PHYLOGENETIC ANALYSIS OF BACTERIAL COMMUNITIES IN KEFIR BY METAGENOMICS

A Thesis Submitted to
the Graduate School of Engineering and Sciences of
İzmir Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

in Molecular Biology and Genetics

by
Burcu ÜNSAL

July 2008
İZMİR

We approve the thesis of **Burcu ÜNSAL**

Assist. Prof. Dr. Alper ARSLANOĞLU
Supervisor

Assist. Prof. Dr. H. Çağlar KARAKAYA
Committee Member

Assist. Prof. Dr. Figen KOREL
Committee Member

4 July 2008

Date

Assoc. Prof. Sami Doğanlar
Head of the Molecular Biology and Genetics Department

Prof. Dr. Hasan BÖKE
Dean of the Graduate School of Engineering and Sciences

ACKNOWLEDGEMENTS

I wish to express my deepest gratitude to my dear supervisor Assist. Prof. Dr. Alper ARSLANOĞLU for giving me opportunity to work with him and also I would like to thank him for his supervision, encouragement, criticism and patience to guide me during my thesis studies.

I also wish to thank Assist. Prof. Dr. H. Çağlar KARAKAYA for his valuable comments and encouragement during my thesis studies.

Other thanks are owed to my dear friend Nesrin HORZUM for her moral support, invaluable friendships and as well as to my other friends Burcu ŞENGEZ, Aysun ADAN who are working with me in the Molecular Microbiology Laboratory for their sincere help offers and friendships. Besides them I also wish to thank Erdal EROĞLU and Hasan TÜRE who have been so wonderful to me at all times in every way.

I would like to thank Zeynep YEĞİN, Ersin AKINCI, E. Selçuk ÜNLÜ and Mert SUDAĞIDAN for their support who shared their experiences with me.

I wish to acknowledge my colleagues in Molecular Biology and Genetics Department in İYTE for their friendship during past three years.

Also, I am very grateful to Salim ÜNAL for being the best ever friend for me. He has been a true and great supporter and he has endless love to me through both my good and bad times during this thesis studies.

Finally, I would especially like to send my deep gratitude and appreciation to my family members Ali ÜNSAL, Fatma ÜNSAL, R. Başar ARTAN, and E. Tolga ARTAN. Their unconditional love, belief and faith in me have been an enormous encouragement that has kept me going during all these years.

ABSTRACT

PHYLOGENETIC ANALYSIS OF BACTERIAL COMMUNITIES IN KEFIR BY METAGENOMICS

Kefir is a traditional fermented milk beverage which is produced by adding kefir grains into milk and is allowed for fermentation. Grains contain vital complex flora of microorganisms (bacteria and yeast) that live in harmony. Since health and food safety of fermented milk products is important, population structure of food-type microbes involve in fermentation should be known very well. Rapid determination of kefir bacterial composition may accelerate the determination of food quality and also may facilitate specification of bioactive products that obtain from kefir. The goal of this thesis was to analysis the genomic structure of bacterial communities of the fermented kefir drink and grains by both culture-dependent and culture-independent methods (metagenomic approach). Total Genomic DNA was purified from each analysis methods and the partial small subunits of 16S rDNA were amplified by a pair of universal bacterial primers. 16S rDNAs fragments were cloned and then sequenced. The vast quantities of data were screened in NCBI database by BLASTN program according to similarity scores with related sequences. 7 different bacteria were identified to species level composed of *Lactococcus lactis subsp. lactis , Lactobacillus kefiranofaciens, Lactobacillus helveticus, Acetobacter lovaniensis, Acetobacter syzygii, Leuconostoc mesenteroides, Enterococcus faecium* and 1 bacteria to genus level named *Lactobacillus kefiri or parabucheri.* The results of this study showed that the combination of both methods is more efficient to identify high percentage of species than using only one of them. Finally, phylogenetic relationships among identified species inferred from partial 16S rRNAs gene sequencing were determined by Neighbor-joining algorithm.

ÖZET

KEFİR BAKTERİ KOMÜNİTELERİNİN METAGENOMİK YÖNTEMLERLE FİLOGENETİK ANALİZİ

Geleneksel ferment bir içecek olan kefir, kefir tanelerinin süte ilave edilip sütün mayalanmaya bırakılmasıyla elde edilir. Kefir taneleri birbirleriyle uyum içinde yaşayan bakteriler ve mayalardan oluşan karışık bir mikrofloraya sahiptir. Fermente süt ürünlerinin gıda ve insan sağlığı açısından güvenliği önem taşıdığı için mayalanmada görevli mikobiyal popülasyonun yapı ve özellikleri de çok iyi bilinmelidir. Kefir bakteriyel kompozisyonun hızlı bir şekilde belirlenmesi gıda güvenliğinin tespitini hızlandırır ve ayni zamanda kefirden elde edilecek biyoaktif maddelerin belirlenmesini kolaylaştırır. Bu çalışmanın amacı fermente kefir içeceğindeki ve kefir tanelerindeki bakteri komünitelerinin kültüre dayalı ve kültüre dayalı olmayan metotlarla (metagenomik yaklaşım) belirlenmesidir. Kefir içeceği ve kefir tanelerinden, her iki analiz metoduyla da toplam genomik DNA izole edilmiş ve kısmi 16S rRNA gen bölgesi üniversal bakteri primerleri ile çoğaltılmıştır. Çoğaltılmış 16S rRNA gen bölgeleri bir klonlama vektörü olan PGEMT-Easy plazmidine aktarılmış ve pozitif klonların DNA dizi analizleri yapılmıştır. Geniş çapta elde edilen DNA dizi analizi verileri NCBI veritabanındaki BLASTN programı kullanılarak taranmış ve benzerlik skorlarına göre de 7 farklı bakteri tür düzeyinde (*Lactococcus lactis subsp. lactis, Lactobacillus kefiranofaciens, Lb. helveticus, Acetobacter lovaniensis, A.syzygii, Leuconostoc mesenteroides, Enterococcus faecium*) *ve* 1 tane bakteride genus düzeyinde belirlenmiştir *(Lactobacillus kefiri or parabucheri*). Belirlenen türlerin filogenetik yakınlıkları 16S rRNA gen bölgeleri kullanılarak Neighbor-joining algoritması ile belirlenmiştir. Bu çalışmanın sonuçları, mikrobial açıdan karışık bir kültürdeki populasyonun büyük çoğunluğunun belirlenmesinde, kültüre dayalı olan ve kültüre dayalı olmayan yöntemlerin birlikte kullanılmasının tek bir metodun kullanılmalarından daha etkili olduğu göstermiştir.

Dedicated to;

the memory of my grandfather

and

my lovely family for being with me in all my life...

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

LIST OF TABLES

Table **Page**

ABBREVIATIONS

IG	Samples obtained from Kefir Grains by culture-independent method
IL	Samples obtained from Kefir Drink by culture-independent method
DG	Samples obtained from Kefir grains by culture-dependent method
DL	Samples obtained from Kefir drink by culture-dependent method
Leu. mesenteroides	*Leuconostoc mesenteroides*
Lc. lactis subsp. lactis	*Lactococcus lactis subsp. lactis*
Lb. helveticus	*Lactobacillus helveticus*
E. faecium	*Enterococcus faecium*
A. lovaniensis	*Acetobacter lovaniensis*
A. syzygii	*Acetobacter syzygii*
Lb. kefiranofaciens	*Lactobacillus kefiranofaciens*
Lb. kefiri	*Lactobacillus kefiri*
dNTP	deoxynucleoside triphosphate
ddTTP	dideoxythymidine-triphosphate
ddCTP	dideoxycytidine -triphosphate
ddATP	dideoxyadenosine -triphosphate
ddGTP	dideoxyguanosine -triphosphate
ddNTP	dideoxynucleoside triphosphate
NJ	Neighbor joining Method
MP	Maximum Parsimony Method
ML	Maximum likelihood
OTU	Operational taxonomic units
BLASTn	Basic local alignment search tool for nucleic acids
rRNA	Ribosomal ribonucleic acid

rDNA	ribosomal deoxyribonucleic acid
SSU	Small subunit
LSU	Large subunit
DGGE	Denaturing Gradient Gel Electrophoresis
UHT	Ultra-high temperature
sp.	species (singular)
bp	Base pair
kb	Kilobase pairs
l	Liter
M	Molar
mM	Millimolar
µM	Micromolar
ml	Milliliter
µl	Microliter
fmol	Fentomole
µg	Microgram
ng	Nanogram
nm	Nanometer
w/v	Weight per volume
UV	Ultraviolet
LB	Luria-Bertani
IPTG	Isopropyl-thio-β-D-galactopyranoside
X-Gal	5-bromo-4-chloro-3 indolyl-B-D-galactoside
DMSO	Dimethylsulphoxide
MRS medium	Man, Rogosa and Sharpe medium
TAE	Tris/acetate/ethylenediamine tetra-acetic acid (buffer)
Taq	Thermus aquaticus (DNA polymerase)
PCR	Polymerase chain reaction
min	Minute
sec	Second
V	Voltage
rpm	Revolutions per minute

CHAPTER 1

INTRODUCTION

1.1.Microbial Diversity and Community Structure of Kefir

Food fermentation is the oldest biotechnology which has been used by humans for centuries. A wide range of microbial and enzymatic activities occurred during the fermentation process which gives desirable properties to the food product by adding flavor, prolonging shelf life, and improving safety. In the fermentation of foods, mostly lactic acid bacteria (LAB), yeast and moulds are involved and they convert the raw food material to fermented product. A variety of fermented food products can be found throughout the world such as cheese, bread, sausages, wines, beers, yoghurt and kefir (Giraffa 2004).

Kefir is a traditional fermented milk beverage with a characteristic viscous, slightly carbonated and acidic taste (Guzel-Seydim, et al. 2000). It is believed to be originated in the Caucasus Mountains of Russia thousand of years ago and is being consumed by people all over the world from past to today (Garrote, et al. 1997).

Kefir starter culture, in the form of grains, is added into milk to accomplish acid-alcoholic fermentation process and it is recovered as a solid grainy matrix after completion of the fermentation to be used for the next fermentation process (Witthuhn, et al. 2004). When active kefir grains are continually cultured in fresh milk, it increases in mass and number, however, it maintains a stable microbial population (Ninane, et al. 2005). Because of the fact that grains contain vital complex flora of microorganisms that live in harmony, they are considered as a dynamic structure. The association of different microorganisms in kefir was developed by continues selection over the centuries and these microorganisms compete with other potential pathogenic microorganisms providing a pathogen free ecology in kefir (Lopitz-Otsoa, et al. 2006). Microbial population of kefir typically produces lactic acid and antimicrobial substances which prevent the proliferation of both spoilage and pathogenic microorganisms in kefir (Farnworth 2005, Santos, et al. 2003).

Microbial population of the grains can grow, multiply and transfer their properties to the following generation of new grains during the fermentation process. Once the grains are inoculated into milk, the microorganisms are dispersed into milk phase (Garrote, et al. 1997). Thus, the presence of microorganisms can be observed in both the kefir grains and the kefir liquid after fermentation process (Chen, et al. 2008).

Since kefir contains mixture of probiotic bacteria and yeast (Lopitz-Otsoa, et al. 2006, Farnworth 2005), it is considered as a functional product that demonstrates many beneficial effects on human health as having a high nutritional, biological and dietetic value. It is also known to lower the risk of chronic diseases and also recommended to patients who suffer from hypertension, allergy, gastrointestinal and metabolic diseases (Zubillaga, et al. 2001, Otles and Cagindi 2003).

The importance of health and food safety of the fermented milk products has been increasing recently. The industrial quality control processes require knowledge about ecological and biochemical factors and the evaluation of the effective composition of food-type microbes' population (Fleet 1999). This type of knowledge enables the assessment of modern quality assurance, predictive fermentation model and risk analysis strategies used for the prevention of food borne pathogens, food spoilage, and increasing aroma and taste of food. The bacterial growth and activities in the fermented food can be effectively manageable, if some information about their physiological, genetic and metabolic properties is known. Information about diversity, taxonomic identity, distribution of microorganisms in fermented products throughout different fermentation stages; and relationships between growth and activity of individual species must be known in order to understand the ecology of fermented foods (Ampe, et al. 1999). Furthermore, extrinsic and intrinsic factors that affect growth and biochemical properties of species in fermented foods must be studied to make fermentation procedure under control (Fleet 1999).

Various types of bacteria and fungi have been isolated and identified from kefir microbial consortia by using selective growth media, morphological and biochemical characteristics (Simova, et al. 2002, Witthuhn, et al. 2005). Witthuhn and his friends concluded that the enumeration values changed between $6.4x10^4$ and $8.5x10^8$ cfu/g for the bacterial species, and $1.5x10^5$ and $3.7x10^8$ cfu/g for the yeast species in kefir grains on different selective media (Witthuhn, et al. 2004).

Kefir grains have a white/yellow, gelatinous and irregular structure varying from cauliflower-like to scroll forms. These structures are elastic and quite tough which are

populated by yeast and lactobacilli. The natural microbiological mass including bacteria and yeasts in kefir grains are surrounded by protein and lipids and soluble polysaccharide matrix named, kefiran (Farnworth 2005). The microflora is embedded in this biopolymer matrix which is constructed by repeating units of branched glucogalactan and its molecular weight is 1,000-4,000 kDa (Yokoi, et al. 1990). The heteropolysaccharide kefiran is produced by several bacteria isolated from kefir including *Lactobacillus kefiranofaciens, L. kefirgranum, L parakefir, L. kefir and L. delbrueckii subsp. bulgaricus, L. bulgaricus* (Vinderola 2006, Frengova, et al. 2002).

The bacterial content of the grains are usually composed of various Gram positive homo- and heterofermentative lactic acid bacteria (LAB) species of *Lactobacillus, Lactococcus, Leuconostoc*, *Enterococcus, Streptococcus;* Gram negative acidic acid bacteria (AAB) species of *Acetobacter*. The yeast content of the grains is usually composed of *Kluyveromyces, Saccharomyces, Candida and Torulopsis* species (Simova, et al. 2002, Guzel-Seydim, et al. 2005, Ninane, et al. 2005). Yeast and bacteria are thought to be in a symbiotic relationship that they share their bioproducts as energy source or growth stimulant to survive or multiply in the same environment (Lopitz-Otsoa, et al. 2006). Yeasts not only provide growth stimulants which enable to activate lactic acid bacteria but also some yeast species metabolize some of the lactic acid in the environment. Further more, ethanol and carbon dioxide are produced by yeasts. Kefir microbiological diversity and majority can change according to cultivation method and the origin of the grains and milk itself (Ninane, et al. 2005, Schoevers and Britz 2003).

Reviewed literature data about the bacterial composition of kefir grains indicated that the identify species varied among different studies. However, some of the bacterial species were common in all the grains and liquid studies (Table 1.1.). To date, many microbial diversity studies on kefir fermentations have been performed by analyzing isolates cultivated on plates. However, some strains in the microbial population of kefir may still remain unidentified due to the lack of factors (such as symbionts or the necessary ingredients) in culture medium used for the isolation of kefir bacteria (Witthuhn, et al. 2005). The inability to isolate some of the bacteria from a complex environment such as kefir can also be due to the stress conditions inflicted upon the bacteria by the unnatural growth conditions. Stress is any change in the genome, proteome or environment that causes the decrease in the growth and survival capacity of living organisms. The viable but non-cultivable (VNC) is a type of stress condition that prevent bacterial colony growth, although, the cell continues its metabolic activity under

this stress condition. Under certain conditions such as limiting nutrition in fermenting food, it is conceivable that food-associated microorganisms may enter the VNC state (Giraffa 2004).

Table 1.1. Bacterial species isolated from Kefir Grains by various research groups

Bacterial Diversity	**References**
Lactic acid bacteria	
Lactococci	
Lactococcus cremoris	Yüksekdağ et al., 2004
Lactococcus lactis subsp.	Witthuhn et al., 2004, Garrote et al., 1996, Simova et al., 2002, Chen et al., 2008, Yüksekdağ et al., 2004, Mainville et al., 2005
Streptococci	
Streptococcus durans	Yüksekdağ et al., 2004
Streptococcus thermophilus	Simova et al., 2002, Yüksekdağ et al., 2004
Lactobacilli	
Lb. fermentum subsp.	Witthuhn et al., 2004
Lb. delbrueckii subsp..	Witthuhn et al., 2004, Simova et al., 2002
Lb. kefiranofaciens	Vancanneyt et al., 2004, Santos et al., 2003, Chen et al., 2008, Garrote et. al., 1996, Mainville et al., 2005
Lb. plantarum	Witthuhn et al., 2004, Santos et al., 2003
Lb. helveticus	Simova et al., 2002
Lb. brevis	Simova et al., 2002, Witthuhn et al., 2004
Lb. casei subsp	Simova et al., 2002
Lb. kefiri	Garbers et al. 2004, Mainville et al., 2005, Heo et al. 2006
Lb. gallinarum	Garbers et al. 2004
Lb. acidophilus	Santos et al., 2003
Lb. curvatus	Witthuhn et al., 2004
Lb. fermentum	Witthuhn et al., 2004
Lb. crispatus	Garbers et al. 2004
Enterococci	
Enterococcus ssp.	Fontan et al., 2005
Acedic acid bacteria	
Acetobacter aceti	Angulo et al., 1993, Marshall 1993
Other Bacteria	
Leuconostoc mesenteroides subsp.	Witthuhn et al., 2004, Chen et al., 2008, Mainville et al., 2005
Leuconostoc lactis subsp.	Witthuhn et al., 2004,

In order to overcome the difficulties and limitations related to cultivation techniques, molecular methods offer new opportunities to capture and identify microbial communities from a complex environment. A DNA based molecular approach, metagenomics, can be used to determine the bacterial populations in mix cultures.

1.2. Metagenomics- Culture-independent Studies of Microorganisms

For decades, the cultivation of microorganisms in pure culture is the only way to characterization and identification of them. However, Pace et al. (1985) represented a molecular based approach that has led us to capture all microbial diversity that even can not be analysis by culture-based approaches. Many recent studies enlightened that many bacterial species which could not be cultured are widely distributed and dominant in number in their environmental niche (Fuhrman 2002).

The genomic investigation of the species into the diversity of niches that have slow-growing nature or need special culture requirements can be achieved by direct isolation of total genomic DNA and DNA sequence analysis bypassing culturing and pure culture isolation steps (Wolfgang and Schmitz 2004).

Current estimate points out that vast majority of microorganisms (more than 99 %) that live in a complex environment, such as biofilm, soil etc., can not be cultured in "unnatural" standard laboratory conditions (Amann, et al. 1995). Soil is the most challenging environmental sample with respect to its microbial population and community structure that it remains grossly unexplored. One gram of cultivated soils can contain an estimated amount of 10^9 prokaryotic cells (Torsvik, et al. 1990). 2,000 to 18,000 distinct prokaryotic genomes per gram of soil are estimated based on DNA isolation from various types of soil (Ranjard, et al. 2000). An interesting point is that 1 gram of soil might have as many as all prokaryotic species identified.

Although, many of the distinct microbial populations can be observed by fluorescence microscopy after staining with a fluorescent dye, only a fraction of these populations can be grown on man made growth media (0.001 % to 1 %) (Amann, et al. 1995). This implies that only a small fraction of microbial diversity has ever been explored by traditional culturing methods (Table 1.2.).

Table 1.2. The cultivability percentage of bacteria in different habitats
(Source: Amann, et al. 1995)

Habitat	**Cultivability [%]a**
Seawater	0.001-0.1
Freshwater	0.25
Mesotrophic lake	0.1-1
Activated sludge	1-15
Sediments	0.25
Soil	0.3

Cultivability was determined as percentage of cultivable cells (cfu) in comparison with total cell counts.

In order to overcome the difficulties and limitations related to cultivation techniques, culture-independent microbial diversity analysis approach, metagenomics, can be used to determine the bacterial populations in mix cultures (Handelsman 2004, Wolfgang and Schmitz 2004, Allen and Banfield 2005).

Metagenomics is a novel and also rapidly developing science which focused on the characterization and analysis of gene pool of microbial communities in their natural environment (Dupre and O'Malley 2007, McHardy and Rigoutsos 2007). The term metagenomics was first defined by Jo Handelsman and et al. (1998). Environmental genomics, community genomics, ecogenomics or microbial population genomics are also used as synonyms for metagenomics.

Advances in amplification technologies, DNA cloning, nucleotide sequencing coupled with genomic tools gave rise to the construction and screening of metagenomics libraries which enable to capture composition of microbial communities and to discover novel genes for bioactive compounds and biocatalysts. Genomic investigations enlighten not only the microbial ecological dynamics but also the evolution of species and their physiological properties (Chivian and Arkin 2006).

When the culture-independent microbial diversity analysis is compared to culture-based microbial analysis, it has major advantage of not requiring pure cultures. However, culture-independent approaches can also be applied to characterize pure cultures and it has numerous advantages. It reduces the potential laboratory errors and is not affected by phenotypic variations (Tang, et al. 1998, Hayden, et al. 2001, Petti, et al. 2005). Because of these reasons, genotypic identifications can offer reliable and accurate results as valuable alternative or complementary methods to phenotypic experiments.

Total environmental genomics DNA isolated from a mix microbial population or a specific fraction of the microbiota or enriched genome may be used to construct the metagenomics libraries. To construct the metagenomics libraries, isolated genomic DNA is either fragmented by restriction digestion method or target partial fragments are amplified by PCR. Prepared DNA fragments are cloned into an expression or a cloning vector upon the aim of the experiment and then they are transformed into competent cells. The next step is the analysis of the constructed DNA libraries. The clones can be analyzed by either one of or both of two major metagenomics approaches, the function-driven or the sequence-driven methods (Figure 1.1.).

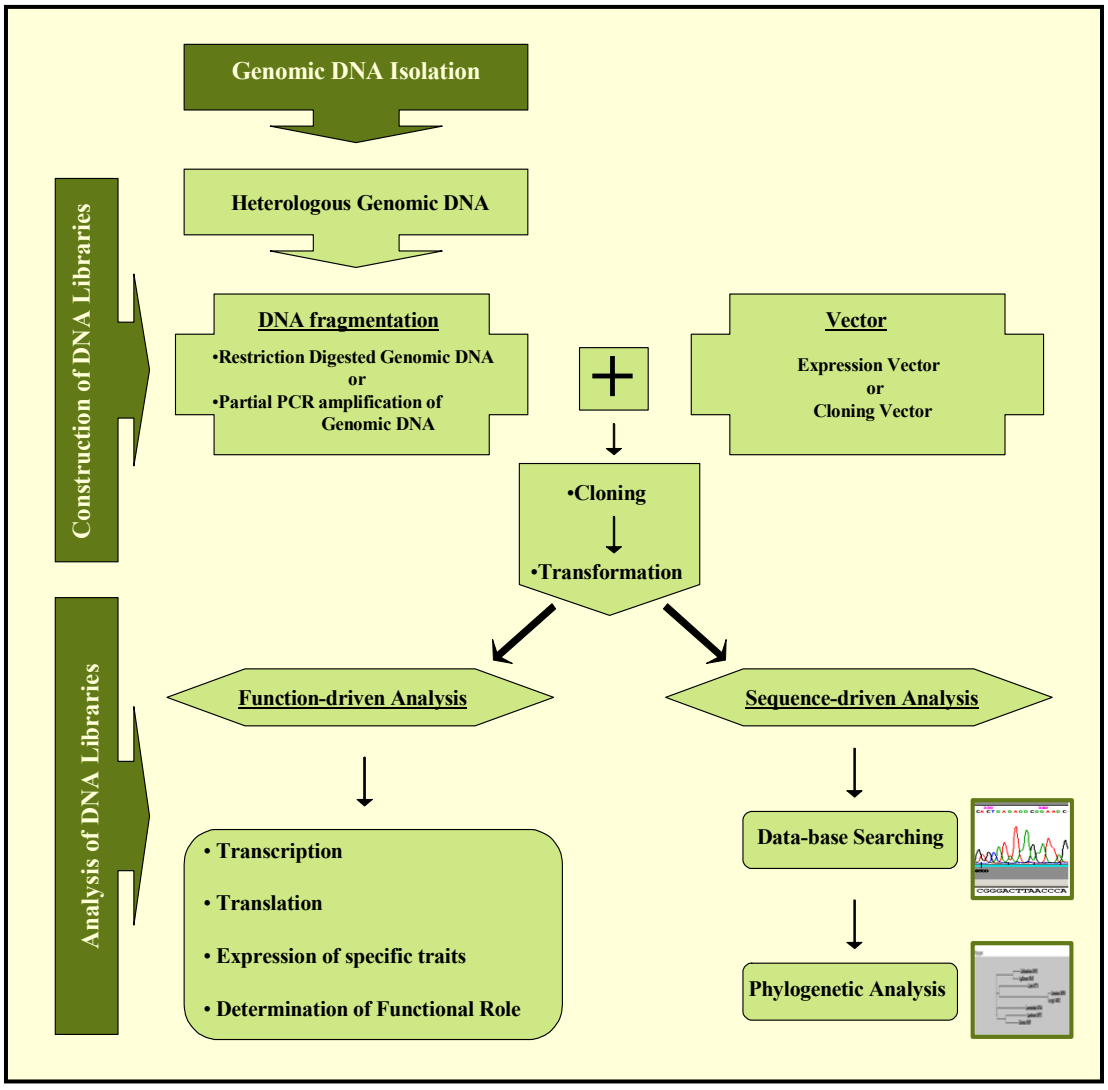


Figure 1.1. The outline of Metagenomics library construction and library screening

1.2.1. The Construction of Metagenomics Libraries

The genomic investigation of the bacterial species uncultivable by routine laboratory cultivation methods can be performed by isolation of direct genomic DNA from environmental samples, construction of metagenomics libraries and finally the analysis of libraries for target genes or sequences (Handelsman 2004, Green and Keller 2006).

The genetic material is extracted by physical (sonication, bead-beating etc.) or chemical methods (alkaline lysis) from the natural environment of microorganisms. The extraction yields of millions of random fragments of genomic DNA. Isolated genomic DNA from mix environmental samples is fragmented by restriction enzymes or mechanical sharing, or a particular region of genomic DNA is amplified by polymerase chain reaction using target gene specific primers. Prepared PCR products or fragmented genomic DNA are then cloned into appropriate cloning or expression vectors to construct the metagenomics libraries (Wolfgang and Schmitz 2004).

The choice of suitable vectors for construction of metagenomics libraries is driven by the desired insert size. Small-insert library vectors such as plasmids or phage vectors (less than 15 kb inserts) are used for the screening of single genes or small operons. Large-insert library vectors such as artificial chromosomes (BACs) (more than 40 kb), cosmids, fosmids (both up to 40 kb) are preferred to recover expression activity of desired gene clusters or complex pathways of large DNA fragments (Figure 1.2.). *E.coli* is usually chosen as cloning or expression host in almost all published studies. However, *Streptomyces lividans* or *Pseudomonas* species or other alternatives can be used as a host cell for cloning or expression of desire gene (Daniel 2005).

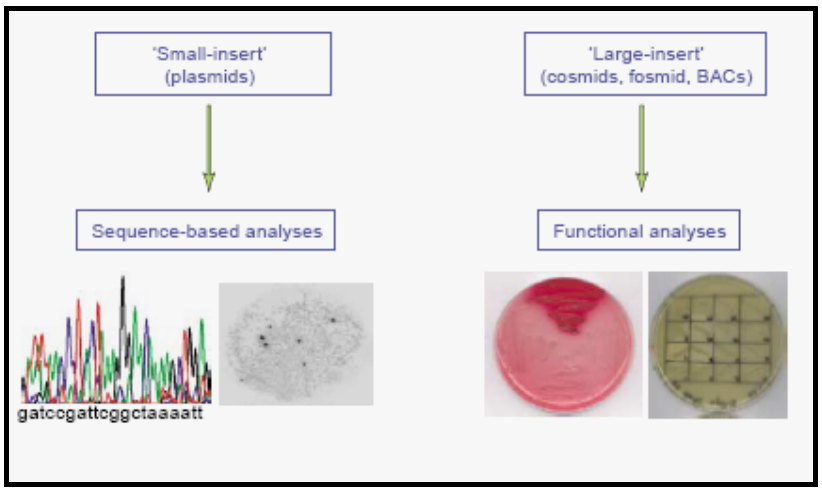


Figure 1.2. The choice of suitable vectors for constructing of Metagenomics libraries (Source: Wolfgang and Schmitz 2004)

The constructed metagenomics libraries are analyzed for the function and/or sequence based information of a microbial community in a target environment by two approaches; the function-driven and the sequence-driven approaches, respectively (Schloss and Handelsman 2003).

In sequence-based approaches, target gene based partial or whole sequence analysis is done. DNA obtained from viruses, sea water, sediment, acid mine drainage, soil, human skin are the samples that sequence-based approaches are applied to analysis of microbial population (Schloss and Handelsman 2008). If the aim is to capture a gene displaying a desired activity DNA fragments are subcloned into an expression vector and are analyzed for expression of the desired gene activity (Schloss and Handelsman 2003).

Sequence-based approach combined with functional analysis was applied to many distinct areas in microbiology to access not only the biodiversity or evolutionary relationship but also to discovery of new functional capacities for industrial and ecological applications (Wolfgang and Schmitz 2004, Schloss and Handelsman 2008, Steele and Streit 2005). Table 1.3. represents the detail application area of metagenomics approaches.

Table 1.3. Applications of Metagenomics analysis (Source: Handelsman, et al. 2004)

Characterization of phylogenetic diversity
Characterization of new genome organization
Elucidation of new biochemical pathways for primary metabolism or energy transduction
Identification of reservoirs of resistance to environmental pollutants
Discovery of enzymes
Discovery of secondary metabolites and other biologically active small molecules
Discovery of polymers

1.2.2. Metagenomics Library Analysis

Community genomics has two general approaches which are based on the identification of microbial population and determination of phylogenetic diversity of them, and screening the metabolic activity of the metagenomics clones. The constructed metagenomic libraries can be screened for phylogenetic markers or for the expression of desired gene enzyme activity, secondary metabolite production, or randomly sequencing (Handelsman 2004).

1.2.2.1. The Sequence-driven Analysis

It is possible to identify microbial populations by approaches which depend on molecular techniques such as comparative sequence analysis of 16S rRNA genes (Mincer 2005, Bosshard, et al. 2003). Broad range PCR primers targeted to the conserved regions of bacterial genomes can be used as phylogenetic markers (16S rRNA, *groEL* (Tsai, et al. 2005), *rpoB* (Drancourt and Raoult 2002), *recA* (Blackwood, et al. 2000), *gyrB* (Dauendorffer, et al. 2003) and hsp 60 heath shock protein (Goh, et al. 1996) which indicate the taxonomic groups can be used in sequence-based analysis to determine the complete sequences of metagenomics clones. Also, it can be used in random sequencing, after the desired target gene is captured phylogenetic markers that flanking DNA are used to determine link between function and phylogeny of the species.

The most large scale project on 16S rRNA sequencing approaches is to identify microbial population, picoplanktons, in the mix environment of Sargasso Sea. rRNA-

targeted oligonucleotide were used to indicate the dominance of members of the domain *Bacteria* in the picoplankton. Subsequently, numerous unknown bacteria are identified by this research (Giovannoni, et al. 1990).

Such methods (molecular identification) also enable to the characterization of pure cultures accurately and rapidly compared to the time consuming biochemical or the other phenotypic characterization tests which results substantial subjective judgment in many cases as a result of variable individual interpretation (Cilia, et al. 1996, Drancourt, et al. 2000, Bosshard, et al. 2003). Furthermore, being a molecular approach, genomic level identification of bacteria, has numerous strengths that is more objective and powerful. It reduces the potential laboratory errors and is not affected by phenotypic variations (Petti, et al. 2005). Because of these reasons, genotypic identifications can offer reliable and accurate results as a valuable alternative or complementary method to phenotypic experiments.

In this experiment we aimed to construct 16S rRNA genes clone libraries of kefir bacterial communities by culture-dependent and culture-independent methods.

Ampe and his friends are the first who applied a molecular technique to a fermented food, *pozol*, to capture the microbial composition (Ampe, et al. 1999). Escalante et al. (2004) applied 16S rDNA sequence analysis to identify genetic diversity of a fermented food (*pulque)* microbial community (Escalante, et al. 2004). Up to date, very few studies on kefir microbial population by molecular techniques have been published. Mainville and her friends used RFLP and 16S rDNA sequencing approaches to identify the kefir bacterial population in 2005 (Mainville, et al. 2006). Takizawa and his friends applied DNA-DNA hybridization and 16S rDNA sequencing methods; and Chen and his friends applied a molecular technique, PCR-DGGE and 16S rDNA sequencing methods to kefir grains to capture the bacterial population of kefir and they were identified *Lactobacillus kefiranofaciens;* and *Lactobacillus kefiri*, *Leuconostoc mesenteroides*, *Lactococcus lactis*, respectively (Chen, et al. 2008, Vancanneyt, et al. 2004).

This study intends to give a vision of the bacterial composition of the kefir grains and liquid by a genomic approach which is chosen as an alternative to phenotypic identification and it involves the oligonucleotide cataloging by phylogenetic anchors. A pair of universal bacterial primer which has broad range was used as phylogenetic anchors. Although they recognize the conserved regions on genomic DNA, small subunit of 16s rRNA gene, their amplified fragments have variable regions that enable to

characterize bacteria at genus or species level (Cai, et al. 2003). Sequence-based analysis was used to obtained the sequence of the clones containing 16s rRNA gene and direct PCR amplification of the kefir bacterial population that provide extensive information about taxonomic groups and the species of mix bacterial environment in kefir.

The purpose of this study was to access and identify the bacterial population in kefir and kefir mother-culture (grains) after 3 days fermentation time by using comparative analysis of culture dependent and independent methods.

1.3. Phenotypic Classification of Microorganisms

Throughout the centuries intense effort to describing the phenotypic diversity of microorganisms has been dedicated by microbiologists. Phenotypic properties of microorganisms are commonly used to classify and determine their taxonomic place from subspecies level to the family level. Throughout the history of microbiology as a science, microbiologists realized that some pitfalls in attempting to classify prokaryotes. Although, many bacterial isolates are identified by their phenotypic characteristics such as morphological, physiological or biochemical properties of cell wall, this process might be usually laborious, when isolates have atypical biochemical or growth patterns (Bergey 2005). In addition, Prokaryotes (Bacteria and Archaea) unlike their multicellular equivalents have very restricted cellular morphological characteristics which serve little use to establish their phylogenetically valid taxonomic level (Woese 1987). Furthermore, many useful fossil records of prokaryotes are absent; however, the existing ones are uninformative (Oren 2004). As a result phenotypic properties are not sufficient enough to classify them systematically and phylogenetically. Similar phylogenetic similarities may be the result of lateral gene transfer or convergent evolution apart from originated from same ancestor (Gillis, et al. 2001).

In recent years, many new molecular methods which have been investigated to analysis of microflora have important roles as guides for the characterization of prokaryotic taxa. Investigating the presence of each bacterial population, their roles and patterns of interactions in environment is important in the identification of microbial community structure (Green and Keller 2006, Steele, et al. 2005). In phylogenetic classification each species should be part of only one genus and this genus should be part

of only one family and so forth as Order, Class, Phylum and Domain, respectively. Hence, phylogenetic classification can be defined as “inclusively hierarchical”.

1.4. Molecular Microbial Phylogenetic

In recent years, molecular studies have proven to be a useful supplement for microbial identification (Bergey 2005). Molecular microbial phylogenetic is interested in molecular sequence patterns of the microorganisms instead of their limited complex morphological, biochemical and physiological characteristics which have restricted success in describing the evolutionary relation between various groups of prokaryotes. Evolution is the theory that organisms have been chancing over time so that descendants differ structurally and functionally from their ancestors. This biological process may be defined a species by inheriting morphological and physiological characters from their ancestors. (Pevsner 2003).

The amount of sequence change that organisms accumulate is related to which “mutations become fixed” and the duration of “time” over the changes have occurred (Woese, et al. 1987). The changes in version of a given sequence can be measured by comparing the sequence with the original sequence to figure out the phylogenetic distance of them. However, the original sequence of an organism, ancestor pattern, is not accessible to scientists. That is why the scientists prefer comparing two (or more) extant representatives’ lineages which last share a common ancestor to measure the sequence differences of them and to place them in the phylogenetic tree (Woese, et al. 1987). Not all sequences have equal value for determining the phylogenetic relationship. A certain chosen molecule as a pyhlogenetic marker has to meet some properties. The sequence changes in it have to occurred as randomly as possible, the size of the molecule has to be large enough to contain sufficient information and it has to take a role as a “chronometer” during the evolution; and the rates of change have to reduce to a common measure with the spectrum of evolutionary distances to be measured (Woese, et al. 1987).

Molecular evolution is a process of mutation with selection. The Phylogenetic analysis of DNA sequences determines the descent of species from a common ancestor or determines the close relatives of extant organisms and places them into a tree representation with placement of similar sequences on neighboring framework.

Sequence similarity is a measurement of the matching characters in an alignment (Pevsner 2003). Molecular phylogenetic studies were first advanced in 1965 and until this time many gene sequence data was collected and become more abundant (Doolittle 1999). Genotypic information, sequence analysis, is more reliable and gives more precise information about evolutionary relationships than phenotypic characterization of organisms. One of the microbiologists working area, microbial ecology, is gained movement after growing phylogenetic studies (Woese, et al. 1987).

Today, living organisms are classified into three "domains" according to construction of Carl Woese and his collaborates' phylogenetics tree structure in 1987 which was established and based on the 16S rDNA gene nucleotide sequence differences and similarities rather than obvious morphological similarities (Woese, et al. 1987). In this way, the relatedness and how species have evolved can be found answer in more reliable way. The first two domains of three belongs to Prokaryotes, Archaea and Bacteria, and the third one corresponds to Eukaryotes which were evolved by separate major evolutionary pathway from a common ancestral form (Figure 1.3.) (Woese, et al. 1987).

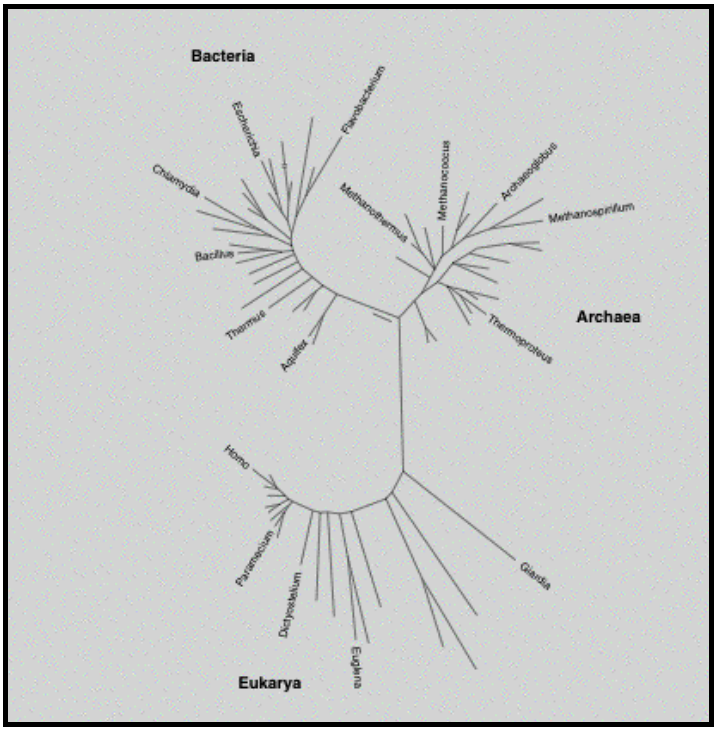


Figure 1.3. A global tree of life based on phylogenetic analysis of 16S rRNA genes (Source: Woese, et al. 1987)

The branching patterns of subordinate taxa, kingdoms and lower divisions of the three domains Archaea, Bacteria, and Eukarya are also supported by SSU rRNA gene sequences. The Bacteria have been grouped into 23 phyla, which are subdivided into 28 classes and the Archaea have been subdivided into two phyla (Figure 1.4.) (Bergey 2005) .

Bacteria which are the products of an estimated 3.8 billion years of evolution are extremely diverse organisms in many aspects of including their physiology, growth and metabolic modes and also their environmental desires such as nutrient contents, oxygen, temperature, pH requirements. Bacterial strains are grouped in a systematic way according to their characteristics that they posses by Systematic Bacteriology. This classified construct has a dynamic structure, because it may usually subject to changes and modification during the time. New species and taxa have been identifying constantly and they are added to list. Some species or taxa are transferred into new places in this list to determine their phylogenetic level more accurately in the tree. Determined higher taxa may be subdivided or some taxa may be merged together to construct more accurate pyhlogenetic tree structure (Bergey 2005).

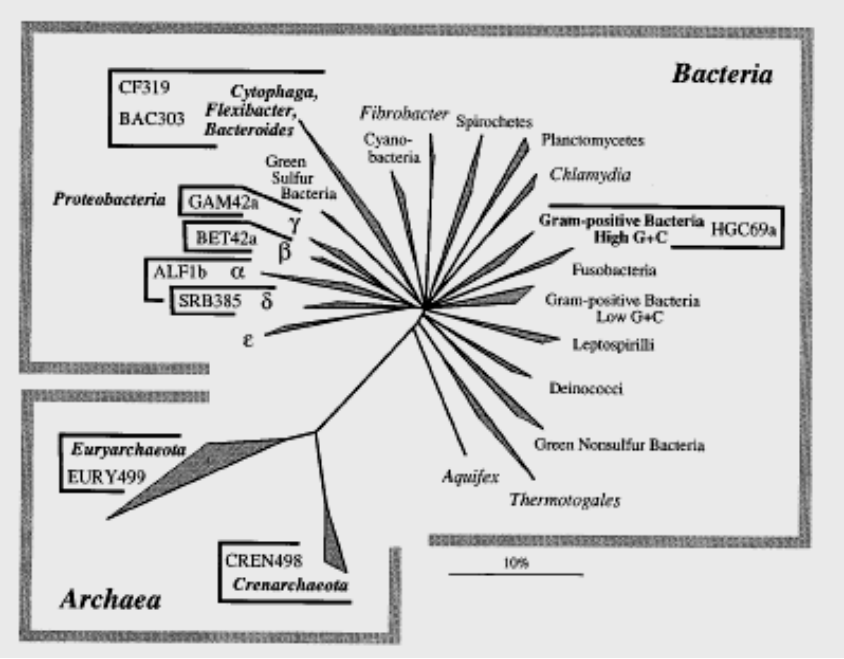


Figure 1.4. The phylogenetic tree of Bacteria and Archaea based on 16S rRNA gene sequences comparision (Source: Amann, et al. 1995)

1.5. 16S rRNA gene as a Phylogenetic Marker

In the mid-1970s, Carl Woese and coworkers began working to assemble 16S rRNA sequence information and sequence database to obtain phylogenetic relationships of microorganisms to form an opinion about their natural classification on the information available. In mid-1980s they defined that rRNA gene as the "evolutionary chronometers". Because all the extant living things have rRNA, it is presumed that rRNA dates back to the earliest forms of life and it can be used to reflect the evolutionary relationships between all species on earth (Pace 1997). Woese considered "chronometer" as a molecule whose sequence changes randomly in time (Woese 1987). The prokaryotes ribosome has two subunits; the small subunit (SSU) which has 16S and the large subunit (LSU) which has 23S and 5S sedimentation coefficients. 16S rRNA or its gene coding sequence that has alternating conserved sequence regions usually preferred to characterize bacteria at genus or species level (Bergey 2005). The ribosomal nucleic acids are found in every living cell and the growing bacterial cells contain about 10^4 to 10^5 copy of 5S, 16S and 23S rRNAs. The 16S and 23S rRNA genes contain many variable conserved sequence motifs that reflect phylogenetic origin of microorganisms (Bergey 2005).

The PCR amplification primers of 16S rRNAs can be designed according to conserved sequence region (Figure 1.5.) and the sequence heterogeneity of this gene identified to deduce the phylogenetic hierarchical differentiation (phylum, family, genus, species) by compared sequences (Bergey 2005). There are ten variable regions (hereafter defined as V) which can be used as promising tool for the detection and identification of microbial species and their phylogenetic relationships much more easily and more accurately than was ever (Woese 1987). The identification and phylogenetic taxonomy of desired bacterial groups are possible by sequencing of partial or total 16S rRNA gene (Cai, et al. 2003).

In this study, partial region of 16S rDNA (V3, V4, V5 and V6) was amplified (~780bp) by E334 and E1115 primer pair to identify the bacterial communities in kefir and the sequence analysis was used to determine their phylogenetic relationships.

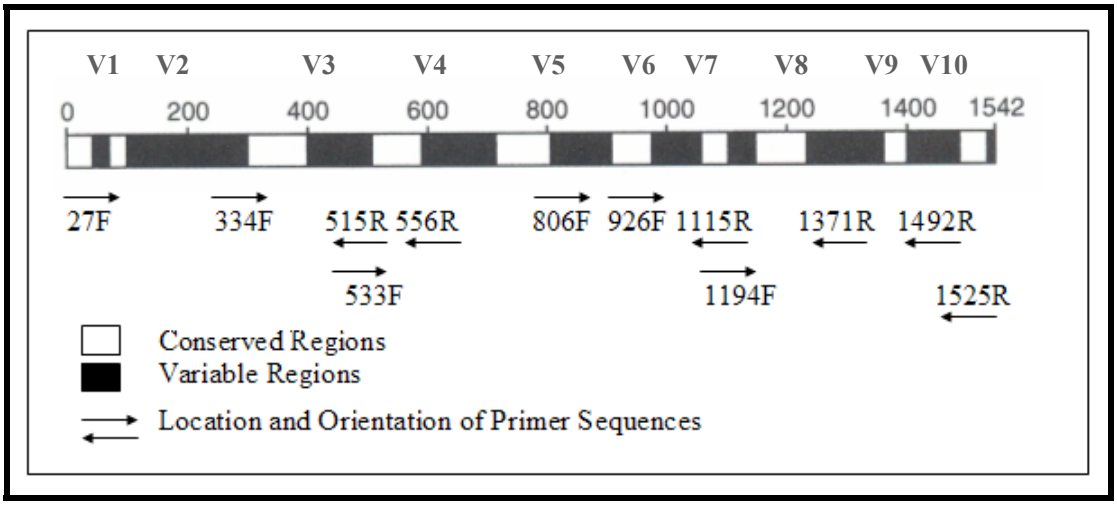


Figure 1.5. A simplified map of the conserved and variable region of 16S rRNA gene and sequencing primer location (Source: Cai, et al. 2003)

About ten years after studies on revolutionary effect of Woese on microbial systematic, Pace and his colleagues started to use 16S ribosomal RNA sequences to determine the phylogenetic characteristics of the species. They were detected the microbial diversity in the environment by direct sequencing of 16S rRNA genes of microorganisms instead of using culturing methods (Handelsman 2004 and Doolittle 1999).

There are many features that explain why rRNAs are confidentially used as "molecular chronometers" for phylogenetic identification of species. They are the most abundant molecules in all extant cells and they have a fundamental function as protein-synthesis. They are ancient molecules that date back to the earliest life of form and they consist of alternating regions that range from universally conserved regions to hyper-variable regions at the nucleotide sequence level across all phylogenetic domains (Amann, et al. 1995). Some parts of these universally conserved regions might be dated back to RNA world where all biochemical reactions were occurred by RNA (Smit, et al. 2007). Homologous sequences that are used as universal signature in phylogenetic analysis are preferred to design universal primers for the three domains or their suborders (Giovannoni, et al. 1988). Because the sequence changing has been occurred at different position on this macromolecule and the changing frequency is very varying through the evolutionary time, it allows measuring a species in phylogenetic relationships. rRNA genes are free from artifacts of the lateral gene transferring through

distinct species that this property gives it a advantage to be used as a true phlylogenetic marker (Amann, et al. 1995). 16S rRNA sequence analysis results can be used not only identification of the slow growing, fastidious and unusual microbial population but also microbial population which are difficult to differentiate by biochemical assays in pure culture (Cai, et al. 2003, Hayden, et al. 2001).

Taken into account, all these features of rRNA genes are made them favorable to be used as a phylogenetic tool to determine relationship between species. The size of 5S rRNA genes is relatively small (~120 nucleotides) that usually could not be preferred to be used as a discriminative phylogenetic marker. The size of 16S rRNA genes is approximately 1,500 nucleotides and 23S rRNA genes length is about 3,000 nucleotides that partially or fully sequence analyzing results are represented reliable information to determine phylogenetic relationships between species. 16S rRNA genes consist of about 50 helical stalks in their secondary structure (Figure 1.6.) and 23S rRNA genes have twice stalks of 16S rRNA genes (Amann, et al. 1995).

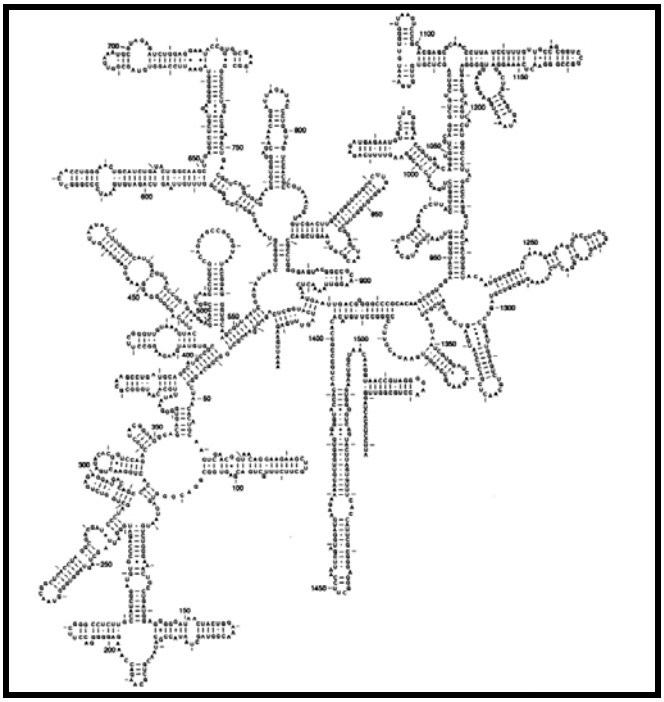


Figure 1.6. Secondary structural diagram for a representative bacterial (*E. coli*) 16S rRNA sequence (Source: Woese, et al. 1990)

Although, 23S rRNA genes have more phylogenetically meaningful region in number than does the shorter 16S rRNA molecule and greater resolution, the major

drawback of this molecule is the currently limited databases (Bergey 2005). *E.coli* is the first microorganisms of which complete nucleotide sequences of SSU and LSU of rRNA gene was determined by the help of Sanger sequencing method in 1978 and 1980, respectively (Brosius, et al. 1978 and Brosius, et al. 1980).

There are also other genetic targets that can be used instead of rRNA genes as indicative phylogenetic macromolecules to molecular determination of the structure of microbial communities. Internal transcribed sequences of 16S-23S rDNA genes; and housekeeping genes such as *groEL* (Tsai, et al. 2005), *rpoB* (Drancourt and Raoult 2002), *recA* (Blackwood, et al. 2000), *gyrB* (Dauendorffer, et al. 2003) and hsp 60 heath shock protein (Goh, et al. 1996) can be used to differentiate bacterial genus or species. The conserved nucleotide sequences in housekeeping genes are very low that causes difficulties while designing PCR primers, however, if it is designed, it can be used to differentiate taxonomically closely related species successfully. Unfortunately, their databases have not grown as fast as SSU rRNA and they have limited sequence information content. It is the fact that the spectrum of LSU rRNA database is superior the all other protein markers mentioned above (Bergey 2005).

1.6. The Reconstruction of Phylogenetic Tree

A phylogeny is a reconstruction of the evolutionary history of the collection of organisms. The phylogenetic trees which are graphical representation of a multiple sequence analysis are constructed to represent the historical relationships of groups of organisms. Each group is called a taxon. Traditionally morphological characteristics data from extant data and fossil records are preferred to draw a phylogenetic tree but today the sequence analysis of proteins or nucleic acids are used almost interchangeably with taxa (very often species) by the advent of molecular sequencing (Hall 2004).

During the last decade molecular microbiological approaches increased very rapidly and one of these approaches is the 16S rRNA sequence analysis which has provided new insight into the microbial ecology to understand the structure of microbial communities and their phylogenetic relationships. It is accepted that the SSU rRNA is currently the most meaningful phylogenetic marker in terms of information capacity, depth of taxonomic resolution and database size (Bergey 2005). As a result, SSU rRNA

is one of the most widely used classification methods in prokaryotic identification and systematic.

The comparative partial sequence of an appropriate region of a gene may be sufficient enough to determine the phylogenetic group of the organisms, their close relatives and to indicate a novel species (Mount 2004). The phylogenetic information content of bacterial 16S rRNA is predicted by using the result of calculated *E.coli* 16S rRNA sequence conserved and variable regions as a reference. There are 974 (63.2%) variable hence informative region and 568 (36.8%) conserved region in *E.coli* 16S rRNA gene. So the assumption is that the reliability of the partial data of SSU rRNA is very high to be used as informative in terms of phylogenic determination of bacteria (Bergey 2005).

The relationships between two given base sequences are determined by the number and character of positional differences of bases that have accumulated during the course of evolution. Then this primary data is used to analysis phylogenetic diversity according to criteria such as variability or likelihood. The results are used to construct trees of which terminal (the organism) and internal nodes (the ancestor) are connected by branches. The path of evolution is indicated by the branching pattern of tree and the phylogenetic distance between two species can be measured by the additive length of peripheral and terminal branches. The number of substitution per site that have take place along a branch is called as phylogenetic distance. The easiest way to find relationship (similarities) between these sequences is to use a program that will search the entire international nucleic acids databases for similar sequences through World Wide Web.

There are different styles of drawing phylogenetic trees. Whereas, a cladogram shows only the branching order of nodes, a phylogram displays both branching order and distance information. In the phylogram, the taxa appear close to each other are more evolutionary related ones. In cladograms, the taxa are spread out that allow visualizing the branch length and labels more clearly (Hall 2004).

There are three types of most commonly used treeing methods from sequence data; distance-based method, distance matrix; and character-based methods, maximum parsimony and maximum likelihood (Bergey 2005).

The “Neighbor joining” (NJ) is a distance method which is based on the number of differences between pairs of sequences. NJ uses evolutionary distances data to reconstruct the phylogenetic tree was developed in 1987 by Saitou and Nei (Saitou and

Nei 1987). The NJ method is a clustering method that produces an unrooted additive tree from series of matrices. Unrooted trees never give information about the direction of evolution- the order of descent. A distance matrix is said to be additive if the distance between the species on the tips of the tree are equal to sum of the length of the branches (Hall 2004). All the OTUs (operational taxonomic units) are first clustered in a star like tree. The two most closely related sequences are then determined by the pair-wise comparisons. These sequences are called as neighbor and they are connected through a single node. These neighbors are connected to the other OTUs by an internal branch. This process is continued progressively until tree topology is constructed. Since the NJ algorithm minimize the sum of branch length at each stage of clustering, NJ is called as "minimum evolution" method. NJ is known to produce true tree with probability as the sequence length goes to infinity (Bergey 2005). Thorough the literature review, it is known that many researches preferred this approach to analysis bacterial population. Hayden et al. used NJ dendogram to analysis cultured-derived spiral bacteria (Hayden, et al. 2001). Song et al. determined phylogenetic relationship of nine clinically significant type strains with their related established species by using NJ (Song, et al. 2003).

The main idea behind the Maximum Parsimony (MP) Method is the minimization of the total number of character changes to explain the data in the alignment (Felsenstein 1988). All parsimony methods that reconstruct hypothetical common ancestral sequences at internal nodes are not always statically consistent unlike NJ. Parsimony is a more powerful method than distance when considered the hierarchical relationships of genes (Hall 2004). Parsimony-informative characters must have at least two states that occur in at least two taxa. The tress which is constructed by using a criterion of parsimony can be drawn as a phylogram or a dendogram. All taxa evolved at the same rate and all characters have same amount of information in terms of the parsimony algorithms. Two taxa which were rapidly evolved are placed together on a tree drawn the by MP not because they are closely related, but because both of them have many mutations. This phenomenon is called as long-branch attraction (Higgs and Attwood 2005).

The other character-based method is maximum likelihood (ML) which looks for the tree that maximized the likelihood of observing the data. Three different types of parameters must be known to calculate the likelihood of a sequence set on a tree; the branch length, tree topology and value of the parameters such as base frequencies etc. All three types of parameters can be changed by many programs to find the ML solution

(Hall 2004). It is possible to estimate the tree topology and branch lengths by allowing the rate parameters to vary. The evolution at different sites and along distant lineages must be statically independent to be analyzed by ML. Maximum likelihood is a particularly suitable choice to analysis the distantly related sequences. ML programs are limited to a small number of sequences. Since it searches all possible tree topology combinations and branch lengths, ML is computationally intense method if there are more than few sequences (Mount 2004).

The flowchart (Figure 1.7.) below indicates the consideration while choosing the most appropriate methods out of three possible ones for phylogenetic prediction.

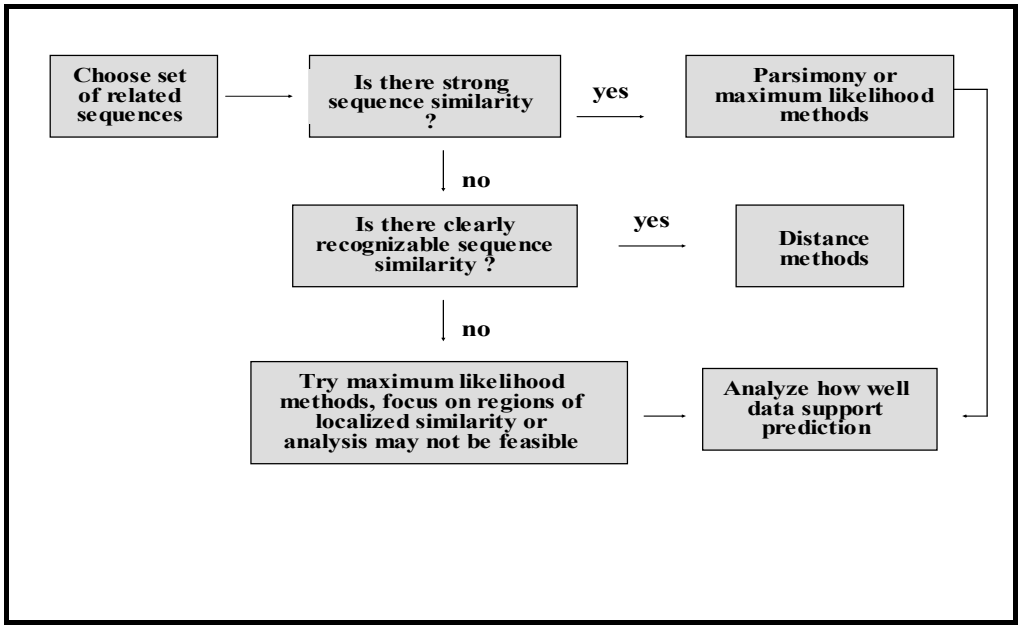


Figure 1.7. The flow chart of the phylogenetic prediction methods
(Source: Mount 2004)

1.7. Bioinformatics

Bioinformatics is a combination of science and technology of which focused on developing researches and application of computational tools for learning and expanding biological data and also including storage, organization, analyzing and visualizing of these data. The discipline of bioinformatics focuses on the analysis of molecular sequences and functional genomics. Computer algorithms and computer databases are the tools for bioinformatics approaches to analysis of molecular sequence data. Internet

is an important source for bioinformatics, because it is a major place to access sequence data, software and also it is a place to integrate different types of resources and information related to molecular biology. There are three main databases which can be publicly accessible and they serve as DNA and protein data storage. The DNA DataBank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL), and GenBank at NCBI (National Center for Biotechnology Information 2008) which are the part of International Nucleotide Sequence Database Collaboration are the three main databases (Table 1.4.). These three organizations exchange sequence data on a daily basis.

Table 1.4. Three main bioinformatics web servers that serve as DNA sequence storage

Resource	**Description**
DNA Databank of Japan (DDBJ)	Associated with the Center for Information Biology
European Bioinformatics Institute (EBI)	EMBL database
Natural Center for Biotechnology Information (NCBI)	GenBank

Natural Center for Biotechnology Information (NCBI) is chosen as a server to analysis this thesis sequence experiment results (National Center for Biotechnology Information 2008).

1.7.1. Base Sequence Analysis

Sequence analysis is the basic tool to determine the relationships between sequences due to their functional, structural or evolutionary properties. Phylogenetic analysis of sequences efforts to reveal evolutionary relationships among a set of similar sequences for classifying them into a tree representation with act of laying most alike sequences on neighboring places. Multiple sequence alignment is a starting point for evolutionary modeling such as most probable phylogenetic tree reconstruction. Three or more sequences are chosen to place in the same column of alignment allowing for mismatches and gabs to perform multiple sequence alignment. For each alignment result, the sequence similarity scores are given which indicate the most closely related sequences. Sequence similarity is the fraction of aligned positions in a sequence

alignment. There are many methods for reconstructing phylogenetic tree as mentioned above and it should be considered that they may give almost the same result. Hence, test of significance should be derived to determine the existence of particular tree (Altschul, et al. 1990).

Sequence databases are the collection of sequences with information about each sequence in each data entry. There are many fast and rapid searching methods. The first important algorithm is FASTA format which found short common patterns between query (sequence that is to be aligned) and target databases sequences. It eliminates unlikely segment from alignment. Then it reports the best-matched sequences and local alignment score. One of the most popular servers on World Wide Web is FASTA Sequence Comparison which is designed by University of Virginia (UVA Fasta Server 2008).

BLAST (Basic Local Alignment Search Tool) is the second algorithm which is very similar to FASTA format. BLAST format is more popular due to its speed and sensivity by searching and availability of the program on the World Wide Web through a large server at the National Center for Biotechnology Information (NCBI) (Wheeler, et al. 2000). Many types of BLAST are possible to compare all combination of protein and nucleic acids queries with databases (protein or nucleotide) (McGinnis and Madden 2004).

1.7.2. Natural Center for Biotechnology Information (NCBI)

NCBI is a public database which manages research in computational biology and produce software tools for analyzing nucleotide and protein sequences. There are seven major categories in NCBI; PubMed, Entrez, BLAST, OMIM, Books, Taxonomy and Structure. BLAST will be mentioned in detail; because the obtained16S rRNA gene sequences of Kefir bacterial population are analysis by BLAST in this thesis study.

1.7.2.1. BLAST

A set of sequence similarity search program of NCBI which is called BLAST (Basic Local Alignment Search Tool) is used to analysis of DNA or protein sequences. The BLAST search identifies similarity of the input query by comparing it with

sequences which are deposited in a database. BLAST which is a family of programs is required four components to perform a BLAST search. Nucleotide BLAST program is chosen for searching a nucleotide database against nucleotide query (Altschul, et al. 1990).

First step is a sequence of interest is chosen and pasted into the BLAST input box. Second step is the chosen of one of BLAST programs (blastp, blastn, blastx, tblastx, tblastn). Blastn program should be used to compare the both stands of nucleotide sequence query against a DNA database. Thirdly, a database is chosen among many of them. Nonredundant (nr) databases chosen for DNA sequence comparison are usually preferred and they contain GenBank, EMBL, DDBJ databases. At the forth step optional parameter can be chosen to modify output such as filtering low-complexity of sequences, restriction of searching to a specific set of organisms (Figure 1.8.). After clicking BLAST button the comparison results are appeared as Figure 1.9. The BLAST queries entered to sever has been increasing very rapidly about 100 000-140 000 per weekday and its computing power is improved day by day (Pevsner 2003).

Hayden et al. used NCBI database and its BLAST algorithm to analysis of cultured-derived spiral bacteria by 16S rRNA gene sequence (Hayden, et al. 2001).

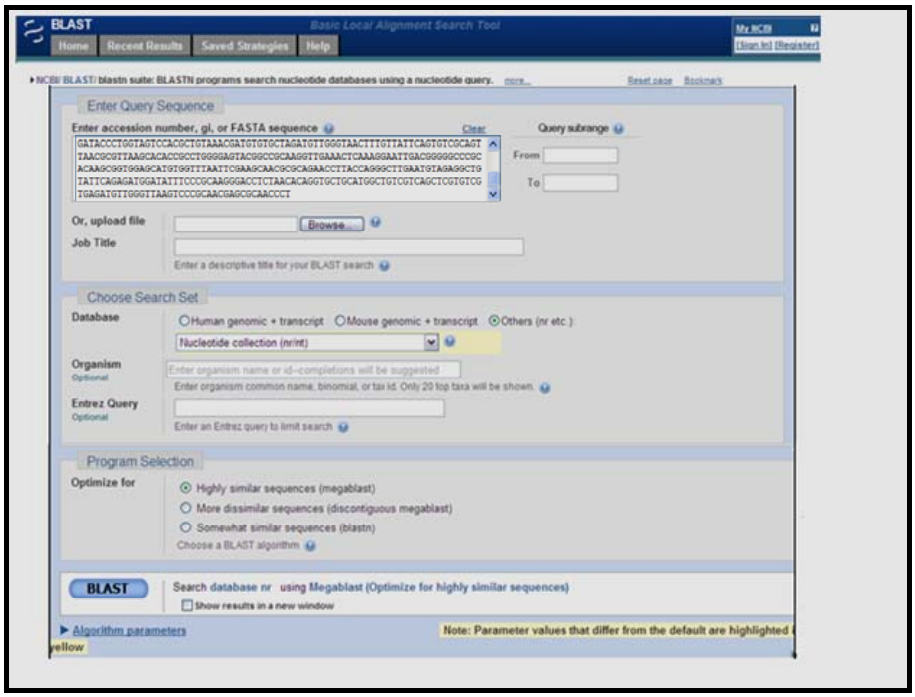


Figure 1.8. Performing a nucleotide BLAST search against a DNA database (Source: Basic Local Alignment Search Tool 2008)

In the top of the BLAST output page, the type of BLAST search, description of the query and the databases that were search are indicated as the summary of the process. Alignment score is overviewed by a graphical representation and each red bar indicate the database nucleic acid sequence that matches the query. The distribution of BLAST hits upon to similarity range means that the most similar hits are placed at the top of this liner map. The score scheme describes the level relatedness between query and subject (particular database match which is aligning to query) (Figure 1.9.). The statistical significance of BLAST program is verified by quantitative measurement of whether the base similarity occurred by chance. Hence, the E- value describes the random background noise. It estimates the false-positive results from BLAST searching. Length of query sequence and length of database is the affecting factor for E value. Lower E-value causes increasing probability of the alignment occurrence (Mount 2004).

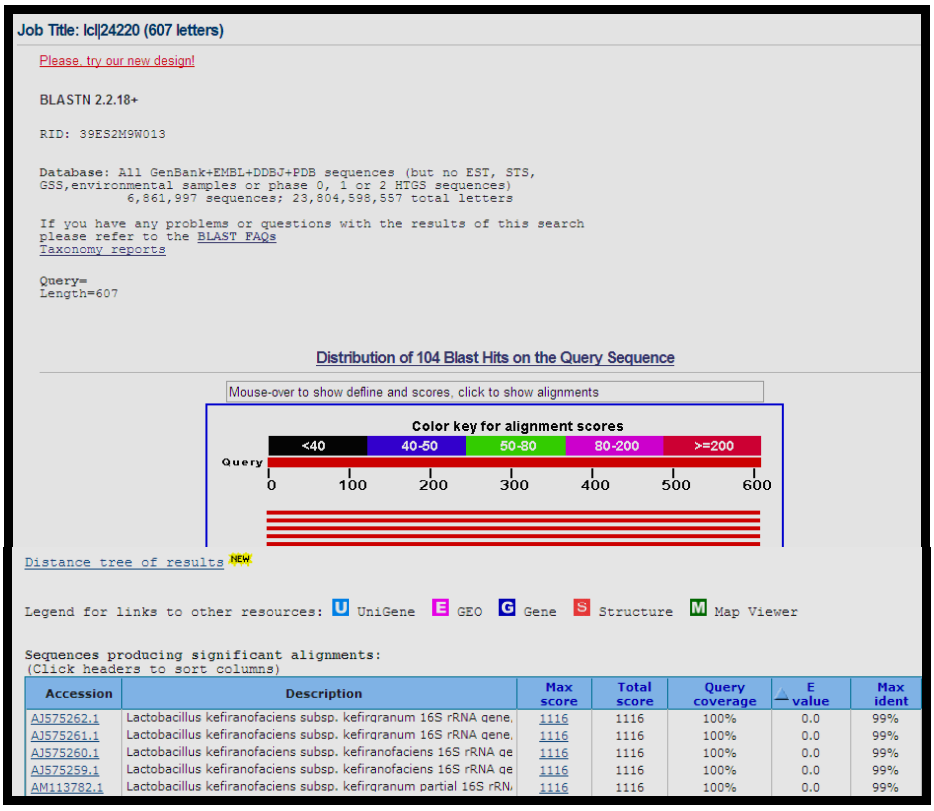


Figure 1.9. Graphical output of a sequence result and list of similar sequences from subject (Source: Basic Local Alignment Search Tool 2008)

1.8. The Aim of this Work

The aim of this work was to identify the bacterial communities in kefir drink and kefir grains; and to determine the phylogenetic relationships between them through the construction and screening of 16s rRNA gene libraries of the potential genomic pool. In this study, 16S rRNA gene sequence analysis was applied. A pair of universal bacterial primers specific to 16S rRNA genes were used to amplify partial region of 16S rDNA of bacterial communities that captured by culture-dependent and independent approaches. The amplified fragments of 16s rDNAs have variable regions that enable to characterize bacteria at genus or species level. Sequence-based analysis which were used to obtain the sequence of the clones containing 16S rDNA of the kefir bacterial population provide extensive information about their phylogenetic relationships, taxonomic groups and the species of mix kefir bacterial communities. The other purpose of this work was to determine the reliability of culture-dependent and independent approaches to capture the bacterial population from a mix microbial environment.

CHAPTER 2

MATERIALS AND METHODS

2.1. Materials

A detailed list of commonly used chemicals, buffers, solutions and their compositions are presented in Appendix C.

2.2. Methods

2.2.1. Kefir Fermentation

Kefir grains used in this study were house hold origin. Fermentation of milk by kefir grains was achieved by adding 50 g kefir grain into 500 ml UHT cow's milk in a sterile jar followed by incubation without shaking for 3 days at 28°C (Figure 2.1.). The kefir grains from the fermented milk were collected by using a sterile sieve and rinsed with sterile water to be used as a starter for the next cycle of fermentation process.

Figure 2.1. 3 days fermented milk by kefir grains in a sterile jar

2.2.2. Homogenization of Kefir Grains

Kefir grains inoculated into milk during 3 days were sieved and washed with sterilized water (Figure 2.2.). 2 gram grains were homogenized in 18 ml sterile distilled water at a setting of 3rd level speed for 4 minutes (IKA Ulltra Turrax Homogenizator, T18 basic, Wilmingtan, USA). Homogenates were used for isolation of single colonies as describe below.

Figure 2.2. Sieved and rinsed kefir grains

2.2.3. Isolation and Maintenance of Bacterial Strains in Kefir Grains and Kefir Liquid

Fermented liquid kefir and their grain homogenates were serially diluted to 10^{-10} and 150 µl from each dilution from $1/10^4$ to $1/10^{10}$ were plated on MRS (MERCK, KGaA Dormstadt Germany) agar medium (Giraffa 2003). Cultured plates were incubated at 28°C in aerophilic and anaerophilic conditions. Anaerophilic conditions were created by an anaerophilic jar and sachets (OXOID Anaero*Gen*^{Tm}, OXOID Anaero*Jar*). The inoculated media plates were placed in the jar and the temperature was kept at 28°C. According to manufacturer's instructions the carbon dioxide level in the jar was between 9% and 13% during incubation time. Following the 3-5 days incubation time for aerophilic and anaerophilic cultivation, observed colonies were chosen according to unique morphological characters and plated on fresh MRS plates to obtain pure cultures.

For each isolated culture, glycerol stocks concentration at 20% (v/v) were prepared in the 1.5 ml vials and frozen at -80°C for long term storage.

2.2.4. Genomic DNA Extraction

2.2.4.1. Genomic DNA Extraction of Isolated Bacterial Cultures

For culture-dependent analysis, genomic DNA samples for using in the construction of 16S rDNA libraries were collected from each individual isolate stored at –80°C from kefir drink and grains as mentioned above. They were activated in 10 ml MRS broth without shaking for 3-5 days at 28°C. Genomic DNA was extracted from each isolate by the Genomic DNA Purification Kit which enables to isolate both Gram-positive and Gram-negative bacteria (Fermentas). 10-20 mg of bacterial cultures was used for the genomic DNA isolation. In this procedure, DNA isolation is based on lysis of the cells with detergent and followed by selective DNA precipitation. The isolation procedure was followed according to instructions of manufacturer. Isolated genomic DNA concentration measured visually after electrophoresis in 1% agarose gel.

2.2.4.2. Total Genomic DNA Extraction from Kefir Grains and Kefir Liquid

For culture-independent analysis, total genomic DNA was isolated from fermented liquid kefir and kefir grains by using a Power Soil DNA Kit (MOBIO) which causes cell lysis by a combination of mechanical methods (bead beating) that was indicated to be very effective at lysing bacterial cells (Schloss, et al. 2005) and chemical methods. After 3 days fermentation time 0.25 mg kefir grains sample which were sieved and 5 times rinsed with sterilized water and 3 days fermented 500 µl liquid kefir sample were loaded into PowerBead Tube for the homogenization and lysis procedure. FastPrep FP120 (Bio101 Savant Instruments, Inc., Holbrook, NY) device at a setting at maximum speed for ten minutes was used for disruption of cell membrane.

The following procedure was applied according to the manufacturer's recommendation.

2.2.5. PCR Amplification of Partial 16S rRNA Gene

Small-subunit (SSU) rRNA genes of kefir bacterial population captured by culture-dependent and culture-independent approaches were amplified from their purified genomic DNA by a pair of universal bacterial primer. A 781 bp region of the 16S rDNA from genomic DNAs were amplified by PCR using these universal bacterial primers, E334F as forward primer and E1115R as reverse primer (Bacer et al., 2003). PCR reactions were performed in a thermocycler, iCycler Thermal Cycler (Bio-Rad Laboratories, Inc.). PCR were performed with PCR Master Mix (2X) (Fermentas) that is a 2X concentrated solution of 0.05 u/µl *Taq* DNA Polymerase (recombinant), reaction buffer, 4mM MgCl2, 0.4 mM of each dNTPs (dATP, dCTP, dGTP, dTTP) except DNA template and primers. *Taq* DNA polymerase in Master Mix (2X) generates PCR products with 3'-dA overhangs which are compatible with TA cloning. 1 µl of 10 times diluted total genomic DNA template, 25 µl of (2X) PCR Master Mixture (Fermentas) and 0.5 µM of each primer were mixed in total of 50 µl. The protocol used consistent of an initial denaturation step at 94°C for 2 minutes, followed by 25 cycles consisting of denaturation at 94°C for 30 seconds, primer annealing at 62°C for 30 seconds, and elongation at 72°C for 1 minute plus an additional final extension at 72°C for 5 minutes to finalize the chain reaction. After these PCR products were visualized by agarose gel electrophoresis and purified by PCR purification Kit they would be used in ligation reaction.

Base sequence of E334F and E1115R primer pairs are given in Appendix B.

2.2.6. Agarose-Gel electrophoresis

The PCR products were visualized by agarose gel electrophoresis by using 1% (w/v) agarose gel in 1X TAE electrophoresis buffer (0.04 µM Tris-acetate and 0.001 M Na2EDTA). 0.5 gram agarose was dissolved in 50 ml 1X TAE buffer in microwave until a clear solution was observed. After gel was cooled, 0.5 µg/ ml EtBr from a 10 mg/ml stock was added to agarose gel. Following the gel pouring into horizontal gel apparatus, combs were placed in it and it was allowed to solidify at room temperature. Combs were removed and DNA samples mixed with 6X loading dye were loaded into wells of gel in 1X TAE electrophoresis buffer. The gel was exposed to an electric constant at 80V for

45-50 minutes until bromophenol blue presence exceeded half of the gel. Finally, gel was exposed to UV in gel documentation system to visualize the DNA bands.

2.2.7. PCR Product Purification of 16S rDNA Genes

The free PCR primers were separated from the PCR products by PCR purification Kit (NucleoSpin Extact, MACHEREY-NAGEL). The quantity of each samples were checked by measuring the ratio of absorbancy at 260 nm and 280 nm. The ratio of A260/A280 was between 1.8 and 1.9 which means it was highly purified DNA sample. NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, Inc., Wilmington, DE, USA) was used to measure the absorbancy.

2.2.8. The Cloning and the Library Construction of 16S Ribosomal SSU Genes

2.2.8.1. Protocol for Ligation Reaction

The amplified and purified 16S rDNA PCR fragments were cloned by using TA cloning vector, pGEM-T Easy Vector System (Promega). All ligation reaction was set up in 10 µl and incubated 1 hour at room temperature. The cocktail contains 1 µl T4 DNA ligase (3 Weiss units/µl), 1 µl pGEM-T Easy vector (50 ng), and 5 µl 2X Rapid ligation Buffer of T4 DNA ligase. The appropriate insert volume was adjusted according to Insert:Vector molar ratio which optimized as 3:1 and it was calculated for each reaction by using the following equation.

$$\frac{\text{ng of vector} \times \text{kb size of insert}}{\text{kb size of vector}} \times \text{insert:vector molar ratio} = \text{ng of insert}$$

2.2.8.2. Bacterial Transformation

2.2.8.2.1. Protocol for Preparation of Competent Host cell, *E.coli*

E.coli Dh5α strain was chosen to prepare the cold treated competent host cell. Glycerol cell culture stock of *E. coli* strain at – 80°C was activated by plating on a LB agar plate and incubated overnight at 37°C. Chosen a single colony from this plate was preincubated in 10 ml LB broth for 6 hours in the incubator-shaker at 200 rpm at 37°C. Following this preincubation step, 10 ml bacterial culture was transferred into 200 ml SOB medium and allowed to incubate in orbital shaker (200 rpm) overnight at 10°C in order to enable the culture reach the mid-log phase. The next day, bacterial culture in SOB medium was placed on ice for 10 minutes. Aliquot cultures in 5 ice-cold 50 ml falcon tubes with the volume of 40 ml were centrifuged at 4000 rpm for 10 minutes at 4°C to pellet the cells. After discarding supernatant completely, the cell pellet was resuspended in 5 ml ice-cold TB and centrifuged (4000 rpm) for 10 minutes at 4°C. Supernatant was removed again and the pellet was gently resuspended in 2.5 ml ice-cold TB and 300 µl DMSO mixtures by pipetting. 100 µl of final cell suspension mixture was divided into sterile eppendorf tubes which already were kept cold on ice.

The prepared competent cells were either use immediately or store at -80°C for long term storage without a significant loose of competence.

2.2.8.2.2. Protocol for Transformation of Cloned pGEM-T Easy Vector into *E.coli* Competent Cells

Previously prepared frozen competent cells were removed from –80°C and placed on ice until just thawed (about 5 minutes). 4µl of each ligation reaction was transferred into 100 µl competent *E.coli* cells and the tube was gently flicked to mix. Then it was placed on ice for 10 minutes. At the next step, the cells were heat-shocked for 30 minutes in a water bath at exactly 42°C followed by incubation on ice for 2 minutes. 250 µl room temperature SOC medium was added to the tubes containing cells transformed with ligation reaction at previous step and allowed to incubation at 37°C for 1 hour with shaking (190 rpm).

2.2.9. Screening Transformants for Inserts Presence

20 µl transformation cultures were plated onto LB plates containing 100 µg/ml ampicillin, 0.5 mM IPTG and 80 µg/ml X-Gal and incubated overnight at 37°C. Because the effect of a successful insert in the pGEMT-Easy vector interrupts the β-galactosidase enzyme synthesis; recombinant clones can be identified by color screening on indicator plates. Clones that contain PCR product produce white colonies in most cases. White colonies were chosen for the following experiment, colony PCR.

2.2.10. Colony PCR

The colony which was supposed as positive clone (white colony) was subjected to colony PCR to verify the insert presence in the vector. Partial 16 rRNA gene fragments being in supposed positive clones were identified and verified by plasmid specific primers SP6 and T7. To set up the PCR reaction, 12.5 µl PCR Master Mix (2X) (Fermentas), 1 µl of SP6 Primer (25µM), 1 µl of T7 Primer (25µM), 10.5 µl dH_2O and white colonies that picked up from plates via pipette tip and suspended into 25 µl PCR reaction mixture employed as template were put into a PCR tube and PCR amplification reaction cycle was designed as following profile: 94°C /7 min; 94°C /1 min, 50°C /1 min, 72°C /1 min for 25 cycles; 72°C /5 min; hold at 4°C.

Base sequence of SP6 Primer and T7 Primer primers are given in Appendix B.

2.2.11. Isolation of Recombinant Plasmid DNA from *E.coli* Host Cells

Recombinant *E.coli* culture was subjected to SDS/alkaline lysis by using GeneJet Plasmid Miniprep Kit (Fermentas) to isolate plasmid DNA.

The single positive colonies whose insert integration into vector which verified by colony PCR were streaked onto fresh LB plate containing selective antibiotic, ampicillin at the concentration of 100 µg/ µl, and they were allowed to incubation at 37°C overnight. The next day, the colonies from an overnight plate were cultured in 10 ml LB medium supplemented with the selective antibiotic at the concentration as mentioned above with shaking (190 rpm) overnight at 37°C.

3 ml of recombinant *E.coli* culture were harvested at 8000 rpm in a microcentrifuged for 2 minutes. Supernatant and all remaining medium was discardedand cell pellet was resuspended by pipetting in 250 µl Resuspension Solution until no cell clumps remain. After addition of 250 µl of the Lysis Solution, tube was inverted 6 times until the solution becomes viscous. 350 µl of the Neutralizaion Solution was added immediately and mix thoroughly by inverting 6 times. Cell debris and chromosomal DNA were pelleted by centrifugation for 5 minutes at 14000 rpm. The supernatant was transferred into GeneJET spin column by pipetting and subjected to centrifugation for 1 minute. Supernatant was discarded and the spin column washed twice with 500 µl Wash Solution. The flow-through was discarded and additional centrifugation was done to remove residual ethanol. GeneJET spin column was transferred into a sterile eppendorf tube and 50 µl sequencing water (Sigma) was added into center of column and after incubation at room temperature the plasmid isolation was eluted by centrifugation at 14000 rpm for 2 minutes.

The quantity and quality of the purified samples was checked in NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, Inc., Wilmington, DE, USA) to be ready to use in sequencing experiment.

2.2.12. Dye terminator Cycle Sequencing Reaction

Some Cloned 16S rRNA gene fragments of kefir bacterial population into pGEMT-Easy Vector were sequenced by CEQ™ 8800 Automated Genetic Analysis System (Beckman Coulter, Inc., Fullerton, CA, USA) and some of them were sequenced by ABI PRISM® 310 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA)

.

2.2.12.1. Plasmid DNAs Preparation as Templates for Sequencing Reaction

The Amount (moles) of template plasmid dsDNA used in the sequencing reaction was determined according to manufacturer recommendation. 60 ng template DNA for 100 fmol concentration was used in the sequencing reaction.

After template DNA diluted to appropriate concentration in deionized water, it was subjected to pre-heat treatment and heated to 96°C for 1 minute to improve both signal strength and current stability.

2.2.12.2. 16S rDNA Gene Sequencing Reaction

The plasmid DNAs with which full-sized insert were partially sequenced by using plasmid specific T7 and SP6 primers.The sequencing reaction Premix (1100 µl) was prepared according to dITP chemistry as follows; 200 µl of 10X Sequencing Reaction Buffer, 100 µl of dNTP mix, 200 µl of ddUTP dye terminator, 100 µl of ddGTP dye terminator, 200 µl of ddCTP dye terminator, 200 µl of ddATP dye terminator and 100 µl of polymerase enzyme. The 180 µl of mixture aliquot was transferred into sterile microcentrifuge tubes and stored at -80°C for long term usage.

DNA sequencing reaction was set up in 20 µl final reaction volume in thin-wall tube and all reagents were kept on ice during experiment. The components; H_2O (to adjust total volume to 20 µl), 1-7 µl (60ng template DNA), 1.6 µM sequencing primer (T7 or SP6 for each sample in separate reaction) and 11 µl premix were added into reaction, respectively. A control reaction was designed and 0.5 µl pUC18 was used as a template and M13 was chosen as primer in this reaction, the other components were same as mentioned above. Thermal cycling program was as follows; 20 seconds denaturation step at 96°C, 20 seconds annealing at 50°C and 4 min extension at 60°C for 30 cycles followed by holding at 4°C.

2.2.12.3. Ethanol Clean-up in Sample Plate

DNA sequencing reaction samples were transferred into 96 well-plates. Prepared fresh stop solution (Appendix C) and 1 µl 20 mg/ ml of glycogen was mixed thoroughly. 5 µl stop solution/glycogen mixture and 60 µl 95% ethanol was added to each sample, respectively. 96 well-plates were sealed with adhesive foil and vortex briefly. Plates were centrifuged at maximum speed for 30 minutes at 4°C. Supernatant was discarded by the action of up-side down of the plate and it was placed on bed of tissue. The plates at that position subjected to centrifugation at 10x g for 20 seconds. After 200 µl 70% ethanol was added onto samples, the plate was centrifuged at maximum speed for 5 minutes at 4°C. Supernatant was removed and ethanol washing step was repeated one

more time. The next step was the plate placed up-side down on bed of tissue and centrifuged at 10x g for 20 seconds. Traces of ethanol remain was removed by SpeedVac System for maximum 10 minutes. Each sample pellet was re-suspended in 40 µl Sample Loading Solution by pipetting. A drop of light mineral oil was added to each sample and plate was loaded into the instrument.

2.2.13. Comparative Sequence Analysis of 16S rRNA Genes

Partial sequence of the 16S rRNA of each clone or isolate generated by each sequencing primers was merged together into a one complete sequence. Sequence results were analyzed by comparing each base sequence of sample with the 16S rDNA sequences present in the nucleotide sequence database, NCBI using the BLAST search program. Each clone was classified and put into a phylogroup according to the homology results obtained from the BLAST search. BLASTN software was used to find the most similar sequence in the database to clones or isolated species and their significant phylogenetic relationship. Genus and species of each clones and isolate were identified at percentage similarity values. The phylogenetic trees were constructed by using Neighbor joining method.

2.2.14. PCR-DGGE Analysis

Approximately 600 base pairs of 16S rRNA genes (V3, V4, V5 variable regions) were amplified using the universal bacterial primers E334R (5` CGC CCG CCG CGC CCC GCG CCC GTC CCG CCG CCC CCG CCC GCC AGA CTC CTA CGG GAG GCA 3`) (the GC clamped sequence is underlined) and E939R (CTTGTGCGGGCCCCCGTCAATTC) (Bacer et al., 2003). PCR was performed in a total reaction volume of 50 µl containing 25 µl 2X PCR Master Mix, 2,5 µl for each primers and 1 µl of 10 times diluted total genomic DNA isolated from kefir grains and kefir liquid by culture-independent method. PCR cyles were carried out using an initial denaturation step of 5 minutes at 94^{o}C followed by denaturation at 94 for 30 seconds. The annealling temperature was 67^{o}C for 30 seconds, the elongation step was conducted at 72^{o}C for 30 seconds. 25 cycle were accomplished. The final chain extantion at 72^{o}C

for 10 minutes was done. Amplified products were run on a 1% agarose gel stained with ethidium bromide and visualized under UV light.

The PCR amplicons were seperated by parallel Denaturing Gradient Gel Electrophoresis performed with DCode Universal Mutation Detection System (Bio-Rad Laboratories, USA) with gel of 16x16x0.01 cm. To visually check yhe formation of the gradient, 100 µl of DCode dye solution per 5 ml high density solution was added. 140 µl ammonium per sulfate (10%) and 17 µl TEMED were added for each 25 ml gel solution.The gel was let to polymerize for about 60 minutes. Fifteen µl of PCR product with same amount of 2x gel loading dye per lane were loaded on a 6% (ml/v) polyacrylamide gel containing and a 40-70% (denaturing solutions) gradient of urea and formamide (100% is 7M urea and 40% (v/v) formamide) increasing in the direction of the DNA migration during a run. Electrophoresis condition was 130V for 4 hours at 65^{o}C in 1X TAE buffer on the DCode apparatus. Gel was stained in 1 µg/ml ethidium bromide for 20 minutes and then visualized by UV transillumination.

CHAPTER 3

RESULTS AND DISCUSSION

In this study, we tried to investigate the composition of bacterial communities in kefir liquid and kefir grains by using both culture dependent and independent methods. To identify the captured bacterial communities, comparative sequence analyses of the partially amplified 16S rRNA genes were used and then taxonomic groups of identified species were determined. Also, the phylogenetic affiliations of identified bacterial species were showed by NJ method.

For culture-independent analysis, direct DNA was isolated from fermented kefir liquid and kefir grains. For culture-dependent analysis, bacteria from kefir liquid and homogenized kefir grains were cultivated under either aerobic or anaerobic conditions on MRS medium and among the resulting colonies, the distinct ones were chosen according to their different colony morphologies. Then the genomic DNA from these species was isolated.

Extracted genomic DNAs from both groups (culture-dependent and independent methods) were used as templates for the amplification of a phylogenetic marker, 16S rRNA gene. Then, the amplified genes were cloned into a cloning vector, pGEMT-Easy and they were transformed into chemically competent cells. The cycle sequencing reaction was applied to constructed cloning vector which isolated from transformed colonies. The raw data was analyzed and the analyses of 16S rDNA sequence similarity data were used as the criterion through the database to determine the isolates at species or genus level. Finally, the phylogenetic affiliation and community structure were determined by the help of analyzed sequence data. Furthermore, the culture-dependent and independent methods were evaluated according to their capability of capturing microbial population from a mix environment (Figure 3.1.).

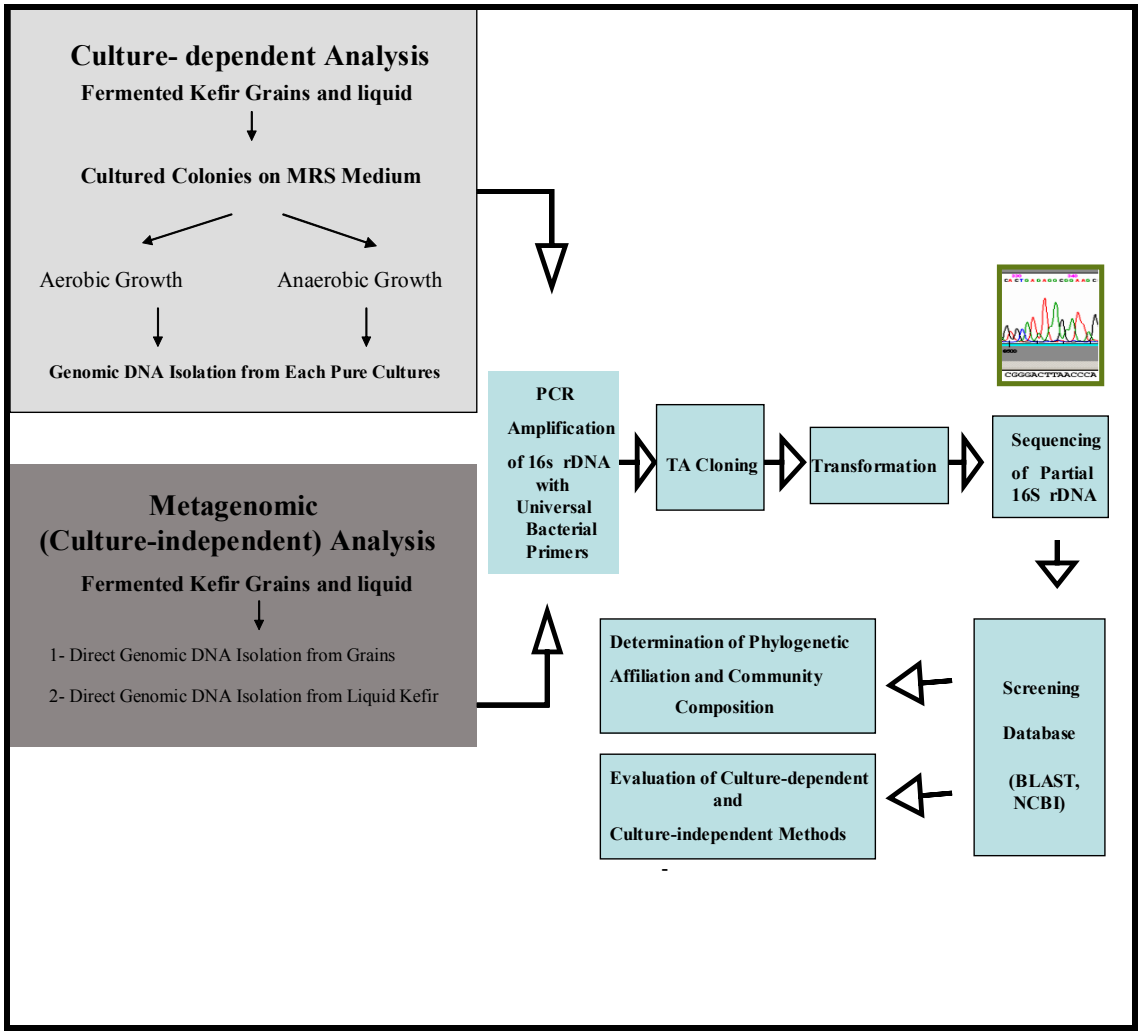


Figure 3.1. Schematic representation of the experiment flow chart

3.1. Capturing Bacterial Communities in Kefir Grains and Kefir Liquid

In this study, both culture-dependent and independent methods were used to capture the bacterial communities in kefir grains and kefir liquid. The total number of screened 16S rDNA clones for sequencing reactions was seventy six**.** Twenty four of sequence species were obtained from culture-dependent method and fifty two of them were obtained from culture-independent method.

3.1.1. The Results of Culture-dependent Method

3.1.1.1. The Isolation of Bacterial Species from both Kefir Grains and Kefir Liquid

Bacteria in kefir liquid and kefir grains were cultured on MRS medium under aerobic and anaerobic conditions (sample preparation was mentioned in Material and Methods). 3 or 5 days later colony formation on culture plates was observed. 7 different colony morphologies were observed on culture mediums (Table 3.1.). By culture-dependent methods, total 24 colonies from both kefir grains and liquid were chosen for identification of captured bacterial species and they were growth on fresh MRS medium as pure cultures to be ready for further experiment, genomic DNA extraction.

Table 3.1. The colony morphology of determined bacterial species from kefir grains and kefir liquid and their growth conditions according to oxygen requirement

		Kefir Liquid		**Kefir Grain**	
Colony Type No	**Colony Morphology on MRS Medium**	**A**	**AN**	**A**	**AN**
Colony 1	White, punctiform, entire, raised, opaque, dry	+	+	+	-
Colony 2	Irregular, undulate, flat, dull, dry	+	+	-	-
Colony 3	White, round, undulate, raised, dull, dry	-	-	+	-
Colony 4	Brown, round, entire, flat, translucent, moist	+	-	+	-
Colony 5	Brown, round, entire, convex, shiny, moist	-	-	+	-
Colony 6	White, round, entire, raised, shiny, moist	+	+	+	+
Colony 7	White, round, entire, convex, shiny, moist	-	-	-	+

A: Aerobic Growth Condition, AN: Anaerobic Growth Condition

10 of 24 bacterial colonies were chosen according to their morphological differences from MRS culture plates on which bacterial populations in kefir grains were grown and 14 of 24 bacterial colonies were chosen from culture plates containing kefir liquid bacterial species. Colony type 2 was only captured from kefir liquid, Colony type

3, 5 and 7 could be determined only from kefir grains, and Colony type 1, 4 and 6 were obtained in both kefir grains and kefir liquid.

Isolated species of Colony type 3, 4 and 5 were only captured from MRS medium cultured in aerobic condition, Colony type 7 was only isolated from anaerobic growth condition, and Colony type 1, 2 and 6 were isolated both from aerobic and anaerobic growth conditions (Table 3.1.).

3.1.1.2. Representation of Gel Electrophoresis Result of Amplified Partial 16S rRNA genes of Bacterial Species

The genomic DNA was extracted from each of isolated bacteria which were captured by culture-dependent methods from Kefir grains (10 isolates) and kefir liquid (14 isolates). The partial regions of 16S rRNA genes were amplified in PCR reaction by using universal bacterial primer pairs, E334 and E1115. The integrity and quantity of amplified 16S rDNA were visualized in 1% agarose gel electrophoresis before ligation into the cloning vector (Figure 3.2. and Figure 3.3.).

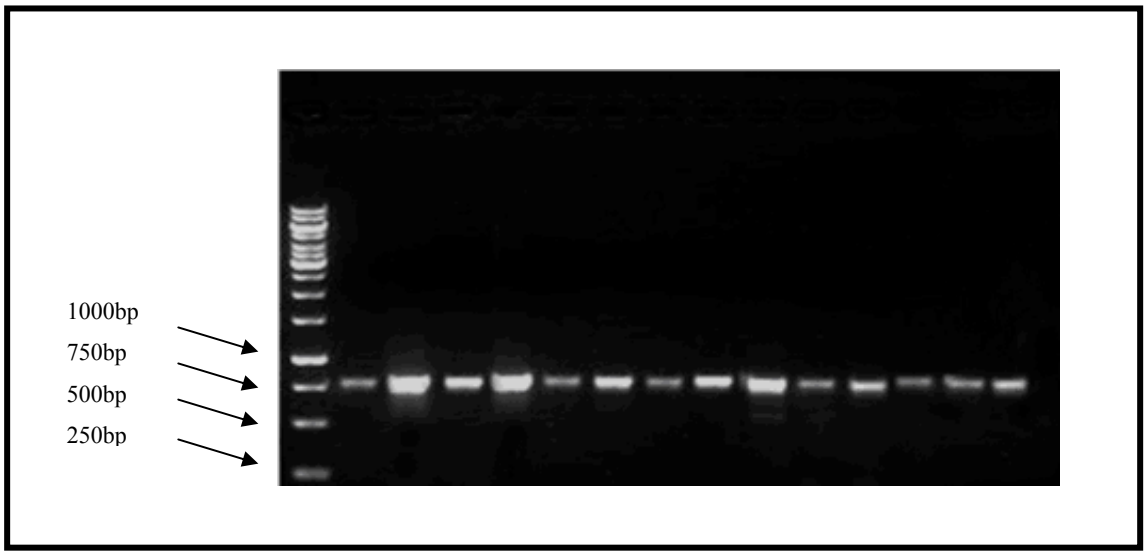


Figure 3.2. Agarose gel picture of PCR amplified partial 16S rRNA gene region of bacterial population that captured by culture-dependent method from kefir liquid

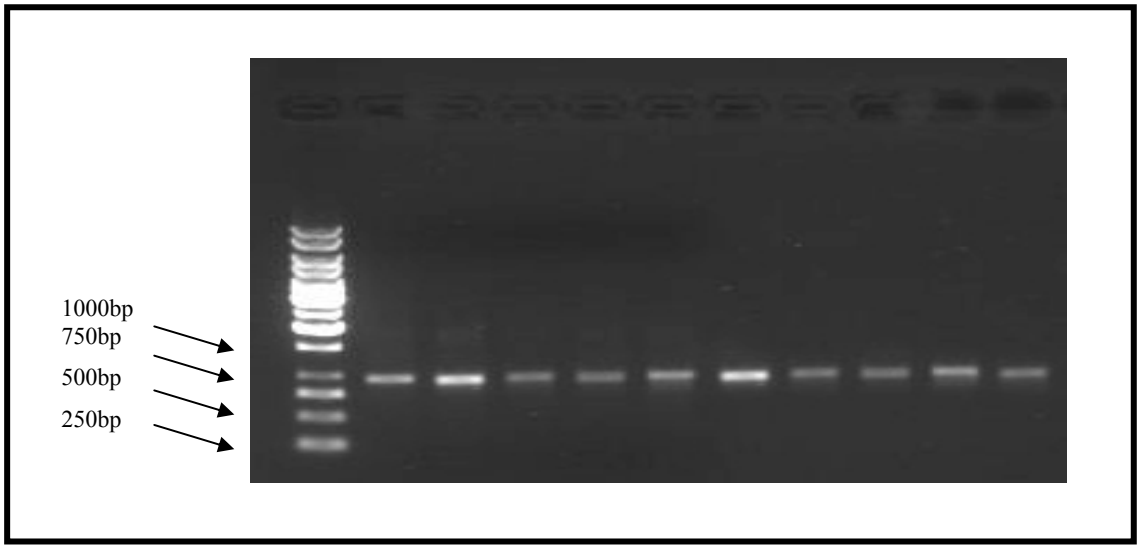


Figure 3.3. Agarose gel picture of PCR amplified partial 16S rRNA gene region of bacterial population that captured by culture-dependent method from kefir grains.

3.1.2. The Results of Culture-independent Method

Metagenome was directly extracted from fermented kefir liquid and kefir grain samples. Isolated heterogeneous genomic DNA was used as template for the amplification of 16S rRNA genes of bacterial communities. The same universal bacterial primer pair was used in PCR reaction as in amplification of 16S rDNA regions of isolated bacterial species. Partially amplified fragments were cloned into cloning vector pGEMT-Easy as mentioned in Material and Methods, and pGEMTeasy-16S rDNA plasmids were obtained (Figure 3.6.). False-positive clones were eliminated by colony PCR using vector specific primers and gel electrophoresis was applied to visualize integrity of the desired gene into the cloning vector. The expected length of fragment (~780 bp) was verified in 1% gel electrophoresis (Figure 3.4. and Figure 3.5.). 26 positive clones from kefir grains and 26 positive clones from kefir liquid were randomly chosen for the following sequencing reaction.

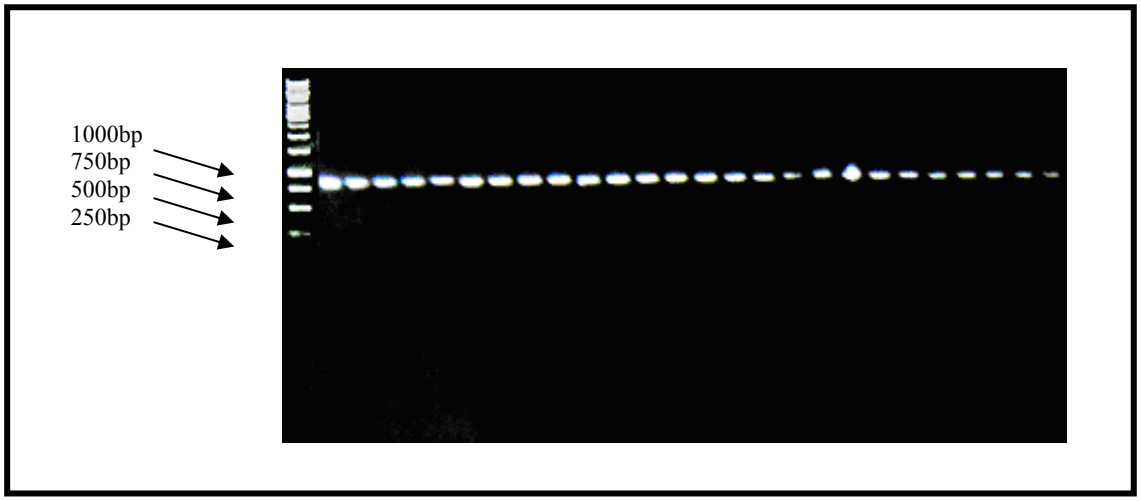


Figure 3.4. Agarose gel picture of PCR amplified partial 16S rRNA gene region of bacterial population that was captured by culture-independent method from kefir grains

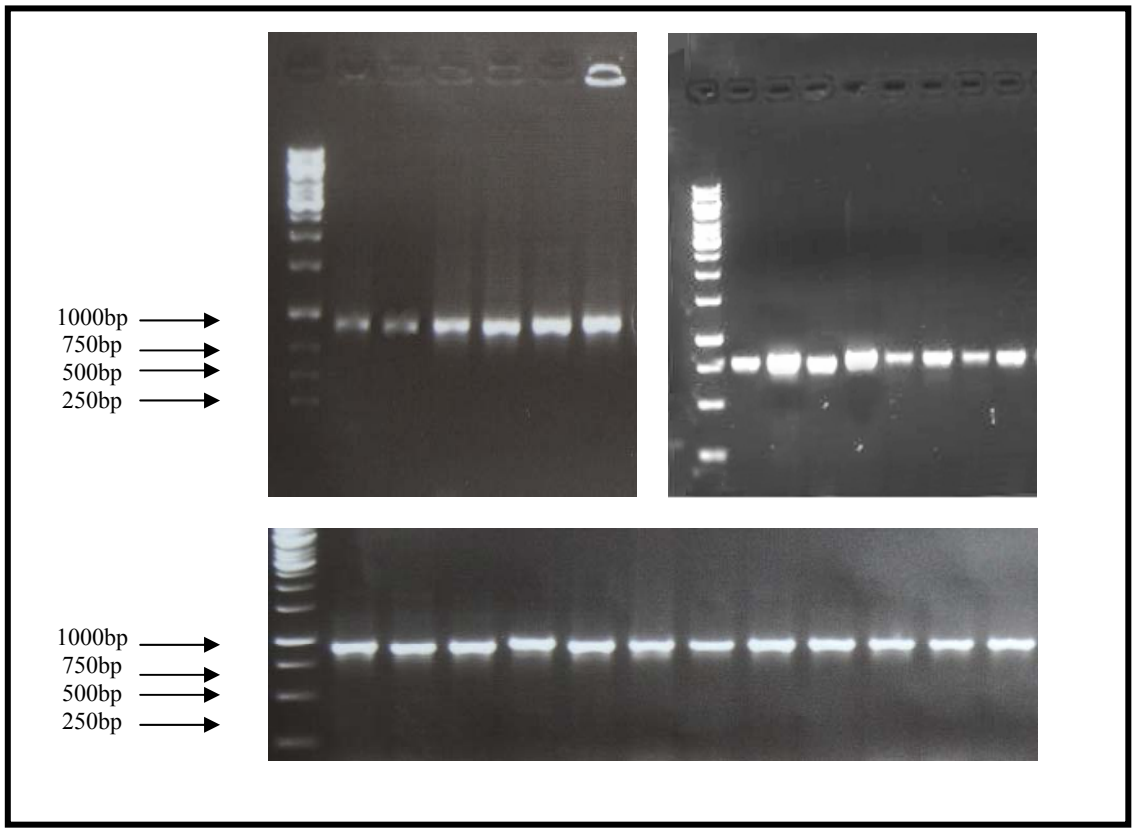


Figure 3.5. Agarose gel picture of PCR amplified partial 16S rRNA gene region of bacterial population that captured by culture-independent method from kefir liquid

3.2. The Construction of "pGEMT-Easy-16S rRNA gene" Cloning Vector

pGEMT-Easy is a TA cloning vector which has compatible ends for fragments which were amplified by Taq polymerase. Ligation reaction was done as mentioned in Material and Methods. The constructed pGEMT-Easy-16S rRNA gene cloning vectors (Figure 3.6.) were transformed into chemically competent cells and blue-white screening was done to choose the positive clones (carrying constructed plasmids).

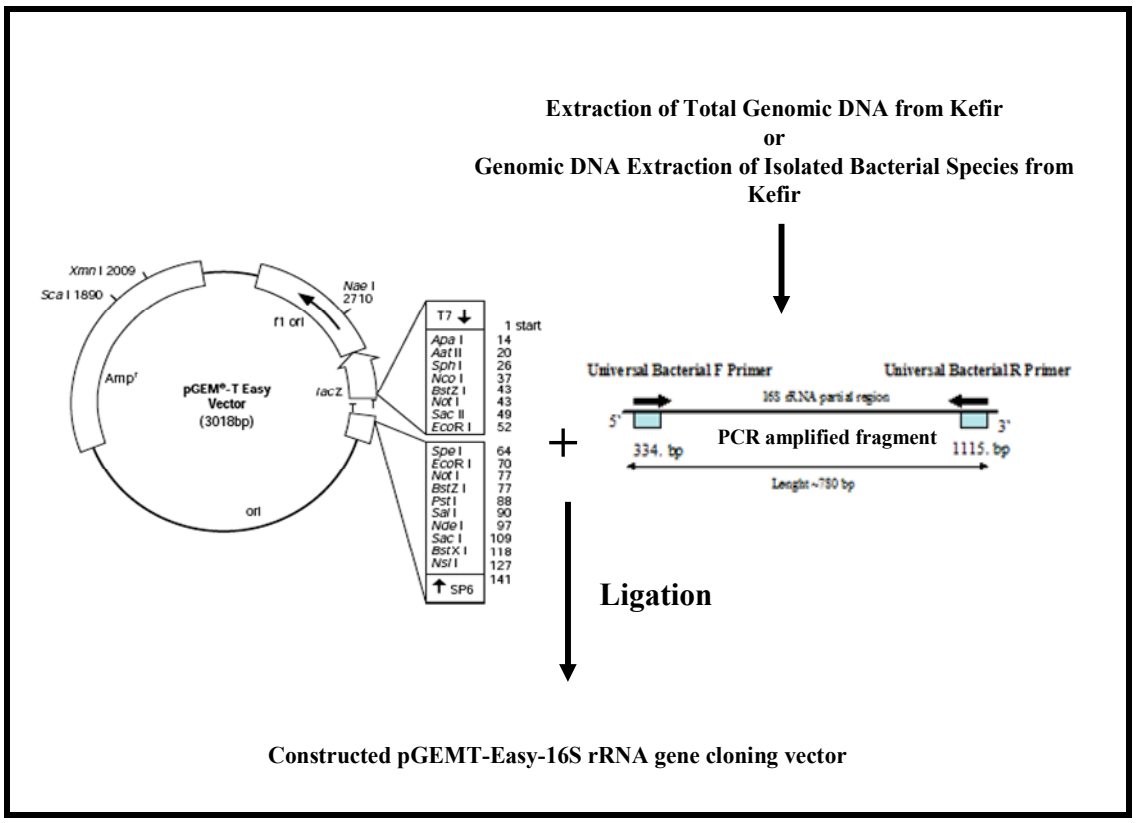


Figure 3.6. The construction of pGEMT-Easy-16S rRNA gene cloning vector

The colony PCR was applied to each chosen white colonies to confirm the insertion of the target fragment into the cloning vector. By this method, false-positive colonies were eliminated. Vector specific T7 and SP6 primer pairs were used in PCR reactions. Each reaction was visualized in agarose gel electrophoresis. The gel results indicated the expected length of desire fragment, partially amplified 16S rRNA gene of bacteria, about 780 bp lengths (Figure 3.4. and Figure 3.5.).

Each positive clone was grown in ampicilin containing liquid LB medium for plasmid isolation and the insert were sequenced in both direction by cycle sequencing using vector specific universal primers T7 and SP6.

3.3. Sequence Analysis

The required amount of input template dsDNA in sequencing reaction was adjusted to 260 ng for 100 fmol. Because sufficient DNA amounts are necessary for efficient sequencing results, defined targets were amplified in vitro by PCR prior the labeling and electrophoresis. The DNA concentration of sequenced sample affects sequence sensibility and readability that means it determines the signal intensity and resolution. The raw data of sequencing reaction for each sample was analyzed by Beckman Coulter Software. Each clone was sequenced in both directions and chromatograms were evaluated to decide whether a base difference caused by a base divergence or a base ambiguity.

A base divergence is a base difference between analyzed sample and the related sequence present in the database. A base ambiguity is caused by any biase in sequencing reaction that the base is not determined definitely.

3.4. Comparative Sequence Analysis of 16S rRNA Genes

Comparative sequence analysis involves the investigation of the sequence homologies between the known bacterial sequences present in the database and the samples under investigation. The level of homology helps to classify the bacterial sample to the different levels of taxonomic tree. High degrees of similarity can classify an organism at the genus or even species level.

In this study, previously described cut-off similarity levels for genus and species were accepted. The identification to the species level and genus level was defined as 99% and 97% similarity of the 16S rDNA gene to relative sequences in database, respectively (Drancourt, et al. 2000). Non redundant (nr) databases containing GenBank, EMBL, DDBJ databases was chosen for DNA sequence comparison in this study.

The numbers of screened 16S rDNA clone libraries of bacterial species that captured by culture-dependent or independent methods from both fermented kefir drink

and kefir grains were showed in Table 3.2. 24 clones from culture-dependent method and 52 clones from culture-independent methods were screened and their base sequences of 16S rDNA genes were determined. 10 colonies from kefir grains and 14 colonies from kefir liquid were chosen according to their observed colony morphologies on MRS medium. The 16S rDNA of each colony was cloned into cloning vector and each of them was screened by sequence-based method. In culture-independent method, 26 positive clones from kefir grains and 26 clones from kefir liquid was randomly chosen for sequence analysis.

In this study, *Acetobacter* species *A. syzygii* and *A. lovaniensis were* identified by both culture-dependent and independent analysis. *Enterococcus faecium* was identified only by culture-dependent approaches. *Lactobacill*i were identified from both kefir grains and kefir liquid. They were *Lb. helveticus*. *Lb. kefiranofaciens* and either *Lb. kefiri strain JCM* 5818 or *Lb. parabuchneri.* Since *Lb. kefiri strain JCM* 5818 or *Lb. parabuchneri* has same base sequences between the region that we amplified, we could not differentiate the species of this genus. All the clones screened from kefir grains by culture-independent method were identified as *Lb. kefiranofaciens Lactococcus lactis subsp. lactis* were determined by both culture-based and culture-independent methods in kefir liquid. *Leuconostoc mesenteroides* species were captured only by culture-dependent method from both grains and the liquid (Table 3.2.).

Table 3.2. The distribution of the identified bacterial isolates captured from kefir grains and kefir liquid by both culture-dependent and independent methods

Determined Species	Number of Isolation and Screened Clones		Number of Screened Clones	
	Culture-dependent		Culture-independent	
	Grains	Liquid	Grains	Liquid
Lb. kefiranofaciens	-	-	26	-
Lc. lactis subsp. lactis	+	+	-	11
Lb. helveticus	-	+	-	5
A.lovaniensis	+	+	-	10
Lb. kefiri/parabuchneri	+	-	-	-
A. syzygii	+	-	-	-

(Cont. on next page)

Table 3.2. (Cont.) The distribution of the identified bacterial isolates captured from kefir grains and kefir liquid by culture-dependent and independent methods

Leu. mesenteroides	+	+	-	-
Enterococcus faecium	+	-	-	-
Total			**26**	**26**

Lc. lactis subsp. lactis, A. lovaniensis and *Lb. helveticus* were captured both by culture-dependent and independent methods. However, *L. kefiranofaciens* was captured only by culture-independent and, *A. syzygii, Leu. mesenteroides, E. faecium and Lb. kefiri/parabuchneri* were captured only by culture-dependent method. Chen et al. published a recent study on kefir population by culture-dependent and culture-independent methods. They only studied kefir grains and identified *Lb. kefiranofaciens, Lb. kefiri, Lc. lactis* subsp. lactis*, Leu. mesenteroides* species (Chen, et al. 2008). However, we also captured *A. syzygii, A. lovaniensis* and *E. faecium* species in addition to *Lb. kefiranofaciens, Lb. kefiri, Lc. lactis* subsp. lactis, *Leu. Mesenteroides species* in kefir grains by applying both culture-dependent and independent methods.

The distribution of kefir bacterial populations in kefir grains and liquid was determined. *L. kefiri/parabuchneri*, *E. faecium, A.syzygii* and *L. kefiranofaciens* were identified in kefir grains but not in kefir liquid. *L.helveticus* was found only in kefir liquid. *L.lactis subsp. lactis, A.lovaniensis* and *L.mesenteroides* were the species which were proved being in both kefir liquid and grains.

The distribution of identified eight bacterial species according to capturing method and source was represented with Table 3.3. In this study, *Lc .lactis subsp. lactis* and *A. lovaniensis* were identified from grains and liquid by culture-dependent method and from liquid by culture-independent method. *L. mesenteroides was* identified only by culture-dependent method and found in both kefir grains and liquid. The distribution of *L. kefiri/parabuchneri*, *E. faecium* and *A. syzygii* were restricted to identification by only culture-dependent method in kefir grains. *L. kefiranofaciens* was the only species which we captured from kefir grains and we could only reach this species by culture-independent method. *L. helveticus* was the species which isolated from kefir liquid by both culture-dependent and independent methods. *L. kefiranofaciens* is only captured from kefir grain by culture-independent method. *Leuconostos mesenteroides* species

were only identified by culture-dependent analysis both from kefir grains and kefir liquid.

Table 3.3. Overview of identified bacterial species according to their capturing methods and capturing sources

Culture-dependent Method		**Culture-independent Method**	
Kefir Liquid	**Kefir Grains**	**Kefir Liquid**	**Kefir Grains**
A.lovaniensis	*A.lovaniensis*	*A.lovaniensis*	*Lb. kefiranofaciens*
Lc. lactis subsp. Lactis	*Lc. lactis subsp. lactis*	*Lc. lactis subsp. lactis*	
Leu. mesenteroides	*Leu. mesenteroides*	*Lb. helveticus*	
Lb. helveticus	*Enterococcus faecium*		
	Lb. kefiri or Lb. parabuchneri		
	A. syzygii		

The detailed sequence analyses of identified bacterial species in kefir liquid and grains by both culture-dependent and independent method were given as follows.

3.4.1. Sequence Analysis Results of Culture-independent Method

3.4.1.1. Bacteria identified from Kefir Grains by Culture-independent Method

26 clones containing 16S rDNA fragments of bacteria that obtained from Kefir grains by culture-independent process were sequenced. Homology with known species was indicated in Table 3.4. Only one species (100%), *Lactobacillus kefiranofaciens,* was isolated from kefir grains by culture-independent method (Figure 3.7.).

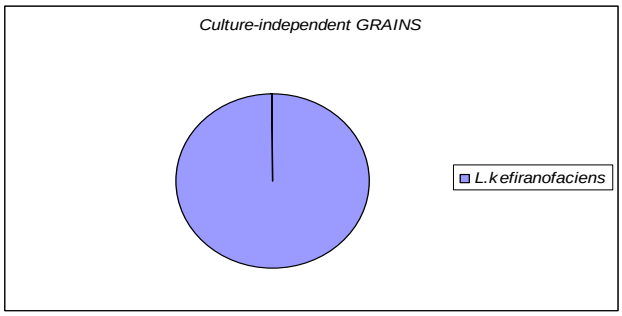


Figure 3.7. The representation of bacterial populations in kefir grains captured by culture-independent method

Lactobacillus kefirgranum was reclassified by Takizawa *et al.* (1994) as *Lactobacillus kefiranofaciens subsp. kefirgranum* (Vancanneyt, et al. 2004). 16 (61.5%) of the clones gave 100% similarity with the known species, *Lb. kefiranofaciens*. 10 (38,4%) of the clones gave the 99% similarity with the known species of *Lb. kefiranofaciens* (Table 3.4.).

Out of those 10 samples, 3 base divergences for 2 two clones (IG-10, IG11); 1 to 2 base divergences and 1 to 2 base ambiguites for 4 clones (IG19,20,21,22), 2 to 5 base ambiguites for 3 clones (IG12,13,15) were identified. The homology results were verified with forward and reverse primers sequencing. The divergence base positions were different in each sample except samples IG-10 and IG-11. They have the same divergence bases in positions 376 and 377 (Table 3.4.).

Table 3.4. The summary of sequence-based analysis of bacterial populations in kefir grains captured by culture-independent method

Sample Sequence No	**Phylogenetic Affiliation (Homology)**	**Sequenced Base Similarity**	**Percentage Similarity**	**Comment Base Ambiguity, Base Divergence and Divergence Position**
IG-1	*Lb. kefiranofaciens*	603/603	100%	
IG-2	*Lb. kefiranofaciens*	603/603	100%	
IG-3	*Lb. kefiranofaciens*	603/603	100%	
IG-4	*Lb. kefiranofaciens*	603/603	100%	
IG-5	*Lb. kefiranofaciens*	603/603	100%	
IG-6	*Lb. kefiranofaciens*	603/603	100%	
IG-7	*Lb. kefiranofaciens*	603/603	100%	
IG-8	*Lb. kefiranofaciens*	603/603	100%	
IG-9	*Lb. kefiranofaciens*	603/603	100%	
IG-10	*Lb. kefiranofaciens*	600/603	99%	3 base-divergence; 375,376,377
IG-11	*Lb. kefiranofaciens*	600/603	99%	3 base-divergence; 376,377,914
IG-12	*Lb. kefiranofaciens*	576/578	99%	2 base-ambiguity
IG-13	*Lb. kefiranofaciens*	502/504	99%	2 base-ambiguity
IG-14	*Lb. kefiranofaciens*	544/544	100%	
IG-15	*Lb. kefiranofaciens*	599/604	99%	5 base-ambiguity
IG-16	*Lb. kefiranofaciens*	779/779	100%	
IG-17	*Lb. kefiranofaciens*	779/779	100%	
IG-18	*Lb. kefiranofaciens*	779/779	100%	
IG-19	*Lb. kefiranofaciens*	776/780	99%	2 base-divergence 820, 893; 2 base- ambiguity
IG-20	*Lb. kefiranofaciens*	779/781	99%	1 base-divergence 615; 1 base- ambiguity
IG-21	*Lb. kefiranofaciens*	778/781	99%	1 base-divergence 768; 2 base- ambiguity
IG-22	*Lb. kefiranofaciens*	776/779	99%	1 base-divergence 453; 2 base- ambiguity
IG-23	*Lb. kefiranofaciens*	603/603	100%	
IG-24	*Lb. kefiranofaciens*	603/603	100%	

(Cont. on next page)

Table 3.4. (Cont.) The summary of sequence-based analysis of bacterial populations in kefir grains captured by culture-independent method

IG-25	*Lb. kefiranofaciens*	779/779	100%	
IG-26	*Lb. kefiranofaciens*	778/779	99%	1 base-divergence 496

IG: Samples obtained from Kefir Grains by culture-independent method. Sequence-based similarity column indicates the base homology ratio of identified species sequence fragments to known species. The numbers in comment column indicates the number of base divergence and their positions, and the number of base- ambiguity

Although, we could not identify *L. kefiranofaciens* species under culture-dependent approaches, culture-independent approaches enabled us to indicate the presence of this species in kefir grains. Takizawa and his collogues identified this bacterium on a selective medium but it was not possible for us to isolate *L. kefiranofaciens* by our culture method. The reason for this can be explained by the possible differences in culturing conditions.

3.4.1.2. Bacteria Identified from Kefir Liquid by Culture-independent Method

26 clones containing 16S rDNA fragments of bacteria that were cloned from liquid Kefir metagenome were sequenced and their similarity identities indicated the presence of three different bacterial species. Their ratio was as follows, 10/26 (38.5%) *Acetobacter lovaniensis*, 11/26 (42%) *Lactococcus lactis subsp. lactis* and 5/26 (19.5%) *Lb. helveticus* (Figure 3.8.).

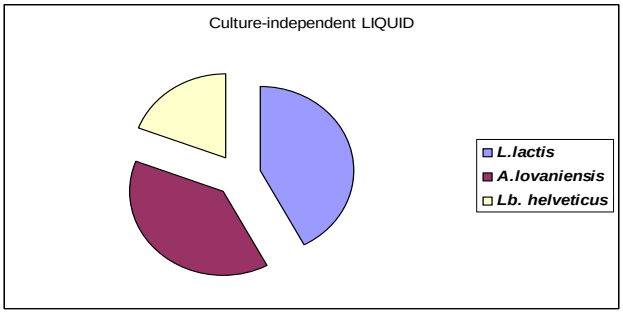


Figure 3.8. The representation of bacterial populations in kefir liquid captured by culture-independent method

Two of the *A. lovaniensis* species identified had 100% similarity (IL3 and IL4) with the known sequences in the database. Six had (IL10, IL12, IL14, IL18, IL20 and IL22) a base divergence at the same base position (687. base). For the remaining other two clones, 2 to 3 base ambiguites were observed. Of the 11 *Lactococcus lactis subsp. lactis* species, two (IL7 and IL8) gave 100% base similarity with the known species in the database and four had base divergences at the different base positions. The remaining five clones had 1-3 base ambiguities. All of the *Lb. helveticus* species identified had same base divergence at the same base position (327. base). However, two of them (IL21 and IL23) had also additional divergences at the base positions 432, 470 and 654 (Table 3.5.).

Table 3.5. The summary of sequence-based analysis of bacterial populations in kefir liquid that captured by culture-independent method

Sample Sequence No	**Phlogenetic Affiliation (Homology)**	**Sequenced Base Similarity**	**Percentage Similarity**	**Comment Base Ambiguities, Base Divergence and Divergence Positions**
IL-1	*Acetobacter lovaniensis*	755/757	99%	2 base- ambiguity
IL-2	*Acetobacter lovaniensis*	754/757	99%	3 base- ambiguity
IL-3	*Acetobacter lovaniensis*	757/757	100%	
IL-4	*Acetobacter lovaniensis*	757/757	100%	
IL10	*Acetobacter lovaniensis*	752/757	99%	3 base-divergence 687,908,970
IL12	*Acetobacter lovaniensis*	750/757	99%	3 base-ambiguity; 4 base divergence, 687, 715, 781, 995
IL14	*Acetobacter lovaniensis*	755/757	99%	2 base- divergence, 434, 687
IL18	*Acetobacter lovaniensis*	755/757	99%	1 base-ambiguity; 1 base-divergence, 687
IL20	*Acetobacter lovaniensis*	754/757	99%	1 base-ambiguity; 2 base-divergence, 687, 775
IL22	*Acetobacter lovaniensis*	756/757	99%	1 base- divergence, 687
IL-5	*Lactococcus lactis subsp. lactis*	773/777	99%	3 base- ambiguity ; 1 base-divergence 785
IL-6	*Lactococcus lactis subsp. lactis* .	775/777	99%	2 base- ambiguity

(Cont. on next page)

Table 3.5. (Cont.) The summary of sequence-based analysis of bacterial populations in kefir liquid that captured by culture-independent method

IL-7	*Lactococcus lactis subsp. lactis* .	777/777	100%	
IL-8	*Lactococcus lactis subsp. lactis* .	777/777	100%	
IL13	*Lactococcus lactis subsp. lactis*	775777	99%	1 base-ambiguity; 1 base-divergence, 375
IL15	*Lactococcus lactis subsp. lactis*	774/777	99%	3 base-ambiguity
IL16	*Lactococcus lactis subsp. lactis*	767/777	99%	3 base-ambiguity
IL17	*Lactococcus lactis subsp. lactis*	774/777	99%	3 base-ambiguity
IL24	*Lactococcus lactis subsp. lactis*	775/777	99%	2 base-ambiguity
IL25	*Lactococcus lactis subsp. lactis*	773/777	99%	3 base-ambiguity; 1 base-divergence; 428
IL26	*Lactococcus lactis subsp. lactis*	774/777	99%	2 base-ambiguity; 1 base-divergence, 607
IL9	*Lactobacillus helveticus*	770/779	98%	4 base- ambiguity; 5 base-divergence 327,823, 1040, 1071, 1082
IL11	*Lactobacillus helveticus*	782/783	99%	1 base-divergence 327
IL19	*Lactobacillus helveticus*	778/783	99%	4 base-ambiguity; 1 base-divergence, 327
IL21	*Lactobacillus helveticus*	776/783	99%	3 base-ambiguity; 4 base-divergence; 327, 432, 470, 654
IL23	*Lactobacillus helveticus*	776/783	99%	3 base- ambiguity; 4 base-divergence; 327, 432, 470, 654

IL: Samples obtained from Kefir Drink by culture-independent method. Sequence-based similarity column indicates the base homology ratio of identified species sequence fragments to known species. The numbers in comment column indicates the number of base divergence and their positions, and the number of base- ambiguity.

3.4.2. Sequence Analysis Results of Culture-dependent Method

3.4.2.1. Bacteria Identified from Kefir Grains by Culture-dependent Method

Bacteria were cultivated from homogenized kefir grains under aerobic or aneorobic conditions. In total, 10 different colonies were selected according to their colony morphologies and they were isolated as pure cultures. Their genomic DNAs were purified and used for the amplification of 16S rRNA genes. The amplified 16S rRNA genes were then cloned in pGEMTeasy vector, sequenced and searched in the database. Of these 10 colonies 3 (DG-3, DG-15, DG-17) were identified as *Leuconostoc mesenteroides,* 3 (DG-34-35-36) were identified as *Acetobacter syzygii,* 1 (DG-1) was identified as *Acetobacter lovaniensis*, 1 (DG-21) was identified as *Enterococcus faecium,* 1 (DG-2) was identified as *Lactobacillus kefiri or parabuchneri*, 1 (DG-5) was identified as *Lactococcus lactis subsp. lactis*.

DG-3 and DG-17 (identified as *Leu.mesenteroides*) have same base divergence in base position 328. The other determined divergences were at different positions. DG-21 16S rDNA fragment sequence gave 99% similarity to *Enterococcus faecium* with 741/743 base homology ratio (Table 3.6.). *E. faecium* was isolated from fermented Turkish kefir and to our knowledge there is no literature that this species were identified in kefir before. However, Yuksekdag *et al.* reported that *E. durans* was isolated from Turkish Kefir (Yuksekdag, et al. 2004).

Table 3.6. The summary of sequence-based analysis of bacterial populations in kefir grains that captured by culture-dependent method

Sample Sequence No	**Phylogenetic Affilation (Homology)**	**Sequenced Base Similarity**	**Percentage Similarity**	**Comment Base Ambiguities, Base Divergence and Divergence Positions**
DG-1	*Acetobacter lovaniensis*	756/757	99%	1 base-divergence 687

(Cont. on next page)

Table 3.6. (Cont.) The summary of sequence-based analysis of bacterial populations in kefir grains that captured by culture-dependent method

DG-34	*Acetobacter syzygii*	754/757	99%	1 base- ambiguity; 2 base-divergence 411,564
DG-35	*Acetobacter syzygii*	754/759	99%	4 base- ambiguity; 1 base-divergence 538
DG-36	*Acetobacter syzygii*	757/759	99%	1 base- ambiguity; 1 base-divergence 667
DG-5	*Lactococcus lactis subsp. lactis*	781/787	99%	5 base- ambiguity ; 1 base-divergence 604
DG-15	*Leuconostoc mesenteroides*	647/650	99%	1 base- ambiguity ; 2 base-divergence 329, 517
DG-3	*Leuconostoc mesenteroides*	780/785	99%	3 base- ambiguity ; 2 base-divergence 625,328
DG-17	*Leuconostoc mesenteroides*	780/785	99%	3 base- ambiguity ; 2 base-divergence 845,328
DG-2	*Lactobacillus kefiri / parabuchneri*	780/783	99%	3 base- ambiguity
DG-21	*Enterococcus faecium*	741/743	99%	2 base- ambiguity

DG: Samples obtained from Kefir grains by culture-dependent methods. Sequence-based similarity column indicates the base homology ratio of identified species sequence fragments to known species. The numbers in comment column indicates the number of base divergence and their positions, and the number of base- ambiguity.

3.4.2.2. Bacteria Identified from Kefir liquid by Culture-dependent Method

Fourteen isolated colonies from fermented liquid kefir were chosen due to their colony morphologies for sequencing analysis. They were isolated as pure cultures under aerobic or aneorobic conditions. Extracted total DNA were amplified with 16SrRNA primers and they were used to construct 16S rRNA gene clones. Sequencing of the clones and database search gave the following results; 2/14 colonies were identified as *Leuconostoc mesenteroides,* 4/14 were identified as *Acetobacter lovaniensis*, 3/14 were identified as *Lactobacillus helveticus,* 5/14 were identified as

Lactococcus lactis subsp. lactis. Two of the *Lb. helveticus* species had identical sequences the same 3 base divergences in position 386, 477, 478 (Table 3.7.).

Table 3.7**.** The summary of sequence-based analysis of bacterial populations in kefir liquid that captured by culture-dependent method

Sample Sequence No	**Phylogenetic Affilation (Homology)**	**Sequenced Base Similarity**	**Percentage Similarity**	**Comment Base Ambiguities, Base Divergence and Divergence Positions**
DL-6	*Leuconostoc mesenteroides*	*603/603*	*100%*	
DL-24	*Leuconostoc mesenteroides*	779/785	99%	4 base- ambiguity
DL-28	*Lactococcus lactis subsp. lactis*	745/746	99%	1 base- ambiguity
DL-30	*Lactococcus lactis subsp. lactis*	773/777	99%	3 base- ambiguity ; 1 base-divergence 493
DL-31	*Lactococcus lactis subsp. lactis*	777/777	100%	
DL-37	*Lactococcus lactis subsp. lactis*	574/580	99%	6 base- ambiguity
DL-49	*Lactococcus lactis subsp. lactis*	534/531	99%	3 base- ambiguity
DL-32	*Lactobacillus helveticus*	776/779	99%	4 base-divergence 386,477,478
DL-38	*Lactobacillus helveticus*	780/785	99%	5 base- ambiguity; 1 base-divergence 693
DL-52	*Lactobacillus helveticus*	779/783	99%	1 base- ambiguity; 3 base-divergence 386,477,478
DL-27	*Acetobacter lovaniensis*	755/757	99%	2 base- ambiguity
DL-46	*Acetobacter lovaniensis*	530/533	99%	3 base- ambiguity

(Cont. on next page)

Table 3.7**.** (Cont.) The summary of sequence-based analysis of bacterial populations in kefir liquid that captured by culture-dependent method

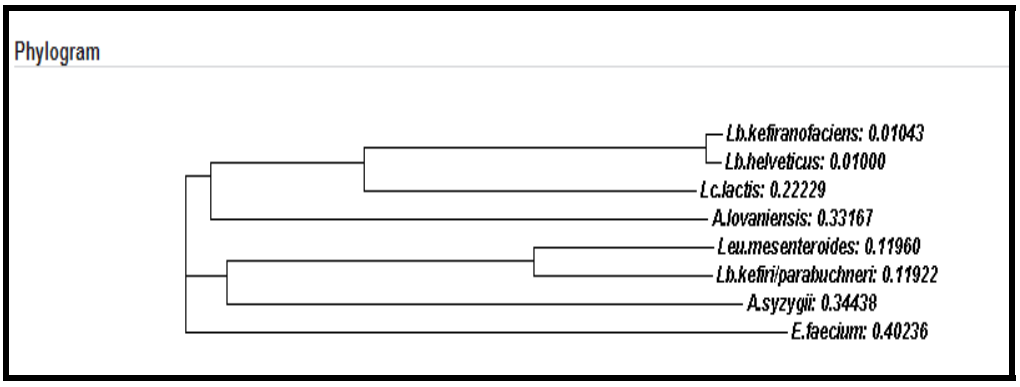
DL-48	*Acetobacter lovaniensis*	569/574	99%	5 base- ambiguity
DL-51	*Acetobacter lovaniensis*	759/761	99%	2 base- ambiguity

DL: Samples obtained from Kefir drink by culture-dependent method. Sequence-based similarity column indicates the base homology ratio of identified species sequence fragments to known species. The numbers in comment column indicates the number of base divergence and their positions, and the number of base- ambiguity

3.5. Phylogenetic Relationship between Identified Bacterial Species in Kefir

The phylogenetic tree which was drawn based on partial 16S rRNA gene sequence homologies among species enables us to understand their evolutionary history. A phylogenetic tree was constructed based on the multiple alignments. Phylogenetic calculation and phylogram drawing of identified kefir bacterial species were based on neighbor-joining method of Saitou and Nei (Saitou and Nei 1987). The scale represents the relative phylogenetic distance. ClustalW2 which is a multiple sequencing alignment program was used to construct the phylogenetic tree of identified bacterial species (European Bioinformatics Institute Server 2008). The distances between the sequences in the alignment were calculated and then it was used by the NJ method to make the phylogenetic tree (Figure 3.9.).

Figure 3.9. The construction of phylogenetic tree based on Neighbor-joining analysis of bacterial population in kefir (Source: European Bioinformatics Institute Server 2008)



Analyzed 16S rDNA sequences of kefir bacteria were grouped into two phylums. All species instead of Acetobacter sp. are belonged to *Firmicutes*. However, Acetobacter sp. represented by *Acetobacter lovaniensis* and *Acetobacter syzygii* are into alfa subgroup of *Proteobacteria* (Figure 3.10.).

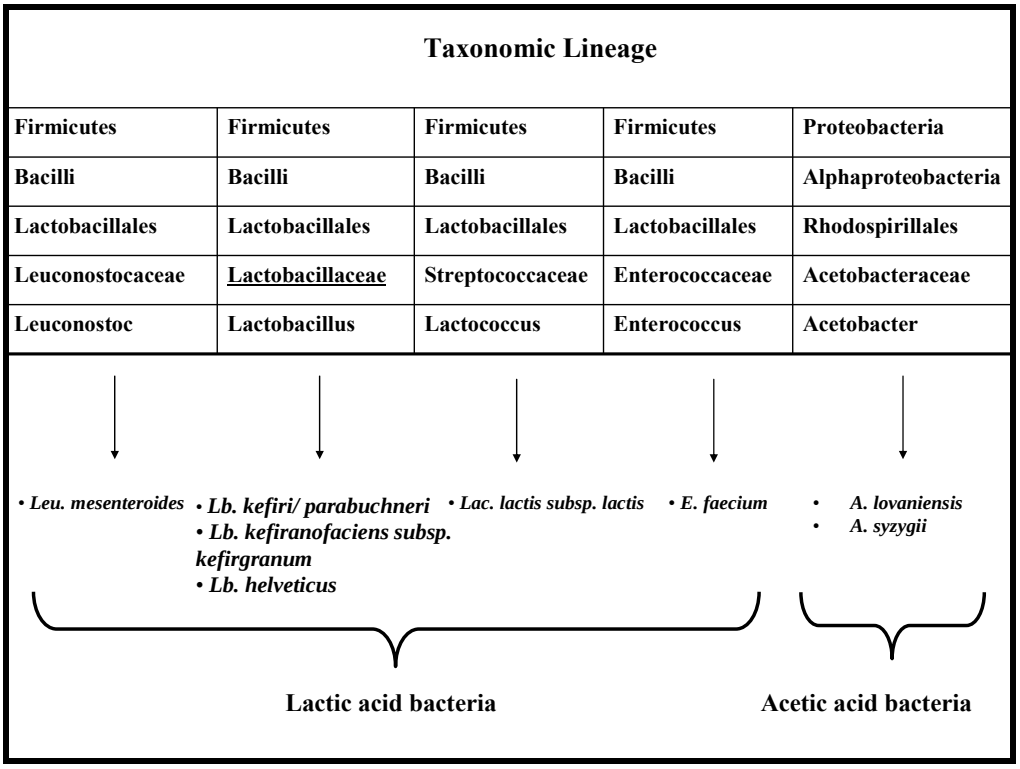


Figure 3.10. Taxonomic lineage of identified bacterial species
(Source: Entrez Taxonomy 2008)

The identified species in kefir, *Lactococcus sp., Leuconostoc sp., Lactobacillus sp. and Enterococcus,* are the lactic acid bacteria. All lactic acid bacteria are the major acid producer in kefir fermentation. *Lb. kefiri* is a heterofermentative bacterium that produces ethanol, acitic acid and carbon dioxide as well as diacetyl, acetoin, 2-3-butanediol and formate additional to lactic acid (Liu 2003). These products determine the flavor characteristics of kefir drink. *Lb. helveticus* is a homofermentative species and produce lactate from sugar as major product of glycolytic pathway. The identified *Lactobacillus kefiranofaciens* (homofermentative) and *Lb. kefiri* has been reported as kefiran producer in kefir (Yokoi, et al. 1991). *Leu.mesenteroides* is a heterofermentative bacterium which increased the viscosity and flavor of kefir. Acetobacter species are acetic acid bacteria and they use ethanol as carbon source and they can survive during fermentation process. Additionally, *E. faecium* also takes role in flavor formation in kefir.

CHAPTER 4

CONCLUSIONS

Kefir is a traditional fermented dairy product which is known as a probiotic beverage, produced and consumed all over the world. Kefir microbial population takes attention for food quality control and also for studies on bioactive products produced by kefir microbial population. Studies on the bacterial communities of kefir have focused on the identification microorganisms in kefir grains and liquid using traditional culture and characterization methods. Culture-based methods have many limitations to capture and identify bacterial populations when studying on mix microbial samples. Therefore, in this study, we used a complementary method, culture-independent method, in addition to isolation of bacterial population on MRS culture medium to study the microbial communities present in kefir.

This study evaluated the utility of 16S rDNA sequence analysis method as a tool for identification of bacterial communities in kefir. For this purpose, 16S rRNA genes were amplified from total genomic DNA of the bacterial communities and from cultured bacterial species present in kefir grains and kefir liquid and then sequenced. 16S rDNA analysis enables us to identify for the first time bacterial species previously undetected in kefir before. This study extends the knowledge of microbial diversity of kefir environment.

In the present study, a diverse spectrum of bacterial genera was determined in kefir samples including lactococci (1 species), lactobacilli (3 species), leuconostoc (1 species), acetobacteria (2 species) and enterococcus (1 species). 7 bacterial isolates, *Lactococcus lactis subsp. lactis, Lactobacillus kefiranofaciens, Lactobacillus Lb. helveticus, Acetobacter lovaniensis, Acetobacter syzygii, Leuconostoc mesenterides, Enterococcus faecium*, were defined at species levels that have 99-100% base homology and 1 isolate *Lactobacillus kefiri or parabucher*i were defined at genus level that have 99% base homology with sequences of known bacteria deposited in NCBI. The sequences of amplified partial region of 16S rRNA genes of identified bacterial species were given in Appendix A. 16S rDNA sequence analysis also provides the information on taxonomic level and relatedness of identified species.

Only 22 of 76 sequence results gave 100% base homology. The others had various base divergences and ambiguities. The base divergence situation can be explained as follows; in bacteria rDNA genes are organized as multigene families, ribosomal RNA (rrn) operons. This means that there are usually more than one rDNA per genome that may contain minor sequence variability (Cilia, et al. 1996 and Acinas, et al. 2004). However, these dissimilarities would not lead to different species identification (Clarridge 2004). Because not all possible rDNA genes for each species have been entered in databases, these base differences may prevent us to obtain 100% base homology (Cilia, et al. 1996). The PCR-induced artifacts could also be observed by *Taq* polymerase (Mincer, et al. 2005). If direct PCR products are sequenced instead of cloned PCR products, some artifacts can be observed due to multi rrn operons such as double peaks in sequencing causing ambiguious base formation. Due to such risks, we preffered to clone 16S rRNA genes into a cloning vector before sequencing the 16S rRNA genes.

In this study, the mismatches between 16S ribosomal universal bacterial primers and the 16S rDNA of various bacterial species were identified (Appendix B). These dissimilarities and also the less abundance of some species in a sample affect the PCR quantitity results which cause minimizing the amplification of some bacterial species 16S rDNA fragments (Bacer, et al. 2003). As a result, dominant species may suppress the amplification of less abundant species. In our experiment, we only identified *Lb. kefiranofaciens* species in kefir grains by culture-independent method. However, many other species (*A. lovaniensis, Lc. lactis subsp. lactis, Leu. mesenteroides, Enterococcus faecium, Lb. kefiri/ Lb. parabuchneri, A. syzygii*) were observed in kefir grains by culture-dependent method. Garrote et al. showed that the number of lactobacilli (10^9cfu/g) was higher than lactococci (10^7cfu/g) in kefir grains (Garrote, et al. 1997). Due to this difference in numbers, the amplification probability of *Lb. kefiranofaciens* DNA would be expected to be several orders of magnitude higher than lactococcal DNA during a PCR reaction. Because of this, rare bacterial genome as a tamplate has little chance to compete with bacteria that are higher amount in samples (Schabereiter-Gurtner, et al. 2001). That can be the explanation to why the present experiment reached only *Lactobacillus kefiranofaciens* from grains by culture-independent approach.

Although, 16S rDNA sequence analysis is a versatile and rapid tool, it has some potential problems, such as base errors, ambiguous base designations, sequence gap and incomplete sequences which cause misidentification. Clean 16S rDNA sequence data are

needed for definite identification. For this purpose, we cloned amplified fragments into cloning vectors and then related regions were sequenced in both direction using forward and reverse primers and analyzed data was merged manually. Additionally, the presence of more accurate and complete genetic database is important to identify the species more accurately. Another complemantery molecular methods to identify the bacterial population in kefir is PCR-based DGGE analysis.

The data in this study shows that a molecular identification method called PCR-based DGGE analysis is able to detect fingerprints of bacterial populations in kefir grains and kefir liquid mix microbial environment. DGGE is a powerful method to identify single base changes in a segment of DNA. Because of that each band observed on the gel represents a different bacterial species in kefir. The addition of GC clamp to one of the PCR primers insures that the screen region is in the lower melting domain and enables DNA double strand remain partially double stranded (Muyzer and Smalla 1998). In DGGE, the denaturing environment is created by a combition of uniform temperature and linear denaturants (urea and formamide). In this experiment, two different denaturing gradients were tried (30-50% and 40-70%). However, the 40-70% gave the best results to separate kefir bacterial populations.

Figure 4.1. indicates the fingerprints of bacterial species in kefir grains (lane 2) and in kefir liquid (lane 3) on parallel DGGE. Although, 3 different fragments were obtained in kefir grains sample on DGGE, we only identified one species (*Lb. kefiranofaciens*) from 16S rRNA sequencing analysis in kefir grains. The thickest band (lane 2) was thought to be the *Lb. kefiranofaciens* species, because its amount was higher than any species, its efficiency of amplification and cloning into vector would be higher than any species in the sample. Although, we found the presence of six other species (*A. lovaniensis, Lc. lactis subsp. lactis, Leu. mesenteroides, Enterococcus faecium, Lb. kefiri/ Lb. parabuchneri, A. syzygii*) in kefir grains by culture-dependent method. It was concluded from DGGE analysis, three most abundant species were amplified in PCR reaction. Because of that we only observed three bands instead of six different fragments.

According to 16S rRNA gene sequencing analysis, four different species were identified (*A. lovaniensis, Lc. lactis subsp. lactis, Lb. helveticus, Leu. mesenteroides*) in kefir liquid by both culture-dependent and independent methods. On DGGE, three bands were observed from kefir liquid derived from culture-independent method. This result

was verified by 16S rRNA gene sequencing analysis. To determine the each species repsented by a fragment on DGGE gel results, a reference markers should be used.

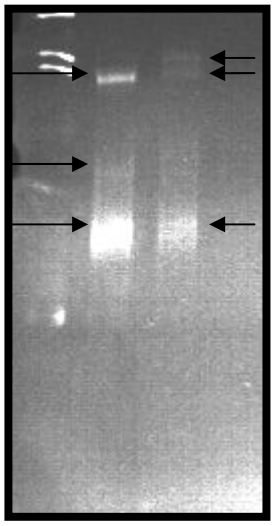


Figure 4.1. Parallel DGGE seperation of PCR amplicon derived from kefir grains and kefir liquid by culture-independent method

Our study shows that both conventional cultivation methods and molecular strategies (culture-independent) have drawbacks, none is sufficient alone or in combination to identify the all bacterial community in complex environmental samples. However, that is the fact that integration of culture-dependent and independent methods give chance to capture high percentage of species by the caused that each of them has different type of determination potential. 75% culture-indeendent species gave positive result with culture-dependent method. However, 42.8% of bacterial species that captured by culture-dependent methods gave positive results with culture-independent method (*Lc. lactis subsp. lactis, A.lovaniensis, Lb. helveticus*). *Lb. kefiranofaciens* gave positive results only with culture-independent method not with culture-dependent method. We concluded that careful consideration should be given to select an appropriate culture medium and culture condition for the identification of all present bacteria in the microbiologically mix samples.

16s rDNA gene is approximately 1,550 bp and the usually 500-1,500 bp fragments sequences can be enough to differentiate bacteria at species or genus level (Cai, et al. 2003). However, for some closely related species partial SSU rDNA sequencing may not be adequate to differentiate two species. From current experiment, we could not identify only one of bacterial isolate to the species level (*Lactobacillus*

kefiri or *parabucheri*). Because, both species have same base sequences at the partial region of 16S rDNA that we amplified in this experiment (V3, V4, V5, and V6). It was found that *Lb. kefiri* and *L. parabuchneri* have % 100 homology along 781 bp. Thus, it could not be possible to identify the right species of *Lactobacillus* by this primer pairs (E334-E1115). Another primer pairs that amplified the out of scanned bases should be used to determine whether the species is *kefiri* or *parabuchneri*. It is an aspect that using other alternative phylogenetic markers (recA, *groEL*, *rpoB*, *gyrB etc.*) is also possible to more definitely identification of microbial population (Schloss and Handelsman, 2003). We could not identify one of bacterial isolate to the species level (*Lactobacillus kefiri* or *parabucheri*). Because, both species have same base sequences at the partial region of 16S rDNA that we amplified in this experiment (V3, V4, V5, and V6). For more definite identification of species, different primer pairs which amplified another regions of 16S rRNA gene or the alternative phylogenetic markers (as mentioned above) should be preferred.

The preferred DNA extraction method (mechanical) in culture-independent approaches may affect the capturing bacterial diversity from kefir grains because it may be better suited for some bacteria and increase their representations. To overcome such biases, several different DNA isolation methods can be chosen (physical, chemical DNA extraction or a combination of both).

According to literature review, this experiment was the first which found the presence of the following species; *A. lovaniensis, A, syszii and E. faecium*. Although, Fontan and et al. found the presence of Enterococcus genus in kefir, they could not identified species of this isolate (Fontan, et al. 2005). Chen and his friends were the first who studied kefir grains by both culture-dependent and independent methods. However, they identified only four species from their studies (*Lb. kefiranofaciens, Lb. kefiri, Leu. mesenteroides, Lc. lactis*) (Chen, et al. 2008). In our study we identified three different species (*A. lovaniensis, A. syzgii and E. faecium*) additional to the ones they identified. Furthermore, our starter sample was not only kefir grains but also the kefir liquid.

In conclusion, both culture-dependent and independent methods should be used as complemantry approaches while studying on a mix microbiological sample. 16S rRNA sequencing and PCR-based DGGE are the most reliable molecular methods for culture-independent approaches.

REFERENCES

Acinas, S.G., Marcelino, L.A., Klepac-Ceraj, V. and Polz, M.F. 2004. Divergence and Redundancy of 16S rRNA Sequences in Genomes with Multiple rrn Operons. *Journal of Bacteriology* 186 (9):2629–2635.

Allen, E.E. and Banfield, J.F. 2005. Community Genomics In Microbial Ecology And Evolution. *Nature Reviews Microbiology* 3:489-498.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol*. 215:403-410.

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.

Ampe, F., Omar, N. and Guyot, J.P. 1999. Culture-independent quantification of physiologically-active microbial groups in fermented foods using rRNA-targeted oligonucleotide probes: application to pozol, a Mexican lactic acid fermented maize dough. *Journal of Applied Microbiology* 87:131-140.

Angulo, L., Lopez, E. and Lema, C. 1993. Microflora present in kefir grains grown in milk and in soy milk. *J dairy Res* 60:263-267.

Angulo, L., Lopez, E., and Lema, C. 1993. Microflora present in kefir grains grown in milk and in soy milk. *J dairy Res* 60:263-267.

Basic Local Alignment Search Tool. 2008. National Center for Biotechnology Information, Nucleotide Basic Local Alignment Search Tool. http://blast.ncbi.nlm.nih.gov/Blast.cgi (accessed February 14, 2008)

Bacer, 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.

Bergey, D.H. 2005. *Bergey's Manual of Systematic Bacteriology*. New York: Springer Press.

Blackwood, K.S., He, C., Gunton, J., Turenne, C.Y., Wolfe, J. and Kabani, A.M**.** 2000. Evaluation of recA Sequences for Identification of Mycobacterium Species. *Journal of Clinical Microbiology* 38:2846-2852.

Bosshard, P.P., Abels, S., Zbinden, R., Bottger, 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* 43:4134–4140.

Brosius, J., Dull, T.J. and Noller, H.F. 1980. Complete nucleotide sequence of a 23S ribosomal RNA gene from Escherichia coli. *Proc. Natl. Acad. Sci. USA* 77:201–204.

Brosius, J., Palmer, M.L., Kennedy, P.J. and Noller, H.F. 1978. Complete nucleotide sequence of a 16S ribosomal RNA gene from Escherichia coli. *Proc. Natl. Acad. Sci. USA* 75:4801–4805.

Cai, H., Archambault, M. and Prescott, J.F. 2003. 16S ribosomal RNA sequence–based identification of veterinary clinical bacteria. *J Vet Diagn Invest* 15:465–469.

Chen, H., Wang, S., and Chen, M. 2008. Microbiological study of lactic acid bacteria in kefir grains by culture-dependent and culture-independent methods. *Food Microbiology* 25:492-501.

Chen, H., Wang, S. and Chen, W. 2008. Microbiological study of lactic acid bacteria in kefir grains by culture-dependent and culture-independent methods. *Food Microbiology* 47:1084-1088.

Chivian, D. and Arkin, A.P. 2006. Genomics for environmental microbiology. *Current Opinion in Biotechnology* 17:229–235.

Cilia, V., Lafay, B. and Christen, R., 1996. Sequence Heterogeneities Among 16s Ribosomal RNA Sequences, and Their Effect on Phylogenetic Analyses at the Species Level. *Molecular Biology and Evolution* 13(3):451-461.

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 (4): 840–862.

Daniel, R. 2005. The metagenomics of soil. *Nature Reviews Microbiology* 3**:**470 478.

Dauendorffer, J.N., Guillemin, I., Aubry, A., Pernot, C.T., Sougakoff, W., Jarlier, V. and Cambau, E. 2003. Identification of Mycobacterial Species by PCR Sequencing of Quinolone Resistance-Determining Regions of DNA Gyrase Genes. *J Clin Microbiol* 41(3): 1311–1315.

Doolittle, W.F. 1999. Phylogenetic Classification and the Universal Tree. *Science* 284:2124-2128.

Drancourt, M., and Raoult, D. 2002. rpoB gene sequence-based identification of Staphylococcus species. *J Clin Microbiol* 40(4):1333-8.

Drancourt, M., Bollet, C., and Carlioz, A. 2000. 16S ribosomal DNA sequence analysis of a large collection of environmental and clinical unidentifiable bacterial isolates. *J Clin Microbiol* 38:3623–3630.

Drancourt, M., Bollet, C., Carlioz, A., Martelin, R., Gayral, J. and Raoult, D. 2000. 16S Ribosomal DNA Sequence Analysis of a Large Collection of Environmental

and Clinical Unidentifiable Bacterial Isolates. *Journal Of Clinical Microbiology* 38:3623–3630.

Dupre, J. and O'Malley, M.A. 2007. Metagenomics and biological ontology. *Stud. Hist. Phil. Biol. & Biomed. Sci*. 38:834–846.

Escalante, A., Rodriguez, M.E., Martinez, A., -Munguia, A.L., Bolivar, F. and Gosset, G. 2004. Characterization of bacterial diversity in Pulque, a traditional Mexican alcoholic fermented beverage, as determined by 16S rDNA analysis. *FEMS Microbiology Letters* 235:273–279.

Entrez Taxonomy. 2008. The NCBI taxonomy database http://www.ncbi.nlm.nih.gov/sites/entrez?db=taxonomy (accessed June 2, 2008).

European Bioinformatics Institute Server. ClustalW2 Multiple Sequence Alignment Program. 2008. http://www.ebi.ac.uk/Tools/clustalw/index.html (accessed June 1, 2008).

Farnworth, E.R. 2005. Kefir – a complex probiotic. *Food Science and Technology Bulletin: Functional Foods* 2 (1):1–17.

Felsenstein, J. 1988. Phylogenies from molecular sequences: Inference and reliability. *Annu. Rev. Genet*. 22:521-565.

Fleet, G.H. 1999. Microorganisms in food ecosystems. *International Journal of Food Microbiology* 50:101–117.

Frengova, G.I., Simova, E.D., Beshkova, D.M. and Simovb, Z.I. 2002. Exopolysaccharides Produced by Lactic Acid Bacteria of Kefir Grains. *Z. Naturforsch.* 57: 805-810.

Fuhrman, J. A. 2002. *Community Structure: Bacteria and Archaea*. United Kingdom: Elsevier Press.

Garbers, I.M., Britz, T.J. and Witthuhn, R.C. 2004. PCR-based denaturing gradient gel electrophoretictypification and identification of the microbial consortium present in kefir grains. *World Journal of Microbiology and Biotechnology* 20:687-693.

Garrote, G.L., Abraham, A.G. and De Antoni, G.L. 1997. Preservation of Kefir Grains, a Comparative Study. *Lebensm.-Wiss. u.-Technol* 30:77–84.

Gillis, M., Vandamme, P., Devos, P., Swings, J. and Kersters, K. 2001. *Polyphasic taxonomy. In Bergerys Manula of systematic Bacteriology*. New York: Springer-Verlag Publishers.

Giovannoni, S.J., Britschgi, T.B., Moyer, C.L. and Field, K.G. 1990. Genetic diversity in Sargasso Sea bacterioplankton. *Nature (London)* 345**:**60–63.

Giovannoni, S.J., Delong, E., Olsen, G.J. and Pace, N.R. 1988. Phylogenetic group-specific oligodeoxynucleotide probes for in situ microbial identification. *J. Bacteriol.* 170:720-726.

Giraffa, G. 2004. Studying the dynamics of microbial populations during food fermentation. *FEMS Microbiology Reviews* 28:251–260.

Goh, S.H., Potter, S., Wood, J.O., Hemmingsen, S.M., Reynolds, R.P. and Chow, A.W. 1996. HSP60 gene sequences as universal targets for microbial species identification: studies with coagulase-negative staphylococci. *J Clin Microbiol*. 34(4): 818–823.

Green, B.D. and Keller, M. 2006. Capturing the uncultivated majority. *Current Opinion in Biotechnology* 17:236–240.

Guzel-Seydim, Z.B., Seydim, A.C., Greene, A.K. and Bodine, A.B. 2000. Determination of Organic Acids and Volatile Flavor Substances in Kefir during Fermentation. *Journal of Food Composition and Analysis* 13:35-43.

Guzel-Seydim, Z., Wyffels, T.J., Seydim, A.C. and Greene, A.K. 2005. Turkish kefir and kefir grains: microbial enumeration and electron microscobic observation. *International Journal of Dairy Technology* 23:58-1.

Hall, B.G. 2004. *Phylogenetic Trees Made Easy*. Sunderland: Sinauer Associates Press.

Handelsman, J. 2004. Metagenomics: Application of Genomics to Uncultured Microorganisms. *Microbiology and Molecular Biology Reviews* 68:669–685.

Hayden, R.T., Kolbert, C.P., Hopkins, M.K. and Persing, D.H. 2001. Characterization of culture-derived spiral bacteria by 16S ribosomal RNA gene sequence analysis. *Diagnostic Microbiology and Infectious Disease* 39:55–59.

Heo, J.C. and Lee, S.H. 2006. Isolation and molecular taxonomy of two predominant types of microflora in Kefir. *J Gen Appl Microbiol*. 52(6):375-379.

Higgs, P.G. and Attwood, T.K., 2005. *Bioinformatics and Molecular Evolution*. New York: Blackwell Pulishing.

Liu, S.Q. 2003. Practical implications of lactate and pyruvate metabolism by lactic acid bacteria in food and beverage fermentations. *Int. J. Food Microbiol*. 83:115–131.

Lopitz-Otsoa, F., Rementeria, A., Elguezabal, N. and Garaziar, J. 2006. Kefir: A symbiotic yeast-bacteria community with alleged healthy capabilities. *Revista Iberoamericana de Micologia* 23: 67-74.

Mainville, I., Robert, N., Lee, B. and Farnworth, E.F. 2006. Polyphasic characterization of the lactic acid bacteria in kefir. *Systematic and Applied Microbiology* 29:59–68.

Maiwald, M., Persing, D.H., Tenover, F.C., Versalovic, J., Tang, Y., Unger, E.R., Relman, D.A. and Whit, T.J. 2004. *Broad-Range PCR for Detection and Identification of Bacteria.* Wahington, DC: ASM Press.

Marshall, V.M. 1993. *In: Encyclopaedia of Food Science and Technology*. Chichester: John Wiley and Sons Press.

McGinnis, S. and Madden, T.L. 2004. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Research* 32:20-25.

McHardy, A.C. and Rigoutsos, I. 2007. What's in the mix: phylogenetic classification of metagenome sequence samples. *Current Opinion in Microbiology* 10:499–503.

Mincer, T.J., Fenical, W. and Jensen, P.R. 2005. Culture-Dependent and Culture-Independent Diversity within the Obligate Marine Actinomycete Genus Salinispora. *Applied and Environmental Microbiology* 71:7019–7028.

Mount, D.W. 2004. *Bioinformatics, Sequence and Genome Analysis*. New York: Cold Spring Harbor Laboratory Press.

National Center for Biotechnology Information. 2008. NCBI Nucleotide Databases. http://www.ncbi.nlm.nih.gov/ (accessed February 15, 2008).

Ninane, V., Berben, G., Romnee, J.M. and Oger, R. 2005. Variability of the microbial abundance of a kefir grain starter cultivated in partially controlled conditions. *Biotechnol. Agron. Soc. Environ* 9 (3):191-194.

Oren, A. 2004. Prokaryote Diversity and Taxonomy: Current Status and Future Challenges Philosophical Transactions. *Biological Sciences, Phil. Trans. R. Soc. Lond*. 359; 623-638 .

Otles, S. and Cagindi, O. 2003. Kefir: A Probiotic Dairy-Composition, Nutritional and Therapeutic Aspects. *Pakistan Journal of Nutrition* 2 (2): 54-59.

Pace, N.R. 1997. A molecular view of microbial diversity and the biosphere. *Science,* 276, 734–740.

Pace, N.R., Stahl, D.A., Lane, D.J. and Olsen, G.J. 1985. Analyzing natural microbial populations by rRNA sequences. *ASM News* 51: 4–12.

Petti, C.A., Polage, C.R. and Schreckenberger, P. 2005. The Role of 16S rRNA Gene Sequencing in Identification of Microorganisms Misidentified by Conventional Methods. *Journal of Clinical Microbiology* 43:6123–6125.

Pevsner, J. 2003. *Bioinformatics and Functional Genomics*. New York: John Wiley and Sons Publishers.

Ranjard, L., Poly, F. and Nazaret, S. 2000. Monitoring complex bacterial communities using culture-independent molecular techniques: application to soil environment. *Res. Microbiol*. 151:167–177.

Saitou, N. and Nei, M. 1987. The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees. *Mol. Biol. Evol*. 4(4):406-425.

Santos, A., San Mauro, M., Sanchez, A., Torres, J.M. and Marquina, D. 2003. The Antimicrobial Properties of Different Strains of Lactobacillus spp. Isolated from Kefir. *System. Appl. Microbiol* 26:434–437.

Schabereiter-Gurtner, C., Maca, S., Rolleke, S., Nigl, K., Lukas, J., Hirschl, A., Lubitz, W. and Barisani-Asenbauer, T. 2001. 16S rDNA-Based Identification of Bacteria from Conjunctival Swabs by PCR and DGGE Fingerprinting. *Investigative Ophthalmology & Visual Science* 42(6):1164-1171.

Schloss, P.D. and Handelsman, J. 2008. A statistical toolbox for metagenomics: assessing functional diversity in microbial communities. *BMC Bioinformatics* 9:34-40.

Schloss, P.D. and Handelsman, J. 2003. Biotechnological prospects from metagenomics. *Current Opinion in Biotechnology* 14:303-310.

Schloss, P.D., Hay, G.A., Wilson, D.B., Gossett, J.M. and Walker, L.P. 2005. Quantifying bacterial population dynamics in compost using 16S rRNA gene probes. *Appl Microbiol Biotechnol* 66: 457–463.

Schoevers, A. and Britz, T.J. 2003. Influence of different culturing conditions on kefir grain increase. *International Journal of Dairy Technology* 56:3-10.

Simova, E., Beshkova, D., Angelov, A., Hristozova, T., Frengova, G. and Spasov, Z. 2002. Lactic acid bacteria and yeasts in kefir grains and kefir made from them. *Journal of Industrial Microbiology & Biotechnology* 28:1–6.

Smit, S., Widmann, J. and Knight, R. 2007. Evolutionary rates vary among rRNA structural Elements. *Nucleic Acids Research* 35:3339–3354.

Song, Y., Liu, C., McTeague, M. and Sydney, M.F. 2003. 16S Ribosomal DNA Sequence-Based Analysis of Clinically Significant Gram-Positive Anaerobic Cocci. *Journal of Clinical Microbiology* 41:1363–1369.

Steele, H.L. and Streit, W.R. 2005. Metagenomics: Advances in ecology and biotechnology. *FEMS Microbiology Letters* 247:105-111.

Tang, Y.W., Ellis, N.M., Hopkins, M.K., Smith, D.H., Dodge, D.E. and Persing, D.H. 1998. Comparison of Phenotypic and Genotypic Techniques for Identification of Unusual Aerobic Pathogenic Gram-Negative Bacilli. *Journal of Clinical Microbiology* 36:3674–3679.

Torsvik, V., Goksøyr, J. and Daae, F.L. 1990. High diversity in DNA of soil bacteria. *Appl. Environ. Microbiol* 56:782–787.

Tsai, J., Hsueh, P.R., Lin, H.M., Chang, H.J., Ho, S.W. and Teng, L.J. 2005. Identification of Clinically Relevant Enterococcus Species by Direct

Sequencing of groES and Spacer Region. *Journal of Clinical Microbiology* 43:235-241.

UVA Fasta Server. FASTA Sequence Comparison at the University of Virginia. http://fasta.bioch.virginia.edu/fasta_www2/fasta_list2.shtml. (accessed April 24, 2008)

Vancanneyt, M., Mengaud, J., Cleenwerck, I., Vanhonacker, K., Hoste, B., Dawyndt, P., Degivry, M.C., Ringuet, D., Janssens, D. and Swings, J. 2004. Reclassification of Lactobacillus kefirgranum Takizawa et al. 1994 as Lactobacillus kefiranofaciens subsp. kefirgranum subsp. nov. and emended description of L. kefiranofaciens Fujisawa et al. 1988. *International Journal of Systematic and Evolutionary Microbiology* 54:551–556.

Vinderola, G., Perdigo´n, G., Duarte, J., Farnworth, E. and Matar, C. 2006. Effects of the oral administration of the exopolysaccharide produced by Lactobacillus kefiranofaciens on the gut mucosal immunity. *Cytokine* 36:254–260.

Watanabe, K., Kodama, Y. and Harayama, S. 2001. Design and evaluation of PCR primers to amplify bacterial 16S ribosomal DNA fragments used for community fingerprinting. *Journal of Microbiological Methods* 44:253-262.

Wheeler, D.L., Chappey, C., Lash, A.E., Leipe, D.D., Madden, T.L., Schuler, G.D., Tatusova, T.A. and Rapp, B.A. 2000. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 28(1):10-14.

Witthuhn, R.C., Schoeman, T. and Britz, T.J. 2004. Isolation and characterization of the microbial population of different South African kefir grains. *International Journal of Dairy Technology* 57:1-8.

Witthuhn, R.C., Schoeman, T. and Britz, T.J. 2005. Characterisation of the microbial population at different stages of Kefir production and Kefir grain mass cultivation. *International Dairy Journal* 15:383–389.

Woese, C.R. 1987. Bacterial Evolution. *Microbiological Reviews*. 34:221-271.

Woese, C.R., Winkers, T.S. and Gutell, R.R. 1990. Architecture of ribosomal RNA: Constraints on the sequence of "tetra-loops. *Proc. Nati. Acad. Sci. USA* 87:8467-8471.

Wolfgang, R.S. and Schmitz, R.A. 2004. Metagenomics – the key to the uncultured microbes. *Current Opinion in Microbiology* 7:492–498.

Yokoi, H., Watanabe, T.S. and Fuji, Y.S. 1990. Isolation and Characterization of Polysaccharide-Producing Bacteria from Kefir Grains. *Journal of Dairy Science* 73:1684-1689.

Yokoi, H., Watanabe, T., Fujii, Y., Mukai, T., Toba, T. and Adachi, S. 1991. Some taxonomical characteristics of encapsulated Lactobacillus sp. KPB-167B isolated from kefir grains and characterization of its extracellular polysaccharide. *Int. J. Food Microbiol*. 13:57–264

Yuksekdag, Z.N., Beyatli, Y. and Aslim, B. 2004. Determination of some characteristics coccoid forms of lactic acid bacteria isolated from Turkish kefirs with natural probiotics. *Lebensm Wiss. Technol Food Sci.Technol*. 37:663-667.

Zubillaga, M., Weill, R., Postaire, E., Goldman, C., Caro, B.R. and Bioch, B.J. 2001. Effect of probiotics and functional foods and their use in different diseases. *Nutrition Research* 21:569–579.

APPENDIX A

PARTIAL 16S RIBOSOMAL RNA SEQUENCES OF IDENTIFIED BACTERIAL COMMUNITIES IN KEFIR

Lactobacillus kefiranofaciens subsp. kefirgranum 16S rRNA gene

CCAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTCCACAATGGACGCAAGTCTGATGGAGC
AACGCCGCGTGAGTGAAGAAGGTTTTCGGACCGTAAAGCTCTGTTGTTGGTGAAGAAGGATA
GAGGTAGTAACTGGCCTTTATTTGACGGTAATCAACCAGAAAGTCACGGCTAACTACGTGCC
AGCAGCCGCGGTAATACGTAGGTGGCAAGCGTTGTCCGGATTTATTGGGCGTAAAGCGAGCG
CAGGCGGAAGAATAAGTCTGATGTGAAAGCCCTCGGCTTAACCGAGGAATTGCATCGGAAAC
TGTTTTTCTTGAGTGCAGAAGAGGAGAGTAGAACTCCATGTGTAGCGGTGGAATGCGTAGAT
ATATGGAAGAATACCAGTGGCGAAGGCGGCTCTCTGGTCTGCAACTGACGCTGAGGCTCGAA
AGCATGGGTAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGATGAGTGCTAA
GTGTTGGGAGGCTTCCGCCTCTCAGTGCTGCAGCTAACGCATTAAGCACTCCGCCTGGGGAGT
ACGACCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGT
GGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCTAGTGCCATTTGTAGA
GATACAAAGTTCCCTTCGGGGACGCTAAGACAGGTGGTGCATGGCTGTCGTCAGCTCGTGTC
GTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCT

Acetobacter lovaniensis gene for 16S ribosomal RNA gene

CCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGGGCAACCCTGATCCAG
CAATGCCGCGTGTGTGAAGAAGGTCTTCGGATTGTAAAGCACTTTCGACGGGGACGATGATG
ACGGTACCCGTAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGG
GGCTAGCGTTGCTCGGAATGACTGGGCGTAAAGGGCGTGTAGGCGGTTTACACAGTCAGATG
TGAAATCCCCGGGCGTAACCTGGGAGCTGCATTTGATACGTGTAGACTAGAGTGTGAGAGAG
GGTTGTGGAATTCCCAGTGTAGAGGTGAAATTCGTAGATATTGGGAAGAACACCGGTGGCGA
AGGCGGCAACCTGGCTCATTACTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATT
AGATACCCTGGTAGTCCACGCTGTAAACGATGTGTGCTAGATGTTGGGTAACTTTGTTATTCA
GTGTCGCAGTTAACGCGTTAAGCACACCGCCTGGGGAGTACGGCCGCAAGGTTGAAACTCAA
AGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGAAGCAACGCGCA
GAACCTTACCAGGGCTTGAATGTAGAGGCTGTATTCAGAGATGGATATTTCCCGCAAGGGAC
CTCTAACACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCG
CAACGAGCGCAACCCT

Lactococcus lactis subsp. lactis 16S ribosomal RNA

CCAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTCGGCAATGGACGAAAGTCTGACCGAG
CAACGCCGCGTGAGTGAAGAAGGTTTTCGGATCGTAAAACTCTGTTGGTAGAGAAGAACGTT
GGTGAGAGTGGAAAGCTCATCAAGTGACGGTAACTACCCAGAAAGGGACGGCTAACTACGT
GCCAGCAGCCGCGGTAATACGTAGGTCCCGAGCGTTGTCCGGATTTATTGGGCGTAAAGCGA
GCGCAGGTGGTTTATTAAGTCTGGTGTAAAAGGCAGTGGCTCAACCATTGTATGCATTGGAA
ACTGGTAGACTTGAGTGCAGGAGAGGAGAGTGGAATTCCATGTGTAGCGGTGAAATGCGTAG
ATATATGGAGGAACACCGGTGGCGAAAGCGGCTCTCTGGCCTGTAACTGACACTGAGGCTCG
AAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTGCT
AGATGTAGGGAGCTATAAGTTCTCTGTATCGCAGCTAACGCAATAAGCACTCCGCCTGGGGA
GTACGACCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCAT
GTGGTTTAATTCGAAGCAACGAAGAACCTTACCAGGTCTTGACATACTCGTGCTATTCCTAGA
GATAGAAGTTCCTTCGGGACACGGGATACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGT
GAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCT

Lactobacillus helveticus 16S ribosomal RNA gene

CCAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTCCGCAATGGACGCAAGTCTGATGGAGC
AACGCCGCGTGAGTGAAGAAGGTTTTCGGATCGTAAAGCTCTGTTGTTGGTGAAGAAGGATA
GAGGTGGTAACTGGCCTTTATTTGACGGTAATCAACCAGAAAGTCACGGCTAACTACGTGCC
AGCAGCCGCGGTAATACGTAGGTGGCAAGCGTTGTCCGGATTTATTGGGCGTAAAGCGAGCG
CAGGCGGAAGAATAAGTCTGATGTGAAAGCCCTCGGCTTAACCGAGGAACTGCATCGGAAA
CTGTTTTTCTTGAGTGCAGAAGAGGAGAGTGGAACTCCATGTGTAGCGGTGGAATGCGTAGA
TATATGGAAGAACACCAGTGGCGAAGGCGACTCTCTGGTCTGCAACTGACGCTGAGGCTCGA
AAGCATGGGTAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGATGAGTGCTA
AGTGTTGGGAGGTTTCCGCCTCTCAGTGCTGCAGCTAACGCATTAAGCACTCCGCCTGGGGA
GTACGACCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCAT
GTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCTAGTGCCATCCTAA
GAGATTAGGAGTTCCCTTCGGGGACGCTAAGACAGGTGGTGCATGGCTGTCGTCAGCTCGTG
TCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCT

Leuconostoc mesenteroides 16S ribosomal RNA gene

CCAGACTCCTATCGGGAGGCAGCAGTAGGGAATCTTCCACAATGGGCGAAAGCCTGATGGAG
CAACGCCGCGTGTGTGATGAAGGCTTTCGGGTCGTAAAGCACTGTTGTATGGGAAGAACAGC
TAGAATAGGAAATGATTTTAGTTTGACGGTACCATACCAGAAAGGGACGGCTAAATACGTGC
CAGCAGCCGCGGTAATACGTATGTCCCGAGCGTGATCCGGGTTTATTGGGCGTAAAGCGAGC
GCAGACGGTTTATTAAGTCTGATGTGAAAGCCCGGAGCTCAACTCCGGAATGGCATTGGAAA
CTGGTTAACTTGAGTGCAGTAGAGGTAAGTGGAACTCCATGTGTAGCGGTGGAATGCGTAGA
TATATGGAAGAACACCAGTGGCGAAGGCGGCTTACTGGACTGCAACTGACGTTGAGGCTCGA
AAGTGTGGGTAGCAAACAGGATTAGATACCCTGGTAGTCCACACCGTAAACGATGAACACTA
GGTGTTAGGAGGTTTCCGCCTCTTAGTGCCGAAGCTAACGCATTAAGTGTTCCGCCTGGGGAG
TACGACCGCAAGGTTGAAACTCAAAGGAATTGACGGGGACCCGCACAAGCGGTGGAGCATG
TGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCCTTTGAAGCTTTTAGA
GATAGAAGTGTTCTCTTCGGAGACAAAGTGACAGGTGGTGCATGGTCGTCGTCAGCTCGTGT
CGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCT

Lactobacillus kefiri or *parabuchneri* 16S ribosomal RNA gene

CCAGACTCCTACGGGAGGCAGCAGTAGTGGAATCTTCCACAATGGACGAAAGTCTGATGGAG
CAACGCCGCGTGAGTGATGAAGGGTTTCGGCTCGTAAAACTCTGTTGTTGGAGAAGAACAGG
TGTCAGAGTAACTGTTGACATCTTGACGGTATCCAACCAGAAAGCCACGGCTAACTACGTGC
CAGCAGCCGCGGTAATACGTAGGTGGCAAGCGTTGTCCGGATTTATTGGGCGTAAAGCGAGC
GCAGGCGGTTTCTTAGGTCTGATGTGAAAGCCTTCGGCTTAACCGGAGAAGTGCATCGGAAA
CCAGGAGGCTTGAGTGCAGAAGAGGGCAGTGGAACTCCATGTGTAGCGGTGAAATGCGTAG
ATATATGGAAGAACACCAGTGGCGAAGGCGGCTGTCTGGTCTGTAACTGACGCTGAGGCTCG
AAAGCATGGGTAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGATGAGTGCT
AAGTGTTGGAGGGTTTCCGCCCTTCAGTGCTGCAGCTAACGCATTAAGCACTCCGCCTGGGG
AGTACGACCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCA
TGTGGTTTAATTCGATGCTACGCGAAGAACCTTACCAGGTCTTGACATCTTCTGCCAACCTAA
GAGATTAGGCGTTCCCTTCGGGGACAGAATGACAGGTGGTGCATGGTTGTCGTCAGCTCGTG
TCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCT

Acetobacter syzygii gene for 16S rRNA gene

CCAGACTCTCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGGGCAACCCTGATCCAG
CAATGCCGCGTGTGTGAAGAAGGTCTTCGGATTGTAAAGCACTTTCGACGGGGACGATGATG
ACGGTACCCGTGGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGG
GGCTAGCGTTGCTCGGAATGACTGGGCGTAAAGGGCGTGTAGGCGGTTTGTACAGTCAGATG
TGAAATCCCCGGGCTTAACCTGGGAGCTGCATTTGATACGTGCAGACTAGAGTGTGAGAGAG
GGTTGTGGAATTCCCAGTGTAGAGGTGAAATTCGTAGATATTGGGAAGAACACCGGTGGCGA
AGGCGGCAACCTGGCTCATTACTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATT
AGATACCCTGGTAGTCCACGCTGTAAACGATGTGTGCTAGATGTTGGGTGACTTTGTCATTCA
GTGTCGCAGTTAACGCGTTAAGCACACCGCCTGGGGAGTACGGCCGCAAGGTTGAAACTCAA
AGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGAAGCAACGCGCA
GAACCTTACCAGGGCTTGAATGTAGAGGCTGTATTCAGAGATGGATATTTCCCGCAAGGGAC
CTCTAACACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCG
CAACGAGCGCAACCAC

Enterococcus faecium gene for 16S ribosomal RNA

AGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATG
CACCACCTGTCACCTTGCCCCCGAAGGGGAAGCTCTATCTCTAGAGTGGTCAAAGGATGTCA
AGACCTGGCAAGGTTCTTCGCGTTGCTTCGAATTAAACCACATGCTCCACCGCTTGTGCGGGC
CCCGTCAATTCCTTTGAGTTTCAACCTTGCGGTCGTACTCCCCAGGCGGAGTGCTTAATGCGT
TAGCTGCAGCACTGAAGGGCGGAAACCCTCCAACACTTAGCACTCATCGTTTACGGCGTGGA
CTACCAGGGTATCTAATCCTGTTTGCTCCCCACGCTTTCGAGCCTCAGCGTCAGTTACAGACC
AGAGAGCCGCCTTCGCCACTGGTGTTCCTCCATATATCTACGCATTTCACCGCTACACATGGA
ATTCCACTCTCCTCTTCTGCACTCAAGTCTCCCAGTTTCCAATGACCCTCCCCGGTTGAGCCGG
GGGCTTTCACATCAGACTTAAGAAACCGCCTGCGCTCGCTTTACGCCCAATAAATCCGGACA
ACGCTTGCCACCTACGTATTACCGCGGCTGCTGGCACGTAGTTAGCCGTGGCTTTCTGGTTAG
ATACCGTCAAGGGATGAACAGTTACTCTCATCCTTGTTCTTCTCTAACAACAGAGTTTTACGA
TCCGAAAACCTTCTTCACTCACGCGGCGTTGCTCGGTCAGACTTTCGTCCATTGCCGAAGATT
CCATACTGCTGCCTCCCGTAGGAGTCTGG

APPENDIX B

THE BASE HOMOLOGY BETWEEN USED PRIMERS AND 16S rRNA GENES OF IDENTIFIED BACTERIAL SPECIES

Table B.1. Base alignment of the used universal bacterial primers and 16S rDNA regions of identified bacterial species in kefir

Identified Species	**Primers (5' → 3')**	
	E334F	**E1115R**
	CCAGACTCCTACGGGAGGCAG	CAACGAGCGCAACCCT
Lb. kefiranofaciens	CCAA**A**CTCCTACGGGAGGCAG	CAACGAGCGCAACCCT
Lb. helveticus	CCAA**A**CTCCTACGGGAGGCAG	CAACGAGCGCAACCCT
Lb. kefiri/parabucneri	CCAA**A**CTCCTACGGGAGGCAG	CAACGAGCGCAACCCT
E. faecium	CCAA**A**CTCCTACGGGAGGCAG	CAACGAGCGCAACCCT
A. lovaniensis	CCAGACTCCTACGGGAGGCAG	CAACGAGCGCAACCC**C**
A .syzygii	CCAGACTCCTACGGGAGGCAG	CAACGAGCGCAACCC**C**
Leu. mesenteroides	CCAA**A**CTCCTACGGGAGGC**T**G	CAACGAGCGCAACCCT
Lc. lactis subsp. lactis	CCAGACTCCTACGGGAGGCAG	CAACGAGCGCAACCC**C**
	* * * _ * * ** ** * * ** * * * * * _ *	* * * * * ** ** * ** * * * * _

Asterisk indicates the base homology between primers and related DNA fragments of determined species. Bold letters show the not conserved bases in each species

APPENDIX C

BUFFERS AND STOCK SOLUTIONS

Luria Bertani (LB) broth, per liter

10 g tryptone, 5 g yeast extract, 5 g NaCl and dH_2O up to 1 L.

Luria Bertani (LB) agar, per liter

10 g tryptone, 5 g yeast extract, 5 g NaCl, 15 g agar and dH_2O up to 1 L.

SOC Medium, per 100 ml

2 g Tryptone, 0.5 g Yeast Extract, 1 ml 1M NaCl, 0.25 ml 1M KCl, 1 ml 2M Mg^{2+} Stock, 1 ml 2M Glucose were dissolved in dH2O up to 100 ml.

SOB Medium, per 100 ml

2 g Tryptone, 0.5 g Yeast Extract, 1 ml 1M NaCl, 0.02 g KCl, 1 ml 1M MgCl, 1 ml 1M $MgSO_4$ were dissolved in deionized dH_2O up to 100 ml.

TB Medium, per 100 ml

0.3 g PIPES, 3 ml 1 M $CaCl_2$ and 1.85 g KCl were dissolved in 100 ml deionized water and the solution pH was adjusted to 6.7 with KOH. Then 1.4 g $MnCl_2$was added and the solution was filtered sterilized.

3M Sodium Acetate (Ph 5.2)

408.1 g sodium acetate.$3H_2O$ was dissolved in 800 ml of deionized water and the pH is adjusted to 5.2 with glacial acetic acid. The volume is adjusted to 1000 ml. Solution was sterilized by autoclaving.

Stop Solution for DNA Sequencing Reaction

2 µl of 3M sodium acetate and 2 µl of 100mM Na_2-EDTA were mixed to prepare fresh stop solution before each reaction.

95% Ethanol/Water, 70% Ethanol/Water

Ultra pure H2O was used to dilute the 100% (molecular biology grade) ethanol to required concentration as mentioned above and the solutions were stored at -20°C.

10% Ammonium Persulfate

0.1 g ammonium persulfate was dissolved in 1.0 ml dH_2O.

2x Gel Loading Dye Reagents

0.25 ml 2% bromophenol blue, 0.25 ml 2% xylene cyanol, 7 ml 100% glycerol and 2.5 ml dH2O were mixed in total volume of 10 ml**.**

DCode Dye Solution Reagents

0.05 g bromophenol blue, 0.05 g xylene cyanol were added in 1X TAE buffer at 10 ml final volume.

40% and 70% Denaturing Solutions

For 40% denaturing solution, 15 ml 40% acrylamide/Bis, 2 ml 50x TAE buffer, 16 ml formamide and 16.8 g urea were dissolved in 100 ml dH_2O**.** For 70% denaturing solution, 15 ml 40% acrylamide/Bis, 2 ml 50x TAE buffer, 28 ml formamide and 29.4 g urea were dissolved in 100 ml dH_2O. It was allowed to degas for 10-15 minutes and filtered through a 0.45 µl filter**.**

40% acrylamide/Bis

38.93 g acrylamide and 1.07 g Bis-acrylamide was dissolved in total 100 ml dH_2O. The solution was filtered through a 0.45 µl filter.

50X TAE Electrophoresis Buffer

242 g Tris base and 37.2 g Na2EDTA (2H2O) was dissolved in 900 ml deionized water. After 57.1 ml glacial acetic acid was added, the volume was adjusted to 1 liter with deionized water

1× TAE

20 ml of 50× TAE buffer was mixed with 980 ml of deionized water and 1× TAE buffer was obtained.

Ethidium Bromide Stock Solution (10 Mg/Ml)

0.2g ethidium bromide (EtBr) was dissolved in 20 ml dH_2O. It was mixed well and stored at room temperature in dark.