA substitution. *Bioinformatics* **14,** 817-818

Swofford, D.L. (2000). PAUP*: Phylogenetic Analysis Using Parsimony and other methods (software). Sinauer Associates, Sunderland, MA.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. & Higgins, D.G. (1997). The ClustalX-Windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucl Acids Res* **25,** 4876-4882

Wilson, K. (1987). Preparation of genomic DNA from bacteria. In *Current Protocols in Molecular Biology*, pp. 2.4.1-2.4.5. Edited by F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith & K. Struhl. New York: Green Publishing and Wiley-Interscience

Table 1: Strains used in this study

LMG = BCCM/LMG Bacteria Collection, Ghent University, Belgium, BCC = Bacterial Culture Collection, Forestry and Agricultural Biotechnology Institute, Pretoria, South Africa, BD = Plant Pathogenic and Plant Protecting Bacteria (PPPPB) Culture Collection, ARC-PPRI, Pretoria, South Africa, CDC = Centres for Disease Control, Atlanta, Georgia, U.S.A, NCPPB = National Collection of Plant Pathogenic Bacteria, York, United Kingdom

Species	**Strain**	**Host**	**Location**
Pantoea agglomerans	LMG 1286^T	Human	Zimbabwe
	LMG 2565	Cereal	Canada
	LMG 2596	Onion	South Africa
	LMG 2660	*Wisteria floribunda*	Japan
Pantoea ananatis	LMG 2665^T	Pineapple	Brazil
	LMG 20103	*Eucalyptus*	South Africa
	LMG 24190	Onion	South Africa
Pantoea stewartii ssp. *stewartii*	LMG 2715^T	Corn	USA
	LMG 2718	Corn	USA
Pantoea stewartii ssp. *indologenes*	LMG 2632^T	Fox millet	India
	LMG 2673	Pineapple	Hawaii
Pantoea dispersa	LMG 2603^T	Soil	Japan
	LMG 2602	Sorghum	India
	LMG 2604	Wild rose	Netherlands
Pantoea citrea	LMG 22049^T	Mandarin orange	Japan
Pantoea punctata	LMG 22050^T	Mandarin orange	Japan
	LMG 23562	Mandarin orange	Japan

Pantoea terrea	LMG 22051^{T}	Soil	Japan
	LMG 23564	Soil	Japan
Pantoea anthophila	LMG 2558^{T} = NCPPB 1682	*Impatiens balsamina*	India
	LMG 2560 = NCPPB 1941	*Tagetes erecta*	Unknown
Pantoea vagens	R-21566^{T} = BCC 105^{T} = LMG 24199^{T}	*Eucalyptus*	Uganda
	R-25484 = BCC 013 = LMG 24195	*Eucalyptus*	Uruguay
	R-25674 = BCC 015 = LMG 24196	*Eucalyptus*	Argentina
	R-30991 = BD 639 = LMG 24201	Maize	South Africa
	R-21559 = BCC 081	*Eucalyptus*	Uruguay
	R-25676 = BCC 072	*Eucalyptus*	Uruguay
	R-30997 = BCC 208	*Eucalyptus*	Uganda
Pantoea eucalypti	R-25679^{T} = BCC 077^{T} = LMG 24198^{T}	*Eucalyptus*	Uruguay
	R-25678 = BCC 076 = LMG 24197	*Eucalyptus*	Uruguay
Pantoea deleyii	R-31523^{T} = BCC 109^{T} = LMG 24200^{T}	*Eucalyptus*	Uganda
Pantoea sp. (Brenner HG II)	LMG 5345 = CDC 3123-70	Human	USA
	R-35488 = CDC 238-70	Human	USA
Pantoea sp. (Brenner HG IV)	LMG 2781 = CDC 1741-71	Human	USA
	R-35491 = CDC 3638-70	Human	USA
Pantoea sp. (Brenner HG V)	LMG 5343 = CDC 3482-71	Human	USA
	R-35494 = CDC 2928-68	Human	USA
	R-35496 = CDC 3527-71	Human	USA

Table 2: Characteristics distinguishing *P. vagens* sp. nov., *P. eucalypti* sp. nov., *P. deleyii* sp. nov. and *P. anthophila* sp. nov. from each other and from their closest phylogenetic neighbours

1 = *P. agglomerans* (3), 2 = *P. ananatis* (4), 3 = *P. vagens* sp. nov. (21), 4 = *P. eucalypti* sp. nov. (2), 5 = *P. deleyii* sp. nov. (1), 6 = *P. anthophila* sp. nov. (2)

+, 90-100 % of strains positive in 1-2 days; (+), 90-100 % of strains positive in 1-4 days; -, 90-100 % of strains negative in 4 days; d, positive in 1- 4 days; (d), positive in 3-4 days

Characteristic	**1**	**2**	**3**	**4**	**5**	**6**
Tween 40	-	+	+	+	+	+
Tween 80	-	+	+	+	+	+
Malonic acid	-	-	d	-	+	+
Lactose	-	+	-	+	-	-
L-ornithine	-	-	+	d	+	+
D-arabitol	+	+	-	-	+	+
L-pyroglutamic acid	-	-	-	+	+	-

Supplementary Table A: DNA-DNA hybridization values amongst strains belonging to the novel species *Pantoea vagens, Pantoea eucalypti, Pantoea deleyii* and *Pantoea anthophila* and reference strains of the seven validly described species of the genus *Pantoea*.

Figure 1: Maximum likelihood tree based on complete 16S rRNA sequences of *Pantoea* species. The tree was generated by the Phyml software using the Tamura-Nei (TN93) model as selected by Modeltest. Bootstrap values after 1000 replicates are expressed as percentages. *Pectobacterium carotovorum* was included as an outgroup.

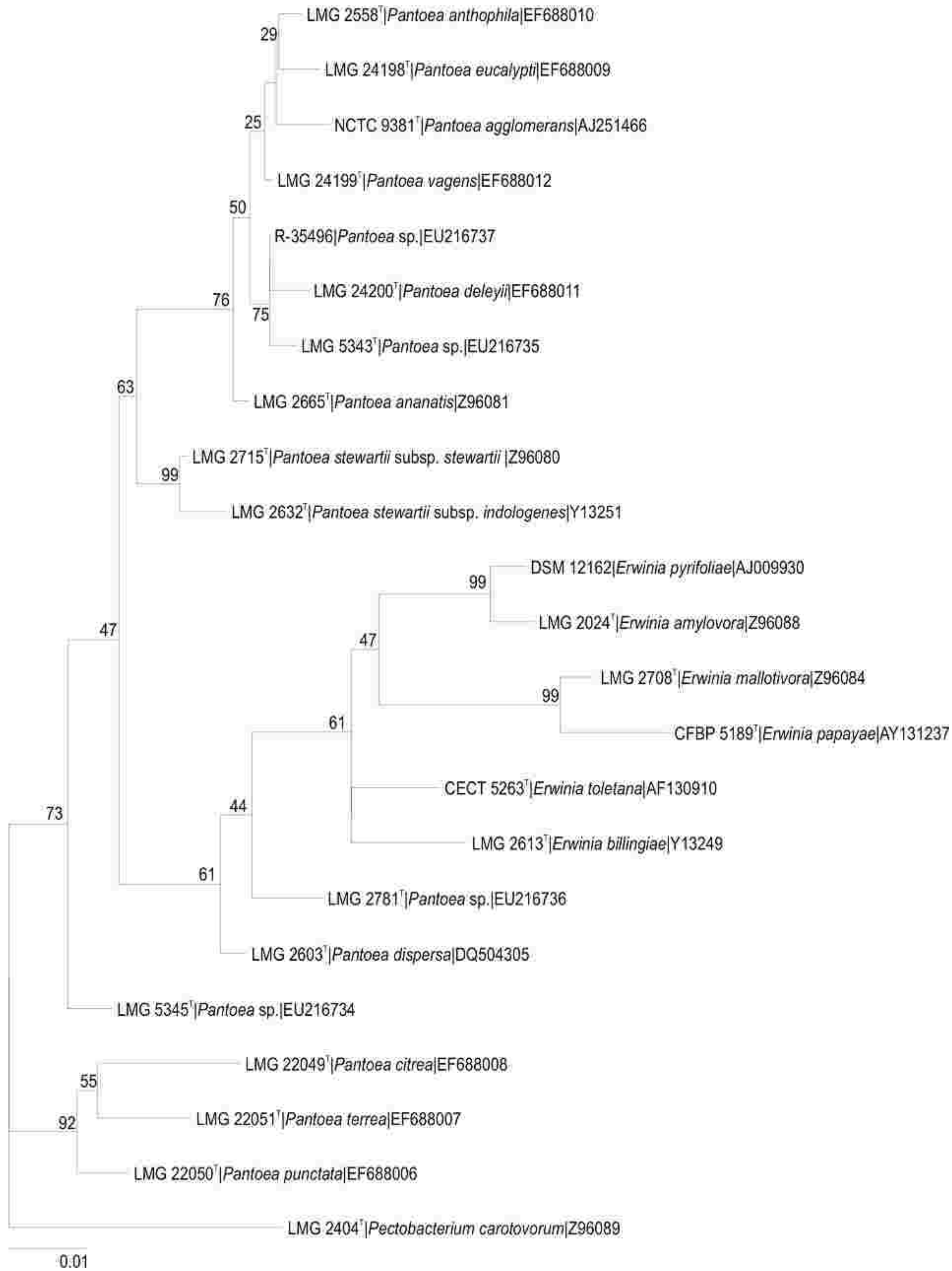
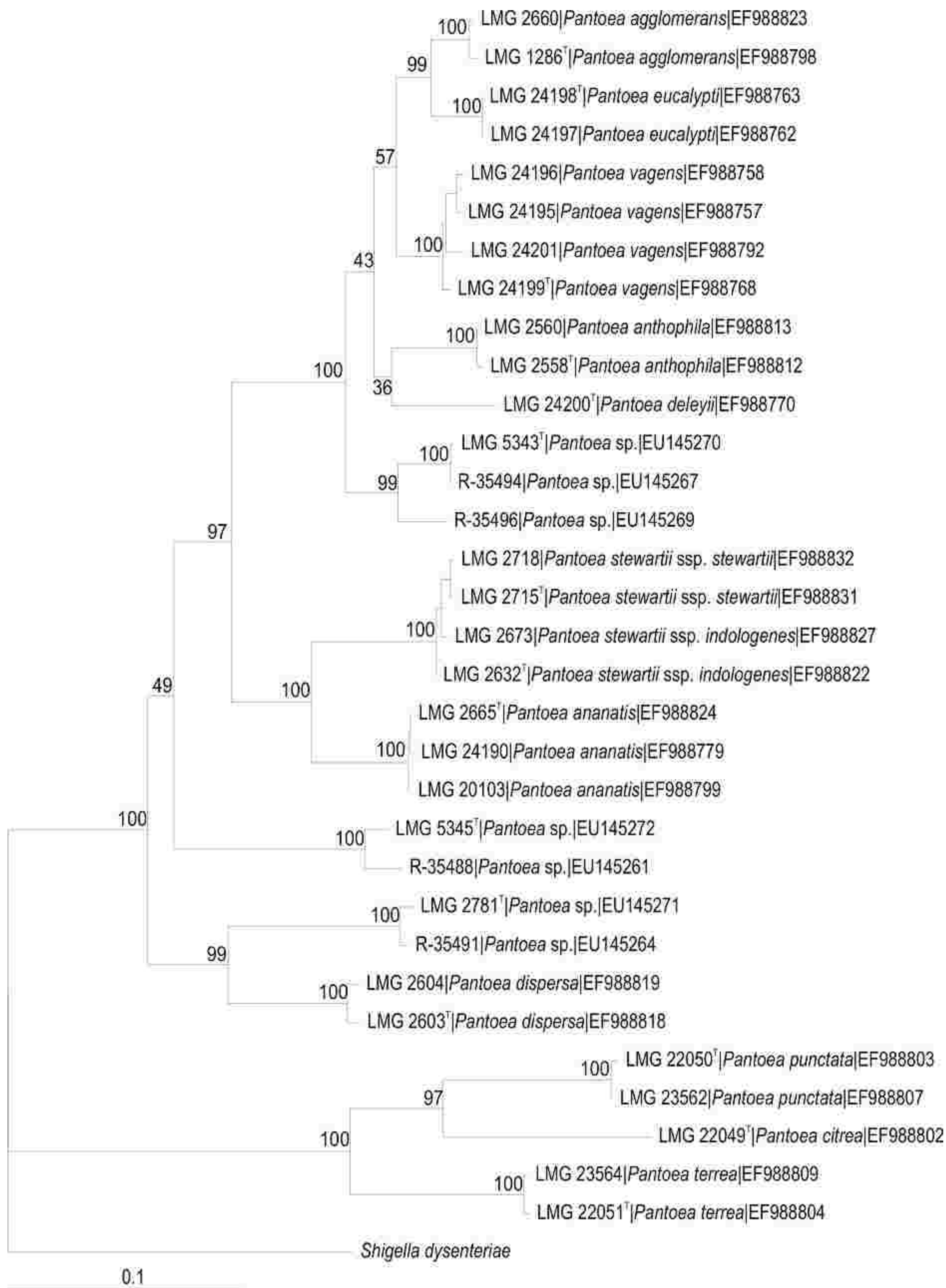
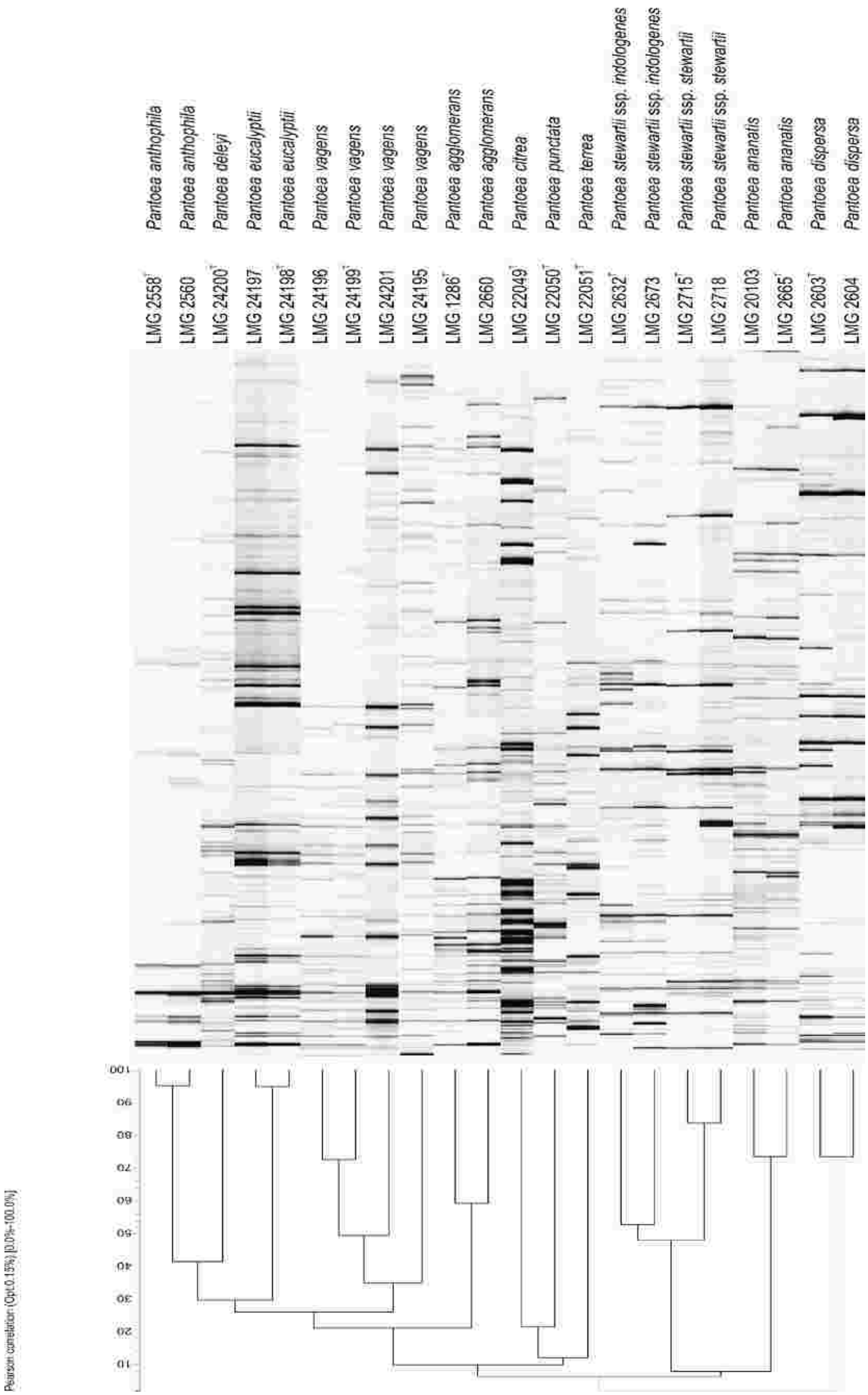

LMG 2558T|Pantoea anthophila|EF688010
LMG 24198T|Pantoea eucalypti|EF688009
NCTC 9381T|Pantoea agglomerans|AJ251466
LMG 24199T|Pantoea vagens|EF688012
R-35496|Pantoea sp.|EU216737
LMG 24200T|Pantoea deleyii|EF688011
LMG 5343T|Pantoea sp.|EU216735
LMG 2665T|Pantoea ananatis|Z96081
LMG 2715T|Pantoea stewartii subsp. stewartii |Z96080
LMG 2632T|Pantoea stewartii subsp. indologenes|Y13251
DSM 12162|Erwinia pyrifoliae|AJ009930
LMG 2024T|Erwinia amylovora|Z96088
LMG 2708T|Erwinia mallotivora|Z96084
CFBP 5189T|Erwinia papayae|AY131237
CECT 5263T|Erwinia toletana|AF130910
LMG 2613T|Erwinia billingiae|Y13249
LMG 2781T|Pantoea sp.|EU216736
LMG 2603T|Pantoea dispersa|DQ504305
LMG 5345T|Pantoea sp.|EU216734
LMG 22049T|Pantoea citrea|EF688008
LMG 22051T|Pantoea terrea|EF688007
LMG 22050T|Pantoea punctata|EF688006
LMG 2404T|Pectobacterium carotovorum|Z96089
29
25
50
76
75
63
99
47
99
47
99
61
44
61
73
55
92
0.01

Figure 2: Maximum likelihood tree based on concatenated partial sequences of *rpoB*, *atpD*, *gyrB* and *infB* of *Pantoea* strains. The tree was generated by the Phyml software using the general time reversible (GTR) model as selected by Modeltest. Bootstrap values after 1000 replicates are expressed as percentages. *Shigella dysenteriae* sequences were obtained from the genome sequencing database of the Sanger Institute (http://www.sanger.ac.uk) and included as an outgroup.


LMG 2660|Pantoea agglomerans|EF988823
LMG 1286^{T}|Pantoea agglomerans|EF988798
LMG 24198^{T}|Pantoea eucalypti|EF988763
LMG 24197|Pantoea eucalypti|EF988762
LMG 24196|Pantoea vagens|EF988758
LMG 24195|Pantoea vagens|EF988757
LMG 24201|Pantoea vagens|EF988792
LMG 24199^{T}|Pantoea vagens|EF988768
LMG 2560|Pantoea anthophila|EF988813
LMG 2558^{T}|Pantoea anthophila|EF988812
LMG 24200^{T}|Pantoea deleyii|EF988770
LMG 5343^{T}|Pantoea sp.|EU145270
R-35494|Pantoea sp.|EU145267
R-35496|Pantoea sp.|EU145269
LMG 2718|Pantoea stewartii ssp. stewartii|EF988832
LMG 2715^{T}|Pantoea stewartii ssp. stewartii|EF988831
LMG 2673|Pantoea stewartii ssp. indologenes|EF988827
LMG 2632^{T}|Pantoea stewartii ssp. indologenes|EF988822
LMG 2665^{T}|Pantoea ananatis|EF988824
LMG 24190|Pantoea ananatis|EF988779
LMG 20103|Pantoea ananatis|EF988799
LMG 5345^{T}|Pantoea sp.|EU145272
R-35488|Pantoea sp.|EU145261
LMG 2781^{T}|Pantoea sp.|EU145271
R-35491|Pantoea sp.|EU145264
LMG 2604|Pantoea dispersa|EF988819
LMG 2603^{T}|Pantoea dispersa|EF988818
LMG 22050^{T}|Pantoea punctata|EF988803
LMG 23562|Pantoea punctata|EF988807
LMG 22049^{T}|Pantoea citrea|EF988802
LMG 23564|Pantoea terrea|EF988809
LMG 22051^{T}|Pantoea terrea|EF988804
Shigella dysenteriae
0.1

Supplementary Figure A: UPGMA dendrogram based on FAFLP analysis of *Pantoea* species using the selective primer combination Eco-C/Mse-GC. The levels of similarity representing the Pearson similarity coefficient, are expressed as percentages. The banding patterns adjacent to each branch are normalised and background-subtracted digitised gel strips processed using Bionumerics.


Pearson correlation (Opt:0.15%) [0.0%-100.0%]
100
90
80
70
60
50
40
30
20
10
LMG 2558T Pantoea anthophila
LMG 2560 Pantoea anthophila
LMG 24200T Pantoea deleyi
LMG 24197 Pantoea eucalyptii
LMG 24198T Pantoea eucalyptii
LMG 24196 Pantoea vagens
LMG 24199T Pantoea vagens
LMG 24201 Pantoea vagens
LMG 24195 Pantoea vagens
LMG 1286T Pantoea agglomerans
LMG 2660 Pantoea agglomerans
LMG 22049T Pantoea citrea
LMG 22050T Pantoea punctata
LMG 22051T Pantoea terrea
LMG 2632T Pantoea stewartii ssp. indologenes
LMG 2673 Pantoea stewartii ssp. indologenes
LMG 2715T Pantoea stewartii ssp. stewartii
LMG 2718 Pantoea stewartii ssp. stewartii
LMG 20103 Pantoea ananatis
LMG 2665T Pantoea ananatis
LMG 2603T Pantoea dispersa
LMG 2604 Pantoea dispersa

CHAPTER 5

Description of four novel *Pantoea* species from human clinical samples, *Pantoea septica* sp. nov., *Pantoea eucrina* sp. nov., *Pantoea brenneri* sp. nov. and *Pantoea conspicua* sp. nov.

As submitted to: International Journal of Systematic and Evolutionary Microbiology

Summary

Bacterial strains belonging to DNA hybridization groups II, IV and V from Brenner *et al*. (1984) were previously suggested as belonging to the genus *Pantoea* but have not been officially described and classified. In this study; the phylogenetic position of these groups was re-examined using molecular techniques. Both 16S rRNA sequencing and multilocus sequence analyses (MLSA) based on *rpoB*, *atpD, gyrB* and *infB* genes revealed that DNA hybridization groups II, IV and V are phylogenetically closely related to *Pantoea agglomerans*, *Pantoea ananatis*, *Pantoea dispersa* and *Pantoea deleyii*. MLSA data together with DNA-DNA hybridization data further proved that Brenner DNA hybridization groups II, IV and V constitute four new species of the genus. Two MLSA groups were found within hybridization group V. Several phenotypic characteristics distinguished these novel species from each other and from their closest phylogenetic neighbours. The names *Pantoea septica* sp. nov. (LMG 5345^T= ATCC 29923^T, DNA group II), *Pantoea eucrina* sp. nov. (LMG 2781^T= ATCC 27998^T, DNA group IV), *Pantoea brenneri* sp. nov. (LMG 5343^T= ATCC 29921^T, DNA group V) and *Pantoea conspicua* sp. nov. (LMG 24534^T= BD 805^T, DNA group V) are proposed.

Since the early 1970s, many attempts have been made to resolve the *Erwinia herbicola-Enterobacter agglomerans* complex (Ewing & Fife, 1972, Gardner & Kado, 1972, Young *et al.*, 1978, Dye, 1981, Brenner *et al.*, 1984, Verdonck *et al.*, 1987 and Beji *et al.*, 1988). The most comprehensive study to date was performed by Brenner *et al.* in 1984, where 124 strains belonging to this complex were separated into 13 groups based on their DNA-DNA hybridization values (Brenner *et al.*, 1984). The majority of strains used in Brenner's study were human clinical strains from a nosocomial septicemia outbreak in 1971 (Maki *et al.*, 1976), although plant-pathogenic strains from the complex were also included.

In the years following Brenner's hybridization study, many of the 13 DNA hybridization groups were further investigated and proposed as new genera or species, or transferred to existing genera. DNA group XIII became *Pantoea agglomerans* (Gavini *et al.*, 1989), DNA group III was renamed as *Pantoea dispersa* (Gavini *et al.*, 1989) and DNA group VI, containing strains of *Erwinia ananas* and its synonym *Erwinia uredovora*, was transferred to the genus *Pantoea* as *Pantoea ananatis* (Mergaert *et al.*, 1993). DNA groups VII – XII were assigned to species within the genera *Enterobacter*, *Rahnella* and *Leclercia*. Beji *et al.* (1988) advised the inclusion of DNA group V in *Enterobacter agglomerans* (now *Pantoea agglomerans*), despite only 62 % DNA homology of the reference strain to the type strain ATCC 27155^{T}. Gavini *et al.* (1989) later rejected this recommendation and proposed that DNA group V and protein profile group VII, from the study of Beji *et al.* (1988) be united as a separate species in a single new genus. This proposal has never been implemented and DNA groups I, II, IV and V have not yet been classified, although the recent edition of Bergey's Manual of Systematic Bacteriology (Grimont & Grimont, 2005) suggests that these four DNA hybridization groups belong to the genus *Pantoea*.

A multilocus sequence analysis (MLSA) study recently identified Beji protein profile group VII as a novel species of *Pantoea* (*P. anthophila*), along with three novel species isolated from *Eucalyptus* and maize (*P. vagens*, *P. eucalypti* and *P. deleyii*) (Brady *et al.*, submitted; Brady *et al.*, submitted). These four novel *Pantoea* species were described using MLSA based on *rpoB*, *atpD*, *gyrB* and *infB* sequences as a supporting phylogenetic technique, since these genes were found to be useful phylogenetic markers for *Pantoea*. By applying the same MLSA scheme and DNA-DNA hybridizations to strains from Brenner's hybridization groups II, IV and V, it became clear that these strains are novel *Pantoea* species.

Strains belonging to Brenner DNA groups II, IV and V were kindly provided by Mrs. Mohr from the Centers for Disease Control (CDC), Atlanta, Georgia, U.S.A. It was not possible to acquire strains from Brenner DNA group I. The strains used in this study are listed in Table 1. An alkalic extraction method was used to isolate genomic DNA from the strains (Niemann *et al.*, 1997) which was stored at -20 °C. The complete 16S rRNA gene was amplified and sequenced for the type strain from each novel species (LMG 5345^{T} = DNA group II, LMG 2781^{T} = DNA group IV, LMG 5343^{T} = DNA group V, LMG 24534^{T} = DNA group V) using the primers and conditions determined by Coenye *et al.* (1999). MLSA based on *rpoB*, *atpD*, *gyrB* and *infB* genes was performed on all strains (Brady *et al.*, submitted). The Genbank accession numbers for *Pantoea septica* sp. nov. (Brenner HG II) LMG 5345^{T}, *Pantoea eucrina* sp. nov. (Brenner HG IV) LMG 2781^{T}, *Pantoea brenneri* sp. nov. (Brenner HG V) LMG 5343^{T} and *Pantoea conspicua* sp. nov. (Brenner HG V) LMG 24534^{T} are EU216734-EU216737, respectively.

The sequences were aligned using ClustalX (Thompson *et al.*, 1997) and the overhangs were trimmed. The Modeltest 3.7 programme (Posada & Crandall, 1998) was then applied to determine the best-fit evolutionary model to apply to each gene. Maximum likelihood and neighbour joining analyses were performed using Phyml (Guindon & Gascuel, 2003) and PAUP 4.0b10 (Swofford, 2000) respectively, by applying the models and parameters determined by Modeltest (only Maximum likelihood phylogenetic trees are shown). Bootstrap analysis with 1000 replicates was performed to assess the support for these clusters.

The type strain of DNA group II (LMG 5345^{T}) shared more than 98 % 16S rRNA sequence similarity with *P. agglomerans*, *P. vagens*, DNA group V (LMG 5343^{T}) and *P.*

eucalypti, whilst the 16S rRNA sequence of the type strain of DNA group IV (LMG 2781^{T}) was more than 98 % similar to those of *P. dispersa* and *P. agglomerans*. The type strain of DNA group V (LMG 5343^{T}) showed high 16S rRNA sequence similarity (> 98 %) to *P. deleyii*, *P. agglomerans*, *P. vagens*, *P. anthophila*, *P. eucalypti*, *P. ananatis* and *P. stewartii* ssp. *stewartii*, as was LMG 24534^{T}, also from DNA group V. These sequence similarities are reflected in the 16S rRNA phylogenetic tree (Fig. 1). The two strains from DNA group V, LMG 5343^{T} and LMG 24534^{T}, cluster closely with the type strain of *P. deleyii*, within the "core" *Pantoea* group with a high bootstrap value. LMG 24534^{T} is situated on a branch separate from LMG 5343^{T}, the acknowledged type strain of DNA group V suggesting that this DNA group contains two species. The type strain of DNA group IV (LMG 2781^{T}) clusters with the type strain of *P. dispersa* with a strong bootstrap value of 81 %. DNA group II's type strain, LMG 5345^{T}, is situated on the border of the "core" *Pantoea* group. The bootstrap values are considerably lower within the *Pantoea* species clusters but this can be explained by the high level of 16S rRNA sequence homogeneity between genera and species belonging to the family *Enterobacteriaceae*.

A similar pattern is seen in the phylogenetic tree based on concatenated sequences of the four genes (Fig. 2), with strains from DNA group V clustering close to *P. agglomerans*, *P. eucalypti*, *P. vagens* and *P. anthophila* with high bootstrap support of 100 %. Strains from DNA group II cluster on the border of the core *Pantoea* species. DNA group IV clusters with *P. dispersa* with a strong bootstrap value in the concatenated and 16S rRNA phylogenetic trees. It was observed in both the 16S rRNA- and MLSA based phylogenetic trees (Figs. 1 & 2), that a single strain from DNA group V (LMG 24534^{T}) always clusters slightly distant to the other strains within this group. Brenner *et al.* (1984) noted that aerogenic and anaerogenic strains were rarely found in the same hybridization groups. Two exceptions were LMG 24527 (CDC 1778-80) and LMG 24534^{T} (CDC 3527-71) which are aerogenic strains belonging to DNA groups II and V respectively, whilst the remainder of the strains in these two DNA groups are anaerogenic. However, strain LMG 24527 clusters closely with the other strains from DNA group II compared to the conspicuous separation of strain LMG 24534^{T} from the rest of DNA group V (Fig. 2).

High quality DNA for DNA-DNA hybridization of strains was prepared by the method of Wilson (1987), with minor modifications (Cleenwerck *et al*., 2002). DNA-DNA hybridizations were performed using the microplate method (Ezaki *et al*., 1989) with some modifications (Cleenwerck *et al*., 2002). The hybridization temperature was 45 °C ± 1 °C and reciprocal reactions were performed with all strains. The type strains from each DNA group were hybridized to those of the phylogenetically related, validly described *Pantoea* species, including the four recently described species (Brady *et al*., submitted) and amongst each other. A complete table of the DNA-DNA hybridization results is available on IJSEM Online as Supplementary Table A.

The type strains of DNA groups II and IV (LMG 5343^T and LMG 2781^T) exhibited 29 and 24 % DNA similarity when hybridized to the type strain of *P. agglomerans* (LMG 1286^T), whilst the type strain of DNA group V (LMG 5343^T) was 48 % similar to LMG 1286^T. These values are significantly lower than those determined by Beji *et al*. when reference strains of Brenner's DNA groups II, IV and V were hybridized to the type strain of "*Enterobacter agglomerans*" (ATCC 27155^T = LMG 1286^T) (Beji *et al*., 1988). The level of DNA-DNA binding between each of the representative strains from DNA groups II, IV and V and the type strain of *P. ananatis* was less than 30 %, less than 15 % to the type strain of *P. stewartii* and less than 31 % to the type strain of *P. dispersa*. The DNA-DNA hybridization values between DNA groups II, IV and V and the recently described *P. vagens*, *P. eucalypti*, *P. deleyii* and *P. anthophila* were in the same range (20 – 43 %).

The DNA hybridization values between the type strains of DNA groups II, IV and V ranged from 31 – 51 %. The latter value (51 %) was found between the aerogenic strain from DNA group V (LMG 24534^T) and the type strain of DNA group V (LMG 5343^T). This value is considerably lower than the 73 % observed by Brenner *et al*. (1984). Based on the clear separation of LMG 24534^T from the rest of DNA group V in both the 16S rRNA- and concatenated-phylogenetic trees, as well as the low DNA-DNA hybridization value it is clear that this strain constitutes a separate species. We propose the names *Pantoea septica* sp. nov. for DNA group II, *Pantoea eucrina* sp. nov. for DNA group IV, *Pantoea brenneri* sp. nov. for DNA group V and *Pantoea conspicua* sp. nov. for LMG 24534^{T} from DNA group V.

The G + C content of the type strains of all four novel species, determined by HPLC as published by Mesbah *et al.* (1989), are as follows: *Pantoea septica* sp. nov., Brenner DNA group II (LMG 5345^{T} = ATCC 29923^{T}) = 59.3 mol %; *Pantoea eucrina* sp. nov., Brenner DNA group IV (LMG 2781^{T} = ATCC 27998^{T}) = 56.5 mol %; *Pantoea brenneri* sp. nov., Brenner DNA group V (LMG 5343^{T} = ATCC 29921^{T}) = 55.4 mol %; *Pantoea conspicua* sp. nov., Brenner DNA group V (LMG 24534^{T} = BD 805^{T}) = 55.7 mol %.

Physiological and biochemical characteristics for DNA groups II, IV and V were obtained from Bergey's Manual of Systematic Bacteriology (Grimont & Grimont, 2005). Additional API 50E and Biotype-100 tests (bioMérieux) were performed on the type strain, and the single aerogenic strain (LMG 24534^{T}), from DNA group V. The four novel species can be distinguished from their closest phylogenetic neighbours, *P. agglomerans*, *P. ananatis* and *P. dispersa*, using the characteristics listed in Table 2. It was observed by Brenner *et al.* (1984) that "DNA groups II through V could not be separated with certainty on the basis of biochemical tests although phenylalanine deaminase, malonate, rhamnose, cellobiose, acetate, and dextrin reactions are helpful in distinguishing them".

Description of *Pantoea septica* sp. nov. (Brenner DNA group II)

Pantoea septica (sép.ti.ca. Gr. adj. *septikos* meaning putrefaction or decay, referring to the septicemia outbreak caused by these strains)

Cells are Gram-negative, short rods (0.9 x 1.5-3.0 μm) occurring singly or in pairs, motile and non-sporeforming. Colonies are beige, round, convex and smooth with entire margins. Facultatively-anaerobic, oxidase negative, glucose dehydrogenase and gluconate dehydrogenase are produced. Indole, H_2S and urease are not produced. The following carbon sources are utilized at 28 °C within three to six days: trans-aconitate, L-arabinose, D-arabitol, D-cellobiose, citrate, dulcitol (weak), L-fucose, D-fructose, D-galactose, D-galacturonate, gentiobiose, D-glucose, inositol, 5-ketogluconate, lactose, lactulose, D-malate, D-maltose, maltotriose, D-mannitol, D-mannose, D-melibiose, D-raffinose, L-rhamnose, D-ribose, D-sorbitol (weak), sucrose, L-tartrate, meso-tartrate, D-trehalose, trigonelline, xylitol and D-xylose. The following carbon sources are not utilized at 28 °C within three to six days: D-adonitol, L-arabitol, betaine, erythritol, glutarate, histamine, 3-0-methyl-D-glucose, propionate, quinate, L-sorbose, D-tagatose, D-tartrate, D-turanose and L-tyrosine.

The G + C content of the type strain is 59.3 mol %. Strains belonging to this species were implicated in a nationwide septicemia outbreak in the USA in 1971. The type strain is LMG 5345^T (=ATCC 29923^T = CDC 3123-70) and was isolated from a human stool sample in New Jersey, USA.

Description of *Pantoea eucrina* sp. nov. (Brenner DNA group IV)

Pantoea eucrina (eú.cri.na. Gr. adj. *eukrines* meaning well-separated, referring to the clear separation of the strains from other species within the genus.)

Cells are Gram-negative, short rods (0.9 x 1.5-3.0 μm) occurring singly or in pairs, motile and non-sporeforming. Colonies are beige, round, convex and smooth with entire margins. Facultatively-anaerobic, oxidase negative, glucose dehydrogenase and gluconate dehydrogenase are produced. Indole, H_2S and urease are not produced. The following carbon sources are utilized at 28 °C within three to six days: trans-aconitate, adonitol, L-arabinose, D-arabitol, L-arabitol, D-cellobiose, citrate, erythritol, D-fructose, D-galactose, D-galacturonate, gentiobiose, D-glucose, inositol, 5-ketogluconate, D-maltose, maltotriose, D-mannitol, D-mannose, L-rhamnose, D-ribose, sucrose, D-trehalose, trigonelline and xylitol. The following carbon sources are not utilized at 28 °C within three to six days: betaine, dulcitol, L-fucose, glutarate, histamine, lactose, lactulose, D-malate, D-melibiose, 3-0-methyl-D-glucose, propionate, quinate, D-raffinose, D-sorbitol, L-sorbose, D-tagatose, D-tartrate, L-tartrate, meso-tartrate, D-turanose, L-tyrosine and D-xylose.

The G + C content of the type strain is 56.5 mol %. Strains belonging to this species were implicated in a nationwide septicemia outbreak in the USA in 1971. The type strain is LMG 2781^T (=ATCC 27998^T = CDC 1741-71) and was isolated from a human trachea in Connecticut, USA.

Description of *Pantoea brenneri* sp. nov. (Brenner DNA group V)

Pantoea brenneri (bren.néri. L. genitive of brenner, named for Don J. Brenner, in recognition of his contribution in resolving the *Erwinia herbicola-Enterobacter agglomerans* complex.)

Cells are Gram-negative, short rods (0.9 x 1.5-3.0 µm) occurring singly or in pairs, motile and non-sporeforming. Colonies are beige, round, convex and smooth with entire margins. Facultatively-anaerobic, oxidase negative, glucose dehydrogenase and gluconate dehydrogenase are produced. Indole, H_2S and urease are not produced. The following carbon sources are utilized at 28 °C within three to six days: trans-aconitate, L-arabinose, D-arabitol, D-cellobiose, citrate, L-fucose, D-fructose, D-galactose, D-galacturonate, D-glucose, inositol, 5-ketogluconate, lactose, lactulose, D-malate, D-maltose, maltotriose, D-mannitol, D-mannose, D-raffinose, L-rhamnose, D-ribose, sucrose, L-tartrate, meso-tartrate, D-trehalose, and D-xylose. The following carbon sources are not utilized at 28 °C within three to six days: D-adonitol, L-arabitol, betaine, dulcitol, erythritol, gentiobiose, glutarate, histamine, D-melibiose, 3-0-methyl-D-glucose, propionate, quinate, D-sorbitol, L-sorbose, D-tagatose, D-tartrate, trigonelline, D-turanose, L-tyrosine and xylitol.

The G + C content of the type strain is 55.4 mol %. Strains belonging to this species were implicated in a nationwide septicemia outbreak in the USA in 1971. The type strain is LMG 5343^T (=ATCC 29921^T = CDC 3482-71) and was isolated from a human urethra in Montana, USA.

Description of *Pantoea conspicua* sp. nov. (Brenner DNA group V)

Pantoea conspicua (con.spí.cúa. L. adj. *conspicuus* meaning conspicuous, referring to the conspicuous separation from other strains within DNA group V)

Cells are Gram-negative, short rods (0.9 x 1.5-3.0 µm) occurring singly or in pairs, motile and non-sporeforming. Colonies are beige, round, convex and smooth with entire margins. Facultatively-anaerobic, aerogenic, oxidase negative, glucose dehydrogenase and gluconate dehydrogenase are produced. Indole, H_2S and urease are not produced. The following carbon sources are utilized at 28 °C within three to six days: trans-aconitate, L-arabinose, D-arabitol, D-cellobiose, citrate, dulcitol, gentiobiose, L-fucose, D-fructose, D-galactose, D-galacturonate, D-glucose, inositol, lactose, D-malate, D-maltose, maltotriose, D-mannitol, D-mannose, L-rhamnose, D-ribose, L-tartrate, D-trehalose, and D-xylose. The following carbon sources are not utilized at 28 °C within three to six days: D-adonitol, L-arabitol, betaine, erythritol, glutarate, histamine, 5-ketogluconate, lactulose, D-melibiose, 3-0-methyl-D-glucose, propionate, quinate, D-raffinose, D-sorbitol, L-sorbose, sucrose, D-tagatose, D-tartrate, meso-tartrate, trigonelline, D-turanose, L-tyrosine and xylitol.

The G + C content of the type strain is 55.7 mol %. The type strain is LMG 24534^T (= BD 805^T = CDC 3527-71) and was isolated from a human blood sample in Paris, France.

Acknowledgements

This study was partially supported by the South African-Flemish Bilateral Agreement, the National Research Foundation (NRF), the Tree Protection Co-operative Programme (TPCP) and the THRIP support programme of the Department of Trade and Industry, South Africa. The BCCM/LMG Bacteria collection is supported by the Federal Public Planning Service-Science Policy, Belgium. The authors wish to acknowledge Katrien Vandemeulebroecke for technical assistance and Mrs. Mohr for providing us with the CDC strains.

References

Beji, A., Mergaert, J., Gavini, F., Izard, D., Kersters, K., Leclerc, H. & De Ley, J. (1988). Subjective synonymy of *Erwinia herbicola*, *Erwinia milletiae*, and *Enterobacter agglomerans* and redefinition of the taxon by genotypic and phenotypic data. *Int J Syst Bacteriol* **38,** 77-88

Brady, C.L., Cleenwerck, I., Venter, S.N., Vancanneyt, M., Swings, J. & Coutinho, T.A. (2008). Phylogeny and identification of *Pantoea* species associated with the environment, humans and plants based on mulitlocus sequence analysis (MLSA). Submitted to *Appl Environ Microbiol*

Brady, C.L., Venter, S.N., Cleenwerck, I., Engelbeen, K., Vancanneyt, M., Swings, J. & Coutinho, T.A. (2008). *Pantoea vagens* sp. nov., *Pantoea eucalypti* sp. nov., *Pantoea deleyii* sp. nov. and *Pantoea anthophila* sp. nov., four novel species belonging to the Genus *Pantoea*. Submitted to *Int J Syst Bacteriol*

Brenner, D.J., Fanning, G.R., Leete Knutson, J.K., Steigerwalt, A.G. & Krichevsky, M.I. (1984). Attempts to classify Herbicola Group-*Enterobacter agglomerans* strains by deoxyribonucleic acid hybridization and phenotypic tests. *Int J Syst Bacteriol* **34,** 45-55

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

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

Dye, D.W. (1981). A numerical taxonomic study of the genus *Erwinia*. *N Z J Agric Res* **24,** 223-229

Ewing, W.H. & Fife, M.A. (1972). *Enterobacter agglomerans* (Beijerinck) comb. nov. (the Herbicola-Lathyri Bacteria). *Int J Syst Bacteriol* **22,** 4-11

Ezaki, T., Hashimoto, Y. & Yabuuchi, E. (1989). Fluorometric deoxyribonucleic acid-deoxyribonucleic acid hybridization in micro-dilution wells as an alternative to membrane filter hybridization in which radioisotopes are used to determine genetic relatedness among bacterial strains. *Int J Syst Bacteriol* **39,** 224-229

Gardner, J.M., & Kado, C.I. (1972). Comparative base sequence homologies of the deoxyribonucleic acids of *Erwinia* species and other *Enterobacteriaceae*. *Int J Syst Bacteriol* **22,** 201-209

Gavini, F., Mergaert, J., Beji, A., Mielcarek, C., Izard, D., Kersters, K. & De Ley, J. (1989). Transfer of *Enterobacter agglomerans* (Beijerinck 1988) Ewing and Fife 1972 to *Pantoea* gen. nov. as *Pantoea agglomerans* comb. nov. and description of *Pantoea dispersa* sp. nov. *Int J Syst Bacteriol* **39,** 337-345

Grimont, P.A.D. & Grimont, F. (2005). Genus: *Pantoea*. In *Bergey's Manual of Systematic Bacteriology*, Volume Two, The *Proteobacteria*, Part B, The *Gammaproteobacteria*, pp. 713-720, Edited by D.J. Brenner, N.R. Krieg & J.T. Staley. 2nd Edition. New York: Springer

Guindon, S. & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52,** 696-704

Maki, D.G., Rhame, F.S., Mackel, D.C. & Bennett, J.V. (1976). Nationwide epidemic of septicemia caused by contaminated intravenous products. Am J Med **60,** 471-485

Mergaert, J., Verdonck, L. & Kersters, K. (1993). Transfer of *Erwinia ananas* (synonym, *Erwinia uredovora*) and *Erwinia stewartii* to the Genus *Pantoea* emend. as *Pantoea ananas* (Serrano 1928) comb. nov. and *Pantoea stewartii* (Smith 1898) comb. nov., Respectively, and Description of *Pantoea stewartii* subsp. *indologenes* subsp.nov. *Int J Syst Bacteriol* **43,** 162-173

Mesbah, M., Premachandran, U. & Whitman, W.B. (1989). Precise measurement of the G + C content of deoxyribonucleic acid by high-performance liquid chromatography. *Int J Syst Bacteriol* **39,** 159-167

Niemann, S., Pühler, 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. *J Appl Microbiol* **82,** 477-484

Posada, D. & Crandall, K.A. (1998). Modeltest: testing the model of DNA substitution. *Bioinformatics* **14,** 817-818

Swofford, D.L. (2000). PAUP*: Phylogenetic Analysis Using Parsimony and other methods (software). Sinauer Associates, Sunderland, MA.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. & Higgins, D.G. (1997). The ClustalX-Windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucl Acids Res* **25,** 4876-4882

Verdonck, L., Mergaert, J., Rijckaert, C., Swings, J., Kersters, K. & De Ley, J. (1987). The genus *Erwinia*: A numerical analysis of phenotypic features. *Int J Syst Bacteriol* **37,** 4-18

Wilson, K. (1987). Preparation of genomic DNA from bacteria. In *Current Protocols in Molecular Biology*, pp. 2.4.1-2.4.5. Edited by F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith & K. Struhl. New York: Green Publishing and Wiley-Interscience

Young, J.M., Dye, D.W., Bradbury, J.F., Panagopoulos, C.G. & Robbs, C.R. (1978). A proposed nomenclature and classification for plant pathogenic bacteria. *N Z J Agric Res* **21,** 153-177

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 1: Strains used in this study

LMG = BCCM/LMG Bacteria Collection, Ghent University, Belgium, CDC = Centers for Disease Control, Atlanta, Georgia, U.S.A, ATCC = American Type Culture Collection, Rockville, Maryland, U.S.A., BD = Plant Pathogenic and Plant Protecting Bacteria (PPPPB) Culture Collection, ARC-PPRI, Pretoria, South Africa

Species	**Strain**	**Host**	**Location**
Pantoea agglomerans	LMG 1286^T	Human	Zimbabwe
	LMG 2660	*Wisteria floribunda*	Japan
Pantoea ananatis	LMG 2665^T	Pineapple	Brazil
	LMG 20103	*Eucalyptus*	South Africa
	LMG 24190	Onion	South Africa
Pantoea stewartii ssp. *stewartii*	LMG 2715^T	Maize	USA
	LMG 2718	Maize	USA
Pantoea stewartii ssp. *indologenes*	LMG 2632^T	Fox millet	India
	LMG 2673	Pineapple	Hawaii
Pantoea dispersa	LMG 2603^T	Soil	Japan
	LMG 2604	Wild rose	Netherlands
Pantoea citrea	LMG 22049^T	Mandarin orange	Japan
Pantoea punctata	LMG 22050^T	Mandarin orange	Japan
	LMG 23562	Mandarin orange	Japan
Pantoea terrea	LMG 22051^T	Soil	Japan
	LMG 23564	Soil	Japan
Pantoea anthophila	LMG 2558^T	*Impatiens balsamina*	India
	LMG 2560	*Tagetes erecta*	Unknown
Pantoea vagens	LMG 24199^T	*Eucalyptus*	Uganda
	LMG 24195	*Eucalyptus*	Uruguay
	LMG 24196	*Eucalyptus*	Argentina
	LMG 24201	Maize	South Africa

Pantoea eucalypti	LMG 24198 ^{T}	*Eucalyptus*	Uruguay
	LMG 24197	*Eucalyptus*	Uruguay
Pantoea deleyii	LMG 24200^{T}	*Eucalyptus*	Uganda
Pantoea septica (Brenner group II)	LMG 5345^{T} = ATCC 29923 = CDC 3123-70	Human, stool	New Jersey, USA
	LMG 24526 = CDC 238-70	Human, blood	New York, USA
	LMG 24527 = CDC 1778-70	Human, blood	Oklahoma, USA
	LMG 24528 = CDC 217-71	Human, skin	Virginia, USA
Pantoea eucrina (Brenner group IV)	LMG 2781^{T} = ATCC 27998 = CDC 1741-71	Human, trachea	Connecticut, USA
	LMG 24529 = CDC 3638-70	Human, cyst	Georgia, USA
	LMG 24530 = CDC 5795-70	Human, urine	Virginia, USA
	LMG 24531 = CDC 6148-70	Human, spinal fluid	Hawaii, USA
Pantoea brenneri (Brenner group V)	LMG 5343^{T} = ATCC 29921 = CDC 3482-71	Human, urethra	Montana, USA
	LMG 24532 = CDC 2928-68	Human, sputum	Wisconsin, USA
	LMG 24533 = CDC 2525-70		Quebec, Canada
Pantoea conspicua (Brenner group V)	LMG 24534^{T} = BD 805^{T} = CDC 3527-71	Human, blood	Paris, France

Table 2: Phenotypic characteristics distinguishing *Pantoea septica* sp. nov., *Pantoea eucrina* sp. nov., *Pantoea brenneri* sp. nov. and *Pantoea conspicua* sp. nov. from each other and from their closest phylogenetic neighbours

1 = *P. agglomerans* (3), 2 = *P. ananatis* (4), 3 = *P. dispersa* (2), 4 = *P. septica* sp. nov. (1) (Brenner II), 5 = *P. eucrina* sp. nov. (1) (Brenner IV), 6 = *P. brenneri* sp. nov. (1) (Brenner V), 7 = *P. conspicua* sp. nov. (1) (Brenner V, LMG 24534^T)

+, 90-100 % of strains positive in 1-2 days; (+), 90-100 % of strains positive in 1-4 days; -, 90-100 % of strains negative in 4 days; d, positive in 1- 4 days; (d), positive in 3-4 days; ND, not determined

Characteristic	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Phenylalanine deaminase	d	-	-	+	d	d	+
Dulcitol	-	-	(d)	(d)	-	-	+
Erythritol	-	-	+	-	d	-	-
Gentiobiose	-	+	+	d	d	-	+
Lactose	-	+	-	d	-	(d)	+
Lactulose	-	+	-	d	-	d	-
Raffinose	-	+	-	d	-	d	-
Sorbitol	-	d	-	(d)	-	-	-
L-Tartrate	-	-	d	d	-	d	+
meso-Tartrate	(+)	d	d	d	-	(+)	-

Supplementary Table A: DNA-DNA hybridization values amongst strains belonging to the novel species *Pantoea septica*, *Pantoea eucrina*, *Pantoea brenneri* and *Pantoea conspicua* and reference strains of the seven validly described species of the genus *Pantoea*.

Figure 1: Maximum likelihood tree based on complete 16S rRNA sequences of *Pantoea* species. The tree was generated by the Phyml software using the Tamura-Nei (TN93) model as selected by Modeltest. Bootstrap values after 1000 replicates are expressed as percentages. *Esherichia coli* was included as an outgroup.

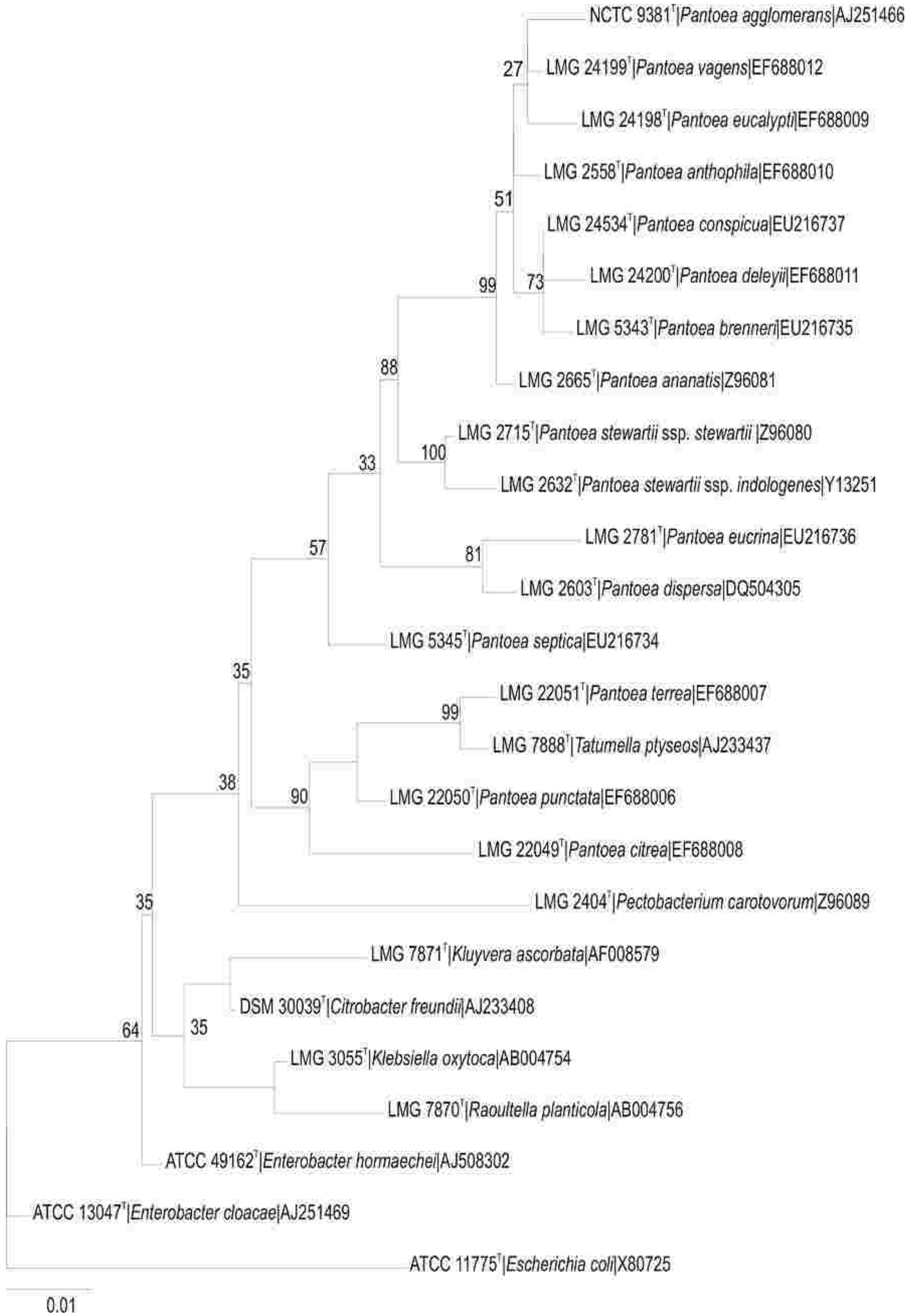
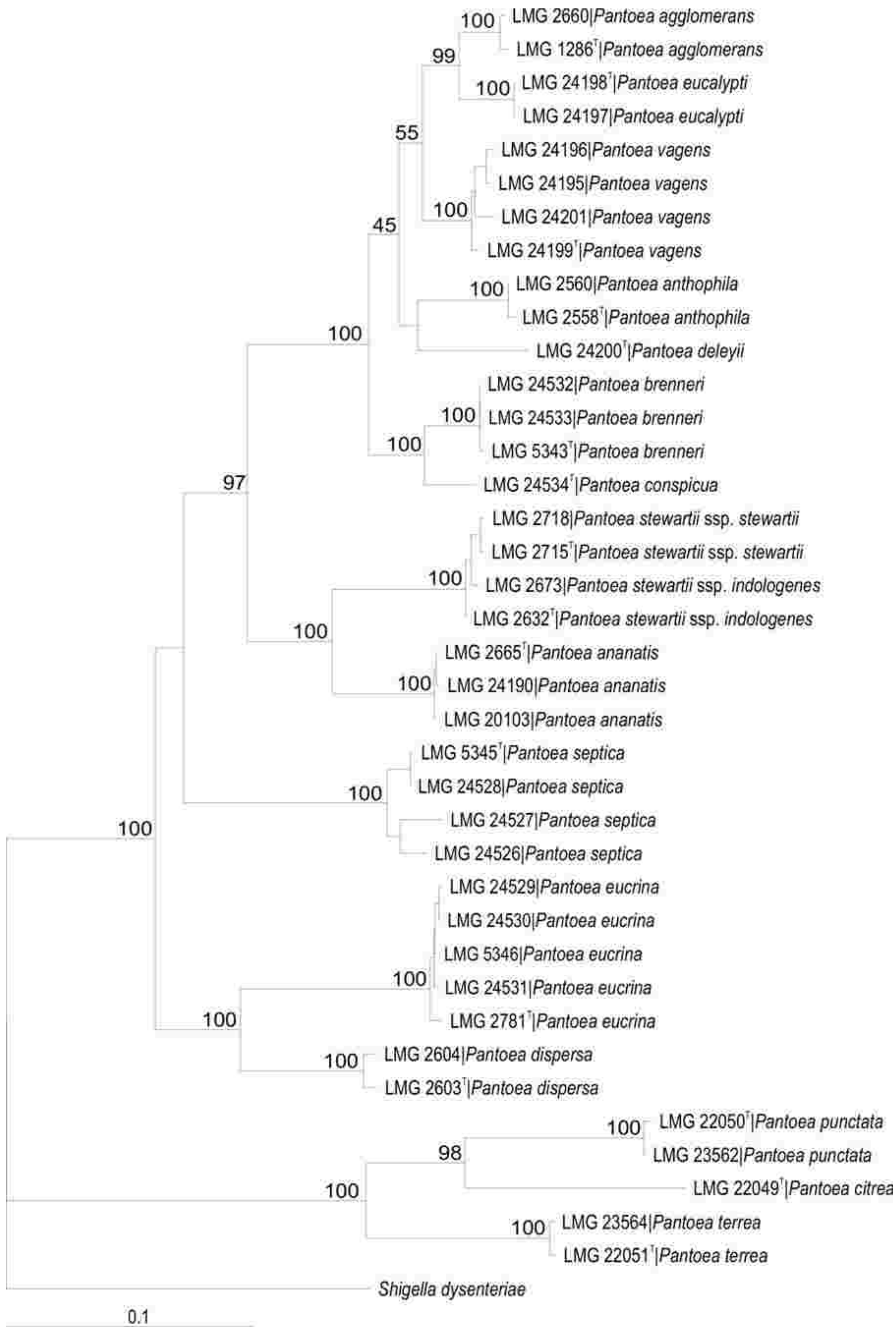

NCTC 9381T|Pantoea agglomerans|AJ251466
LMG 24199T|Pantoea vagens|EF688012
LMG 24198T|Pantoea eucalypti|EF688009
LMG 2558T|Pantoea anthophila|EF688010
LMG 24534T|Pantoea conspicua|EU216737
LMG 24200T|Pantoea deleyii|EF688011
LMG 5343T|Pantoea brenneri|EU216735
LMG 2665T|Pantoea ananatis|Z96081
LMG 2715T|Pantoea stewartii ssp. stewartii |Z96080
LMG 2632T|Pantoea stewartii ssp. indologenes|Y13251
LMG 2781T|Pantoea eucrina|EU216736
LMG 2603T|Pantoea dispersa|DQ504305
LMG 5345T|Pantoea septica|EU216734
LMG 22051T|Pantoea terrea|EF688007
LMG 7888T|Tatumella ptyseos|AJ233437
LMG 22050T|Pantoea punctata|EF688006
LMG 22049T|Pantoea citrea|EF688008
LMG 2404T|Pectobacterium carotovorum|Z96089
LMG 7871T|Kluyvera ascorbata|AF008579
DSM 30039T|Citrobacter freundii|AJ233408
LMG 3055T|Klebsiella oxytoca|AB004754
LMG 7870T|Raoultella planticola|AB004756
ATCC 49162T|Enterobacter hormaechei|AJ508302
ATCC 13047T|Enterobacter cloacae|AJ251469
ATCC 11775T|Escherichia coli|X80725
27
51
73
99
88
100
33
81
57
35
99
90
38
35
35
64
0.01

Figure 2: Maximum likelihood tree based on concatenated partial sequences of *rpoB*, *atpD*, *gyrB* and *infB* of *Pantoea* strains. The tree was generated by the Phyml software using the general time reversible (GTR) model as selected by Modeltest. Bootstrap values after 1000 replicates are expressed as percentages. *Shigella dysenteriae* sequences were obtained from the genome sequencing database of the Sanger Institute (http://www.sanger.ac.uk) and included as an outgroup.


LMG 2660|Pantoea agglomerans
LMG 1286T|Pantoea agglomerans
LMG 24198T|Pantoea eucalypti
LMG 24197|Pantoea eucalypti
LMG 24196|Pantoea vagens
LMG 24195|Pantoea vagens
LMG 24201|Pantoea vagens
LMG 24199T|Pantoea vagens
LMG 2560|Pantoea anthophila
LMG 2558T|Pantoea anthophila
LMG 24200T|Pantoea deleyii
LMG 24532|Pantoea brenneri
LMG 24533|Pantoea brenneri
LMG 5343T|Pantoea brenneri
LMG 24534T|Pantoea conspicua
LMG 2718|Pantoea stewartii ssp. stewartii
LMG 2715T|Pantoea stewartii ssp. stewartii
LMG 2673|Pantoea stewartii ssp. indologenes
LMG 2632T|Pantoea stewartii ssp. indologenes
LMG 2665T|Pantoea ananatis
LMG 24190|Pantoea ananatis
LMG 20103|Pantoea ananatis
LMG 5345T|Pantoea septica
LMG 24528|Pantoea septica
LMG 24527|Pantoea septica
LMG 24526|Pantoea septica
LMG 24529|Pantoea eucrina
LMG 24530|Pantoea eucrina
LMG 5346|Pantoea eucrina
LMG 24531|Pantoea eucrina
LMG 2781T|Pantoea eucrina
LMG 2604|Pantoea dispersa
LMG 2603T|Pantoea dispersa
LMG 22050T|Pantoea punctata
LMG 23562|Pantoea punctata
LMG 22049T|Pantoea citrea
LMG 23564|Pantoea terrea
LMG 22051T|Pantoea terrea
Shigella dysenteriae
100
99
100
55
100
45
100
100
100
100
97
100
100
100
100
100
100
100
100
98
100
100
0.1

CHAPTER 6

Transfer of *Pantoea citrea*, *Pantoea punctata* and *Pantoea terrea* to the genus *Tatumella* emend. as *Tatumella citrea* comb. nov., *Tatumella punctata* comb. nov., and *Tatumella terrea* comb. nov. (Kageyama *et al*., 1992) and the description of *Tatumella morbirosei* sp. nov.

As submitted to: International Journal of Systematic and Evolutionary Microbiology

Summary

Three *Pantoea* species were described for strains isolated from fruit and soil originating in Japan, namely, *Pantoea citrea*, *Pantoea punctata* and *Pantoea terrea*. These three "Japanese" species have been shown to be phylogenetically distant to the remaining species of the genus *Pantoea*. It has previously been observed using multilocus sequence analysis (MLSA) of *Pantoea* strains that the "Japanese" species consistently form a distinct clade with an extended branch length, casting doubt on the inclusion of these species within the genus. Furthermore, the "Japanese" species cluster closely with *Tatumella ptyseos*, strains of which originate from human clinical specimens. DNA-DNA hybridization and phenotypic tests confirmed the phylogenetic distance of *P. citrea*, *P. punctata* and *P. terrea* from the genus *Pantoea* and the affiliation of these species with *Tatumella*. In addition, strains causing pink disease of pineapple, previously identified as *P. citrea*, were proven to be separate species by 16S rRNA, MLSA and DNA-DNA hybridization data. The name *Tatumella morbirosei* sp. nov. (LMG 23360^T = NCPPB 4036^T = CMC6) is proposed for the causal agent of pink disease of pineapple. The new combinations *Tatumella citrea* (Kageyama *et al*. 1992) comb. nov. (LMG 22049^T = ATCC 31623^T = SHS 2003^T), *Tatumella punctata* (Kageyama *et al*. 1992) comb. nov. (LMG 22050^T = ATCC 31626^T = SHS 2006^T) and *Tatumella terrea* (Kageyama *et al*. 1992) comb. nov. (LMG 22051^T = ATCC 31628^T = SHS 2008^T) are proposed for *P. citrea, P. punctata* and *P. terrea*, respectively.

Following the proposal of the novel genus *Pantoea* in 1989 (Gavini *et al*., 1989), but preceding the transfer of *Erwinia ananas* and *Erwinia stewartii* to *Pantoea* (Mergaert *et al*., 1993), three novel *Pantoea* species were described from fruit and soil samples in Japan (Kageyama *et al*., 1992). *Pantoea citrea*, *Pantoea punctata* and *Pantoea terrea* all produced 2,5-diketo-D-gluconic acid (DKGA) from D-glucose and were included in the genus *Pantoea* based on phenotypic data and DNA-DNA hybridization values, despite the inability of other *Pantoea* species to produce DKGA. Until recently, no phylogenetic study had been performed on all validly published *Pantoea* species, giving no reason to doubt the inclusion of *P. citrea*, *P. punctata* and *P. terrea* in the genus *Pantoea*. However, the most recent edition of Bergey's Manual of Systematic Bacteriology (Grimont & Grimont, 2005) states that more taxonomic work is required to justify the assignment of these three species to the genus *Pantoea*.

A recent phylogenetic study of the *Enterobacteriaceae* revealed an *atpD* sequence indel which is specific to *Pantoea* and *Tatumella* (Paradis *et al*., 2005), indicating a close phylogenetic relationship between these two genera. This was in agreement with the initial suggestion of P. Grimont that the "Japanese" species may be more similar to *Tatumella ptyseos* than *Pantoea* in their nutritional patterns (Kageyama *et al*., 1992). The single species genus *Tatumella* was created for clinical strains isolated in North and South America between 1960 and 1980 (Hollis *et al*., 1981). A MLSA scheme based on *rpoB*, *atpD*, *gyrB* and *infB* genes was recently performed on 102 *Pantoea* strains including the "Japanese" species and *Tatumella ptyseos* (Brady *et al*., submitted). A concatenated tree constructed from the sequences of the four genes was found to be the most reliable method for determining phylogenetic relationships amongst *Pantoea* strains. The MLSA study indicated a clear phylogenetic division of *P. citrea*, *P. punctata* and *P. terrea* from the remainder of the *Pantoea* species.

Strains used in this study were obtained from the BCCM/LMG Bacteria Collection (http://www.belspo.be/bccm) and the Centers for Disease Control, Atlanta, Georgia, U.S.A., and are listed in Table 1. An alkalic extraction method was used to isolate genomic DNA from the strains (Niemann *et al*., 1997) which was stored at -20 °C. The complete 16S rRNA gene was amplified and sequenced for the type strains of *P. citrea* (LMG 22049^{T}), *P. punctata* (LMG 22050^{T}) and *P. terrea* (LMG 22051^{T}), as well as one of the additional *P. citrea* strains (LMG 23360) found to cause pink disease of pineapple (Cha *et al*., 1997) using the primers and conditions determined by Coenye *et al*. (1999). MLSA was performed on all strains (Brady *et al*., submitted).

The GenBank/EMBL accession numbers for the 16S rRNA gene sequences for *Tatumella ptyseos* LMG 7888T, *Tatumella punctata* LMG 22050^T, *Tatumella terrea* LMG 22051^T, *Tatumella citrea* LMG 22049^T and *Tatumella morbirosei* LMG 23360^T are EU344770, EF688006-EF688008 and EU344769, respectively.

The sequences were aligned using ClustalX (Thompson *et al.*, 1997) and the overhangs were trimmed. The Modeltest 3.7 programme (Posada & Crandall, 1998) was then applied to determine the best-fit evolutionary model. Maximum likelihood and neighbour joining analyses were performed using Phyml (Guindon & Gascuel, 2003) and PAUP 4.0b10 (Swofford, 2000) respectively, by applying the models and parameters determined by Modeltest (only Maximum likelihood phylogenetic trees are shown). Bootstrap analysis with 1000 replicates was performed to assess the support for these clusters.

In the 16S rRNA tree (Fig. 1) the genus *Pantoea* is phylogenetically split into two sublineages. The majority of the *Pantoea* "core" species are contained in a cluster supported by high bootstrap support, whilst the "Japanese" species are situated on a distinctly separate branch, also with high bootstrap support of 89 %. These findings support those of Grimont & Grimont, based on 16S rRNA and *rpoB* sequence comparisons, that the "Japanese" species cluster at a lower level to the *Pantoea* "core" clade (2005). Interestingly, the type strain of *Tatumella ptyseos* (LMG 7888^T) clusters closely with the type strain of *Pantoea terrea*, within the "Japanese" species clade. LMG 23360, one of the *P. citrea* strains causing pink disease of pineapple, does not cluster with the type strain of this species (LMG 22049^T), but is found on a separate branch. The 16S rRNA sequence similarity of LMG 23360 was greater than 98 % to the type strains of *P. punctata*, *P. citrea*, *P. septica* and *T. ptyseos*. It is also interesting to note that *P. septica*, *P. eucrina* and *P. dispersa*, which are considered "core" species, cluster separately from the remaining *Pantoea* species in Fig. 1. A similar observation was made in a recent study examining the reliability of 16S rRNA sequence for phylogenetic analysis of the *Enterobacteriaceae* (Naum *et al*., 2008). It was observed that adding additional animal pathogen sequences to a 16S rRNA phylogenetic analysis would affect the taxon placement of the phytopathogenic enterobacteria, particularly *Erwinia*, *Brenneria* and *Pectobacterium*. This indicated that 16S rRNA gene sequences may be inadequate for determining the true phylogeny of phytopathogenic *Enterobacteriaceae* at the genus level. The discrepancies seen in the 16S rRNA phylogeny can be resolved with the assistance of MLSA.

In Fig. 2, a MSLA phylogenetic tree based on the concatenated sequences of *rpoB*, *atpD*, *gyrB* and *infB* genes, *Pantoea* strains form a monophyletic cluster within the *Enterobacteriaceae*. This cluster contains two sublineages which are supported by high bootstrap values. One of the sublineages consists of the "Japanese" species and the type strain of *T. ptyseos*. The type strain of *T. ptyseos* clusters closely with *P. terrea* strains, specifically with LMG 23565 and its synonym CCUG 30163. This result indicates that *P. terrea* might be a subjective synonym of *T. ptyseos* or another possibility is that only LMG 23565 (syn. CCUG 30163) belongs to *T. ptyseos*. The type strain of *P. citrea* (LMG 22049^T) clusters with the other two *P. citrea* strains causing pink disease of pineapple (LMG 23359 and LMG 23360), but is situated on a separate branch. The distance between the *P. citrea* type strain and the pink disease-causing strains is comparable to the one found between *P. agglomerans* and the newly described *P. eucalypti* (Brady *et al.*, submitted), and suggests that LMG 23359 and LMG 23360 do not belong to *P. citrea*.

High quality DNA for DNA-DNA hybridization of strains was prepared by the method of Wilson (1987), with minor modifications (Cleenwerck *et al*., 2002). DNA-DNA hybridizations were performed using the microplate method (Ezaki *et al*., 1989) with some modifications (Cleenwerck *et al*., 2002). The hybridization temperature was 45 °C ± 1 °C and reciprocal reactions were performed with all strains. Representative strains from each "Japanese" *Pantoea* species were selected for hybridization based on the 16S rRNA- and MLSA- phylogenetic trees. The type strains of the "Japanese" species were hybridized amongst each other, and with the type strain of *Tatumella ptyseos*, the phylogenetically closest neighbour, as well as with the type strains of *P. agglomerans*, *P. ananatis*, *P. vagens*, *P. stewartii* ssp. *stewartii* and *P. dispersa*. A summary of the hybridization results is presented in Table 2. The type strains of *P. citrea*, *P. punctata* and *P. terrea* (LMG 22049^T, LMG

22050^{T}, LMG 22051^{T}) exhibited 13-21 % DNA similarity when hybridized amongst each other, which is considerably lower than the 28-43 % observed by Kageyama *et al*. (1992) and could be due to different hybridization methods used, namely, the S1 nuclease procedure as described by Johnson (1981). LMG 23360, a strain causing pink disease of pineapple (Cha *et al*., 1997) displayed only 42 % DNA similarity with the type strain of *P. citrea* (LMG 22049^{T}). This result together with the 16S rRNA and MLSA data proves that LMG 23360 and LMG 23359 do not belong to *P. citrea*. *Pantoea citrea* and *P. punctata* displayed 14 and 21 % DNA relatedness when hybridized to the type strain of *T. ptyseos* (LMG 7888^{T}), respectively. In contrast, *T. ptyseos* shared 66 % DNA relatedness with the type strain of *P. terrea*, and 87 % DNA relatedness with strain LMG 23565, thought to be *P. terrea*. LMG 23565 demonstrated only 55 % DNA similarity, when hybridized to the type strain of the type strain of *P. terrea*. These results prove that LMG 23565 (syn. CCUG 30163) must be re-classified as *T. ptyseos* and confirm the close phylogenetic relationship between *P. terrea* and *T. ptyseos*, observed in both the 16S rRNA- and MLSA-phylogenetic trees (Figs. 1 & 2).

The DNA relatedness between the type strains of the "Japanese" *Pantoea* species and the type strains of *P. agglomerans*, *P. ananatis*, *P. vagens*, *P. stewartii* and *P. dispersa* was below 10 %. These hybridization values are again lower than those observed by Kageyama *et al*. (1992) between the "Japanese" species and *P. agglomerans* and *P. dispersa* (19 to 22 %). The DNA relatedness values obtained provide further evidence for the exclusion of the "Japanese" species from the genus *Pantoea*.

Based on the close relationship between *T. ptyseos* and *P. terrea*, the consistent clustering of the "Japanese" species with the type strain of *T. ptyseos*, and the DNA hybridization data we propose to transfer *P. citrea*, *P. punctata* and *P. terrea* to the genus *Tatumella* as *Tatumella citrea* comb. nov., *Tatumella punctata* comb. nov. and *Tatumella terrea* comb. nov. We further propose *Tatumella morbirosei* sp. nov. for strains LMG 23359 and LMG 23360, the causal agent of pink disease of pineapple.

The G + C content of the type strains of all four "Japanese" species and *Tatumella ptyseos*, determined by HPLC as published by Mesbah *et al*. (1989), are as follows:

Tatumella citrea comb. nov. (LMG 22049^T) = 49.8 mol %; *Tatumella morbirosei* sp. nov. (LMG 23360^T) = 50.2 mol %; *Tatumella punctata* comb. nov. (LMG 22050^T, LMG 22098) = 50.7 mol %, *Tatumella terrea* comb. nov. (LMG 22051^T, LMG 23564) = 52.6-52.8 mol % and *Tatumella ptyseos* (LMG 7888T, LMG 23565) = 51.7-52.1 mol %.

Physiological and biochemical characteristics for the "Japanese" species were obtained from Bergey's Manual of Systematic Bacteriology (Grimont & Grimont, 2005). Additional tests, including API 50E and Biotype-100 tests (bioMérieux) were performed on the type strains. A summary of distinguishing characteristics of species of the genus *Tatumella* are listed in Table 3. The most important characteristic distinguishing *T. citrea*, *T. morbirosei* sp. nov., *T. punctata*, *T. terrea* and *T. ptyseos* from *Pantoea* species is their ability to produce 2-ketogluconate dehydrogenase which oxidizes 2-ketogluconate to 2,5-diketo-D-gluconic acid (DKGA). Noticeably, not one *Pantoea* species has this ability but there are other species belonging to the *Enterobacteriaceae* which can produce DKGA from D-glucose, for example, *Pectobacterium cypripedii*, *Ewingella americana*, *Rahnella aquatilis* and *Serratia marcescens* (Bouvet *et al*., 1989).

In a personal communication to Kageyama *et al*., P. Grimont suggested that the "Japanese" species were phenotypically similar to *Tatumella*, however Kageyama *et al*. (1992) listed several characteristics in which the "Japanese" species differed from *T. ptyseos*. These included acid production from D-xylose and L-arabinose, arginine dihydrolase activity, methyl red reaction, Voges-Proskauer reaction, esculin hydrolase activity and citrate utilization. A closer examination of *T. ptyseos* revealed that this species is in fact positive for the above characteristics listed by Kageyama *et al.* (1992), as are the "Japanese" species. Another reason given by Kageyama *et al.* (1992) not to include the "Japanese" species in the genus *Tatumella* was the G + C content of *T. ptyseos* (Hollis *et al*., 1981), which is 53-54 mol %. They felt that it was too high when compared to the "Japanese" species which range from 49.7 to 51.9 mol %. However, the G + C contents of *P. agglomerans* and *P. dispersa* are 54.9 and 56.9 mol %, respectively which is even higher than that of *T. ptyseos* (Gavini *et al*., 1989).

Emended description of the genus *Tatumella* Hollis, Hickman, Fanning, Farmer, Weaver & Brenner 1981

(Ta.tum.el'la. M.L. dim. Neut. –*ella* ending; M.L. fem. n. *Tatumella* named to honour Harvey Tatum, an American bacteriologist who made many contributions to our understanding of the classification and identification of fermentative and nonfermentative bacteria of medical importance.) The description below is based on the data of Hollis *et al*. (1981), Kageyama *et al*. (1992) and this paper.

Gram-negative, non-capsulated, non-sporeforming small rods that are 0.6-1.2 x 0.9-3.0 µm in size. Cells are motile by means of polar, subpolar or lateral flagella or non-moltile, can be non-motile at 36 °C. Facultatively anaerobic, fermentative, catalyse positive (weak and slow), oxidase negative. Non-pigmented, or pale beige to pale orange. Glucose dehydrogenase, gluconate dehydrogenase and 2-ketogluconate dehydrogenase are produced. Reduce nitrate to nitrite. Indole, urease and gelatin tests are negative. Positive for Voges-Proskauer (Coblentz) and citrate (Simmons), phenylalanine, L-arginine dihydrolase and ONPG tests are variable. Negative for H_2S (TSI), lysine decarboxylase, ornithine decarboxylase, tryptophan deaminase, KCN test, lipase and DNase. Acid is produced from D-glucose, D-trehalose, sucrose, D-cellobiose, ribose, L-arabinose, D-arabitol, glycerol, *myo*-inositol, D-mannitol and L-malate, but not from L-sorbose, D-melibiose, D-raffinose, glycogen, histamine, glutarate and propionate. Susceptible to many antibiotics. Isolated from human clinical samples, fruit and soil. The G + C contents range from 49.8 to 53 mol %. The type species is *Tatumella ptyseos* Hollis, Hickman, Fanning, Farmer, Weaver & Brenner 1981.

Description of *Tatumella citrea* (Kageyama, Nakae, Yagi & Sonoyama 1992) comb. nov.

Cells are Gram-negative, short rods (0.8-1.2 x 1.0-3.0 µm) occurring singly or in pairs, non-motile and non-sporeforming. Colonies are pale beige to pale orange, round, convex and smooth with entire margins. Nicotinic acid or nicotinamide are required for growth. Facultatively-anaerobic, oxidase negative, catalase positive, glucose dehydrogenase, gluconate dehydrogenase and 2-ketogluconate dehydrogenase are produced. Indole, urease and phenylalanine deaminase are negative. Reduce nitrate to nitrite. The following carbon sources are utilized at 28 °C within three to six days: trans-aconitate, D-adonitol (weak), L-arabinose, D-arabitol, L-arabitol (weak), D-cellobiose (weak), citrate, erythritol, D-fructose, D-galactose, D-galacturonate, D-glucose, inositol, 5-ketogluconate, lactose (weak), lactulose, D-malate, D-maltose, maltotriose, D-mannitol, quinate, L-rhamnose, D-ribose, sucrose, D-tagatose, L-tartrate, meso-tartrate, D-trehalose, xylitol (weak) and D-xylose. The following carbon sources are not utilized at 28 °C within three to six days: L-alanine, betain, dulcitol, L-fucose, gentiobiose, glutarate, histamine, D-mannose, D-melibiose, propionate, D-raffinose, D-sorbitol, L-sorbose, D-tartrate, trigonelline, L-tryptophan, D-turanose, L-tyrosine.

The G + C content of the type strain is 49.8 mol %. The type strain is LMG 22049^{T} (= ATCC 31623^{T} = SHS 2003^{T}) and was isolated from mandarin orange in Japan.

Description of *Tatumella punctata* (Kageyama, Nakae, Yagi & Sonoyama 1992) comb. nov.

Cells are Gram-negative, short rods (1.1-1.2 x 1.3-2.3 µm) occurring singly or in pairs, non-motile and non-sporeforming. Colonies are pale beige to pale orange, round, convex and smooth with entire margins. Nicotinic acid or nicotinamide are required for growth. Facultatively-anaerobic, oxidase negative, catalase positive, glucose dehydrogenase, gluconate dehydrogenase and 2-ketogluconate dehydrogenase are produced. Indole, urease and phenylalanine deaminase are negative. Reduce nitrate to nitrite. The following carbon sources are utilized at 28 °C within three to six days: trans-aconitate, L-alanine, L-arabinose, D-arabitol, D-cellobiose, citrate, dulcitol (weak), erythritol, D-fructose, D-galactose, D-galacturonate, gentiobiose, D-glucose, inositol, 5-ketogluconate, lactulose (weak), D-malate (weak), D-maltose, maltotriose, D-mannitol, D-mannose, quinate (weak), L-rhamnose, D-ribose, D-sorbitol (weak), sucrose, D-tagatose (weak), L-tartrate, meso-tartrate, D-trehalose, D-turanose (weak) and D-xylose. The following carbon sources are not utilized at 28 °C within three to six days: D-adonitol, L-arabitol, betain, L-fucose, glutarate, histamine, lactose, D-melibiose, propionate, D-raffinose, L-sorbose, D-tartrate, trigonelline, L-tryptophan, L-tyrosine and xylitol.

The G + C content of the type strain is 50.7 mol %. The type strain is LMG 22050^{T} (= ATCC 31626^{T} = SHS 2006^{T}) and was isolated from mandarin orange in Japan.

Description of *Tatumella terrea* (Kageyama, Nakae, Yagi & Sonoyama 1992) comb. nov.

Cells are Gram-negative, short rods (0.8-0.9 x 1.2-2.0 µm) occurring singly or in pairs, motile by means of one or two lateral flagella and non-sporeforming. Colonies are pale beige to pale orange, round, convex and smooth with entire margins. Nicotinic acid or nicotinamide are required for growth. Facultatively-anaerobic, oxidase negative, catalase positive, glucose dehydrogenase, gluconate dehydrogenase and 2-ketogluconate dehydrogenase are produced. Indole, urease and phenylalanine deaminase are negative. Reduce nitrate to nitrite. The following carbon sources are utilized at 28 °C within three to six days: trans-aconitate (weak), D-adonitol (weak), L-arabinose, D-arabitol, D-cellobiose, citrate, erythritol (weak), D-fructose, D-galactose, D-galacturonate (weak), gentiobiose, D-glucose, inositol, 5-ketogluconate, lactulose, D-malate, D-maltose, maltotriose, D-mannitol, D-mannose, quinate (weak), L-rhamnose, D-ribose, sucrose, D-tagatose, meso-tartrate (weak), D-trehalose and D-xylose. The following carbon sources are not utilized at 28 °C within three to six days: L-alanine, L-arabitol, betain, dulcitol, L-fucose, glutarate, histamine, lactose, D-melibiose, propionate, D-raffinose, D-sorbitol, L-sorbose, D-tartrate, L-tartrate, trigonelline, L-tryptophan, D-turanose, L-tyrosine, xylitol.

The G + C content of the type strain is 52.8 mol %. The type strain is LMG 22051^T = ATCC 31628^T = SHS 2008^T) and was isolated from soil in Japan.

Description of *Tatumella morbirosei* sp. nov.

Tatumella morbirosei (mor.bi.ró.se.i. L. n. *morbus* meaning disease and L. adj. *roseus* meaning rosy, pink. N.L. gen. N. *morbirosei*, of the pink disease, referring to the causal agent of pink disease of pineapple)

Cells are Gram-negative, short rods (0.8-1.2 x 1.0-3.0 μm) occurring singly or in pairs, non-motile and non-sporeforming. Colonies are pale beige, round, convex and smooth with entire margins. Facultatively-anaerobic, oxidase negative, catalase positive, glucose dehydrogenase, gluconate dehydrogenase and 2-ketogluconate dehydrogenase are produced. Indole and urease are negative, phenylalanine deaminase is weakly positive. Reduce nitrate to nitrite. The following carbon sources are utilized at 28 °C within three to six days: trans-aconitate, D-adonitol, L-arabinose, D-arabitol, D-cellobiose, citrate, erythritol, D-fructose, D-galactose, D-galacturonate, D-glucose, inositol, 5-ketogluconate, lactulose, D-malate, D-maltose, maltotriose, D-mannitol, D-mannose, quinate (weak), L-rhamnose, D-ribose, sucrose, D-tagatose, meso-tartrate (weak), D-trehalose, trigonelline, L-tyrosine and D-xylose. The following carbon sources are not utilized at 28 °C within three to six days: L-alanine, L-arabitol, betain, dulcitol, L-fucose, gentiobiose, glutarate, histamine, lactose, D-melibiose, propionate, D-raffinose, D-sorbitol, L-sorbose, D-tartrate, L-tartrate, L-tryptophan, D-turanose and xylitol.

The G + C content of the type strain is 50.2 mol %. The type strain is LMG 23360^{T} (= NCPPB 4036 = CMC6) and was isolated from pineapple in the Philippines.

Acknowledgements

This study was partially supported by the South African-Flemish Bilateral Agreement, the National Research Foundation (NRF), the Tree Protection Co-operative Programme (TPCP) and the THRIP support programme of the Department of Trade and Industry, South Africa. The BCCM/LMG Bacteria collection is supported by the Federal Public Planning Service-Science Policy, Belgium. The authors wish to thank Katrien Engelbeen for technical assistance and Dr J.P. Euzeby for suggesting the name "morbirosei".

References

Bouvet, O.M.M., Lenormand, P. & Grimont, P.A.D. (1989). Taxonomic diversity of the D-glucose oxidation pathway in the *Enterobacteriaceae*. *Int J Syst Bacteriol* **39,** 61-67

Brady, C.L., Cleenwerck, I., Venter, S.N., Vancanneyt, M., Swings, J. & Coutinho, T.A. (2008). Phylogeny and identification of *Pantoea* species associated with the environment, humans and plants based on mulitlocus sequence analysis (MLSA). Submitted to *Appl Environ Microbiol*

Brady, C.L., Venter, S.N., Cleenwerck, I., Engelbeen, K., Vancanneyt, M., Swings, J. & Coutinho, T.A. (2008). *Pantoea vagens* sp. nov., *Pantoea eucalypti* sp. nov., *Pantoea deleyii* sp. nov. and *Pantoea anthophila* sp. nov., four novel species belonging to the Genus *Pantoea*. Submitted to *Int J Syst Bacteriol*

Cha, J.-S., Pujol, C., Ducusin, A.R., Macion, E.A., Hubbard, C.H. & Kado, C.I. (1997). Studies on *Pantoea citrea*, the causal agent of pink disease of pineapple. *J Phytopathol* **145,** 313-319

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

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

Ezaki, T., Hashimoto, Y. & Yabuuchi, E. (1989). Fluorometric deoxyribonucleic acid-deoxyribonucleic acid hybridization in micro-dilution wells as an alternative to membrane filter hybridization in which radioisotopes are used to determine genetic relatedness among bacterial strains. *Int J Syst Bacteriol* **39,** 224-229

Gavini, F., Mergaert, J., Beji, A., Mielcarek, C., Izard, D., Kersters, K. & De Ley, J. (1989). Transfer of *Enterobacter agglomerans* (Beijerinck 1988) Ewing and Fife 1972 to *Pantoea* gen. nov. as *Pantoea agglomerans* comb. nov. and description of *Pantoea dispersa* sp. nov. *Int J Syst Bacteriol* **39,** 337-345

Grimont, P.A.D. & Grimont, F. (2005). Genus: *Pantoea*. In *Bergey's Manual of Systematic Bacteriology*, Volume Two, The *Proteobacteria*, Part B, The *Gammaproteobacteria*, pp. 713-720, Edited by D.J. Brenner, N.R. Krieg & J.T. Staley. 2nd Edition. New York: Springer

Guindon, S. & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52,** 696-704

Hollis, D.G., Hickman, F.W., Fanning, G.R., Farmer III, J.J., Weaver, R.E. & Brenner, D.J. (1981). *Tatumella ptyseos* gen. nov., sp. nov., a member of the family *Enterobacteriaceae* found in clinical samples. *J Clin Microbiol* **14,** 79-88

Johnson, J.L. (1981). Genetic characterization. In *Manual of methods for general bacteriology*, pp. 450-472. Edited by P. Gerhardt, R.G.E. Murray, R.N. Costilow, E.W. Nester, W.A. Wood, N.R. Krieg & G.B. Phillips. American Society for Microbiology, Washington D.C.

Kageyama, B., Nakae, M., Yagi, S. & Sonoyama, T. (1992). *Pantoea punctata* sp. nov., *Pantoea citrea* sp. nov., and *Pantoea terrea* sp. nov. isolated from fruit and soil samples. *Int J Syst Bacteriol* **42,** 203-210.

Mergaert, J., Verdonck, L. & Kersters, K. (1993). Transfer of *Erwinia ananas* (synonym, *Erwinia uredovora*) and *Erwinia stewartii* to the Genus *Pantoea* emend. as *Pantoea ananas* (Serrano 1928) comb. nov. and *Pantoea stewartii* (Smith 1898) comb. nov., Respectively, and Description of *Pantoea stewartii* subsp. *indologenes* subsp.nov. *Int J Syst Bacteriol* **43,** 162-173

Mesbah, M., Premachandran, U. & Whitman, W.B. (1989). Precise measurement of the G + C content of deoxyribonucleic acid by high-performance liquid chromatography. *Int J Syst Bacteriol* **39,** 159-167

Naum, M., Brown, E.W. & Mason-Gamer, R.J. (2008). Is 16S rDNA a reliable phylogenetic marker to characterize relationships below the family level in the *Enterobacteriaceae*? *J Mol Evol* doi 10.1007/s00239-008-9115-3

Niemann, S., Pühler, 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. *J Appl Microbiol* **82,** 477-484

Paradis, S., Biossinot, M., Paquette, N., Bélanger, S.D., Martel, E.A., Boudreau, D.K., Picard, F.J., Ouellette, M., Roy, P.H. & Bergeron, M.G. (2005). Phylogeny of the *Enterobacteriaceae* based on genes encoding elongation factor Tu and F-ATPase β-subunit. *Int J Syst Evol Microbiol* **55,** 2013-2025

Posada, D. & Crandall, K.A. (1998). Modeltest: testing the model of DNA substitution. *Bioinformatics* **14,** 817-818

Swofford, D.L. (2000). PAUP*: Phylogenetic Analysis Using Parsimony and other methods (software). Sinauer Associates, Sunderland, MA.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. & Higgins, D.G. (1997). The ClustalX-Windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucl Acids Res* **25,** 4876-4882

Wilson, K. (1987). Preparation of genomic DNA from bacteria. In *Current Protocols in Molecular Biology*, pp. 2.4.1-2.4.5. Edited by F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith & K. Struhl. New York: Green Publishing and Wiley-Interscience

Table 1: Strains of *Pantoea* and *Tatumella* used in this study, LMG = BCCM/LMG Bacteria Collection, Ghent University, Belgium, ATCC = American Type Culture Collection, Rockville, Maryland, U.S.A., CCUG = Culture Collection, University of Göteborg, Sweden, CDC = Centres for Disease Control, Atlanta, Georgia, U.S.A.

^{T} = type strain

Species	**Strain**	**Host**	**Location**
Pantoea agglomerans	LMG 1286^{T}	Human	Zimbabwe
	LMG 2660	*Wisteria floribunda*	Japan
Pantoea ananatis	LMG 2665^{T}	Pineapple	Brazil
	LMG 20103	*Eucalyptus*	South Africa
Pantoea stewartii ssp. *stewartii*	LMG 2715^{T}	Maize	USA
	LMG 2718	Maize	USA
Pantoea stewartii ssp. *indologenes*	LMG 2632^{T}	Fox millet	India
	LMG 2673	Pineapple	Hawaii
Pantoea dispersa	LMG 2603^{T}	Soil	Japan
	LMG 2604	Wild rose	Netherlands
Pantoea anthophila	LMG 2558^{T}	*Impatiens balsamina*	India
	LMG 2560	*Tagetes erecta*	Unknown
Pantoea vagens	LMG 24199^{T}	*Eucalyptus*	Uganda
	LMG 24201	Maize	South Africa
Pantoea eucalypti	LMG 24198^{T}	*Eucalyptus*	Uruguay
	LMG 24197	*Eucalyptus*	Uruguay
Pantoea deleyii	LMG 24200^{T}	*Eucalyptus*	Uganda
Pantoea septica	LMG 5345^{T}	Human, stool	New Jersey, USA
	LMG 24526	Human, blood	New York, USA
Pantoea eucrina	LMG 2781^{T}	Human, trachea	Connecticut, USA
	LMG 24529	Human, cyst	Georgia, USA

Pantoea brenneri	LMG 5343^T	Human, urethra	Montana, USA
	LMG 24532	Human, sputum	Wisconsin, USA
Pantoea conspicua	LMG 24534^T	Human, blood	Paris, France
Tatumella citrea	LMG 22049^T = SHS 2003	Mandarin orange	Japan
Tatumella morbirosei	LMG 23360^T = NCPPB 4036^T	Pineapple	Philippines
	LMG 23359 = NCPPB 4035	Pineapple	Philippines
Tatumella punctata	LMG 22050^T = SHS 2006	Mandarin orange	Japan
	LMG 22097 = SHS 2004	Mandarin orange	Japan
	LMG 22098 = SHS 2005	Persimmon	Japan
	LMG 23562 = SHS 2004	Mandarin orange	Japan
	LMG 23563 = SHS 2007	Mandarin orange	Japan
	CCUG 30157 = SHS 2004	Mandarin orange	Japan
	CCUG 30160 = SHS 2007	Mandarin orange	Japan
Tatumella terrea	LMG 22051^T = SHS 2008	Soil	Japan
	LMG 23564 = SHS 2009	Soil	Japan
	CCUG 30162 = SHS 2009	Soil	Japan
Tatumella ptyseos	LMG 7888^T = ATCC 33301^T	Human	USA
	LMG 23565 = SHS 2010	Soil	Japan
	CCUG 30163 = SHS 2010	Soil	Japan

Table 2: DNA-DNA hybridization values between *T. citrea* comb. nov. (LMG 22049ᵀ) *T. morbirosei* sp. nov. (LMG 23360), *T. punctata* comb. nov. (LMG 22050ᵀ, LMG 22098), *T. terrea* comb. nov. (LMG 22051ᵀ, LMG 23564), *T. ptyseos* (LMG 7888ᵀ, LMG 23565), *P. agglomerans* (LMG 1286ᵀ), *P. ananatis* (LMG 2665ᵀ), *P. vagens* (LMG 24199ᵀ), *P. stewartii* ssp. *stewartii* (LMG 2715ᵀ) and *P. dispersa* (LMG 2603ᵀ).

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
1. LMG 22049ᵀ	100												
2. LMG 23360	42	100											
3. LMG 22050ᵀ	21	17	100										
4. LMG 22098	16		93	100									
5. LMG 22051ᵀ	13	11	17	18	100								
6. LMG 23564					82	100							
7. LMG 23565	15	13	20		55	54	100						
8. LMG 7888ᵀ	14	15	21		66		87	100					
9. LMG 1286ᵀ	6		8		8				100				
10. LMG 2665ᵀ	5		6		6				21	100			
11. LMG 24199ᵀ	5		7		8				65	20	100		
12. LMG 2715ᵀ	2		3		3				6	20	9	100	
13. LMG 2603ᵀ	7		7	7	8				24	20	19	22	100

Table 3: Phenotypic characteristics distinguishing *T.citrea* comb. nov., *T. morbirosei* sp. nov., *T. punctata* comb. nov. and *T.terrea* comb. nov. from each other and from *T. ptyseos*

1 = *T. citrea*, 2 = *T. morbirosei*, 3 = *T. punctata*, 4 = *T. terrea*, 5 = *T. ptyseos*

+, 90-100 % of strains positive in 1-2 days; (+), 90-100 % of strains positive in 1-4 days; -, 90-100 % of strains negative in 4 days

Characteristic	**1**	**2**	**3**	**4**	**5**
Phenylalanine deaminase	-	+	-	-	+
Arginine dihydrolase	+	+	+	-	(+)
Lactose	+	-	-	-	+
Gentiobiose	-	-	+	+	(+)
D-Xylose	+	-	+	+	+
L-Rhamnose	+	-	+	+	+
L-Tartrate	+	-	+	-	+
Trigonelline	-	+	-	-	-
D-Arabinose	-	+	+	-	-

Figure 1: Maximum likelihood tree based on complete 16S rRNA sequences of *Tatumella* and *Pantoea* species. The tree was generated by the Phyml software using the Tamura-Nei (TN93) model as selected by Modeltest. Bootstrap values after 1000 replicates are expressed as percentages. *Escherichia coli* was included as an outgroup.

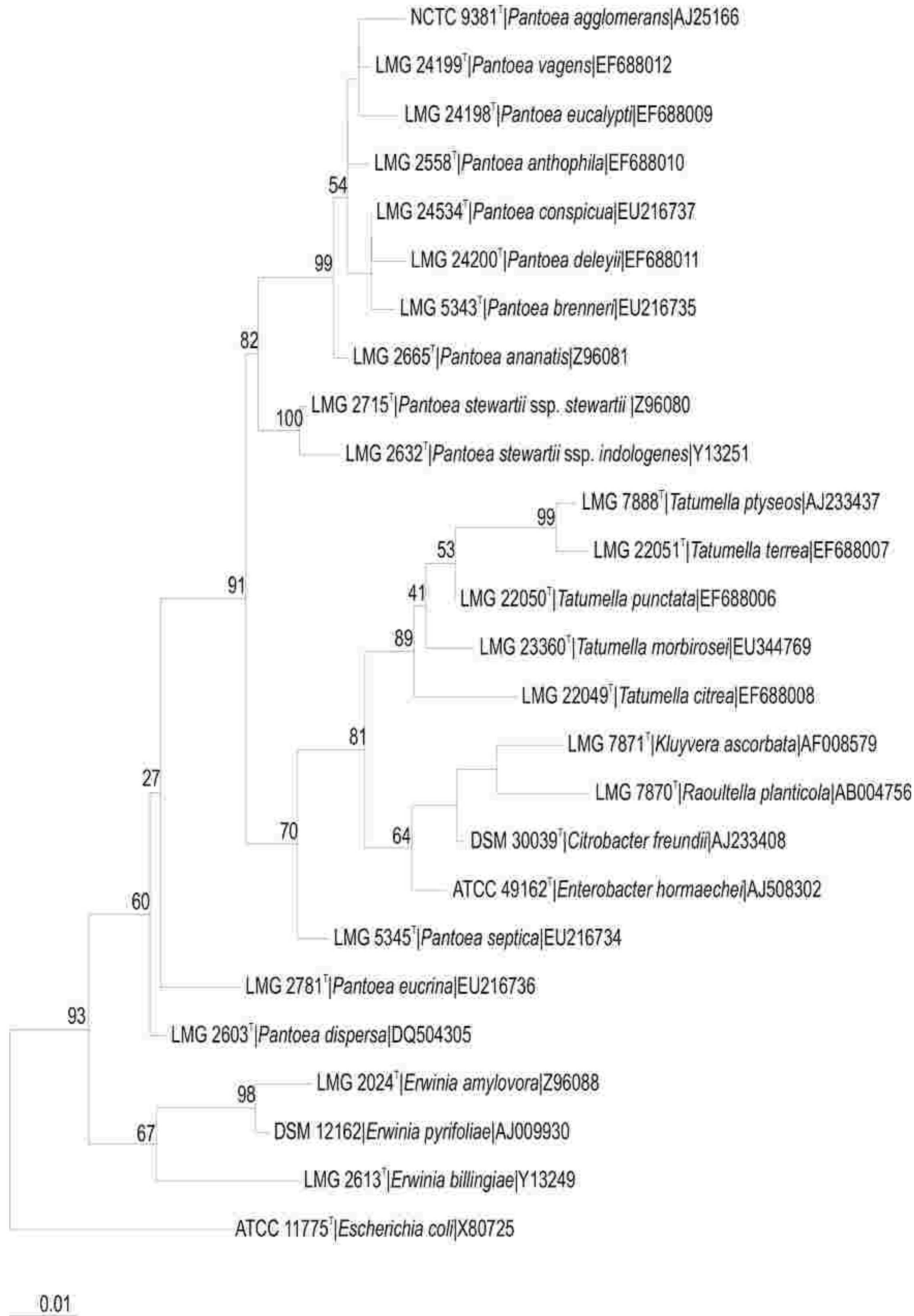
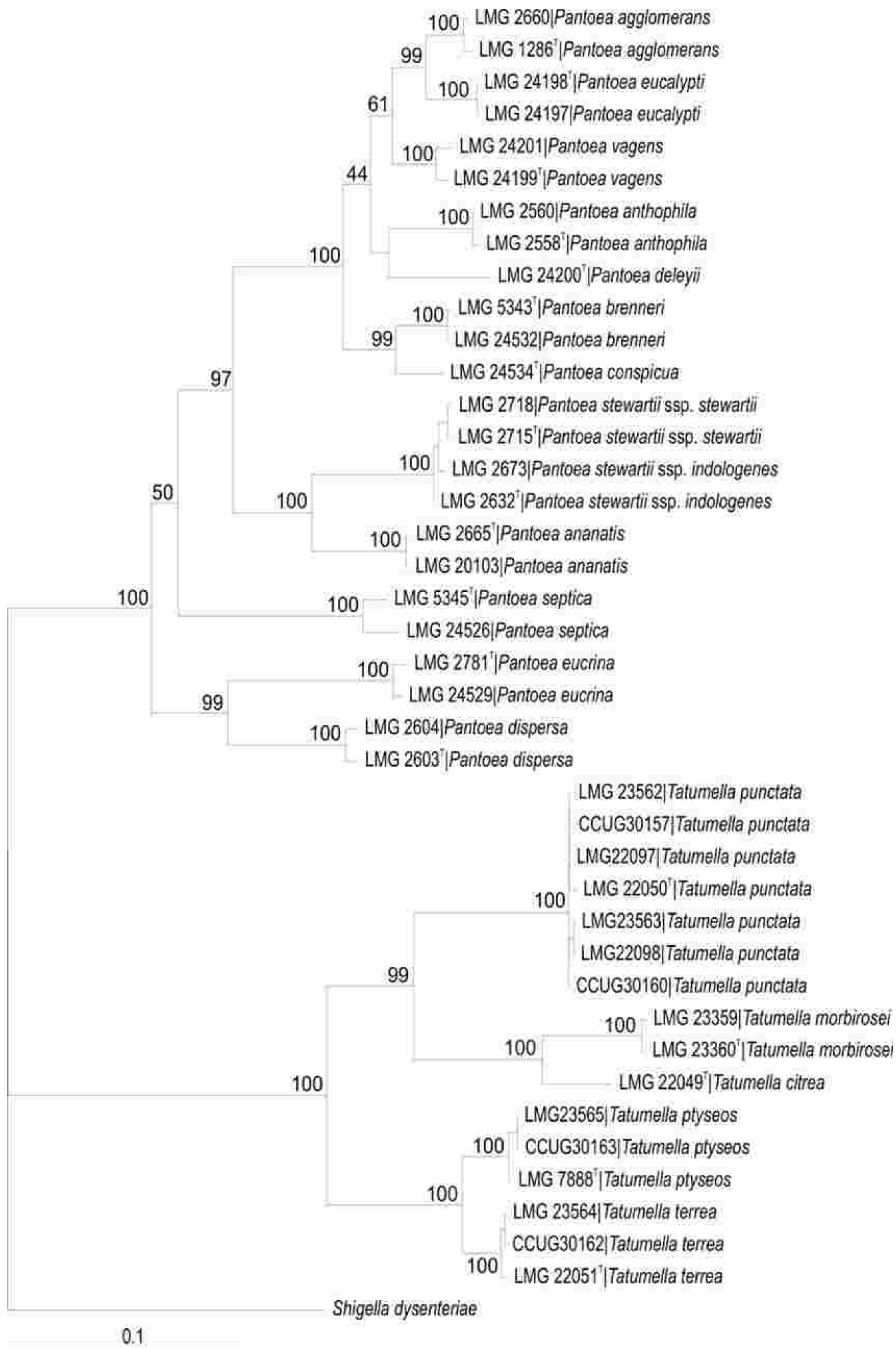

NCTC 9381T|Pantoea agglomerans|AJ25166
LMG 24199T|Pantoea vagens|EF688012
LMG 24198T|Pantoea eucalypti|EF688009
LMG 2558T|Pantoea anthophila|EF688010
LMG 24534T|Pantoea conspicua|EU216737
LMG 24200T|Pantoea deleyii|EF688011
LMG 5343T|Pantoea brenneri|EU216735
LMG 2665T|Pantoea ananatis|Z96081
LMG 2715T|Pantoea stewartii ssp. stewartii |Z96080
LMG 2632T|Pantoea stewartii ssp. indologenes|Y13251
LMG 7888T|Tatumella ptyseos|AJ233437
LMG 22051T|Tatumella terrea|EF688007
LMG 22050T|Tatumella punctata|EF688006
LMG 23360T|Tatumella morbirosei|EU344769
LMG 22049T|Tatumella citrea|EF688008
LMG 7871T|Kluyvera ascorbata|AF008579
LMG 7870T|Raoultella planticola|AB004756
DSM 30039T|Citrobacter freundii|AJ233408
ATCC 49162T|Enterobacter hormaechei|AJ508302
LMG 5345T|Pantoea septica|EU216734
LMG 2781T|Pantoea eucrina|EU216736
LMG 2603T|Pantoea dispersa|DQ504305
LMG 2024T|Erwinia amylovora|Z96088
DSM 12162|Erwinia pyrifoliae|AJ009930
LMG 2613T|Erwinia billingiae|Y13249
ATCC 11775T|Escherichia coli|X80725
54
99
82
100
99
53
41
89
91
81
64
70
27
60
93
98
67
0.01

Figure 2: Maximum likelihood tree based on concatenated partial sequences of *rpoB*, *atpD*, *gyrB* and *infB* of *Tatumella* and *Pantoea* strains. The tree was generated by the Phyml software using the general time reversible (GTR) model as selected by Modeltest. Bootstrap values after 1000 replicates are expressed as percentages. *Shigella dysenteriae* sequences were obtained from the genome sequencing database of the Sanger Institute (http://www.sanger.ac.uk) and included as an outgroup.


LMG 2660|*Pantoea agglomerans*
LMG 1286^{T}|*Pantoea agglomerans*
LMG 24198^{T}|*Pantoea eucalypti*
LMG 24197|*Pantoea eucalypti*
LMG 24201|*Pantoea vagens*
LMG 24199^{T}|*Pantoea vagens*
LMG 2560|*Pantoea anthophila*
LMG 2558^{T}|*Pantoea anthophila*
LMG 24200^{T}|*Pantoea deleyii*
LMG 5343^{T}|*Pantoea brenneri*
LMG 24532|*Pantoea brenneri*
LMG 24534^{T}|*Pantoea conspicua*
LMG 2718|*Pantoea stewartii* ssp. *stewartii*
LMG 2715^{T}|*Pantoea stewartii* ssp. *stewartii*
LMG 2673|*Pantoea stewartii* ssp. *indologenes*
LMG 2632^{T}|*Pantoea stewartii* ssp. *indologenes*
LMG 2665^{T}|*Pantoea ananatis*
LMG 20103|*Pantoea ananatis*
LMG 5345^{T}|*Pantoea septica*
LMG 24526|*Pantoea septica*
LMG 2781^{T}|*Pantoea eucrina*
LMG 24529|*Pantoea eucrina*
LMG 2604|*Pantoea dispersa*
LMG 2603^{T}|*Pantoea dispersa*
LMG 23562|*Tatumella punctata*
CCUG30157|*Tatumella punctata*
LMG22097|*Tatumella punctata*
LMG 22050^{T}|*Tatumella punctata*
LMG23563|*Tatumella punctata*
LMG22098|*Tatumella punctata*
CCUG30160|*Tatumella punctata*
LMG 23359|*Tatumella morbirosei*
LMG 23360^{T}|*Tatumella morbirosei*
LMG 22049^{T}|*Tatumella citrea*
LMG23565|*Tatumella ptyseos*
CCUG30163|*Tatumella ptyseos*
LMG 7888^{T}|*Tatumella ptyseos*
LMG 23564|*Tatumella terrea*
CCUG30162|*Tatumella terrea*
LMG 22051^{T}|*Tatumella terrea*
Shigella dysenteriae
0.1

CHAPTER 7

Isolation of *Enterobacter cowanii* from *Eucalyptus* showing symptoms of bacterial blight and dieback in Uruguay

As prepared for: Letters in Applied Microbiology

ABSTRACT

Aims: This study was performed to identify bacterial strains isolated simultaneously with *Pantoea* species from *Eucalyptus* trees showing symptoms of bacterial blight and dieback in Uruguay.

Methods and Results: Several molecular techniques including 16S rRNA- and *rpoB*-gene sequencing and DNA-DNA hybridization were used to characterize the Gram-negative, facultatively-anaerobic, slime-producing bacterial strains isolated along with *Pantoea* species from *Eucalyptus*. Hypersensitivity reactions and pathogenicity tests were performed on tobacco and *Eucalyptus* seedlings, respectively. The isolates clustered closely with the type strain of *E. cowanii* in both phylogenetic trees constructed. The DNA-DNA similarity between the isolates and the type strain of *E. cowanii* ranged from 88-92 %. A positive hypersensitivity reaction was observed on the tobacco seedlings, but no disease symptoms were visible on the inoculated *Eucalyptus* seedlings.

Conclusions: *E. cowanii* was isolated from trees with symptoms of bacterial blight, although strains of this bacterial species do not appear to be the causal agent of the disease.

Significance and Impact of Study: This study provides the first report of *E. cowanii* isolated from *Eucalyptus*. Its presence in *Eucalyptus* tissue suggests that it is an endophyte in trees showing symptoms of blight.

INTRODUCTION

Eucalyptus grandis trees in Uruguay commonly exhibit symptoms of leaf blight and dieback disease. The cause of this disease is unknown, but symptoms such as water-soaked lesions with a greasy appearance are typical of bacterial infections. In 2002, symptomatic leaves and shoots were collected and isolations were made from the infected tissue. Gram-negative, facultatively-anaerobic bacteria were consistently isolated from the diseased material. The majority of the strains were yellow-pigmented and were thought to belong to *P. ananatis*, the causal agent of bacterial blight and dieback on *Eucalyptus* in South Africa (Coutinho *et al*., 2002). Several non-pigmented, slime-producing strains were also isolated from the diseased material. The yellow-pigmented strains were subsequently identified as representing three novel species belonging to the genus *Pantoea* using multilocus sequence analysis (MLSA) based on *rpoB*, *atpD*, *gyrB* and *infB* gene sequences as a supporting technique (Brady *et al*., submitted). The aim of this study was to identify the non-pigmented, slime-producing bacterial strains isolated together with *Pantoea* species from *E. grandi*s leaves and shoots in Uruguay. The strains were identified using 16S rRNA- and *rpoB* gene-sequence comparisons as well as DNA-DNA hybridization. In addition pathogenicity tests were performed on *Eucalyptus* seedlings to consider their possible role in causing disease, as it was not clear whether the *Pantoea* strains or the non-pigmented strains were responsible for the leaf- and shoot-blight symptoms observed.

MATERIALS AND METHODS

Bacterial strains and DNA extraction

Five non-pigmented slime-producing strains were isolated from *Eucalyptus* leaves showing typical bacterial blight symptoms including leaf spots and water-soaked lesions. The leaves were surface-disinfected, crushed in sterile water and the resulting suspension was streaked on nutrient agar and incubated at 28 °C for three days. Single colonies were obtained by re-streaking and incubation under the same conditions. Genomic DNA was extracted from each of the bacterial strains using an alkalic extraction method (Niemann *et al.*, 1997) and stored at -20 °C. Strains used in this study are listed in Table 1 and are maintained at the Forestry and Agricultural Biotechnology Institute (FABI).

16S rRNA and *rpoB* gene sequencing and analysis

The complete 16S rRNA sequence was determined for two representative strains using the primers and conditions determined by Coenye *et al.* (1999). The representative strains were selected from clusters (data not shown) generated by an AFLP technique developed for the genus *Pantoea* (Brady *et al.*, 2007). *rpoB* gene sequencing was performed on all five strains using the method described in the MLSA scheme developed for the genus *Pantoea* (Brady *et al.*, submitted). The sequences were aligned using ClustalX (Thompson *et al.*, 1997) and the overhangs trimmed. The Modeltest 3.7 programme (Posada and Crandall, 1998) was then applied to the data sets to determine the best-fit evolutionary model to apply to each gene. Maximum likelihood analysis was performed using Phyml (Guindon and Gascuel, 2003), by applying the models and parameters determined by Modeltest. Bootstrap analysis with 1000 replicates was performed on the trees to assess the reliability of the clusters.

DNA-DNA hybridization and G + C content

High quality DNA for DNA-DNA hybridization of strains was prepared using the method of Wilson (1987), with minor modifications (Cleenwerck *et al.*, 2002). DNA-DNA hybridizations were performed using the microplate method (Ezaki *et al.*, 1989) with some modifications (Cleenwerck *et al.*, 2002). The hybridization temperature was 45 °C ± 1 °C and reciprocal reactions were performed with DNA from all strains. The type strain of *Enterobacter cowanii* (LMG 23569^{T}) was hybridized to BCC 009 and BCC 078, and BCC 009, BCC 011 and BCC 078 were hybridized amongst each other. The G + C contents of the strains were determined by HPLC as published by Mesbah *et al.* (1989).

Pathogenicity tests

Hypersensitivity reaction (HR) tests were conducted on four tobacco seedlings (*Nicotiana tabacum*) by injecting a bacterial suspension of 10^{8} CFU/ml of strains BCC 008, BCC 011, BCC 074 and BCC 078 into the intercellular spaces of the leaves with a fine needle and syringe. Pathogenicity tests were performed on 12 cuttings of a susceptible *E. grandis* x *E. nitens* clone as previously described (Coutinho *et al*., 2002). Seedling leaves were inoculated with sterile water as a negative control and with LMG 20103 (*Pantoea ananatis* pathogenic on *Eucalyptus*) as a positive control. The seedlings were covered with plastic bags to induce humid conditions and incubated for two weeks. Cutting were assessed by using a 0 to 3 scale (0 = no disease, 3 = lesion larger than 1 cm).

RESULTS

Sequence analyses

The GenBank/EMBL accession numbers for the 16S rRNA gene sequences for BCC 009 and BCC 078 are EU629163 and EU629164, respectively; and EU629165-EU629169 for the *rpoB* genes for strains BCC 008, BCC 009, BCC 011, BCC 074 and BCC 078. The 16S rRNA sequences of strains BCC 009 and BCC 078 was greater than 99.8 % similar to *E. cowanii* and greater than 98 % similar to *E. cloacae*, *E. radicincitans*, *E. asburiae* and *E. cancerogenus*. In the 16S rRNA phylogenetic tree (Fig. 1), BCC 009 and BCC 078 clustered with the type strain of *E. cowanii* with a strong bootstrap support of 79 %. All five strains clustered closely with the type strain of *E. cowanii* in the *rpoB* phylogenetic tree with high bootstrap support of 100 % (Fig. 2). The topologies of the 16S rRNA- and *rpoB*- trees were similar to those of Stephan *et al.* (2007).

DNA-DNA hybridization and G + C content

When hybridized to BCC 009 and BCC 078, the type strain of *E. cowanii* (LMG 23569^T) exhibited 92 % and 88 % DNA similarity, respectively. The DNA similarity amongst strains BCC 009, BCC 011 and BCC 078 ranged from 76 to 92 %. The G + C contents for strains BCC 008, BCC 009, BCC 011, BCC 074 and BCC 078 ranged from 55.8 to 56.6 mol %, which is similar to the 53 mol % of the type strain of *E. cowanii*, LMG 23569^T published by Grimont and Grimont (2005a).

Pathogenicity tests

A positive hypersensitivity reaction was observed on the tobacco seedlings inoculated with strains BCC 008, BCC 011, BCC 074 and BCC 078, which was demonstrated by the complete collapse of the leaf tissue after 24 hours. The inoculated *Eucalyptus* seedlings and the negative water control displayed no symptoms during the two weeks in which the leaves were examined (score obtained = 0). In contrast, the leaves inoculated with *P. ananatis* developed necrotic lesions within five days of inoculation (score obtained = 3).

DISCUSSION

Results of this study demonstrated that the non-pigmented, slime-producing bacterial strains isolated from the internal parts of *Eucalyptus* tissue together with *Pantoea* species, are *E. cowanii*. This identification was clear from the phylogenetic trees based on 16S rRNA- and *rpoB*-gene sequence comparisons, both of which were strongly supported by high bootstraps.The sequencing results were confirmed by DNA-DNA hybridization data and G + C content of the strains. The DNA-DNA similarity values of BCC 009 and BCC 078 with LMG 23569^T, the type strain of *E. cowanii* and between strains BCC 009, BCC 011 and BCC 078, are both well above the recommended species definition cut-off of 70 % (Wayne *et al.*, 1987). Additionally, the G + C contents of the strains are within the 5 mol % difference range for species delineation (Rosselló-Mora and Amann, 2001).

Species belonging to the genus *Enterobacter* are typically associated with the environment. Some cause diseases of trees, while others are opportunistic human pathogens (Grimont and Grimont, 2006). The occurrence of *E. cowanii* on *Eucalyptus* in this study was interesting as the bacterium has never previously been isolated from this tree, where it evidently can live internally in leaf tissue. *Enterobacter cowanii* was described for a group of clinical strains which were previously identified as *E. agglomerans* (*Pantoea agglomerans*) in routine diagnostic laboratories (Inoue *et al*., 2000). Of the 15 recognized species belonging to *Enterobacter*, only three have previously been isolated from diseased trees, *E. cancerogenus, E. nimipressuralis* and *E. pyrinus* (Grimont and Grimont, 2005a). In this regard, the occurrence of the bacterium on *Eucalyptus* is perhaps not unusual.

The pathogenicity tests indicated that *E. cowanii* is unlikely to have played a part in the bacterial blight and dieback of *Eucalyptus* in Uruguay, although the hypersensitivity reactions on tobacco seedlings were positive. It is possible that *E. cowanii* can contribute to disease under certain environmental conditions favourable to the bacterium. However, it is likely that *E. cowanii* isolated in this study is an endophyte that was coincidentally isolated together with *Pantoea* species which may cause the leaf symptoms observed.

The view that *Enterobacter cowanii* isolated in this study is most likely an endophyte of the *Eucalyptus* tissue from which it was isolated in this study, is consistent with the ecology of other *Enterobacteriaceae*. For example, *E. cloacae* has been found as an endophytic symbiont of corn (Hinton and Bacon, 1995) and papaya (Pious *et al*., 2007) and as an obligatory endophyte of Mediterranean pines (Madmony *et al*., 2005). *E. asburiae* is a well known endophyte (Quadt-Hallmann *et al*., 1997) and *E. gergoviae* is an opportunistic endophyte of maize (An *et al*., 2007). These three *Enterobacter* species are known for their regular isolation from clinical samples, in addition to *E. cloacae* causing nosocomial infections (Grimont and Grimont, 2006). There are other examples within the family *Enterobacteriaceae* of species causing human disease or being isolated from clinical samples but also existing as phytopathogens or endophytes, especially for species residing in the genus *Pantoea*. *P. agglomerans* is considered a rare opportunistic pathogen but also causes disease on plant hosts and *P. ananatis* causes a range of plant and agricultural diseases but has also been isolated from septic patients (Grimont and Grimont, 2005b). This emphasizes the ubiquitous nature of both *Enterobacter* and *Pantoea* species.

This study represents the first report of *E. cowanii* isolated from diseased *Eucalyptus*. The bacterium does not appear to be involved in the bacterial blight and dieback in Uruguay. Nonetheless, its occurrence in *Eucalyptus* tissue is intriguing and it will be interesting to know whether it is present in this plantation tree elsewhere in South America, as well as other parts of the world.

ACKNOWLEDGEMENTS

This study was partially supported by the South African-Flemish Bilateral Agreement, the National Research Foundation (NRF), the Tree Protection Co-operative Programme (TPCP) and the THRIP support programme of the Department of Trade and Industry, South Africa. The BCCM/LMG Bacteria collection is supported by the Federal Public Planning Service-Science Policy, Belgium. The authors wish to acknowledge Katrien Vandemeulebroecke for technical assistance and Mike Wingfield for the collection of the diseased plant material.

REFERENCES

An, Q., Dong, Y., Wang, W., Li, Y. and Li, J. (2007) Constitutive expression of the *nifA* gene activates associative nitrogen fixation of *Enterobacter gergoviae* 57-7, an opportunistic endophytic diazotroph. *J Appl Microbiol* **103,** 613-620.

Brady, C., Venter, S., Cleenwerck, I., Vancanneyt, M., Swings, J. and Coutinho, T. (2007) A FAFLP system for the improved identification of plant-pathogenic and plant-associated species of the genus *Pantoea*. *Syst Appl Microbiol* **30,** 413-417.

Brady, C.L., Cleenwerck, I.,Venter, S.N., Vancanneyt, M., Swings, J. and Coutinho, T.A. (2008) Phylogeny and identification of *Pantoea* species associated with the environment, humans and plants based on multilocus sequence analysis (MLSA). *Submitted to Appl Environ Microbiol*

Brady, C.L., Venter, S.N., Cleenwerck, I., Engelbeen, K., Vancanneyt, M., Swings, J. and Coutinho, T.A. (2008) *Pantoea vagens* sp. nov., *Pantoea eucalypti* sp. nov., *Pantoea deleyii* sp. nov. and *Pantoea anthophila* sp. nov., four novel species belonging to the Genus *Pantoea*. Submitted to *Int J Syst Bacteriol*

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

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

Coutinho, T.A., Preisig, O., Mergaert, J., Cnockaert, M.C., Riedel, K.-H., Swings, J. and Wingfield, M.J. (2002) Bacterial blight and dieback of *Eucalyptus* species, hybrids, and clones in South Africa. *Plant Dis* **86,** 20-25.

Ezaki, T., Hashimoto, Y. and Yabuuchi, E. (1989) Fluorometric deoxyribonucleic acid-deoxyribonucleic acid hybridization in micro-dilution wells as an alternative to membrane filter hybridization in which radioisotopes are used to determine genetic relatedness among bacterial strains. *Int J Syst Bacteriol* **39,** 224-229.

Grimont, F. and Grimont, P.A.D. (2006) The Genus *Enterobacter* In *The Prokaryotes: Proteobacteria: Gamma Subclass*, eds. M. Dworkin, S. Falkow, E. Rosenberg, K.H. Schleifer and E. Stackebrandt, 3rd edn, Springer, New York, 197-214.

Grimont, P.A.D. and Grimont, F. (2005a) Genus: *Enterobacter* In Volume Two: The *Proteobacteria*, Part B: The *Gammaproteobacteria* In *Bergey's Manual of Systematic Bacteriology*, eds. D.J. Brenner, N.R. Krieg and J.T. Staley, 2nd edn, Springer, New York, 661-669.

Grimont, P.A.D. and Grimont, F. (2005b) Genus: *Pantoea* In Volume Two: The *Proteobacteria*, Part B: The *Gammaproteobacteria* In *Bergey's Manual of Systematic Bacteriology*, eds. D.J. Brenner, N.R. Krieg and J.T. Staley, 2nd edn, Springer, New York, 713-720.

Guidon, S. and Gascual, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52,** 696-704.

Hinton, D.M. and Bacon, C.W. (1995) *Enterobacter cloacae* is an endophytic symbiont of corn. *Mycopathologia* **129,** 117-125.

Inoue, K., Sugiyama, K., Kosako, Y., Sakazaki, R. and Yamai, S. (2000) *Enterobacter cowanii* sp. nov., a new species of the family *Enterobacteriaceae*. *Curr Microbiol* **41,** 417-420.

Kämpfer, P., Ruppel, S. and Remus, R. (2005) *Enterobacter radicincitans* sp. nov., a plant growth promoting species of the family *Enterobacteriaceae*. *Syst Appl Microbiol* **28,** 213-221.

Madmony, A., Chernin, L., Pleban, S., Peleg, E. and Riov, J. (2005) *Enterobacter cloacae*, an obligatory endophyte of pollen grains of Mediterranean Pines. *Folia microbiol* **50,** 209-216.

Mesbah, M., Premachandran, U. and Whitman, W.B. (1989) Precise measurement of the G + C content of deoxyribonucleic acid by high-performance liquid chromatography. *Int J Syst Bacteriol* **39,** 159-167.

Niemann, S., Pühler, A., Tichy, H.-V., Simon, R. and Selbitschka, W. (1997) Evaluation of the resolving power of three different DNA fingerprinting methods to discriminate among isolates of a natural *Rhizobium meliloti* population. *J Appl Microbiol* **82,** 477-484.

Posada, D. and Crandall, K.A. (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14,** 817-818.

Quadt-Hallmann, A., Hallmann, J. and Kloepper, J.W. (1997) Bacterial endophytes in cotton: Location and interaction with other plant-associated bacteria. *Can J Microbiol* **43,** 254-259.

Rosselló-Mora, R. and Amann, R. (2001) The species concept for prokaryotes. *FEMS Microbiol Rev* **25,** 39-67.

Stephan, R., van Trappen, S., Cleenwerck, I., Vancanneyt, M., de Vos, P. and Lehner, A. (2007) *Enterobacter turicensis* sp. nov. and *Enterobacter helveticus* sp. nov., isolated from fruit powder. *Int J Syst Evol Microbiol* **57,** 820-826.

Swofford, D.L. (2000) PAUP*: Phylogenetic Analysis Using Parsimony and other methods, Sinauer Associates, Sunderland, MA.

Thomas, P., Kumari, S., Swarna, G.K. and Gowda, T.K.S. (2007) Papaya shoot tip associated endophytic bacteria isolated from in vitro cultures and host-endophyte interaction in vitro and in vivo. *Can J Microbiol* **53,** 380-390.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997) The ClustalX-Windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucl Acids Res* **25,** 4876-4882.

Wayne, L.G., Brenner, D.J., Colwell, R.R., and 9 other authors. (1987) Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int J Syst Bacteriol* **37,** 463-464.

Wilson, K. (1987) Preparation of genomic DNA from bacteria. In *Current Protocols in Molecular Biology*, eds. F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith and K. Struhl. Green Publishing and Wiley-Interscience, New York, 2.4.1-2.4.5

Table 1: Strains of *Enterobacter cowanii* included in this study, LMG = BCCM/LMG Bacteria Collection, Ghent University, BCC = Bacterial Culture Collection, Forestry and Agricultural Biotechnology Institute, Pretoria, South Africa. ^{T} = type strain

Species	Strain	Host	Location
Enterobacter cowanii	LMG 23569^T = CCUG 45998^T = CIP 107300^T	Blood culture	Japan
	BCC 008 = R-25669	*Eucalyptus*	Uruguay
	BCC 009 = R-25670	*Eucalyptus*	Uruguay
	BCC 011 = R-25672	*Eucalyptus*	Uruguay
	BCC 074 = R-21554	*Eucalyptus*	Uruguay
	BCC 078 = R-25680	*Eucalyptus*	Uruguay

Figure 1: Maximum likelihood tree based on complete 16S rRNA sequences of *Enterobacteriacae* species. The tree was generated by the Phyml software using the general time reversible (GTR) model as selected by Modeltest. Bootstrap values after 1000 replicates are expressed as percentages. *Serratia marcescens* was included as an outgroup.

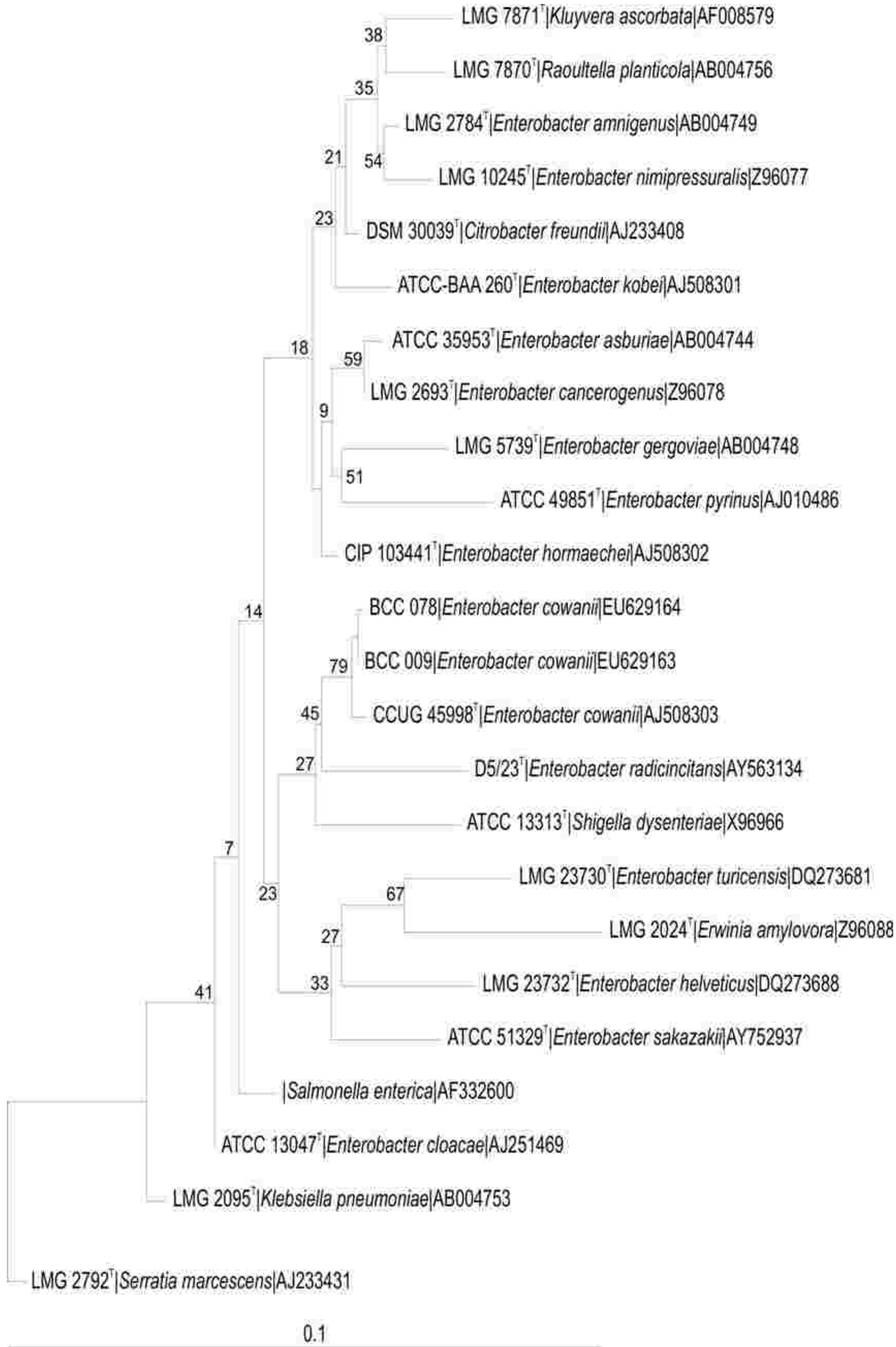
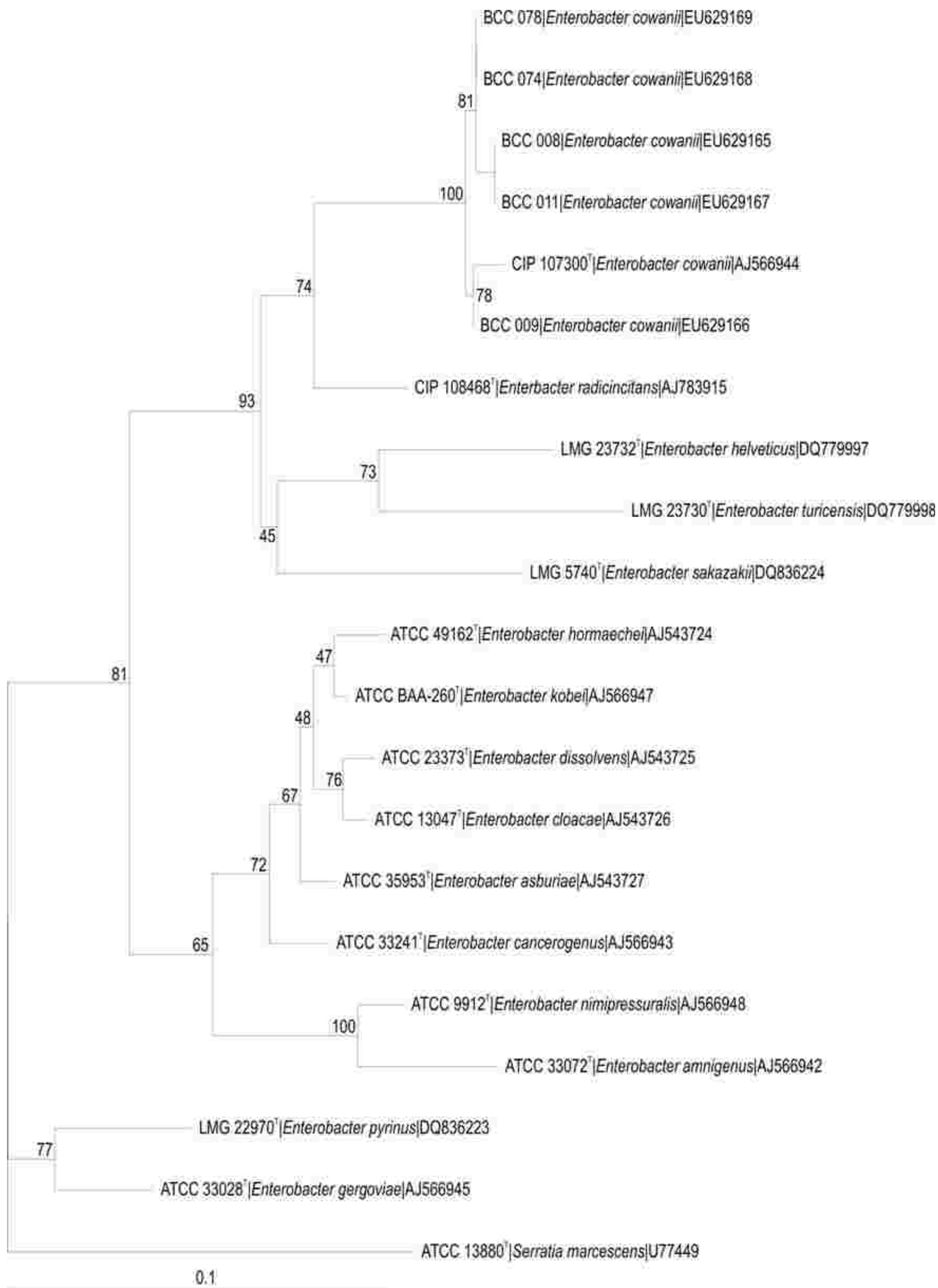

LMG 7871^{T}|Kluyvera ascorbata|AF008579
LMG 7870^{T}|Raoultella planticola|AB004756
LMG 2784^{T}|Enterobacter amnigenus|AB004749
LMG 10245^{T}|Enterobacter nimipressuralis|Z96077
DSM 30039^{T}|Citrobacter freundii|AJ233408
ATCC-BAA 260^{T}|Enterobacter kobei|AJ508301
ATCC 35953^{T}|Enterobacter asburiae|AB004744
LMG 2693^{T}|Enterobacter cancerogenus|Z96078
LMG 5739^{T}|Enterobacter gergoviae|AB004748
ATCC 49851^{T}|Enterobacter pyrinus|AJ010486
CIP 103441^{T}|Enterobacter hormaechei|AJ508302
BCC 078|Enterobacter cowanii|EU629164
BCC 009|Enterobacter cowanii|EU629163
CCUG 45998^{T}|Enterobacter cowanii|AJ508303
D5/23^{T}|Enterobacter radicincitans|AY563134
ATCC 13313^{T}|Shigella dysenteriae|X96966
LMG 23730^{T}|Enterobacter turicensis|DQ273681
LMG 2024^{T}|Erwinia amylovora|Z96088
LMG 23732^{T}|Enterobacter helveticus|DQ273688
ATCC 51329^{T}|Enterobacter sakazakii|AY752937
|Salmonella enterica|AF332600
ATCC 13047^{T}|Enterobacter cloacae|AJ251469
LMG 2095^{T}|Klebsiella pneumoniae|AB004753
LMG 2792^{T}|Serratia marcescens|AJ233431
38
35
54
21
23
18
59
9
51
14
79
45
27
7
23
67
27
33
41
0.1

Figure 2: Maximum likelihood tree based on partial *rpoB* sequences of *Enterobacter* strains. The tree was generated by the Phyml software using the Tamura-Nei (TN93) model as selected by Modeltest. Bootstrap values after 1000 replicates are expressed as percentages. *Serratia marcescens* was included as an outgroup.


BCC 078|Enterobacter cowanii|EU629169
BCC 074|Enterobacter cowanii|EU629168
81
BCC 008|Enterobacter cowanii|EU629165
100
BCC 011|Enterobacter cowanii|EU629167
CIP 107300T|Enterobacter cowanii|AJ566944
74
78
BCC 009|Enterobacter cowanii|EU629166
CIP 108468T|Enterbacter radicincitans|AJ783915
93
LMG 23732T|Enterobacter helveticus|DQ779997
73
LMG 23730T|Enterobacter turicensis|DQ779998
45
LMG 5740T|Enterobacter sakazakii|DQ836224
ATCC 49162T|Enterobacter hormaechei|AJ543724
47
81
ATCC BAA-260T|Enterobacter kobei|AJ566947
48
ATCC 23373T|Enterobacter dissolvens|AJ543725
76
67
ATCC 13047T|Enterobacter cloacae|AJ543726
72
ATCC 35953T|Enterobacter asburiae|AJ543727
65
ATCC 33241T|Enterobacter cancerogenus|AJ566943
ATCC 9912T|Enterobacter nimipressuralis|AJ566948
100
ATCC 33072T|Enterobacter amnigenus|AJ566942
LMG 22970T|Enterobacter pyrinus|DQ836223
77
ATCC 33028T|Enterobacter gergoviae|AJ566945
ATCC 13880T|Serratia marcescens|U77449
0.1

Conclusions

Taxonomy and characterization of species belonging to the former *Erwinia herbicola-Enterobacter agglomerans* complex has become difficult due to the continuous rearrangement of species within the predominantly plant-pathogenic genera *Erwinia, Enterobacter*, *Pantoea*, *Pectobacterium*, *Brenneria* and *Dickeya*. The taxonomic issues within these genera have resulted in difficulties with identification and classification and consequently numerous mis-identified and unidentified strains, especially in the genus *Pantoea*. There are very few unique phenotypic differences between genera of the plant-pathogenic *Enterobacteriaceae* and the species belonging to these genera share exceedingly high 16S rRNA sequence similarity. Therefore, a technique was required which could differentiate between *Pantoea* species and closely-related *Enterobacteriaceae*, be used for rapid identification of *Pantoea* strains and hopefully resolve several taxonomic issues within the genus.

Multilocus sequence analysis (MLSA) was selected as the technique of choice for a taxonomic evaluation of the genus *Pantoea*. An MLSA scheme was developed based on the four housekeeping genes *rpoB*, *atpD*, *gyrB* and *infB*. These four housekeeping genes were found to be reliable genetic markers for the identification and classification of *Pantoea* species. In all phylogenetic trees constructed, the seven species of *Pantoea* could be clearly delineated and ten potential new species identified. Furthermore, the MLSA scheme revealed a phylogenetic division of the genus *Pantoea* into the core species and the "Japanese" species, namely *P. citrea*, *P. punctata* and *P. terrea*, and improved our understanding of the relationships of *Pantoea* with its closest phylogenetic neighbours. MLSA has demonstrated the ability to overcome the inconsistencies observed in 16S rRNA phylogeny. In this study, a concatenated data set based on several conserved protein-encoding genes has proven to be a robust and reliable means to resolve taxonomically complex groups of bacteria.

One of the primary objectives of this study was to conclusively identify *Pantoea*-like strains isolated from *Eucalyptus* trees showing symptoms of bacterial blight and dieback in Argentina, Colombia, Uruguay and Uganda and from maize infected with brown stalk rot in South Africa. In a previous study, based on AFLP analysis, it was observed that these strains belonged to the genus *Pantoea*. However, the AFLP clusters observed were not consistent with 16S rRNA sequencing data. A phylogenetic tree based on the concatenated sequences of the four housekeeping genes used in the MLSA scheme clearly differentiated between three well-supported clusters of strains from *Eucalyptus* and maize, and the *Pantoea* core species of *P. agglomerans*, *P. ananatis*, *P. stewartii* and *P. dispersa*. A high correlation was observed between the MLSA data and the DNA-DNA hybridization data of the three clusters. Using MLSA as a supporting technique, four novel species were proposed: *P. vagens* sp. nov. for strains isolated from *Eucalyptus* and maize, *P. eucalypti* sp. nov. and *P. deleyii* sp. nov. for strains isolated only from *Eucalyptus* and *P. anthophila* sp. nov. for strains belonging to protein profile group VII (Beji *et al*., 1988) which were previously allocated to *P. agglomerans*.

The latest edition of Bergey's Manual of Systematic Bacteriology brought attention to four DNA hybridization groups of clinical strains belonging to the former *Erwinia herbicola-Enterobacter agglomerans* complex, which were never assigned to a genus nor described as novel species (Grimont and Grimont, 2005). It was proposed that DNA hybridization groups I, II, IV and V (Brenner *et al*., 1984) should be transferred to the genus *Pantoea* as novel species. Based on this suggestion, strains from DNA hybridization groups II, IV and V were included in the MLSA scheme. Three well-supported clusters grouped closely with the *Pantoea* core species in a concatenated phylogenetic tree, correlating to the three DNA hybridization groups. Additionally, it was noted that the cluster corresponding to DNA hybridization group V contained two sublineages. DNA-DNA hybridization was used to confirm the existence of two novel species within DNA hybridization group V. Consequently, four novel clinical *Pantoea* species were proposed: *P. septica* sp. nov., *P. eucrina* sp. nov., *P. brenneri* sp. nov. and *P. conspicua* sp. nov.

It has been stated that the genus *Pantoea* can be divided into two groups of species: the core group of *P. agglomerans*, *P. ananatis*, *P. stewartii* and *P. dispersa* and the

"Japanese" group of *P. citrea*, *P. punctata* and *P. terrea* (Grimont and Grimont, 2005). This phylogenetic division of *Pantoea* was also observed in the concatenated MLSA tree (Chapter 3). The "Japanese" *Pantoea* species are morphologically and metabolically different to the *Pantoea* core species, and were described based on phenotypic tests and DNA-DNA hybridization data. In the concatenated MLSA tree, the "Japanese" *Pantoea* species formed a tight, well-supported cluster with the type strain of *Tatumella ptyseos* suggesting an affiliation of these species. DNA-DNA hybridization confirmed the close phylogenetic relationship of *T. ptyseos* to the "Japanese" species. The DNA-DNA similarity values between *T. ptyseos* and the "Japanese" species were considerably higher than those between the "Japanese" species and *P. agglomerans* or *P. dispersa*. The "Japanese" species were also found to be more metabolically similar to *Tatumella* than to *Pantoea*. Therefore, it was proposed to transfer the "Japanese" *Pantoea* species to the genus *Tatumella* emended as *Tatumella citrea* comb. nov., *Tatumella punctata* comb. nov. and *Tatumella terrea* comb. nov. The MLSA data also revealed the presence of two species within the *T. citrea* cluster: the *T. citrea* type strain and strains causing pink disease of pineapple. DNA-DNA hybridization validated the existence of two species within *T. citrea*. The strains causing pink disease of pineapple were described as a novel species, *Tatumella morbirosei* sp. nov.

Several non-pigmented, slime-producing strains were isolated simultaneously with *Pantoea* strains from *Eucalyptus* showing symptoms of bacterial blight in Uruguay. The *Pantoea* strains were later proposed as three novel species (Chapter 4). The non-pigmented, slime-producing strains showed highest 16S rRNA sequence similarity to *Enterobacter cowanii*, a primarily clinical species. *rpoB* sequencing placed the strains in a well-supported cluster with the type strain of *E. cowanii* and DNA-DNA hybridization was used to confirm their identity. Pathogenicity tests revealed that *E. cowanii* is not the causal agent of bacterial blight in Uruguay. This is the first report of *E. cowanii* being isolated as an endophyte from *Eucalyptus*.

The MLSA scheme developed for *Pantoea* species raised some questions regarding the relationship of the genus to phylogenetically-related members of the *Enterobacteriaceae*. The genus *Pantoea* appears to be more closely-related to the genera *Tatumella* and *Erwinia* than to the remaining plant-pathogenic genera

Pectobacterium, *Brenneria* and *Dickeya*. However, is has been observed in a previous study that the plant pathogen *Pectobacterium cypripedii* may be associated with the genus *Pantoea* (Young and Park, 2007). The same observation was made in this study when additional phytopathogenic enterobacteria were added to the phylogenetic trees (data not shown). Strains of *Pe. cypripedii* clustered on the border of the genus *Pantoea* with *P. dispersa* and *P. eucrina* in the MLSA phylogenetic trees, distant to the remaining species of the genus *Pectobacterium*. It was also observed in the MLSA trees, that three *Erwinia* species (*E. mallotivora*, *E. psidii* and *E. tracheiphila*) clustered slightly distant to the type species, *E. amylovora* (data not shown). To further investigate the relationship between *Pe. cypripedii* and *Pantoea*, and the possible division within the genus *Erwinia*, more strains should be added to the MLSA scheme to generate stable clusters.

There are still numerous misidentified or unidentified *Pantoea* strains in culture collections and consequently wrongly named sequences in comparison databases. These strains and sequences should classified correctly using MLSA, in order to further improve the rapid identification of environmental *Pantoea* species. Because the genus *Pantoea* is ubiquitous, it is highly likely that many more novel species exist in nature. The sampling strategy should be expanded to include a wider host range and geographical area which should ensure the isolation of *Pantoea* species. The more species and strains which can be added to the MLSA scheme, the clearer the phylogeny of the genus *Pantoea* within the *Enterobacteriaceae* will be.