TOWARDS UNDERSTANDING IN MOLECULAR TAXONOMY USING AN *In Silico* APPROACH: A CASE STUDY IN LACTIC ACID BACTERIA

Komsit Wisitrasamewong[1], Jedsadapa Jongmahasavat[1], Saisamorn Lumyong[2] and Ekachai Chukeatirote[1*]

Received: Nov 1, 2008; Revised: Jan 13, 2009; Accepted: Jan 15, 2009

Abstract

In this paper, we propose an *in silico* analysis using the lactic acid bacteria (LAB) to gain understanding in their diversity, detection, identification and classification. Initially, the 16S rRNA gene sequences of 90 LAB species retrieved from the GenBank database were analysed in terms of their similarity and variation using the ClustalW software. The phylogenetic tree was then constructed in order to investigate their evolutionary relationship. It was found that, according to the tree, the LAB could be grouped into 6 distinct clades similar to the genus classification except the genus *Lactobacillus*. In addition, these results obtained were in agreement with the data derived from traditional identification. The probes were also identified based on the multiple sequence alignment results to identify the LAB at different levels such as group, genus and species. Furthermore, the *in silico* PCR-RFLP was performed to reveal their taxonomic classification. Various kinds of restriction endonucleases were selected and used to restrict the LAB 16S rRNA genes to observe their restriction profiles. It was found that *Bsa*WI, *Bst*DSI and *Dsa*I produced distinct restriction patterns suitable for genus and species determination.

Keywords: *in silico*, molecular taxonomy, lactic acid bacteria

Introduction

Lactic acid bacteria (LAB), albeit used as a loosely defined term, are referred to a related group of bacteria that share the property of producing lactic acid as the principal end-product from hexoses. Nevertheless, it is agreeable that LAB are Gram-positive, non-spore forming, catalase negative which consist of both cocci and bacilli forms. Typical LAB are acid tolerant, fastidious and grow under microaerophilic to obligately anaerobic conditions. Nowadays, it is generally acknowledged that LAB consist of a wide range of genera including *Bifidobacterium*, *Lactobacillus*, *Lactococcus*, *Enterococcus*, *Streptococcus*, *Pediococcus*, *Leuconostoc*, and *Weissella* (Stiles and Holzapfel, 1997). They can be isolated from soil, water, silage, and fermented food products. In addition, several LAB species are part of the microbiota in the gastro-intestinal

[1] *School of Science, Mae Fah Luang University, Chiang Rai ,57100, Thailand. E-mail: ekachai@mfu.ac.th*
[2] *Department of Biology, Faculty of Science, Chiang Mai University, Chiang Mai, 50200, Thailand.*
[*] *Corresponding author*

tract of man and animal. Particular attention has been focused on LAB because the group has been shown to produce distinct metabolites that can be applied in various applications including food, medicine and industry. Due to their long history of safe use, known as GRAS (Generally Regarded As Safe) status, a number of LAB species have been used as starter culture in many kinds of fermented foods (Lee, 1997; Awad *et al*., 2007; de Vuyst *et al*., 2008). Besides, a group of metabolites produced by LAB which include organic acids, diacetyl, acetoin, hydrogen peroxide, antibiotics and bacteriocins has been shown to contribute in improving the quality of fermented foods possibly by preventing pathogens, extending shelf life and improving sensory qualities (Hannon *et al*., 2007; Gerez *et al*., 2009). Some strains of LAB are also currently used as probiotics (Chukeatirote, 2003; Saito, 2004; Ljungh and Wadstrom, 2006). Last but not least, biodegradable plastic of polylactide (PLA) can be generated from food waste using LAB and this achievement would definitely make a major impact on environmental issues (Sakai *et al*., 2004).

Owing to their significance, identification of the correct LAB species (or even a particular strain) is thus crucial especially when considered from the safety and legislative (i.e., labelling and patenting) issues. However, routine identification of LAB in most laboratories still depends on traditional techniques. This includes morphological, physiological and biochemical assays in which the results obtained are often ambiguous, laborious, time-consuming and, more importantly, may lead to misidentifications (Bonomo *et al*., 2008; Fontana *et al*., 2005). During the past two decades, molecular means especially those relied on rRNA gene sequence have been shown to be a powerful tool for detection and identification (Amann *et al*., 1995; Head *et al*., 1998). These techniques are not only simple and rapid but have also successfully overcome the limitations of the traditional methods. It is therefore not surprising that a number of articles have been reported on molecular identification and characterisation of LAB (reviewed in Amor *et al*., 2007; Ehrmann and Vogel, 2005; Temmerman *et al*., 2004). Unfortunately, these published results appear to be specific and focus on probiotic species (including those are of great importance in industry only). It should also be noted that the molecular techniques used for taxonomic identification still remains unclear to most biologists. Additionally, regarding molecular taxonomy, some scientists tend to think of advanced technology that has to deal with special expertise and state-of-art scientific instruments. As a result, these molecular techniques although originated nearly thirty years ago are not widespread and thus limited to certain groups of scientists. This paper is therefore an attempt to introduce the concept of molecular taxonomy using the LAB group as a case study. A series of molecular techniques were performed *in silico* to deal with the LAB diversity and their phylogenetic relationship.

Methodology

As shown in Table 1, representatives of LAB species were collected from the DSMZ's catalogue (DSMZ, 2008). Their 16S rRNA gene sequences were then retrieved from the GenBank database (Benson *et al*., 2008) and aligned by ClustalW software in which the phylogenetic tree used to classify the evolutionary relationships was also generated (Larkin *et al*., 2007). Further analysis was then performed by applying these data to the principles of molecular techniques related to taxonomic identification. Sequences of highly variable regions were screened and selected to design probes specific at different levels (i.e., species and genus). The probe specificity was evaluated *in silico* using the Probe Match software available from the Ribosomal Database Project (RDP-II) (Cole *et al*., 2005). In addition, the PCR technique was also introduced by locating the identical areas at the 5'- and 3'-ends of the 16S rRNA sequences. Such identical areas referred to 'designed' primers were flanked the nucleotide regions from which their whole internal sequences were represented as 'amplified' fragments. For *in silico* PCR-RFLP (polymerase chain reaction-restriction fragment

length polymorphism), such 'amplified' fragments were subjected to various kinds of restriction endonuclease enzymes using the Webcutter 2.0 software (Heiman, 1997). The RFLP diagram was then established using Microsoft Excel.

Results and Discussion

The LAB group is heterogenous consisting of several genera. To reveal their diversity and evolutionary relationship, we attempted to use molecular tools based on sequence comparisons of the 16S rRNA genes. The use of rRNA gene in studying phylogenetic relationship of living organisms is now widely approved because of several beneficial features such as presence in all cellular life forms, conserved sequence domains and functions and availability of databases (Amann *et al*., 1995). It should be also mentioned that, in this study, an *in silico* approach was performed and thus not required any expensive materials (i.e., chemicals and enzymes) and analytical instruments.

Initially, ninety LAB species were selected and their 16S rRNA gene sequences were retrieved from the GenBank database as shown in Table 1. The 16S rRNA gene sequences of *Escherichia coli* and *Pseudomonas fluorescens* were also used for outgroup analysis. These 16S rRNA sequences were then subjected to multiple sequence analysis using ClustalW software in which the dendogram was also generated (Figure 1). According to Figure 1, the LAB species could be grouped into six different clades: L1 (*Lactobacillus* spp.), L2 (*Bifidobacterium* spp.), L3 (*Lactococcus* spp.), L4 (*Leuconostoc* spp.), L5 (*Weissella* spp.) and L6. It should be noted however that the clade L6 was polyphyletic consisting of the genera *Lactobacillus* and *Pediococcus*. These data were then used to determine if there was any relationship to morphological, physiological and biochemical data as illustrated in Table 2. The interpretation of both data sets appeared to be in agreement.

An *in situ* hybridisation technique was then introduced to further explain the principles and applications of molecular taxonomy in bacterial identification. Although the 16S rRNA gene sequence is well known to have highly conserved structures, it contains some variable regions that can be used as specific probes for detection and identification. Such sequences are useful and can be applied for genus and/or species identification. In this study, a number of probes could be designed and they were useful at the group (3 probes), genus (30 probes), and species (209 probes) levels (not all data were shown). Some examples of these specific probes are shown in Table 3.

Additionally, we also used PCR-RFLP with an expectation that this technique could help generate distinct 'amplified' DNA profiles for LAB discrimination. For this, the universal primers previously reported (Marchesi *et al*., 1998) were used to *in silico* amplify the internal sequence. The amplicons obtained were then subjected to digestion by several restriction endonucleases using the Webcutter software. A few examples of such PCR-RFLP profiles are shown in Figure 2. These *in silico* results are in agreement with previously published work (Aymerich *et al*., 2006; Rantsiou *et al*., 2006).

Similar to other microbes, the identification of LAB species has previously been performed using traditional approach which is based on their morphological and biochemical characteristics. However, it has been shown that these analyses are inconsistent and often give different biochemical results depending on strains and variations. Currently, several molecular techniques such as RAPD (Random Amplificaton of Polymorphic DNA), RFLP (Restriction Fragment Length Polymorphism) and ARDRA (Amplified rDNA Restriction Analysis) have been established and used in microbial identification and characterisation. The advantages and disadvantages between the two approaches have been extensively discussed (Amor *et al*., 2007; Temmerman *et al*., 2004). For example, traditional biochemical tests can be carried out easily in any laboratory but the technique is laborious (due to many media and biochemical reactions) and time-consuming. Besides, variations in biochemical profiles within

Table 1. Lists of LAB species used in this study

Genus	Species	Strain	Accession no.
Lactobacillus	*Lb. acetotolerans*	DSM 20749	M58801
	Lb. acidophilus	BCRC 10695	AY773947
	Lb. alimentarius	DSM 20249	M58804
	Lb. amyloticus	LA 5	Y17361
	Lb. amylophilus	DSM 20533	M58806
	Lb. amylovorus	DSM 20531	M58805
	Lb. animalis	DSM 20602	M58807
	Lb. bifermentans	DSM 20003	M58809
	Lb. brevis	ATCC 14869	M58810
	Lb. casei	DSM 20021	M58815
	Lb. coryneformis	DSM 20001	M58813
	Lb. curvatus	DSM 20010	AJ270951
	Lb. delbrueckii	DSM 20074	M58814
	Lb. fermentum	ATCC 14931	M58819
	Lb. fructivorans	DSM 20203	M58818
	Lb. gastricus	LMG 22113T	AY253658
	Lb. helveticus	NCDO 2712	X61141
	Lb. intestinalis	DSM 6629T	AJ306299
	Lb. kefiranofaciens subsp. *kefirgranum*	CECT 5919T	AJ575742
	Lb. kefiranofaciens subsp. *kefiranofaciens*	R-14703	AJ575260
	Lb. kefiri	NRIC 1693	AB024300
	Lb. kimchii	AP1077	AF183558
	Lb. malefermentans	CECT 5928T	AJ575743
	Lb. manihotivorans	YAM I	AF000163
	Lb. mucosae	DSM 13345	AF126738
	Lb. oligofermentans	533	AY733085
	Lb. Paracasei	LPC1	AY675255
	Lb. parakefiri	LMG 15133T	AY026750
	Lb. paralimentarius	ACA-DC 3415	AJ422036
	Lb. Paraplantarum	DSM 10667T	AJ306297
	Lb. pentosus	JCM 1558	D79211
	Lb. plantarum	DSM 20205	M58827
	Lb. rhamnosus	LR 2	AY675254
	Lb. ruminis	DSM 20403	M58828
	Lb. sakei	DSM 20017	M58829
	Lb. salivarius subsp. *salicinius*	DSM 20554	M59054
	Lb. salivarius subsp. *salivarius*	ATCC 11741	AF089108
	Lb. sanfranciscensis	ATCC 27651	X76327
	Lb. thermotolerans	G22	AF308147
	Lb. vaganalis	NCTC 12197	X61136
Bifidobacterium	*B. angulatum*	ATCC 27535	D86182
	B. animalis	Bb 12	AB027536
	B. asteroids	ATCC 29510	M58730
	B. bifidum	BF2	AY694148
	B. boum	JCM 1211	D86190
	B. breve	BR2	AY735402
	B. catenulatum	ATCC 27539	M58732
	B. choerinum	ATCC 27686	D86186
	B. coryneforme	ATCC 25911	M58733
	B. cunniculi	ATCC 27916	M58734
	B. dentium	ATCC 15423	M58735
	B. gallicum	JCM 8224	D86189
	B. gallinarum	JCM 6291	D861191
	B. indicum	ATCC 25912	M58737
	B. longum	ATCC 15707	M58739
	B. magnum	ATCC 27540	M58740
	B. mininum	ATCC 27538	M58741

Genus	Species	Strain	Accession no.
	B. pseudolonggum	ATCC 25526	M58742
	B. thermophilum	ATCC 25525	AB026246
Leuconostoc	*Ln. argentinum*	DSM 8581	AF175403
	Ln. carnosum	NCFB 2776	X95977
	Ln. citreum	NCFB 2787	X53963
	Ln. fallax	DSM 20189	S63851
	Ln. ficulneus	FS-1	AF360736
	Ln. fructosum	NCDO 2345	X61140
	Ln. Garlicum	-	AY456086
	Ln. gasicomitatum	LMG 18811	AF231131
	Ln. gelidum	JCM 10093	AB004661
	Ln. inhae	LM 2630	AY675244
	Ln. kimchii	IH25	AF173986
	Ln. lactis	DSM 20202	M23031
	Ln. mesenteroides	DSM 20343	M23035
	Ln. pseudoficulneum	LC51	AY169967
	Ln. pseudomesenteroides	NCDO768	X95979
Pediococcus	*Pc. acidilactici*	NCDO 2767	X95976
	Pc. damnosus	DSM 20331	AJ318414
	Pc. dextrinicus	JCM 5887	D87679
	Pc. inopinatus	DSM 20285	AJ271383
	Pc. parvulus	JCM 5889	D88528
	Pc. pentosaceus	RO95	AF515227
	Pc. rinaeequi	IFO12173	D87677
Lactococcus	*Lc. garvieae*	MR1	AF283499
	Lc. raffinolactis	NCDO 2118	X54260
	Lc. piscium	HR1A-68	X53905
	Lc. Lactis	NCDO 617	X54261
Weissella	*W. cibaria*	PL9001	AF477495
	W. confuse	LMG 17699T	AJ295989
	W. halotolerans	DSM 20190	M23037
Streptococcus	*S. thermophilus*	DSM 20617	X68418

bacterial strains may lead to misidentification. In contrast, the molecular techniques offer a rapid means and the results are derived from the genetic information which is not affected by environmental factors (i.e., culture conditions).

Conclusions

This present paper illustrates the use of molecular techniques for LAB identification. Interestingly, a series of molecular means introduced in this study appear to be a good case providing a general and detailed background for biologists who have little experience in molecular techniques. Due to several beneficial properties, the rRNA gene sequence is selected and used as representatives in this study. However, other housekeeping genes (i.e., RecA, gyrD and 6-phosphogluconate dehydrogenase) can also be used in introducing the principle of molecular techniques and it is interesting to compare such results in terms of the LAB phylogenetic relationship. Furthermore, it should be emphasised that an *in silico* approach is performed and thus overcomes the limitation of expensive chemicals and advanced analytical instruments.

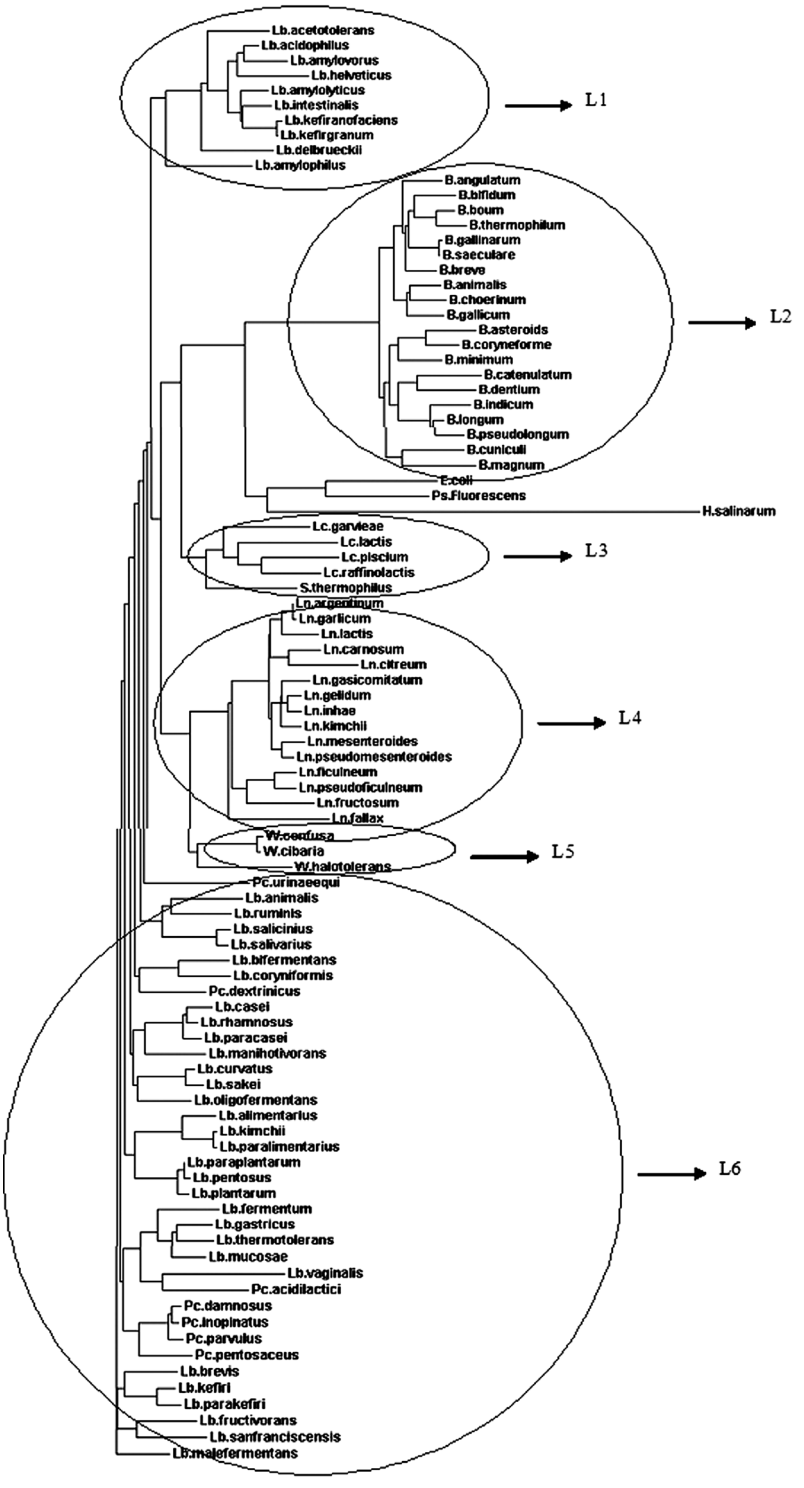

Figure 1. Phylogenetic tree of LAB species based on their 16S rRNA gene similarity

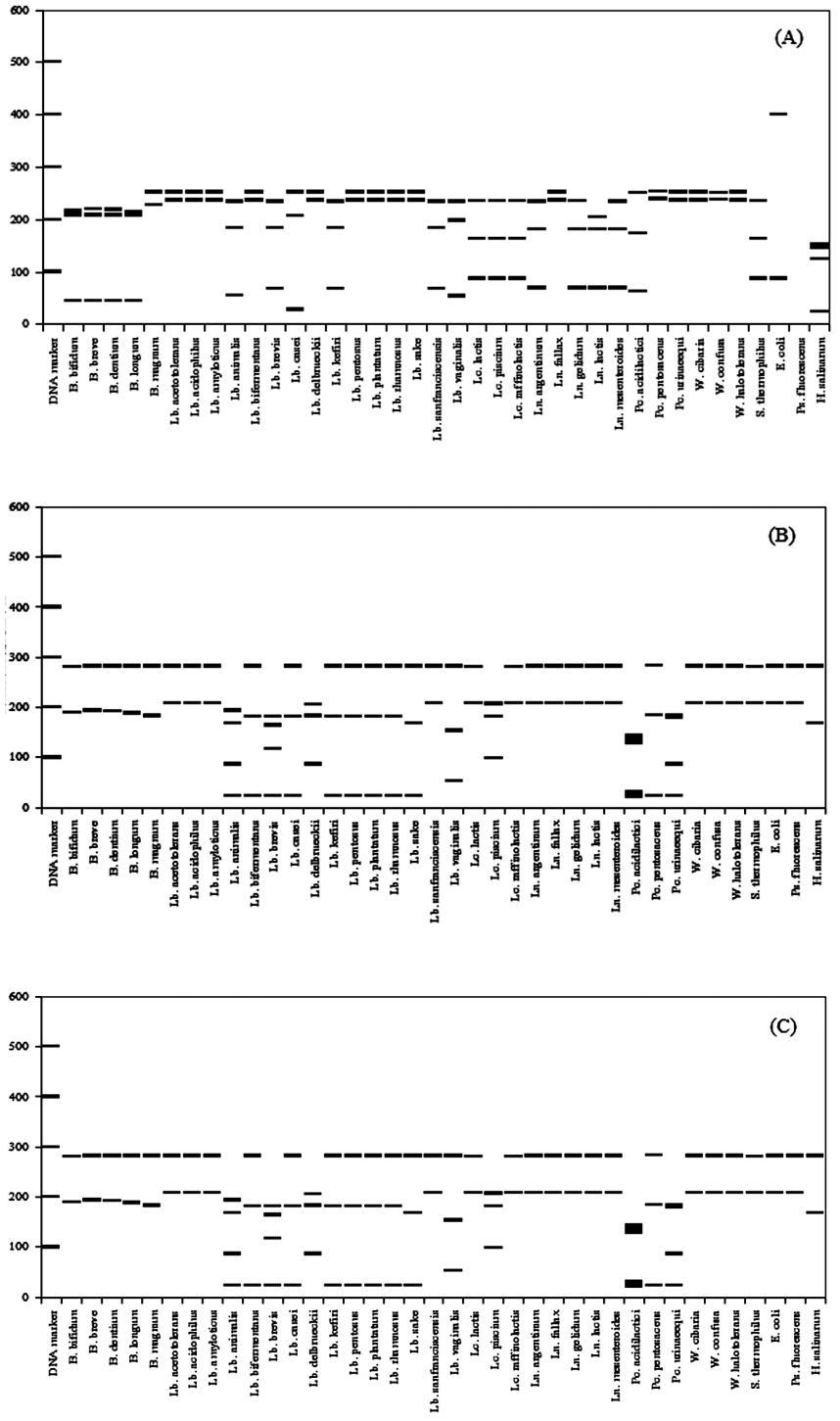

Figure 2. Representatives of *in silico* analysis of LAB 16S rRNA genes using *Bsa*WI (A), *Bst*DSI (B) and *Dsa*I (C)

Table 2. Some morphological, physiological and biochemical features of LAB species used in this study

Characteristics	L1	L2	L3	L4	L5	L6
Cell morphology	Rod	Rod	Cocci	Lenticular cocci	Irregular cocci / rod	Rod / cocci
%G+C	32 - 51	55 - 67	37 - 44	38 - 44	44 - 45	34 - 54[a]
Optimal T for growth	28 - 45	37 - 41	30 - 45	20 - 30	30 - 37	30 - 45
Motility	-	-	-	-	-	-[b]
Oxygen requirement[c]	F	A	F	F	F	F
Fermentation[d]	FH, HO	HE	HO	HE	HE	FH, HO, HE
Lactate isomer	D / L / DL	D / L	L	D	D / DL	D / L / DL
Catalase test	-	-[e]	-	-	-	-[f]
Nitrate reduction	-	-	-	-	-	-

Notes: [a]%G+C of *Lb. fermentum* is higher than 50% (52 - 54).
[b]*Lb. ruminis* is an exception possessing peritrichous flagella.
[c]F = Facultative anaerobe; A = Obligate anaerobe.
[d]FH = Facultative heterofermentative; HE = Heterofermentative; HO = Homofermentative.
[e]*B. indicum* and *B. asteroids* are excluded.
[f]*Pc. acidilactici* and *Pc. pentosaceus* are excluded.

Table 3. Representatives of specific probes of LAB species

Code no.	Probe sequence (5'-3')	Nucleotide position[a]	Genus[b]		
			Group	Specificity	Species[c]
KC001	CACCGCTACACATGGAGTTCCAC	664-686	+	-	-
KC007	GCGATGGACTTTCACACC	600-617	-	B	-
KC013	CCATACAACAGT	434-445	-	Ln	-
KC024	CCACTCTCACAC	454-465	-	S	-
KC031	CCGATGCACTTC	627-638	-	Lb	-
KC035	ATCGGGATGTCAAGAGG	986-1001	-	-	+
KC059	CCTTTTATAAGCTGA	191-202	-	-	+
KC098	TCATCCAGAAGTGATAG	836-847	-	-	+
KC216	CCCCACCGTCAAGCTG	244-259	-	-	+

Notes: [a]Positions of nucleotides in accordance with those of *Escherichia coli* 16S rRNA gene.
[b]B = *Bifidobacterium*; Ln = *Leuconostoc*; S = *Streptococcus*; Lb = *Lactobacillus*.
[c]KC035 for *S. thermophilus* DSM 20617; KC059 for *Lb. acidophilus* NCDO 1748 and ATCC 4356; KC098 for *Lactobacillus brevis* ATCC 1486, NRIC 1684, str. K9 and L63; KC216 for *B. indicum* ATCC 25912

References

Amann, R.I., Ludwig, W., and Schleifer, K.H. (1995). Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. Microbiological Reviews, 59:143-169.

Amor, K.B., Vaughan, E.E., and de Vos, W.M. (2007). Advanced molecular tools for the identification of lactic acid bacteria. Journal of Nutrition, 137:741S-747S.

Awad, S., Ahmed, N., and Soda, M.E. (2007). Evaluation of isolated starter lactic acid bacteria in Ras cheese ripening and flavour development. Food Chemistry, 104:1,192-1,199.

Aymerich, T., Martin, B., Garriga, M., Vidal-Carou, M.C., Bover-Cid, S., and Hugas, M. (2006). Safety properties and molecular strain typing of lactic acid bacteria from slightly fermented sausages. Journal of Applied Microbiology, 100:40-49.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Wheeler, D.L. (2008). GenBank. Nucleic Acids Research, 36: D25-D30.

Bonomo, M.G., Ricciardi, A., Zotta, T., Parente, E., and Salzano, G. (2008). Molecular and technological characterization of lactic acid bacteria from traditional fermented sausages of Basilicata region (Southern Italy). Meat Science, 80:1,238-1,248.

Chukeatirote, E. (2003). Potential use of probiotics. Songklanakarin Journal of Science and Technology, 25:275-282.

Cole, J.R., Chai, B., Farris, R.J., Wang, Q., Kulam, S.A., McGarrell, D.M., Garrity, G.M., and Tiedje, J.M. (2005). The Ribosomal Data base Project (RDP-II): sequences and tools for high-throughput rRNA analysis. Nucleic Acids Research, 33:D294-D296.

de Vuyst, L., Falony, G., and Leroy, F. (2008). Probiotics in fermented sausages. Meat Science, 80:75-78.

DSMZ. Available at http://www.dsmz.de/micro organisms/bacteria_catalogue.php (Last accessed 31 October 2008).

Ehrmann, M.A. and Vogel, R.F. (2005). Molecular taxonomy and genetics of sourdough lactic acid bacteria. Trends in Food Science and Technology, 16:31–42.

Fontana, C., Cocconcelli, P.S. and Vignolo, G. (2005). Monitoring the bacterial population dynamics during fermentation of artisanal Argentinean sausages. International Journal of Food Microbiology, 103:131-142.

Gerez, C.L., Torino, M.I., Rollan, G., and de Valdez, G.F. (2009). Prevention of bread mould spoilage by using lactic acid bacteria with antifungal properties. Food Control, 20:144-148.

Hannon, J.A., Kilcawley, K.N., Wilkinson, M.G., Delahunty, C.M., and Beresford, T.P. (2007). Flavour precursor development in Cheddar cheese due to lactococcal starters and the presence and lysis of *Lactobacillus helveticus*. International Dairy Journal,17:316-327.

Head, I.M., Saunders, J.R., and Pickup, R.W. (1998). Microbial evolution, diversity, and ecology: a decade of ribosomal RNA analysis of uncultivated microorganism. Microbial Ecology, 35:1-21.

Heiman, M. (1997). Webcutter 2.0. Available at http://rna.lundberg.gu.se/cutter2 (Last accessed 31 October 2008).

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., and Higgins, D.G. (2007). ClustalW and ClustalX version 2. Bioinformatics, 23:2,947-2,948.

Lee, C.H. (1997). Lactic acid fermented foods and their benefits in Asia. Food Control, 8:259-269.

Ljungh, A. and Wadstrom, T. (2006). Lactic acid bacteria as probiotics. Current Issues in Intestinal Microbiology, 7:73-89.

Marchesi, J.R., Sato, T., Weightman, A.J., Martin, T.A., Fry, J.C., Hiom, S.J., and Wade, W.G. (1998). Design and evaluation of useful bacterium-specific PCR primers that amplify genes coding for bacterial 16S rRNA. Applied and Environmental Micro biology, 64:795-799.

Rantsiou, K., Drosinos, E.H., Gialitaki, M., Metaxopoulos, I., Comi, G., and Cocolin, L. (2006). Use of molecular tools to characterize *Lactobacillus* spp. isolated from Greek traditional fermented sausages. International Journal of Food Microbiology, 112:215-222.

Saito, T. (2004). Selection of useful probiotic lactic acid bacteria from the *Lactobacillus acidophilus* group and their applications to functional foods. Animal Science Journal, 75:1-13.

Sakai, K., Taniguchi, M., Miura, S., Ohara, H., Matsumoto, T. and Shirai, Y. (2004). Making Plastics from Garbage. Journal of Industrial Ecology, 7:63-74.

Stiles, M.E. and Holzapfel, W.H. (1997). Lactic acid bacteria of foods and their current taxonomy. International Journal of Food Microbiology, 36:1-29.

Temmerman, R., Huys, G., and Swings, J. (2004). Identification of lactic acid bacteria: culture-dependent and culture-indepen dent methods. Trends in Food Science and Technology, 15:348-359.