International Journal of Systematic and Evolutionary Microbiology (2007), **57**, 2777–2789 **DOI** 10.1099/ijs.0.64711-0

Identification of lactobacilli by *pheS* and *rpoA* gene sequence analyses

Sabri M. Naser,[1] Peter Dawyndt,[2,4] Bart Hoste,[3] Dirk Gevers,[2,5] Katrien Vandemeulebroecke,[3] Ilse Cleenwerck,[3] Marc Vancanneyt[3] and Jean Swings[2,3]

Correspondence
Sabri M. Naser
Sabri-Naser@najah.edu

[1]Department of Biology and Biotechnology, Faculty of Sciences, An-Najah National University, Nablus, Palestine

[2]Laboratory of Microbiology, Ghent University, K.L. Ledeganckstraat 35, Ghent 9000, Belgium

[3]BCCM™/LMG Bacteria Collection, Ghent University, K.L. Ledeganckstraat 35, Ghent 9000, Belgium

[4]Department of Applied Mathematics, Biometrics and Process Control, Ghent University, Coupure links 653, Ghent 9000, Belgium

[5]Bioinformatics and Evolutionary Genomics, Ghent University/VIB, Technologiepark 927, Ghent 9052, Belgium

The aim of this study was to evaluate the use of the phenylalanyl-tRNA synthase alpha subunit (*pheS*) and the RNA polymerase alpha subunit (*rpoA*) partial gene sequences for species identification of members of the genus *Lactobacillus*. Two hundred and one strains representing the 98 species and 17 subspecies were examined. The *pheS* gene sequence analysis provided an interspecies gap, which in most cases exceeded 10 % divergence, and an intraspecies variation of up to 3 %. The *rpoA* gene sequences revealed a somewhat lower resolution, with an interspecies gap normally exceeding 5 % and an intraspecies variation of up to 2 %. The combined use of *pheS* and *rpoA* gene sequences offers a reliable identification system for nearly all species of the genus *Lactobacillus*. The *pheS* and *rpoA* gene sequences provide a powerful tool for the detection of potential novel *Lactobacillus* species and synonymous taxa. In conclusion, the *pheS* and *rpoA* gene sequences can be used as alternative genomic markers to 16S rRNA gene sequences and have a higher discriminatory power for reliable identification of species of the genus *Lactobacillus*.

INTRODUCTION

Lactic acid bacteria (LAB) belonging to the genus *Lactobacillus* comprise the largest group of Gram-positive, rod-shaped and catalase-negative organisms (Hammes & Vogel, 1995) with *Lactobacillus delbrueckii* as the type species (Kandler & Weiss, 1986). Species of the genus *Lactobacillus* form part of the normal flora of the gastrointestinal tract, vagina and oral cavity of humans and animals (Hammes & Vogel, 1995; Klein *et al.*, 1998). Lactobacilli are of great economic importance for the dairy and other fermented food industries, where they are used as starter cultures for fermenting raw materials of vegetable or animal origin. *Lactobacillus* species are claimed to have health-promoting (probiotic) properties and some pharmaceutical preparations contain viable *Lactobacillus* strains (Holzapfel *et al.*, 2001; Reid, 1999; Stiles & Holzapfel, 1997). In this context, the accurate identification of members of the genus *Lactobacillus* remains a point of crucial importance.

Several methods have been used for the identification of lactobacilli to the species level, e.g. SDS-PAGE of whole-cell proteins, randomly amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), rep-PCR and ribotyping (Daud Khaled *et al.*,

Abbreviations: FAFLP, fluorescent amplified fragment length polymorphism; LAB, lactic acid bacteria; OTU, operational taxonomic unit.

The GenBank/EMBL/DDBJ accession numbers for the sequences reported in this paper are AM087677–AM087773, AM263502–AM263510, AM157783–AM157787, AM168426–AM168429, AM159098–AM159099, AM236139–AM236143, AM284176–AM284250, AM694185, AM694187 (*pheS* partial gene sequences) and AM087774–AM087869, AM263511–AM263518, AM157775, AM157777–AM157780, AM168431–AM168433, AM236144–AM236148, AM284251–AM284315, AM694186, AM694188 (*rpoA* partial gene sequences).

Neighbour-joining phylogenetic trees constructed using the *pheS* and *rpoA* gene sequences of the type strains of species of the genus *Lactobacillus* are available with the online version of this paper.

1997; Gancheva *et al.*, 1999; Gevers *et al.*, 2001; Massi *et al.*, 2004; Pot *et al.*, 1993; Yansanjav *et al.*, 2003). Although useful, there are some pitfalls associated with the use of these methods concerning portability, inter-laboratory reproducibility and time efficacy. Informational genes such as the 16S rRNA gene are commonly considered as reliable phylogenetic markers for assigning evolutionary relationships among species of the genus *Lactobacillus* (Schleifer & Ludwig, 1995). However, 16S rRNA gene sequence data do not allow the identification of closely related species. The use of housekeeping genes is emerging as an alternative to overcome these problems (Santos & Ochman, 2004; Stackebrandt *et al.*, 2002). Recent *in silico* studies based on complete genomes have provided the basis for establishing sets of housekeeping genes that can accurately predict genome relatedness and improve the accuracy of species identification. The need for alternative genomic markers that provide higher levels of discrimination than the 16S rRNA gene has led to a more systematic sequencing of housekeeping genes (Coenye *et al.*, 2005; Gevers *et al.*, 2005; Konstantinidis & Tiedje, 2005; Naser *et al.*, 2005a, b; Thompson *et al.*, 2005; Zeigler, 2003).

To be useful for species discrimination, genes must ideally be present in a single copy, evolve more rapidly than rRNA genes and be widely distributed among bacterial genomes. Those genes in which recombination might confer a selective advantage, or closely linked genes, should be avoided. Furthermore, these genes should be informative with an adequate degree of resolution and provide sufficient variability to differentiate species of a particular genus (Zeigler, 2003).

The use of the housekeeping genes that code for the α-subunit of bacterial phenylalanyl-tRNA synthase (*pheS*) and the α-subunit of RNA polymerase (*rpoA*) has proven to be a robust system for the identification of all the recognized species of the genus *Enterococcus* (Naser *et al.*, 2005b). As it is our intention to extend the application of these protein-coding loci to all other LAB genera, the present study was aimed at evaluating the usefulness of *pheS* and *rpoA* gene sequences as alternative genomic tools for the identification of species of the genus *Lactobacillus*. We compared the sequence data of the *pheS* and *rpoA* genes with the available 16S rRNA gene sequences. In addition, a software tool, named TaxonGap, was developed during this study to enable a straightforward evaluation of the discriminatory power of the individual genes in the *Lactobacillus* identification scheme.

METHODS

Two hundred and one well-characterized *Lactobacillus* strains representing 98 species and 17 subspecies of the genus *Lactobacillus* isolated from humans, animals or food products were analysed in this study (Table 1). Strains were grown on MRS agar media (Oxoid) at 37 °C for 48 h. All strains included in this study have been deposited in the BCCM/LMG Bacteria Collection at Ghent University (Ghent, Belgium). Bacterial genomic DNA was extracted as described by Gevers *et al.* (2001) or DNA alkaline extract was used (Niemann *et al.*, 1997). The amplification and sequencing of *pheS* and *rpoA* genes were as described by Naser *et al.* (2005a, b) with the following modifications: where an amplicon was not obtained with the referred conditions, the primer combination rpoA-21-F/rpoA-22-R (5′-ATGATYGARTTTGAAAAACC-3′/5′-ACYTTVATCATNTCWGVY-TC-3′) was used for the amplification of the *rpoA* gene and/or the Failsafe PCR system (Epicenter).

Consensus sequences were determined as described by Naser *et al.* (2005a, b). The CLUSTAL_X program was used for multiple sequence alignment. Consequently, the aligned sequences were imported into BioNumerics software version 4.5 (Applied Maths) for the calculation of similarity matrices and neighbour-joining trees (Saitou & Nei, 1987). The reliability of hierarchical clustering was determined by using the bootstrapping method with 1000 resamplings. The 16S rRNA gene sequence data of the *Lactobacillus* type strains were obtained from EMBL.

TaxonGap software tool. When evaluating multiple genes as candidate biomarkers for the identification of different operational taxonomic units (OTUs) (Sneath & Sokal, 1973), one is intuitively looking for molecular markers that show the least amount of heterogeneity within OTUs and also result in maximal separation between the different OTUs. The first requirement must guarantee that members of the same OTU have the same (or at least similar) biomarkers, so that they can easily be grouped together based on those markers. The second requirement is that members of different OTUs must have sufficiently different biomarkers so that an evaluation of these markers cannot erroneously suggest assignment of the members to the same OTU. The TaxonGap software tool was specially designed to produce a compact representation of the resolution of the biomarkers within and between taxonomic units, allowing easy and reliable inspection of the data for evaluations across the different OTUs and the different biomarkers.

For a given set of OTUs O_1, O_2, . . ., O_n, the *s*-heterogeneity within the taxon O_i (i=1, . . ., n) is defined as $\max_{x,y \in Oi,\ x \neq y} d_s(x, y)$. Herein, $d_s(x, y)$ represents the distance between the (different) members x and y of the taxon O_i as measured from the biomarker *s*. Likewise, the *s*-separability of the taxon O_i (i=1, . . ., n) is defined as $\min_{x \in Oi, y \notin Oi} d_s(x, y)$. The taxon containing y, for which the minimum distance is reached during the calculation of the *s*-separability, is called the closest neighbour of the taxon O_i. Note, however, that the closest neighbour relationship is not necessarily symmetric; given that O_i is the closest neighbour of O_j, it does not automatically follow that O_j is also the closest neighbour of O_i. The calculation of the *s*-heterogeneity and the *s*-separability are schematically represented in Fig. 1 for a taxon A and its closest neighbouring taxon B.

The TaxonGap software tool calculates the matrix of *s*-heterogeneity and *s*-separability values with the different OTUs as the matrix rows and the different biomarkers as the matrix columns. Headers are placed to the left and on top of the matrix. The rows and columns of the matrix can be placed in any order. However, to improve interpretability of the resulting representation, we have included the option to present the OTUs according to their position in a phylogenetic tree as an alternative to listing them in alphabetical order. Again, with the aim of improving the visual inspection and interpretation of the data, the TaxonGap software tool presents the *s*-heterogeneity and *s*-separability values as light grey and dark grey horizontal bars, respectively. The same scaling is used for plotting the *s*-heterogeneity and *s*-separability bars for the individual biomarkers in order to support optimal comparability of the values across the biomarkers. The name of the closest neighbour is attached to the right side of the dark grey bar. Light grey bars are printed on top of the dark grey bars and are made slightly thinner than the dark grey bars to improve visualization even when the light bars grow larger than the

Table 1. Details of the *Lactobacillus* species and strains that were analysed in this study

Species name	Strain number	Other strain numbers	Source
L. acetotolerans	LMG 10751^T	ATCC 43578^T, CCUG 32229^T	Spoiled rice vinegar broth
L. acidifarinae	LMG 22200^T	CCM 7240^T, CCUG 50162^T	Wheat sourdough
L. acidipiscis	LMG 19820^T	CCUG 46556^T, DSM 15836^T, FS60-1^T	Fermented fish
L. acidipiscis	LMG 23135		Fermented fish
L. acidophilus	LMG 9433^T	ATCC 4356^T, CCRC 10695^T	Human
L. acidophilus	LMG 8151	CCUG 12853, PRSF-L 133	Milk
L. agilis	LMG 9186^T	CCUG 31450^T, DSM 20509^T	Municipal sewage
L. agilis	LMG 11398	DSM 20508, Weiss 123	Municipal sewage
L. agilis	LMG 11399	DSM 20510, Weiss 298	Municipal sewage
L. algidus	LMG 19872^T	ATCC BAA-482^T, DSM 15638^T	Vacuum-packaged refrigerated beef
L. alimentarius	LMG 9187^T	ATCC 29643^T, CCUG 30672^T, DSM 20249^T	Marinated fish product
L. alimentarius	LMG 9188	ATCC 29647, DSM 20181	Marinated fish product
L. amylolyticus	LMG 18796^T	LA5^T, CCUG 39901^T, DSM 11664^T	Acidified beer wort
L. amylolyticus	LMG 18797	Bohak LA13	Acidified beer wort
L. amylolyticus	LMG 18804	Bohak LA44	Acidified beer wort
L. amylophilus	LMG 6900^T	ATCC 49845^T, CCUG 30137^T	Swine waste-corn fermentation
L. amylotrophicus	LMG 11400^T	DSM 20534^T, NRRL B-4436^T	Swine waste-corn fermentation
L. amylotrophicus	NRRL B-4435		Swine waste-corn fermentation
L. amylophilus	NRRL B-4438		Swine waste-corn fermentation
L. amylophilus	NRRL B-4439		Swine waste-corn fermentation
L. amylophilus	NRRL B-4440		Swine waste-corn fermentation
L. amylovorus	LMG 9496^T	ATCC 33620^T, CCUG 27201^T	Cattle waste-corn fermentation
L. amylovorus	LMG 18180	JCM 1032, KCTC 3149	Pig, intestine
L. amylovorus	LMG 9434	ATCC 33198, CCUG 37571	Pig, small intestine
L. animalis	LMG 9843^T	ATCC 35046^T, CCUG 33906^T	Baboon, dental plaque
L. animalis	LMG 17195	Devriese TA 44	
L. antri	LMG 22111^T	CCUG 48456^T, DSM 16041^T	
L. aviarius subsp. *aviarius*	LMG 10753^T	ATCC 43234^T, CCUG 32230^T, DSM 20655^T	Chicken, faeces
L. bifermentans	LMG 9845^T	ATCC 35409^T, CCUG 32234^T, DSM 20003^T	Blown Dutch cheese
L. bifermentans	LMG 11431	NCFB 1231	Cheese
L. bifermentans	LMG 11432	CCUG 42896, NCFB 1232	Cheese
L. brevis	LMG 6906^T	ATCC 14869^T, CCUG 30670^T	Human, faeces
L. brevis	LMG 11435	NCDO 473, NCFB 473	Silage
L. buchneri	LMG 6892^T	ATCC 4005^T	Tomato pulp
L. buchneri	LMG 11439	ATCC 9460, NCFB 111	
L. casei	LMG 6904^T	ATCC 393^T, CCM 7088^T	Cheese
L. coleohominis	LMG 21591^T	CCUG 44007^T, DSM 14060^T	31-year-old healthy woman, vagina
L. collinoides	LMG 9194^T	ATCC 27612^T, CCUG 32259^T	Fermenting apple juice
L. collinoides	LMG 9195	NCFB 2149	Cider and apple juices
L. collinoides	LMG 18850		Distillation cider
L. coryniformis subsp. *coryniformis*	LMG 9196^T	ATCC 25602^T, CCUG 30666^T	Silage
L. coryniformis subsp. *torquens*	LMG 9197^T	ATCC 25600^T, CCUG 30667^T	Air of dairy barn
L. crispatus	LMG 9479^T	ATCC 33820^T, CCUG 30722^T	Eye
L. curvatus	LMG 19715		Blood culture
L. curvatus	LMG 12006	C14/7, NCFB 1039	Italian hard cheese
L. curvatus	LMG 12007	C4/1, NCFB 1041	English hard cheese
L. curvatus	LMG 9198^T	ATCC 25601^T, CCUG 30669^T	Milk
L. curvatus	LMG 17299	CCUG 31333, Reuter Rv40a	Raw sausage
L. cypricasei	LMG 21592^T	CCUG 42961^T, DSM 15353^T	Cheese
L. cypricasei	CCUG 42959	CCUG 42959, LMK1	Cheese
L. cypricasei	CCUG 42960	CCUG 42960, LMK2	Cheese
L. cypricasei	CCUG 42962	CCUG 42962, LMD 2	Cheese
L. delbrueckii subsp. *bulgaricus*	LMG 6901^T	ATCC 11842^T, CCM 7190^T	Bulgarian yoghurt
L. delbrueckii subsp. *bulgaricus*	LMG 12168	PRSF-L 144, Topisirovic BGPF 1	Homemade yoghurt
L. delbrueckii subsp. *delbrueckii*	LMG 6412^T	ATCC 9649^T, CCM 7191^T	Distillery sour grain mash
L. delbrueckii subsp. *delbrueckii*	LMG 22235	CCUG 29179	Human, urine
L. delbrueckii subsp. *delbrueckii*	LMG 22236	CCUG 47846, KARL B 30110/03	83-year-old woman

Table 1. cont.

Species name	Strain number	Other strain numbers	Source
L. delbrueckii subsp. *indicus*	LMG 22083^T	DSM 15996^T, NCC725^T	Indian diary products
L. delbrueckii subsp. *lactis*	LMG 7942^T	ATCC 12315^T, CCUG 31454^T	Emmental cheese
L. delbrueckii subsp. *lactis*	LMG 6401	ACM 3573, ATCC 7830	
L. diolivorans	LMG 19667^T	DSM 14421^T, JCM 12183^T	Maize silage
L. durianis	LMG 19193^T	CCUG 45405^T, CIP 107501^T, DSM 15802^T	Tempoyak
L. durianis	LMG 19196	MC13-1	Tempoyak
L. equi	LMG 21748^T	CCUG 47129^T, DSM 15833^T	Horse, faeces
L. farciminis	LMG 9200^T	ATCC 29644^T, CCUG 30671^T, DSM 20184^T	Sausage
L. farciminis	LMG 17703	Leisner II-8-50	Marinated meat product
L. fermentum	LMG 6902^T	ATCC 14931^T, CCM 7192^T	Fermented beets (*Beta vulgaris*)
L. fermentum	LMG 8902	CCM 2481, NCFB 2341	
L. fermentum	LMG 8154	CCUG 2231, La45	
L. fructivorans	LMG 9201^T	ATCC 8288^T, DSM 20203^T	
L. fructivorans	LMG 9202	NCFB 2166, NCIMB 5223, strain W1	
L. frumenti	LMG 19473^T	DSM 13145^T, TMW 1.666^T	Rye-bran sourdough
L. fuchuensis	LMG 21669^T	CCUG 47133^T, DSM 14340^T	Vacuum-packaged refrigerated beef
L. gallinarum	LMG 9435^T	ATCC 33199^T, CCUG 30724^T, DSM 10532^T	Chicken, crop
L. gallinarum	LMG 14751	CCUG 31412, Fujisawa T-50, JCM 8782	Chicken, faeces
L. gallinarum	LMG 14755	Fujisawa TFC3, JCM 8786	Chicken, faeces
L. gasseri	LMG 9203^T	ATCC 33323^T, CCUG 31451^T	Human
L. gasseri	LMG 13134	ATCC 9857, CIP 62.18	Vaginal tract
L. gasseri	LMG 11413	NCIMB 8819, PRSF-L 146	Human, saliva
L. gasseri	LMG 18176	JCM 1025, PRSF-L 150	Human, intestine
L. gasseri	LMG 10771	CCUG 25736	Wine
L. gasseri	LMG 13047	ATCC 19992, CCUG 39972, DSM 20077	Human, faeces
L. gasseri	LMG 18177	JCM 1026	
L. gasseri	LMG 11478	ATCC 4963, JCM 5343	Human
L. gastricus	LMG 22113^T	CCUG 48454^T, DSM 16045^T, Kx156A7^T	Human stomach mucosa
L. graminis	LMG 9825^T	ATCC 51150^T, CCUG 32238^T	Grass silage
L. hammesii	LMG 23074^T		French wheat sourdough
L. hamsteri	LMG 10754^T	ATCC 43851^T, DSM 5661^T	Hamster, faeces
L. helveticus	LMG 6413^T	ATCC 15009^T, BCRC 12936^T, CCM 7193^T	Swiss Emmental cheese
L. helveticus	LMG 11445	ATCC 521, BCRC 14026, CCM 1751	
L. helveticus	LMG 11447	ATCC 10812, BCRC 14021	
L. helveticus	LMG 13522	ATCC 12046, BCRC 12259, JCM 1554	
L. helveticus	LMG 18225	ATCC 8001, NCFB 103	
L. helveticus	LMG 22464	CCUG 50205, SA, type strain of *L. suntoryeus*	Malt whisky fermentation
L. helveticus	LMG 22465	M4	Malt whisky fermentation
L. hilgardii	LMG 6895^T	ATCC 8290^T, CCUG 30140^T	Wine
L. hilgardii	LMG 11964	CECT 4681, Couto 28	Port wine
L. hilgardii	LMG 11966	CECT 4682, Couto 30	Port wine machinery
L. homohiochii	LMG 9478^T	ATCC 15434^T, CCUG 32247^T, DSM 20571^T	Spoiled sake
L. iners	LMG 18914^T	CCUG 28746^T, DSM 13335^T	36-year-old woman, urine
L. iners	LMG 18915	CCUG 37287, strain 8	Medical care product
L. iners	LMG 18916	CCUG 38673	Healthy 28-year-old woman, cervix
L. ingluviei	LMG 20380^T	CCUG 45722^T, KR3^T, JCM 12531^T	Pigeon, crop
L. ingluviei	LMG 22056	DSM 14792, type strain of *L. thermotolerans*	Chicken, faeces
L. intestinalis	LMG 14196^T	ATCC 49335^T, CCUG 30727^T, DSM 6629^T	Rat, intestine
L. intestinalis	LMG 11462	NCFB 2176, strain HE1	
L. jensenii	LMG 6414^T	ATCC 25258^T, BCRC 12939^T	Human, vaginal discharge
L. johnsonii	LMG 9436^T	ATCC 33200^T, CCUG 30725^T, DSM 10533^T	Human, blood
L. johnsonii	LMG 18206	CCUG 31413, JCM 8793	Pig, faeces
L. johnsonii	LMG 9437	ATCC 11506	
L. johnsonii	LMG 11468	ATCC 332, CCUG 44520	Human
L. johnsonii	LMG 18175	JCM 1022, PRSF-L 156	Human, intestine
L. johnsonii	LMG 18193	CCRC 14037, JCM 5812, PRSF-L 154	Pharmaceutical preparation

Table 1. cont.

Species name	Strain number	Other strain numbers	Source
L. johnsonii	LMG 18195	JCM 5814	Chicken, faeces
L. johnsonii	LMG 18204	CM 8791, PRSF-L 153	Mouse, faeces
L. johnsonii	LMG 18205	Fujisawa F133, JCM 8792	Calf, faeces
L. kalixenisis	LMG 22115^T	CCUG 48459^T, DSM 16043^T, Kx127A2^T	Human stomach mucosa
L. kefiranofaciens subsp. *kefiranofaciens*	LMG 19149^T	ATCC 43761^T, CCUG 32248^T	Kefir grains
L. kefiranofaciens subsp. *kefirgranum*	LMG 15132^T	ATCC 51647^T, CCUG 49353^T	Kefir grains
L. kefiri	LMG 9480^T	ATCC 35411^T, BCRC 14011^T, CCUG 30673^T	Kefir grains
L. kefiri	LMG 11496	NCFB 2090	Kefir grains
L. kefiri	LMG 11453	NCFB 2132, strain X6	Kefir grains
L. kimchii	LMG 19822^T	CCUG 45370^T, DSM 13961^T	Kimchi, a Korean fermented food
L. kitasatonis	LMG 23133^T		Chicken, intestine
L. kunkeei	LMG 18925^T	ATCC 700308^T, DSM 12361^T, YH-15^T	Fermented grape juice
L. lindneri	LMG 14528^T	DSM 20690^T, JCM 11027^T	Spoiled beer
L. malefermentans	LMG 11455^T	ATCC 49373^T, CCUG 32206^T	Beer
L. mali	LMG 6899^T	ATCC 27053^T, CCM 2878^T	Apple juice from cider press
L. manihotivorans	LMG 18010^T	CCUG 42894^T, DSM 13343^T	Cassava sour starch
L. mindensis	LMG 21932^T	CCUG 48642^T, DSM 14500^T	Sourdough fermentation
L. mucosae	LMG 19534^T	CCUG 43179^T, DSM 13345^T, strain S32^T	Pig, intestine
L. mucosae	LMG 19536	CCUG 43181, strain 1028	Pig, intestine
L. murinus	LMG 14189^T	ATCC 35020^T, CCUG 33904^T	Rat, digestive tract
L. nagelii	LMG 21593^T	ATCC 700692^T, CCUG 43575^T	Partially fermented wine
L. oris	LMG 9848^T	ATCC 49062^T, CCUG 37396^T	Italian human, saliva
L. panis	LMG 21658^T	CCUG 37482^T, DSM 6035^T	Rye sourdough
L. pantheris	LMG 21017^T	DSM 15945^T, JCM 12539^T	Jaguar, faeces
L. parabrevis	LMG 11984^T	ATCC 53295^T	Wheat
L. parabrevis	LMG 11494	Hayward 8/3, NCFB 1058	Farmhouse red Cheshire cheese
L. parabuchneri	LMG 11457^T	ATCC 49374^T, CCUG 32261^T	Human, saliva
L. parabuchneri	LMG 11987	ATCC 12936, Rogosa 708B	Oral
L. parabuchneri	LMG 22038	JCM 12511, type strain of *L. ferintoshensis*	Malt whisky fermentation
L. paracasei subsp. *paracasei*	LMG 13729	Patarata 56	Young red table wine
L. paracasei subsp. *paracasei*	LMG 13087^T	ATCC 25302^T, CCM 1753^T	
L. paracasei subsp. *paracasei*	LMG 10774	CCUG 27320	Cerebrospinal fluid
L. paracasei subsp. *paracasei*	LMG 11965	Couto 29	Port wine machinery
L. paracasei subsp. *paracasei*	LMG 8157	CCUG 17717	
L. paracasei subsp. *tolerans*	LMG 9191^T	ATCC 25599^T, CCUG 34829^T	Pasteurized milk
L. paracollinoides	LMG 22473^T		Brewery environment
L. parakefiri	LMG 15133^T	ATCC 51648^T, CCUG 39468^T, DSM 10551^T	Kefir grains
L. paralimentarius	LMG 19152^T	CCUG 43349^T, DSM 13238^T, JCM 10415^T	Sourdough
L. paraplantarum	LMG 16673^T	ATCC 700211^T, CCUG 35983^T	Beer
L. paraplantarum	LMG 18398	ATCC 10776, DSM 10641	
L. paraplantarum	LMG 21638	ATCC 700210	Human, stool
L. pentosus	LMG 10755^T	ATCC 8041^T	
L. pentosus	LMG 9210	CCM 4619	Liquor waste fermentation
L. pentosus	LMG 17677	Leisner 13-16	Chili bo
L. perolens	LMG 18936^T	L 532^T, DSM 12744^T, JCM 12534^T	
L. perolens	LMG 18937	Bohak L48	
L. perolens	LMG 18939	Bohak L426	
L. plantarum	LMG 6907^T	ATCC 14917^T, CCM 7039^T	Pickled cabbage
L. plantarum	LMG 11405	DSM 2648	Silage
L. plantarum	LMG 18404	ATCC 8008, L14	
L. plantarum	LMG 19807	CCUG 45396, type strain of *L. arizonensis*	Jojoba meal fermentation
L. plantarum subsp. *argentoratensis*	LMG 9205^T	CCUG 50787^T, DSM 16365^T	Fermented corn product (Ogi)
L. pontis	LMG 14187^T	ATCC 51518^T, DSM 8475^T	Rye sourdough
L. pontis	LMG 14188	ATCC 51519, DSM 8476	Rye sourdough
L. psittaci	LMG 21594^T	CCUG 42378^T, DSM 15354^T	Parrot, lung
L. reuteri	LMG 9213^T	ATCC 23272^T, CCRC 14625^T, CCUG 33624^T	Adult, intestine

Table 1. cont.

Species name	Strain number	Other strain numbers	Source
L. reuteri	LMG 18238	ATCC 55148, Bio Gaia AB 11284	Chicken
L. reuteri	LMG 13090	CCUG 42759, PRSF-L 164, strain A1	Rat
L. rhamnosus	LMG 6400^T	ACM 539^T, ATCC 7469^T	
L. rhamnosus	LMG 12166	Topisirovic BGEN1	Homemade hard cheese
L. rhamnosus	LMG 10775	CCUG 27333, PRSF-L 173	Human, clinical sample
L. rhamnosus	LMG 18030	El Soda 42, PRSF-L 169	Zabady (yoghurt)
L. rossiae	LMG 22972^T	DSM 15814^T	Wheat sourdoughs
L. ruminis	LMG 10756^T	ATCC 27780^T, DSM 20403^T	Bovine, rumen
L. ruminis	LMG 11461	ATCC 27781, DSM 20404	Bovine, rumen
L. saerimneri	LMG 22087^T	CCUG 48462^T, DSM 16049^T, GDA154^T	Pig, faeces
L. saerimneri	LMG 22088	CCUG 48463, DSM 16027, GDA164	Pig, faeces
L. sakei subsp. *carnosus*	LMG 17305	CCUG 32077	Human, blood
L. sakei subsp. *carnosus*	LMG 17306	CCUG 32584, Kalmar B2571	Human with endocarditis, blood
L. sakei subsp. *carnosus*	LMG 17302^T	CCUG 31331^T, DSM 15831^T	Pig, faeces
L. sakei subsp. *sakei*	LMG 9468^T	AS 1.2142^T, ATCC 15521^T, CCUG 30501^T	Sake starter (Moto)
L. sakei subsp. *sakei*	LMG 7941	DSM 20198	
L. salivarius	LMG 14476	Devriese 94/438	Cat with myocarditis
L. salivarius	LMG 14477	Devriese 94/428	Parakeet with sepsis
L. salivarius subsp. *salicinius*	LMG 9476^T	ATCC 11742^T, CCUG 39464^T	Saliva
L. salivarius subsp. *salivarius*	LMG 9477^T	ATCC 11741^T, DSM 20555^T	Saliva
L. sanfranciscensis	LMG 16002^T	ATCC 27651^T, DSM 20451^T	San Francisco sour dough
L. satsumensis	LMG 22973^T		Shochu mashes
L. sharpeae	LMG 9214^T	ATCC 49974^T, DSM 20505^T	Municipal sewage
L. spicheri	LMG 21871^T	DSM 15429^T, LTH 5753^T	Rice sourdough
L. suebicus	LMG 11408^T	ATCC 49375^T, DSM 5007^T	Apple mash
L. ultunensis	LMG 22117^T	CCUG 48460^T, DSM 16047^T, Kx146C1^T	Human stomach mucosa
L. vaccinostercus	LMG 9215^T	ATCC 33310^T, DSM 20634^T	Cow dung
L. vaginalis	LMG 12891^T	ATCC 49540^T, DSM 5837^T	Vagina
L. versmoldensis	LMG 21929^T		Raw fermented sausage
L. vitulinus	LMG 18931^T	ATCC 27783^T, DSM 20405^T	Calf, rumen
L. zeae	LMG 17315^T	ATCC 15820^T, DSM 20178^T	Corn steep liquor
L. zymae	LMG 22198^T	CCM 7241^T, CCUG 50163^T	Wheat sourdough

dark bars. The latter only occurs in the rare occasion when, for a given biomarker, members in a taxon are more distant to each other than a member of the taxon is to a member of another taxon. Although not a strict requirement, it is advised that the same OTUs are used for the evaluation of different biomarkers. Missing biomarker data for a given OTU leads to holes in the TaxonGap output matrix. There is no requirement to use the same OTU members for measuring different biomarkers.

Distances used for the calculation of the *s*-heterogeneity and *s*-separability values were determined using pairwise nucleotide sequence alignments with the Needleman-Wunsch algorithm as implemented in the BioNumerics 4.5 software package.

RESULTS AND DISCUSSION

Application of TaxonGap for the evaluation of *pheS* and *rpoA* gene sequences as biomarkers for species identification

Fig. 2 shows the TaxonGap output for the *Lactobacillus* identification scheme discussed in the present study. The OTUs subjected to the TaxonGap analysis were the different species of the genus *Lactobacillus*. Cases where species synonymy has been reported in the literature were regarded as a single species during the TaxonGap analysis. The biomarkers were the *pheS*, *rpoA* and 16S rRNA genes. The *s*-heterogeneity is a measure of the heterogeneity observed in the biomarker *s* among the different strains of the same *Lactobacillus* species (subsequently referred to as intraspecies heterogeneity). The *s*-separability is a measure of the divergence between the different *Lactobacillus* species (subsequently referred to as interspecies divergence). Subspecies were not taken into account during this analysis as it was evident from the data that few subspecies could be separated by the biomarkers studied. Where a given gene was able to make clear separation between subspecies, it is indicated in the discussion of the different phylogenetic groups below.

The members of the genus *Lactobacillus* were ordered according to their phylogenetic positioning in a neighbour-joining tree calculated from the 16S rRNA gene sequences of their type strains. The different *Lactobacillus* species groups are delineated on the left of the neighbour-joining

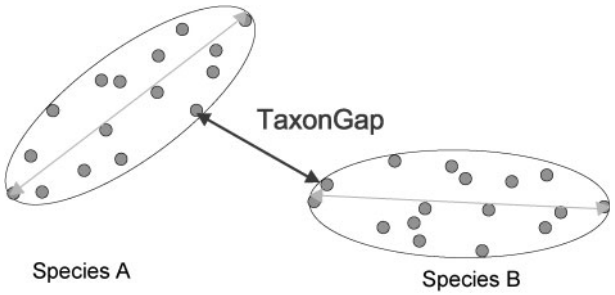

Fig. 1. TaxonGap versus intraspecies diversity. A schematic representation of diversity within and between two species A and B. Dots represent operational taxonomic units (OTUs) in evolutionary space, with the distance between dots relative to the distance derived from sequence information. The intraspecies diversity is indicated by the light grey arrows and represents the maximum sequence distance between strains of the species (corresponds to the light grey bars in Fig. 2). The TaxonGap, indicated by the dark grey arrow, represents the distance between species A and its closest neighbouring species, species B (corresponds to the dark grey bars in Fig. 2).

tree. Although heterogeneity could not be estimated for the 16S rRNA gene as sequence data were only available for the type strains, the separability of the *Lactobacillus* species based on the 16S rRNA gene was added as the first column of the TaxonGap output matrix. This allows better evaluation of the discriminatory power of the 16S rRNA gene for species identification when compared with the other genes included in the identification scheme. The *pheS* and *rpoA* genes formed the second and third biomarker columns in the TaxonGap output matrix. In order to guide the readership in the interpretation of the TaxonGap output in the following discussion, we focus on the first row of Fig. 2. From this row, we can determine the *s*-heterogeneity and *s*-separability values for the *L. agilis* species. For this species, the observed *pheS*-heterogeneity was 1.5 % (see light grey bar), whereas the *rpoA* heterogeneity only reached 0.3 % for the same species. Likewise, one can see that the closest neighbour of the *L. agilis* species is estimated differently for the 16S rRNA gene (*L. equi*; 4.9 %), the *pheS* gene (*L. animalis*; 17.3 %) and the *rpoA* gene (*L. acidipiscis*; 15.7 %). However, it should be noted that all of these species belong to the same *L. salivarius* species group. This is an example of the general trend observed in the dataset: that when species have different closest neighbours for the genes in the identification scheme, these species all belong to the same species group.

The representation produced by the TaxonGap software tool offers a number of advantages over comparing individual trees for the different gene sequences included in polygenic identification studies. First of all, a separate row is reserved in the TaxonGap output matrix for the heterogeneity and separability values of the different genes for each species, which is not the case when comparing phylogenetic trees. Even after the tedious process of swapping branches, it is not always possible to draw phylogenetic trees in a way that enables clear visual comparisons to be made. This is especially the case when trees for multiple genes need to be compared. In addition, TaxonGap uses the same scaling for depicting the distance values based on the different gene sequences. Few software tools for drawing phylogenetic trees allow precise control over the scaling. Both placement and scaling improve the comparability of the heterogeneity and separability for individual species. Secondly, we want to point out that phylogenetic trees present approximations of the underlying distance values whereas the TaxonGap filters out original similarity values instead of approximations by using minimum and maximum as aggregation operators. This is important when comparing *s*-heterogeneity and *s*-separability for all species for a given gene *s*. To underscore the overall success rate of the individual genes to discriminate between species of the genus *Lactobacillus*, we have depicted the overall heterogeneity (light grey) and separability (dark grey) per species as vertical lines for each gene in Fig. 2. Finally, the graphical output of TaxonGap remains compact, even for datasets where the number of OTU members grows large. This is because the software has a built-in aggregation based on the individual OTUs. Representing phylogenetic trees with over a few hundred entries would be almost impossible in printed format.

The TaxonGap software tool thus allows for a more straightforward evaluation of the discriminatory power of the individual genes in the *Lactobacillus* species identification scheme, as opposed to the need to compare separate gene trees drawn for each of the genes in the scheme.

Robustness of *pheS* and *rpoA* partial gene sequences for *Lactobacillus* species identification

The success of any bacterial species identification system depends on accuracy. Accuracy allows the distinction between intraspecific variation and interspecific divergence in the selected loci. The less overlap there is between genetic variation within species and divergence from species, the more effective the system becomes (Meyer & Paulay, 2005).

Both the *pheS* (382–455 nt) and *rpoA* (402–694 nt) partial gene sequences were applied as alternative genomic markers for the identification of *Lactobacillus* at the species level. Two hundred and one well-characterized *Lactobacillus* strains representing 98 species and 17 subspecies of the genus *Lactobacillus* from different origins were analysed in this study (Table 1). The strains were selected on the basis of previous polyphasic classification using AFLP, RAPD-PCR and SDS-PAGE of whole-cell proteins and represent the known heterogeneity of *Lactobacillus* species. In order to evaluate the *pheS* and

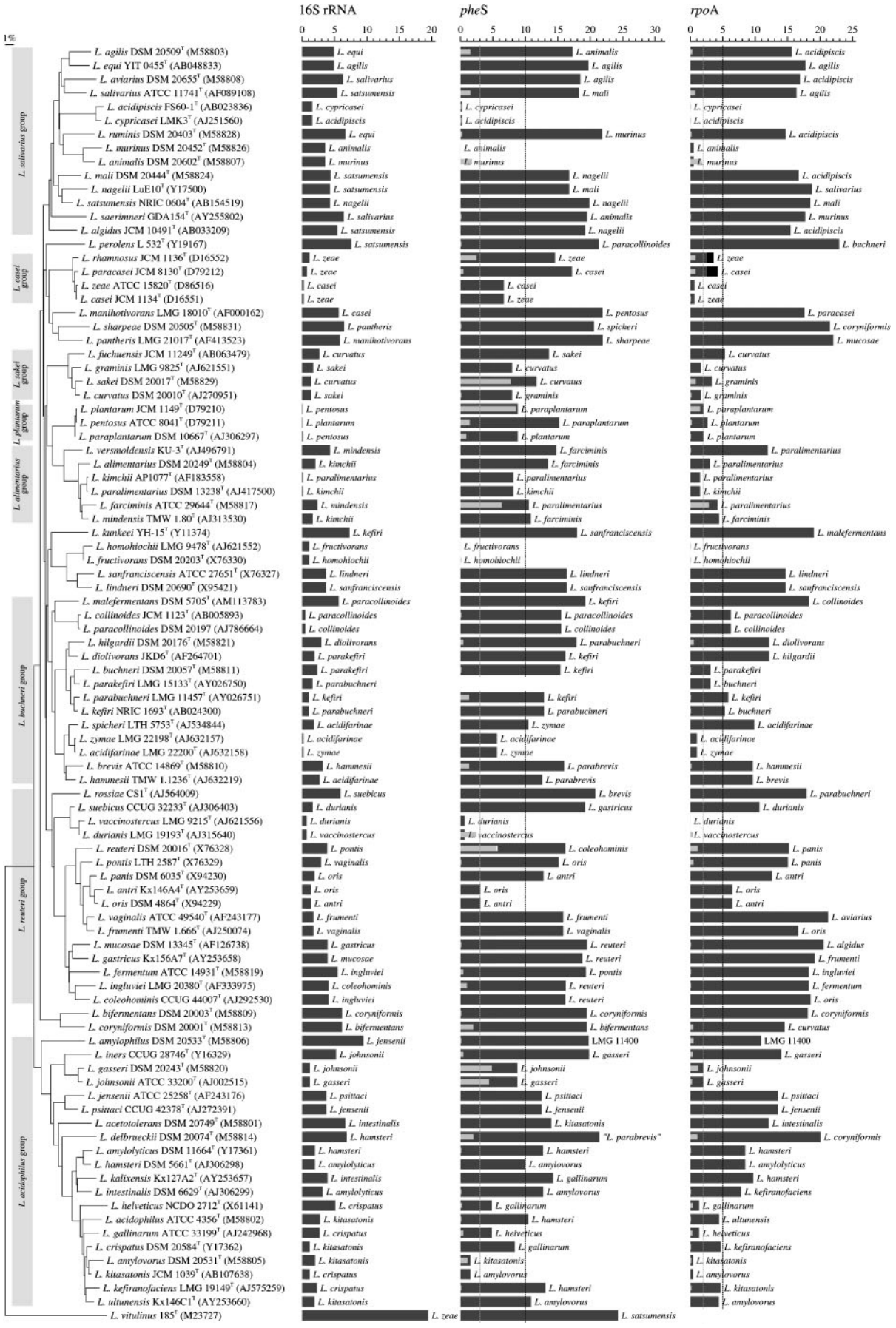
16S rRNA
pheS
rpoA
1%
L. salivarius group
L. casei group
L. sakei group
L. plantarum group
L. alimentarius group
L. buchneri group
L. reuteri group
L. acidophilus group
L. agilis DSM 20509T (M58803)
L. equi YIT 0455T (AB048833)
L. aviarius DSM 20655T (M58808)
L. salivarius ATCC 11741T (AF089108)
L. acidipiscis FS60-1T (AB023836)
L. cypricasei LMK3T (AJ251560)
L. ruminis DSM 20403T (M58828)
L. murinus DSM 20452T (M58826)
L. animalis DSM 20602T (M58807)
L. mali DSM 20444T (M58824)
L. nagelii LuE10T (Y17500)
L. satsumensis NRIC 0604T (AB154519)
L. saerimneri GDA154T (AY255802)
L. algidus JCM 10491T (AB033209)
L. perolens L 532T (Y19167)
L. rhamnosus JCM 1136T (D16552)
L. paracasei JCM 8130T (D79212)
L. zeae ATCC 15820T (D86516)
L. casei JCM 1134T (D16551)
L. manihotivorans LMG 18010T (AF000162)
L. sharpeae DSM 20505T (M58831)
L. pantheris LMG 21017T (AF413523)
L. fuchuensis JCM 11249T (AB063479)
L. graminis LMG 9825T (AJ621551)
L. sakei DSM 20017T (M58829)
L. curvatus DSM 20010T (AJ270951)
L. plantarum JCM 1149T (D79210)
L. pentosus ATCC 8041T (D79211)
L. paraplantarum DSM 10667T (AJ306297)
L. versmoldensis KU-3T (AJ496791)
L. alimentarius DSM 20249T (M58804)
L. kimchii AP1077T (AF183558)
L. paralimentarius DSM 13238T (AJ417500)
L. farciminis ATCC 29644T (M58817)
L. mindensis TMW 1.80T (AJ313530)
L. kunkeei YH-15T (Y11374)
L. homohiochii LMG 9478T (AJ621552)
L. fructivorans DSM 20203T (X76330)
L. sanfranciscensis ATCC 27651T (X76327)
L. lindneri DSM 20690T (X95421)
L. malefermentans DSM 5705T (AM113783)
L. collinoides JCM 1123T (AB005893)
L. paracollinoides DSM 20197 (AJ786664)
L. hilgardii DSM 20176T (M58821)
L. diolivorans JKD6T (AF264701)
L. buchneri DSM 20057T (M58811)
L. parakefiri LMG 15133T (AY026750)
L. parabuchneri LMG 11457T (AY026751)
L. kefiri NRIC 1693T (AB024300)
L. spicheri LTH 5753T (AJ534844)
L. zymae LMG 22198T (AJ632157)
L. acidifarinae LMG 22200T (AJ632158)
L. brevis ATCC 14869T (M58810)
L. hammesii TMW 1.1236T (AJ632219)
L. rossiae CS1T (AJ564009)
L. suebicus CCUG 32233T (AJ306403)
L. vaccinostercus LMG 9215T (AJ621556)
L. durianis LMG 19193T (AJ315640)
L. reuteri DSM 20016T (X76328)
L. pontis LTH 2587T (X76329)
L. panis DSM 6035T (X94230)
L. antri Kx146A4T (AY253659)
L. oris DSM 4864T (X94229)
L. vaginalis ATCC 49540T (AF243177)
L. frumenti TMW 1.666T (AJ250074)
L. mucosae DSM 13345T (AF126738)
L. gastricus Kx156A7T (AY253658)
L. fermentum ATCC 14931T (M58819)
L. ingluviei LMG 20380T (AF333975)
L. coleohominis CCUG 44007T (AJ292530)
L. bifermentans DSM 20003T (M58809)
L. coryniformis DSM 20001T (M58813)
L. amylophilus DSM 20533T (M58806)
L. iners CCUG 28746T (Y16329)
L. gasseri DSM 20243T (M58820)
L. johnsonii ATCC 33200T (AJ002515)
L. jensenii ATCC 25258T (AF243176)
L. psittaci CCUG 42378T (AJ272391)
L. acetotolerans DSM 20749T (M58801)
L. delbrueckii DSM 20074T (M58814)
L. amylolyticus DSM 11664T (Y17361)
L. hamsteri DSM 5661T (AJ306298)
L. kalixensis Kx127A2T (AY253657)
L. intestinalis DSM 6629T (AJ306299)
L. helveticus NCDO 2712T (X61141)
L. acidophilus ATCC 4356T (M58802)
L. gallinarum ATCC 33199T (AJ242968)
L. crispatus DSM 20584T (Y17362)
L. amylovorus DSM 20531T (M58805)
L. kitasatonis JCM 1039T (AB107638)
L. kefiranofaciens LMG 19149T (AJ575259)
L. ultunensis Kx146C1T (AY253660)
L. vitulinus 185T (M23727)

Fig. 2. Representation of the discriminatory power of the genes for species identification of the genus *Lactobacillus*. The left panel shows a neighbour-joining tree of the complete 16S rRNA gene sequences of the *Lactobacillus* type strains, including species groups. EMBL accession numbers of the 16S rRNA gene sequences are indicated in parentheses. For each of the species in the phylogenetic tree, the right panel depicts the intraspecies variability for *pheS* and *rpoA* genes and interspecies variability for the 16S rRNA, *pheS* and *rpoA* genes as horizontal light grey and dark grey bars, respectively. Overall distance gaps between species are represented in the graphic as lines. The right panel also contains the names of the closest relatives as estimated from the different loci.

rpoA gene sequence variations at the intraspecies level, we included several representative strains for each *Lactobacillus* species. In general, the *pheS* and *rpoA* gene sequences showed intraspecies variations up to 3 % and 2 %, respectively (Fig. 2).

The differentiating power of the *pheS* and *rpoA* partial gene sequences was examined for *Lactobacillus* species at the subspecies level. In general, the subspecies of *Lactobacillus* were highly related, having 98–100 % *pheS* and *rpoA* gene sequence similarities. This shows that the discriminatory power of the investigated loci to differentiate between the subspecies of most lactobacilli is low. However, *pheS* gene sequences could differentiate between the subspecies of *L. sakei* and *L. plantarum* (see below).

The analysis of *pheS* and *rpoA* partial gene sequences clearly differentiates the members of the genus *Lactobacillus* (see also Supplementary Figs S1 and S2 available in IJSEM Online). In comparison with the 16S rRNA gene, our data clearly indicate that *pheS* and *rpoA* genes provide higher resolution for differentiating *Lactobacillus* species. As shown in Fig. 2, both *pheS* and *rpoA* partial gene sequences provide alternative reliable genomic markers to differentiate the members of the genus *Lactobacillus*. However, it should be mentioned here that both *pheS* and *rpoA* partial gene sequences showed a variable discriminatory power for identifying different species of the genus *Lactobacillus*. An example that illustrates the variation of the *pheS* and *rpoA* partial gene sequences in their degree of resolution is shown in Fig. 2 between the type strains of *L. acidifarinae* and *L. zymae* (*L. buchneri* group).

The *pheS* gene sequence analysis provided the highest discrimination for the identification of different species of lactobacilli. The case of *L. antri* and *L. oris* (*L. reuteri* group) is an exception here where the *rpoA* gene provided more resolution than the *pheS* gene in differentiating the two species. The *pheS* gene sequence analysis provided an interspecies gap, which normally exceeds 10 % divergence and an intraspecies variation up to 3 %. The *rpoA* gene sequences revealed a somewhat lower resolution with an interspecies gap normally exceeding 5 % and an intraspecies variation up to 2 %.

It should be mentioned that the variation of the investigated genes in their discriminatory power, together with the fact that different genes might provide different closest neighbours or topologies without hampering their use to unambiguously circumscribe bacterial species, validated the necessity for the simultaneous analysis of several protein-coding loci for a robust taxonomic analysis at the species and genus levels.

Species groups based on 16S rRNA gene similarity

The currently recognized phylogenetic relationships within the genus *Lactobacillus* have been determined by comparative analysis of their 16S rRNA gene sequences (Schleifer & Ludwig, 1995). Based on these data, different phylogenetic species groups have been distinguished: the *L. acidophilus*, *L. reuteri*, *L. buchneri*, *L. alimentarius*, *L. plantarum*, *L. sakei*, *L. casei* and *L. salivarius* species groups.

On the basis of *pheS* gene sequence analysis, members of the *L. reuteri*, *L. alimentarius*, *L. plantarum*, *L. sakei* and *L. casei* species groups clustered together in clades corresponding with the 16S rRNA gene based phylogeny (see Supplementary Fig. S1 in IJSEM Online), whereas members of the *L. acidophilus*, *L. buchneri*, and *L. salivarius* species groups are clustered in two separate clades. On the basis of *rpoA* gene sequence analysis, the *L. acidophilus*, *L. reuteri*, *L. alimentarius*, *L. plantarum*, *L. sakei*, *L. casei* and *L. salivarius* species groups clustered together in clades corresponding with the 16S rRNA gene based phylogeny whereas the *L. buchneri* species group clustered in two separate clades (see Supplementary Fig. S2 in IJSEM Online).

In subsequent sections, we will discuss and compare our data and the data from the literature for all species of the genus *Lactobacillus* on the basis of the species groups delineated by the 16S rRNA gene phylogeny.

L. acidophilus species group

Within the *L. acidophilus* species group, the *pheS* and *rpoA* gene sequence data clearly differentiate the members of the *L. acidophilus* group with a maximum of 94 % and 98 % *pheS* and *rpoA* gene sequence similarities, respectively, except for *L. kitasatonis* and *L. amylovorus* (with 98.5 % and 99 % *pheS* and *rpoA* gene sequence similarities, respectively). At the intraspecies level, strains of same species were highly related (>98 % *pheS* and *rpoA* gene sequence similarities). However, as an exception, the neighbour-joining tree based on *pheS* gene sequences revealed distinct subclusters among strains of the species *L. gasseri* (8 strains) having 95 % *pheS* gene sequence similarity and among strains of the species *L. johnsonii*

(9 strains) having 96 % *pheS* gene sequence similarity (results not shown). The heterogeneity within *L. gasseri* strains was also observed by comparing the fluorescent amplified fragment length polymorphism (FAFLP) fingerprints of these strains with reference profiles of lactic acid bacteria taxa (unpublished data).

The neighbour-joining trees derived from the *pheS* and *rpoA* gene sequences revealed close relatedness between *L. helveticus* and *L. suntoryeus*, with at least 99.5 % *pheS* and *rpoA* gene sequence similarities (see Supplementary Figs S1 and S2). In addition, sequence analysis of the gene that codes for the α-subunit of ATP synthase (*atpA*) also showed a high relatedness between the two species. Further genomic data derived from DNA–DNA hybridization unambiguously demonstrated that *L. suntoryeus* is a later synonym of *L. helveticus* (Naser *et al.*, 2006a).

The *pheS* and *rpoA* partial gene sequences revealed heterogeneity among culture collection strains of *L. amylophilus* described by Nakamura & Crowell (1979). Strains LMG 11400 and NRRL B-4435 represent a separate lineage that is distantly related to the type strain of *L. amylophilus* LMG 6900^T and to three other strains of the species (NRRL B-4438, NRRL B-4439 and NRRL B-4440). The *pheS* and *rpoA* gene sequence data showed that strains LMG 11400 and NRRL B-4435 constituted a distinct cluster, showing 100 % *pheS* and *rpoA* gene sequence similarities. The other reference strains clustered together with the type strain of *L. amylophilus* LMG 6900^T and were clearly differentiated from strains LMG 11400 and NRRL B-4435 (80 % and 89 % *pheS* and *rpoA* gene sequence similarities, respectively). Further phenotypic and genotypic research confirmed that both strains represent a novel taxon, for which the name *Lactobacillus amylotrophicus* has been proposed (Naser *et al.*, 2006b).

L. alimentarius species group

Within the *L. alimentarius* group, the *pheS* gene sequence similarity between *L. kimchii* and *L. paralimentarius* is 92 %, whereas on the basis of *rpoA* gene sequences, the two species show high relatedness, having 98.5 % *rpoA* gene sequence similarity. The *pheS* gene reflects a fast-evolving evolutionary clock that shows a finer resolution than the *rpoA* gene at both the intraspecies and interspecies levels in most cases. In support of the distinct genomic relatedness between *L. kimchii* LMG 19822^T and *L. paralimentarius* LMG 19152^T, De Vuyst *et al.* (2002) reported a DNA–DNA reassociation value of 68 %. Such a hybridization value is considered to be at the borderline for species delineation. The *pheS* gene sequence data indicates that *L. kimchii* and *L. paralimentarius* are separate species.

L. buchneri species group

Both *pheS* and *rpoA* gene sequence analyses showed that the members of *L. buchneri* species group are clustered in two subclades (see Supplementary Figs S1 and S2). An interesting relationship confirmed by the simultaneous analysis of *pheS* and *rpoA* gene sequences is the high genomic relatedness of *L. parabuchneri* LMG 11457^T and *L. ferintoshensis* LMG 22038^T (100 % *pheS* and *rpoA* gene sequence similarities). Recently published data are in complete accordance with the *pheS* and *rpoA* gene sequence data. Vancanneyt *et al.* (2005) confirmed this finding and demonstrated that these taxa are synonymous species, based on a polyphasic study.

Representative strains of *L. brevis*, LMG 6906^T, LMG 11435, LMG 7761, LMG 11494 and LMG 11984, were investigated. The *pheS* gene sequence analysis showed that strains LMG 11494 and LMG 11984 constituted a distinct cluster separated from the type strain of *L. brevis* with a sequence similarity of less than 82 % (see Supplementary Figs S1 and S2). 16S rRNA gene sequence analysis showed that both strains belong to the *L. buchneri* group with nearest neighbours *L. hammesii* and *L. brevis* (sequence similarities of 99.2 and 98.1 %, respectively). Strains LMG 11494 and LMG 11984, isolated from cheese and wheat, respectively, showed 99.9 % *pheS* gene sequence similarity. It has recently been confirmed that both strains represent a novel taxon, for which the name *L. parabrevis* was proposed (Vancanneyt *et al.*, 2006).

L. casei species group

Difficulties in the accurate identification of species belonging to the *L. casei* species group have been reported (Tynkkynen *et al.*, 1999; Zhong *et al.*, 1998). A study by Mori *et al.* (1997) found high 16S rRNA gene sequence similarity between the members of *L. casei* species group (>99 %). In the present study, *L. rhamnosus*, *L. casei* and *L. paracasei* were clearly distinguished on the basis of *pheS* and *rpoA* genes. Apart from *L. casei* and *L. zeae* (see below), these species have a maximum of 84 % and 95 % *pheS* and *rpoA* gene sequence similarities, respectively. This result further emphasizes the discriminatory power of the housekeeping genes investigated in this study.

Within the *L. casei* species group, the *pheS* gene sequence similarity between *L. casei* LMG 6904^T (=ATCC 393^T) and *L. zeae* LMG 17315^T (=ATCC 158520^T) was 93 %, whereas on the basis of *rpoA* gene sequences, the two species were more highly related, having 99 % gene sequence similarity. In addition, the sequence analysis of the gene that codes for the α-subunit of ATP synthase (*atpA*) also showed a high relatedness (96 %) between the two species (data not shown). Data from the literature were in complete accordance with the present data and supported the high relatedness found between these two taxa. Further genomic data derived from *recA* gene sequence analysis and high DNA–DNA reassociation values (80 %) demonstrated that both species are members of the same species (Dicks *et al.*, 1996; Felis *et al.*, 2001) and supported the reclassification of *L. casei* as *L. zeae* (Dellaglio *et al.*, 2002). This example strongly supports the simultaneous use of multiple loci.

L. plantarum species group

16S rRNA gene sequences are not suitable for definitive differentiation of the members of *L. plantarum* species group due to the high gene sequence similarity (>99 %) between *L. plantarum, L. paraplantarum* and *L. pentosus* (Collins *et al.*, 1991; Torriani *et al.*, 2001). Our data clearly showed that *pheS* and *rpoA* gene sequences had a high discriminatory power in differentiating *L. plantarum, L. paraplantarum* and *L. pentosus* with a maximum 90 % and 98 % *pheS* and *rpoA* gene sequence similarities, respectively. At the subspecies level, the neighbour-joining tree based on the *pheS* gene sequences showed that *L. plantarum* subsp. *plantarum* and *L. plantarum* subsp. *argentoratensis* were clearly differentiated from each other (91 % *pheS* gene sequence similarity) (see Supplementary Fig. S1). *L. plantarum* LMG 6907^T and *L. arizonensis* LMG 19807^T were highly related with >99.5 % *pheS* and *rpoA* gene sequence similarity. Kostinek *et al.* (2005) showed that *L. arizonensis* is a later heterotypic synonym of *L. plantarum* because the type strain of *L. arizonensis* NRRL B-14768^T (=DSM 13273^T) is not distinguishable from the *L. plantarum* type strain DSM 20174^T on the basis of ribotyping patterns, rep-PCR fingerprinting patterns, 16S rRNA gene sequences or DNA–DNA hybridization data.

L. reuteri species group

Within this species group, high degrees of similarity exist between *L. ingluviei* LMG 20380^T and *L. thermotolerans* LMG 22056^T (99 % and 100 % *pheS* and *rpoA* gene sequence similarities, respectively) as well as between *L. durianis* LMG 19193^T and *L. vaccinostercus* LMG 9215^T (99 % and 98 % *pheS* and *rpoA* gene sequence similarities, respectively). A study recently conducted by Felis *et al.* (2006) confirmed that *L. thermotolerans* is a later synonym of *L. ingluviei*. Representative strains of *L. durianis* and *L. vaccinostercus* were further investigated. Genomic data derived from FAFLP and DNA–DNA hybridizations, respectively, has provided evidence for the reclassification of *L. durianis* as *L. vaccinostercus* (Dellaglio *et al.*, 2006).

On the other hand, the neighbour-joining tree based on *pheS* gene sequences revealed heterogeneity between strains of *L. reuteri*. As mentioned earlier, the *pheS* gene reflects a fast-evolving evolutionary clock that shows a finer resolution, in most cases, than the *rpoA* gene at both the intraspecies and interspecies levels.

L. sakei species group

L. sakei and *L. curvatus* have >99 % 16S rRNA gene sequence similarity; the corresponding *pheS* and *rpoA* gene sequence similarities were 88 % and 96 %. At the subspecies level, the neighbour-joining tree based on the *pheS* gene sequences showed that *L. sakei* subsp. *sakei* and *L. sakei* subsp. *carnosus* were clearly differentiated from each other (92 % *pheS* gene sequence similarity) (see Supplementary Fig. S1).

L. salivarius species group

The *pheS* neighbour-joining tree split this species group into two subclusters (see Supplementary Fig. S1). An interesting relationship detected by the simultaneous analysis of *pheS* and *rpoA* gene sequences is the high genomic relatedness of the *L. cypricasei* and *L. acidipiscis* type strains. *L. acidipiscis* strains (LMG 19820^T and LMG 23135) and the strains of *L. cypricasei* (LMG 21592^T, CCUG 42959, CCUG 42960 and CCUG 42962) revealed 99.8–100 % *pheS* and *rpoA* gene sequence similarities. Sequence analysis of the *atpA* gene also showed a high relatedness (>99 %) between the two species (data not shown). High DNA–DNA reassociation values confirmed that *L. cypricasei* is a later synonym of *L. acidipiscis* (Naser *et al.*, 2006c).

In addition, whereas the type strains of *L. animalis* and *L. murinus* are separated by their 16S rRNA gene sequences, these two species are highly related on the basis of their *pheS* and *rpoA* gene sequences (Fig. 2). The type strains of *L. animalis* and *L. murinus* occupied a distinct subcluster having 98.5 % *pheS* and *rpoA* gene sequence similarities.

Other *Lactobacillus* species

The type strains of *L. fructivorans* and *L. homohiochii* showed a high degree of similarity (100 % *pheS* and *rpoA* gene sequence similarities). Further taxonomical studies are needed to clarify their relatedness.

Conclusions

It is now generally accepted that a correct classification should reflect the natural relationships as encoded in the DNA and consequently genotypic methods are considered of paramount importance to modern taxonomy. The use of several housekeeping genes in bacterial taxonomy is best suited for analysis at the species and genus levels as it integrates the information of different molecular clocks around the bacterial chromosome (Gevers *et al.*, 2005; Stackebrandt *et al.*, 2002; Zeigler, 2003).

Our data convincingly prove that the simultaneous analysis of *pheS* and *rpoA* partial gene sequences provide an alternative tool for the rapid and reliable identification of different species of the genus *Lactobacillus*. The analysis of *pheS* and *rpoA* gene sequences effectively allows closely related *Lactobacillus* species to be differentiated at a higher discrimination level than that possible with 16S rRNA gene sequence comparisons.

The fact that within species groups, different genes may yield different tree topologies does not hamper their use to unambiguously assign isolates to a particular species. Several factors account for the different topologies determined for different housekeeping genes, i.e. the level of the information content, the different rates of evolution due to different selection forces on various genes and the length of the partial sequences that are compared

(Christensen *et al.*, 2004). The variation in the discriminatory power of the investigated genes, together with the fact that different genes might provide different closest neighbours or tree topologies, has highlighted the necessity for simultaneous analysis of several protein-coding loci for a robust identification analysis.

We intend to contribute to the present identification system by the construction of a central, curated database in which data can be stored and accessed freely online. This is expected to contribute in the long run to the improvement of a better species definition for the genus *Lactobacillus*. The system is rapid, highly reproducible, portable and provides adequate resolution power. In addition, we further intend to extend this system to include all other genera of LAB.

ACKNOWLEDGEMENTS

S. M. N. acknowledges a PhD scholarship from the Ministry of Education and Higher Education, Palestine. J. S. and D. G. acknowledge grants from the Fund for Scientific Research (FWO), Belgium. We thank Leentje Christiaens and Marjan De Wachter for their technical assistance.

REFERENCES

Christensen, H., Kuhnert, P., Olsen, J. E. & Bisgaard, M. (2004). Comparative phylogenies of the housekeeping genes *atpD, infB, rpoB* and the 16S rRNA gene within the *Pasteurellaceae. Int J Syst Evol Microbiol* **54**, 1601–1609.

Coenye, T., Gevers, D., Van de Peer, Y., Vandamme, P. & Swings, J. (2005). Towards a prokaryotic genomic taxonomy. *FEMS Microbiol Rev* **29**, 147–167.

Collins, M. D., Rodrigues, U., Ash, C., Aguirre, M., Farrow, J. A. E., Martinez-Murcia, A., Phillips, B. A., Williams, A. M. & Wallbanks, S. (1991). Phylogenetic analysis of the genus *Lactobacillus* and related lactic acid bacteria as determined by reverse transcriptase sequencing of 16S rRNA. *FEMS Microbiol Lett* **77**, 5–12.

Daud Khaled, D., Neilan, B. A., Henriksson, A. & Conway, P. L. (1997). Identification and phylogenetic analysis of *Lactobacillus* using multiplex RAPD-PCR. *FEMS Microbiol Lett* **153**, 191–197.

De Vuyst, L., Schrijvers, V., Paramithiotis, S., Hoste, B., Vancanneyt, M., Swings, J., Kalantzopoulos, G., Tsakalidou, E. & Messens, W. (2002). The biodiversity of lactic acid bacteria in Greek traditional wheat sourdoughs is reflected in both composition and metabolite formation. *Appl Environ Microbiol* **68**, 6059–6069.

Dellaglio, F., Felis, G. E. & Torriani, S. (2002). The status of the species *Lactobacillus casei* (Orla-Jensen 1916) Hansen and Lessel 1971 and *Lactobacillus paracasei* Collins *et al.* 1989. Request for an opinion. *Int J Syst Evol Microbiol* **52**, 285–287.

Dellaglio, F., Vancanneyt, M., Endo, A, Vandamme, P., Felis, G. E., Castioni, A., Fujimoto, J., Watanabe, K. & Okada, S. (2006). *Lactobacillus durianis* Leisner *et al.* 2002 is a later heterotypic synonym of *Lactobacillus vaccinostercus* Kozaki and Okada 1983. *Int J Syst Evol Microbiol* **56**, 1721–1724.

Dicks, L. M., Du Plessis, E. M., Dellaglio, F. & Lauer, E. (1996). Reclassification of *Lactobacillus casei* subsp. *casei* ATCC 393 and *Lactobacillus rhamnosus* ATCC 15820 as *Lactobacillus zeae* nom. rev., designation of ATCC 334 as the neotype of *L. casei* subsp. *casei*, and rejection of the name *Lactobacillus paracasei. Int J Syst Bacteriol* **46**, 337–340.

Felis, G. E., Dellaglio, F., Mizzi, L. & Torriani, S. (2001). Comparative sequence analysis of a *recA* gene fragment brings new evidence for a change in the taxonomy of the *Lactobacillus casei* group. *Int J Syst Evol Microbiol* **51**, 2113–2117.

Felis, G. E., Vancanneyt, M., Snauwaert, C., Swings, J., Torriani, S. & Castioni, A.andDellaglio, F. (2006). Reclassification of *Lactobacillus thermotolerans* Niamsup *et al.* 2003 as a later synonym of *Lactobacillus ingluviei* Baele *et al.* 2003. *Int J Syst Evol Microbiol* **56**, 793–795.

Gancheva, A., Pot, B., Vanhonacker, K., Hoste, B. & Kersters, K. (1999). A polyphasic approach towards the identification of strains belonging to *Lactobacillus acidophilus* and related species. *Syst Appl Microbiol* **22**, 573–585.

Gevers, D., Huys, G. & Swings, J. (2001). Applicability of rep-PCR fingerprinting for identification of *Lactobacillus* species. *FEMS Microbiol Lett* **205**, 31–36.

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

Hammes, W. P. & Vogel, R. F. (1995). The genus *Lactobacillus*. In *The Genera of Lactic Acid Bacteria*, pp. 19–54. Edited by B. J. B. Wood & W. H. Holzapfel. London: Blackie Academic & Professional.

Holzapfel, W. H., Haberer, P., Geisen, R., Bjorkroth, J. & Schillinger, U. (2001). Taxonomy and important features of probiotic microorganisms in food and nutrition. *Am J Clin Nutr* **73**, 365S–373S.

Kandler, O. & Weiss, N. (1986). Genus *Lactobacillus* Beijerinck 1901, 212. In *Bergey's Manual of Systematic Bacteriology,* vol. 2, pp. 1209−1234. Edited by P. H. A. Sneath, N. S. Mair & J. G. Holt. Baltimore, USA: The Williams and Wilkins Co.

Klein, G., Pack, A., Bonaparte, C. & Reuter, G. (1998). Taxonomy and physiology of probiotic lactic acid bacteria. *Int J Food Microbiol* **41**, 103–125.

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

Kostinek, M., Pukall, R., Rooney, A. P., Schillinger, U., Hertel, C., Holzapfel, W. H. & Franz, C. M. A. P. (2005). *Lactobacillus arizonensis* is a later heterotypic synonym of *Lactobacillus plantarum. Int J Syst Evol Microbiol* **55**, 2485–2489.

Massi, M., Vitali, B., Federici, F., Matteuzzi, D. & Brigidi, P. (2004). Identification method based on PCR combined with automated ribotyping for tracking probiotic *Lactobacillus* strains colonizing the human gut and vagina. *J Appl Microbiol* **96**, 777–786.

Meyer, C. P. & Paulay, G. (2005). DNA barcoding: error rates based on comprehensive sampling. *PLoS Biol* **3** (*e422*).

Mori, K., Yamazaki, K., Ishiyama, T., Katsumata, M., Kobayashi, K., Kawai, Y., Inoue, N. & Shinano, H. (1997). Comparative sequence analyses of the genes coding for 16S rRNA of *Lactobacillus casei*-related taxa. *Int J Syst Bacteriol* **47**, 54–57.

Nakamura, L. K. & Crowell, C. D. (1979). *Lactobacillus amylophilus*, a new starch-hydrolyzing species from swine waste-corn fermentation. *Dev Ind Microbiol* **20**, 532–540.

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., Hagen, K. E., Vancanneyt, M., Cleenwerck, I., Swings, J. & Tompkins, T. A. (2006a). *Lactobacillus suntoryeus* Cachat and Priest 2005 is a later synonym of *Lactobacillus helveticus* (Orla-Jensen 1919) Bergey *et al.* 1925 (Approved Lists 1980). *Int J Syst Evol Microbiol* **56**, 355–360.

Naser, S. M., Vancanneyt, M., Snauwaert, C., Vrancken, G., Hoste, B., De Vuyst, L. & Swings, J. (2006b). Reclassification of *Lactobacillus amylophilus* LMG 11400 and NRRL B-4435 as *Lactobacillus amylotrophicus* sp. nov. *Int J Syst Evol Microbiol* **56**, 2523–2527.

Naser, S. M., Vancanneyt, M., Hoste, B., Snauwaert, C. & Swings, J. (2006c). *Lactobacillus cypricasei* Lawson *et al.* 2001 is a later synonym of *Lactobacillus acidipiscis* Tanasupawat *et al.* 2000. *Int J Syst Evol Microbiol* **56**, 1681–1683.

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

Pot, B., Hertel, C., Ludwig, W., Descheemaeker, P., Kersters, K. & Schleifer, K. H. (1993). Identification and classification of *Lactobacillus acidophilus, L. gasseri* and *L. johnsonii* strains by SDS-PAGE and rRNA-targeted oligonucleotide probe hybridization. *J Gen Microbiol* **139**, 513–517.

Reid, G. (1999). The scientific basis for probiotic strains of *Lactobacillus. Appl Environ Microbiol* **65**, 3763–3766.

Saitou, N. & Nei, M. (1987). The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406–425.

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

Schleifer, K. H. & Ludwig, W. (1995). Phylogeny of the genus *Lactobacillus* and related genera. *Syst Appl Microbiol* **18**, 461–467.

Sneath, P. H. A. & Sokal, R. R. (1973). Numerical Taxonomy. *The Principles and Practice of Numerical Classification.* San Francisco: W. H. Freeman and Co.

Stackebrandt, E., Frederiksen, W., Garrity, G. M., Grimont, P. A., Kämpfer, P., Maiden, M. C., Nesme, X., Rossello-Mora, R., Swings, J. & 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.

Stiles, M. E. & Holzapfel, W. H. (1997). Lactic acid bacteria of foods and their current taxonomy. *Int J Food Microbiol* **36**, 1–29.

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.

Torriani, S., Felis, G. E. & Dellaglio, F. (2001). Differentiation of *Lactobacillus plantarum, L. pentosus*, and *L. paraplantarum* by *recA* gene sequence analysis and multiplex PCR assay with *recA* gene-derived primers. *Appl Environ Microbiol* **67**, 3450–3454.

Tynkkynen, S., Satokari, R., Saarela, M., Mattila-Sandholm, T. & Saxelin, M. (1999). Comparison of ribotyping, randomly amplified polymorphic DNA analysis, and pulsed-field gel electrophoresis in typing of *Lactobacillus rhamnosus* and *L. casei* strains. *Appl Environ Microbiol* **65**, 3908–3914.

Vancanneyt, M., Engelbeen, K., De Wachter, M., Vandemeulebroecke, K., Cleenwerck, I. & Swings, J. (2005). Reclassification of *Lactobacillus ferintoshensis* as a later heterotypic synonym of *Lactobacillus parabuchneri. Int J Syst Evol Microbiol* **55**, 2195–2198.

Vancanneyt, M., Naser, S. M., Engelbeen, K., De Wachter, M., Van der Meulen, R., Cleenwerck, I., Hoste, B., De Vuyst, L. & Swings, J. (2006). Reclassification of *Lactobacillus brevis* LMG 11494 and LMG 11984 as *Lactobacillus parabrevis* sp. nov. *Int J Syst Evol Microbiol* **56**, 1553–1557.

Yansanjav, A., Svec, P., Sedlacek, I., Hollerova, I. & Nemec, M. (2003). Ribotyping of lactobacilli isolated from spoiled beer. *FEMS Microbiol Lett* **229**, 141–144.

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

Zhong, W., Millsap, K., Bialkowska-Hobrzanska, H. & Reid, G. (1998). Differentiation of *Lactobacillus* species by molecular typing. *Appl Environ Microbiol* **64**, 2418–2423.