MOLECULAR IDENTIFICATION AND TYPING OF *LACTOBACILLUS DELBRUECKII* SUBSPECIES *BULGARICUS* AND *STREPTOCOCCUS THERMOPHILUS*

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

AYSUN CEBECİ AYDIN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
BIOCHEMISTRY

FEBRUARY 2008

Approval of the thesis:

MOLECULAR IDENTIFICATION AND TYPING OF *LACTOBACILLUS DELBRUECKII* SUBSPECIES *BULGARICUS* AND *STREPTOCOCCUS THERMOPHILUS*

submitted by **AYSUN CEBECİ AYDIN** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Biochemistry Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen
Dean, Graduate School of **Natural and Applied Sciences** _____

Assoc. Prof. Dr. Nursen Çoruh
Head of Department, **Biochemistry** _____

Assoc. Prof. Dr. G. Candan Gürakan
Supervisor, **Food Engineering Dept., METU** _____

Examining Committee Members:

Prof. Dr. Mesude İşcan
Biology Dept., METU _____

Assoc. Prof. Dr. G. Candan Gürakan
Food Engineering Dept., METU _____

Prof. Dr. Zümrüt B. Ögel
Food Engineering Dept., METU _____

Prof. Dr. Mahinur Akkaya
Chemistry Dept., METU _____

Prof. Dr. Barbaros Özer
Food Engineering Dept., Harran University _____

Date: 08.02.2008

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : Aysun, Cebeci Aydın

Signature :

ABSTRACT

MOLECULAR IDENTIFICATION AND TYPING OF *LACTOBACILLUS DELBRUECKII* SUBSPECIES *BULGARICUS* AND *STREPTOCOCCUS THERMOPHILUS*

Cebeci Aydın, Aysun

Ph.D., Department of Biochemistry

Supervisor: Assoc. Prof. Dr. G. Candan Gürakan

February 2008, 143 pages

Lactic acid bacteria are associated with preservation of foods, including milk, meat and vegetables. Yoghurt is produced by the cooperative action of two starter bacteria; *S. thermophilus* and *L. delbrueckii* subsp. *bulgaricus.* In this study, identification and typing of yoghurt starter bacteria were aimed. Traditional home made yoghurts were collected from different areas of Turkey, identification of those isolates at species and subspecies level and typing at strain level were achieved using PCR based methods.

In our study, identification of yogurt starter bacteria was studied using species specific primers and ARDRA. These methods were inefficient in identification of yoghurt starter bacteria, at species and subspecies level. Consequently, a reliable and quick method for accurate identification of yoghurt starter bacteria was developed. The new method focuses on amplification of methionine biosynthesis genes, for selective identification of yoghurt starter bacteria together with some cheese starters. Further discrimination by ARDRA enabled differentiation of

yoghurt starter bacteria from cheese starters. Confirmation of the proposed method has been accomplished by partial sequencing of the 16S rRNA gene.

After correct identification of starter bacteria had been achieved, the strains were typed at strain level using RAPD-PCR and MLST. RAPD-PCR with primer 1254 resulted better fingerprints, compared to primer M13 at strain level. Comparisons of the two typing methods showed that RAPD-PCR revealed strain diversity better than MLST, however MLST was a more robust and reliable method and resulted in clustering of the strains depending on the isolation source.

Keywords: *L. delbrueckii* subsp. *bulgaricus*, *S. thermophilus,* RAPD-PCR, ARDRA, MLST

ÖZ

LACTOBACILLUS DELBRUECKII ALTTÜR *BULGARICUS* VE *STREPTOCOCCUS THERMOPHILUS* MOLEKÜLER TANILARI VE TİPLENDİRİLMELERİ

Cebeci Aydın, Aysun

Doktora, Biyokimya Bölümü

Tez Yöneticisi: Doç. Dr. G. Candan Gürakan

Şubat 2008, 143 sayfa

Süt, et ve sebze gibi gıda ürünlerinin muhafaza edilmesinde laktik asit bakterilerinden faydalanılmaktadır. Yoghurt; *S. thermophilus* ve *L. delbrueckii* alttür *bulgaricus* starter kültürlerinin birikte çalışması sonucunda üretilmektedir. Bu çalışmada, yoğurt starter bakterilerinin tanımlandırılması ve tiplendirilmesi hedeflenmiştir. PZR bazlı metodlar kullanılarak Türkiye'nin çeşitli bölgelerinden toplanan geleneksel ev yapımı yoğurt örneklerinden izole edilen bakteriler tür ve alttür seviyesinde tanımlanmış ve suş seviyesinde tiplendirilmiştir.

Bu çalışmada yoğurt starter bakterilerinin tanımlanması amacıyla literatürden elde edilen türe özel primerler ve ARDRA kullanılmıştır. Ancak yoğurt starter bakterilerinin tanımlanmasında bu metodların tür ve alttür seviyesinde yetersiz oldukları saptanmıştır. Sonuç olarak yoğurt starter bakterilerinin doğru tanımlanması için güvenilir ve hızlı bir metod geliştirilmiştir. Yeni metod metiyonin biyosentezi geninin amplifikasyonunu hedefleyerek yoğurt starter ve bazı peynir starter bakterilerinin seçici tanısını sağlamaktadır. ARDRA kullanılarak yoğurt ve peynir starter bakterileri arasında daha ileri düzeyde ayrım

yapılabilmiştir. Önerilen metodun doğruluğu 16S rRNA geninin kısmi dizileme sonuçlarıyla kanıtlanmıştır.

Starter bakterilerinin doğru tanımlanması sağlandıktan sonra RAPD-PZR ve MLST kullanılarak suş düzeyinde tiplendirme yapılmıştır. 1254 primeriyle yapılan RAPD-PZR M13 primeriyle yapılan çalışmalara göre suş seviyesinde daha iyi parmakizi sonuçları vermiştir. İki tiplendirme metodunun kıyaslanması sonucunda RAPD-PZR'nin suş çeşitliliğini ortaya çıkarmakta MLST'den daha iyi sonuç verdiği görülmüştür. Ancak, MLST'nin daha güçlü ve güvenilir bir metod olduğu ve suşların gruplandırılmasını izolasyon kaynağına göre gerçekleştirdiği saptanmıştır.

Anahtar Sözcükler: *L. delbrueckii* alttür *bulgaricus*, *S. thermophilus*, RAPD-PZR, ARDRA, MLST

TO MY FAMILY

ACKNOWLEDGEMENTS

I would like to express my gratitude to people who were with me during this work.

I would like to express my sincere gratitude to my supervisor, Assoc. Prof. Dr. G. Candan Gürakan for her understanding, support, and advises which drove my motivation not only through this thesis but in all respects of my life.

I would like to thank the members of my jury, Prof. Dr. Mesude İşcan, Prof. Dr. Zümrüt B. Ögel, Prof. Dr. Mahinur Akkaya, and Prof. Dr. Barbaros Özer who kindly agreed to participate in my jury and shared their valuable contributions, suggestions and comments on my thesis.

I deeply appreciate my beloved friends Suzan Tireki, Beray Gençsoy, Dr. Birsen Can Demirdöğen and Selda Türkoğlu for their intelligence, support, and sincere friendships, Dr. Gökhan Duruksu for his kindness and advices and Dr. Ayşegül Ersayın for her joy. My previous lab mates Firdevs Yalçın, Gamze Aydın, Zeynep Sönmezalp, Hümeyra Bahçeci will remain my friends forever. Finally, I thank to my lab mates Neslihan Altay, Esra Uçkun and Houman Jabbari for their friendship and collaboration through our studies.

My beloved parents and brother were always with me through my thesis, without their care, support, understanding and patience I could not complete this thesis. My dear husband Erkin Aydın always supported me with love, and his presence is very valuable to me. My deepest admiration and love belongs to Yağız Eray, he always experienced a busy and tired mother, and showed a great patience at long days, weekends and holidays I could not be with him. He is my source of happiness and joy.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVATIONS

LAB: Lactic Acid Bacteria

PCR: Polymerase Chain Reaction

ssu: small subunit

rRNA: ribosomal RNA

ITS: Intergenic Transcribed Spacer

ARDRA: Amplified Ribosomal DNA Restriction Analysis

RFLP: Restriction Fragment Length Polymorphism

RAPD: Random Applied Polymorphic DNA

MLST: Multilocus Sequence Typing

HGT: Horizontal Gene Transfer

β-*gal*: beta-galactosidase gene

*phe*S: phenylalanyl-tRNA synthase

*rpo*A: RNA polymerase alpha subunit

UPGMA: Unweighted Pair Group Method of Arithmetic mean

EDTA: Ethlenediamineetetraacetic acid

EtBr: Ethidium Bromide

bp: basepair

rpm: Revolution per minute

UV: Ultraviolet

NA: not applied

CHAPTER 1

INTRODUCTION

1.1. Introduction to lactic acid bacteria

Lactic acid bacteria (LAB) are generally chained cocci or rod shaped gram (+), nonmotile, nonsporulating bacteria that produce lactic acid as a major or sole product of fermentative metabolism. LAB ferment lactose to lactic acid and this process contributes to flavour development and has a major role in preventing the growth of spoilage and pathogenic microorganisms.

LAB live in habitats where pH of the medium is low. There are eight genera that produce lactic acid: *Lactobacillus, Streptococcus, Lactococcus, Leuconostoc, Pediococcus, Carnobacterium, Enterococcus and Sporolactobacillus*. All LAB are anaerobes, however they are facultative anaerobes, and they can grow in the presence of oxygen. Some strains produce H_2O_2 through flavoprotein oxidase systems, and eliminate H_2O_2 by either catalase or peroxidase. LAB use lactose as their main source of carbon to produce energy. Lactose (4-O-β-D-galactopyranosyl-D-glucose) is a disaccharide sugar composed of glucose and galactose. It is unique to mammalian milks, lowest in marine mammals and highest in humans. Lactose itself is a fermentable substrate, first being hydrolyzed by facultative or anaerobic microorganisms in the small intestine, allowing for anaerobic metabolism of the resultant simple sugars. If dietary lactose can not be hydrolyzed into its component simple sugars in the small intestine, it is absorbed across the intestine and fermented by the colonic microflora. This condition leads to lactose intolerance.

Members of the LAB lack porphyrins and cytochromes, do not carry out oxidative phosphorylation, and hence obtain energy only by substrate level phosphorylation. LAB that only produce lactic acid as an end product are called homofermentative; those that also produce acetic acid, ethanol and carbon dioxide are termed heterofermentive (Table 1.1.). The differences observed in the fermentation products are determined by the presence or absence of the enzyme aldolase, one of the key enzymes in glycolysis. Heterofermentors lack aldolase, and cannot break down fructose bisphosphate to triose phosphate, however they have transketolase to produce lactic acid, acetaldehyde, ethanol, and CO_2 (Madigan et al., 1997).

Table 1.1. Differentiation of principal genera of LAB (Madigan et al., 1997)

Genus	**Cell form and arrangement**	**Fermentation**
Streptococcus	Cocci in chains	Homofermentative
Leuconostoc	Cocci in chains	Heterofermentative
Pediococcus	Cocci in tetrads	Homofermentative
Lactobacillus	(1) Rods, usually in chains	Homofermentative
	(2) Rods, usually in chains	Heterofermentative
Enterococcus	Cocci in chains	Homofermentative
Lactococcus	Cocci in chains	Homofermentative

LAB can grow in most of our common food raw materials. They constitute part of their natural microflora, can be used in spontaneous fermentation, and can also be added as starters (Table 1.2.). LAB need some sugar for fermentation, either naturally present or added. Except for some fruits, the pH of raw materials is seldomly low enough to inhibit the growth of these bacteria. The members of the genus *Lactobacillus* are also found among the gastrointestinal microflora.

From a taxonomic perspective, the streptococci, the lactobacilli and the leuconostocs are grouped with the other LAB and *Bacillus* species within a supercluster of the clostridial sub-branch of the gram-positive eubacteria, characterized by a low (<55mol%) genomic GC content.

Bifidobacteria are not true LAB in the sense of a *Lactococcus* or *Pediococcus,* it has a high GC content, however, it is long associated with the LAB group. Bifidobacteria produce both acetic and lactic acids. In addition, small quantities of formic acid and ethanol are often produced (Hughes and Hoover, 1991). Bifidobacteria are over shaded by the members of Lactobacilli. The reason for this is because Bifidobacteria are difficult organisms for research, for growth they need anaerobic conditions and a growth factor present in human milk is required.

There are numerous application areas for use of LAB in preservation of foods and use as probiotics.

1.2. Yoghurt

Yoghurt is a popular fermented milk product in Turkey. It is produced by the cooperative action of yoghurt starter bacteria; *S. thermophilus* and *L. delbrueckii* subsp*. bulgaricus*. Their major contribution in yoghurt production is fermentation of lactose to lactic acid, and result in coagulation (De Brabandere and De Baerdemaeker, 1999). However, in addition to lactic acid, free fatty acids, acetaldehyde, diacetyl, and acetoin are also produced, and these compounds give the distinct flavor to the yoghurt (Bonczar et al., 2002). The acidification and aroma forming abilities of starter cultures are strain dependent (Beshkova et al., 1998, Marilley and Casey, 2004). Production of acetaldehyde, diacetyl and acetoin is supplied probably by *L. delbrueckii* subsp. *bulgaricus* strains, while acid production is faster in *S. thermophilus* and production of free fatty acids are obtained by the synergistic actions of the two bacteria (Beshkova et al., 1998).

Table 1.2: Foods and their associated LAB (Stiles 1996)

Food types	**LAB**
Milk and dairy foods	
Hard cheeses without eye formation	*Lc. lactis* subsp. *cremoris* and subsp. *lactis*
Cottage cheeses and cheeses with a few or small eyes (Edam)	*Lc. lactis* subsp. *cremoris* and subsp. *lactis* and *Le. mesenteroides* subsp. *cremoris*
Cultured butter, buttermilk cheeses with round eyes (Gouda)	*Lc. lactis* subsp. *cremoris* and subsp. *lactis* and subsp. *lactis* biovar *diacetylactis*, and *Le. mesenteroides* subsp. *cremoris*
Swiss type cheeses	*L. delbrueckii* subsp. *bulgaricus*, *L. helveticus.*
Dairy foods in general	*L. brevis, L. buchneri, L. casei, L. paracasei, L. fermentum, L. plantarum, Le. mesenteroides* subsp. *cremoris* and *Le. lactis*
Fermented milks	
-yogurt	*Streptococcus thermophilus* and *L. delbrueckii* subsp. *bulgaricus, Lc. lactis* subsp. *lactis* biovar *diacetylactis*
-acidophilus milk	*L. acidophilus*
-kefir	*L. kefir* and *L. kefiranofaciens*
Sourdough bread	*L. sanfransisco, L. farciminis, L. fermentum, L. brevis, L. plantarum, L. amylovorus, L. reuteri*
Soy sauce	*Tetragenococcus (Pediococcus) halophilus*
Wine (malo-lactic fermented)	*Le. oenos*

Table 1.2: Foods and their associated LAB (Stiles 1996) (cont'd)

Fermented vegetables	*P. acidilactici, P. pentosaceus, L. fermentum, L.plantarum, L. sake, L. buchnerii*
Cucumbers, sauerkraut	*Le. mesenteroides* subsp. *mesenteroides, L. bavaricus, L. brevis, L. sake, L. plantarum*
Olives	*Le. mesenteroides, L. pentosus*

C: Carnobacterium, L: Lactobacillus, Lc: Lactococcus, Le: Leuconostoc, P: Pediococcus

1.2.1. *L. delbrueckii* subsp. *bulgaricus*

Lactobacillus delbrueckii subsp. *bulgaricus* is gram-positive, facultatively anaerobic, non-motile and non-spore-forming, rod-shaped bacteria (Figure 1.1.). Like *S. thermophilus*, *L. delbrueckii* subsp. *bulgaricus* is a thermophilic starter culture with optimum growth at 42 °C. Together with *Lactobacillus delbrueckii* subsp. *bulgaricus, L. delbrueckii* species contains two other subspecies, *L. delbrueckii* subsp. *delbrueckii*, and *L. delbrueckii* subsp. *lactis*. Phylogenetically, *L. delbrueckii* subsp. *bulgaricus* is closely related to *L. acidophilus* group; *L. amylovorus, L. acidophilus, L. helveticus, L. rhamnosus, L. gasseri,* and *L. amylophilus*.

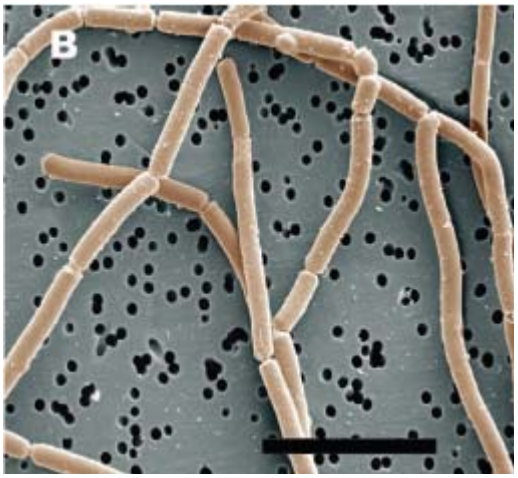


Figure 1.1. Scanning electron micrograph of *L. delbrueckii* subsp. *bulgaricus* (Image by Jeff Broadbent)

1.2.2. *S. thermophilus*

Streptococcus thermophilus is a dairy starter bacterium in cheese and yogurt production. It is regarded as the second most important industrial dairy starter after *Lactococcus lactis*. *S. thermophilus* is the only food species among commensal and pathogenic *Streptococcus* genus. Isolates identified as anaerobic, aerotolerant, cocci shaped, catalase-negative and gram-positive, and able to grow at 45 °C, hence termed thermotolerant. *S. thermophilus* is used alone or in combination with several lactobacilli and lactococci, but for yogurt it is always used with *Lactobacillus delbrueckii* subsp. *bulgaricus*. The role of *S. thermophilus* in milk fermentation is due to its rapid acidification ability and the production of metabolites important for technological properties (Delorme, 2007).

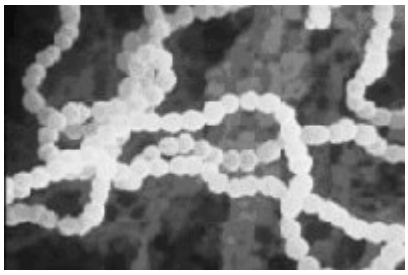


Figure 1.2. Electron micrograph of *S. thermophilus* (Image by Micheline Rousseau ©INRA)

1.3. Identification of bacteria using phenotypic and biochemical characteristics

There are basically two methods of identifying bacteria; one is based on phenotypic methods and the other is based on differences on the genetic material of bacteria.

Phenotype is the observable expression of genotype and conventional methods of bacterial identification rely on morphological, physiological and biochemical properties of the organism. Before the advance in molecular techniques, identification of bacteria was based only on phenotypic features, which require use of pure laboratory cultures.

Traditional phenotypic tests used for LAB in classical microbiological laboratories include growth on specific media, gram-staining, catalase production, investigation of cell morphology under microscope, investigation of colony morphology on agar, growth at different temperatures, pH and salt, and fermentation of carbohydrates and/or API kits. The main advantage of using these methods is that they usually do not require expensive laboratory equipment. API kits are ready-to-inoculate identification kits (such as API 50 CH, LRA Zym and API Zym enzymatic tests) and can be used for the rapid and theoretically reproducible

phenotypic identification of pure cultures. The main disadvantage of phenotypical identification methods is that, they are time consuming and labor intensive. Moreover, the reliability of these tests has been questioned. An example is presented by Andrighetto et al., (1998). They used API 50 CH to analyze 25 strains of thermophilic lactobacilli (*L. delbrueckii* subsp. *lactis* and subsp. *bulgaricus*, *L. helveticus* and *L. acidophilus*), however clear assignment to a particular species or subspecies was not possible for most of the strains. It was stated by the authors that this was resulted from difficulties in interpretation of sugar fermentation profiles. In addition, the manufacturer's database was reported as not up-to-date and some *Lactobacillus* species were missing. Another major disadvantage with phenotypic methods is that since gene expression is related to growth conditions, and same organism might show different phenotypic characters in different environmental conditions. To overcome this difference, phenotypic data must be compared with similar set of data from type strain of closely related organisms. Reproducibility of results between different laboratories is another problem; therefore, standardized procedures should be used during experiments (Prakash et al., 2007).

1.4. Identification based on protein fingerprinting

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) of whole-cell proteins is a fingerprinting method used for identification of lactobacilli at species level. If a bacterial strain can be grown under standardized conditions, it always produces the same set of proteins and fingerprints. SDS-PAGE separates proteins exclusively according to molecular weight, while non-denaturing (native) PAGE can be used as a complementary technique, separating cell proteins according to their charge and size, providing high resolution and good band definition. Under standardized conditions, reproducible patterns can be obtained that are suitable for digital analysis. Protein profiles can be stored in database format and may be routinely used to confirm the identity of *Lactobacillus* strains, to differentiate between unknown isolates and to evaluate classification schemes, at species level. De Angelis et al., (2001) isolated nonstarter LAB from 12 Italian

ewe's milk cheeses. Most of the species studied gave specific protein profiles, except *L. plantarum* and *L. pentosus*, which were grouped in the same cluster. Gancheva et al., (1999), used SDS-PAGE to analyze the cellular proteins of a set of 98 strains belonging to nine species of the *Lactobacillus* acidophilus rRNA group (*L. acidophilus*, *L. amylolyticus*, *L. crispatus*, *L. johnsonii*, *L. gasseri*, *L. gallinarum*, *L. helveticus*, *L. iners* and *L. amylovorus*). Most of these species can be differentiated by SDS-PAGE, but poor discrimination was obtained between *L. johnsonii* and *L. gasseri* strains, and between some strains of *L. amylovorus* and *L. gallinarum*.

1.5. Identification of bacteria using genotypic methods and importance 16S rRNA gene in bacterial identification

The ribosomal RNAs stand at the center of protein synthesis. They carry out key reactions in translation. A major function is to ensure the correct structure of DNA, thereby allowing its tight packing around the ribosome's active core. All organisms possess ribosomal RNA whose numbers vary from 20,000 to 50,000 depending on the intensity of protein synthesis. In prokaryotes, ribosomal RNA is composed of 30S (small subunit) and 50S (large subunit), and in eukaryotes it is composed of 40S and 60S. Small subunit RNA contain one RNA (18S in eukaryotes, 16S in prokaryotes), while large subunit contains 2 RNA (5S and 23S) in prokaryotes, and 3 RNA in eukaryotes (5.8S, 5S and 25/28S) (Figure 1.3.).

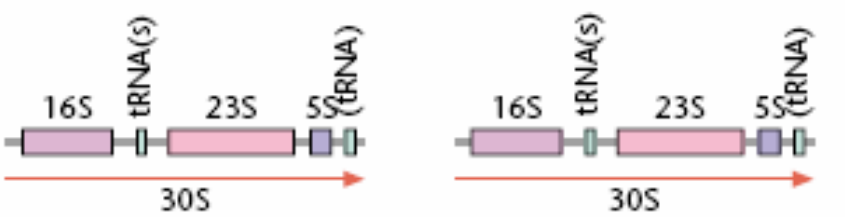


Figure 1.3. rRNA gene organization in the example of *E. coli* (Lafontaine and Tollervey 2001).

In bacteria and archea, the organization of rRNA includes 16S, 5S and 23S rRNA with a tRNA gene located between 16S and 23S, within the internal or 3'-terminal of the spacer, and one or more tRNA gene may be present to the 3' of 5S. Ribosomal RNA is generated from pre-rRNA by posttranscriptional processing. Transcription of a ribosomal operon produces an RNA molecule about 5,000 nucleotides long. Such transcript is processed to become fully functional. Before or during the processing, certain residues in the sequences of spacers, rRNA, and tRNA are specifically modified. Endonucleolytic cleavage of the primary transcript releases mature rRNA fragments, which then undergo additional modification, fold into a defined three-dimensional structure, and bind the set of ribosomal proteins to assemble into mature ribosomal subunits (Tourova, 2003). Most of the organisms containing multiple ribosomal operons carry three to eight copies. Comparison of closely related organisms reveals that they usually have similar numbers of ribosomal genes (Fogel et al., 1999), making it possible to extrapolate the data from one species to a related species for which the number of copies of the rRNA genes is not known. Such an extrapolation is often helpful, for example, when estimating the number of organisms in environmental samples. The functional role of the multiple copies of ribosomal operons is explained by two mechanisms. In the first mechanism, it is suggested that the existence of multiple copies or ribosomal genes in prokaryotes is needed to maintain the high rate of rRNA synthesis and thus the large number of ribosomes required for fast cell growth and division. In the second mechanism, it was suggested as an adaptive mechanism. Support for the latter idea comes from the observation that, the genomes of symbiotic or parasitic species present in very stable environments contain only a small (one or two) number of ribosomal operons (Tourova, 2003)

Inspection of the structure of rRNAs reveals that despite substantial differences in primary sequence, both the small subunit rRNA (SSU-rRNA) and large subunit rRNA (LSU-rRNA) display remarkable conservation of their secondary, and probably tertiary, structures. An example of 16S rRNA gene secondary structure is presented in Figure 1.4.

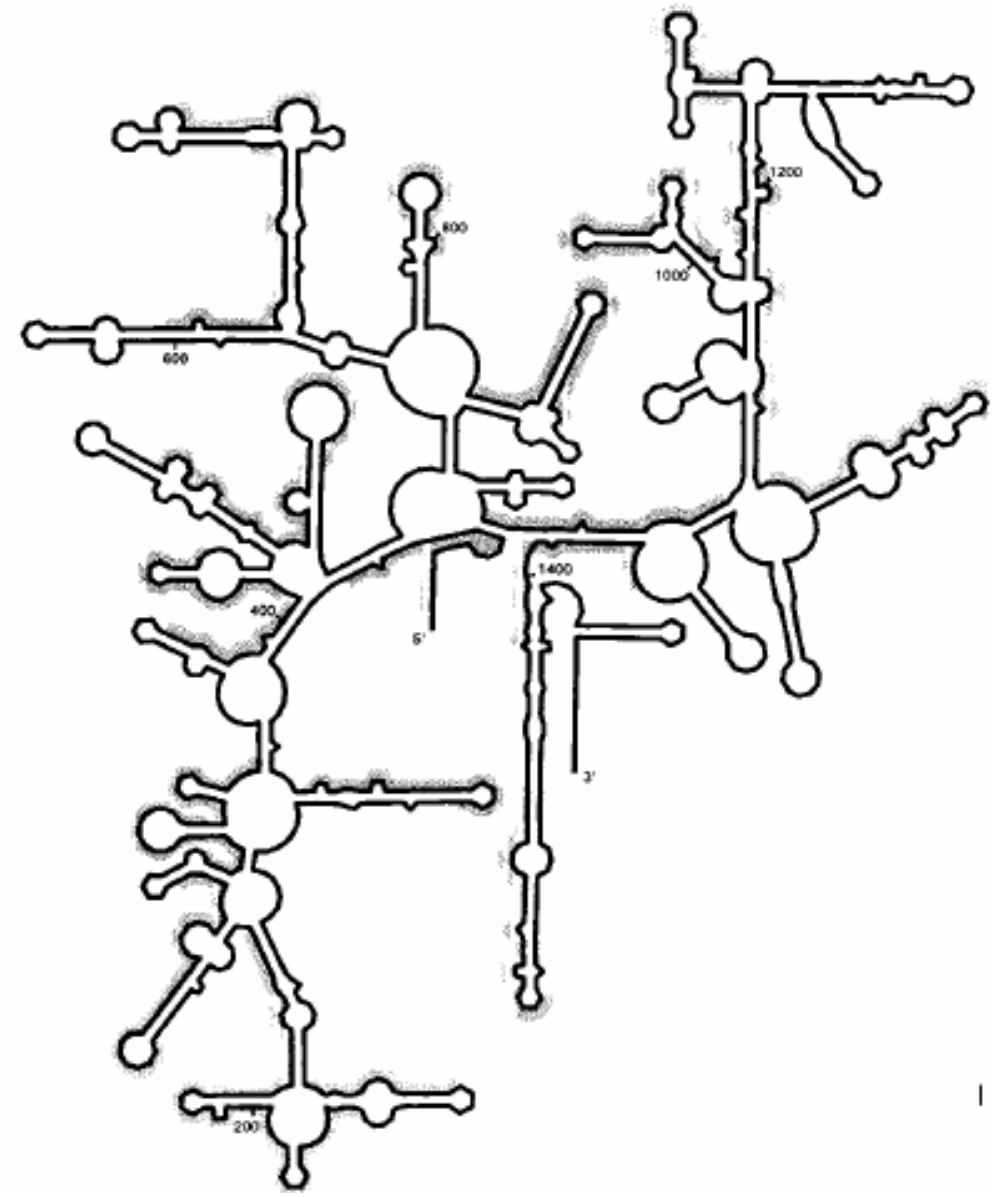


Figure 1.4. Secondary structure of the 16S rRNA. (Amann et al, 1995)

Table 1.3. Functional domains within the *E. coli* rRNAs (Lafontaine and Tollervey 2001)

Functional domain	Region of rRNA	Major functions	Functionally related antibiotics
C1400 region	16S rRNA (1400–1500)	Decoding Translocation	Paromomycin
530 loop	16S rRNA (500–545)	Decoding EF-Tu binding	Streptomycin
Helix 34	16S rRNA	Decoding EF-G function and translocation	Spectinomycin
912 region	16S rRNA (885–912)	Translational accuracy	Streptomycin
Helix 45 (colicin fragment)	16S rRNA (1494–3′end)	Shine–Dalgarno interaction Initiation factor binding Subunit interface	Kasugamycin
Domain II (GTPase region)	23S rRNA	EF-G-dependent GTP hydrolysis ppGpp synthesis in stringent response	Thiostrepton Micrococcin
Domain V (PT centre)	23S rRNA	Peptide bond formation Interaction with tRNA 3′ ends Translational accuracy	Chloramphenicol Erythromycin
Domain IV (1916 loop)	23S rRNA	Interaction with anticodons and tRNA 3′ ends Translational accuracy Subunit interface	
α sarcin loop	23S rRNA (2653–2667)	EF-Tu and EF-G binding site	Sarcin, ricin

From Table 1.3, it is clear that the specific recognition of the codon by the tRNA of the ribosome lies within its small subunit and the peptidyltransferase activity is carried out by the large subunit. The accuracy of translation is determined by components of both subunits, probably reflecting interactions of the tRNAs with both ribosomal subunits. In bacteria, the 3' end of the 16S rRNA base pairs with the mRNA which is called as Shine–Dalgarno interaction and is crucial for translation initiation. In all organisms, the mRNA and tRNA interact in the codon–anticodon recognition, which is repeated throughout the whole translational process. This interaction, which involves only three base-paired nucleotides, is stabilized by a number of interactions between the tRNA and the rRNAs. Interactions important for the catalysis of peptide bond formation are provided by the recognition of the universally conserved 3' CCA end of aminoacyl-tRNA substrates by 23S rRNA. In *E. coli*, an rRNA–mRNA interaction is proposed to play a role in the recognition of the termination triplet.

5S, 16S and 23S rRNA molecules and spacers between these can be used for phylogenetic analyses but the small and large sizes of 5S rRNA gene (120 bp) and 23S rRNA gene (3000 bp) have restricted their use. The small subunit rRNA gene is highly conserved in structure and presumed function across all of evolution, and has become the most commonly used marker for establishing phylogenetic relationships between organisms mainly because of the suitable size of about 1600bp. In the genome, rRNA genes exist in high copy number, thus mutations present in the rRNA cannot be observed easily. Also, mutations in the conserved core regions of the rRNA are heavily biased toward nucleotide substitution, rather than deletion and insertion. Together with the highly conserved secondary structure, this property eases alignment procedures.

Comparative analysis of ribosomal genes, or ribosomal phylogenetics, is currently widely used to infer evolutionary relationships between species, especially prokaryotic species. Woese and Fox (1977) studied comparative analysis of the 16S rRNA genes, and their study resulted in three primary lineages, eukarya, bacteria, and archaea (Figure 1.5). The underlying principle of ribosomal phylogenetics assumes that ribosomal genes can serve as molecular clocks. The choice of ribosomal genes as a molecular clock was strongly influenced by the fact that, selection pressures work to stabilize the genes. Such genes inferring molecular clocks can be used for reconstructions of evolutionary history and for building a universal tree of living organisms. There exists additional advantages of using ribosomal genes as phylogenetic markers; these are a) their universal distribution, b) high sequence conservation, and c) functional stability.

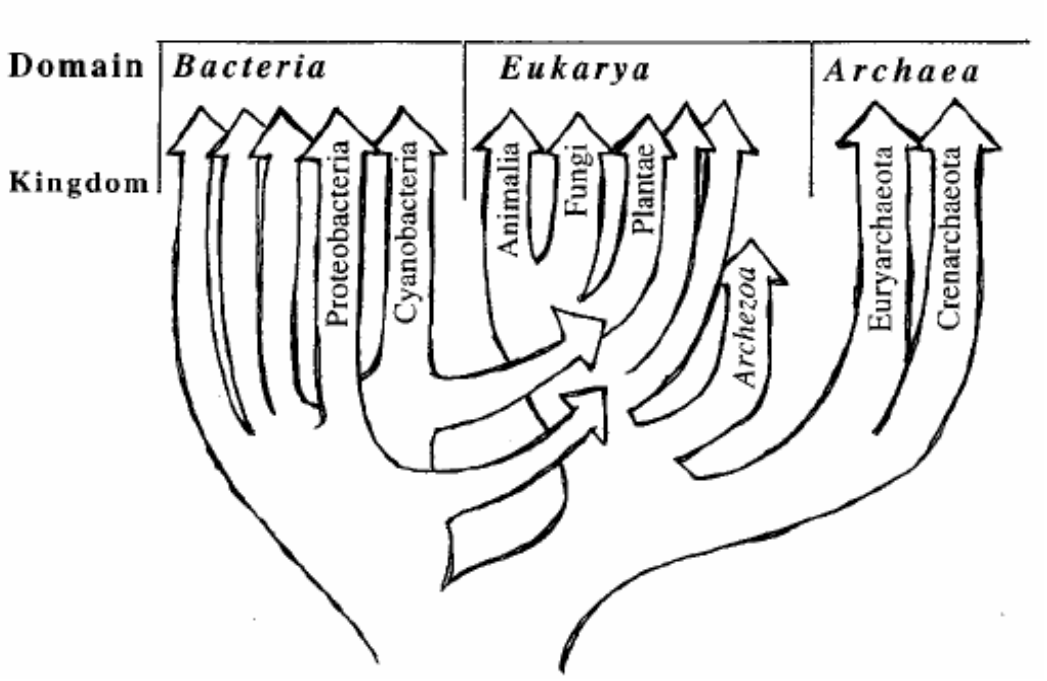


Figure 1.5. A simplified sketch of the universal 16S rRNA gene tree. The tree is commonly taken as a representation of organism phylogeny (Doolittle 1999).

Following work of Woese, *Bergey's Manual of Systematic Bacteriology* used 16S rRNA gene sequences to classify prokaryotes (Boone et al., 2001). Figure 1.6 is a dendrogram generated by almost complete (1400bp) 16S rRNA gene sequences of lactobacilli and related genera.

Figure 1.6. Phylogenetic tree based on 16S rRNA gene sequences of lactobacilli and related genera. Multifurcated branches indicate a topology that could not be unambiguously resolved. *One of the major representatives of the group (Satokari et al., 2003).

In general, a bacterial species is expected to show 70% of DNA–DNA binding with less than 5% difference in their melting temperature (ΔTm) and over 97% of 16S ribosomal RNA (rRNA) gene-sequence identity among its strains (Stackebrandt and Goebel, 1994). DNA-DNA homology criterion is not

determined theoretically, however, it was adjusted empirically to match phenotype-based species (Cohan, 2002). While having less than 97% 16S rRNA similarity has less than 70% DNA-DNA homology, samples having more than 97% homology may not have 70% DNA-DNA homology treshold value for species assignment (Fox et al., 1992, Stackebrandt and Goebel, 1994). To avoid misidentification, some researchers prefer to use more stringent identification limits, for example, Bosshard et. al., (2003) used 99% similarity to assign a species in a clinical laboratory while using 16S ribosomal DNA.

The 16S rRNA gene sequence is about 1,550 bp long and is composed of both variable and conserved regions. Universal primers are usually chosen as complementary to the conserved regions at the beginning of the gene and at either the 540-bp region or at the end of the whole sequence (about the 1,550-bp region), and the sequence of the variable region in between is used for the comparative taxonomy. Although 500bp and 1,500bp are common lengths to sequence and compare, sequences in databases can be of various lengths. In a study, Clarridge (2004) compared dendrograms produced by 500bp of 16S rRNA gene with the 1500bp of 16S rRNA gene (Figure 1.7). The resulting dendrograms were similar but not identical. Using 500bp of 16S rRNA gene sequences is favored by practical and economical reasons.

The intergenic transcribed spacer (ITS) of 16S-23S rRNA can be used for characterizing bacteria below species level, since this region has high polymorphism and can serve as an important tool for phylogenetic analysis.

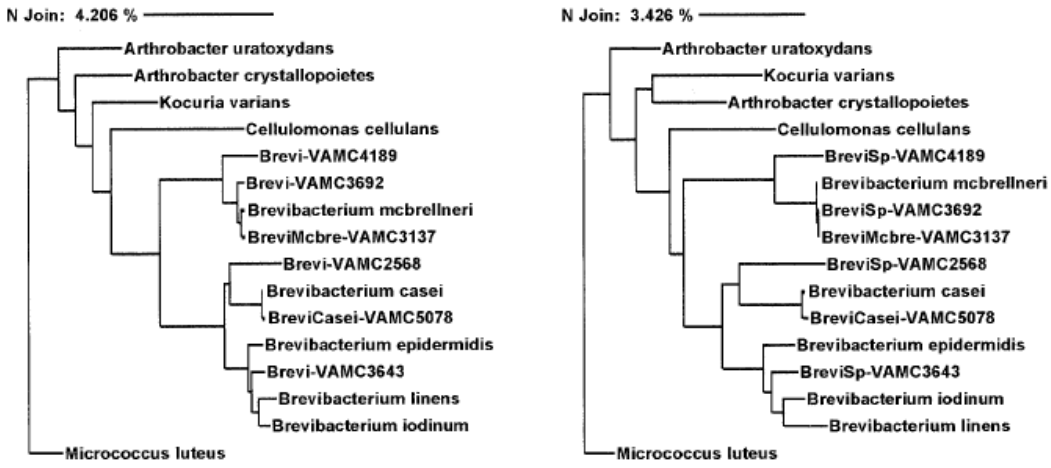


Figure 1.7. A comparison of dendrograms generated using either the 1,500-bp 16S rRNA gene sequence (left) or the 500-bp 16S rRNA gene sequence (right) of a group of clinical and type strains of *Brevibacterium* (Clarridge, 2004).

Although accepted by many scientists as a primary method in identification of species, in some species its use is limited in resolutionary power and reliability. Doolittle (1999), states that although ribosomal RNA is an important tool, it can mislead to false evolutionary trees because of horizontal gene transfer (HGT). It is argued that HGT is very common among bacteria and archea, with the examples of HGTs of ribosomal protein genes, and frequent between-domain exchanges of genes encoding aminoacyl tRNA synthetases. There are several examples supporting this view, a) horizontal gene transfer of partial or entire rRNA genes between different organisms (Ueno et al., 2007, Yap et al., 1999). Moreno et al., (2002) studied analysis of the variations in 16S rRNA genes of the genus *Vibrio*. The presence of short segments with a high number of base variations could be explained by recombination of short segments with laterally transferred rRNA genes (Wang & Zhang, 2000, Moreno et al., 2002), and b) divergent rRNA operons can coexist in the same genome (Reischl, et al., 1998).

Though disadvantages of 16S rRNA gene sequencing are presented above, they are still indispensable tools to study bacteria, mainly because there is no other

current method to replace analysis of 16S rRNA gene. In addition, 16S rRNA gene sequence has been determined for a large number of strains. GenBank, the largest databank of nucleotide sequences, has over 61 million deposited sequences to-date (January 2008), of which about 760,000 are of 16S rRNA gene. However, in specialized issues like naming of a new isolate, or in very closely related bacteria where 16S rRNA gene is insufficient to identify or differentiate, a more detailed analysis of the sample is necessary. The preferred route is to use a polyphasic approach, explained as the use of several methods together to obtain a reliable result. The typing of bacteria at strain level also requires such approach. The methods may include phenotypic or genotypic approaches, or both, including sequencing of SDS-PAGE of total cell proteins, or proteins of cell wall, Pulsed Field Gel Electrophoresis (PFGE), Multi Locus Sequence Typing (MLST), or Random Amplified Polymorphic DNA (RAPD) (Sanchez et al., 2004).

1.6. Identification of bacteria at genus level

The genus *Lactobacillus* is heterogeneous, with a G+C content of the DNA of its species varying from 32 to 54% (Nour, 1998). However, within a well-defined genus the G+C content should vary by no more than 10% range (Vandamme et al, 1996). A genus-specific primer was developed by Dubernet et al., (2002) through analyses of similarities between the nucleotide sequences of the spacer regions between the 16S and 23S ribosomal RNA genes of *Lactobacillus*. The specificity of this genus-specific primer was tested against 23 strains of lactobacilli of various origins, and resulted positive in only *Lactobacillus* strains among other LAB.

1.7. Identification of bacteria at species level

1.7.1. DNA-DNA hybridization

DNA-DNA hybridization or reassociation technique is based on a comparison between whole genome of two bacterial species. According to DNA-DNA hybridization method in bacterial species classification, a bacterial species have

70% or greater DNA-DNA hybridization with 5°C or less ΔTm values. However, it must be noted that this technique gives the relative % of similarity but not the actual sequence identity.

The DNA-DNA hybridization technique is based on denaturation and renaturation of DNA. It is based on three parameters i.e., i) G + C mol % , ii) the ionic strength of the solution and iii) the melting temperature of DNA hybrid (Tm). Tm is the only variable parameter out of three (as ionic strength can be kept constant). Therefore, more the similarity between the heteroduplex molecules, more temperature will be required to separate it (high Tm value). Till date more than 5000 bacterial species have been successfully delineated on the basis of this technique (Stackebrandt, 2003). DNA-DNA hybridization is the "gold standard" for proposed new species and for the definitive assignment of a strain with ambiguous properties to the correct taxonomic unit. However there exist significant disadvantages of DNA hybridization, like its being time-consuming, labor-intensive, and expensive to perform.

1.7.2. Hybridization

A nucleic acid probe is a fragment (20-30bp) of a single-stranded nucleic acid fragment that specifically hybridizes to complementary regions of the target nucleic acid. It can be used directly on a colony, or after DNA/RNA extraction. The target nucleic acid consists of single-stranded DNA or RNA molecules. Molecular probes may be labeled radioactively or non-radioactively. Radioactive labeling involves the phosphorylation of the 5' terminus of the probe with [32P] ATP. Non-radioactive labeling may be direct, using alkaline phosphatase or peroxidase, or indirect, by attachment of a ligand-protein or a hapten-antibody. Fluorescent probes (FISH: fluorescent in situ hybridization) may also be used. The extensive use of multiple oligonucleotide probes has become possible after important developments were achieved in sequencing of rRNA genes. Depending on the level of detection (genus or species), different regions of the genome might be used as targets. The sequences of 16S and 23S rRNA molecules contain highly

conserved regions that are common to all eubacteria, and also highly variable regions unique for the particular species. Thus, nucleic acid probes, in particular probes targeting rRNA gene sequences, have been extensively used in the analysis of dairy products (Ehrmann et al., 1994, Sghir et al., 1998). However, such rRNA gene probes cannot be used for closely related species due to the high level of similarity between their rRNA gene sequences. For example, such probes cannot distinguish *L. casei* from *L. paracasei* or *L. rhamnosus* (Hertel et al., 1993).

1.7.3. DNA-DNA microarray

To overcome the limitations of DNA-DNA hybridization, DNA microarray can be used. In this method, instead of whole genomic DNA, fragmented DNA can be used. DNA microarray technology can detect and measure thousands of distinct DNA sequences and used for high throughput, quantitative, systematic and detailed studies of microbial communities. DNA microarrays are basically glass surfaces spotted with numerous covalently linked DNA fragments (probes) that are available for hybridization. Current applications of DNA microarrays include monitoring gene expression (transcriptional profiling) or detecting DNA sequence polymorphisms or mutations in genomic DNA. This method gives resolution up to strain level and has been used in detecting virulence among the strains of pathogenic bacteria by identifying the strain-specific unique regions (Broekhuijsen et al., 2003). However, applications of this technology to microbial communities are still limited, because of the unknown composition of these samples (Call et al., 2003). DNA microarrays can identify and quantify bacteria in complex samples, based on the detection of taxonomic markers such as small subunit rRNA or small subunit rRNA gene (Desantis et al., 2005). These microarray studies use PCR amplification of rRNA gene with universal primers prior to hybridization (Wu et al., 2001). Use of randomly designed oligonucleotide probes is another approach. Dubois et al. (2004) developed a prototype microarray with 50-mer oligonucleotide probes designed from several bacterial taxonomic, functional and pathogenic genes. Hybridization of genomic DNA led to distinct hybridization

patterns for each group of product, but did not allow identification of the organisms.

1.7.4. 16S rRNA Gene Fingerprinting

Cloning and sequencing of 16S rRNA gene sequences is suitable for analysis of bacterial species in a culture-independent way, but this approach is unsuitable for monitoring communities, because the analysis is very laborious and expensive. Denaturing Gradient Gel Electrophoresis (DGGE) is a method of fingerprinting of 16S rRNA gene to monitor community shifts. The members of the bacterial community are often amplified using primers corresponding to the 16S rRNA gene sequence. DGGE method is based on the lower electrophoretic mobility of a partially melted double-stranded DNA molecule in polyacrylamide gels containing a linear gradient of DNA denaturants (urea and formamide). Once the domain with the lowest melting temperature reaches its melting temperature (T_m), in the denaturing or temperature gradient gel, the molecule undergoes a transition from a helical to a partially melted structure, and its migration stops. Species can then be distinguished by comparing the migration distance of the PCR amplicons in gels with those of reference strains. Muyzer et al., (1993) applied this method to study the bacterial diversity in a marine ecosystem. After this pioneering study, a variety of microbial ecosystems have been analyzed using DGGE or Temperature Gradient Gel Electrophoresis (TGGE). It has been reported that DGGE or TGGE are sensitive enough to represent bacteria that constitute up to 1% of the total bacterial community (Muyzer et al., 1993; Zoetendal et al., 1998). This means that only the most dominant bacteria will be represented in the profiles when domain-specific primers are used. On the other hand, *Bifidobacterium* and *Lactobacillus* group-specific PCR-DGGE approaches have also been developed to specifically amplify monitor these groups (Simpson et al., 2000; Satokari et al., 2001a,b; Walter et al., 2001). It was demonstrated that probiotics could be detected and monitored in feces using group-specific PCR-DGGE, which is difficult using general bacteria-specific primers (Simpson et al., 2000, Walter et al., 2001, Heilig et al., 2002). In TGGE, separation of fragments is based on electrophoretic

mobility of DNA subjected to a linear temperature gradient. Single strand conformation polymorphism (SSCP) is another method in analysis of microbial ecology by nonculture based methods (Godon et al., 2001). SSCP detects sequence variations between DNA fragments, usually amplified by PCR from variable regions of the 16S rRNA gene and uses neutral, nondenaturing polyacrylamide gels. First, double-stranded short DNA fragments are amplified, and then they are mixed with formamide and heated to create single-stranded DNA. The two complementary strands of DNA will migrate differently and will therefore separate during gel electrophoresis.

1.7.5. PCR-RFLP and ARDRA

PCR-RFLP (PCR-restriction fragment length polymorphism) analysis involves the amplification of a specific region, followed by restriction enzyme digestion. Giraffa et al, (2003) studied PCR-RFLP of three protein coding genes, and following restriction digestion, the researchers were able to differentiate three subspecies of *L. delbrueckii*. If the amplified gene is rRNA gene, the method is described as ARDRA (amplified ribosomal DNA restriction analysis). The discriminatory power of the method is dependent on the section and length of the amplified fragment and restriction enzymes. ARDRA patterns are highly reproducible and comparable between laboratories. However, some bacterial species show high rRNA gene sequence similarity and therefore, it may be difficult to select restriction endonucleases that produce distinct restriction patterns for closely related species (Roy and Sirois, 2000). Succesful results have been obtained by Rodas et al., (2003) in identification of LAB associated with wine at species level. Typing of *L. acidophilus*, *L. helveticus* and *L. delbrueckii* by ARDRA had been achieved by Giraffa et al., (1998a) at the species but not at the strain level. Bouton et al., (2002) confirmed by PCR-ARDRA strains isolated from Comté cheese belonged to *L. delbrueckii* subsp. *lactis*. For six probable *L. helveticus* strains, *Eco*RI was unable to digest rRNA gene, although fermentation profiles together with RAPD analysis suggested that all strains belonged to *L. helveticus*. Similar results were observed within by Andrighetto et al., (1998) in *L.*

helveticus strains. Rearrangements of the chromosome or cross-protection by endogenous methylation were proposed to explain the inability of *Eco*RI to cut the rRNA gene. Later, it was represented by Giraffa et al., (2000) that *L. helveticus* strains undigested by *Eco*RI had a single nucleotide substitution (C instead of T), located at the *Eco*RI site of the 16S rRNA (Table 3.1).

1.7.6. Ribotyping

Ribotyping is a variation of the restriction fragment length polymorphism (RFLP) analysis of the genomic DNA, where certain fragments were probed to obtain fingerprints. In ribotyping, DNA is transferred to a membrane for hybridization with a labeled 23S, 16S or 5S rRNA gene probe. The labeled probes hybridize with several fragments of multiple copies of rRNA operons. Fingerprint patterns consist of chromosomal DNA fragments that are derived from the rRNA operon and its adjacent regions which hybridize to the rRNA probe. In general, the fingerprint patterns obtained by this method are more stable and easier to interpret than those obtained by restriction enzyme analysis. The probes used in ribotyping vary from partial sequences of the rRNA gene or their spacer regions (Brandt et al., 2001) to the whole rRNA operon. If probes are designed from conserved regions of rRNA gene, they can be applied to a wide range of bacteria, even those that are phylogenetically distant. The discriminatory power of the technique is dependent on the size of the probe, but also on the restriction enzyme(s) used. A fully automated ribotyping system, the RiboPrinter® microbial characterization system, has been developed for identification at the genus, species and strain levels. This method is easy to carry out but the equipment and expendables are quite expensive. Giraffa et al, (1998b) used ribotyping to divide 26 *L. helveticus* strains into five, nine and ten ribotypes using restriction enzymes *Eco*RI, *Pvu*II and *Mlu*I, and the intergenic 16S–23S rRNA gene was used as the probe. Moreover, ribotyping has been shown to be a useful tool in differentiating human intestinal lactobacilli and bifidobacteria at both the species and strain level (Kimura et al., 1997, Tynkkynen et al., 1999).

1.7.7. Sequencing

Nucleic acid sequencing methods have undergone tremendous advances over the past decade. These rapid advances have made it possible for a small laboratory to determine the sequence of millions of base pairs of DNA per year. In addition to the speed of the sequencing, the quality of sequence data has also improved. In 1980 in the Approved Lists, 1,791 valid names were recognized at the rank of species. In January 2008, this number has increased to 8,651 species, a 483% increase (http://www.bacterio.cict.fr/number.html#total). The explosion in the number of recognized taxa is directly attributable to the ease in performance of 16S rRNA gene sequencing compared to the more cumbersome manipulations involving DNA-DNA hybridization. Clarridge (2004) calculated that in a clinical microbiological laboratory, 60 samples can be analyzed in a week. The period includes procedures starting from harvesting, to DNA isolation, PCR amplification, sequencing, up to analysis of sequence for species assignment

Cycle sequencing is similar to PCR in that it uses DNA (purified products of the first PCR cycle) as the template. Both the forward and reverse sequences are used as the template in separate reactions in which only the forward or reverse primer is used. Cycle sequencing differs from PCR in that no new template is formed (the same template is re-used for as many cycles as programmed, usually 25 cycles) and the product is a mixture of DNA of various lengths. This is achieved by adding specially labeled bases called dye terminators (along with unlabeled bases), which, when they are randomly incorporated in this second cycle, terminate the sequence. Thus, fragments of every size are generated. As each of the four added labeled terminator bases has different fluorescent dye, each of which absorbs at a different wavelength, the terminal base of each fragment can be determined by a fluorometer. The products are purified to remove unincorporated dye terminators, and the length of each is determined using capillary electrophoresis or gel electrophoresis. An electropherogram, which is a tracing of the detection of the separated fragments as they elute from the column (or are separated in the gel) where each base is represented by a different color, can be manually or

automatically edited. It is possible to have the fragments of various lengths so well separated that every base of a 500-bp sequence can be determined. When ambiguities occur, most of them can be resolved by visual re-editing of the electropherogram. In contrast to the accuracy achievable nowadays, with the excellent equipment and reagents available, some of the sequences deposited in GenBank, particularly those derived over 10 years ago, are not very accurate, indicated by presence of many ambiguous bases noted as N, R, Y, W, M, S, or K (meaning that the base is unknown, A or G, C or T, A or T, A or C, G or C, G or T, respectively). It is also possible that intracellular polymorphisms might cause difficulties in obtaining an easily interpretable sequence; i.e., since there are multiple copies of the 16S rRNA gene within a single-cell genome, there could be several different sequences and thus there could be two different base pairs at a given location. The existence of variant 16S rRNA gene alleles in a single genome has been clearly demonstrated in several reports (Ninet et al., 1996, Reischl et al., 1998, Ueda et al., 1999).

Well-known databases of 16S rRNA gene sequences that can be consulted via the World Wide Web are; GenBank (http://www.ncbi.nlm.nih.gov/, 10-01-2008), the Ribosomal Database Project (RDP-II) (http://rdp.cme.msu.edu/html/, 10-01-2008), the European Molecular Biology Laboratory (http://www.ebi.ac.uk/embl/, 10-01-2008) and National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/About/tools/index.html, 10-01-2008).

1.8. Analysis at strain level

Identification of starter strains is an important preliminary step for the selection of starter cultures, because technological, probiotic, antimicrobial and sensorial attributes are strain-specific and it may help to distinguish strains with particular technological properties. The typing of starter cultures may also be used for monitoring patented strains or distinguishing probiotic strains from natural isolates of the host gastrointestinal tract.

1.8.1. Restriction Enzyme Analysis (REA)

Restriction enzyme analysis (REA) involves digestion of chromosomal DNA with restriction endonucleases. The number of bands obtained generally ranged between 1,000 and 20,000 bp in size, and are dependent on the restriction enzymes used. The complexity of the banding pattern necessitates the use of computer-assisted analysis. Zhong et al., (1998) used this method to separate 64 strains of Lactobacilli, and discriminated strains, but the produced patterns were very complex.

1.8.2. Pulsed-field gel electrophoresis (PFGE)

Pulsed-field gel electrophoresis is stduy of whole genome with rare cutter restriction enzymes digestion like *Apa*I and *Not*I. The restriction fragments are resolved by pulsed-field gel electrophoresis which applies an alternating electric field from different directions. Using PFGE it is able to separate very large fragments, thus analysis the whole genome can be obtained. This method is highly discriminatory and reproducible. However, it is labor-intensive, time consuming, and equipment and expendables are expensive, and thus only a limited number of samples can be analyzed. Sanchez et al, (2004) studied PFGE and RAPD-PCR to assess the genetic diversity of LAB from Almagro eggplants. Their study suggested that combined analysis of RAPD-PCR and PFGE results in a better discrimination than when the methods were applied alone. Such analyses requires use of softwares like GelCompar (Applied-Maths, Belgium). Tynkkynen et al., (1999) compared PFGE, RAPD and ribotyping to 24 *L. rhamnosus* and *L. paracasei* strains and among these methods, PFGE was described as the most discriminatory method; it revealed 17 genotypes for the 24 strains studied.

1.8.3. RAPD-PCR

Randomly Amplified Polymorphic DNA (RAPD)-PCR is a common method in typing of LAB. The method is also known as arbitrarily primed PCR (AP-PCR). The method is commonly used to examine diversity in eukaryotes, including fungi, plants and animals, with particular utility in the field of population genetics (Perez et al., 1998). Differential display is a variation of this method, in that RNA is used to produce the fingerprints. The RAPD technique employs approximately 10 base pair random primers to locate random segments of genomic DNA to reveal polymorphisms. The primers adhere to a specific nucleotide segment of the genomic DNA, and amplification will occur if a suitable distance is present between priming sites (Figure 1.8).

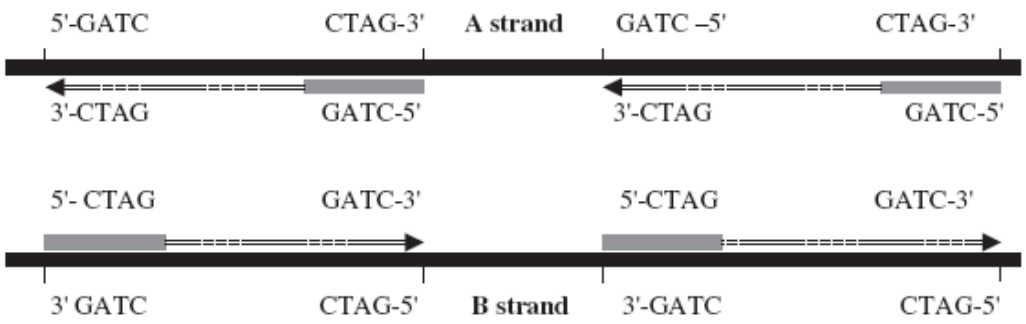


Figure 1.8. Random amplification of polymorphic sequences (Levin, 2003)

The DNA is cut into many segments of a specific length, which can be measured using gel electrophoresis. Differences in DNA fingerprints reflect either polymorphisms that have accumulated over evolutionary time or mutations that have developed during the life of the organism. For a mutation to change the RAPD pattern, it must occur either in the priming region or must change the length of the DNA between priming regions. Thus, RAPD analysis can provide a simple

and reliable method for measuring genomic variation. RAPD-PCR is a simple, rapid and sensitive method and does not require previous knowledge of the bacteria of interest. The main disadvantage of the method is that; it requires stringent conditions, since variable results between laboratories are common. This method is used in general to allow strain discrimination of previously identified species, however, RAPD-PCR can further be used to construct a fingerprint library using bioinformatics tools, and then serve as a reference for further studies, as long as highly standardized conditions are set (Rosetti and Giraffa, 2005). The discriminatory power of RAPD-PCR is compared in typing *Listeria monocytogenes*, and it was stated that if two or more primers are used, discriminatory power of RAPD-PCR equals to that of PFGE (Levin, 2003).

RAPD-PCR is a valuable method for the species whose genome has not been sequenced yet. Although several LAB species' genome have been sequenced, the vast majority of them is still unknown, and RAPD-PCR can be used to produce species specific probes and fingerprints, or species specific fingerprints may be used for unknown species assignment. In the meantime, it is a good method for discrimination purposes at the strain level for all LAB (Tilsala-Timisjarvi and Alatossava, 1998, Rodas et al., 2005). Table 1.4 compares RAPD-PCR with some phenotypic and genotypic methods.

The majority of amplified bands in RAPD-PCR usually originate from unique sequences. Intensity of a particular band between different samples is proportional to its corresponding template, thus RAPD-PCR can be regarded as a semi quantitative method (Welsh et al., 1995). It is strongly recommended that each genomic fingerprint be generated at several DNA concentrations, to reveal any products that show a significant dependence on template concentration.

Table 1.4. Criteria of methods suitable for identification of lactobacilli (Reuter et. al., 2002)

Method	Work load			Precision for identity of		
	Investive costs	Material	Time	Genus	Species	Strain
Phenotypic						
Morphological						
Physiological	±	±	±	++	+	±
Biochemical						
SDS-PAGE	+	+	+	++	++	+
Genotypic						
RAPD-PCR	+	+	++	*	+	++
Plasmid demonstration	*	*	*	*	±	+
REA	±	±	±	+	+	+
Pulsed field gel electrophoresis	++	++	++	+	+	++
DNA–DNA hybridization (dot blot)	+	+	± / +	++	++	*

++: very high, +: high; +/-: moderate; *: low.

RAPD-PCR was used to discriminate lactobacilli which could not be differentiated on the basis of (L (+)- nicotinamide adenine dinucleotide-dependent lactic dehydrogenase) electrophoretic profiles. Du Plessis and Dicks (1995), used RAPD-PCR to differentiate *Lactobacillus acidophilus, L. crispatus, L. amylovorus, L. gallinarum, and L. johnsonii* strains with identical L-NDH profiles, and stated that these strains could be distinguished on the basis of their RAPD-PCR profiles.

Several researchers make use of RAPD-PCR in discrimination of strains isolated from dairy products. Coppola et al., (2006) studied microbial diversity during the

manufacture of Fior di Latte di Agerola cheese, Kenny et al., (2005) studied Cheddar cheese, Sanchez et. al., (2005) studied goat milk cheese and Giraffa and Rosetti (2004) studied cheese starter cultures with RAPD-PCR to verify the microbiological intraspecific composition and to detect strain shifts in blended cultures. These studies require previous identifications mainly by using phenotypical methods. Afterwards, genotypic characterization using RAPD-PCR was performed.

Fermented foods are of high interest in dairy industry, and discovery of new species, subspecies and strains would be of special interest. Besides dairy products, other fermented foods are frequently studied for their microbial content, and one of these products is sourdough. Several species of LAB and yeasts contribute to sourdough fermentation. Sourdough can be defined as "a dough whose microorganisms originate from sourdough or a sourdough starter and re-metabolically active or can be reactivated, upon addition of flour and water they continue to produce acid". Sourdough LAB usually belong to the genus *Lactobacillus*, but occasionally, *Leuconostoc* species and *Enterococcus* species are found. *L. sanfranciscensis* is considered one of the most important species in sourdough, although *Lactobacillus pontis*, *Lactobacillus fructivorans*, *Lactobacillus fermentum*, *Lactobacillus brevis*, *Lactobacillus reuteri* and *Lactobacillus amylovorans* and also *Lactobacillus panis*, *Lactobacillus johnsonii*, *Lactobacillus acidophilus*, *Lactobacillus delbrueckii*, *Lactobacillus brevis*, *Weissella* and *Lactococcus lactis*. Commonly found yeasts include *Saccharomyces cerevisiae*, *Saccharomyces exiguus*, *Candida milleri* and *Candida krusei* (Catzeddu et. al., 2006).

Further, RAPD-PCR applications can be used in GMO (genetically modified organism) studies. Yoke-Kqueen and Radu (2006) successfully differentiated (GMO) from non-GMO ones using two arbitrary primers. The study investigated maize and soybean samples, and analysis of data enabled clustering of GMO and non-GMO samples on different clusters.

1.8.4. Multi Locus Sequence Typing (MLST)

Although sequencing of ribosomal RNA gene has been accepted as a reliable technique in bacterial identification, there exist some problems. The most important problem can be the highly conserved 16S rRNA gene, which may not be able to correctly identify closely related species. Another problem might result from the fact that rRNA gene does not encode a protein, and insertions and deletions in the sequence can cause problems in alignments. In the literature there exist different species with identical 16S rRNA sequences. Thus, methods that can be applied for classification below the species level can provide more detailed information of isolates.

Figure 1.9 suggests that 16S rRNA gene sequencing should be restricted to comparisons between genera in highly uniform species, because analysis of 16S rRNA gene analysis has high uniformity in these species. It is suggested that, in these species MLST is able to reveal clonal complexes, when hypervariable loci are investigated. However, in moderate to diverse species, 16S rRNA gene can be applied for species identification.

MLST can be used to define species, subspecies, and even strains. The method was first developed to characterize pathogenic microorganism *Neisseria meningitidis* (Maiden et al., 1998), and extended to several other pathogens including *Streptococcus pneumoniae, Yersinia* species*, Campylobacter jejuni and Helicobacter pylori* on MLST databases, mainly http://pubmlst.org/ and http://www.mlst.net/. Maiden (2006) provides outline of MLST in Figure 1.10. Although developed for pathogenic bacteria, MLST can be applied to nonpathogenic bacteria; namely lactobacilli, *Leuconostoc*, *Lactococcus* and *Pediococcus* as well as fungi (de las Rivas et al., 2006, Gil-Lamaignere et al., 2003).

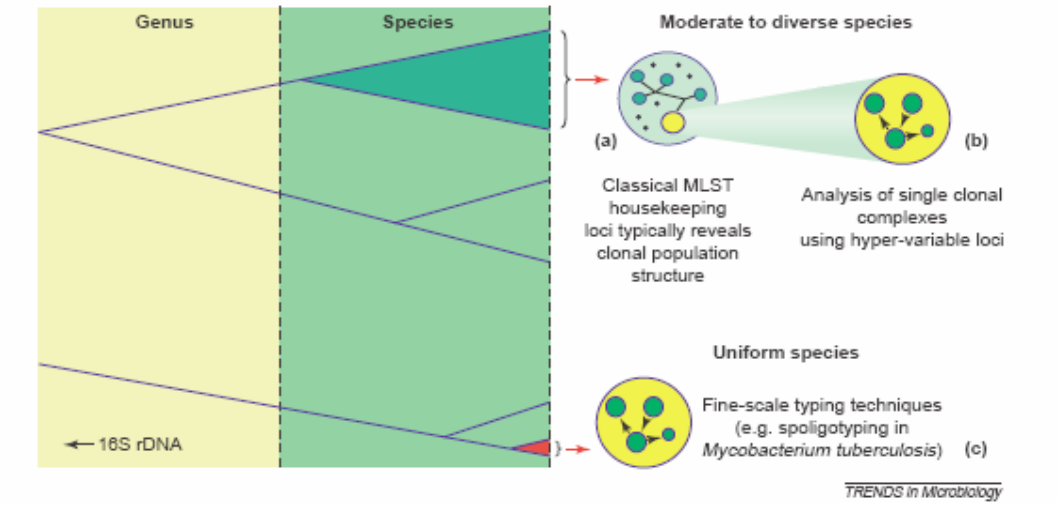


Figure 1.9. The different levels of resolution afforded by 16S rRNA gene sequencing and MLST (Cooper and Feil, 2004).

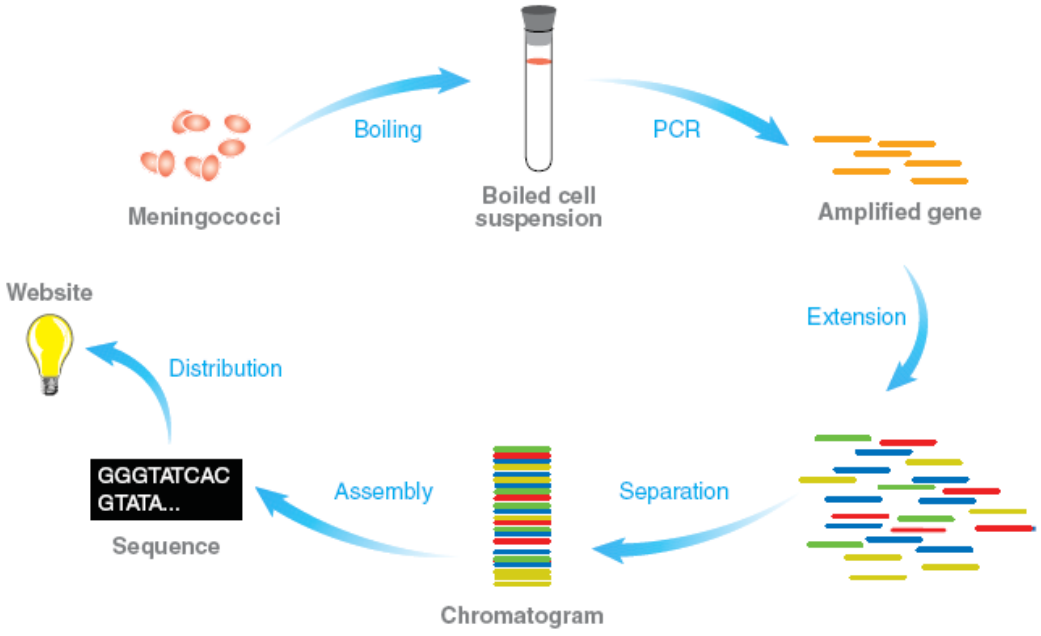


Figure 1.10. The outline of the MLST proposed by EU-MenNeT consortium (Maiden, 2006). DNA is extracted by simple boiling, and subjected to PCR analysis. Amplified gene is sequenced by high throughput methods, and evaluated by using several bioinformatics tools.

MLST has become a universal approach by the following advances in technology, a) whole genome sequences of bacteria has been released, leading to a better and deeper understanding of organisms, b) availability of high-throughput sequencing resulted in lower cost and increased speed sequencing, and c) internet became available worldwide by accessible to people from around the world, and provides easy transport of the information (Maiden, 2006).

The idea of MLST was developed on principles of MLEE (Multi Locus Enzyme Electrophoresis). MLEE was method of choice for identification of pathogens, mainly because at that times DNA sequencing was a time-consuming method. Nowadays, Sanger dideoxy termination method allows high throughput sequencing, allowing MLST feasible. MLEE analyses the electrophoretic motilities' of housekeeping enzymes on starch gels and equates the different charge variants of each enzyme with alleles at the underlying genetic locus. The method assumes that electrophoretic type of each bacterial clone is stable over time, since the number of nucleotide substitutions required to change the electrophoretic mobility of an allelic variant. MLEE requires analysis of 20 or more loci, resulting from the fact that only a small number of variants are detected at each locus (Enright and Spratt, 1999). Study of high number of loci in MLEE means a labor-intensive and time-consuming procedure, and furthermore, the resulting fingerprinting data is difficult to compare between laboratories.

MLST is a definite procedure for characterizing isolates of bacterial species using the sequences of house-keeping genes. In MLST, instead of evaluating enzyme fingerprints on gels, sequence analyses of the chosen genes are used. In this method, usually 450bp of the housekeeping genes are sequenced and mismatches at loci is designated as a different allele. Maiden et al., (1998) started MLST by studying 6 loci; however, it was extended to 7 loci for identification of invasive strains of meningicocci. The number of analyzed loci in MLST differs from one-to-seven (Gevers et al., 2005). For the analysis of MLST, each gene is designated as a locus, and for each gene, even a single nucleotide difference are assigned as

different alleles. The alleles at the seven loci provide an allelic profile (in the format of 1-3-2-1-4-3-2), defining the sequence type ST (for example, ST-2) of each isolate. The matrix of pair-wise differences between allelic profiles of isolates is done, and a dendrogram is constructed on this matrix. By using MLST, universal database based on gene sequence could be constructed with far more variation, and tracking of pathogens could be performed worldwide (Maiden et. al., 1998, Brehony et al., 2007) (Figure 1.11).

Definition at interspecies and intraspecies level may require use of different sets of genes, and a different term is suggested for interspecies level; Multi Locus Sequence Analysis (MLSA) (Gevers et al., 2005).

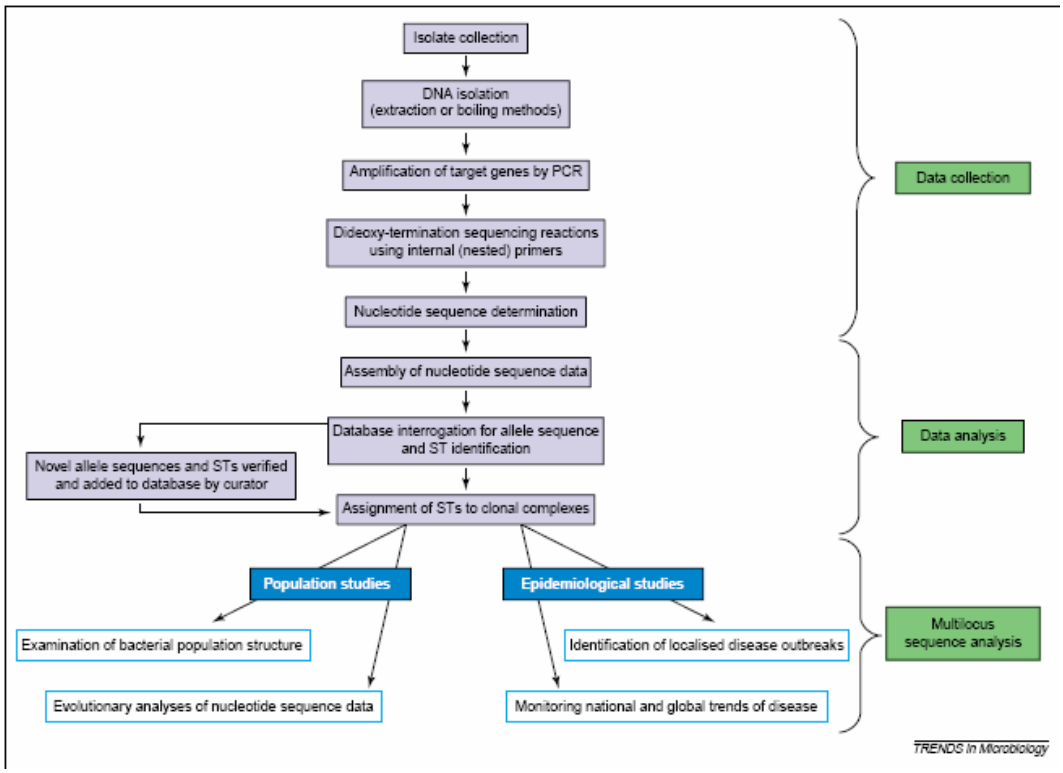


Figure 1.11. Flowchart of MLST analysis of bacterial pathogens. MLST is used in population and epidemiological studies (Urwin and Maiden 2003).

The key point of MLST lies in the choice of genes to be analyzed. Correctly chosen genes will successfully result in proper identification and differentiation of the species and subspecies. The most appropriate genes are the housekeeping genes. Reason for this choice is explained by selective evolutionary pressures upon genes. Housekeeping genes are under stabilizing selection to conserve their metabolic functions (Urwin and Maiden, 2003). Stabilizing selection results in decreasing the rate of nonsynonymous base substitutions (base substitutions that cause changes in the amino acid), while in diversifying selection nonsynonymous substitutions are favored. Examples to genes that are under diversifying selection are antigen coding genes of pathogens, because nonsynonymous base substitutions confer advantage to them.

Choice of housekeeping genes should be made on the basis of (Gevers et. al., 2005, Zeigler, 2003);

- presence in most/all of samples so that it can be applied to species whose identity is not fully established,
- presence in single copy, without close paralogues that could confuse analysis,
- gene should be long enough to contain useful information, as well as short enough for easy sequencing, (approx. 450bp)
- selected gene should be devoid of recombination and gene-linkage.

Evaluation of MLST data can be performed by several ways. For publicly available species of pathogens, results of the sequencing data may be submitted to database either by querying at a) single locus (a single sequence or a batch of sequences for a single locus to be compared with all known alleles), b) multiple locus (sequences of all seven loci for a single strain is submitted), or c) batch query (sequences of all seven loci for a batch of strains can be submitted). Submission of query will result in assignment of allelic number, or it may be related to the closest allele by showing percentage homology. In the latter case, if a new allele has been discovered, a curator of the database will assign a new allele

number. Each of allele numbers is combined to determine the allelic profile of the sample. This allelic profile will then be used to determine the sequence type (ST) of the sample, to describe the strain (Aanensen and Spratt, 2005).

Alternatively, all of the loci studied can be analyzed on the basis of nucleotide sequences. This approach is easy to perform, and can be for the species that are not present in publicly available databases. In this approach, all of the loci under investigation are concatenated in an order. Providing the order of gene sequences are the same for all sample, a neighbor joining or UPGMA tree is constructed using software such as MEGA4 (Tamura et al., 2007, Hanage et. al, 2005, Ahmed et. al, 2006, Aanensen and Spratt, 2005).

1.9. Availability of whole genome sequencing and age of bioinformatics

Since the first whole bacterial genome sequencing of *Haemophilus influenzae* Rd (Fleischmann et al., 1995) was completed in 1995, accelerating numbers of genome sequences are available. Genome sequencing of LAB will speed up the application of these bacteria in both traditional and non-traditional arenas. Although the genome size of LAB are small, (2–4 Mb), the LAB had been isolated from a diverse number of enviroments (foods, gastrointestinal and vaginal tracts, plants) suggesting that considerable genetic adaptation has occurred during their evolution. Comparative genomics among the microbes sequenced thus far has already illustrated that essential housekeeping gene functions are widely conserved among microbes and horizontal gene transfer commonly occurs (Klaenhammer et al., 2002). Whole genome sequencing provides a deeper understanding of LAB, since metabolic pathways can be elucidated, for use in both dairy industry and human health. Dairy industry especially focuses on metabolic pathways involved in fermentation ability, flavour development, probiotic properties, production of bacteriocins and resistance to bacteriophages. The latter case is especially important, since vulnerability of starter cultures to bacteriophages is an important economical problem in dairy industry. In this field, considerable development has been achieved, through the use of clustered regularly interspaced short palindromic

repeats (CRISPRs). These regions are composed of successive repeats of 24-47bp repeated sequences, separated from each other by spacers. The spacer sequences have been found to be similar to phage DNA sequences, and a direct correlation was obtained between numbers of identical spacers in phage resistant organisms to the increased exposure to phages. It is stated that bacteria have developed immunization to phage attack by using extrachromosomal elements from past invasions (Bolotin et al., 2005). Thus CRISPRs can be used to predict sensitivity to phages, or to prevent acquisition of unwanted genes like antibiotic resistance genes (Barrangou et al., 2007). In addition to deeper understanding of the particular organisms, genome sequencing provides evidence for a more reliable taxonomy of prokaryotes.

From an evolutionary approach, there exist several events in shaping of genomes. These include; gene duplications, gene loss, chromosomal rearrangements and horizontal gene transfer.

1. Gene duplication: whereas several gene duplication events was thought as the main force in evolution of prokaryotes, it is now clear that small gene duplications are more common. Evidence on mycobacterial genome sequencing reveals that gene duplication is especially important for adaptation to constantly changing environments (Coenye et al., 2005).

2. Gene loss: bacterial genomes counteract gene duplication and horizontal gene transfer by gene loss. Bolotin et al., (2004) states that 10% of *S. thermophilus* genes are pseudogenes which are inactive because of mutations, deletions and truncations. Most interesting gene families of this kind include transport proteins and carbohydrate metabolism. These gene families lost their function probably because milk provides a stable niche. In addition to energy related genes, *S. thermophilus* lost its virulence genes for the same reason. Sequencing of *L. delbrueckii* subsp. *bulgaricus* reveal similar gene loss events, these include transposases, protein coding genes, genes involved in carbohydrate metabolism, amino acid and cofactor

biosynthesis, and competence development. There exists differences in gene loss of *L. delbrueckii* subsp. *bulgaricus* and *S. thermophilus*, in that *L. delbrueckii* subsp. *bulgaricus* has lost most of its amino acid biosynthesis ability, while *S. thermophilus* retains, probably because *S. thermophilus* does not have extracellular proteases to utilize milk proteins (van de Guchte et al., 2007).

3. Horizontal gene transfer: Horizontal gene transfer is defined as the movement of genetic material between bacteria other than by descent. In contrast, vertical transfer occurs when an organism receives genetic material from its ancestor, e.g. its parent or a species from which it evolved. Horizontal gene transfer may be mediated through transformation, transduction or bacterial conjugation. Although the phenomenon of horizontal gene transfer was known, it was known to be present in a few cases; however genome sequencing has proven that it is a common way for adaptation to a new environment (Coenye et al., 2005). HGT results in mosaic species, organisms in which different portions of the genome have different histories, and evolutionary trees must take such transfer into account (Eisen, 2000). In LAB, horizontal gene transfer is mediated through IS-elements, bacteriophages, and mobile genetic elements, and result in unique and beneficial properties of LAB. In *S. thermophilus*, more than 50 insertion sequences have been found, associated with adaptation to milk. An interesting example of horizontal gene transfer is presented in Figure 1.12 (Bolotin et al., 2004), representing 95% homology between *S. thermophilus* and *L. delbrueckii* subsp. *bulgaricus.* Methionine is a rare amino acid in milk, and ability to produce methionine would confer advantage to the bacteria. It is suggested that close association of two species facilitated gene transfer Figure 1.13.

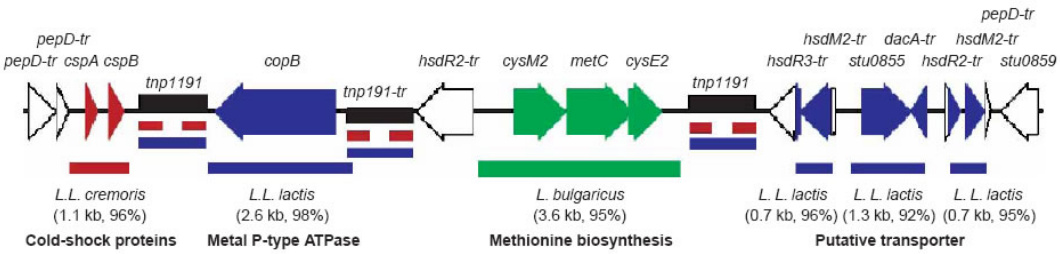


Figure 1.12. Horizontal gene transfer between *S. thermophilus* and other starter bacteria (Bolotin et al., 2004).

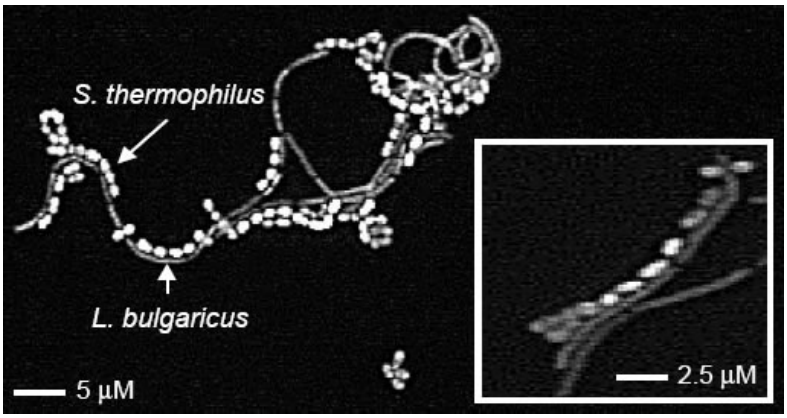


Figure 1.13. Adhesion of *S. thermophilus* and *L. delbrueckii* subsp. *bulgaricus* (Bolotin et al., 2004).

4. Chromosomal rearrangements: Together with gene loss, duplication and transfer events, genome rearrangements are important for evolution. Analysis of the first two sequenced genomes, *H. influenzae* and *M. genitalium*, revealed that, gene order was not conserved throughout bacterial evolution (Mushegian and Koonin, 1996). In prokaryotes, gene order is conserved to a much lesser extent than average protein sequence (Wolf et al., 2001). Vandamme et al., (2005) state that IS elements play

important roles in genome rearrangements, since they offer similar sequences where recombination can start. Chromosomal rearrangements may have a tremendous effect on banding patterns generated by PFGE or PCR based methodologies, since the loss or gain of restriction sites and/or primer-binding sites could result in altered patterns.

Understanding whole genome sequences is necessary to study bacterial systematics in the age of high throughput systems. The accelated speed of whole genome sequences can only be interpreted by using several bioinformatics tools. These tools provide analysis of genomes; make comparisons between them, and provide re-evaluation of bacterial systematics and phylogenetics. The basic tools can be performed on-line, provided mainly by databases. Most importantly, databases are available to the public, enabling information to laboratories all over the world. The best known websites for nucleotide sequences and analysis tools are NCBI (National Center for Biotechnology Information), European Bioinformatics Institute (EBI), the DNA Data Bank of Japan (The International Nucleotide Sequence Database Collaboration), Integrated Microbial Genomes (IMG) system, Ribosomal Database Project (RDPII). Study of dairy starter bacteria in literature requires software to analyze data using different methods, and the most widely used softwares are PHYLIP, MEGA4 (Tamura et al., 2007), START2 (Jolley, 2001) which are freeware, while sophisticated softwares capable of performing diverse applications, like QuantityOne (BioRad, USA) and Bionumerics (Applied-Maths, Belgium) are quite expensive, especially the latter.

1.10. Scope of the study

Yoghurt is a popular milk product in Turkey, and commercial yoghurt has been produced by many local, national, and international companies. The estimated production was 1 million tons in 2006. In addition to commercial production, many people prefer traditional yogurt, in which fermentation is achieved by addition of previous yoghurt to milk. In industry, yoghurt is fermented by synergistic action of yoghurt starter bacteria, *S. thermophilus* and *L. delbrueckii*

subsp. *bulgaricus*. The starter cultures are imported from European countries, and the market of starter cultures is big, considering high amount of consumption of yoghurt. Yoghurt quality depends largely on properties of milk, and starter culture (Bonczar et al., 2002), and starter industry is in search for new strains, to produce healtier, better tasted, phage resistant with longer shelf life foods. Identification of new starter cultures can be obtained by using phenotypic or genotypic methods. Phenotypic methods are time consuming, labor intensive, and may misidentify isolates. On the other hand, genotypic methods are quick, sensitive and reliable. In our study, it is aimed to identify isolates using genotypic methods, which were previously identified by phenotypic methods. The literature survey in genotypic identification of yoghurt starter bacteria led to insufficient or ambigious results, since identification at subspecies level could not be achieved by single PCR, and can only be achieved after species identification, however, species identification led to erroneous results in our laboratory. The requirement of a new method for identification of starter cultures was concluded. Identification of *S. thermophilus, L. delbrueckii* subsp. *bulgaricus, L. delbrueckii* subsp. *lactis* and *L. helveticus* was achieved using a new PCR based method. The method identified yoghurt starter isolates as *S. thermophilus* and *L. delbrueckii* subsp. *bulgaricus*, and results were confirmed by sequencing of 16S rRNA genes. In addition to successful identification of starters at species and subspecies level, the method enables selective identification of yoghurt starter bacteria from mixed cultures. Furthermore, the method is able to identify important cheese starters; *L. delbrueckii* subsp. *lactis* and *L. helveticus* from mixed cultures. After successful identification of yoghurt starters at species and subspecies level has been achieved, the strains were typed at strain level, since properties of starter cultures are strain dependent. Typing of yoghurt starter bacteria at strain level has been performed using RAPD and MLST.

CHAPTER 2

MATERIALS AND METHODS

2.1. Bacterial strains and growth conditions

A total of 100 bacterial strains were used in this study including reference strains, bacteria isolated from commercial starter cultures and from traditional home-made yoghurts (Table 2.1). Traditional yoghurts were collected from Mediterranian Region of Turkey, and specific designation could not be assigned, except for strains starting with K1, from Kemah (Erzincan).

The reference strains NRRL were kindly provided by Dr. Alejandro Rooney of Agricultural Research Service Culture Collection (ARS, US Department of Agriculture), reference strains LMG 18311 and CNRZ 1066 were kindly provided by INRA (France), and strain ATCC BAA-365 were kindly provided by Dr. J. Steele, as freeze dried cultures.

Freze dried cultures were activated 3 times at MRS and M17 broth (Merck) at 37°C (lactobacilli other than yoghurt starter bacteria, lactococci and streptococci) and at 42°C (*L. delbrueckii* subsp. *delbruekii* and *S. thermophilus*) for use, and maintained on Microbanks (Pro-Lab Diagnostics) and 20% glycerol stocks at -80°C.

8 commercial strains were isolated from commercial starter cultures and 38 samples of traditional home-made yoghurt were collected from rural areas of

Table 2.1. The bacterial strains used in the study

Species or subsp.	Strain number	Source
L. fermentum	NRRL-B1840	ARSCC[1]
L. paraplantarum	NRRL-B23115	ARSCC[1]
L. casei subsp. *casei*	NRRL-B1922	ARSCC[1]
L. gasseri	NRRL-B4240	ARSCC[1]
L. rhamnosus	NRRL-B442	ARSCC[1]
L. helveticus	NRRL-B4526	ARSCC[1]
L. paracesi subsp. *paracasei*	NRRL-B4560	ARSCC[1]
L. reuteri	NRRL-B14171	ARSCC[1]
L. salivarius subsp. *salivarius*	NRRL-B1949	ARSCC[1]
L. johnsonii	NRRL-B2178	ARSCC[1]
L. pentosus	NRRL-B227	ARSCC[1]
L. amylovorous	NRRL-B4540	ARSCC[1]
L. brevis	NRRL-B4527	ARSCC[1]
L. acidophilus	ATCC 4356	ATCC[2]
L. delbrueckii subsp. *delbruekii*	NRRL-B763	ARSCC[1]
L. delbrueckii subsp. *lactis*	NRRL-B4525	ARSCC[1]
L. delbrueckii subsp. *bulgaricus*	NRRL-B548	ARSCC[1]
L. delbrueckii subsp. *bulgaricus*	DSM20081	DSMZ[3]
L. delbrueckii subsp. *bulgaricus*	ATCC BAA-365	University of Wisconsin, Madison
S. dysgalactiae	NRRL-B688	ARSCC[1]
S. equinus	NRRL-B3573	ARSCC[1]
S. infantarius	NRRL-B41208	ARSCC[1]
S. thermophilus	LMG18311	INRA[4]
S. thermophilus	CNRZ1066	INRA[4]
Lc. lactis	3113	JC17[5]
Lc. lactis	2088	-
Lc. lactis	2910	-
Lc. lactis	2911	-
L. delbrueckii subsp. *bulgaricus*	B1000-1	Visby
L. delbrueckii subsp. *bulgaricus*	B1000-2	Visby
L. delbrueckii subsp. *bulgaricus*	B1000-3	Visby
L. delbrueckii subsp. *bulgaricus*	Yo-mix 410-1	Danisco
L. delbrueckii subsp. *bulgaricus*	LB340-2	Danisco
L. delbrueckii subsp. *bulgaricus*	LB340-3	Danisco
S. thermophilus	Ta 040-1	Danisco
S. thermophilus	Ta 040-2	Danisco
S. thermophilus	B1000-3	Visby
S. thermophilus	Yo-mix 410-3	Danisco
L. delbrueckii subsp. *bulgaricus*	M2-5	Isolate from traditional yoghurt
L. delbrueckii subsp. *bulgaricus*	M2-14	Isolate from traditional yoghurt
L. delbrueckii subsp. *bulgaricus*	M2-18	Isolate from traditional yoghurt
L. delbrueckii subsp. *bulgaricus*	K1-10	Isolate from traditional yoghurt
L. delbrueckii subsp. *bulgaricus*	K1-16	Isolate from traditional yoghurt
L. delbrueckii subsp. *bulgaricus*	K1-29	Isolate from traditional yoghurt
L. delbrueckii subsp. *bulgaricus*	K1-33	Isolate from traditional yoghurt
L. delbrueckii subsp. *bulgaricus*	K1-38	Isolate from traditional yoghurt
L. delbrueckii subsp. *bulgaricus*	K1-43	Isolate from traditional yoghurt

Table 2.1. The bacterial strains used in the study (cont'd)

Species or subsp.	Strain number	Source
L. delbrueckii subsp. *bulgaricus*	K1-44	Isolate from traditional yoghurt
L. delbrueckii subsp. *bulgaricus*	M23-2	Isolate from traditional yoghurt
L. delbrueckii subsp. *bulgaricus*	M23-3	Isolate from traditional yoghurt
L. delbrueckii subsp. *bulgaricus*	M23-4	Isolate from traditional yoghurt
L. delbrueckii subsp. *bulgaricus*	M23-13	Isolate from traditional yoghurt
L. delbrueckii subsp. *bulgaricus*	N2-2	Isolate from traditional yoghurt
L. delbrueckii subsp. *bulgaricus*	N2-4	Isolate from traditional yoghurt
L. delbrueckii subsp. *bulgaricus*	N2-5	Isolate from traditional yoghurt
L. delbrueckii subsp. *bulgaricus*	N3-2	Isolate from traditional yoghurt
L. delbrueckii subsp. *bulgaricus*	N4-3	Isolate from traditional yoghurt
L. delbrueckii subsp. *bulgaricus*	N6-2	Isolate from traditional yoghurt
L. delbrueckii subsp. *bulgaricus*	K2-1	Isolate from traditional yoghurt
L. delbrueckii subsp. *bulgaricus*	K2-2	Isolate from traditional yoghurt
L. delbrueckii subsp. *bulgaricus*	K2-3	Isolate from traditional yoghurt
L. delbrueckii subsp. *bulgaricus*	K2-4	Isolate from traditional yoghurt
L. delbrueckii subsp. *bulgaricus*	K2-5	Isolate from traditional yoghurt
L. delbrueckii subsp. *bulgaricus*	M21-3	Isolate from traditional yoghurt
L. delbrueckii subsp. *bulgaricus*	M21-4	Isolate from traditional yoghurt
L. delbrueckii subsp. *bulgaricus*	M2-8	Isolate from traditional yoghurt
L. delbrueckii subsp. *bulgaricus*	M2-12	Isolate from traditional yoghurt
L. delbrueckii subsp. *bulgaricus*	M2-16	Isolate from traditional yoghurt
L. delbrueckii subsp. *bulgaricus*	M2-17	Isolate from traditional yoghurt
L. delbrueckii subsp. *bulgaricus*	M2-20	Isolate from traditional yoghurt
L. delbrueckii subsp. *bulgaricus*	M2-21	Isolate from traditional yoghurt
L. delbrueckii subsp. *bulgaricus*	K1-19	Isolate from traditional yoghurt
S. thermophilus	N2-1	Isolate from traditional yoghurt
S. thermophilus	N2-3	Isolate from traditional yoghurt
S. thermophilus	N2-4	Isolate from traditional yoghurt
S. thermophilus	N3-1	Isolate from traditional yoghurt
S. thermophilus	N6-1	Isolate from traditional yoghurt
S. thermophilus	N6-3	Isolate from traditional yoghurt
S. thermophilus	N5-4	Isolate from traditional yoghurt
S. thermophilus	N5-7	Isolate from traditional yoghurt
S. thermophilus	N-1	Isolate from traditional yoghurt
S. thermophilus	N7-4	Isolate from traditional yoghurt
S. thermophilus	N8-2	Isolate from traditional yoghurt
S. thermophilus	S1-3	Isolate from traditional yoghurt
S. thermophilus	K1-1	Isolate from traditional yoghurt
S. thermophilus	K1-7	Isolate from traditional yoghurt
S. thermophilus	K1-9	Isolate from traditional yoghurt
S. thermophilus	K1-12	Isolate from traditional yoghurt
S. thermophilus	K1-13	Isolate from traditional yoghurt
S. thermophilus	K1-14	Isolate from traditional yoghurt
S. thermophilus	K1-15	Isolate from traditional yoghurt
S. thermophilus	K1-16	Isolate from traditional yoghurt
S. thermophilus	K1-18	Isolate from traditional yoghurt
S. thermophilus	K1-19	Isolate from traditional yoghurt

Table 2.1. The bacterial strains used in the study (cont'd)

Species or subsp.	Strain number	Source
S. thermophilus	K1-23	Isolate from traditional yoghurt
S. thermophilus	K1-24	Isolate from traditional yoghurt
S. thermophilus	K1-27	Isolate from traditional yoghurt
S. thermophilus	K1-28	Isolate from traditional yoghurt
S. thermophilus	K1-29	Isolate from traditional yoghurt
S. thermophilus	K1-30	Isolate from traditional yoghurt

[1] Agricultural Research Service Culture Collection, USA
[2] American Type Culture Collections
[3] German Resource Centre for Biological Material
[4] National Institute of Agronomical Research, France
[5] Piard et al. 1993

Turkey and phenotypic identifications of isolates were performed by Neslihan Altay according to colony morphology on agar, microscopic examination, gram staining and catalase production. Gram positive and catalase negative strains were further investigated for growth at different pH values (pH 2, pH 10) and different temperatures (10°C, 45°C) and carbohydrate fermentation profiles were studied using microtiter plates.

2.2. Chemicals and Enzymes

The list of chemicals and enzymes used and their suppliers were provided in Appendix A.

2.3. Growth Media, Buffers and Solutions

The preparation of the growth media, buffers and solutions used were provided in Appendix B.

2.4. DNA isolation

Preparation of genomic DNA was performed either by manual isolation using phenol-chloroform or by using Tissue and Cell Genomic DNA Isolation Kit (Genemark, Taiwan). Manual DNA isolation was performed as follows; 2ml of overnight grown culture was centrifuged at 15000 rpm (Andreas Hettich, Germany) for 2 min, and washed twice in TE buffer (10mM TrisHCl (AppliChem), 1mM EDTA (AppliChem), pH 8.0). The pellet was dissolved in 300µl TE buffer and 5µl lysozyme (AppliChem) (50mg/ml) was added to lyse the cells. The solution was incubated at 37°C for 45min. Thereafter, 20µl of EDTA (Merck) (0.25M), 25µL SDS (Merck) (10%), and 4µl of Proteinase K (Fermentas) (20mg/ml) was added to the solution and incubated at 60°C for 60min. Once digestion was complete, samples were extracted with phenol-chloroform three times, and ethanol precipitated with the addition of 0.1V 3M sodium acetate (Merck) (pH 5.5). DNA has been precipitated; samples were washed with 70% ethanol (Merck), air-dried for 20min and dissolved in distilled H_2O. RNase (Fermentas) was used to digest RNA, and samples were stored at -20°C until use.

DNA isolations of bacteria isolated from traditional yoghurts were performed using Genemark DNA isolation kit. Briefly, 2ml of overnight grown culture was centrifuged at 15.000rpm for 2min. The pellet was suspended in 180µl of lysozyme buffer. Bacterial suspension was incubated at 37°C for 30min and 200µl extraction solution was added. Addition of 20µl Proteinase K (Fermentas) was followed by incubation at 56°C for 1h with occasional vortexing. 4µl of RNase A (Fermentas) was added and incubated at room temperature for 10min. Afterwards, 200µl of binding solution was added, mixed by vortex and incubated at 70°C for 10min. 200µl of ethanol was added, vortexed and the resulting solution was applied to the spin column with collection tube. The mixture was centifuged at 15.000rpm for 1min and flow through was discarded. The spin column was washed with 300µl of binding buffer, flow through was discarded, and the column was further washed with 650µl wash solution. The flow through was discarded and the spin column was further centrifuged for 5min to remove residual ethanol.

Collection tube was discarded and the spin column was placed into a new 1.5ml tube. Into this column, 100-200µl of elution buffer at 70°C was added, incubated for 2min, and DNA was eluted by centrifugation at 15.000rpm for 1min. The DNA was stored at -20°C.

2.5. Determination of DNA concentration

For reference and commercial strains the quality and quantity of DNA isolations were tested both with A_{260} /A_{280} ratios and on agarose gels with comparison to different concentrations of λ DNA, and for traditional yoghurt isolates it was tested on agarose gels with comparison to different concentrations of λ DNA.

The concentration of DNA in agarose gels was calculated according to the following formula 1.1:

$$\frac{L_{FRAGMENT}}{TL_{MARKER}} \times C_{MARKER} \times V_{MARKER} \times I_{BAND} \times \frac{1}{V_{DNA}} \quad (1.1)$$

Where $L_{FRAGMENT}$ was the length of the fragment, TL_{MARKER} was the total length of the marker, C_{MARKER} was the concentration of marker; V_{MARKER}, was the volume of marker, I_{BAND} was the intensity of the band and $V_{DNA}$was the volume of loaded DNA. The sizes and concentrations of the markers were displayed in Appendix C.

The concentrations of DNA with A260 /A280 ratios were calculated as follows; absorbance was measured at 260 nm and 280 nm. The ratio of A260/A280 represented the purity of the sample (ratios of 1.8=pure, 2.0≥RNA contamination, 1.6≤ protein contamination). Samples having ratios between 1.8 and 2.0 were used, and DNA concentrations were expressed as ng/µl.

DNA concentration= OD_{260} X 50 x Dilution factor

(OD260 stands for optical density in A_{260} and 50 represent the A_{260} unit of dsDNA).

2.6. Visualization of Gels

Electrophoresis was carried out on a horizontal electrophoresis apparatus, GelDocXR (BioRad, USA). For visualizing genomic DNA on agarose gel, 0.8 % agarose gel was prepared in 1X TAE Buffer (Appendix B). The gel was placed into the electrophoresis tank filled with 1X TAE buffer for covering the gel. Electrophoresis was performed at 80 volt for 60 minutes. After gels were run, post-staining of gels was performed in 1.0 μg/ml EtBr solution for 20 minutes. De-staining was perfomed to wash excess EtBr from gels. Finally, the gel was placed to GelDocXR (BioRad, USA) and illuminated under UV and photographed.

2.7. Primer design for Polymerase Chain Reaction (PCR)

Specific primers were designed to identify bacteria using Primer 3 (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3, 26-06-2007) and oligocalc (http://www.basic-northwestern.edu/biotools/oligocalc.html,26-06-2007). Analysis of primer binding sites and comparisons of different species were performed using ClustalW (http://www.ebi.ac.uk/Tools/clustalw2/index.html, 06-12-2008). Specifities of primers were checked using BLAST (http://www.ncbi.nlm.nih.gov/BLAST/, 06-12-2008). All of the primers were ordered from Iontek (Istanbul). Primer binding sites of specific primers and amplification regions of genes were shown in Appendix D.

2.7.1. Species specific primers for *L. delbrueckii* and *S. thermophilus*

Primer pair DelI/II were used for specific identification of *L. delbrueckii*, and primer pair ThI/II were used for specific identification of *S. thermophilus* species

(Tilsala-Timisjarvi A. and Alatossava T., 1997). Expected product sizes were 210bp and 260bp, respectively (Table 2.2).

2.7.2. Primers for restriction analysis of 16S rRNA gene and ITS region (ARDRA)

The ARDRA analyses performed in this section used primer pairs forward A (Mora et al. 1998) and reverse L1 (Jensen et al. 1993, Yavuz et al., 2004). The primers cover 16S rRNA gene and ITS regions, and expected product size was 1700bp (Table 2.2).

ARDRA with *Eco*RI was performed using primers 9699 and 9700 (Delley and Germond, 2002). The expected product size was 1500bp.

2.7.3. Methionine Biosynthesis Gene Specific Primers

Several primers were designed to achieve specific primers for detection of *L. delbrueckii* subsp. *delbruekii* and *S. thermophilus,* and primer pair cysmet2 was chosen, because its specificity was higher than other primers. Expected product size of cysmet2 was 750 bp, covering the partial gene sequences of cysteine synthase and cystathionine beta-lyase (Table 2.2).

2.7.4. Primers for sequencing analysis of 16S rRNA gene

Primers for partial sequencing of 16S rRNA gene were forward primer A (Mora et al. 1998) and reverse primer U926 (Baker et al., 2003).

2.7.5. Primers for RAPD analysis

Primers M13 and 1254 were suitable for typing of LAB from food products (Delfederico et al., 2006).

Table 2.2. Primers used in the study

Primers	Target Gene	Primer Sequence	Product Length	References
DelI DelII	ITS (Intergenic spacer region)	ACGGATGGATGGAGAGCAG GCAAGTTTGTTCTTTCGAACTC	210bp	Tilsala-Timisjarvi A. and Alatossava T., 1997
ThI ThII	ITS (Intergenic spacer region)	ACGGAATGTACTTGAGTTTC TTTGGCCTTTCGACCTAAC	260bp	Tilsala-Timisjarvi A. and Alatossava T., 1997
Cysmet2f Cysmet2r	Methionine biosynthesis	GGAACCTGAAGGCTCAAT GTCAACCACGGTAAAGGTC	750bp	This study
9699 (forward) 9700 (reverse)	16S rRNA	ATCCGAGCTCAGAGTTTGATCCTGGC TCAGGTCGACGCTACCTTGTTACGAC	1500bp	Delley and Germond, 2002
A (forward) L1 (reverse) U926R (reverse)	16S rRNA Intergenic spacer region 16S rRNA	AGAGTTTGATCCTGGCTCAG CAAGGCATCCACCGT CCGTCAATTCCTTTRAGTTT	1700bp 926bp	Mora et al., 1998 Jensen et al., 1993 Baker et al., 2003
bgalac2f bgalac2r	Beta-galactosidase	GATTGAAAGCCGGATGTATG CAAAACGGAGTCTTCCTTGA	850bp	This study
rpoA-21-F rpoA-23-R	RNA polymerase alpha subunit	ATGATYGARTTTGAAAAACC ACHGTRTTRATDCCDGCRCG	800bp	Naser et al., 2005

Table 2.2. Primers Used in The Study (cont'd)

Primers	Target Gene	Primer Sequence	Product Length	References
pheS-21-F pheS-22-R	phenylalanyl-tRNA synthase	CAYCCNGCHCGYGAYATGC CCWARVCCRAARGCAAARCC	500bp	Naser et al., 2005
M13 1254	Whole genome Whole genome	GAGGGTGGCGGTTC CCGCAGCCAA		Delfederico et al., 2006

2.7.6. Primers for MLST analysis

Analysis of *pheS* gene was obtained using primers pheS-21-F and pheS-21-R, with an expected amplification product of 500bp (Naser et al., 2005). Analysis of *rpoA* gene was performed using primers rpoA-21-F and rpoA-23-R, with expected amplicon size of 800bp (Naser et al., 2005). Analysis of beta-galactosidase gene was performed using primers bgalac2f and bgalac2r. The primers were designed using Primer3 and oligocalc programs, with an amplification product of 800bp.

Primer binding sites and primers' sequence were supplied in Appendix D and Table 2.2, respectively.

2.8. Polymerase Chain Reaction

The polymerase chain reactions (PCR) were carried out on made on MJMini thermal cycler (BioRad, USA) machine. For each series, a master mix was prepared. Each PCR reaction mix contained 1X PCR Buffer (Fermentas, Lithuania), $MgCl_2$ (Fermentas, Lithuania), dNTP (Fermentas, Lithuania), Taq Polymerase (Fermentas, Lithuania), and ddH_2O. Reactions were conducted with 500ng template DNA and in a total of 50µl. In order to check out reagent contamination, one tube was prepared with ddH_2O instead of template DNA.

2.8.1. Species specific PCR for *L. delbrueckii*

Primer pair DelI/II were used for specific identification of *L. delbrueckii* species (Tilsala-Timisjarvi A. and Alatossava T., 1997). Reaction mixture and amplification conditions were displayed in Tables 2.3 and 2.4.

Table 2.3. PCR reaction mixture for specific identification of *L. delbrueckii*

Reaction Components	Final Concentration
Sterile ddH_2O	-
$MgCl_2$	1.5mM
PCR reaction buffer	1X
dNTP	200µM of each dNTP
Forward primer	1µM
Reverse primer	1µM
Taq DNA polymerase	0.5U
DNA	500ng

Table 2.4. Amplification conditions for specific identification of *L. delbrueckii*

	Cycles	Temperature	Time
Initial denaturation		95°C	2 min
Denaturation	30	95°C	30 sec
Annealing		62°C	30 sec
Extension		72°C	30 sec
Final extension		72°C	10 min

Reference *L. delbrueckii* strains; *L. delbrueckii* subsp. *bulgaricus* DSM20081, *L. delbrueckii* subsp. *lactis* NRRL-B4525, *L. delbrueckii* subsp. *delbrueckii* NRRL-B763 were used as positive controls. *L. helveticus* NRRL-B4526 and *S. thermophilus* LMG18311 were used as negative controls.

Amplification products were loaded to 1.5 % agarose gels, run on electrophoresis at 80V for 45 min (BioRad), post-stained with EtBr, de-stained, and visualized under UV.

2.8.2. Species specific PCR for *S. thermophilus*

Primer pair ThI/II were used for species specific PCR (Tilsala-Timisjarvi A. and Alatossava T., 1997). Reaction mixture was as same as for *L. delbrueckii* (Table 2.3) and amplification conditions were same as displayed in table 2.4, except that annealing temperature was raised from 55°C to 62°C.

2.8.3. ARDRA for identification of *L. delbrueckii* subsp. *bulgaricus*

For restriction analysis with *Eco*RI, the amplification of 16S rRNA gene was performed using 9699-9700 primers (Delley et al. 2002) and for restriction analysis with *Taq*I and *Hpa*II, the amplification of 16S rRNA gene was performed using forward A (Mora et al. 1998) and reverse L1 (Jensen et al. 1993, Yavuz et al., 2004). Reaction mixtures and amplification conditions for ARDRA were listed in Tables 2.5 and 2.6.

Table 2.5. PCR reaction mixture for ARDRA analysis for specific identification of *L. delbrueckii* subsp. *delbrueckii* and *S. thermophilus*

Reaction Components	**Final Concentration**
Sterile ddH_2O	-
$MgCl_2$	1.5mM
PCR reaction buffer	1X
dNTP	200µM of each dNTP
Forward primer	1µM
Reverse primer	1µM
Taq DNA polymerase	0.5U
DNA	500ng

Table 2.6. PCR reaction conditions for methionine biosynthesis gene

	Cycles	Temperature	Time
Initial denaturation		95°C	2 min
Denaturation	35	95°C	1 min
Annealing		56°C	1 min
Extension		72°C	1 min
Final extension		72°C	10 min

Amplification products were loaded to 1.5 % agarose gels, run on electrophoresis at 80V for 45 min (BioRad), post-stained with EtBr, de-stained, and visualized under UV.

2.8.4. Restriction analysis of 16S rRNA gene

Restriction was carried on using *Eco*RI, *Taq*I or *Hpa*II as the enzyme (Delley and Germond, 2002). The reaction mixture was listed in Table 2.7.

Table 2.7. Reaction components of ARDRA

Reaction Components	Final Volume
Sterile ddH_2O	-
Restriction enzyme buffer	1X
PCR product	8,5µl
Restriction enzyme	0.5µl

The restriction products were run on 1.5% agarose gels, electrophoresed at 90V for 60 min, post-stained with ethidium bromide, de-stained and visualized under UV using GelDocXR (BioRAD).

2.8.5. Methionine Biosynthesis Gene Specific PCR

For amplification of methionine biosynthesis gene, primer pair cysmet2 (Table 2.2) was used. Genbank accession numbers of methionine biosynthesis genes were; YP_141251 and YP_141252, for cysteine synthase and cystathionine beta-lyase, respectively. Table 2.8 displays reaction mixture for amplification.

Table 2.8. PCR reaction mixture for methionine biosynthesis gene

Reaction Components	**Final Concentration**
Sterile ddH_2O	-
$MgCl_2$	1.5mM
PCR reaction buffer	1X
dNTP	200µM of each dNTP
Forward primer	2µM
Reverse primer	2µM
Taq DNA polymerase	0.5U
DNA	500ng

Reference strains (*L. delbrueckii* subsp. *delbruekii* DMS20081 and *S. thermophilus* LMG18311) were used as positive control. Negative control has no template DNA, ddH_2O was used as the control. PCR amplification was performed using the following conditions (Table 2.9).

Table 2.9. PCR reaction conditions for methionine biosynthesis gene

	Cycles	Temperature	Time
Initial denaturation		94°C	2 min
Denaturation	45	94°C	30 sec
Annealing		54°C	40 sec
Extension		72°C	45 sec
Final extension		72°C	10 min

Amplification products were loaded to 1.5 % agarose gels, run on electrophoresis at 80V for 45 min (BioRad), post-stained with EtBr, de-stained, and visualized under UV.

2.8.6. Sequencing of 16S rRNA gene

Sequencing of the 16S rRNA gene was performed using forward primer L1 (Mora et al. 1998) and reverse primer U926R (Baker et al., 2003). Amplification conditions and reaction components were same as displayed in tables 2.7 and 2.8. Amplified fragments were run on 1% agarose gels and electrophoresed at 80V for 45 mins. The gels were post-stained with EtBr, destained and visualized under UV. In the presence of non-specific bands, the specific fragments were extracted from agarose gels using DNA extraction kit (Fermentas). In the absence of non-specific bands, the amplification products were sent to Iontek (Istanbul), where purification from residual reaction components and sequencing was performed.

2.8.7. Extraction of DNA fragments from agarose gels

Extractions were performed according to manufacturer's instructions with minor changes. Briefly; gel slice containing the DNA band was excised with minimum UV exposure. Approximate volume of gel slice was determined by weight (1g = 1ml) and the slice was placed into a eppendorf tube. For 1 volume of gel 3 volumes of binding solution was added. Tube was incubated 5 minutes at 55°C to

dissolve agarose. 5µl of silica powder suspension was added to the tube, and incubated for 5 minutes at 55°C. Infrequent mixing (2–3 times) was done to keep silica powder in suspension. Silica powder/DNA complex was spinned for 3 seconds to form a pellet and supernatant was removed. 500µl of ice cold wash buffer was added, vortexed and spinned for 3 sec. and supernatant was poured off. The procedure was repeated three times. After the supernatant from the last wash had been removed, the tube was spinned again and the remaining liquid was removed with pipette. DNA was eluted into TE buffer and incubated at 55°C for 5 minutes. The tube was spinned and the supernatant was removed into a new tube avoiding the pellet. The elution was repeated with another aliquot of TE to remove small amounts of the silica powder and the tube was spinned again for 30sec in centrifuge. The supernatant was transferred into a new tube. Quantity of extracted DNA was determined on agarose gel as described in section 2.6. Eluted DNA was stored at -20°C.

2.8.8. Amplification conditions and reaction mixture for RAPD-PCR

Typing analysis of *L. delbrueckii* subsp. *delbruekii* and *S. thermophilus* strains were obtained using primers M13 and 1254 (Delfederico et al., 2006). Amplification reactions were performed at the following reaction mixture and amplification conditions (Tables 2.10 and 2.11).

Table 2.10. PCR reaction mixture for RAPD-PCR analysis

Reaction Components	**Final Concentration**
Sterile ddH_2O	-
$MgCl_2$	3.0mM
PCR reaction buffer	1X
dNTP	200µM of each dNTP
Primer	1µM
Taq DNA polymerase	0.5U
DNA	500ng

Table 2.11. Amplification conditions for RAPD analysis for yoghurt starter bacteria

	Cycles	Temperature	Time
Initial denaturation		94°C	2 min
Denaturation	40	94°C	1 min
Annealing		42°C	20 sec
Extension		72°C	2 min
Final extension		72°C	10 min

Amplification products were loaded onto 1.5% agarose gels, electrophoresed at 90V for 80 min, post-stained with ethidium bromide, de-stained and visualized under UV using GelDocXR (BioRAD).

RAPD profiles were analysed by QuantityOne software (BioRad, USA). Similarities between strains were estimated using the Dice coefficient. The dendrograms showing the relationships between starter strains were obtained using the unweighted pair group method with average linkage (UPGMA). An 80% similarity was arbitrarily selected as a threshold for the definition of the homogeneous RAPD-based clusters (Aquilanti et al., 2007, Kenny et al., 2005, Rossetti and Giraffa, 2005).

2.8.9. Amplification conditions and reaction mixture for MLST

Analysis of MLST was performed using *rpoA*, *pheS* and β-*gal* genes.

Analysis of *pheS* gene was obtained using primers pheS-21-F and pheS-21-R, with an expected amplification product of 500bp (Naser et al., 2005).

Analysis of *rpoA* gene was performed using primers rpoA-21-F and rpoA-23-R, with expected amplicon size of 800bp (Naser et al., 2005).

Reaction mixes and amplification conditions of *rpoA* and *pheS* genes were similar, except their annealing temperatures (Tables 2.10 and 2.11). For *rpoA*, T_a was 42°C, while for *pheS* gene, T_a was 46°C.

Analysis of β-galactosidase gene was performed using primers bgalac2f and bgalac2r. The primers were designed using Primer3 and oligocalc programs, with an amplification product of 800bp. Amplification conditions and reaction conditions of beta-galactosidase gene were same as methionine biosynthesis gene, and performed according to Tables 2.5 and 2.6. Primer binding sites and primers' sequence were supplied in Appendix D and Table 2.2, respectively.

Table 2.12. PCR reaction mixture for *rpoA* and *pheS* genes

Reaction Components	**Final Concentration**
Sterile ddH_2O	-
$MgCl_2$	1.5mM
PCR reaction buffer	1X
dNTP	200µM of each dNTP
Forward primer	1µM
Reverse primer	1µM
Taq DNA polymerase	0.5U
DNA	500ng

Table 2.13. Amplification conditions for *rpoA* and *pheS* genes

	Cycles	**Temperature**	**Time**
Initial denaturation		95°C	5 min
Denaturation	3	95°C	1 min
Annealing		42-46°C	2 min 15 sec
Extension		72°C	1 min 15 sec
Denaturation	30	95°C	35 sec
Annealing		42-46°C	1 min 15 sec
Extension		72°C	1 min 15 sec
Final extension		72°C	7 min

Amplified fragments were run on 1% agarose gels and electrophoresed at 80V for 45 mins. The gels were post-stained with EtBr, destained and visulized under UV. In the presence of non-specific bands, the specific fragments were extracted from agarose gels using DNA extraction kit (Fermentas) as explained in section 2.8.5. In the absence of non-specific bands, the amplification products were sent to Iontek (Istanbul), where purification from residual reaction components and sequencing was performed.

2.9. Analysis of experiments using bioinformatics

Analyses of the experimental results require use of several tools, except analysis of methionine biosynthesis for specific identification of yoghurt starter cultures. Sequencing, RAPD, and MLST require different tools to analyze experimental results.

2.9.1. Analysis of 16S rRNA sequencing

Chromatographs of sequencing results (.abI) were visualized and analyzed using MEGA4 (Tamura et al., 2007). MEGA4 offers tools for visualization of chromatograms, alignment of sequences by ClustalW, and produce trees to confer relationships among samples.

2.9.2. Analysis of RAPD

RAPD analysis produce complex patterns and interpretation of these patterns were obtained using QuantityOne (BioRad). QuantityOne can detect bands on the gel photographs with file extensions .sc1, the primary format of gel pictures taken by GelDoc (BioRad) systems. Presence / absence of bands were compared, matching bands were detected and, dendrograms were produced. Detection of bands and matching were done automatically, however, manual detection can also be performed. In this study, cluster analysis was performed by using the Dice similarity coefficient and the unweighted pair-group method with arithmetic

averages (UPGMA). An 80% similarity was arbitrarily selected as a threshold for the definition of RAPD-based clusters.

2.9.3. Analysis of MLST

Chromatograms of sequencing results were visualized using MEGA4 (Tamura et al., 2007) and START2 (Jolley et al., 2001) softwares. The genes *rpoA*, *pheS* and β-*gal* were aligned using ClustalW. Nonuniform sequences at the beginning and at the end of the genome were removed. For the analysis of MLST, each gene was designated as a locus, and for each gene, nucleotide differences were assigned as different alleles (even single nucleotide differences). The alleles at the three loci were used to construct an allelic profile (in the format of 1-2-4), to define the sequence type ST (as an example, ST-2) for each isolate. Phylogenetic trees were compiled with START2 software, and dendrograms were constructed by the unweighted pair group method with arithmetic averages (UPGMA) by using profiles option.

CHAPTER 3

RESULTS AND DISCUSSION

3.1. Experimental strategy

In this study, our aim was to identify and differentiate yoghurt starter bacteria, *L. delbrueckii* subsp. *bulgaricus,* and *S. thermophilus,* from bacteria inhabiting similar milk environments, by using PCR based methods. Yoghurt is a favourite milk product in Turkey, and traditional methods of yogurt making are preferred by many people. Traditional home-made yoghurts are possible reservoirs of strains that are comparable or even better than the current commercial starters. Although there are methods for identification of these bacteria, these methods either require pre-analysis of cultures by biochemical and microbiological methods, and/or may result in misidentification at species and subspecies level. In our study, identification of yoghurt starter bacteria was studied first by using species specific primers, and then by ARDRA. Results of the experiments revealed the need for a quick and reliable method for selective identification of yoghurt starter bacteria. A new PCR based method was developed and successful results were obtained for joint identification of *L. delbrueckii* subsp. *bulgaricus* and *S. thermophilus* strains from traditional home-made yoghurts without preliminary identification by phenotypic methods. In addition to yoghurt starters, cheese starter bacteria; *L. delbrueckii* subsp. *lactis* and *L. helveticus* could also be identified by the proposed method. The reliability of the developed method had been confirmed by sequencing of 16S rRNA genes of isolates. After successful identification of

yoghurt starter cultures, strain specific fingerprints were obtained using RAPD and MLST. The experimental strategy was explained in a flowchart in Figure 3.1.

3.2. Identification of *L. delbrueckii* and *S. thermophilus* by species specific PCR

Identification of LAB can be obtained by PCR based methods. The methods used in the literature include species specific PCR, ARDRA and PCR-RFLP. PCR-RFLP was performed by amplification of housekeeping genes followed by restriction digestion (Giraffa et al., 2003). ARDRA is a modification of PCR-RFLP, in ARDRA ribosomal DNA is used instead of housekeeping genes.

3.2.1. Identification of *L. delbrueckii* by species specific PCR

Identification of lactobacilli is generally focused on cheese starter bacteria, mainly *Lactococcus lactis*, *L. helveticus* and *S. thermophilus*. Identification of *L. delbrueckii* at subspecies level is a difficult task, since *L. delbrueckii* and its two other subspecies (*delbrueckii* and *lactis*) share over 80% DNA similarity, and can be found in fermented milk products (Giraffa et al., 2003), and genotypic methods for identification of these bacteria are rare. Tilsala-Timisjarvi and Alatossava (1997) developed species specific primers for identification of LAB, namely *L. paracasei, L. rhamnosus, L. delbrueckii, L. acidophilus, L. helveticus* and *S. thermophilus.* The *L. delbrueckii* and *S. thermophilus* specific primers were suitable for our purpose, and used in our study.

In our study, *L. delbrueckii* specific primer DelI/II (Tilsala-Timisjarvi A. and Alatossava T., 1997) was used to test specificity of primer for several *L. delbrueckii* strains. 15 presumptive and 2 reference *L. delbrueckii*, 1 reference *L. helveticus*, 4 presumptive and 2 reference *S. thermophilus* species were used to test primer specificity. Figure 3.2 represents results of the experiment.

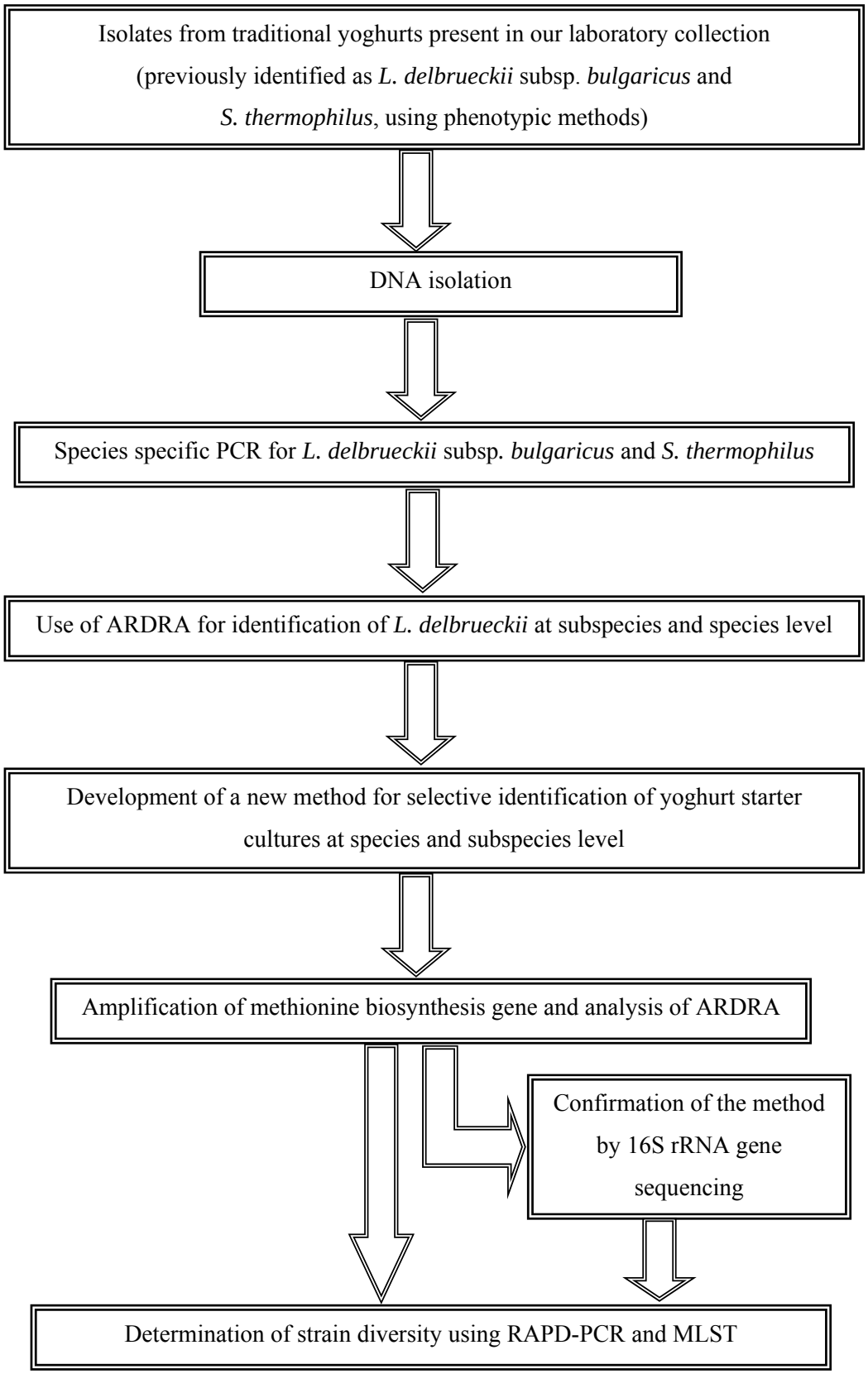


Figure 3.1. Flow chart of the experimental strategy

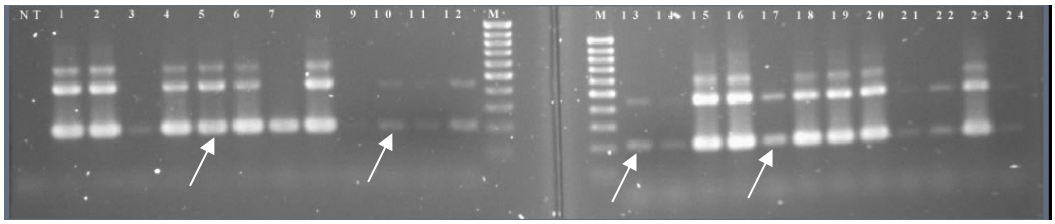


Figure 3.2. *L. delbrueckii* species specific PCR. M: 100bp ladder (Fermentas), lanes 1-8 presumptive *L. delbrueckii* (1: K1-10, 2: K1-16, 3: K1-23, 4: K1-29, 5: K1-33, 6: K1-38, 7: K1- 43, 8: K1-44), lanes 15-16, 18-23 *L. delbrueckii* reference and commercial strains (15: DSM20081, 16: NRRL B-4525, 18: B1000-1, 19: B1000-2, 20: B1000-3, 21: LB340-2, 22: LB340-3, 23: Yo-mix 410-1), 17: *L. helveticus* NRRL B-4526, lanes 9-14 and 24 *S. thermophilus* reference and commercial strains (9: CNRZ1066, 10: CECT986, 11: LMG18311, 12: B1000-3, 13: Ta040-1, 14: Ta040-2, 24: Yo-mix 410-3), NT: Negative control (no DNA template).

In the figure, *L. delbrueckii* specific amplicon was indicated by arrows. The specific amplification product was about 200bp, and produced by all *L. delbrueckii* species. Presumptive *L. delbrueckii* strains LB340-2 and LB340-3 produced faint bands, and their identity remained unclear until sequencing of 16S rRNA gene. Partial sequencing of the 16S rRNA gene revealed that LB340-2 and LB340-3 actually belong to species *L. acidophilus*. The specific 200 bp amplicon was produced by reference *L. helveticus,* though unexpected (lane 16). PCR was optimized to eliminate *L. delbrueckii* specific band in *L. helveticus* as suggested by authors (Tilsala-Timisjarvi and Alatossava, 1997), however, the amplicon was still produced. The *L. delbrueckii* specific bands were observed even in several strains of *S. thermophilus*, although fainter bands were observed, though in lanes 12 and 13, strains B1000-3 and Ta040-1 produced a clear band. The negative control did not produce the amplicon, however, the experiment was repeated four times to check the possibility of contamination, and the problem was persistent.

According to the result of the experiments, *L. delbrueckii* species specific primer was not suitable for species specific identification. In addition, the method does

not discriminate *L. delbrueckii* at subspecies level, as stated by Tilsala-Timisjarvi and Alatossava, (1997).

3.2.2. Identification of *S. thermophilus* by species specific PCR

S. thermophilus specific primer ThI/II (Tilsala-Timisjarvi A. and Alatossava T., 1997) was used to identify *S. thermophilus* strains. Figure 3.3 displays the results of species specific PCR experiment.

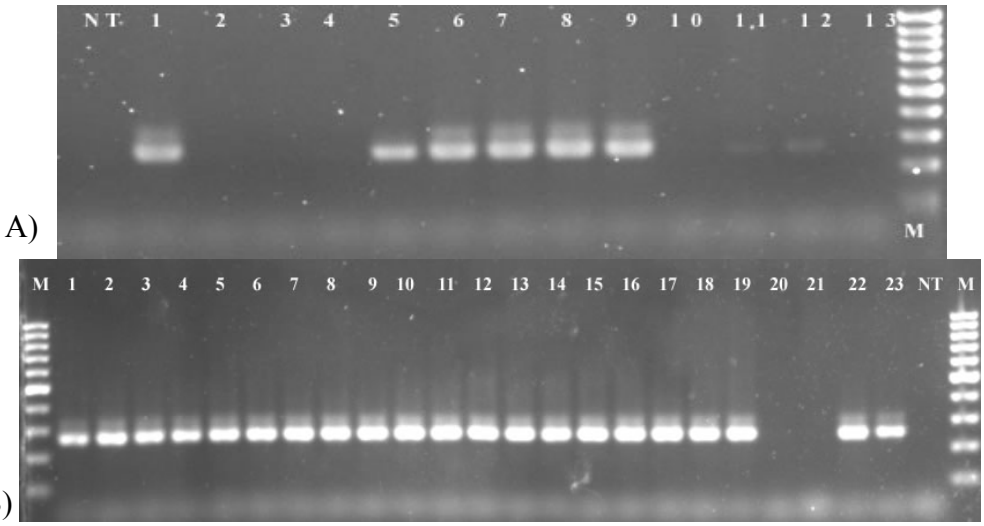


Figure 3.3. *S. thermophilus* species specific PCR. A) for refererence and commercial strains M: 100bp ladder (Fermentas), lanes 2-4, 10, 11,13 reference and commercial *L. delbrueckii* (2: B1000-1, 3: B1000-2, 4: 1000-3, 10: DSM20081, 11: NRRL B-4525, 13: Yo-mix 410-1), lane 12: L. helveticus NRRL B-4526, lanes 1 and 5-9 reference and commercial *S. thermophilus* strains (5: CECT986, 6: LMG18311, 7: Ta040-1, 8: Ta040-2, 9: B1000-3), NT: Negative control (no DNA template), B) for strains from traditional yoghurt. M: 100bp ladder, lane 1: LMG18311, 2: Ta040-1, 3: B1000-3, 4: K1-16, 5: K1-7, 6: N6-1, 7: N6-3, 8: N2-2, 9: K1-14, 10: N2-1, 11: K1-12, 12: N2-3, 13: N5-7, 14: K1-18, 15: K2-4, 16: K1-27, 17: K1-30, 18: N8-2, 19: K1-28, 20: N7-4, 21: N7-1, 22: N3-1, 23: N5-4, NT: Negative control (no template DNA), M: 100bp ladder.

S. thermophilus specific PCR produced 260bp amplicons, which were specific to *S. thermophilus* in discrimination from lactobacilli (Figure 3.3a). The specific amplification products were observable in all reference and commercial *S. thermophilus* strains (Figure 3.3b, lanes 1-3). For traditional yoghurt isolates, 18 strains produced *S. thermophilus* species specific amplicons, but two strains failed to produce the amplicon. These two strains were N7-1 and N7-4, and sequencing of 16S rRNA gene identified these strains as *S. bovis*. While strain N5-4 was identified also as *S. bovis, S. thermophilus* specific PCR resulted in amplification in this strain. Since the method produced ambigious results for yogurt isolates, it was not used for identification of *S. thermophilus* species.

3.3. Identification of *L. delbrueckii* at species and subspecies level by ARDRA

Specific PCR for identification of *L. delbrueckii* at species level was not successful, and a reliable method was investigated for this purpose. Identification of *L. delbrueckii* could be performed by ARDRA as stated by researchers (Giraffa et al., 1998, Delley and Germond, 2002, Miteva et al., 2001). In ARDRA, ribosomal DNA was amplified and digested by restriction enzymes. The method was used for *L. delbrueckii* strains which were either obtained from culture collections (Roy et al., 2001, Delley and Germond, 2002, Miteva et al., 2001), or isolated from dairy products (Collado and Hernandez, 2007). In ARDRA, method is performed to differentiate closely related species, thus, previous identification by phenotypic methods were necessary. Previous identification is useful to reduce number of isolates, in this way, screening for the target band pattern will be performed on a smaller number of isolates, and use of restriction enzymes will be limited.

L. delbrueckii has been identified at genus level as lactobacilli, among probiotic products containing bifidobacteria and streptococci, using *Mwo*I (Collado and Hernandez, 2007), at species level by *Eco*RI (Delley and Germond, 2002, Miteva et al., 2001, Giraffa et al., 1998) by *Cfo*I (Roy et al., 2001), and by *Taq*I (Yavuz et al., 2004).

Differentiation of *L. delbrueckii* subsp. *lactis* from subspecies *bulgaricus* represented contradictory results in the article of Roy et al., (2001). In the article, one *L. delbrueckii* subsp. *bulgaricus* strain (ATCC21815) exhibited a band pattern similar to subspecies *lactis*. In addition, banding patterns of subspecies *bulgaricus* and subspecies *delbrueckii* was same. The contradictions in banding patterns *Cfo*I prevented reliable use of this enzyme.

Finally, considering the literature survey, *Taq*I, *Eco*RI and *Hpa*II was chosen for identification of *L. delbrueckii.*

3.3.1. ARDRA analysis of *L. delbrueckii* by *Taq*I

Species identification of *L. delbrueckii* by *Taq*I was studied by Yavuz et al., (2004). The 16S rRNA gene and ITS (intergenic spacer) regions were covered in the study in order to have increased discriminatory power. It was reported that species specific profiles were obtained for seven species of lactobacilli. Two subspecies of *L. debrueckii*; subsp. *bulgaricus* NRRL-B548 and subsp. *delbrueckii* NRRL-B443 produced distinct band patterns, and discriminated *L. delbrueckii* from six *Lactobacillus* species.

In our study, the primer pair used in Yavuz et al., (2004) was utilized. The amplification product was 1700bp, covering 16S rRNA gene-ITS region. The *L. debrueckii* subsp. *bulgaricus* NRRL-B548 was kindly obtained from the research group of Yavuz et al., (2004) and included into our *L. delbrueckii* subsp. *bulgaricus* collection.

Our results produced the expected banding pattern for *L. delbrueckii* subsp. *bulgaricus* NRRL-B548 (Figure 3.4). However, this pattern was different than the three type strains of *L. delbrueckii* from our culture collection. The produced pattern was similar to the pattern produced by *L. helveticus* (Figure 3.4a, lane 10).

In addition, our study demonstrated that ARDRA with *Taq*I could produce species specific pattern for *L. delbrueckii* species, and a similar pattern had been produced by *L. gasseri*. In addition, *Taq*I was able to differentiate *L. casei, L. reuteri, L. helveticus* and *L. rhamnosus* (Figure 3.4a). The method could not distinguish between *L acidophilus* and *L. plantarum*. Figure 3.4b displays a clear picture of *Taq*I digestion of 16S rRNA gene and ITS region in *L. delbrueckii* species. Overall, ARDRA with *Taq*I could be used to produce species-specific patterns for *L. delbrueckii*, while it could not discminate *L. delbrueckii* at subspecies level.

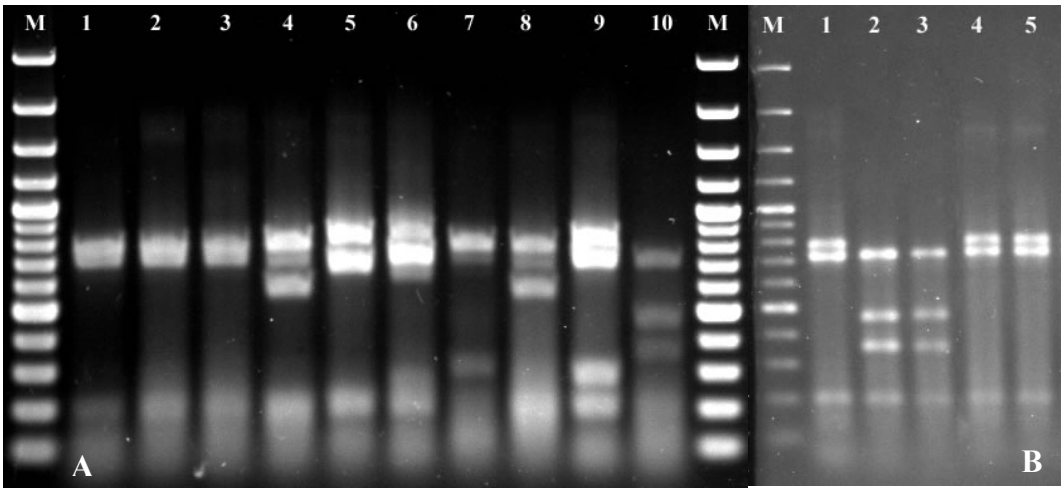


Figure 3.4. Identification of *L. delbrueckii* species by ARDRA with *Taq*I. A) M: Marker 100bp ladder plus, lanes 1-3 *L. delbrueckii* (1: subsp. *bulgaricus* DSM20081, 2: subsp. *lactis* NRRL-B4525, 3: subsp. *delbrueckii* NRRL- B763), 4: *L. acidophilus* ATCC4356, 5: *L. casei* NRRL-B1922, 6: *L. reuteri* NRRL-B14171, 7: *L. gasseri* NRRL-B4240, 8: *L. plantarum* DSM20246, 9: *L. rhamnosus* NRRL-B442, 10: *L. helveticus* NRRL-B4526. B) M: Marker 100bp ladder plus, 1: *L. delbrueckii* subsp. *bulgaricus* DSM20081, 2: *L. delbrueckii* subsp. *bulgaricus* NRRL-B548*, 3: LB340-3**, 4: B1000-1, 5: B1000-2.
*Identified as *L. helveticus* according to ARDRA by *Eco*RI.
**Identified as *L. acidophilus* according to partial sequencing of 16S rRNA gene.

3.3.2. ARDRA analysis of *L. delbrueckii* by *Eco*RI

Although identification of *L. delbrueckii* at subspecies level by *Eco*RI was reported by some researchers (Miteva et al., 2001), others state that it can not differentiate at subspecies level, since this enzyme cannot differentiate between all three subspecies *delbrueckii, lactis,* and *bulgaricus* (Giraffa et al., 1998, Delley and Germond, 2002).

In our study, *Eco*RI enabled differentiation of *L. delbrueckii* subsp. *bulgaricus* from subsp. *lactis* and subspecies *delbrueckii,* as stated by earlier researches (Giraffa et al., 1998, Delley and Germond, 2002, Miteva et al., 2001). In addition, *L. helveticus* produced a distinct digestion pattern. However, *Eco*RI could not discriminated subspecies *lactis* and subspecies *delbrueckii* from each other (Figure 3.5). This discrimination would be of great importance, since subspecies *lactis* is an important dairy starter for dairy industry, while subspecies *delbrueckii* is not. The differentiation between these two subspecies had later been achieved by a new method using methionine biosynthesis genes.

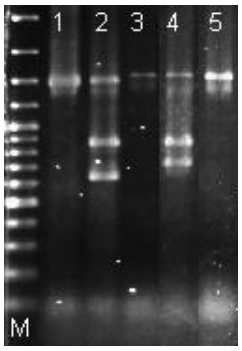


Figure 3.5. Identification of *L. delbrueckii* species by ARDRA with *Eco*RI. M: Marker 100bp ladder plus, lane 1: *L. delbrueckii* subsp. *lactis* NRRL-B4525, 2: *L. delbrueckii* subsp. *bulgaricus* DSM20081, 3: *L. acidophilus* ATCC4356, 4: *L. helveticus* NRRL-B4526, 5: *L. delbrueckii* subsp. *delbrueckii* NRRL- B763.

3.3.3. ARDRA analysis of *L. delbrueckii* by HpaII

Identification of *L. delbrueckii* by ARDRA was then studied with *Hpa*II. *Hpa*II enzyme was selected after performing online preliminary analysis of 16S rRNA gene sequences of *L. acidophilus* group species. The 16S rRNA gene and ITS regions were covered in the study. The selected strains belong to *L. delbrueckii* subsp. *bulgaricus*, *L. delbrueckii* subsp. *lactis, L. delbrueckii* subsp. *delbrueckii, L. acidophilus, L.casei, L. reuteri, L. gasseri, L. rhamnosus, L. plantarum, and* presumed *L. delbrueckii* subsp. *bulgaricus* LB340-3. The banding patterns obtained with this enzyme were represented in figure 3.6.

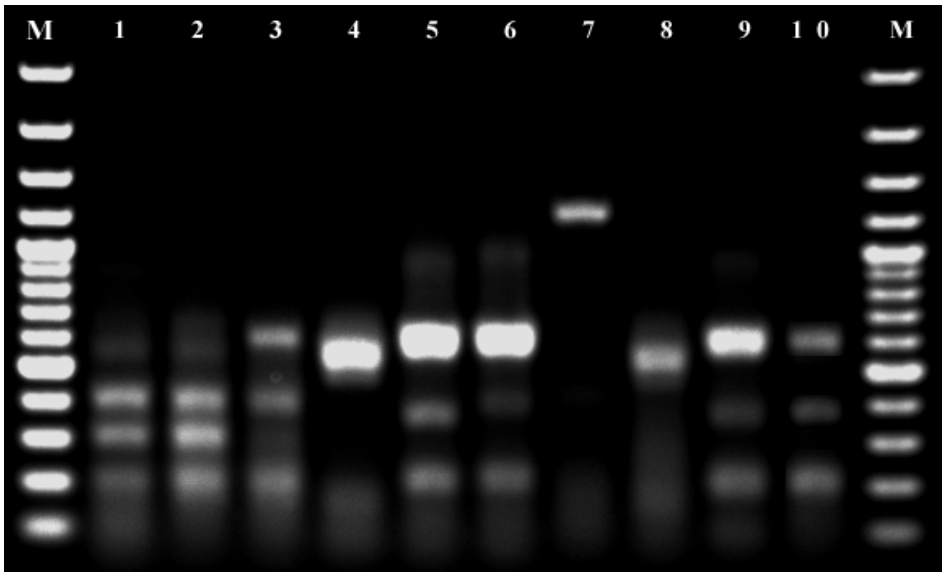


Figure 3.6. Identification of *L. delbrueckii* species by ARDRA with *Hpa*II. M: Marker 100bp ladder plus, lanes 1-3 *L. delbrueckii* (1: subsp. *bulgaricus* DSM20081, 2: subsp. *lactis* NRRL-B4525, 3: subsp. *delbrueckii* NRRL- B763), 4: *L. acidophilus* ATCC4356, 5: *L. casei* NRRL-B1922, 6: *L. reuteri* NRRL-B14171, 7: *L. gasseri* NRRL-B4240, 8: *L. plantarum* DSM20246, 9: *L. rhamnosus* NRRL-B442, 10: *L. helveticus* NRRL-B4526.

ARDRA with *Hpa*II for identification of *L. delbrueckii* species revealed identical patterns with subspecies *lactis* and subspecies *bulgaricus* (lanes 1 and 2). Subspecies *delbrueckii* pattern was similar to *L. reuteri* and, species *L. casei*, *L. rhamnosus* and *L. helveticus* produced the same pattern. Overall, digestion of 16S rRNA and ITS region with *Hpa*II enabled discrimination of *L. delbrueckii* subsp. *lactis* and subspecies *bulgaricus* from the species tested.

On the whole, ARDRA analysis with enzyme *Taq*I produced species specific pattern for *L. delbrueckii*. ARDRA analysis with enzyme *Eco*RI produced specific pattern for *L. delbrueckii* subsp. *bulgaricus*, however, it could not differentiated between the two other subspecies (*lactis* and *delbrueckii*). Finally, ARDRA analysis with enzyme *Hpa*II produced a specific pattern for *L. delbrueckii* subsp. *bulgaricus* and subspecies *lactis*. Thus, of the restriction enzymes tested, no single enzyme could produce a pattern for discrimination of the three subspecies.

Consequently, ARDRA with *Eco*RI could be used to differentiate important starters, namely, *L. delbrueckii* subsp. *bulgaricus* and *L. delbrueckii* subsp. *lactis*. However, it should be noted that preliminary analysis should be performed before ARDRA to reduce diversity and number of bacteria isolated from milk products. This step was necessary because ARDRA had been reliable for identifying and discriminating closely related bacteria, but for identification of diverse species, use of high numbers of restriction enzymes was necessary. An example was represented by Moreira et al., (2005); the researchers used 11 different restriction enzymes to identify 20 different lactobacilli at species level. In that study, it was also demonstrated that in order to differentiate *L. delbrueckii* from 19 other lactobacilli, 3 different enzymes were required.

Comparing the experimental results, it was concluded that an easy, reliable and reproducible method would be of great use to identify and differentiate yoghurt starter bacteria at species and subspecies level.

3.4. Development of a new method for identification of yoghurt starter bacteria

Identification of lactobacilli at genus, species, subspecies and strain level by genotypic methods has been achieved in literature (Giraffa et al. 2003, Dubernet et al. 2002, Rosetti and Giraffa 2005). However, these methods require previous identification by microbiological methods. The need to use microbiological identification tests prior to genetic identification arises from the fact that, milk and milk products contain diverse number of bacterial species, which are difficult to differentiate. Furthermore, although there exists species specific primers for these bacteria, it has resulted in contradictory results in our laboratory. To correctly identify yoghurt starter bacteria using genetical methods -which were identified formerly by phenotypic methods-, a new method was developed.

The developed method was PCR based; at first, amplification of methionine biosynthesis gene was performed to identify yoghurt starter bacteria, and second, ARDRA (Amplified Ribosomal DNA Restriction Analysis) analysis was performed for subspecies level identification.

3.4.1. Amplification of methionine biosynthesis gene

In the article published in Nature Biotechnology (2004), Bolotin et al., presented the complete sequence of two *S. thermophilus* strains: CNRZ 1066 and LMG 18311. The researchers also investigated comparative genomics of the two strains, in following respects; gene decay, pseudogenes, genes involved in metabolism, and virulence related genes. It was interesting that genus streptococci comprises important pathogens, while species *S. thermophilus* having GRAS (Generally Recognized as Safe) status. In the article, it was concluded that together with gene decay and pseudogenes, horizontal gene transfer is very important in the shaping of *S. thermophilus* genome and several genes that have been transferred laterally were listed. Among the horizontally transferred genes, methionine biosynthesis genes were important to our study. Presence of methionine biosynthesis genes is

important in survival of *S. thermophilus* and *L. delbrueckii* subsp. *bulgaricus* in milk, since methionine concentration in milk is low. Investigation of methionine biosythesis pathways were performed by using KEGG (Kyoto Encyclopedia of Genes and Genomes). Comparative analysis of the pathway in KEGG revealed that the synthesis pathways are more complete in *S. thermophilus* than *L. delbrueckii* subsp. *bulgaricus*.

The methionine biosynthesis gene cluster was found to be 95% identical to *L. delbrueckii* subsp. *bulgaricus*. Sequences of genes involved in methionine biosynthesis were retrieved from IMG (Integrated Microbial Genomes) database, and orthologous genes were searched on the genome to reveal percent similarity. The *met*B gene was only 73.51% identical to *Lactobacillus plantarum* WCFS1 and 60.42% identical to *Lactobacillus acidophilus* NCFM. Considering the high similarity between *S. thermophilus* and *L. delbrueckii* subsp. *bulgaricus* it was assumed that these genes may be used for specific identification of *L. delbrueckii* subsp. *bulgaricus* and *S. thermophilus*.

The methionine biosynthesis gene sequences were used to construct primers using Primer3. 10 pairs of primers were designed, and primer pair cysmet2 was better in specific identification of *L. delbrueckii* subsp. *bulgaricus* and *S. thermophilus*. To ensure that the primer was specific, blast analysis was performed, and analysis resulted in homology only with *S. thermophilus* and *L. delbrueckii* subsp. *bulgaricus*.

Specificity of primer was tested against 19 reference strains, mostly type strains from genus lactobacilli, streptococci and lactococcus. No amplification was observed in reference strains (Figure 3.7). Although not included in the figure, *L. acidophilus* ATCC 4356 did not produce the band, either.

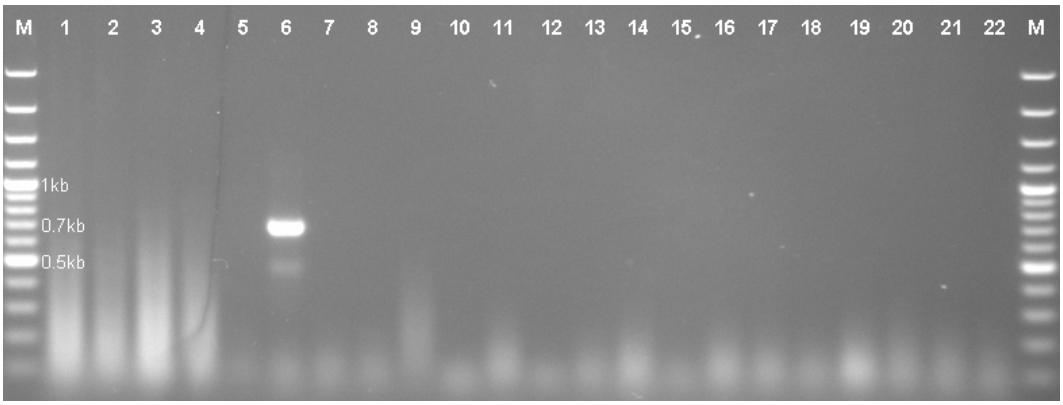


Figure 3.7. Amplification of the methionine biosynthesis gene in reference strains. M: 100bp ladder, *Lc. lactis* (1: 2910, 2: 2911, 3: 2088, 4: 3113), 5: Negative control (no DNA control), 6: *L. delbrueckii* subsp. *bulgaricus* DSM20081, 7: *L. reuteri* NRRLB-14171, 8: *L. pentosus* NRRLB-227, 9: *L. brevis* NRRLB-4527, 10: *L. gasseri* NRRLB-4240, 11: *L. amylovorous* NRRLB-4540, 12: *L. casei* subsp. *casei* NRRLB-1922, 13: *L. johnsonii* NRRLB-2178, 14: *L. salivarius* subsp. *salivarius* NRRLB-1949, 15: *L. paraplantarum* NRRLB-23115, 16: *L. fermentum* NRRLB-1840, 17: *L. rhamnosus* NRRLB-442, 18: *L. paracasei* subsp. *paracasei* NRRLB-4560, 19: *S. dysgalactiae* NRRLB 688, 20: *S. equinus* NRRLB-3573, 21: *S. infantarius* NRRLB-41208, 22: *L. delbrueckii* subsp. *delbruekii* NRRLB-763, M: 100bp ladder.

After specificity of the primer was achieved, amplification of the methionine biosynthesis gene was performed in yoghurt starters, *L. delbrueckii* subsp. *bulgaricus* and *S. thermophilus.* The expected fragment of about 750bp was amplified in all *L. delbrueckii* subsp. *bulgaricus* and *S. thermophilus* strains (Figure 3.8). Methionine biosynthesis gene amplification product was amplified only in *S. thermophilus*, but not in *S. equinus, S. galactiae, S infantarius,* and *Lactococcus lactis* strains (Table 3.1). Thus, selective amplification of *S. thermophilus* species from other streptococci and lactococci was achieved successfully. In addition, methionine biosynthesis gene sequences of bacteria commonly present in milk products; *Enterococcus faecium, E. faecalis, Oenococcus oeni,* and *Leuconostoc mesenteroides* were obtained from http://img.jgi.doe.gov. Multiple sequence alignments of these species displayed

heterogeneous sequences (Appendix D), and primer binding sites differed in 3-9bp for each primer. It should also be noted that *L. plantarum* gene sequence was the closest relative to gene sequences of yoghurt starter bacteria, and its gene was not amplified by the primer, thus, annealing of the primer and amplification of the gene product in the *Enterococcus faecium, E. faecalis, Oenococcus oeni,* and *Leuconostoc mesenteroides* species were not expected.

On the other hand, although blast analysis of the primer confirmed specificity in only *L. delbrueckii* subsp. *bulgaricus* and *S. thermophilus*, however, the methionine biosynthesis gene product was also amplified in *L. delbrueckii* subsp. *lactis* and *L. helveticus.* To check this result, sequencing of the methionine biosynthesis gene from *L. helveticus*, *L. delbrueckii* subsp. *bulgaricus* and *S. thermophilus* was performed, and the result confirmed that *L. helveticus* had the identical gene (Appendix D).

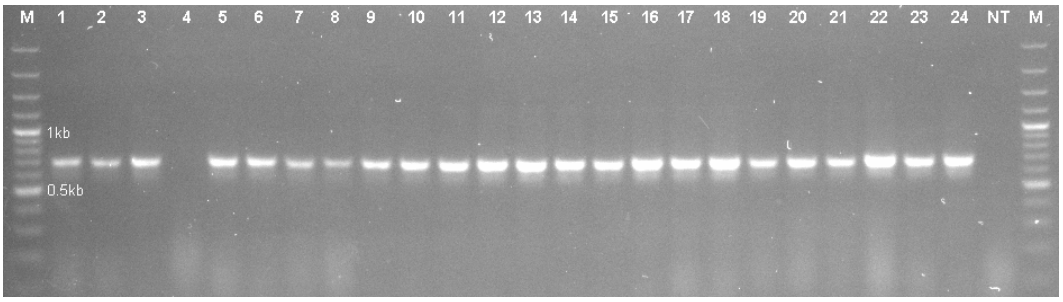


Figure 3.8. Amplification of the methionine biosynthesis gene in *L. delbrueckii* subsp. *bulgaricus* and *S. thermophilus* strains. M: 100bp ladder plus (Fermentas), lanes 1-2, 5-18: *L. delbrueckii* subsp. *bulgaricus* (1: DSM20081, 2:ATCC BAA-365, 5: B1000-1, 6: B1000-2, 7: B1000-3, 8: Yo-mix 410-1, 9: M2-5, 10: M2-14, 11: M2-18, 12: K1-10, 13: K1-16, 14: K1-29, 15: K1-33, 16: K1-38, 17: K1-43, 18: K1-44), 3: *L. delbrueckii* subsp. *lactis* NRRL-B4525, 4: *L. delbrueckii* subsp. *delbrueckii* NRRL-B763, 19: *L. helveticus* NRRL B-4526, lanes 20-24: *S. thermophilus* (20: LMG18311, 21: CNRZ1066, 22: B1000-3, 23: Yo-mix 410-3, 24: Ta 040-1), NT: Negative control (no DNA control), M: 100bp ladder.

From figure 3.8, it was seen that methionine biosynthesis gene was absent in *L. delbrueckii* subsp. *delbrueckii*. Although the gene was amplified in two other subspecies; *L. delbrueckii* subsp. *lactis* and *L. delbrueckii* subsp. *bulgaricus*, the absence in *L. delbrueckii* subsp. *delbrueckii,* could be explained by the fact this subspecies adapted itself to plant environment, rather than milk. This result was important because while present methods in the literature were unable to discriminate between subspecies *lactis* and subspecies *delbrueckii*, in our study, it was possible to discriminate between those two subspecies.

The amplification of gene fragment in *L. delbrueckii* subsp. *lactis* and *L. helveticus* was unexpected, however, the two species are important starters for cheese, together with *S. thermophilus*, and during history in cheese fermentation, and horizontal gene transfer might have been occurred between these species.

The aim of our study was joint identification and differentiation of yoghurt starters. Differentiation of *S. thermophilus* from lactobacilli can be achieved by simple microscopic examination. *S. thermophilus* appear as cocci in chains while *L. delbrueckii* subsp. *bulgaricus* are rod shaped (Figure 1.12). However, in order to discriminate *L. delbrueckii* subsp. *bulgaricus* from *L. delbrueckii* subsp. *lactis* and *L. helveticus*, another method was necessitated. The restriction analysis of 16S rRNA gene with *Eco*RI, which has been reported by Giraffa et al., (1998), Delley and Germond, (2002), and Miteva et al., (2001), appeared as a promising method.

3.4.2. Analysis by ARDRA

ARDRA (Amplified Ribosomal RNA Restriction Analysis) has been used as a powerful method in differentiating *L. delbrueckii* subsp. *bulgaricus*, *L. delbrueckii* subsp. *lactis* and *L. helveticus* (Delley and Germond, 2002, Giraffa et al., 1998). In our study, the method was used to discriminate the isolates that amplified methionine biosynthesis gene.

Analysis of ARDRA was performed by amplification of 16S rRNA gene, and subsequent analysis using *Eco*RI. Figure 3.9 represents 16S rRNA gene amplifications of some starter strains, where a common product of approximately 1500bp was obtained.

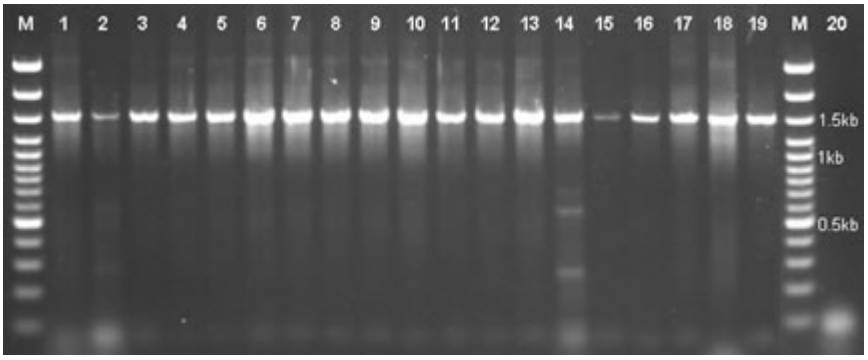


Figure 3.9. Amplification of 16S rRNA gene from isolates that amplified methionine biosynthesis gene product. M: 100bp DNA ladder, lanes 1-2, 4-16, 20: *L. delbrueckii* subsp. *bulgaricus* (1: DSM20081, 2: ATCCBAA-365, 4: M2-14, 5: M2-18, 6: K1-10, 7: K1-16, 8: K1-29, 9: K1-33, 10: K1-38, 11: K1-43, 12: K1-44, 13: M2-5, 14: B1000-1, 15: B1000-2, 16: B1000-3, 21: Yo-mix 410-1), 3: *L. delbrueckii* subsp. *lactis* NRRL-B4525, 17: *L. helveticus* NRRL B-4526, M: 100bp DNA ladder, 20: Negative control (no DNA control).

After successful amplification of 16S rRNA gene, amplicons of *L. delbrueckii* subsp. *bulgaricus*, *L. delbrueckii* subsp. *lactis* and *L. helveticus,* were digested by *Eco*RI. It was shown by Delley and Germond (2002), that restriction enzyme *Eco*RI cut *L. delbrueckii* subsp. *bulgaricus* at two sites, *L. delbrueckii* subsp. *lactis* and subsp. *delbrueckii* and *L. helveticus* at only one site (Figure 3.5).

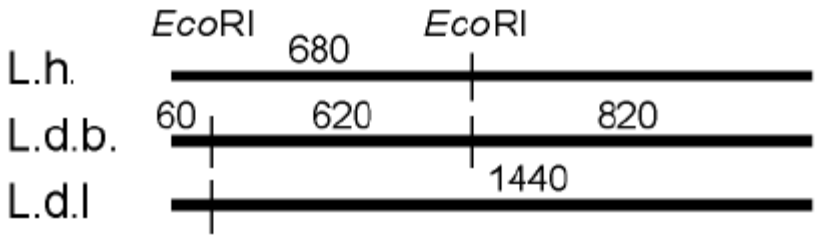


Figure 3.10. Restriction sites of *L. helveticus* (L.h.), *L. delbrueckii* subsp. *bulgaricus* (L.d.b.), and *L. delbrueckii* subsp. *lactis* and *L. delbrueckii* subsp. *delbrueckii* (L.d.l.) by *Eco*RI.

Figure 3.11 displays results of ARDRA analysis for *L. delbrueckii* subsp. *bulgaricus* isolated from yoghurt. The restriction digestion patterns of *L. delbrueckii* subsp. *bulgaricus* and *L. delbrueckii* subsp. *lactis* strains used in our experiment confirm findings of Delley and Germond (2002). It is clear that restriction patterns produced by *Eco*RI were different in *L. delbrueckii* subsp. *bulgaricus* and *L. delbrueckii* subsp. *lactis,* and *L. helveticus.* However, differing to their findings, it was observed that 16S rRNA gene of *L. helveticus* was only partially digested, since the undigested 1500bp band still exists together with faint restricted bands. The result has been confirmed several times on the same species, and also on another *L. helveticus* starter strain from Cristian and Hansen (data not shown). Thus, it could be concluded that restriction of 16S rRNA gene of *L. helveticus* by *Eco*RI might not always result in complete digestion of the gene. Findings of Miteva et al., (2001) support similar unexpected results in restriction analysis of 16S rRNA gene. The reason for this situation was explained by the presence of several 16S rRNA gene copies in a single genome. If some gene copies of 16S rRNA gene lack *Eco*RI recognition site, then these sites cannot be cut by the restriction enzyme, and the amplified gene would remain undigested. The supporting evidence was presented by Giraffa et al., (2000), that *L. helveticus* strains undigested by *Eco*RI had a single nucleotide substitution (C instead of T), located at the *Eco*RI site of the 16S rRNA. In order to clarify the problem,

ClustalW alignment of *L. helveticus* strains was performed. All *L. helveticus* strains having nucleotide sequences longer than 1200bp in Ribosomal Database Project II (RDPII) were used for this purpose. As a result, it was clear that, a nucleotide substitution from T to C prevents *Eco*RI from restriction (Table 3.1). Although a difference was presented on the digestion patterns of *L. helveticus*, it could still be differentiated from *L. delbrueckii* subsp. *bulgaricus* and *L. delbrueckii* subsp. *lactis.* No differences were encountered in digestion of *L. delbrueckii* subsp. *bulgaricus,* and consequently the restriction analysis of 16S rRNA gene was suitable for identification of *L. delbrueckii* subsp. *bulgaricus*.

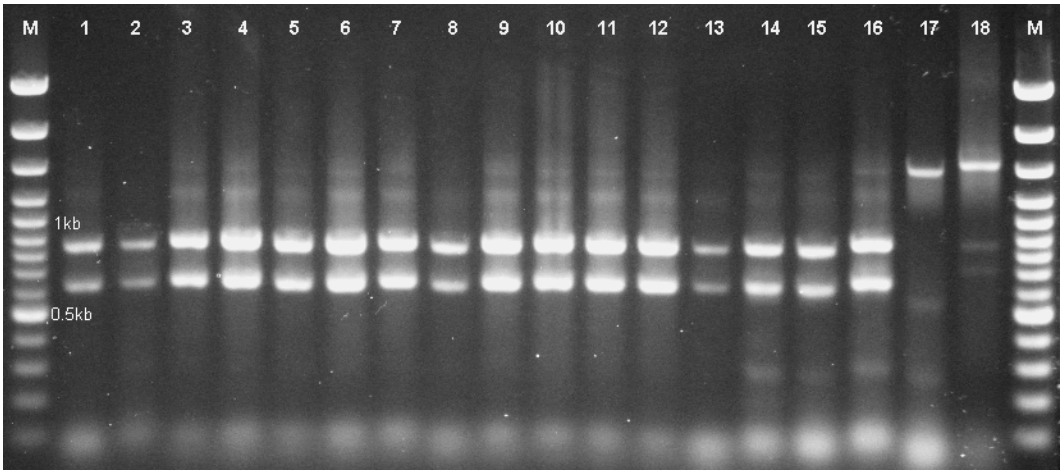


Figure 3.11. Restriction analysis of 16S rRNA gene with *Eco*RI. M: 100bp DNA ladder, lanes 1-16: *L. delbrueckii* subsp. *bulgaricus* (1: DSM20081, 2: ATCC BAA-365, 3: K1-10, 4: K1-16, 5: K1-29, 6: K1-33, 7: K1-38, 8: K1-43, 9: K1-44, 10: M2-5, 11: M2-14, 12: M2-18, 13: Yo-mix 410-1, 14: B1000-1, 15: B1000-2, 16: B1000-3), 17: *L. delbrueckii* subsp. *lactis* NRRL-B4525, 18: *L. helveticus* NRRL B-4526, M: 100bp DNA ladder

Table 3.1. Multiple sequence alignment of *L. helveticus* 16S rRNA gene sequences

S000859943_helveticus	TTTCTTGAGTGCAGAAGAGGAGAGTGGAACTCCATGTGTAGCGGTGGAAT	600
S000422101_helveticus	TTTCTTGAGTGCAGAAGAGGAGAGTGGAACTCCATGTGTAGCGGTGGAAT	600
S000633952_helveticus	TTTCTTGAGTGCAGAAGAGGAGAGTGGAACTCCATGTGTAGCGGTGGAAT	600
S000422102_helveticus	TTTCTTGAGTGCAGAAGAGGAGAGTGGAACTCCATGTGTAGCGGTGGAAT	600
S000422099_helveticus	TTTCTTGAGTGCAGAAGAGGAGAGTGGAACTCCATGTGTAGCGGTGGAAT	600
S000333719_helveticus	TTTCTTGAGTGCAGAAGAGGAGAGTGGAACTCCATGTGTAGCGGTGGAAT	600
S000473948_helveticus	TTTCTTGAGTGCAGAAGAGGAGAGTGGAACTCCATGTGTAGCGGTGGAAT	600
S000942615_helveticus	TTTCTTGAGTGCAGAAGAGGAGAGTGGAACTCCATGTGTAGCGGTGGAAT	600
S000966634_helveticus	TTTCTTGAGTGCAGAAGAGGAGAGTGGAACTCCATGTGTAGCGGTGGAAT	600
S000968187_helveticus	TTTCTTGAGTGCAGAAGAGGAGAGTGGAACTCCATGTGTAGCGGTGGAAT	600
S000965195_helveticus	TTTCTTGAGTGCAGAAGAGGAGAGTGGAACTCCATGTGTAGCGGTGGAAT	600
S000965200_helveticus	TTTCTTGAGTGCAGAAGAGGAGAGTGGAACTCCATGTGTAGCGGTGGAAT	600
S000965193_helveticus	TTTCTTGAGTGCAGAAGAGGAGAGTGGAACTCCATGTGTAGCGGTGGAAT	600
S000941666_helveticus	TTTCTTGAGTGCAGAAGAGGAGAGTGGAATTCCATGTGTAGCGGTGGAAT	600
S000633951_helveticus	TTTCTTGAGTGCAGAAGAGGAGAGTGGAATTCCATGTGTAGCGGTGGAAT	600
S000144437_helveticus	TTTCTTGAGTGCAGAAGAGGAGAGTGGAATTCCATGTGTAGCGGTGGAAT	600
S000608874_helveticus	TTTCTTGAGTGCAGAAGAGGAGAGTGGAATTCCATGTGTAGCGGTGGAAT	600
S000008891_helveticus	TTTCTTGAGTGCAGAAGAGGAGAGTGGAATTCCATGTGTAGCGGTGGAAT	599
	********** **********	

Overall, the proposed method for identification of yoghurt starter bacteria was composed of two parts. In the first part, for identification of *L. delbrueckii* subsp. *bulgaricus* species, amplification of methionine biosynthesis gene resulted in specific identification with closely related starter bacteria, i.e. *L. delbrueckii* subsp. *bulgaricus, L. delbrueckii* subsp. *lactis,* or *L. helveticus*. In the second part, differentiation of *L. delbrueckii* subsp. *bulgaricus* from *L. delbrueckii* subsp. *lactis* and *L. helveticus* was achieved by ARDRA.

For presumptive *S. thermophilus* isolates, methionine biosynthesis gene resulted in specific identification for 25 out of 28 isolates. Three of presumptive *S. thermophilus* isolates; N5-4, N7-1 and N7-4 did not amplify the desired methionine biosynthesis gene (Table 3.2). These species were determined as *S. thermophilus* by phenotypic and biochemical tests, and species specific PCR resulted contradictory. Strain N5-4 produced the species specific amplicon, while

speices N7-1 and N7-4 did not. For these isolates, analysis of 16S rRNA gene was necessary to assign species identification.

The proposed method confirmed species and subspecies identification of yoghurt starter bacteria, *L. delbrueckii* subsp. *bulgaricus* and *S. thermophilus*, previously identified by phenotypical methods, for 60 out of 63 strains. To confirm the methods' reliability and sensitivity, sequence analysis of 16S rRNA gene was performed.

3.5. Sequencing of 16S rRNA gene for all yoghurt isolates

Confirmation of the developed method was performed by partial sequencing of 16S rRNA gene. The sequenced region covers the variable region of 16S rRNA gene, corresponding to 800bp from 5'- downstream. The sequences were visualized and CLUSTALW analyses were performed using MEGA4. For *L. delbrueckii* subsp. *bulgaricus* strains all of the isolates produced the same sequence, and isolate K2-1 was selected for blast analysis (Appendix E). Result of the blast analysis verified results of the developed method, in that all of the isolates were *L. delbrueckii* subsp. *bulgaricus* strains (Appendix F).

For *S. thermophilus* isolates, same primers were used for sequencing. Isolates N7-1, N7-4 and N5-4 did not produce specific amplicon for methionine biosynthesis gene, and sequence analysis of 16S rRNA gene represented different sequences, confirming the sensitivity of method (Appendix E). Blast analysis of these isolates resulted that they belong to species *Streptococcus bovis* (Appendix F). *S. thermophilus* species specific primer misidentified strain N5-4 as *S. bovis*, while in our study; it was not misidentified, representing the sensitivity and reliability of the method. The blast analyses of remaining (25 out of 28) presumptive *S. thermophilus* isolates' identification by analysis of methionine biosynthesis gene were confirmed by partial sequencing of the 16S rRNA (Appendix F). The results of the specific identification method and sequencing of 16S rRNA gene for all isolates were displayed in Table 3.2.

Table 3.2. Results of experiments for identification of yoghurt starter bacteria

Species or subsp.	Strain number	Methionine biosynthesis	Identification by ARDRA	Identification by 16S rRNA sequencing
L. fermentum	NRRL-B1840	-	NA	NA
L. paraplantarum	NRRL-B23115	-	NA	NA
L. casei subsp. *casei*	NRRL-B1922	-	NA	NA
L. gasseri	NRRL-B4240	-	NA	NA
L. rhamnosus	NRRL-B442	-	NA	NA
L. helveticus	NRRL-B4526	+	*L. helveticus*	NA
L. paracasei	NRRL-B4560	-	NA	NA
L. reuteri	NRRL-B14171	-	NA	NA
L. salivarius	NRRL-B1949	-	NA	NA
L. johnsonii	NRRL-B2178	-	NA	NA
L. pentosus	NRRL-B227	-	NA	NA
L. amylovorous	NRRL-B4540	-	NA	NA
L. brevis	NRRL-B4527	-	NA	NA
L. acidophilus	ATCC 4356	-	NA	NA
L.d. delbruekii	NRRL-B763	-	NA	NA
L. d. lactis	NRRL-B4525	+	*L. lactis*	NA
*L. d. bulgaricus**	NRRL-B548	+	*L. helveticus*	NA
L. d. bulgaricus	DSM20081	+	*L. bulgaricus*	*L. bulgaricus*
L. d. bulgaricus	ATCCBAA-365	+	*L. bulgaricus*	NA
S. dysgalactiae	NRRL-B688	-	NA	NA
S. equinus	NRRL-B3573	-	NA	NA
S. infantarius	NRRL-B41208	-	NA	NA
S. thermophilus	LMG18311	+	NA	*S. thermophilus*
S. thermophilus	CNRZ1066	+	NA	NA
Lc. lactis	3113	-	NA	NA
Lc. lactis	2088	-	NA	NA
Lc. lactis	2910	-	NA	NA
Lc. lactis	2911	-	NA	NA
L. d. bulgaricus	B1000-1	+	*L. bulgaricus*	*L. bulgaricus*
L. d. bulgaricus	B1000-2	+	*L. bulgaricus*	*L. bulgaricus*
L. d. bulgaricus	B1000-3	+	*L. bulgaricus*	*L. bulgaricus*
L. d. bulgaricus	Yo-mix 410-1	+	*L. bulgaricus*	*L. bulgaricus*
*L. d. bulgaricus***	LB340-2	+	NA	*L. acidophilus*
*L. d. bulgaricus***	LB340-3	+	NA	*L. acidophilus*
S. thermophilus	Ta 040-1	+	NA	*S. thermophilus*
S. thermophilus	Ta 040-2	+	NA	*S. thermophilus*
S. thermophilus	B1000-3	+	NA	*S. thermophilus*
S. thermophilus	Yo-mix 410-3	+	NA	*S. thermophilus*
L. d. bulgaricus	M2-5	+	*L. d. bulgaricus*	*L. d. bulgaricus*
L. d. bulgaricus	M2-14	+	*L. d. bulgaricus*	*L. d. bulgaricus*
L. d. bulgaricus	M2-18	+	*L. d. bulgaricus*	*L. d. bulgaricus*
L. d. bulgaricus	K1-10	+	*L. d. bulgaricus*	*L. d. bulgaricus*
L. d. bulgaricus	K1-16	+	*L. d. bulgaricus*	*L. d. bulgaricus*
L. d. bulgaricus	K1-29	+	*L. d. bulgaricus*	*L. d. bulgaricus*
L. d. bulgaricus	K1-33	+	*L. d. bulgaricus*	*L. d. bulgaricus*
L. d. bulgaricus	K1-38	+	*L. d. bulgaricus*	*L. d. bulgaricus*
L. d. bulgaricus	K1-43	+	*L. d. bulgaricus*	*L. d. bulgaricus*
L. d. bulgaricus	K1-44	+	*L. d. bulgaricus*	*L. d. bulgaricus*
L. d. bulgaricus	M23-1	+	*L. d. bulgaricus*	*L. d. bulgaricus*
L. d. bulgaricus	M23-2	+	*L. d. bulgaricus*	*L. d. bulgaricus*

Table 3.1. Results of experiments for identification of yoghurt starter bacteria (cont'd)

Species or subsp.	Strain number	Methionine biosynthesis	Identification by ARDRA	16S rRNA sequencing
L. d. bulgaricus	M23-3	+	*L. d. bulgaricus*	*L. d. bulgaricus*
L. d. bulgaricus	M23-4	+	*L. d. bulgaricus*	*L. d. bulgaricus*
L. d. bulgaricus	M23-13	+	*L. d. bulgaricus*	*L. d. bulgaricus*
L. d. bulgaricus	N2-4	+	*L. d. bulgaricus*	*L. d. bulgaricus*
L. d. bulgaricus	N2-5	+	*L. d. bulgaricus*	*L. d. bulgaricus*
L. d. bulgaricus	N3-2	+	*L. d. bulgaricus*	*L. d. bulgaricus*
L. d. bulgaricus	N4-3	+	*L. d. bulgaricus*	*L. d. bulgaricus*
L. d. bulgaricus	N6-2	+	*L. d. bulgaricus*	*L. d. bulgaricus*
L. d. bulgaricus	K2-1	+	*L. d. bulgaricus*	*L. d. bulgaricus*
L. d. bulgaricus	K2-2	+	*L. d. bulgaricus*	*L. d. bulgaricus*
L. d. bulgaricus	K2-3	+	*L. d. bulgaricus*	*L. d. bulgaricus*
L. d. bulgaricus	K2-4	+	*L. d. bulgaricus*	*L. d. bulgaricus*
L. d. bulgaricus	K2-5	+	*L. d. bulgaricus*	*L. d. bulgaricus*
L. d. bulgaricus	M21-3	+	*L. d. bulgaricus*	*L. d. bulgaricus*
L. d. bulgaricus	M21-4	+	*L. d. bulgaricus*	*L. d. bulgaricus*
L. d. bulgaricus	M2-8	+	*L. d. bulgaricus*	*L. d. bulgaricus*
L. d. bulgaricus	M2-12	+	*L. d. bulgaricus*	*L. d. bulgaricus*
L. d. bulgaricus	M2-16	+	*L. d. bulgaricus*	*L. d. bulgaricus*
L. d. bulgaricus	M2-17	+	*L. d. bulgaricus*	*L. d. bulgaricus*
L. d. bulgaricus	M2-20	+	*L. d. bulgaricus*	*L. d. bulgaricus*
L. d. bulgaricus	M2-21	+	*L. d. bulgaricus*	*L. d. bulgaricus*
L. d. bulgaricus	K1-19	+	*L. d. bulgaricus*	*L. d. bulgaricus*
S. thermophilus	N2-1	+	NA	*S. thermophilus*
S. thermophilus	N2-3	+	NA	*S. thermophilus*
S. thermophilus	N2-4	+	NA	*S. thermophilus*
S. thermophilus	N3-1	+	NA	*S. thermophilus*
S. thermophilus	N6-1	+	NA	*S. thermophilus*
S. thermophilus	N6-3	+	NA	*S. thermophilus*
S. thermophilus ***	N5-4	-	NA	*S. bovis*
S. thermophilus	N5-7	+	NA	*S. thermophilus*
S. thermophilus ***	N7-1	-	NA	*S. bovis*
S. thermophilus ***	N7-4	-	NA	*S. bovis*
S. thermophilus	N8-2	+	NA	*S. thermophilus*
S. thermophilus	S1-3	+	NA	*S. thermophilus*
S. thermophilus	K1-1	+	NA	*S. thermophilus*
S. thermophilus	K1-7	+	NA	*S. thermophilus*
S. thermophilus	K1-9	+	NA	*S. thermophilus*
S. thermophilus	K1-12	+	NA	*S. thermophilus*
S. thermophilus	K1-13	+	NA	*S. thermophilus*
S. thermophilus	K1-14	+	NA	*S. thermophilus*
S. thermophilus	K1-15	+	NA	*S. thermophilus*
S. thermophilus	K1-16	+	NA	*S. thermophilus*
S. thermophilus	K1-18	+	NA	*S. thermophilus*
S. thermophilus	K1-19	+	NA	*S. thermophilus*
S. thermophilus	K1-23	+	NA	*S. thermophilus*
S. thermophilus	K1-24	+	NA	*S. thermophilus*
S. thermophilus	K1-27	+	NA	*S. thermophilus*
S. thermophilus	K1-28	+	NA	*S. thermophilus*
S. thermophilus	K1-29	+	NA	*S. thermophilus*
S. thermophilus	K1-30	+	NA	*S. thermophilus*

* Analysis by ARDRA revealed that the strain actually belongs to species *L. helveticus.*

** Blast analysis of 16S rRNA gene sequencing revealed that the strains belong to *L. acidophilus*

***Blast analysis of 16S rRNA gene sequencing revealed that the strains belong to *S. bovis*.

In conclusion, the proposed method was successful in joint identification of yoghurt starter bacteria at species and subspecies level. It became possible to selectively identify *L. delbrueckii* subsp. *bulgaricus* and *S. thermophilus* strains either from culture collections, or fermented dairy products and traditional yoghurts. The method was also able to discriminate yoghurt starter bacteria from *L. helveticus* and *L. delbrueckii* subsp. *lactis*.

Confirmation of the method had been performed using partial sequencing of the 16S rRNA gene. Differentiation at the strain level had been studied by RAPD-PCR and MLST analysis. Typing of starter bacteria was necessary because properties of yoghurt starters are strain dependent, therefore they must be identified at the strain level, as well.

3.6. Use of RAPD-PCR for typing of yoghurt starter bacteria

Following identification of yoghurt starters, strains were discriminated by RAPD-PCR for selected *S. thermophilus* and *L. delbrueckii* subsp. *bulgaricus* strains. The selection criteria were based on technological properties of strains. After a detailed literature survey, 14 preliminary primers were selected; those were D8635, coc, qln07, M13, leuc, cc1, tr99, pl1, eric2, opl5, cld06, 1254, arg-dei-for and primm239. Comparison of each of these primers was performed using 5-6 samples, including reference strains. Fingerprints of each primer were compared and primers M13 and 1254 were selected for complete RAPD analysis, because they produced the highest number of fragments and they were able to differentiate among strains. Remaining 12 primers produced either a low number of bands, or they produced similar bands, which could not be used to discriminate isolates at strain level, thus they were excluded.

For analysis of the RAPD amplification patterns, the clearest and most reproducible bands were chosen for determination of their presence or absence in each strain, and bands whose sizes fell between 200 and 2000 bp were considered only, since bands outside this range had variations between different runs. Bands

were considered present or absent regardless of their intensities. In addition, faint bands which could not be systematically visualized were not taken into account.

The *S. thermophilus* and *L. delbrueckii* subsp. *bulgaricus* strains were grouped into clusters at the 80% similarity level, which was stated as the minimum value for consistent and reliable grouping of clusters (Rossetti and Giraffa, 2005, Kenny et al, 2005).

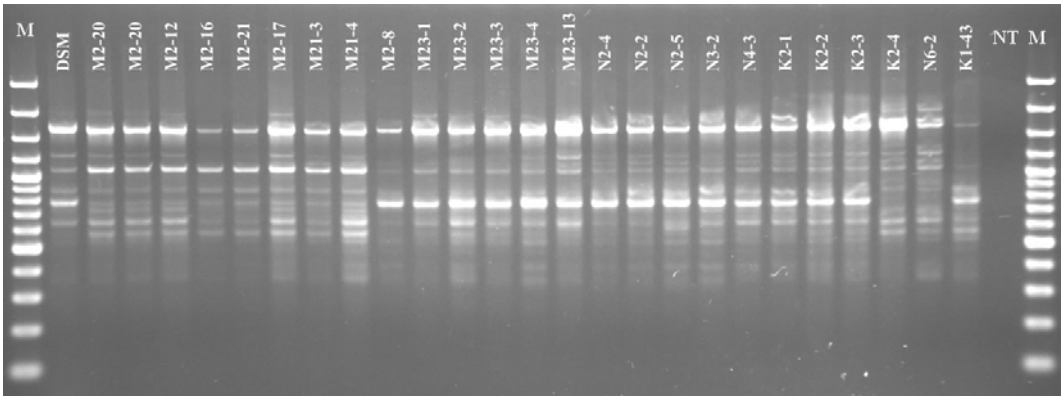
RAPD-PCR analysis is an easy-to-perform analysis; however, stringent conditions were required. The disadvantages of the method were; a) high sensitivity to components of reaction mixture and amplification conditions, b) sensitivity to contamination and c) low reproducibility among different laboratories.

3.6.1. RAPD analysis with primer M13

M13 analyses of home-made yoghurt isolates with RAPD-PCR were applied to *S. thermophilus* and *L. delbrueckii* subsp. *bulgaricus* strains (Figures 3.12a and 3.13a). In both of the figures, it was clear that the strains isolated from the same source (i.e. isolates starting with the same initials) tend to cluster together, as stated by other researchers (Corsetti et al., 2003).

L. delbrueckii subsp. *bulgaricus* isolates were grouped into five clusters at the 80% similarity level. The bands produced by M13 ranged from 500bp to 2000bp, and the number of bands differed between 6 to 9. Several common bands were present, although in some isolates these bands were fainter. The common bands include 1700bp, 1100bp and 650bp. Analysis of M13 UPGMA tree for *L. delbrueckii* subsp. *bulgaricus* revealed that strains isolated from the same source tend to cluster separately, however, several isolates were clustered independent of their source (Figure 3.12b).

A)



B)

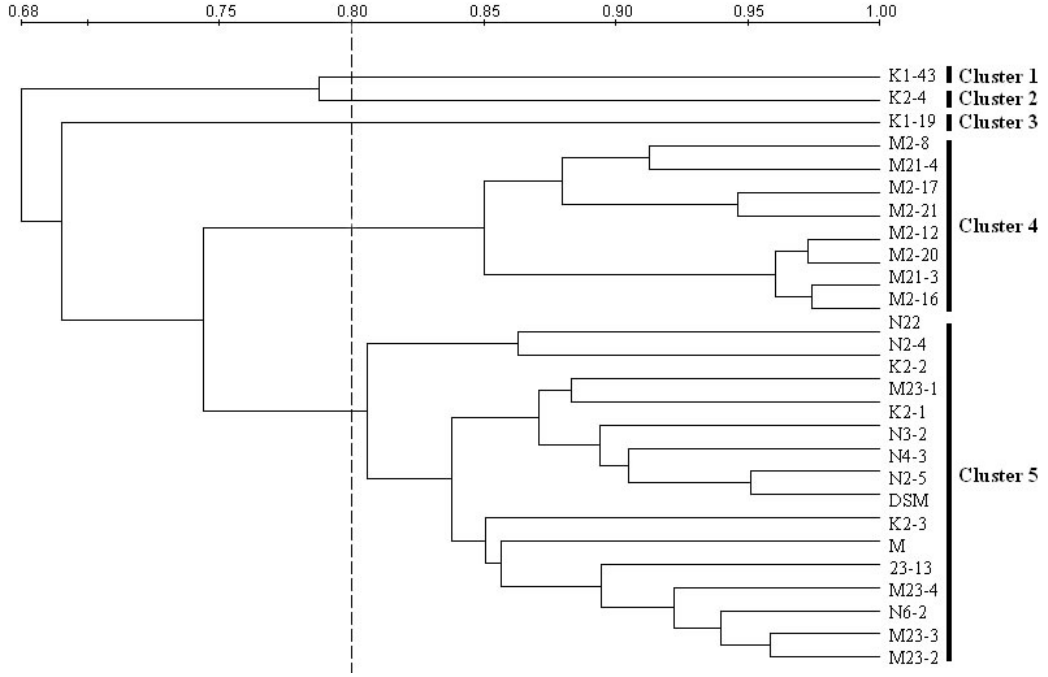
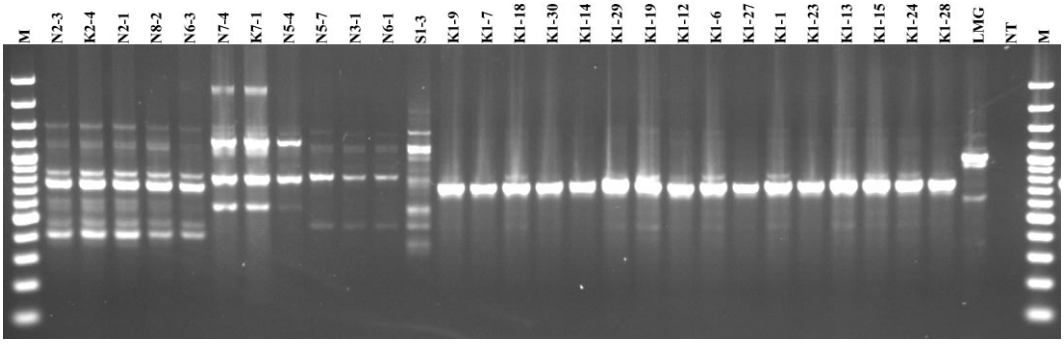


Figure 3.12. M13 analysis for typing of *L. delbrueckii* subsp. *bulgaricus* A) Gel photograph of RAPD with primer M13. M: 100bp ladder plus, DSM: *L. delbrueckii* subsp. *bulgaricus* DSM20081, NT: Negative control (no DNA template), B) Dendrogram of RAPD-PCR patterns of the *L. delbrueckii* subsp. *bulgaricus* strains after their numerical analysis using UPGMA, clustering was performed at 80% similarity level.

A)



B)

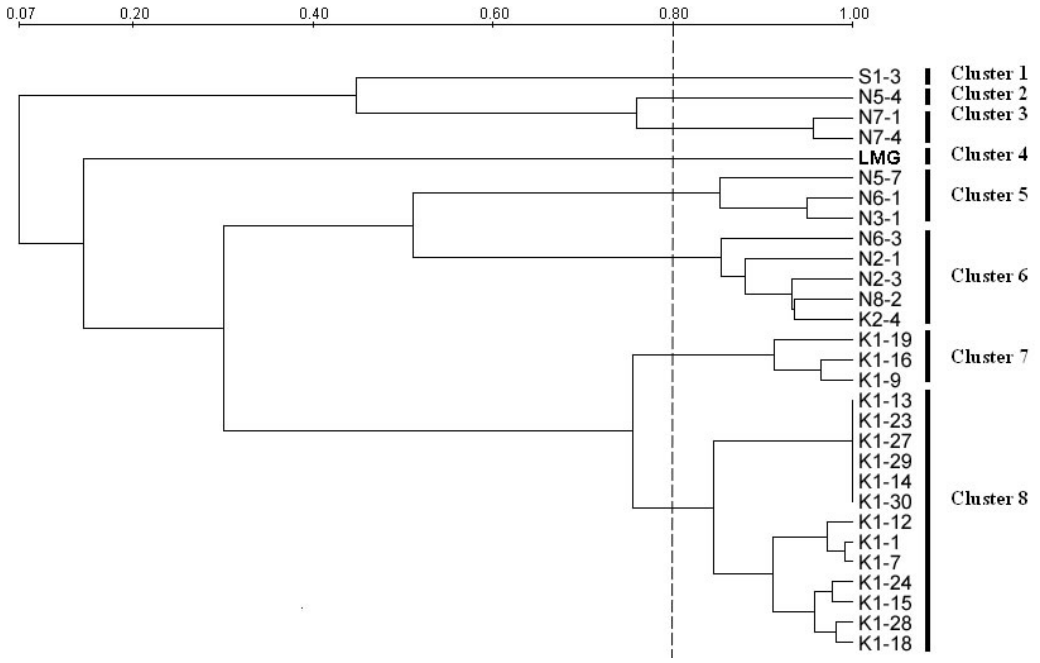


Figure 3.13. M13 analysis for typing of *S. thermophilus*. A) Gel photograph of M13 fingerprinting. M: 100bp ladder plus, LMG: *S. thermophilus* LMG18311, NT: negative control (no DNA template). B) UPGMA dendrograms derived from comparisons of the RAPD-PCR patterns obtained with primer M13. Clustering was performed at 80% similarity level.

For *S. thermophilus* isolates, at the similarity level of 80%, 29 strains were grouped into eight clusters. The bands range from 400bp to 2500bp, and the number of bands differs between 1-7 bands. M13 fingerprinting resulted in no

common band in *S. thermophilus* isolates, however, a band at 750bp was visible for most of the isolates, except in isolates N5-7, N3-1, N6-1 and in the reference strain. Isolates starting from K1 (K1-9 to K1-30) lanes 13 to 28 in the figure displayed only one or two bands, polymorphism in these strains were revealed at limited level by primer M13. In the UPGMA dendrogram of *S. thermophilus* isolates, strains S1-3, N7-1, N7-4 and N5-4 were clustered separately from the rest of the isolates (Figure 3.13b). The strains N7-1, N7-4 and N5-4 were identified as *S. bovis* by sequencing of 16S rRNA gene, and included to observe differences produced by primer M13. The lowest similarity values for *S. thermophilus* isolates was 14%; lower than those of *L. delbrueckii* subsp. *bulgaricus* (68%), indicating that there exist higher genetic diversity among *S. thermophilus* isolates than in *L. delbrueckii* subsp. *bulgaricus* isolates.

Analysis of percent similarity values between *L. delbrueckii* subsp. *bulgaricus* and *S. thermophilus* isolates showed that genetic diversity among *S. thermophilus* isolates was higher in comparison to *L. delbrueckii* subsp. *bulgaricus.*

RAPD analysis of commercial strains with primer M13 was performed on a separate agarose gel, apart from the isolates from traditional yoghurts (Figure 3.14).

Analysis of commercial strains presumed to be *L. delbrueckii* subsp. *bulgaricus* had shown a homology in banding patterns. Strains B1000-1, B1000-2, B1000-3 and Yo-mix 410-1 had identical banding patterns, and were very similar to the reference strain DSM20081. Strains B1000-1, B1000-2 and B1000-3 were isolated from the same commercial starter, and the same band patterns could be expected, however their strong resemblance to the reference strain and to Yo-mix 410-3 was considered important. On the other hand, strains LB340-2 and LB340-3 had very similar banding patterns in between them, however, similarity to the other strains differed, in that, the common band above 1200bp was absent in these two strains. Indeed sequence analysis of 16S rRNA genes of these two strains revealed that they belong to *L. acidophilus* species.

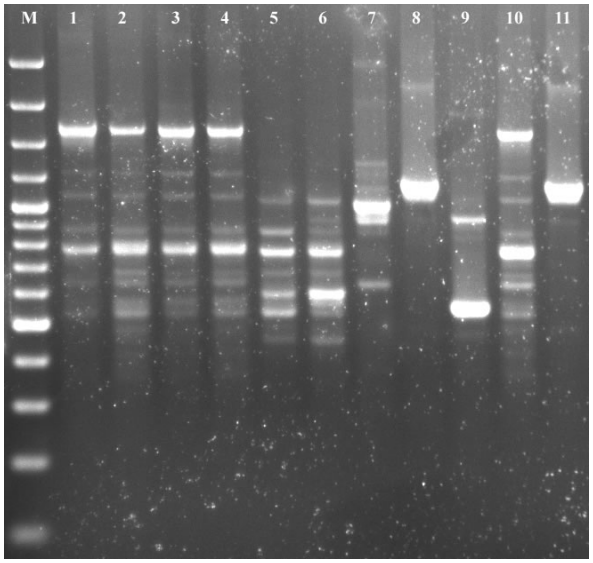


Figure 3.14. RAPD-PCR analysis of commercial yoghurt starters with primer M13. M: 100bp ladder plus, lanes 1-6 and 10 *L. delbrueckii* subsp. *bulgaricus* (1: DSM20081, 2: B1000-1, 3: B1000-2, 4: B1000-3, 5: LB340-2, 6: LB340-3, 10: Yo-mix 410-1), lanes 7-9 and 11 *S. thermophilus* (7: LMG18311, 8: Ta040-2, 9: Ta040-1, 11: Yo-mix 410-3).

3.6.2. RAPD analysis with primer 1254

RAPD analyses with primer 1254 were studied on *L. delbrueckii* subsp. *bulgaricus* and *S. thermophilus* strains. The analysis with primer 1254 revealed a clear banding pattern for all the isolates. The banding patterns obtained with primer M13 resulted in bands that were difficult to distinguish by QuantityOne software (BioRad), and manual corrections had to be done. However, in 1254, manual corrections were performed only for few numbers of bands. Thus, 1254 banding patterns were easier to interpret than M13 banding patterns.

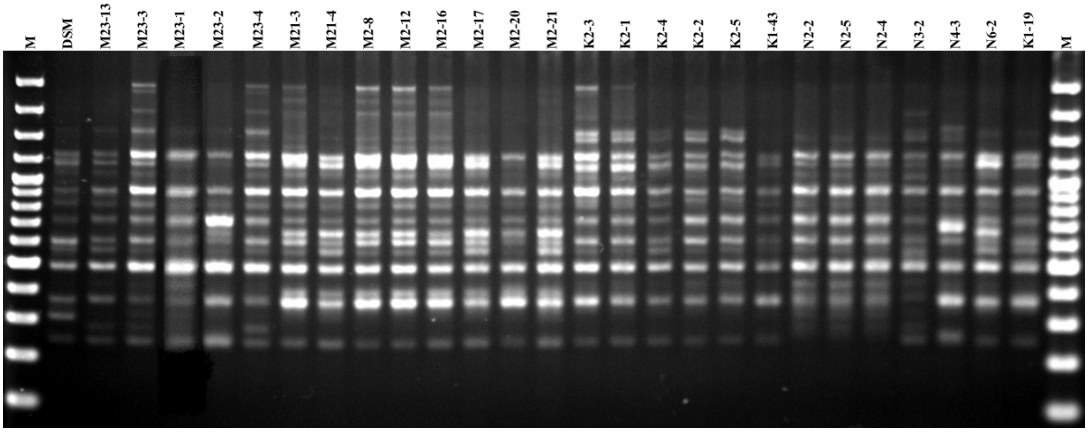
Primer 1254 grouped 27 *L. delbrueckii* subsp. *bulgaricus* strains into seven clusters. The number of bands obtained with RAPD analysis with primer 1254 in *L. delbrueckii* subsp. *bulgaricus* isolates differed between 10-15 bands, and four

common bands could be observed; 500bp, 700bp, 900bp, 1200bp (Figure 3.15a). The bands ranged from 200bp to 3000bp. Grouping of isolates were more source dependent in 1254 fingerprinting than obtained with M13 fingerprinting (Figure 3.15b). Isolates with names starting with capitals N and K tend to cluster closely, while M23- isolates were clustered with the subspecies *bulgaricus* type strain in a separate cluster, and M21- and M2- isolates clustered separately. The lowest similarity values obtained from primer M13 and 1254 were similar for *L. delbrueckii* subsp. *bulgaricus* isolates, 68% and 64%, respectively.

Fingerprinting with 1254 primer revealed clear banding patterns in *S. thermophilus* isolates. Higher numbers of bands were produced by primer 1254, compared to primer M13, the number of bands differs in between 5-8, and bands ranged from 200bp to 2500bp. Two common bands, one at 700bp and the other at 1700bp were observed in all *S. thermophilus* strains (Figure 3.16a).

S. thermophilus isolates were grouped into 4 clusters at the 80% similarity level with primer 1254. Among *S. thermophilus* isolates from traditional yoghurts, strains N7-1, N7-4 and N5-4, which were identified as *S. bovis* by sequencing of 16S rRNA genes, were clustered separately from remaining strains (Figure 3.16b). Strains from K1-1 to K1-30 were clustered in a separate group. Fingerprinting with 1254 resulted quite similar to M13 fingerprinting, however, an important difference in 1254 fingerprinting was that strain S1-3 had been clustered within *S. thermophilus* group, while in M13 fingerprinting; this strain was closely grouped to *S. bovis* (N5-4, N7-1 and N7-4) with similarity values higher than 40%. Another important difference with 1254 was that reference strain (LMG 18311) had been closely related to *S. thermophilus* group, while M13 analysis clustered the reference strain as an outgroup. The percent similarity values obtained in *S. thermophilus* strains (excluding *S. bovis* isolates) by primer 1254 (45%) were higher than those with primer M13 (14%).

A)



B)

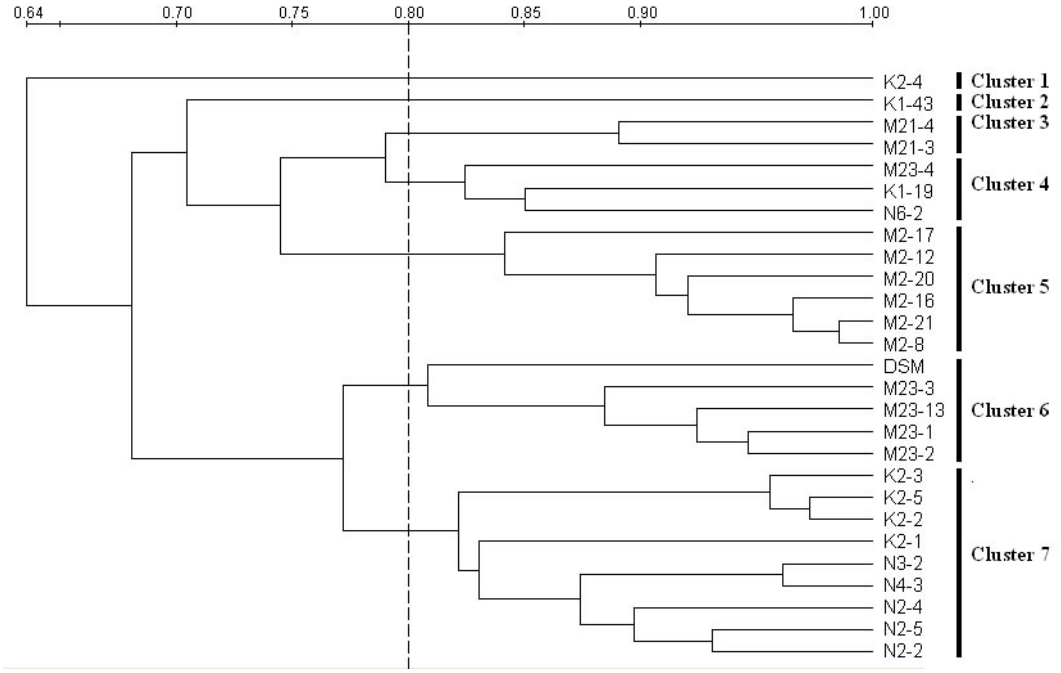
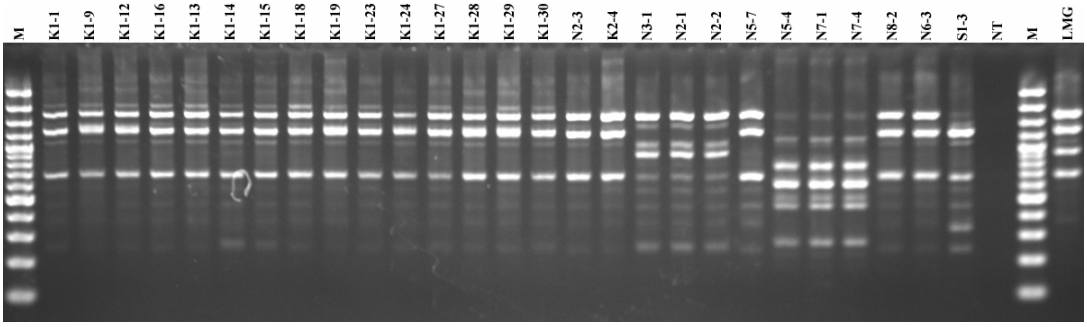


Figure 3.15. 1254 analysis for typing of *L. delbrueckii* subsp. *bulgaricus.* A) Gel photograph of RAPD with primer 1254. M: 100bp ladder plus, DSM: *L. delbrueckii* subsp. *bulgaricus* DSM20081, B) UPGMA dendrogram drawn from comparisons of the RAPD-PCR patterns by QuantityOne software (BioRad). Clusters were grouped at 80% similarity level.

A)



B)

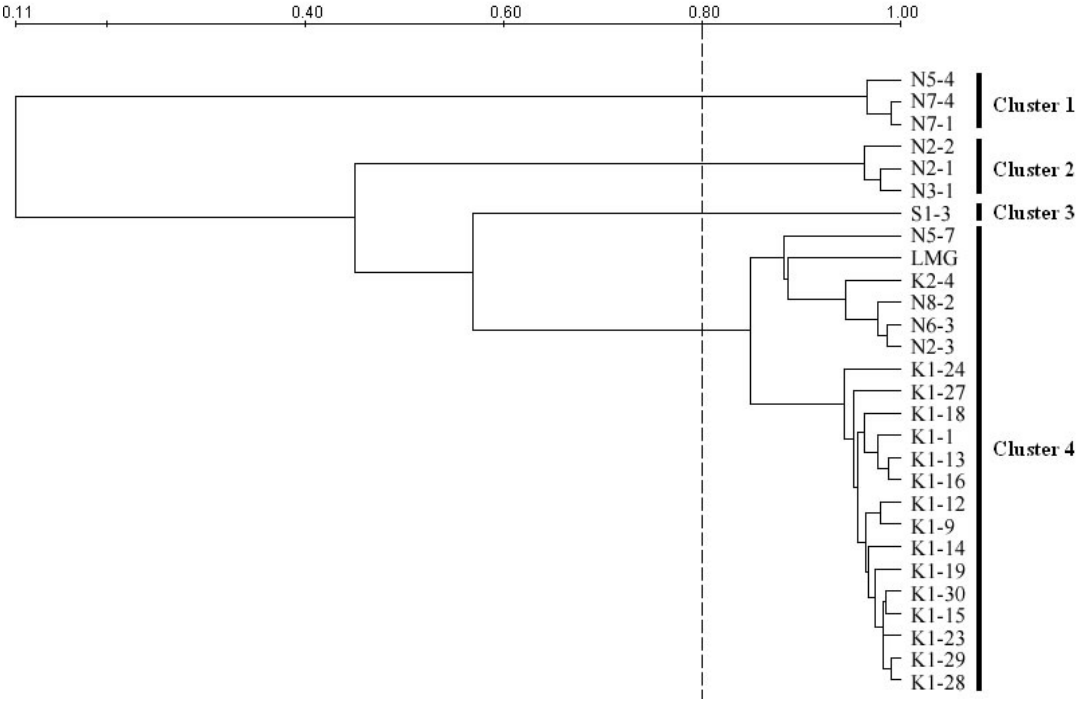


Figure 3.16. RAPD analysis with primer 1254 for typing of *S. thermophilus.* A) Gel photograph of 1254 fingerprinting. M: 100bp ladder plus, NT: negative control (no DNA template), LMG: *S. thermophilus* LMG18311. B) UPGMA dendrograms derived from comparisons of the RAPD-PCR patterns obtained with primer 1254. Clusters were grouped at 80% similarity level.

RAPD analysis of commercial strains with primer 1254 was also performed (Figure 3.17). Among the presumed *L. delbrueckii* subsp. *bulgaricus* commercial strains, Yo-mix 410-1B1000 strain displayed the closest pattern to reference strain *L. delbrueckii* subsp. *bulgaricus* DSM20081. B1000-1, B1000-2 and B1000-3 strains produced the same banding patterns, and they contained four common bands with the reference strain. On the other hand, LB340-2 and 340-3 strains had different banding patterns than the reference strain, only one common band at 500bp was observed in the figure. These LB strains had later been identified as *L. acidophilus* by partial sequencing of the 16S rRNA gene (Appendix F).

Among the presumed *S. thermophilus* commercial strains, the reference strain (*S. thermophilus* LMG18311), and three commercial strains, Ta040-1, Ta 040-2 and Yo-mix 410-3 contained common bands around 700bp and 1500bp.

On the other hand, the heterogeneity obtained with commercial strains of *S. thermophilus* were higher when compared to *L. delbrueckii* subsp. *bulgaricus* commercial strains, though, this coincided well with the heterogeneity obtained with the isolates from traditional yoghurts.

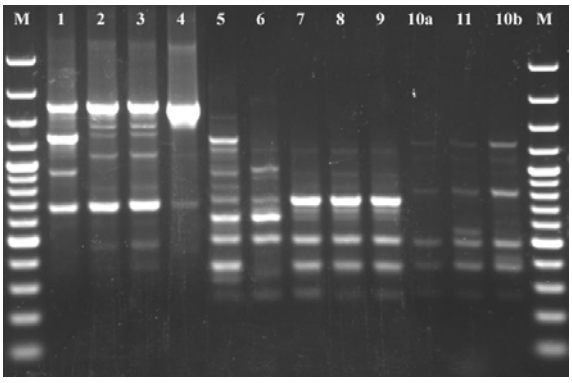


Figure 3.17. RAPD analysis of commercial yoghurt starters with primer 1254. M: 100bp ladder plus, lanes 1-4 *S. thermophilus* (1: LMG18311, 2: Ta040-1, 3: Ta040-2, 4: Yo-mix 410-3), lanes 5-11 *L. delbrueckii* subsp. *bulgaricus* (5: DSM20081, 6: Yo-mix 410-1, 7: B1000-1, 8: B1000-2, 9: B1000-3, 10a,b: LB340-2, 11: LB340-3).

Taken as a whole, typing of yoghurt starter bacteria by RAPD analysis gave good resolution at strain level. It had been observed from the dendrograms that, for both of the primers tested; isolates from the same source tend to cluster together. The cluster analysis of the isolates were done at a similarity level of 80%, for *L. delbrueckii* subsp. *bulgaricus* isolates, primer M13 grouped strains into five clusters while 1254 grouped isolates into seven clusters. For *S. thermophilus* isolates, primer M13 grouped isolates into eight clusters, while 1254 resulted in four clusters. Primer 1254 produced clearer bands and the produced bands were distributed on a wider range than primer M13. In addition, groupings of the isolates according to source were displayed better by primer 1254 for both starter strains. Consequently, the bands produced by 1254 were easier to interpret for both starters. When dendrograms of the two starters were compared, it had been seen that, for both primers, *L. delbrueckii* subsp. *bulgaricus* isolates had higher percent similarity values, while similarity values of *S. thermophilus* isolates were lower, suggesting the presence of a higher degree of DNA polymorphism.

3.7. Analysis of yoghurt starter bacteria by MLST

MLST is used for differentiation of isolates at strain level. The method is based on differences present in protein coding genes' DNA sequences. Differences in nucleotide sequences were interpreted as distinct alleles, and alleles at each locus determine the sequence type (ST) of an organism. The analysis of MLST was applied to isolates identified as *L. delbrueckii* subsp. *bulgaricus* and S. *thermophilus*.

3.7.1. Analysis of *L. delbrueckii* subsp. *bulgaricus* by MLST

The selected genes for MLST analysis of *L. delbrueckii* subsp. *bulgaricus* strains were β-*gal*, *phe*S and *rpo*A. The amplification products of β-*gal* (800bp), *phe*S (500bp) and *rpo*A (850bp) genes were displayed at figure 3.18.

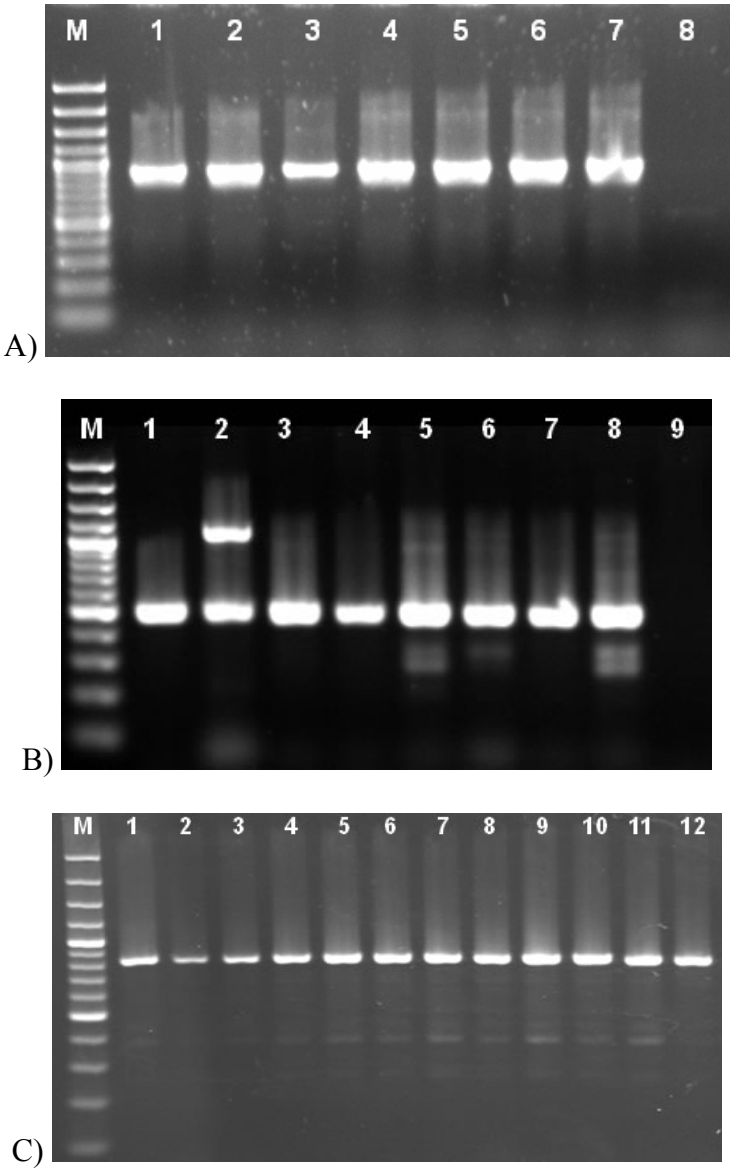


Figure 3.18. Gel photographs of A) β-*gal* gene products, M: 100bp DNA ladder plus, lanes 1-7 *L. delbrueckii* subsp. *bulgaricus,* 1: DSM20081, 2: M21-4, 3: N6-2, 4: M2-16, 5:M2-14, 6: K2-3, 7: N2-5, 8: Negative control, B) *pheS* gene products M: 100bp DNA ladder plus, lanes 1-8 *L. delbrueckii* subsp. *bulgaricus*, 1: DSM20081, 2: K1-43, 3: M21-3, 4: M2-16, 5:M2-21, 6: K2-3, 7: N2-5, 8: M23-1, 9: Negative control, C) *rpoA* gene products, M: 100bp DNA ladder plus, lanes 1-12 *L. delbrueckii* subsp. *bulgaricus*, 1: DSM20081, 2: K1-19, 3: M21-3, 4: M2-16, 5:M2-20, 6: K2-2, 7: K2-3, 8: N2-5, 9: M23-1, 10: M23-1, 11: M2-8, 12: N2-4.

After amplification of gene products was achieved, the products were purified from the agarose gel to remove unspecific bands. The purified products were subjected to sequencing, and evaluations of the sequences were performed by MEGA4 and by visual comparisons. The sequences at the beginning and at the end were trimmed, since the chromatograms displayed unclear peaks at these sites. Afterwards, the sequences of all strains were aligned, and the sequences were adjusted to same length. The procedure was applied to β-*gal*, *pheS* and *rpoA* genes.

For each gene, sequences were analyzed and assigned sequence types. The profile frequencies (Table 3.3) of *L. delbrueckii* subsp. *bulgaricus* strains revealed that the most frequent profile was ST-2 (18,52%), followed by ST-5 (14,81%) and ST-13 (11,11%). It was seen that isolates from the same yoghurt samples either had identical alleles for three out of three genes or two out of three genes. The reference strain *L. delbrueckii* subsp. *bulgaricus* DSM20081 were given allele number (1-1-1) and ST-1. The sequence types for all strains used were displayed in Table 3.4. It had been seen that 15 sequence types were assigned to 25 strains.

Table 3.3. Profile frequencies of *L. delbrueckii* subsp. *bulgaricus* strains

ST	Strain numbers	Profile			Frequency	%
		β-*gal*	*phe*S	*rpo*A		
1	1	1	1	1	1	3,7
2	2, 3, 4, 5, 6	6	1	2	5	18,52
3	7	3	2	1	1	3,7
4	8, 26	3	1	1	2	7,41
5	9, 12, 14, 15	5	2	3	4	14,81
6	10	6	2	3	1	3,7
7	11	5	6	3	1	3,7
8	13	5	3	3	1	3,7
9	16	5	1	3	1	3,7
10	17, 19	6	2	1	2	7,41
11	18	4	1	1	1	3,7
12	20, 21	6	1	1	2	7,41
13	22, 23, 24	3	4	1	3	11,11
14	25	1	1	4	1	3,7
15	27	2	5	1	1	3,7

Table 3.4. Sequence type (ST) assignment for *L. delbrueckii* subsp. *bulgaricus* strains

			Allele number at locus		
Strain no	Strain	ST	β-*gal*	*phe*S	*rpo*A
1	DSM20081	1	1	1	1
2	K2-1	2	6	1	2
3	K2-2	2	6	1	2
4	K2-3	2	6	1	2
5	K2-4	2	6	1	2
6	K2-5	2	6	1	2
7	K1-19	3	3	2	1
8	K1-43	4	3	1	1
9	M21-3	5	5	2	3
10	M21-4	6	6	2	3
11	M2-8	7	5	6	3
12	M2-12	5	5	2	3
13	M2-16	8	5	3	3
14	M2-17	5	5	2	3
15	M2-20	5	5	2	3
16	M2-21	9	5	1	3
17	M23-1	10	6	2	1
18	M23-2	11	4	1	1
19	M23-3	10	6	2	1
20	M23-4	12	6	1	1
21	M23-13	12	6	1	1
22	N2-2	13	3	4	1
23	N2-4	13	3	4	1
24	N2-5	13	3	4	1
25	N3-2	14	1	1	4
26	N4-3	4	3	1	1
27	N6-2	15	2	5	1

The UPGMA dendrogram of MLST analysis of *L. delbrueckii* subsp. *bulgaricus* strains (Figure 3.20) displayed two main clusters. Most of the strains were placed into distinct branches; however strains having identical STs were placed at the same external nodes (strain numbers 22-23-24, 26-8, 2-3-4-5-6, 20-21, 9-12-14-15).

MLST analysis of *L. delbrueckii* subsp. *bulgaricus* using genes β-*gal*, *phe*S and *rpo*A genes revealed differentiation at strain level. MLST enabled a deeper understanding of the *L. delbrueckii* subsp. *bulgaricus* strains, through analysis of

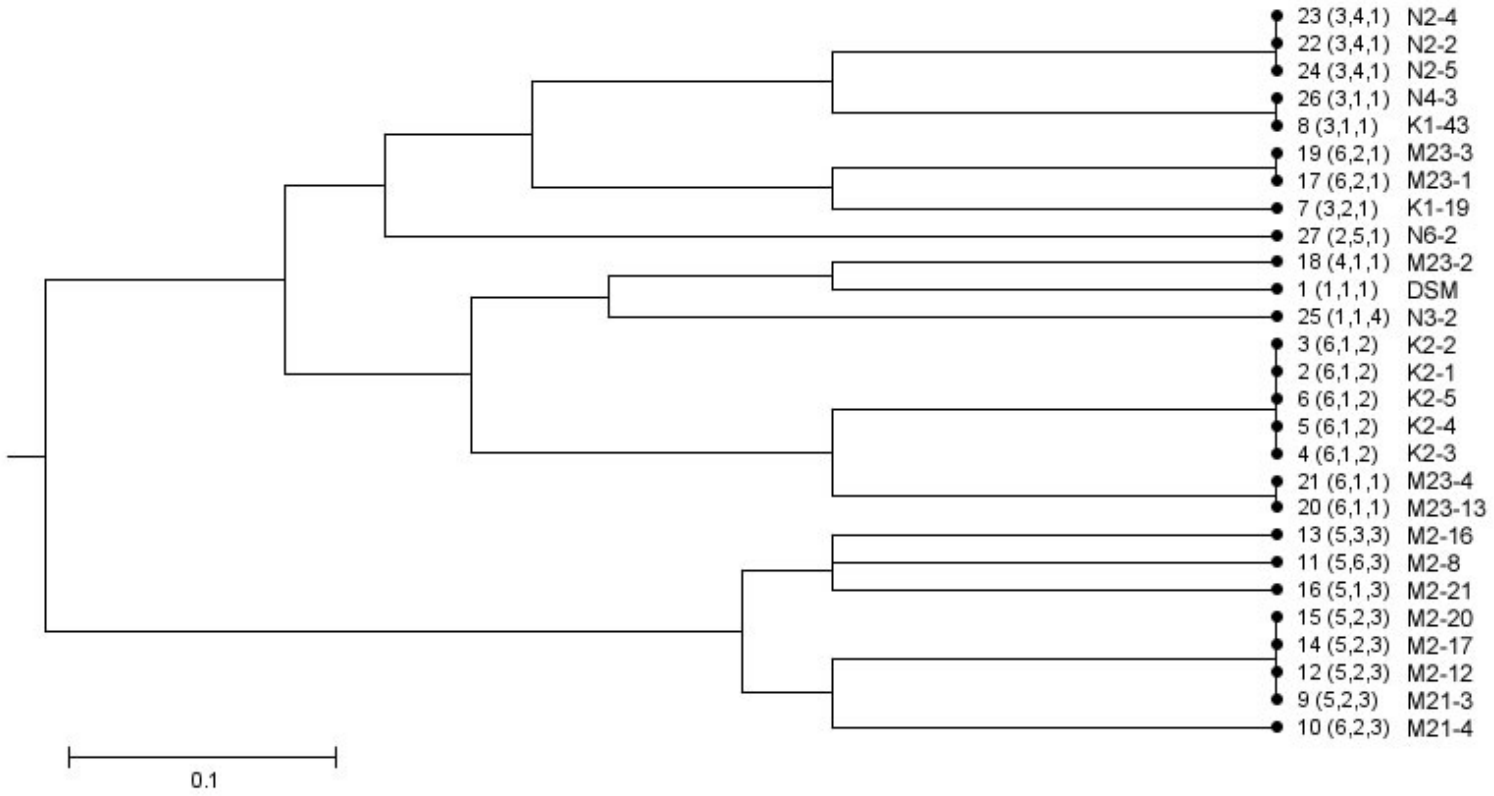


Figure 3.19. MLST analysis of *L. delbrueckii* subsp. *bulgaricus* strains displayed in an UPGMA tree. ST: sequence type, numbers in parenthesis refer to allele numbers for each sequence type.

alleles and sequence types. The determination of allele numbers eased comparisons of strains on a reliable basis. Increasing the number of loci may help differentiating strains to a higher extent. However, the analysis of new loci would be additive only if hypervariable loci was used. Meanwhile, sequencing analysis of high numbers of strains was expensive, and the number of loci should be chosen carefully.

3.7.2. Analysis of *S. thermophilus* by MLST

The selected genes for MLST analysis of *S. thermophilus* strains were *phe*S and *rpo*A. The genes were amplified, and the products were purified from agarose gels for sequencing. Amplified genes produced bands of 500bp for *phe*S and and 850bp for *rpo*A genes, same as those produced by *L. delbrueckii* subsp. *bulgaricus* strains (Figure 3.18). DNA sequences of the genes were obtained, and evaluated by MEGA4. The sequences at the beginning and at the end were cut because of the uncertanities visible at the chromatograms. Afterwards, the sequences of all strains were aligned, and the sequences were adjusted to same length. The procedure was applied to both *phe*S and *rpo*A genes.

For each gene, sequences were analyzed and assigned sequence types. The reference strain *S. thermophilus* LMG18311 were given allele numbers (1-1) and ST-1. The sequence types for all strains used were displayed in Table 3.5. It was seen from the table that only seven sequence types were assigned to 25 strains. The profile frequencies (Table 3.6) of *S. thermophilus* strains revealed that the most frequent profile was ST-2 (40%), followed by ST-1 (28%). Yoghurt starters from K1- sample generally had the allele numbers (2-1). For *phe*S gene, polymorphism was low; only three alleles were present, while *rpo*A gene displayed higher polymorphism, in that six alleles were present. The sequence differences in each allele were displayed in Appendix H. The mean GC% value for *phe*S gene was 43.6, while for *rpo*A gene, it was 41.0%. Considering that *S. thermophilus* has an average GC% of 39 (Bolotin et al., 2004), *phe*S gene had a high GC content.

Table 3.5. Sequence type (ST) assignment for *S. thermophilus* strains

			Allele number at locus	
Strain no	Strain	ST	*rpo*A	*phe*S
1	LMG 18311	1	1	1
2	K1-1	1	1	1
3	K1-6	2	6	1
4	K1-7	2	6	1
5	K1-9	2	6	1
6	K1-12	2	6	1
7	K1-13	3	2	1
8	K1-14	1	1	1
9	K1-15	1	1	1
10	K1-18	2	6	1
11	K1-19	2	6	1
12	K1-23	2	6	1
13	K1-24	1	1	1
14	K1-27	4	3	1
15	K1-28	1	1	1
16	K1-29	1	1	1
17	K1-30	2	6	1
18	K2-4	3	2	1
19	N2-1	5	4	1
20	N2-3	2	6	1
21	N5-7	6	5	3
22	N6-1	6	5	3
23	N8-2	2	6	1
24	N8-3	5	4	1
25	S1-3	7	5	2

Table 3.6. Profile frequencies of *S. thermophilus* strains

ST	Strain numbers	Profile		Frequency	%
		*rpo*A	*phe*S		
1	1, 2, 8, 9, 13, 15, 16	1	1	7	28
2	3, 4, 5, 6, 10, 11, 12, 17, 20, 23	6	1	10	40
3	7, 18	2	1	2	8
4	14	3	1	1	4
5	19, 24	4	1	2	8
6	21, 22	5	3	2	8
7	25	5	2	1	4

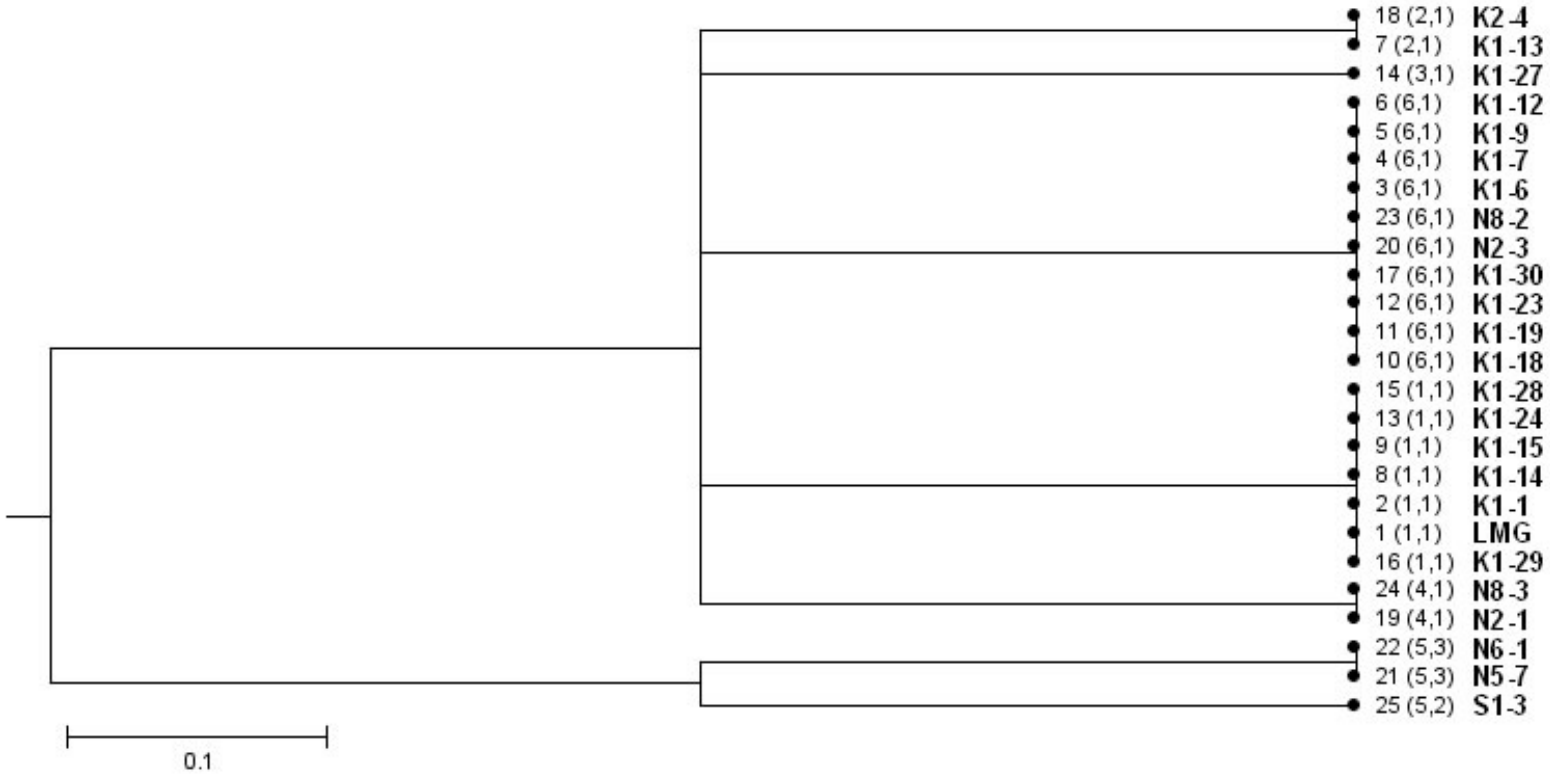


Figure 3.20. MLST analysis of *S. thermophilus* strains displayed in an UPGMA tree. ST: sequence type, numbers in parenthesis refer to allele numbers for each sequence type.

The UPGMA dendrogram of MLST analysis of *S. thermophilus* strains (Figure 3.19) displayed two main clusters. Strains N5-7, N6-1 and S1-3 were grouped in one cluster, while rest of the strains was grouped into a second cluster, where strain diversity was not well reflected. This resulted from the fact that the latter cluster contained strains having similar or identical allele numbers. This outcome could be expected in that, only two genes were examined. The total number of genes in MLST analysis of *S. thermophilus* strains will be six, and the rest of the analysis will be performed by our colleagues at University of Wisconsin, Madison.

Overall, MLST analysis is a reliable and robust method for typing of yoghurt starter bacteria. Analysis of the *L. delbrueckii* subsp. *bulgaricus* strains using genes β-*gal*, *phe*S and *rpo*A resulted in grouping of strains reflective of isolation source. MLST analysis of *S. thermophilus* strains were incomplete, and number of genes to be analysed will be increased to six. After that, a more through analysis can be depicted for *S. thermophilus*.

CHAPTER 4

CONCLUSION

Yoghurt is an important fermented milk product in Turkey, with high consumption levels. Production of yoghurts by traditional methods at rural areas led to the opinion that Turkey could be good source for starter strains with different and/or improved yoghurt production abilities. The literature surveys about identification of yoghurt starters as well as other LAB rely mainly on phenotypic and biochemical methods. The disadvantages of phenotypic methods were discussed. Moreover, genotypic identification methods of yoghurt bacteria were discussed.

Our study aimed identification of yoghurt starter cultures by PCR-based methods, and further characterization of them at strain level. In this study, identification of yoghurt starters was first performed using methods present in the literature. For this reason isolates were studied with species specific PCR, however, misidentifications were encountered for both yoghurt starters. Later, ARDRA was performed to correctly identify *L. delbrueckii* subsp. *bulgaricus* strains. This method was able to differentiate *L. delbrueckii* subsp. *bulgaricus* and *L. delbrueckii* subsp. *lactis* at subspecies level, however, the method still required preliminary identification of isolates by using phenotypic and biochemical methods. ARDRA with increased number of restriction enzymes could be the other choice, however, considering the high number of isolates from traditional yoghurt samples, this would increase the cost and additional experimental steps would be required. Overall, results of the experiments with species specific PCR revealed ambigious results and ARDRA alone was not able to differentiate yoghurt starters at subspecies level.

Consequently, a quick and reliable method was necessary for selective identification of yoghurt starter bacteria, and a PCR based method was developed for joint identification of them. The method focused on amplification of methionine biosynthesis genes and resulted in selective identification of yoghurt starter bacteria. By using this method, yoghurt starters (*L. delbrueckii* subsp. *bulgaricus* and *S. thermophilus),* together with cheese starters (*L. helveticus*, and *L. delbrueckii* subsp. *lactis*) produced the specific amplicon. Differentiation between these starter bacteria could be achieved as follows; lactobacilli and streptococci could be distinguished by simple microscopic examination, while differentiation among lactobacilli could be achieved using ARDRA with *Eco*RI. The validation of the developed method was done by partial sequencing of 16S rRNA gene.

The characterizations of the isolates at strain level were studied by RAPD and MLST. RAPD analysis with primer 1254 was found better than primer M13, for both of the yoghurt starter bacteria. The bands produced by primer 1254 were brighter, easier to interpret, and a higher number of bands were produced. In addition, clusters produced by primer 1254 were grouped according to the source of isolation, as expected. MLST analyses were performed for the first time in yoghurt starter bacteria, as a robust and reliable method of differentiating strains. MLST analysis of *L. delbrueckii* subsp. *bulgaricus* was performed using three genes β-*gal*, *phe*S and *rpo*A and strains were successfully characterized. The MLST method can be used on higher number of genes, to increase resolution, and can be performed when required. The method was reproducible and data exchange between laboratories would be performed on a reliable platform. In addition to *L. delbrueckii* subsp. *bulgaricus,* MLST analyses were performed for *S. thermophilus* strains, in a joint project. A total of six genes will be studied, analysis of two of the genes were completed, and presented in this thesis, while the remaining four genes will be studied by the other research group.

In conclusion, yoghurt starter bacteria were successfully identified at species, subspecies and strain level using PCR based methods and bioinformatics tools.

REFERENCES

Aanensen D.M. and Spratt B.G. (2005) The multilocus sequence typing network: mlst.net. Nucleic Acids Research 33: W728-733.

Ahmed N., Devi S.M., de los Á Valverde M., Vijayachari P., Machangu R.S., Ellis W.A., and Hartskeerl R.A. (2006) Multilocus sequence typing method for identification and genotypic classification of pathogenic *Leptospira* species. Annals of Clinical Microbiology and Antimicrobials 5: 28-33.

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.

Andrighetto C., De Dea R., Lombardi A., Neviani E., Rossetti L. and Giraffa G. (1998) Molecular identification and cluster analysis of homofermentative thermophilic lactobacilli isolated from dairy products. Research in Microbiology 149: 631-643.

Aquilanti L., Zannini E., Zocchetti A., Osimani A. and Clementi F. (2007) Polyphasic characterization of indigenous lactobacilli and lactococci from PDO Canestrato Pugliese cheese. LWT 40: 1146–1155.

Baker G.C., Smith J.J., and Cowan D.A. (2003) Review and re-analysis of domain-specific 16S primers. Journal of Microbiological Methods 55: 541-555.

Barrangou R., Fremaux C., Deveau H., Richards M., Boyaval P., Moineau S., Romero D.A., and Horvath P. (2007) CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. Science 315: 1709-1712.

Beshkova D., Simova E., Frengova G., and Simov Z. (1998) Production of flavour compounds by yogurt starter cultures. Journal of Industrial Microbiology and Biotechnology 20: 180-186.

BLAST program, 06-12-2008. http://www.ncbi.nlm.nih.gov/BLAST/

Bolotin A., Quinquis B., Renault P., Sorokin A., Ehrlich S. D., Kulakauskas S., Lapidus A., Goltsman E., Mazur M., Pusch G. D., Fonstein M., Overbeek R., Kyprides N., Purnelle B., Prozzi D., Ngui K., Masuy D., Hancy F., Goffeau A., and Hols P. (2004) Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. Nature Biotechnology 22: 1554-1558.

Bolotin A., Quinquis B., Sorokin A., and Dusko Ehrlich, S. (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology 151: 2551-2561

Bonczar G., Wszolek M., and Siuta A. (2002) The effects of certain factors on the properties of yoghurt made from ewe's milk. Food Chemistry 79: 85–91.

Boone D.R., Castenholz R.W., and Garrity G.M. (editors) (2001) Bergey's Manual of Systematic Bacteriology, 2nd edition. Springer, New York.

Bosshard P.P., Abels S., Zbinden R., Böttger E.C., and Altwegg M. (2003) Ribosomal DNA sequencing for identification of aerobic gram-positive rods in the clinical laboratory (An 18-month Evaluation). Journal of Clinical Microbiology 41: 4134–4140

Bouton Y., Guyot P., Beuvier E., Tailliez P., and Grappin R. (2002) Use of PCR-based methods and PFGE for typing and monitoring homofermentative lactobacilli during Comté cheese ripening. International Journal of Food Microbiology 76: 27–38.

Brandt K., Tilsala-Timisjarvi A., and Alatossava T. (2001) Phage-related DNA polymorphism in dairy and probiotic *Lactobacillus*. Micron 32: 59–65.

Brehony C., Jolley K.A., and Maiden M.C.J. (2007) Multilocus sequence typing for global surveillance of meningococcal disease. FEMS Microbiology Reviews 31: 15–26.

Broekhuijsen M., Larsson P., Johansson A., Byström M., Eriksson U., Larsson E., Prior R.G., Sjöstedt A., Titball R.W., and Forsman M. (2003) Genome-wide DNA microarray analysis of *Francisella tularensis* strains demonstrates extensive genetic conservation within the species but identifies regions that are unique to the highly virulent *F. tularensis* subsp. *tularensis*. Journal of Clinical Microbiology 41: 2924–2931.

Bulut C., Gunes H., Okuklu B., Harsa S., Kilic S., Coban H.S., Yenidunya A.F. (2005) Homofermentative lactic acid bacteria of a traditional cheese, Comlek peyniri from Cappadocia region. Journal of Dairy Research 72: 19-24.

Call, D.R., Borucki, M.K., and Loge, F.J. (2003) Detection of bacterial pathogens in environmental samples using DNA microarrays. Journal of Microbiological Methods 53: 235-24.

Catzeddu P., Mura E., Parente E., Sanna M., and Farris G.A. (2006) Molecular characterization of lactic acid bacteria from sourdough breads produced in Sardinia (Italy) and multivariate statistical analyses of results. Systematic and Applied Microbiology 29: 38–144.

Clarridge J.E. (2004) Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. Clinical Microbiology Reviews 17: 840–862.

ClustalW program, 06-12-2008. http://www.ebi.ac.uk/Tools/clustalw2/index.html

Coenye, T., Gevers, D., Van De Peer, Y., Vandamme, P. and Swings, J. (2005) Towards a prokaryotic genomic taxonomy. FEMS Microbiology Reviews 29: 147-167.

Coenye T., and Vandamme, P. (2005) Organisation of the *S10*, *spc* and alpha ribosomal protein gene clusters in prokaryotic genomes. FEMS Microbiology Letters 242: 117-126

Cohan F.M. (2002) What are bacterial species? Annual Reviews of Microbiology 56: 457–487.

Collado C. and Hernandez M.M. (2007) Identification and differentiation of *Lactobacillus, Streptococcus* and *Bifidobacterium* species in fermented milk products with bifidobacteria. Microbiological Research 162: 86-92.

Cooper J. E. and Feil E. J. (2004) Multilocus sequence typing – what is resolved? Trends in Microbiology 12: 373-377.

Coppola S., Fusco V., Andolfi R., Aponte M., Blaiotta G., Ercolini D. and Moschetti G. (2006) Evaluation of microbial diversity during the manufacture of Fior di Latte di Agerola, a traditional raw milk pasta-filata cheese of the Naples area. Journal of Dairy Research 73: 264-272.

De Angelis M., Corsetti A., Tosti N., Rossi, J., Corbo M.R., and Gobbetti M. (2001) Characterization of non-starter lactic acid bacteria from Italian ewe cheeses based on phenotypic, genotypic, and cell wall protein analyses. Applied and Environmental Microbiology 67: 2011–2020.

De Brabandere, A.G and De Baerdemaeker, J.G. (1999) Effects of process conditions on the pH development during yogurt fermentation. Journal of Food Engineering 41: 221-227.

De las Rivas B., Marcobal A., and Munoz R. (2006) Development of a multilocus sequence typing method for analysis of *Lactobacillus plantarum* strains. Microbiology 152: 85–93.

Delfederico L., Hollmann A., Martinez M., Iglesias N.G., De Antoni G., and Semorile, L. (2006) Molecular identification and typing of lactobacilli isolated from kefir grains. Journal of Dairy Research 73: 20-27.

Delley M. and Germond J.-E. 2002 Differentiation of Lactobacillus helveticus, *Lactobacillus delbrueckii* subsp. *bulgaricus*, subsp. *lactis* and subsp. *delbrueckii* using physiological and genetic tools and reclassification of some strains from the ATCC collection system. Applied Microbiology 25 228–231

Desantis, T.Z., Stone, C.E., Murray, S.R., Moberg, J.P., and Andersen, G.L. (2005) Rapid quantification and taxonomic classification of environmental DNA from both prokaryotic and eukaryotic origins using a microarray. FEMS Microbiology Letters 245: 271-278.

Delorme C. (2007) *Streptococcus thermophilus*. International Journal of Food Microbiology doi:10.1016/j.ijfoodmicro.2007.08.014

Doolittle W.F. (1999) Phylogenetic classification and the universal tree. Science 284: 2124-2128.

Du Plessis E. and Dicks L.M.T. (1995) Evaluation of Random Amplified Polymorphic DNA (RAPD)-PCR as a Method to Differentiate *Lactobacillus acidophilus, Lactobacillus crispatus, Lactobacillus amylovorus, Lactobacillus gallinarum, Lactobacillus gasseri,* and *Lactobacillus johnsonii*. Current Microbiology 31: 114-118.

Dubernet S., Desmasures N., and Guéguen M. (2002) A PCR-based method for identification of lactobacilli at genus level. FEMS Microbiology Letters 214: 271-275.

Dubois J.W., Hill S., England L.S., Edge T., Masson L., Trevors J.T., and Brousseau R. (2004) The development of a DNA microarray-based assay for the characterization of commercially formulated microbial products. Journal of Microbiological Methods 58: 251–262.

Ehrmann M., Ludwig W., and Schleifer K.H. (1994) Reverse dot blot hydridization: a useful method for the direct identification of lactic acid bacteria in fermented food. FEMS Microbiology Letters 117: 143–150.

Eisen J.A. (2000) Horizontal gene transfer among microbial genomes: new insights from complete genome analysis. Current Opinion in Genetics & Development 10: 606–611.

Enright M.C. and Spratt B.G. (1999) Multilocus sequence typing. Trends in Microbiology 7: 482-487.

European Molecular Biology Laboratory, 10-01-2008. http://www.ebi.ac.uk/embl/

Fleischmann R. D., Adams M. D., and White O. & 37 other authors (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269: 496–512.

Fogel G.B., Collins C.R., Li J., and Brunk C.F. (1999) Prokaryotic genome size and SSU rDNA copy number: estimation of microbial relative abundance from a mixed population. Microbial Ecology 38: 93–113.

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

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

GenBank Database, 10-01-2008. http://www.ncbi.nlm.nih.gov/

Gevers D., Cohan F.M., Lawrence J.G., Spratt B.G., Coenye T., Feil E.J., Stackebrandt E., Van de Peer Y., Vandamme P., Thompson F.L., and Swings J. (2005) Re-evaluating prokaryotic species. Nature Reviews: Microbiology 3: 733-739.

Gil-Lamaignere C., Roilides E., Hacker J., and Müller F.-M.C. (2003) Molecular typing for fungi - a critical review of the possibilities and limitations of currently and future methods. Clinical Microbiology and Infection 9: 172-185.

Giraffa G., De Vecchi P., and Rossetti L. (1998a) Identification of *Lactobacillus delbrueckii* subspecies *bulgaricus* and subspecies *lactis* dairy isolates by amplified rDNA restriction analysis. Journal of Applied Microbiology 85: 918–924.

Giraffa G., De Vecchi P., Rossi P., Nicastro G., and Fortina M.G. (1998b) Genotypic heterogeneity among *Lactobacillus helveticus* strains isolated from natural cheese starters. Journal of Applied Microbiology 85: 411–416.

Giraffa G., Gatti M., Rossetti L., Senini L., and Neviani E.(2000) Molecular diversity within *L. helveticus* as revealed by genotypic characterization, Applied and Environmental Microbiology 65 1259–1265.

Giraffa, G., Lazzi, C., Gatti, M., Rossetti, L., Mora, D., and Neviani, E. (2003) Molecular typing of *Lactobacillus delbrueckii* of dairy origin by PCR-RFLP of protein-coding genes. International Journal of Food Microbiology 82, 163-172.

Giraffa G. and Rossetti L. (2004) Monitoring of the bacterial composition of dairy starter cultures by RAPD-PCR. FEMS Microbiology Letters 237: 133–138.

Godon J.J., Duthoit F., Delbès C., Millet L., Montel M.C. (2001) Use of molecular fringerprint for the study of complex microbial ecosystem: Application to AOC Salers cheese. Lait 81: 257–262.

Hanage W.P., Kaijalainen T., Herva E., Saukkoriipi A., Syrjanen R., and Spratt B.G. (2005) Using multilocus sequence data to define the *Pneumococcus*. Journal of Bacteriology 187: 6223–6230.

Heilig, H.G.H.J., Zoetendal, E.G., Vaughan, E.E., Marteau, P., Akkermans, A.D.L. and De Vos, W.M. (2002) Molecular diversity of *Lactobacillus* spp. and other lactic acid bacteria in the human intestine as determined by specific amplification of 16S ribosomal DNA. Applied and Environmental Microbiology 68: 114-123.

Hertel C., Ludwig W., Pot B., Kersters K., and Schleifer K.H. (1993) Differentiation of lactobacilli occurring in fermented milk products by using oligonucleotide probes and electrophoretic protein profiles. Systematics in Applied Microbiology 16: 463–467.

Hughes D.B. and D. G. Hoover (1991) Bifidobacteria: Their Potential for use in American Dairy Adjuncts. Food Technology. 45, 4, 74-83.

Jensen, M.A., Webster, J.A. and Strauss, N. 1993 Rapid identification of bacteria on the basis of polymerase chain reaction-amplified ribosomal DNA spacer polymorphisms. Applied and Environmental Microbiology 59: 943–952.

Jolley K.A., Feil E.J., Chan M.S. and Maiden M.C. (2001) Sequence type analysis and recombinational tests (START). Bioinformatics 17:1230-1231.

Kenny O.M., Fitzgerald R.J., O'Cuinn G., Beresford T.P., and Jordan K.N. (2005) Comparative analysis of the autolytic potential of *Lactobacillus helveticus* strains during Cheddar cheese ripening. International Journal of Dairy Technology 58: 207-213.

Kimura K., Mc Cartney, A.L., McConnell M.A., and Tannock G.W. (1997) Analysis of populations of bifidobacteria and lactobacilli and investigation of the immunological responses of their human hosts to the predominant strains. Applied and Environmental Microbiology 63: 3394–3398.

Klaenhammer T., Altermann E., Arigoni F., Bolotin A.S, Breidt F., Broadbent J., Cano R., Chaillou S., Deutscher J., Gasson M., van de Guchte M., Guzzo J., Hartke A., Hawkins T., Hols P., Hutkins R., Kleerebezem M., Kok J., Kuipers O., Lubbers M., Maguin E., McKay L., Mills D., Nauta A., Overbeek R., Pel H., Pridmore D., Saier M., van Sinderen D., Sorokin A., Steele J., O'Sullivan D., de Vos W., Weimer B., Zagorec M. and Siezen R. (2002) Discovering lactic acid bacteria by genomics. Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology 82: 29-58.

Lafontaine D.L.J. and Tollervey D. (2001) Ribosomal RNA; in Encyclopedia of Life Sciences.

Levin R.E. (2003) Application of random amplified polymorphic DNA (RAPD) and pulsed-field gel electrophoresis (PFGE) analysis to *Listeria monocytogenes*: A review of methodology and results. Food Biotechnology 17: 217–236.

Madigan M. T., J. M. Martinko, J. Parker (1997) Biology of microorganisms. Prentice Hall International, Inc.

Maiden M.C.J., Bygraves J.A., Feil E., Morelli G., Russell J.E., Urwin R., Zhang Q., Zhou J., Zurth K., Caugant D.A., Feavers I.M., Achtman M., and Spratt B.G. (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. PNAS USA 95: 3140-3145.

Maiden M.C.J. (2006) Multi Locus Sequence Typing of Bacteria. Annual Reviews in Microbology 60: 561-580.

Marilley L., Casey M.G. (2004) Flavours of cheese products: metabolic pathways, analytical tools and identification of producing strains. International Journal of Food Microbiology 90: 139– 159.

Miteva V., Boudakov I., Ivanova-Stoyancheva G., Marinove B., Mitev V., and Mengaud J. (2001) Differentiation of *Lactobacillus delbrüeckii* subspecies by ribotyping and amplified ribosomal DNA restriction analysis (ARDRA). Journal of Applied Microbiology 90: 909-918.

Mora B., Fortina M.G., Nicastro G., Parini C. and Manachini P.L. (1998) Genotypic characterisation of thermophilic bacilli: a study on new soil isolates and several reference strains. Research in Microbiology 149: 711–722.

Moreno C., Romero J., and Espejo R.T. (2002) Polymorphism in repeated 16S rRNA genes is a common property of type strains and environmental isolates of the genus *Vibrio*. Microbiology 148: 1233–1239.

Mushegian, A.R. and Koonin, E.V. (1996) Gene order is not conserved in bacterial evolution. Trends in Genetics 12: 289–290.

Muyzer, G., De Waal, E.C., and Uitterlinden, A.G. (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Applied and Environmental Microbiology 59: 695-700.

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

National Center for Biotechnology Information, 10-01-2008. http://www.ncbi.nlm.nih.gov/About/tools/index.html

Ninet B., Monod M., Emler S., Pawlowski J., Metral C., Rohner P., Auckenthaler R., and Hirschel B. (1996) Two different 16S rRNA genes in a mycobacterial strain. Journal of Clinical Microbiology 34: 2531–2536.

Nour, M. (1998) 16S-23S and 23S-5S intergenic spacer regions of lactobacilli: Nucleotide sequence, secondary structure and comparative analysis. Research in Microbiology 149: 433-448.

Oligocalc program, 26-06-2007. http://www.basic-northwestern.edu/biotools/oligocalc.html

Perez, T., Albornoz, J. and Dominguez, A. (1998) An evaluation of RAPD fragment reproducibility and nature. Molecular Ecology: 1347-1357.

Prakash O., Verma M., Sharma P., Kumar M., Kumari K., Singh A., Kumari H., Jit S., Gupta S.K., Khanna M., and Lal R. (2007) Polyphasic approach of bacterial classification – An overview of recent advances. Indian Journal of Microbiology 47: 98–108.

Primer 3 program, 26-06-2007. http://www-genome.wi.mit.edu/cgi-bin/primer/primer3

Reischl U., Feldmann K., Naumann L., Gaugler B.J., Ninet B., Hirschel B., and Emler S. (1998) 16S rRNA sequence diversity in *Mycobacterium celatum* strains caused by presence of two different copies of 16S rRNA gene. Journal of Clinical Microbiology 36: 1761–1764.

Reuter G., Klein G., and Goldberg M. (2002) Identification of probiotic cultures in food samples. Food Research International 35: 117–124.

Ribosomal Database Project-II, 10-01-2008. http://rdp.cme.msu.edu/html/

Rodas A.M., Ferrer S., and Pardo I. (2003) 16S-ARDRA, a tool for identification of lactic acid bacteria isolated from grape must and wine system. Applied Microbiology 26: 412–422.

Rodas A.M., Ferrer S. and Pardo I. (2005) Polyphasic study of wine *Lactobacillus* strains: taxonomic implications. International Journal of Systematic and Evolutionary Microbiology 55: 197–207.

Rosetti L. and Giraffa G. (2005) Rapid identification of dairy lactic acid bacteria by M13-generated, RAPD-PCR fingerprint databases. Journal of Microbiological Methods 63: 135-144.

Roy D. and Sirois S. (2000) Molecular differentiation of *Bifidobacterium* species with amplified ribosomal DNA restriction analysis and alignment of short regions of the ldh gene. FEMS Microbiology Letterts 191: 17–24.

Sanchez I., Sesen S. and Palop L.L. (2004) Polyphasic study of the genetic diversity of lactobacilli associated with 'Almagro' eggplants spontaneous fermentation, based on combined numerical analysis of randomly amplified polymorphic DNA and pulsed-field gel electrophoresis patterns. Journal of Applied Microbiology 97: 446–458.

Sanchez I., Sesena S., Poveda J.M., Cabezas L., and Palop L. (2005) Phenotypic and genotypic characterization of lactobacilli isolated from Spanish goat cheeses. International Journal of Food Microbiology 102: 355– 362.

Satokari R.M., Vaughan E.E., Akkermans A.D.L., Saarela M., and De Vos W.M. (2001a) Bifidobacterial diversity in human feces detected by genus-specific PCR and denaturing gradient gel electrophoresis. Applied and Environmental Microbiology 67: 504-513.

Satokari R.M., Vaughan E.E., Akkermans A.D.L., Saarela M., and De Vos W.M. (2001b) Polymerase chain reaction and denaturing gradient gel electrophoresis monitoring of fecal *Bifidobacterium* populations in a prebiotic and probiotic feeding trial. Systematic and Applied Microbiology 24: 227-231.

Satokari R.M., Vaughan E.E., Smidt H., Saarela M., Matto J. and De Vos W.M. (2003) Molecular Approaches for the Detection and Identification of Bifidobacteria and Lactobacilli in the Human Gastrointestinal Tract. Systematic and Applied Microbiology 26: 572-584.

Simpson J.M., McCracken V.J., Gaskins H.R., and Mackie R.I. (2000) Denaturing gradient gel electrophoresis analysis of 16S ribosomal DNA amplicons to monitor changes in fecal bacterial populations of weaning pigs after introduction of *Lactobacillus reuteri* strain MM53. Applied and Environmental Microbiology 66: 4705-4714.

Sghir A., Antonopoulos D., and Mackie R.I. (1998) Design and evaluation of *Lactobacillus* group-specific ribosomal RNA targeted hybridization probe and its application to the study of intestinal microecology in pigs. Systematics in Applied Microbiology 21: 291–296.

Stackebrandt E. and Goebel B.M.. (1994) A place for DNA-DNA re-association and 16S rRNA sequence analysis in the present species definition in bacteriology. International Journal of Systematic Bacteriology 44: 846–849.

Stackebrandt E. (2003) The richness of prokaryotic diversity: there must be a species somewhere. Food Technology and Biotechnology 41: 7–22.

Stiles, M.E. (1996) Biopreservation by lactic acid bacteria. Antonie van Leeuwen-hoek, International Journal of General and Molecular Microbiology 70: 331-345.

Tamura K., Dudley J., Nei M., and Kumar S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24: 1596-1599.

Tilsala-Timisjarvi A. and Alatossava T. (1997) Development of oligonucleotide primers from the 16S-23S rRNA intergenic sequences for identifying different dairy and probiotic lactic acid bacteria by PCR. International Journal of Food Microbiology 35: 49-56.

Tilsala-Timisjarvi A. and Alatossava T. (1998) Strain-specific identification of probiotic *Lactobacillus rhamnosus* with randomly amplified polymorphic DNA-derived PCR primers. Applied and Environmental Microbiology 64: 4816–4819.

Tourova T.P. (2003) Copy number of ribosomal operons in prokaryotes and its effect on phylogenetic analyses. Mikrobiologiya 72: 432-452.

Tynkkynen S., Satokari R., Saarela M., Mattila Sandholm T., and Saxelin M. (1999) Comparison of ribotyping, randomly amplified polymorphic DNA analysis, and pulsedfield gel electrophoresis in typing of *Lactobacillus rhamnosus* and *L. casei* strains. Applied and Environmental Microbiology 65: 3908–3914.

Ueda K., Seki T., Kudo T., Yoshida T., and Kataoka M.. (1999) Two distinct mechanisms cause heterogeneity of 16S rRNA. Journal of Bacteriology 181: 78–82.

Ueno R., Huss V.A.R., Urano N., and Watabe S. (2007) Direct evidence for redundant segmental replacement between multiple 18S rRNA genes in a single *Prototheca* strain. Microbiology 153: 3879–3893.

Urwin R. and Maiden M.C.J. (2003) Multi-locus sequence typing: a tool for global epidemiology. Trends in Microbiology 11: 479-487.

Vandamme P., Pot B., Gillis M., De Vos P., Kersters K., and Swings J. (1996) Polyphasic taxonomy, a consensus approach to bacterial systematics. Microbiology Reviews 60: 407–438.

Van De Guchte, M., Penaud, S., Grimaldi, C., Barbe, V., Bryson, K., Nicolas, P., Robert, C., Oztas, S., Mangenot, S., Couloux, A., Loux, V., Dervyn, R., Bossy, R., Bolotin, A., Batto, J.-M., Walunas, T., Gibrat, J.-F., Bessieres, P., Weissenbach, J., Ehrlich, S.D. and Maguin, E. (2006) The complete genome sequence of *Lactobacillus bulgaricus* reveals extensive and ongoing reductive evolution. Proceedings of the National Academy of Sciences of the United States of America 103: 9274-9279.

Walter J., Hertel, C., Tannock G.W., Lis C.M., Munro K., and Hammes W.P. (2001) Detection of *Lactobacillus, Pediococcus, Leuconostoc*, and *Weissella* Species in Human Feces by Using Group-Specific PCR Primers and Denaturing Gradient Gel Electrophoresis. Applied and Environmental Microbiology 67: 2578-2585.

Wang Y. and Zhang Z. (2000) Comparative sequence analyses reveal frequent occurrence of short segments containing an abnormally high number of non-random base variations in bacterial rRNA genes. Microbiology 146: 2845-2854.

Welsh J., Rampino N., McClelland M. and Perucho, M. (1995) Nucleic acid fingerprinting by PCR-based methods: Applications to problems in aging and mutagenesis. Mutation Research - DNAging Genetic Instability and Aging 338: 215-229.

Williams J.G.K., Kubelik A.R., Livak K.J., Rafalski J.A., and Tingey S.V. (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Research 18, 6531–6535.

Wee Y.-J., Kim J.-N., and Ryu H.-W. (2006) Biotechnological production of lactic acid and its recent applications. Food Technology and Biotechnology 44: 163–172.

Woese C.R. and Fox G.E. (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. PNAS USA 74: 5088-5090.

Wolf Y.I., Rogozin I.B., Kondrashov A.S., and Koonin E.V. (2001) Genome alignment, evolution of prokaryotic genome organization, and prediction of gene function using genomic context. Genome Research 11: 356–372.

Wu L., Thompson D.K., Li G., Hurt R.A., Tiedje J.M., and Zhou J. (2001) Development and evaluation of functional gene arrays for detection of selected genes in the environment. Applied and Environmental Microbiology 67: 5780–5790.

Yap W.H., Zhang Z., and Wang Y. (1999) Distinct types of rRNA operons exist in the genome of the actinomycete *Thermomonospora chromogena* and evidence for horizontal transfer of an entire rRNA operon. Journal of Bacteriology 181: 5201–5209.

Yoke-Kqueen C. and Radu S. (2006) Random amplified polymorphic DNA analysis of genetically modified organisms. Journal of Biotechnology 127: 161–166.

Zeigler D.R. (2003) Gene sequences useful for predicting relatedness of whole genomes in bacteria. International Journal of Systematic and Evolutionary Microbiology 53: 1893–1900.

Zhong W., Millsap K., Bialkowska-Hobrzanska H., and Reid G. (1998) Differentiation of *Lactobacillus* species by molecular typing, Applied and Environmental Microbiology 64: 2418–2423.

Zoetendal E.G., Akkermans A.D.L., and De Vos W.M. (1998) Temperature gradient gel electrophoresis analysis of 16S rRNA from human fecal samples reveals stable and host-specific communities of active bacteria. Applied and Environmental Microbiology 64: 3854-3859.

APPENDIX A

CHEMICALS, ENZYMES AND SUPPLIERS

Agar	Merck, Germany
MRS broth	Merck, Germany
M17 broth	Merck, Germany
NaOH	Merck, Germany
HCl	Merck, Germany
Agarose	Sigma, USA
Chloroform	Merck, Germany
Phenol	Merck, Germany
EDTA (Ethylenediaminetetraacetate)	Merck, Germany
EtBr	Sigma, USA
Ethanol	Merck, Germany
Sodium Acetate	Merck, Germany
Ficol 400	Sigma, USA
Bromphenol Blue	Sigma, USA
Glacial Acetic Acid	Applichem, Germany
Tris HCl	Sigma, USA
Taq DNA Polymerase	Fermentas, Lithuania
Ribonuclease A (RNase)	Fermentas, Lithuania
Proteinase K	Fermentas, Lithuania
DNA ladders	Fermentas, Lithuania
Deoxynucleotide Triphosphates (dNTPs)	Fermentas, Lithuania
*Taq*I	Fermentas, Lithuania
*Eco*RI	Fermentas, Lithuania
*Hpa*II	Fermentas, Lithuania

APPENDIX B

PREPARATIONS OF GROWTH MEDIA, BUFFERS AND SOLUTIONS

1. Agarose Gel 0.8% (w/v)

0.8 g agarose is dissolved in 100 ml 1X TAE buffer and melted in microwave oven.

2. Agarose Gel 1.5% (w/v)

1.5 g agarose is dissolved in 100 ml 1X TAE buffer and melted in microwave oven.

3. EDTA (0.25M, pH 8.0)

93 g of ethylenedinitrilotetraacetic acid disodium salt dihydrate is added to 800 ml of distilled water. The solution is stirred vigorously on a magnetic stirrer while the pH is adjusted to 8.0 with 10N NaOH. The solution is dispensed into aliquots and sterilized by autoclaving.

4. Tris HCl buffer (1M)

39.5 g of Tris HCl is dissolved in 250ml distilled water.

5. TE buffer

10mM TrisHCl and 1mM EDTA is mixed, the pH of the solution is adjusted to 8.0, and sterilized by autoclaving.

6. TAE buffer

2M Tris

1M Acetic Acid

100 mM Na_2EDTA

48. 44g, 11.8g and 7.45g respectively for 200 ml solution were dissolved in dH_2O and pH was adjusted to 8.0. The solution was diluted 50 times before using.

7. Lysozyme (50mg/ml)

7.5 mg of lysozyme is dissolved in 150µl ddH_2O.

8. Ethanol (70%, 100 ml)

70 ml 99.5% ethanol mixed with 30 ml distilled sterile water.

9. MRS Broth (per Liter)

52.2 g MRS broth is dissolved in 1 liter of distilled water and pH is adjusted to 5.4 with HCl and autoclaved. The medium is stored at 4 °C.

10. MRS Agar (per Liter)

To 52.2 g MRS broth, 20 h bacteriological agar is added. Final volume is adjusted to 1 liter with distilled water and pH is adjusted to 5.4 with HCl and autoclaved. The plates are stored at 4 °C.

11. M17 Broth (per Liter)

42 g MRS broth is dissolved in 1 liter of distilled water and pH is adjusted to 5.4 with HCl and autoclaved. The medium is stored at 4 °C.

12. MRS Agar (per Liter)

To 40 g MRS broth, 20 g bacteriological agar is added. Final volume is adjusted to 1 liter with distilled water and pH is adjusted to 5.4 with HCl and autoclaved. The plates are stored at 4 °C.

13. NaAC (3 M, per Liter)

408.3 g of NaAC.3H_2O is dissolved in 800 ml of H_2O. The pH is adjusted to pH5.5 with glacial acetic acid. The volume of the solution is then adjusted to 1 liter with distilled water, and sterilized by autoclaving.

14. NaOH (10 N, 100 ml)

40g of NaOH pellets is added slowly to 80 ml of H_2O. When the pellets are dissolved completely, the volume is adjusted to 1 liter with H_2O. The solution is stored at room temperature.

15. Loading Dye for Electrophoresis

0.25 % Bromophenol Blue

0.25 % Xylene cyanol

15 % Ficol 400

40% (w/v) Sucrose

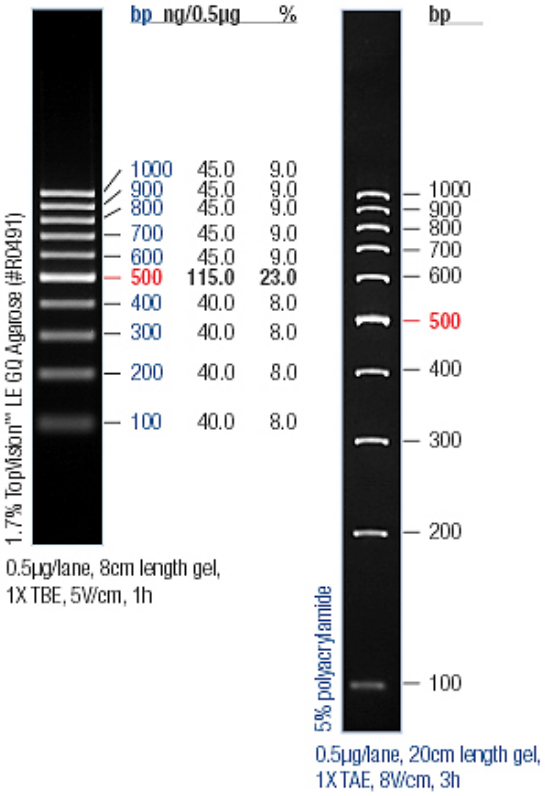
Dissolved in dH_2O.

16. Ethidium Bromide Solution

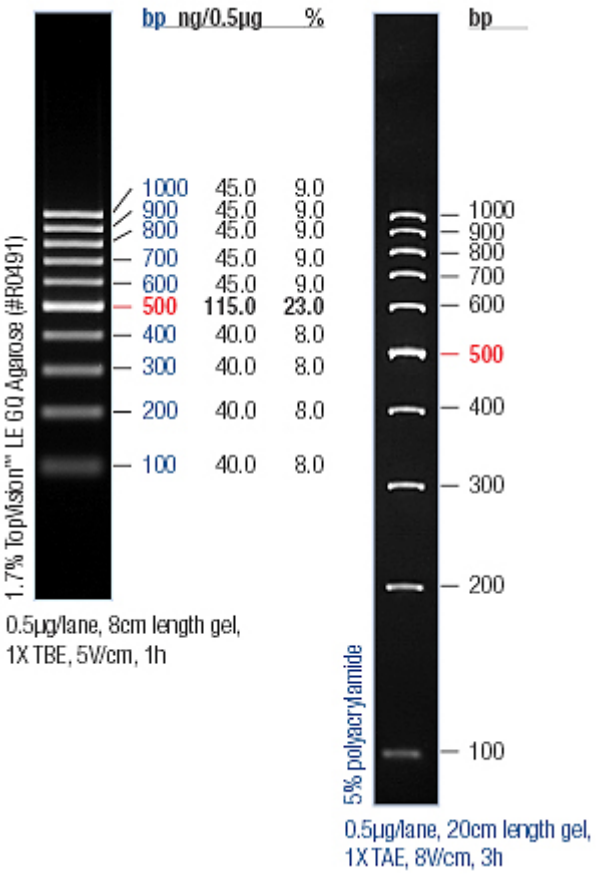
10 mg/ml EtBr was prepared with dH_2O.

APPENDIX C

DNA LADDERS



100bp DNA ladder (Fermentas)



100bp DNA ladder plus (Fermentas)

APPENDIX D

MULTIPLE SEQUENCE ALIGNMENT OF METHIONINE BIOSYNTHESIS GENES

CLUSTAL W multiple sequence alignment for methionine biosynthesis genes (cysteine synthase and cystathionine beta-lyase) of *L. delbrueckii* subsp. *bulgaricus* ATCC11842, *L. helveticus* CNRZ32, *S. thermophilus* LMG18311, *L. plantarum* WCFS1, *Leuconostoc mesenteroides* subsp. *mesenteroides* ATCC8293, *Oenococcus oeni* PSU-1, *Enterococcus faecium* DO, *Enterococcus faecalis* V583 (primer binding sites of cysmet2 are highlighted).

```
bulgaricus      TCAAACCATCGTTGTGGAACCTGAAGGCTCAATTCTAAATGGTGGGCCTGCCCATCCCCA 647
helveticus      TCAAACCATCGTTGTGGAACCTGAAGGCTCAATTCTAAATGGTGGGCCTGCCCATCCCCA 647
thermophilus    TCAAACCATCGTTGTGGAACCTGAAGGCTCAATTCTAAATGGTGGGCCTGCCCATCCCCA 647
plantarum       CAAAGCCGTAGTCGTTGAGCCAGAAGGTTCCATTTTAAATGGTGGACCAGCTCACGCACA 644
mesenteroides   AAAAAATATTGTTGTAGAACCCGAAGGATCTATTTTAAATGGTGGCCCAGCACATGCTCA 647
oeni            GCAGGCCGTGACTGTCGAACCGGAAGGTTCAATTCTAAACGGGGGACCAGAACATCCCCA 647
faecium         TCGAATAATCGGTGTAGAACCAGAGGGCTCTGTGTTAAATGGTGGTGATCCAGCCCCTCA 653
faecalis        GCAATTATGGGGCGTGGAGCCAGAAGGTTCAATTTTAAATGGCGGACCTGCACACGGACA 653
                              ** ** ** ** ** **  *  **** ** **           **

bulgaricus      TCGAACTGAAGGCATCGGTGTTGAATTTGTCCCACCATTTTTTAAAGATGTTCAAATCGA 707
helveticus      TCGGACTGAAGGCATCGGTGTTGAATTTGTCCCACCATTTTTTAAAGATGTTCAAATCGA 707
thermophilus    TCGGACTGAAGGCATCGGTGTTGAATTTGTCCCACCATTTTTTAAAGATGTTCAAATCGA 707
plantarum       TCGTACCGAGGGAATCGGCGTCGAATTCATCCCACCATTTTTCGATCAGGTTCGCATCGA 704
mesenteroides   CCGTACAGAAGGTATTGGTGTTGAATTTGTACCACCATTCTTCGAATCAATTCACATTGA 707
oeni            CCGAATCGAGGGAATCGGCGTTGAATTTGTTCCGCCTTTCTTCGATCAGATTCGTGTGGA 707
faecium         CGAAATTGAGGGTATCGGTGTAGAATTCATTCCGCCGTTTCTTTCTCCCCTTTCAATCAA 713
faecalis        TGAAATCGAAGGCATTGGCGTGGAGTTCGTGCCGCCTTTTCTTAAACAGATTCCTGTGGA 713
                    *  ** ** ** ** ** ** **  * ** ** **  *        **    *  *

bulgaricus      CAAAACGTTAACAATTGCCGATGCCGACGCCTTTCATCAAGTTAAGGAGCTTGCCAAGAA 767
helveticus      CAAAACGTTAACAATTTCCGATGCCGACGCCTTTCATCAAGTTAAGGAGCTTGCCAAGAA 767
thermophilus    CAAAACGTTAACAATTGCCGATGCCGACGCCTTTCATCAAGTTAAGGAGCTTGCCAAGAA 767
plantarum       TCAGACACTGACCATCGCGGACAATGATGCCTTTGCCCAAGTTCGACACCTAGCCCGTGA 764
mesenteroides   CAAAACATTAACTATTAGCGATGACAATGCCTTTCAGCAAGTTCGATATGCTGCCGAGAA 767
oeni            CCAGGTTAAAACAATCAGTGACGATCGGGCTTTTGAATATGTTAAATGGTTGGCTGCGCA 767
faecium         TCAGATCGAGACAATCTCTGATGTAGAAGGGTTTAACTATACTAGGCAGTTAGCACGAGA 773
faecalis        CGGCTTTTATACCATTTCAGACGAGGATGGTTTTTTCTGGGTCAAACAATTAGCCAAGAA 773
                          ** **    **       *  ***                 **     *
```

```
bulgaricus      GCAGGGATTATTCATTGGCAGTTCAAGCGGGGCTGCCTTGGCTGCCAGCTTAAAGGTTGC 827
helveticus      GCAGGGATTATTCATTGGCAGTTCAAGCGGGGCTGCTTTGGCTGCCAGCTTAAAGGTTGC 827
thermophilus    GCAGGGATTATTCATTGGCAGTTCAAGCGGGGCTGCTTTGGCTGCCAGCTTAAAGGTTGC 827
plantarum       CCACGGTTTACTGATTGGCAGTTCTAGCGGCGCTGCGCTAGCCGCTAGTTTACAACTTGC 824
mesenteroides   CCTAGGACTTTTTATCGGTAGTTCAAGTGGCGCGGCCTTGGCTGCTAGCCTACAGATGGC 827
oeni            CGTCGGCCTGTTCGCCGGCAGTTCAAGTGGCGCTGCTTTAGCAGCAAGTTTACAGCTGGC 827
faecium         ACAAGGACTATTAGTTGGAAGTTCGAGTGGTGCGGCTTTTGCAGCAGCATTAAGAGAAAT 833
faecalis        AAGCGCTTTGCTGGTGGGAAGCTCCAGCGGCGCAGCATTTGCGGCGGCATTAAGAGAGGC 833
                    *   *  *    ** ** ** ** ** ** **  * ** **     **

bulgaricus      CAAGGCACTTCCCGAGAATAGTACCATTGTCACAATTTTTCCAGATAGCAGTGAACGTTA 887
helveticus      CAAGGCACTTCCCGAGAATAGTACCATTGTCACAATTTTTCCAGATAGCAGTGAACGTTA 887
thermophilus    CAAGGCACTTCCCGAGAATAGTACCATTGTCACAATTTTTCCAGATAGCAGTGAACGTTA 887
plantarum       CACCAACCTGCCTGCCAATAGTCATATCGTGACCATCTTTCCAGACAGTAGTGAGCGTTA 884
mesenteroides   AGAAACTTTGCCGCCACAAAGTAATATTGTCACGATATTCCCAGACAGCAGCGAACGTTA 887
oeni            AAAGGAGTTGCCGTCCGATTCAACCGTTGTGACTGTTTTTCCGGATTCAAGCGAACGTTA 887
faecium         TCGTCGTTTACCACCTGGTCATCGTGTCGTTACGATTTTTCCAGACGCAGCAGACCGCTA 893
faecalis        GCGCCGCTTACCAGCAGGCAGCACAATTGTAACTATTTTTCCTGACGGTAGTGATCGTTA 893
                       * **             * ** **  * ** ** **       ** ** **

bulgaricus      CATGAGTGAAAATATTTATGATTAAAAATAAAGTGAGGTTATTAAAATGGAATTTGATAC 947
helveticus      CATGAGTGAAAATATTTATAATTAAA--------------------TGGAATTTGATAC 926
thermophilus    CATGAGTGAAAATATTTATGATTAAA--------------------TGGAATTTGATAC 926
plantarum       TCTGAGCCAAAAAATCTAT-ACGAAATG-----------------AATGAAATTTGAAAC 926
mesenteroides   CATGAGCACAAACATTTAT--GGAAACT----------------GAATGAAATTTGATAC 929
oeni            TTTAAGTAAAGGTATCTACGACTAAA--------------------TGAAATTCAATAC 926
faecium         TTTATCAAAAAACATTTACTTATAGA--------------------TGCACATTCAAAC 932
faecalis        TTTATCAAAAAATATTTATTCTTAGA--------------------TGAAATTCAATAC 932
                  *      *   ** **     * *                     ** *  *  * **

bulgaricus      AAAATTAATTCATGGTGGTATTAGCGAAGATAAAGCAACCGGAGCAGTTTCGGTGCCGAT 1007
helveticus      AAAATTAATTCATGGTGGTATTAGCGAAGATAAAGCAACCGGAGCAGTTTCGGTGCCGAT 986
thermophilus    AAAATTAATTCATGGTGGTATTAGCGAAGATAAAGCAACCGGAGCAGTTTCGGTGCCGAT 986
plantarum       CCAATTAATTCACGGTGGTATCAGTGAGGATGCCACTACTGGCGCGACTTCGGTACCCAT 986
mesenteroides   ACAACTTATTCATGGTGGCATTAGTGTTGACCAATCAACTGGCGCCGTATCTGTCCCTAT 989
oeni            AAAACTTATTCATGGCGGTATTAGCGAAGATTCATCAACCGGGGCAGTTTCAATCCCTAT 986
faecium         AAAATTGATCCACGGAGGCATCAGTGAAGATCCCACAACAGGAGCTGTCAGTGTACCTAT 992
faecalis        CAAATTAATTCATGGCGGCATTAGTAAAGACACAACAACAGGTGCAGTGAGTGTTCCTAT 992
                  ** * ** ** ** ** ** **    **     * ** ** **        * ** **

bulgaricus      TTATATGGCATCAACTTTTCATCAGCAAAAAATCGGTGAAAAT---CAATATGAGTATTC 1064
helveticus      TTATATGGCATCAACTTTTCATCAGCAAAAAATCGGTGAAAAT---CAATATGAGTATTC 1043
thermophilus    TTATATGGCATCAACTTTTCATCAGCAAAAAATTGGTGAAAAT---CAATATGAGTATTC 1043
plantarum       CTACATGGCCTCGACCTTCCGCCAAACAAAAATCGGTCAAAAT---CAATACGAATATTC 1043
mesenteroides   TCATATGGCTTCAACCTTTAAGCAAACTAAAATTGGCGAGGCA---AAATATGAATATTC 1046
oeni            CTATCGTTCTTCGACTTTTCATCAAAACAAGGTCGCTGGAAATGCAAAGTGGGAATACGG 1046
faecium         TTATCAAACATCCACTTATCGGCAAGATGGCGTCGGCCAGCCTAAACAATATGAATATTC 1052
faecalis        TTATCAAACGTCAACCTTTGAACAAAATGGCGTGGGCCAGCCAAAAGAATATGAATATTC 1052
                  *     * ** ** *     **        * *            * *  ** **

bulgaricus      TCGTTCTGGAAACCCAACTCGAGAAGCAGTTGAAAAACTAATTGCCGAGTTGGAAGGCGG 1124
helveticus      TCGTTCTGGAAACCCAACTCGAGAAGCAGTTGAAAAACTAATTGCCGAGTTGGAAGTCGG 1103
thermophilus    TCGTTCTGGAAACCCAACTCGAGAAGCAGTTGAAAAACTAATTGCCGAGTTGGAAGGCGG 1103
plantarum       ACGGACGGGAAATCCAACCCGGGCCGCCGTCGAAGCATTAATTGCCACCCTCGAACATGG 1103
mesenteroides   AAGATCTGGCAACCCAACCCGTGAAGCCGTAGAAAGCTTAATTGCAGACTTAGAAAATGG 1106
oeni            GCGCAGTGGAAATCCAACCCGTGCGGCTTTGGAAAAACTGATTGCCGATTTAGAAGAAGG 1106
faecium         AAGATCTGGTAATCCTACTCGTTTTGCATTGGAAGAACTGATTGCTGATTTGGAAGGCGG 1112
faecalis        TCGTTCAGGTAATCCCACACGTCACGCGTTAGAAACCTTAATCGCTGAGTTAGAAGGTGG 1112
                  *    ** ** ** ** **    **  * ***    * ** **     * ***   **

bulgaricus      CACAGCCGGGTTTGCATTTGCTTCCGGTTCAGCTGCAATCGACACAGTCTTCTCAATGTT 1184
helveticus      CACAGCCGGGTTTGCATTTGCTTCCGGTTCAGCTGCAATTGACACAGTCTTCTCAATGTT 1163
thermophilus    CACAGCCGGGTTTGCATTTGCTTCCGGTTCAGCTGCAATCTACACAGTCTTCTCAATGTT 1163
plantarum       CAGCGCTGGCTTCGCATTTGCTTCTGGCTCCGCTGCCATTAATACCGTCTTCTCACTATT 1163
mesenteroides   TACTGCTGGCTTTGCTTTTGCATCAGGATCTGCGGCGATAAGTACTATTTTTTCACTATT 1166
oeni            GAAAGCCGGTTTTGCTTTTGCCTCCGGTTCGGCGGCGATTCATGCGGTTTTTTCATTGTT 1166
faecium         TGTTCGAGGATTCGCTTTTTCTTCTGGTCTTTCTGGGATTCATGCCGTTTTTTCTCTTTT 1172
faecalis        GAGCCATGGCTTTGCGTTTAGCTCAGGCTTAGCTGGTATCCATGCAATCATTTCTATGTT 1172
                       ** ** ** ***   ** **     * *  **     *  *  * **  * **
```

```
bulgaricus      CTCCGCCGGCGACCATTTCGTCATTGGCAATGATGTCTACGGTGGTACCTTCCGATTGAT 1244
helveticus      CTCCGCCGGCGACCATTTCGTCATTGGCAATGATGTCTACGGTGGTACCTTCCGATTGAT 1223
thermophilus    CTCCGCCGGCGACCATTTCGTCATTGGCAATGATGTCTACGGTGGTACCTTCCGATTGAT 1223
plantarum       CTCGGCTGGTGATCACATTATTGTGGGAAATGATGTCTACGGTGGCACCTTCCGCTTGAT 1223
mesenteroides   TTCATCTGGGGATCACATTATTGTTGGAAACGATGTTTATGGTGGTACATTTAGGCTAAT 1226
oeni            CTCTTCCGGTGATCACATTGTTGTTGGCGATGATGTTTACGGAGGTACTTTCCGTTTGAT 1226
faecium         TCAAGCTGGCGATCATATCTTATTAGGAGATGATGTCTATGGAGGAACATTCCGTCTATT 1232
faecalis        TGGACCCGAAGATCATATTTTATTAGGTGACGATGTCTATGGGGGGAGTTTCCGTTTATT 1232
                     * *  ** **  *  *  * **  * ***** ** ** ** *  **  *  *  *

bulgaricus      TGACGCCGTTTTAAAACGCTTTGGCATGACCTTTACCGTGGTTGACACCCGCGACCTGGC 1304
helveticus      TGACGCCGTTTTAAAACGCTTTGGCATGACCTTTACCGTGGTTGACACCCGCGACCTGGC 1283
thermophilus    TGACGCCGTTTTAAAACGCTTTGGCATGACCTTTACCGTGGTTGACACCCGCGACCTGGC 1283
plantarum       CGACGCCGTTTTGAAACACTTTGGCATGACTTTTACAGCCGTAGATACGCGTGACTTGGC 1283
mesenteroides   TGACAATGTCCTAAAAAGAACAGGTCAGACATTTACAATTGTCGATACCCGTGATTTATC 1286
oeni            AGATCAGGTGTTAAAACGCTTCGGTTTGGAATTTACTGTTGTTGATACCCGAGATCTGTC 1286
faecium         TGACAAAGTTCTTACTAAAAATGGTCTCGAGTACACGATCATAGATACGAGTAATCTTGA 1292
faecalis        GGATAAAGTATTTGTAGCAAATGGGTTAAGCTACACGATTGTAAACGCCAGCGACTTATC 1292
                 **    **  *          **       *  **     *  *  *  *  *  *
```

APPENDIX E

16S rRNA GENE SEQUENCE OF REPRESENTATIVE ISOLATES

REPRESENTATIVE *L. delbrueckii* subsp. *bulgaricus*

>K2-1-U926

NNNCCCCGTGCGCTCCAGCGGACGCTTATGCGTTTGCTGCGGCACTGAGGACCGGAAA
GTCCCCAACACCTAGCGCTCATCGTTTACGGCATGGACTACCAGGGTATCTAATCCTGT
TCGCTACCCATGCTTTCGAGCCTCAGCGTCAGTTGCAGACCAGAGAGCCGCCTTCGCC
ACTGGTGTTCTTCCATATATCTACGCATTCCACCGCTACACATGGAATTCCACTCTCCT
CTTCTGCACTCAAGAATGACAGTTTCCGATGCAGTTCCACGGTTGAGCCGTGGGCTTTC
ACATCAGACTTATCATTCCGCCTGCGCTCGCTTTACGCCCAATAAATCCGGACAACGCT
TGCCACCTACGTATTACCGCGGCTGCTGGCACGTAGTTAGCCGTGACTTTCTGGTTGAT
TACCGTCAAATAAAGACCAGTTACTGCCTCTATCCTTCTTCACCAACAACAGAGCTTTA
CGATCCGAAAACCTTCTTCACTCACGCGGCGTTGCTCCATCAGACTTGCGTCCATTGTG
GAAGATTCCCTACTGCTGCCTCCCGTAGGAGTTTGGGCCGTGTCTCAGTCCCAATGTGG
CCGATCAGTCTCTCAACTCGGCTACGCATCATTGCCTTGGTAGGCCTTTACCCCACCAA
CTAGCTAATGCGCCGCGGGCTCATCCTAAAGTGACAGCTTACGCCGCCTTTCAAACTT
GAATCATGCGATTCATGTTGTTATCCGGTATTAGCACCTGTTTCCAAGTGGTATCCCAG
TCTTTAGGGCAGATTGCCCACGTGTTACTCACCCATCCGCCGCTAGCGTCCAACAAATC
ATCCCGAAGGAATCTTTGAATTCAGCTCGCT

REPRESENTATIVE *S. thermophilus*

>K1-30-U926

NTTTTCCGATGCGCTCCAGAGGANTGCTTATGCGTTAGCTGCGGCACTGAATCCCGGA
AAGGATCCAACACCTAGCACTGCATCGTTTACGGCGTGGACTACCAGGGATATCTAAT
CCAGCTTCGGTCCCCACGCATTTCGAGCCTCAGCGTCAGTTACAGACCAGAGAGCCGT
CATTTCGCCACCGGTGTTCCTCCATATATCTACGCATTTCACCGCTACACATGGAATTC
CACTCTCCCCTTCTGCACTCAAGTTTGACAGTTTCCAAAGCGAACTATGGTTGAGCCAC
AGCCTTTAACTTCAGACTTATCAAACCGCCTGCGCTCGCTTTACGCCCAATAAATCCGG
ACAACGCTCGGGACCTACGTATTACCGCGGCTGCTGGCACGTAGTTAGCCGTCCCTTTC
TGGTAAGCTACCGTCACAGTGTGAACTTTCCACTCTCACACCCGTTCTTGACTTACAAC
AGAGCTTTACGATCCGAAAACCTTCTTCACTCACGCGGCGTTGCTCGGTCAGGGTTGCC
CCCATTGCCGAAGATTCCCTACTGCTGCCTCCCGTAGGAGTCTGGGCCGTGTCTCAGTC
CCAGTGTGGCCGATCACCCTCTCAGGTCGGCTATGTATCGTCGCCTAGGTGAGCCATTA
CCTCACCTACTAGCTAATACAACGCAGGTCCATCTTGTAGTGGAGCAATTGCCCCTTTC
AAATAAATGACATGTGTCATCCATTGTTATGCGGTATTAGCTATCGTTTCCAATAGTTA
TCCCCCGCTACAGGGCAGGTTACCTACACGTTACTCACCCGTTCGCAACTCATCCAAG
AAGAGCAAGCTCCTCTCTTCAGCGTTCTAC

PRESUMPTIVE *S. thermophilus*

>N5-4-U926

NNNCCATGCGACCCCAGCGGGTGCTTAATGCGTTAGCTGCGGCACTAAGCCCCGGAAA
GGGCCTAACACCTAGCACTCATCGTTTACGGCGTGGACTACCAGGGTATCTAATCCTG
TTTGCTCCCCACGCTTTCGAGCCTCAGCGTCAGTTACAGACCAGAGAGCCGCTTTCGCC
ACCGGTGTTCCTCCATATATCTACGCATTTCACCGCTACACATGGAATTCCACTCTCCC
CTTCTGCACTCAAGTCTAACAGTTTCCAAAGCGAACAATGGTTAAGCCACTGCCTTTAA
CTTCAGACTTATTAAACCGCCTGCGCTCGCTTTACGCCCAATAAATCCGGACAACGCTC
GGGACCTACGTATTACCGCGGCTGCTGGCACGTAGTTAGCCGTCCCTTTCTGGTAAGTT
ACCGTCACTGTGTGAACTTTCCACTCTCACACACGTTCTTCTCTTACAACAGAGCTTTA
CGATCCGAAAACCTTCTTCACTCACGCGGCGTTGCTCGGTCAGGGTTGCCCCCATTGCC
GAAGATTCCCTACTGCTGCCTCCCGTAGGAGTCTGGGCCGTGTCTCAGTCCCAGTGTGG
CCGATCACCCTCTCAGGTCGGCTATGTATCGTCGCCTTGGTGAGCCGTTACCTCACCAA
CTAGCTAATACAACGCAGGTCCATCTACTAGTGAAGCAATTGCTCCTTTCAAGCATCTA
ACATGGGGTAAATGCTGTTATGCGGTATTAGCTATCGTTTCCAATAGATATCCCCCGCT
AGTAGGCAGGTTACCTACGCGTTACTCACCCGTTCGCAACTCTTCCAACTTTAGCAAGG
TAAGGTTTTCAGCGTTCTACTTGCAT

APPENDIX F

BLAST ANALYSIS OF REPRESENTATIVE STARTER STRAINS

Blast analysis of representative *L. delbrueckii* subsp. *bulgaricus*

```
>gb|CP000412.1| L. delbrueckii subsp. bulgaricus ATCC BAA-365, complete genome
Length=1856951

Features in this part of subject sequence:
   rRNA-16S ribosomal RNA

 Score = 1502 bits (813),  Expect = 0.0
 Identities = 831/839 (99%), Gaps = 3/839 (0%)
 Strand=Plus/Minus

Query: strain K2-1
Subject: L. delbrueckii subsp. bulgaricus ATCC BAA-365

Query  4       CCA-GCGGA-CGCTT-ATGCGTTTGCTGCGGCACTGAGGACCGGAAAGTCCCCAACACCC  60
               ||| ||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||
Sbjct  684166  CCAGGCGGAGCGCTTAATGCGTTTGCTGCGGCACTGAGGACCGGAAAGTCCCCAACACCT  684107

Query  61      ACCGCTCATCGTTTACGGCATGGACTACCAGGGTATCTAATCCTGTTCGCTACCCATGCT  120
               | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  684106  AGCGCTCATCGTTTACGGCATGGACTACCAGGGTATCTAATCCTGTTCGCTACCCATGCT  684047

Query  121     TTCGAGCCTCAGCGTCAGTTGCAGACCAGAGAGCCGCCTTCGCCACTGGTGTTCTTCCAT  180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  684046  TTCGAGCCTCAGCGTCAGTTGCAGACCAGAGAGCCGCCTTCGCCACTGGTGTTCTTCCAT  683987

Query  181     ATATCTACGCATTCCACCGCTACACATGGAATTCCACTCTCCTCTTCTGCACTCAAGAAT  240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  683986  ATATCTACGCATTCCACCGCTACACATGGAATTCCACTCTCCTCTTCTGCACTCAAGAAT  683927

Query  241     GACAGTTTCCGATGCAGTTCCACGGTTGAGCCGTGGGCTTTCACATCAGACTTATCATTC  300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  683926  GACAGTTTCCGATGCAGTTCCACGGTTGAGCCGTGGGCTTTCACATCAGACTTATCATTC  683867

Query  301     CGCCTGCGCTCGCTTTACGCCCAATAAATCCGGACAACGCTTGCCACCTACGTATTACCG  360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  683866  CGCCTGCGCTCGCTTTACGCCCAATAAATCCGGACAACGCTTGCCACCTACGTATTACCG  683807

Query  361     CGGCTGCTGGCACGTAGTTAGCCGTGACTTTCTGGTTGATTACCGTCAAATAAAGACCAG  420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  683806  CGGCTGCTGGCACGTAGTTAGCCGTGACTTTCTGGTTGATTACCGTCAAATAAAGACCAG  683747

Query  421     TTACTGCCTCTATCCTTCTTCACCAACAACAGAGCTTTACGATCCGAAAACCTTCTTCAC  480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  683746  TTACTGCCTCTATCCTTCTTCACCAACAACAGAGCTTTACGATCCGAAAACCTTCTTCAC  683687

Query  481     TCACGCGGCGTTGCTCCATCAGACTTGCGTCCATTGTGGAAGATTCCCTACTGCTGCCTC  540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  683686  TCACGCGGCGTTGCTCCATCAGACTTGCGTCCATTGTGGAAGATTCCCTACTGCTGCCTC  683627
```

```
Query  541     CCGTAGGAGTTTGGGCCGTGTCTCAGTCCCAATGTGGCCGATCAGTCTCTCAACTCGGCT  600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  683626  CCGTAGGAGTTTGGGCCGTGTCTCAGTCCCAATGTGGCCGATCAGTCTCTCAACTCGGCT  683567

Query  601     ACGCATCATTGCCTTGGTAGGCCTTTACCCCACCAACTAGCTAATGCGCCGCGGGCTCAT  660
               |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Sbjct  683566  ACGCATCATTGCCTTGGCAGGCCTTTACCCCACCAACTAGCTAATGCGCCGCGGGCTCAT  683507

Query  661     CCTAAAGTGACAGCTTACGCCGCCTTTCAAACTTGAATCATGCGATTCATGTTGTTATCC  720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  683506  CCTAAAGTGACAGCTTACGCCGCCTTTCAAACTTGAATCATGCGATTCATGTTGTTATCC  683447

Query  721     GGTATTAGCACCTGTTTCCAAGTGGTATCCCAGTCTTTAGGGCAGATTGCCCACGTGTTA  780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  683446  GGTATTAGCACCTGTTTCCAAGTGGTATCCCAGTCTTTAGGGCAGATTGCCCACGTGTTA  683387

Query  781     CTCACCCATCCGCCGCTAGCGTCCAACAAATCATCTCGAACGAATCTTTGAATTCAGCT  839
               |||||||||||||||||||||||||||||||||| |||| ||||||||||||||||||
Sbjct  683386  CTCACCCATCCGCCGCTAGCGTCCAACAAATCATCCCGAAGGAATCTTTGAATTCAGCT  683328
```

Blast analysis of strain LB340-2

>gb|CP000033.3| *L. acidophilus* NCFM, complete genome
Length=1993560

Features in this part of subject sequence: rRNA-16S ribosomal RNA

Score = 1509 bits (817), Expect = 0.0
Identities = 822/824 (99%), Gaps = 1/824 (0%)
Strand=Plus/Minus

Query: strain LB340-2
Subject: *L. acidophilus* NCFM

```
Query  1      CGTT-GCTGCAGCACTGAGAGGCGGAAACCTCCCAACACTTAGCACTCATCGTTTACGGC  59
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  60143  CGTTAGCTGCAGCACTGAGAGGCGGAAACCTCCCAACACTTAGCACTCATCGTTTACGGC  60084

Query  60     ATGGACTACCAGGGTATCTAATCCTGTTCGCTACCCATGCTTTCGAGCCTCAGCGTCAGT  119
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  60083  ATGGACTACCAGGGTATCTAATCCTGTTCGCTACCCATGCTTTCGAGCCTCAGCGTCAGT  60024

Query  120    TGCAGACCAGAGAGCCGCCTTCGCCACTGGTGTTCTTCCATATATCTACGCATTCCACCG  179
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  60023  TGCAGACCAGAGAGCCGCCTTCGCCACTGGTGTTCTTCCATATATCTACGCATTCCACCG  59964

Query  180    CTACACATGGAGTTCCACTCTCCTCTTCTGCACTCAAGAAAAACAGTTTCCGATGCAGTT  239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  59963  CTACACATGGAGTTCCACTCTCCTCTTCTGCACTCAAGAAAAACAGTTTCCGATGCAGTT  59904

Query  240    CCTCGGTTAAGCCGAGGGCTTTCACATCAGACTTATTCTTCCGCCTGCGCTCGCTTTACG  299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  59903  CCTCGGTTAAGCCGAGGGCTTTCACATCAGACTTATTCTTCCGCCTGCGCTCGCTTTACG  59844

Query  300    CCCAATAAATCCGGACAACGCTTGCCACCTACGTATTACCGCGGCTGCTGGCACGTAGTT  359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  59843  CCCAATAAATCCGGACAACGCTTGCCACCTACGTATTACCGCGGCTGCTGGCACGTAGTT  59784

Query  360    AGCCGTGACTTTCTGGTTGATTACCGTCAAATAAAGGCCAGTTACTACCTCTATCCTTCT  419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  59783  AGCCGTGACTTTCTGGTTGATTACCGTCAAATAAAGGCCAGTTACTACCTCTATCCTTCT  59724

Query  420    TCACCAACAACAGAGCTTTACGATCCGAAAACCTTCTTCACTCACGCGGCGTTGCTCCAT  479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  59723  TCACCAACAACAGAGCTTTACGATCCGAAAACCTTCTTCACTCACGCGGCGTTGCTCCAT  59664

Query  480    CAGACTTTCGTCCATTGTGGAAGATTCCCTACTGCTGCCTCCCGTAGGAGTTTGGGCCGT  539
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  59663  CAGACTTTCGTCCATTGTGGAAGATTCCCTACTGCTGCCTCCCGTAGGAGTTTGGGCCGT  59604

Query  540    GTCTCAGTCCCAATGTGGCCGATCAGTCTCTCAACTCGGCTATGCATCATTGCCTTGGTA  599
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  59603  GTCTCAGTCCCAATGTGGCCGATCAGTCTCTCAACTCGGCTATGCATCATTGCCTTGGTA  59544

Query  600    GGCCGTTACCCTACCAACTAGCTAATGCACCGCGGGGCCATCCCATAGCGACAGCTTACG  659
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  59543  GGCCGTTACCCTACCAACTAGCTAATGCACCGCGGGGCCATCCCATAGCGACAGCTTACG  59484

Query  660    CCGCCTTTTATAAGCTGATCATGCGATCTGCTTTCTTATCCGGTATTAGCACCTGTTTCC  719
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  59483  CCGCCTTTTATAAGCTGATCATGCGATCTGCTTTCTTATCCGGTATTAGCACCTGTTTCC  59424

Query  720    AAGTGGTATCCCAGACTATGGGGCAGGTTCCCCACGTGTTACTCACCCATCCGCCGCTCG  779
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  59423  AAGTGGTATCCCAGACTATGGGGCAGGTTCCCCACGTGTTACTCACCCATCCGCCGCTCG  59364

Query  780    CGTTCCCAACGTCATCACCCAAGTGAATCTGTTGGTTCAGCTCG  823
              |||||||||||||||||| |||||||||||||||||||||||||
Sbjct  59363  CGTTCCCAACGTCATCACCGAAGTGAATCTGTTGGTTCAGCTCG  59320
```

Blast analysis of representative *S. thermophilus*

```
>gb|CP000419.1| S. thermophilus LMD-9, complete genome
Length=1856368

Features in this part of subject sequence: rRNA-16S ribosomal RNA

 Score = 1491 bits (807),  Expect = 0.0
 Identities = 822/829 (99%), Gaps = 3/829 (0%)
 Strand=Plus/Minus

Query: strain K1-30
Subject: S. thermophilus LMD-9

Query  21     GGANTGCTT-ATGCGTTAGCTGCGGCACTGAATCCCGGAAAGGATCCAACACCTAGCACT  79
              ||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  67693  GGAGTGCTTAATGCGTTAGCTGCGGCACTGAATCCCGGAAAGGATCCAACACCTAGCACT  67634

Query  80     GCATCGTTTACGGCGTGGACTACCAGGGATATCTAATCCTGTTCGGTCCCCACGCTTTCG  139
               |||||||||||||||||||||||||| ||||||||||||||| ||||||||||||||
Sbjct  67633  -CATCGTTTACGGCGTGGACTACCAGGG-TATCTAATCCTGTTCGCTCCCCACGCTTTCG  67576

Query  140    AGCCTCAGCGTCAGTTACAGACCAGAGAGCCGCTTTCGCCACCGGTGTTCCTCCATATAT  199
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  67575  AGCCTCAGCGTCAGTTACAGACCAGAGAGCCGCTTTCGCCACCGGTGTTCCTCCATATAT  67516

Query  200    CTACGCATTTCACCGCTACACATGGAATTCCACTCTCCCCTTCTGCACTCAAGTTTGACA  259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  67515  CTACGCATTTCACCGCTACACATGGAATTCCACTCTCCCCTTCTGCACTCAAGTTTGACA  67456

Query  260    GTTTCCAAAGCGAACTATGGTTGAGCCACAGCCTTTAACTTCAGACTTATCAAACCGCCT  319
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  67455  GTTTCCAAAGCGAACTATGGTTGAGCCACAGCCTTTAACTTCAGACTTATCAAACCGCCT  67396

Query  320    GCGCTCGCTTTACGCCCAATAAATCCGGACAACGCTCGGGACCTACGTATTACCGCGGCT  379
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  67395  GCGCTCGCTTTACGCCCAATAAATCCGGACAACGCTCGGGACCTACGTATTACCGCGGCT  67336

Query  380    GCTGGCACGTAGTTAGCCGTCCCTTTCTGGTAAGCTACCGTCACAGTGTGAACTTTCCAC  439
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  67335  GCTGGCACGTAGTTAGCCGTCCCTTTCTGGTAAGCTACCGTCACAGTGTGAACTTTCCAC  67276

Query  440    TCTCACACCCGTTCTTGACTTACAACAGAGCTTTACGATCCGAAAACCTTCTTCACTCAC  499
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  67275  TCTCACACCCGTTCTTGACTTACAACAGAGCTTTACGATCCGAAAACCTTCTTCACTCAC  67216

Query  500    GCGGCGTTGCTCGGTCAGGGTTGCCCCCATTGCCGAAGATTCCCTACTGCTGCCTCCCGT  559
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  67215  GCGGCGTTGCTCGGTCAGGGTTGCCCCCATTGCCGAAGATTCCCTACTGCTGCCTCCCGT  67156

Query  560    AGGAGTCTGGGCCGTGTCTCAGTCCCAGTGTGGCCGATCACCCTCTCAGGTCGGCTATGT  619
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  67155  AGGAGTCTGGGCCGTGTCTCAGTCCCAGTGTGGCCGATCACCCTCTCAGGTCGGCTATGT  67096

Query  620    ATCGTCGCCTAGGTGAGCCATTACCTCACCTACTAGCTAATACAACGCAGGTCCATCTTG  679
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  67095  ATCGTCGCCTAGGTGAGCCATTACCTCACCTACTAGCTAATACAACGCAGGTCCATCTTG  67036

Query  680    TAGTGGAGCAATTGCCCCTTTCAAATAAATGACATGTGTCATCCATTGTTATGCGGTATT  739
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  67035  TAGTGGAGCAATTGCCCCTTTCAAATAAATGACATGTGTCATCCATTGTTATGCGGTATT  66976

Query  740    AGCTATCGTTTCCAATAGTTATCCCCCGCTACAGGGCAGGTTACCTACACGTTACTCACC  799
              |||||||||||||||||||||||||||||||| ||||||||||||| ||||||||||||
Sbjct  66975  AGCTATCGTTTCCAATAGTTATCCCCCGCTACAAGGCAGGTTACCTACGCGTTACTCACC  66916

Query  800    CGTTCGCAACTCATCCAAGAAGAGCAAGCTCCTCTCTTCAGCGTTCTAC  848
              |||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  66915  CGTTCGCAACTCATCCAAGAAGAGCAAGCTCCTCTCTTCAGCGTTCTAC  66867
```

Blast analysis of strain N5-4

```
>dbj|AB002481.1|  S. bovis DNA for 16S rRNA, strain ATCC 27960
Length=1500

 Score = 1513 bits (819), Expect = 0.0
 Identities = 834/841 (99%), Gaps = 2/841 (0%)
 Strand=Plus/Minus

Query: strain N5-4
Subject: S. bovis ATCC 27960

Query  9    CCCCA-GCGG-GTGCTTAATGCGTTAGCTGCGGCACTAAGCCCCGGAAAGGGCCTAACAC  66
            ||||| |||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  882  CCCCAGGCGGAGTGCTTAATGCGTTAGCTGCGGCACTAAGCCCCGGAAAGGGCCTAACAC  823

Query  67   CTAGCACTCATCGTTTACGGCGTGGACTACCAGGGTATCTAATCCTGTTTGCTCCCCACG  126
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  822  CTAGCACTCATCGTTTACGGCGTGGACTACCAGGGTATCTAATCCTGTTTGCTCCCCACG  763

Query  127  CTTTCGAGCCTCAGCGTCAGTTACAGACCAGAGAGCCGCTTTCGCCACCGGTGTTCCTCC  186
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  762  CTTTCGAGCCTCAGCGTCAGTTACAGACCAGAGAGCCGCTTTCGCCACCGGTGTTCCTCC  703

Query  187  ATATATCTACGCATTTCACCGCTACACATGGAATTCCACTCTCCCCTTCTGCACTCAAGT  246
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  702  ATATATCTACGCATTTCACCGCTACACATGGAATTCCACTCTCCCCTTCTGCACTCAAGT  643

Query  247  CTAACAGTTTCCAAAGCGAACAATGGTTAAGCCACTGCCTTTAACTTCAGACTTATTAAA  306
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  642  CTAACAGTTTCCAAAGCGAACAATGGTTAAGCCACTGCCTTTAACTTCAGACTTATTAAA  583

Query  307  CCGCCTGCGCTCGCTTTACGCCCAATAAATCCGGACAACGCTCGGGACCTACGTATTACC  366
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  582  CCGCCTGCGCTCGCTTTACGCCCAATAAATCCGGACAACGCTCGGGACCTACGTATTACC  523

Query  367  GCGGCTGCTGGCACGTAGTTAGCCGTCCCTTTCTGGTAAGTTACCGTCACTGTGTGAACT  426
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  522  GCGGCTGCTGGCACGTAGTTAGCCGTCCCTTTCTGGTAAGTTACCGTCACTGTGTGAACT  463

Query  427  TTCCACTCTCACACACGTTCTTCTCTTACAACAGAGCTTTACGATCCGAAAACCTTCTTC  486
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  462  TTCCACTCTCACACACGTTCTTCTCTTACAACAGAGCTTTACGATCCGAAAACCTTCTTC  403

Query  487  ACTCACGCGGCGTTGCTCGGTCAGGGTTGCCCCCATTGCCGAAGATTCCCTACTGCTGCC  546
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  402  ACTCACGCGGCGTTGCTCGGTCAGGGTTGCCCCCATTGCCGAAGATTCCCTACTGCTGCC  343

Query  547  TCCCGTAGGAGTCTGGGCCGTGTCTCAGTCCCAGTGTGGCCGATCACCCTCTCAGGTCGG  606
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  342  TCCCGTAGGAGTCTGGGCCGTGTCTCAGTCCCAGTGTGGCCGATCACCCTCTCAGGTCGG  283

Query  607  CTATGTATCGTCGCCTTGGTGAGCCGTTACCTCACCAACTAGCTAATACAACGCAGGTCC  666
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  282  CTATGTATCGTCGCCTTGGTGAGCCGTTACCTCACCAACTAGCTAATACAACGCAGGTCC  223

Query  667  ATCTACTAGTGAAGCAATTGCTCCTTTCAAGCATCTAACATGGGGTAAATGCTGTTATGC  726
            |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Sbjct  222  ATCTACTAGTGAAGCAATTGCTCCTTTCAAGCATCTAACATGGGTTAAATGCTGTTATGC  163

Query  727  GGTATTAGCTATCGTTTCCAATAGATATCCCCCGCTAGTAGGCAGGTTACCTACGCGTTA  786
            |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Sbjct  162  GGTATTAGCTATCGTTTCCAATAGTTATCCCCCGCTAGTAGGCAGGTTACCTACGCGTTA  103

Query  787  CTCACCCGTTCGCAACTCTTCCAACTTTAGCAAGGTAAGGTTTTCAGCGTTCTACTTGCA  846
            |||||||||||||||||||||||||||||||| ||| || ||||||||||||||||||||
Sbjct  102  CTCACCCGTTCGCAACTCTTCCAACTTTAGCAAGCTAAAGTCTTCAGCGTTCTACTTGCA  43

Query  847  T  847
            |
Sbjct  42   T  42
```

APPENDIX G

MLST ANALYSIS OF *L. DELBRUECKII* SUBSP. *BULGARICUS*

β-*gal*: 6 alleles, 8 polymorphic sites

```
  1       10        20        30        40        50        60
  tttatctcagttgaatacgctcacgccatgggcaactccgtcggtgacctggccgcctac

 61       70        80        90       100       110       120
  acggccctggaaaaataccccactaccagggcggcttcatctgggactggattgaccaa
       t

121      130       140       150       160       170       180
  ggactggaaaaagacgggcacctgctttatgggggcgacttcgatgaccggccaaccgac

181      190       200       210       220       230       240
  tatgaattctgcgggaacggcctggtctttgctgaccggactgaatcgccgaaactggct
                 g

241      250       260       270       280       290       300
  aatgtcaaggccctttacgccaaccttaagttagaagtaaaagatgggcagctcttcctc
             t

301      310       320       330       340       350       360
  aaaaacgacaatttatttaccaacagctcatcttactacttcttgactagtcttttggtc

361      370       380       390       400       410       420
  gatggcaagttgacctaccagagccggcctctgacctttggcctggagcctggcgaatcc
                           a               t

421      430       440       450       460       470       480
  gggacctttgccctgccttggccggaagtcgctgatgaaaaaggagaggtcgtctaccgg
             t                 t              g

481      490       500       510       520       530       540
  gtaacggcccacttaaaagaagacttgccttgggcggatgagggcttcactgtggctgaa

541      550       560       570       580       590
  gcagaagaagtagctcaaaagctgccggaatttaagccggaagggcggccagattta
```

Table G.1. Nucleotide differences of β-*gal* gene between alleles and isolates

Position	a alleles	c alleles	g alleles	t alleles	a isolates	c isolates	g isolates	t isolates
66		5		1 (bgal-2)		14		1 (ST-15)
196	4		2		7		8	
252		2		4		3		12
386	2		4		8		7	
403		3		3		6		9
432		3		3		6		9
450		5		1 (bgal-6)		11		4
465	4		2		7		8	

*phe*S: 6 alleles, 11 polymorphic sites

```
  1       10        20        30        40        50        60
  cggagccagacttcaccagttcaagcccggaccatggaaaagcatgacttcgataaaggg

 61       70        80        90       100       110       120
  gacctcaagatgatctccccaggcaaggtttacagaagagacgatgacgacgccacccac
    t                                 g

121      130       140       150       160       170       180
  tcccaccagttcatgcagatggaagggctggttgtcggcaagaacatctccttgagtgac

181      190       200       210       220       230       240
  cttaaggggaccttggaactggtggccaagcacgaattcggccaggaccgggaaacccgc

241      250       260       270       280       290       300
  ttgcggccaagctacttcccatttactgaaccatcagttgaaatggacgtttcttgcttt
             a                  t g

301      310       320       330       340       350       360
  gaatgcggcggcaagggctgcgcgatctgcaagaacaccggctggatcgaagttctgggt
  a                    c   c                     a

361      370       380       390       400       410       420
  gccgggatcgttcacccgaatgttttgtctgccgccggcattgacccaagcgtctactct
            c                    t
```

Table G.2. Nucleotide differences of *phe*S gene between alleles and isolates

Position	a alleles	c alleles	g alleles	t alleles	a isolates	c isolates	g isolates	t isolates
63		3		3		9		6
94	5		1 (phes-5)		14		1 (ST-15)	
252	1 (phes-6)	5			1 (ST-7)	14		
271		5		1 (phes-3)		14		1 (ST-8)
273	5		1 (phes-4)		14		1 (ST-13)	
301	1 (phes-6)		5		1 (ST-7)		14	
322		1 (phes-4)	5			1 (ST-13)	14	
326		1 (phes-4)		5		1 (ST-13)		14
349	1 (phes-6)		5		1 (ST-7)		14	
371		1 (phes-3)		5		1 (ST-8)		14
392		5		1 (phes-3)		14		1 (ST-8)

*rpo*A: 4 alleles, 3 polymorphic sites

```
  1       10        20        30        40        50        60
  ccgctggagcgtgggtttggtactaccttaggtaattcactgcgtcgggttttgctgact

 61       70        80        90       100       110       120
  tctgtcccagggaccggtttggtgaaggtgaagatcgatggtatcttgcacgaattcact
                                                 g
121      130       140       150       160       170       180
  actgttcccggtgttaaagaagacgtaaccaagatcatcttgaacctgaagaagcttgaa
                                                              t
181      190       200       210       220       230       240
  ctccgggcctacactgaagaagtaaagacgatcgaactcgatgttgaaggtccagctacg

241      250       260       270       280       290       300
  gtaactgctgaagatttgaaggctgatgctgatgttgaagtcttgaatcctgaccaatac

301      310       320       330       340       350       360
  atttgtaccatcgctcaaggtggccacctgcacatgtggattgatgtctgcaacggccgg
                                   a
361      370       380       390       400       410       420
  ggctacgtaccagccagcgaaaacaagactgctgaaatgtccatcggcgacattccagtt

421      430       440       450       460
  gactcacttttctcaccaatcgaaaaggtcaactaccaagttgaa
```

Table G.3. Nucleotide differences of *rpo*A gene between alleles and isolates

Position	a alleles	c alleles	g alleles	t alleles	a isolates	c isolates	g isolates	t isolates
103	2		2		13		2	
179	3			1 (rpoa-3)	10			5
330	1 (rpoa-4)		3		1 (ST-14)		14	

APPENDIX H

MLST ANALYSIS OF *S. THERMOPHILUS*

rpoa: 6 alleles, 6 polymorphic sites

```
  1       10        20        30        40        50        60
  tttgtcatcgaaccactagaacgtggttatggtacaactttgggtaactctcttcgtcgt
                             c     g

 61       70        80        90       100       110       120
  gtactcttgtcatcacttccaggtgctgccgtaacatcaattaaaattgatggagtgctc
                           c                                a

121      130       140       150       160       170       180
  cacgagtttgataccgtaccaggcgtccgtgaagacgtgatgcaaattatccttaatatt

181      190       200       210       220       230       240
  aagggacttgctgtaaaatcttacgtcgaagacgaaaaaacaattgaacttgatgttcaa

241      250       260       270       280       290       300
  ggtccagctgaggttactgctggagatatcttgactgatagtgacattgaaattgttaac

301      310       320       330       340       350       360
  cctgatcattatctttttactatcgctgagggtgcttcattccaagcaacaatgacagta
       a

361      370       380       390       400       410       420
  tctacaaaccgtggttatgttccagcagaagaaaataaaaaagatgatgcaccagtggga

421      430       440       450       460       470       480
  actttggctgtagattcaatctacacaccagtgaaaaaagtcaattatcaagttgaacct

481      490       500       510       520       530       540
  gctcgtgtaggtagcaatgatggttttgataaattgacaatcgaaatcgtgacaaacgga
                                     g

541      550       560       570       580
  acaatcatcccagaagatgccctcggtctctcggctcgaattttg
```

Table H.1. Nucleotide differences of *rpo*A gene between alleles and isolates

Position	a alleles	c alleles	g alleles	t alleles	a isolates	c isolates	g isolates	t isolates
27		2		4		4		21
33			1 (rpoa-4)	5			2	23
84		1 (rpoa-4)		5		2		23
116	2			4	3			22
306	4			2	15			10
513	5		1 (rpoa-5)		22		3	

phes: 3 alleles, 2 polymorphic sites

```
  1       10        20        30        40        50        60
  agtcctgtccaagctcgtacacttgataaacatgatttttctaaaggtcctcttaagatg

 61       70        80        90       100       110       120
  atctcaccaggacgtgttttccgtcgtgataccgatgatgcgactcacagccaccagttt

121      130       140       150       160       170       180
  caccaaatcgaaggtttggtcgttggtaaaaacatctcaatgggtgatctgaagggaacg
                                                        a

181      190       200       210       220       230       240
  cttgagatgattattcaaaaaatgtttggtgcagaacgtcaaatccgtttgcgtccttct

241      250       260       270       280       290       300
  tacttcccattcactgaaccttctgttgaggttgacgtgtcatgcttcaagtgtggtggt
                           c

301      310       320       330       340       350       360
  aaaggatgtaacgtatgcaagaagacaggttggattgagatccttggtgctggtatggtt

361      370       380       390       400
  cacccacaagtgcttgagatgtcaggtgttgattctgaagaatattca
```

Table H.2. Nucleotide differences of *phe*S gene between alleles and isolates

Position	a alleles	c alleles	g alleles	t alleles	a isolates	c isolates	g isolates	t isolates
171	1 (phes-2)		2		1 (ST-25)		24	
264		2		1 (phes-1)		3		22

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Cebeci Aydın, Aysun
Nationality: Turkish (TC)
Date and Place of Birth: 19 April 1978, Sinop
Marital Status: Married
email: aysuncebeci@yahoo.com

EDUCATION

Degree	Institution	Year of Graduation
MSc	METU Biochemistry	2002
BS	METU Biology	1999
High School	Ankara Atatürk Lisesi	1994

WORK EXPERIENCE

Year	Place	Enrollment
2001-2006	METU Department of Biochemistry	Research Assistant
2000-2001	Gazi University Hospital	Biologist
1998 July	BİFA (Bayer) Pharmaceutical Company	Intern Student

FOREIGN LANGUAGES

Advanced English, Beginners French

PUBLICATIONS

1. Cebeci A. and Gürakan C. (2003) Properties of potential probiotic *Lactobacillus plantarum* strains. Food Microbiology 20: 511-518.
2. Cebeci A. and Gürakan G.C. (2008) Rapid joint identification of *Lactobacillus bulgaricus* and *Streptococcus thermophilus* isolates using molecular methods. Journal of Dairy Research (in revision)
3. Cebeci A., Altay N., and Gürakan G.C. Identification and characterization of *Lactobacillus delbrueckii* subsp. *bulgaricus* and *Streptococcus thermophilus* isolated from home-made Turkish yogurts using biochemical and molecular tools. Biomicroworld2007, (November 2007) Sevilla, Spain.
4. Altay N, Cebeci A. Gürakan G.C. Characterization and identification of wild strains isolate from traditionally prepared Turkish yoghurts. 5th NIZO conference: Prospects for flavor formation and perception (June 2007) Papendal, Netherlands.
5. Gürakan G.C. and Cebeci Aydın A. Development of DNA probe using RAPD-PCR fort he rapid identification of *Lactobacillus plantarum*. 8th Symposium on Lactic Acid Bacteria, (August 2005), Egmond Aan Zee, Netherlands.
6. Cebeci A. and Gürakan G. C. Probiotic Characterization of *Lactobacillus plantarum* strains. XIII. Biotechnology Congress, (August 2003), Çanakkale, Turkey.

HOBBIES

Reading, listening to music, biking