CHARACTERIZATION OF THE INTESTINAL MICROBIOME AND LACTOBACILLI COMMUNITY: INFLUENCE OF DIETARY AND ENVIRONMENTAL FACTORS

BY

LONG CHEN

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Food Science and Human Nutrition
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Doctoral Committee:

Professor Sharon Donovan, Chair
Assistant Professor Michael Miller, Director of Research
Professor Bruce Chassy
Assistant Professor Yong-su Jin

ABSTRACT

The human gastrointestinal (GI) tract is colonized by a dense and diverse bacterial community called the commensal microbiota, which plays an important role in the overall health of individuals. The complexity of the GI microbiota composition and the limitations of experimental tools presented many challenges throughout this study. Identification and comparison of individual members of the microbiota and determination of the impacts of multiple factors on the development of microbiota were difficult. Therefore, this report is a comprehensive study utilizing traditional culture-dependent methods along with next generation molecular biological tools. This study was an investigation of the neonatal piglets' intestinal microbiome and lactobacilli community structure. In Chapter 2, an up to date comparison of the current methods used in microbiota research followed by a discussion of the known impact of different environmental and dietary factors. In Chapter 3, a new protocol of terminal restriction fragment length polymorphism (T-RFLP) analysis was developed to facilitate in depth analysis of the lactobacilli community structure. In Chapter 4, both culture-dependent methods and the novel T-RFLP analysis were used to reveal the impact of piglet age and route of delivery on ileal *Lactobacillus* diversity. In Chapter 5, pyrosequencing, a next generation molecular biological tool, was used to investigate how route of delivery and nutrition altered the piglet ileal microbiota. Multiple important conclusions were made including: (1) our novel lactobacilli T-RFLP protocol can characterize the lactobacilli community in complex samples; (2) route of delivery did affect the ileal lactobacilli community; and (3) piglet nutrition is a major shaping force of bacterial microbiota at genus and species level. The conclusions from this study are important additions to the field of Gastrointestinal Microbiology and can have long-lasting impacts on infant nutrition study in the future.

Keywords: microbiota, microbiome, pyrosequencing, T-RFLP, lactobacilli, GI tract

ACKNOWLEDGEMENTS

It is with great pleasure to thank many people who were instrumental in the completion of my Doctoral degree and my thesis project.

When I was applying to graduate schools in 2007, Dr. Bruce Chassy gave me one of the most important suggestions in my life. He said, “Choosing the right advisor is more important than choosing your wife”. At first, I thought he was joking. Throughout the years, I gradually understood and agreed. I am so happy with my choice for advisor.

It is difficult to state my gratitude to my doctoral advisor, Dr. Michael Miller. Dr. Miller’s passion, guidance, and discipline have been indispensable to my growth as a scientist and as a person over these past five years. I am especially grateful to Dr. Miller for his devotion to his students’ education and their success. He has always encouraged me to make decisions regarding my career instead of making those decisions for me. Throughout my time at Illinois, Dr. Miller gave me the strength to continue even when facing challenges while preparing for exams, working the in lab and writing manuscripts and my thesis.. His qualities as a doctoral advisor is an example of how an advisor should go above and beyond to make sure their students are prepared for whatever the next step in their journeys may be.

I also would like to thank my advisory committee who guided me through all these years. Thank you to Dr. Sharon Donovan, Dr. Bruce Chassy and Dr. Yong-su Jin for being my mentors. A special thanks to Dr. Donovan for generously providing animals, equipment and human resources (Mei and Min) for my research. I am very grateful to Dr. Chassy who offered great assistance in my application for University of Illinois and for introducing Dr. Miller to me.

I need to make a separate paragraph here for my friend, Jennifer Hoeflinger. I am ever so grateful to her for all the things she taught me throughout these years: the American culture, English words, lab protocols, statistics, time management and much much more. Jeni is kind, reliable and always

willing to help. It's hard to imagine that I could graduate without her great help; to lose her would be like losing an arm, maybe two. I have had the great fortune during my graduate career of having developed such a good friendship with Jeni.

Special thanks to my labmates, Dr. Taksawan Thongaram, Maxwell Van Tassell, Dr. Melissa Kaczmarczyk, Dimitri Kashtanov, Matt Teasdale, Chris Kasia, Claire Zoellner, Kevin McDonald and Alyssa Francl. It was them who shared very important moments with me during the best ages of my life.

I also would like to thank Dr. Mei Wang, Dr. Min Li, Dr. Carl Yeoman, Dr. Marcia Monaco, Dr. Brittany Vester, Dr. Gregory Freund, and Dan Hoeflinger. All these individuals had a great impact on my research and made my life much easier.

Lastly, and most importantly, I would thank my parents, my father Zichang Chen and my mother Yajun Ni. They bore me, raised me, supported me, taught me, and loved me. To them I dedicate this thesis.

TABLE OF CONTENTS

CHAPTER 1 Introduction……1

1.1 Motivation……1

1.2 Specific Aims……3

CHAPTER 2 Literature Review……4

2.1 Commensal Microbiota……4

2.2 *Lactobacillus* spp. ……4

2.3 Methods for Investigating Human GI Microbiota ……6

2.4 Methods for Isolating and Identifying *Lactobacillus* spp. in GI Tract……13

2.5 Factors Impacting Early Colonization of GI Microbiota……19

2.6 Summary……23

CHAPTER 3 Development of a *Lactobacillus* Specific T-RFLP Method to Determine Lactobacilli Diversity in Complex Samples……24

3.1 Abstract……24

3.2 Introduction……24

3.3 Methods and Materials……26

3.4 Results……33

3.5 Discussion……38

CHAPTER 4 Impact of Piglet Age and Route of Delivery on Ileal *Lactobacillus* Diversity……41

4.1 Abstract ……41

4.2 Introduction ……41

4.3 Methods and Materials……44

4.4 Results……49

4.5 Discussion……57

CHAPTER 5 Route of Delivery and Nutrition Alter the Ileal Microbiota……61

5.1 Abstract……61

5.2 Introduction……62

5.3 Methods and Materials……64

5.4 Results……71

5.5 Discussion……81

CHAPTER 6 Overall Conclusions and Future Direction ……84

REFERENCES……86

APPENDIX ……98

Table S1……98

Table S2……101

Code S1……102

Code S2……108

Code S3……111

CHAPTER 1. INTRODUCTION

1.1 Motivation

The bacterial microbiota in human gastrointestinal (GI) tract plays an important role in overall health (Cummings and Macfarlane, 1997). It aids in digestion and the extraction of energy from the diet, it provides nutrition for enterocytes, contributes to the development of the host's immune system, and acts against pathogen invasion (Isolauri *et al.*, 2002; Vaughan *et al.*, 2002). Alterations of this complex ecosystem have been associated with various diseases in humans, including inflammatory bowel disease (IBD) (Sokol *et al.*, 2006; Frank *et al.*, 2007), irritable bowel syndrome (IBS) (Kassinen *et al.*, 2007), and obesity (Ley *et al.*, 2006; Zhang *et al.*, 2009). Specifically, lactobacilli are one of the most important bacterial genera in the GI tract. Lactobacilli are a diverse assemblage of 140 species, many of which have been isolated from human and animal fecal sources (Wall *et al.*, 2007). Lactobacilli have been shown to benefit the host's health (Fuller 1989; Giraffa *et al.*, 2010). Some species have been widely exploited as probiotics and are extensively used in consumer products (Euzeby *et al.*, 1997; Sanders *et al.*, 2003).

It is known that the development of the GI microbiota is influenced by several factors, including host genotype and physiology, environmental exposure to microbes, antibiotic use, route of delivery (Mackie *et al.*, 1999), and the types of nutrition after birth (breast-fed vs. formula-fed)(Herfel *et al.*, 2009; Stark *et al.*, 1982). It has been reported that infants born by caesarean section seem to have delayed bacterial colonization and altered infant intestinal microbiota (Eggesbo *et al.,* 2003). This pattern of colonization may increase the chance of morbidities in the neonate, which may persist through the life cycle (Sjögren *et al.,* 2009). It is also believed that vaginal delivery affects lactobacilli diversity in the neonate that confer functional benefits compared to infants with C-section delivery, including protection against GI infection and reduction of diet related allergies (Murgas *et al.,* 2011). Short term impact on lactobacilli composition by

route of delivery has been observed, whether the route of delivery can impact the long-term development of intestinal lactobacilli community is still unclear (Matsumiya *et al.*, 2002).

Diet has been regarded as a major factor driving the microbial successional patterns after acquisition (Mackie *et al.*, 1999). Human milk oligosaccharides (HMO), which are unique components in breast milk, have been shown to increase the number of some beneficial species in the gut (Tuohy *et al.*, 2010). Prebiotic supplementation of infant formula with oligosaccharides such as galactooligosaccharide (GOS) and polydextrose (PDX) has a similar effect (Ribeiro *et al.*, 2011; Scalabrin *et al.*, 2012). Introduction of solid food at weaning diminish differences between infants delivered by different routes, and the microbiota becomes more complex (Mackie *et al.*, 1999).

The difficulty in accurately describing a bacterial community in the GI tract has been a major challenge in studies of microbial development. Standard culture-based methods, such as plating with selective agents such as bile, esculin or antibiotics, can be used for selective enrichment for certain species (O'Sullivan *et al.*, 1999.). Recently, the movement away from culture-dependent methods to culture-independent methods has allowed rapid growth in the field of GI microbiota research. Molecular biological tools such as bacterial Tag-Encoded FLX Amplicon Pyrosequencing (bTEFAP) and Terminal Restriction Fragment Length Polymorphism (T-RFLP) have allowed investigators to study the impact of various dietary and environmental factors on microbial development.

The overall objective of this dissertation was to determine the impact of nutrition, age and route of delivery on the ileal microbiota using multiple approaches. Our central hypothesis are: [1] Nutrition, age and route of delivery impact the ileal microbiota of infants; [2] There is significant bacterial diversity at sub-genus level in the ileal microbiota; [3] Newly developed T-RFLP and bTEFAP are capable to characterize complex sample and to identify potentially novel taxa from cultivable cells. In order to test these hypotheses, the following specific aims were undertaken:

1.2 Specific Aims

Specific Aim 1:

Develop a *Lactobacillus* specific T-RFLP protocol based on the *Lactobacillus hsp60* gene diversity. The primary objective of this aim was to develop a T-RFLP protocol using the *hsp60* gene rather than 16S rDNA as a template. This modification offers higher resolution to identify very closely related *Lactobacillus* spp. We optimized conditions for PCR and restriction enzyme digestion in order to minimize sampling, extraction and template biases.

Specific Aim 2:

Utilize the newly-developed *Lactobacillus hsp60*-based T-RFLP, together with the traditional culture-dependent methods including carbohydrate utilization tests, to characterize the composition of *Lactobacillus* spp. in piglet ileal content samples.

Specific Aim 3:

Investigate the piglet ileal bacterial population by 16S rDNA-based Bacterial Tag-Encoded FLX Amplicon Pyrosequencing (bTEFAP), a high-throughput sequencing tool. Determine the impact of route of delivery and nutrition on bacterial composition, and more specifically, the lactobacilli composition at the genus and species level.

CHAPTER 2. LITERATURE REVIEW

2.1 Commensal Microbiota

The gastrointestinal (GI) tract of mammals is considered sterile at birth (Walter *et al.*, 2010). However, it is colonized rapidly by a dense and diverse bacterial community, which is called the commensal microbiota. A previous study in the Miller lab showed that more than 900 species of bacteria reside in the gut of a healthy adult person (Fernandez-Raudales et al., 2012). The total intestinal bacterial population often exceeds 10^{11} CFU/g of feces (Hungate *et al.*, 1966). These microorganisms come from a variety of sources including the mother, diet, and the surrounding environment (Rebecca *et al.*, 2006). The compositions of bacterial microbiota are very different between individuals (Hayashi *et al.,* 2002; Zoetendal *et al.,* 1998), and are modified by diet, habits and the environment, throughout the person's lifespan (Dominguez-Bello *et al.*, 2011). **Fig. 2.1** shows a typical human adult gut microbiota composition in different sections of GI tract.

2.2 *Lactobacillus* spp.

Lactobacilli are one of the most important bacterial genera in the GI tract. Lactobacilli are gram-positive, catalase negative, non-motile, non-sporulating, facultative anaerobes which grow under microaerophilic to strictly anaerobic conditions (Klein *et al.*, 1998). The genus *Lactobacillus* contains a diverse assemblage of 140 species (Euzeby *et al.*, 1997); many of which have been shown to benefit the host's health and are extensively used in consumer products. Lactobacilli have been included as probiotics, which is defined as living microorganisms that, when administered in adequate amounts, confer a health benefit for the host (Sanders *et al.*, 2003). The probiotic effects of some *Lactobacillus* spp. make it an important bacterial genus for new applications, specifically in the food and health industries (Claesson *et al.*, 2006).

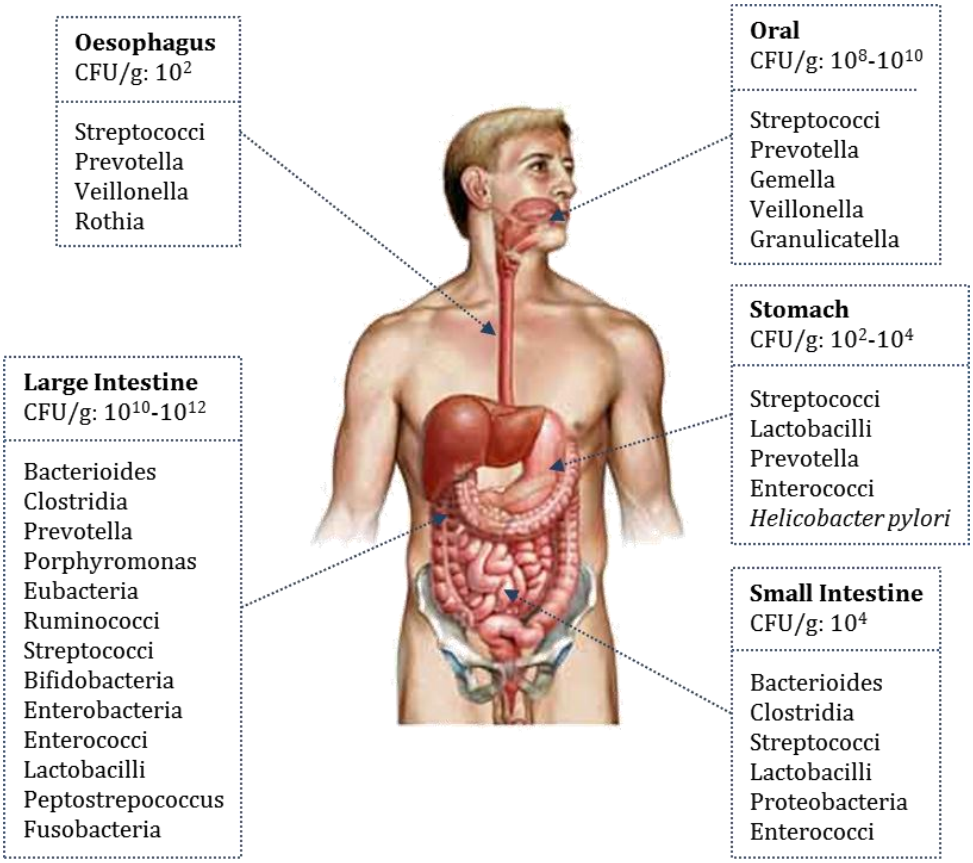


Fig. 2.1 Schematic representation of the human adult gastrointestinal microbiota composition (the relative abundance decrease in each box). Many of these have recently been identified in the gut microbiota however their activity is still largely unknown (Manson *et al.*, 2008; Roberfroid *et al.*, 2010).

Community taxonomic research, which studies the composition of bacterial species in the community, is usually the first step of GI tract lactobacilli research. The taxonomy of lactobacilli associated with the human GI tract has been studied extensively (Salminen *et al.,* 1998). However, the high degree of diversity and complex phylogeny of *Lactobacillus* spp. presents a challenge when characterizing or exploiting individual strains (Schleifer and Ludwig, 1995). Some closely related

species have been re-classified, and a few new species have recently been found using novel molecular technologies like RFLP (Blaiotta *et al.*, 2008) or traditional methods like the API 50 test (Pang *et al.* 2011). Application of more accurate molecular biological tools is necessary for a better understanding of their diversity and functionality.

2.3 Methods for Investigating the Human GI Microbiota

Although investigating the composition of the bacterial community is the foundation of mammalian gut microbiota research, there were many challenges arise when describing the bacterial species present. Hundreds of protocols were developed to separate, isolate, cultivate, identify and compare individual members of the microbiota, and monitor the overall change of the entire communities. Generally, the current methods used to determine the bacterial diversity can be classified into two major categories: culture-dependent and culture-independent methods.

2.3.1 Culture-dependent Methods

Culture-dependent methods rely on the isolation and the cultivation of individual microorganisms using synthetic media which provide the required nutrients to permit growth. Briefly, bacterial colonies are isolated using standard agar plating procedures. The bacterial colony-forming units (CFU) can be calculated by counting the number of colonies on the agar plates with the assumption that each colony was formed by an individual cell (Silvio *et al.*, 2012). The distinct individual colonies can be then isolated using a sterilized loop and cultivated in medium broth prior to analysis.

The major problem of isolating bacteria in this way is the bias of cultivation. The use of artificial media containing selective agents may inhibit the growth of various bacterial species even those of interest (Head *et al.*, 1998). Previously, the majority of the bacterial species in the GI were considered as non-cultivable due to the difficulty in growing them under standard laboratory

conditions (Cuppels *et al.,* 1973; Stackebrandt *et al.*, 1995). Recently, Goodman and colleagues (2011) demonstrated that the most of the species present in the human GI tract can be recovered after extensive dilution and long term cultivation in multiple media. However, this method is not practical for most studies due to the large scale cultivation that is extremely time consuming and expensive (Goodman *et al.*, 2011).

Investigators began combat the inability to cultivate a large portion of the GI bacterial community with the use of selective media. Biochemical selective agents such as bile, esculin or antibiotics are used for selective enrichment (O'Sullivan *et al.,* 1999) and to encourage growth of only desired taxa. However, challenges still exist because many taxa do not have a selective media with proper specificity and are not cultivable. Therefore, if taxa-specific selective media is not available, it may be difficult to isolate these taxa from more abundant cultivable microbiota.

Another drawback associated with cultivation is that some bacterial species present in low relative abundance are easily outnumbered by other species (Hugenholtz *et al.*, 1998). Usually, investigators are interested in specific organisms with a comparatively low abundance but high economic or health benefits. However, recovering these species requires a large sample size to increase the likelihood for detection. For example, to obtain 100 lactobacilli isolates from feces, more than 10,000 bacterial colonies need to be analyzed, since lactobacilli is usually less than 1% of the total population in microbiota.

After cultivation, the individual bacterial isolates can be identified by their phenotypic or genotypic features. Phenotypic classification can be done using various methods such as serotyping, carbohydrate utilization, etc. Alternatively, DNA extracted from the bacterial isolate can be sequenced to determine its taxonomic identity. Recently, identification of bacterial species by molecular biological methods, especially 16S rDNA sequencing, has become the gold standard.

Methods that have been widely applied in the identification of GI bacteria are listed in **Table 2.1**.

Table 2.1 List of methods for identifying or comparing bacterial isolates after selective cultivation

Methods	Description	Reference
ARDRA: Amplified Ribosomal DNA Restriction Analysis	ARDRA involves a PCR amplification and restriction digestion of 16S rDNA of the species. The pattern obtained is representative of the species and can be used to phylogenetically characterize cultured isolates.	Girafa *et al.,* 1998
MLST: Multilocus Sequence Typing	MLST is a technique for the typing of multiple loci. The procedure characterizes isolates of bacterial species using the DNA sequences of internal fragments of multiple housekeeping genes.	de las Rivas *et al.,* 2006 Calmin *et al.,* 2008 Picozzi *et al.,* 2010
PFGE: Pulse Field Gel Electrophoresis	PFGE is one technique used for mapping of genomes of whole organism, which allows the electrophorestic separation of DNA molecules several megabases in size.	Jacobsen *et al.,* 1999
T-RFLP: Terminal Restriction Fragment Length Polymorphism	TRFLP is a molecular biology technique for profiling of microbial communities based on the position of a restriction site closest to a labeled end of an amplified gene.	Christensen *et al.,* 2004
RAPD-PCR: Random Amplification of Polymorphic DNA-PCR	RAPD-PCR is a type of PCR reaction that the segments of DNA are amplified randomly. By resolving the resulting patterns, a semi-unique profile can be gleaned from a RAPD reaction.	Hayford *et al.,* 1999 Nigatu *et al.,* 2001 Roy *et al.,* 2000
ERIC-PCR/rep-PCR/BOX-PCR	Bacterial fingerprint techniques based on the different number of repetitive sequence elements in bacterial genomes	Gevers *et al.,* 2001 Ventura *et al.,* 2002
SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis	A molecular biological method which can separate proteins according to their electrophoretic mobility. SDS gel electrophoresis of samples having identical charge per unit mass due to binding of SDS results in fractionation by size.	Pot *et al.,* 1993
qPCR: Quantitative PCR	qPCR is a technique used to amplify and simultaneously quantify a targeted DNA molecule. It enables both detection and quantification of one or more specific sequences in a DNA sample.	Ladero *et al.,* 2010
DNA Microarray	A DNA microarray is a multiplex technology which consists of an arrayed series of thousands of probes to determine relative abundance of nucleic acid sequences in the target.	Preza *et al.,* 2009 Weckx *et al.,* 2010 Wang *et al.,* 2004

2.3.2 Culture-independent Methods

Culture-independent methods were developed to address the weaknesses of culture-dependent methods. Several culture-independent methods include a polymerase chain reaction (PCR) step in order to amplify total DNA. Generally, the total DNA of bacterial community is directly extracted from the substrate without any cultivation steps. The generated amplicons from different species are then separated using gel or capillary separation or by hybridization to specific probes. The separated DNA segments can be sequenced to obtain nucleotide sequencing information and to identify bacterial species. A few culture-independent methods applied to studies examining the bacterial community of the human GI tract are listed in **Table 2.2**.

Culture-independent methods also provide numerous community level tools which are based on the direct analysis of DNA. As they are fast and potentially more exhaustive, these methods are well suited for analyzing microbial communities over time and may provide the possibility of exploring microbiota dynamics in detail. Specific features of culture-dependent and culture-independent methods are compared in **Table 2.3**.

In conclusion, the culture-independent methods provide greater sensitivity, convenience, and repeatability for studies analyzing the bacterial community associated with the human GI tract. These methods are particularly useful in monitoring the whole community dynamics, identifying bacterial species which are difficult to cultivate and species present in the GI tract at low relative abundance. The movement away from culture-dependent methods to culture-independent methods has resulted in rapid growth in the field of GI microbiota research.

Table 2.2 List of culture-independent methods for comparing bacterial communities or directly detecting species in complex samples

Methods	Communities Comparison	Species Detection
ARDRA		Guan *et al.,* 2003 Markiewicz *et al.,* 2010
RFLP/AFLP		Blaiotta *et al*: 2008 Claisse *et al.,* 2007
DGGE/TGGE	Jiang *et al.,* 2010 Fu-gui *et al.,* 2007 Petersson *et al.,* 2009	Guan *et al.,* 2003 Endo *et al.,* 2009 Fu-gui *et al.,* 2007 Petersson *et al.,* 2009
PFGE		Tynkkynen *et al.,* 1999 Markiewicz *et al.,* 2010
SSCP	Samelis *et al.,* 2011 Peu *et al.,* 2006 Ott *et al.,* 2004	Samelis *et al.,* 2011 Ott *et al.,* 2004
T-RFLP	Nieminen *et al.,* 2011 Christensen *et al.,* 2004 Davis *et al.,* 2010 Coolen *et al.,* 2005	Christensen *et al.,* 2004 Davis *et al.,* 2010 Coolen *et al.,* 2005
RAPD-PCR		Song *et al.,* 2000 Tynkkynen *et al.,* 1999 Markiewicz *et al.,* 2010
ERIC-PCR/rep-PCR/BOX-PCR		Markiewicz *et al.,* 2010
SDS-PAGE	Ying *et al.,* 2004	Teanpaisan *et al.,* 2006
Pyrosequencing	Dominguez-Bello *et al.,* 2010	Tarnberg *et al.,* 2002 Dominguez-Bello *et al.,* 2010 Roesch *et al.,* 2009
FISH	Skowronska *et al.,* 2009 Mare *et al.,* 2006 Harmsen *et al.,* 1999 Collado *et al.,* 2007	Skowronska *et al.,* 2009 Mare *et al.,* 2006 Harmsen *et al.,* 1999
qPCR	Ladero *et al.,* 2010 Fujimoto *et al.,* 2008 Lubbs *et al.,* 2009	Ladero *et al.,* 2010 Lubbs *et al.,* 2009
DNA Microarray	Weber *et al.,* 2008	Weber *et al.,* 2008 Preza *et al.,* 2009 Weckx *et al.,* 2010 Wang *et al.,* 2004

Table 2.3 Comparison of Features of Culture-dependent and Culture-independent Methods

	Culture dependent	Culture independent
Major sources of bias	Bias of selective enrichment media	PCR primer bias, probes bias
Specificity	Low, no perfect selective media	High, primers or probes can be designed to target specific species
Detected cell types	Only living bacterial cells	Both living and dead cells
Able to identify phenotypes?	Yes	No
Level of taxonomy	Individual isolates	Whole community, groups of species, individual isolates
Sensitivity to low abundant species	Low	High
Time and labor consuming	High	Comparatively low
Repeatability	Low	Comparatively High
Popular tools	API50, Carbohydrate Utilization Test, Antibiotic Resistant Test	T-RFLP, DGGE, REP-PCR, FISH, Pyrosequencing

2.3.3 Novel Molecular Biologic Methods

Pyrosequencing is a new culture-independent method for high-throughput DNA sequencing, and is an alternative to the traditional Sanger sequencing method. The typical protocol of pyrosequencing involves four stages: [1] target DNA segments are amplified using PCR; [2] double-stranded DNA is converted to single-stranded DNA templates; [3] oligonucleotide primers are hybridized to a complementary sequence of interest; and [4] the pyrosequencing reaction itself, in which a reaction mixture of enzymes and substrates catalyzes the synthesis of complementary nucleotides. Data are shown as a collection of signal peaks in a pyrogram and can be converted to multiple nucleotide sequences (Clarke *et al.*, 2005). A chart explaining the major steps in pyrosequencing is shown in **Fig. 2.2**.

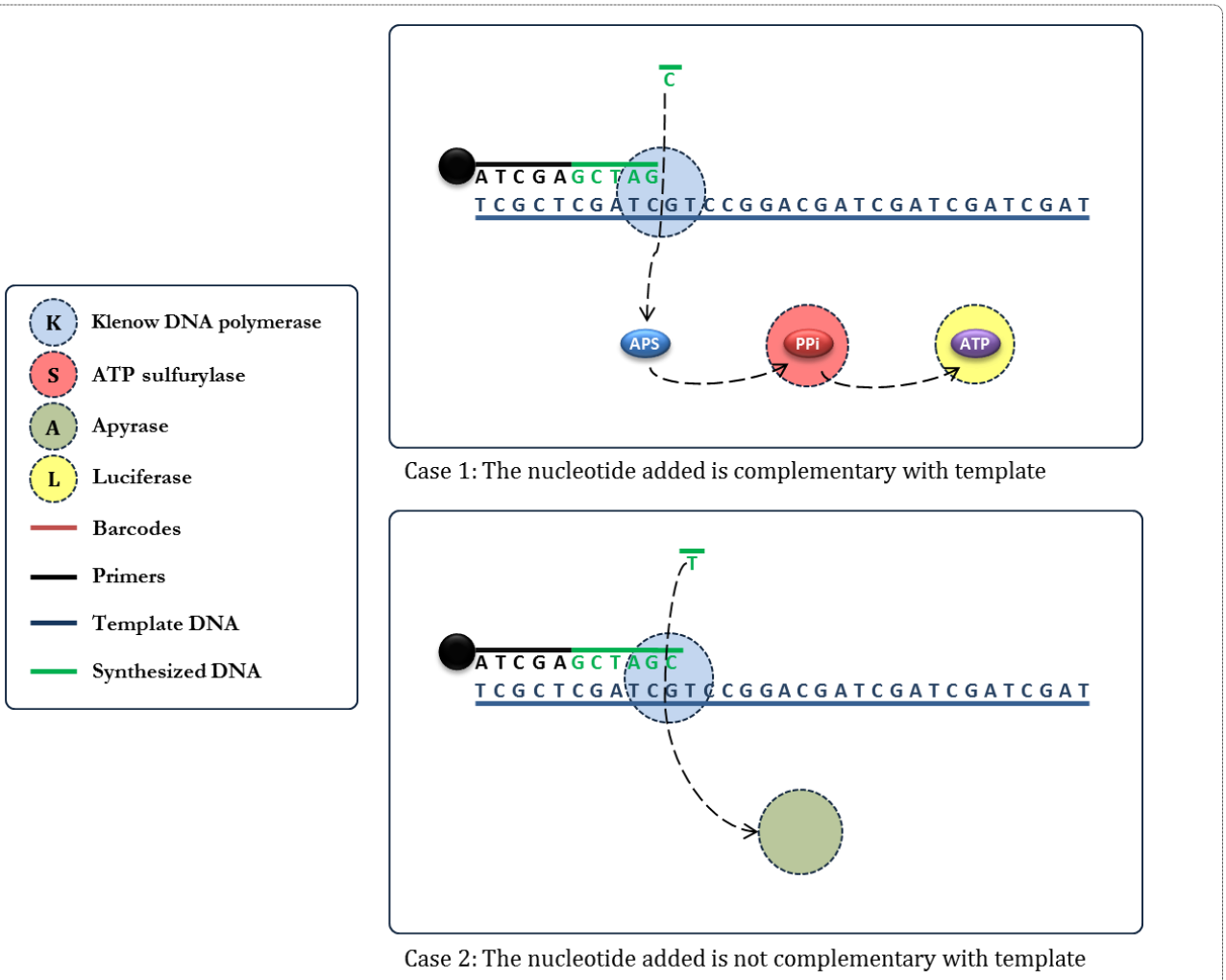


Fig. 2.2 The major steps in bacterial Tag-Encoded FLX Amplicon Pyrosequencing (bTEFAP) analysis. In case 1, APS released when complementary nucleotides added in to the system. It is converted to ATP and utilized by luciferase; the light signal emited was recorded by a camera. In case 2, mismatched nucleotides were taken away from the system by apyrase.

The extremely high-throughput nature and comparatively low cost are the most significant benefits of various pyrosequencing technologies. Pyrosequencing is able to generate millions of readable sequences (~400 bp each) from a single reaction with the average cost of less than one cent per readable sequence. This allows the researchers to rapidly conduct a comprehensive census of whole microbial communities, offering an opportunity to perform studies which were unfeasible

in the past. Pyrosequencing operation has recently become fully automated, thus enabling the high-throughput analysis of samples (Clarke, 2005). Consequently, it is widely used in the detection, identification of bacterial strains, and comparison of complex bacterial communities.

In addition, some new protocols have made pyrosequencing even more inexpensive and efficient. Bacterial Tag-Encoded FLX Amplicon Pyrosequencing (bTEFAP) is a pyrosequencing protocol that makes use of a novel tag priming method and an efficient bioinformatics pipeline. Reference genes used in identifying bacteria were amplified in a PCR reaction using a uniquely designed primer which has an oligonucleotide tag attached at the 5' end. After pyrosequencing, samples can be retrieved by comparing the tags at the beginning of the sequences; in other words the "name" of each sequence is identified. This innovation allows analysis of multiple samples in just one reaction (multiplexing) (Hamady *et al.*, 2008) and automatic computer analysis (Dowd *et al.*, 2008). This unique step makes bTEFAP ultra-efficient in terms of both lab operation and data processing. Recently, bTEFAP has been used as a novel tool in many microbial ecology studies including earth, fecal materials and clinic samples (Jones *et al.*, 2009; Costello *et al.*, 2009; Callaway *et al.*, 2009).

2.4 Methods for Isolating and Identifying *Lactobacillus* spp. in GI Tract.

2.4.1 Methods Based On Selective Media and Phenotypic Tests

During the last several decades culture-dependent methods were the gold standards for the isolation, detection, identification and comparison of *Lactobacillus* spp. Synthetic media with selective agents were used for selective enrichment of bacterial species. *Lactobacillus* Selection (LBS) agar and Rogosa (Rogosa SL) agar are the two most commonly used media for selective cultivation of lactobacilli, though they have significant differences in the degree of selectivity. These two media do not exclusively select for lactobacilli from complex microbial communities.

Examination reveals that Rogosa is more selective for Gram positive rods than LBS. Only 55% of the colonies on LBS were Gram positive rods, but 80% of the colonies on Rogosa agar were Gram positive rods. Of the identified colonies, 31% from LBS agar were lactobacilli compared to 24% from Rogosa agar (Nelson *et al.*, 1995). It appears that LBS agar is more suitable than Rogosa agar in isolating lactobacilli from complex samples since it has less false positives. In fact, previous research in the Miller lab demonstrated that the lactobacilli species detected using culture-independent methods match the species identified from LBS isolates. A strong correlation was seen between the numbers of lactobacilli (CFU/g) cultivated on LBS agar and lactobacilli-specific qPCR. It was concluded that LBS provides acceptable selectivity with only a few non-lactobacilli recovered from the LBS plates including the closely related *Weissella* spp. and *Pediococcus* spp.

Identification of isolated *Lactobacillus* spp. by their phenotype is notoriously difficult due to the ability to transfer genes by plasmids. Protocols of identification would require several bio-molecular assays. In order to make this procedure easier and more comprehensive, various schemes have been proposed based on distinguishable phenotypic, biochemical and physiological characteristics. The API 50 test, which is a widely accepted scheme to identify *Lactobacillus* spp. and extensively used over the past decades (Chou *et al.,* 2004), has a combination of bio-chemical and fermentation tests that are used to identify microorganisms. However, the repeatability of this test is still poor. In fact, a previous studies in the lab (Francl *et al.*, 2008), three *Lactobacillus gasseri* strains which were identified using 16S rDNA sequencing, showed different phenotypic profiles in the API 50 test and would be mis-classified as other species based on the API 50 test results. Therefore, identification by phenotype in a single assay is insufficient for inter- and intra-species differentiation and need to be supplemented with more sensitive molecular methods to obtain more repeatable results.

2.4.2 Culture-independent Methods

More recently, various culture-independent methods based on 16S rDNA PCR and sequencing have been developed which were able to more rapidly characterize the lactobacilli community (Jiang *et al.,* 2010; Samelis *et al.*, 2011; Coolen *et al.*, 2005) and detect specific *Lactobacillus* spp. from diverse sources (Nieminen *et al.*, 2011; Christensen *et al.*, 2004; Davis *et al.*, 2010). Since these molecular methods easily bypass the challenges and inconsistencies of selective cultivation and phenotype based identification, they have become increasingly popular. Currently, the most well established, culture-independent methods used for lactobacilli analysis include T-RFLP, DGGE, FISH and qPCR.

2.4.2.1 Terminal Restriction Fragment Length Polymorphism (T-RFLP)

Terminal restriction fragment length polymorphism (T-RFLP) is a popular molecular approach which can determine taxa dominance and richness within samples (Engelbrektson *et al.,* 2006; Kitts *et al.,* 2001). Briefly, T-RFLP uses a set of fluorescently labeled primers in a PCR reaction to amplify partial 16S rDNA sequences from bacterial community DNA. The mixed amplicons are then digested with restriction enzymes to create a mixture of fluorescence-tagged terminal restriction fragments (TRFs), the exact sizes of each TRF is determined by standard capillary electrophoresis and the retrieved data of fragment sizes can be used to in community level analysis methods such as agglomerative hierarchal clustering (AHC) and principle component analysis (PCA). T-RFLP based on *Lactobacillus* 16S rDNA polymorphism has been used in *Lactobacillus* community research for many years (Nieminen *et al.*, 2011; Christensen *et al.*, 2004; Davis *et al.*, 2010; Coolen *et al.*, 2005). In some of these studies, group or species-specific primers and enzymes were applied for detecting certain *Lactobacillus* spp. by seeking their unique TRF profiles (Christensen *et al.*, 2004; Davis *et al.*, 2010; Coolen *et al.*, 2005).

2.4.2.2 Denaturing Gradient Gel Electrophoresis (DGGE)

Denaturing Gradient Gel Electrophoresis (DGGE) is another widely used molecular method. DGGE uses a set of CG-clamped primers in a PCR reaction to amplify partial 16S rDNA sequences from bacterial community DNA. The mixed amplicons are separated based on the decreased electrophoretic mobility of a partially melted double-stranded DNA (dsDNA) molecule using polyacrylamide gels containing a linear gradient of DNA denaturants (a mixture of urea and formamide). Once a dsDNA molecule reaches its lowest melting concentration at a particular position in the denaturing gradient gel, a transition of a helical to a partially melted molecule occurs, and migration of the molecule will practically halt. Sequence variation within different molecules causes the melting position to differ, and molecules with different sequences will stop migrating at different positions in the gel (Muyzer *et al.*, 1997).

DGGE protocols using *Lactobacillus* species-specific primers were developed and applied in GI microbiota research and has been shown to be a useful tool in monitoring lactobacilli community change (Endo *et al.,* 2009). DGGE also has the potential to separate DNA fragments from different members of the microbiota, and upon sequencing can provide the species identification (Guan *et al.*, 2003; Endo *et al.*, 2009; Fu-gui *et al.*, 2007; Petersson *et al.*, 2009). In contrast T-RFLP does not provide sequence information due to the destructive restriction enzyme digestion of the PCR amplicons. However, DGGE is less reliable and less repeatable than T-RFLP, this is due to two reasons:

1. The operation protocol of DGGE involves more sensitive steps and complex equipment than T-RFLP, which could introduce variation. In previous research, obvious variation occurred when samples were separated with duplicate gels at the same time. Unexpected disturbing factors (temperature change, voltage instability, shaking of electrophoresis device) could impact the final result, making it harder to compare two individual gels without computational calibration.

2. Capillary electrophoresis in TRF detection has a higher resolution than acrylamide gel imaging devices. Capillary electrophoresis used in T-RFLP is used in DNA sequencing equipment which can distinguish fragments as small as one base pair difference (Blaiotta *et al.*, 2008). While DGGE uses regular gel imaging device which has a much bigger variation in imaging (Brody *et al.* 2004).

Additionally, DGGE requires expensive, specially-designed gel casting tools and electrophoresis tanks, as well as well-trained personnel. All of which could greatly increase the cost of research projects that are studying bacterial composition.

2.4.2.3 Fluorescent *in situ* Hybridization (FISH)

Fluorescent *in situ* Hybridization (FISH) is a probe-based method that avoids potential PCR-bias, while providing detailed information about the spatial distribution of gut microorganisms. FISH uses a fluorescently-tagged probe that has a single stranded DNA oligonucleotide complementary to desired DNA fragments located on bacterial chromosomes. The bacteria bound to the probes can be detected using fluorescent microscopy. Quantitative FISH analyzes images of fixed cells on glass slides under fluorescent-microscopy (Langendijk *et al,.* 1995). Another unique benefit of FISH is that it provides the information of microorganism distribution in that a discrete fluorescent signal is visible at the site of probe hybridization (Trask *et al.,* 1991).

FISH can be used to detect bacteria on different phylogenetic levels; i.e. phyla, genera and species depending on the probes chosen. Specific oligonucleotide probes and primers have been designed for many bacterial taxa that are known to be present in the human GI tract, including lactobacilli (Hensiek *et al.*, 1992) and bifidobacterium (Langendijk *et al.,* 1995).

The unique feature of FISH than other molecular tools is that there is no PCR step, which in theory can avoid the PCR-bias present in PCR-based culture-independent methods. However, FISH can be problematic when duplex of DNA/RNA occurs and when probes are mismatched with non-

target rRNA (Yilmaz *et al.,* 2008). Due to the highly accurate method of duplicating DNA segments by PCR, hybridization using amplicons from PCR reaction has high repeatability and stability. The lack of a PCR step in screening the targets which bind to probes can lead to inaccuracies and instabilities when using FISH. Actually, the mismatching of probes and target DNAs/RNAs in FISH is considered as systemic error and is complicated by multiple factors involving probe design (Ludwig *et al.*, 2004) and hybridization conditions (Pozhitkov *et al.*, 2006).

2.4.2.4 Quantitative Real-Time Polymerase Chain Reaction (qPCR)

Quantitative Real-Time Polymerase Chain Reaction (qPCR) is a molecular biological tool used to determine the quantity of specific bacterial taxa present in a bacterial community. A benefit of qPCR over other technologies in microbiota research is its ability to detect bacterial species of certain specificity and allow for their quantification at the same time. Briefly, genomic DNA is isolated from a complex bacterial community and amplified in real-time using primers designed for detection of the desired bacterial taxa. The quantity of a bacterial gene copy in the reaction system will be reflected proportionally by a fluorescence signal emitted and is calculated by comparison to a standard curve. qPCR primers with different specificities have been developed to target genus, species and even strain level in GI tract lactobacilli research (Ladero *et al.*, 2010; Lubbs *et al.,* 2009). qPCR using *Lactobacillus* species-specific primers can be used to quantify the *Lactobacillus* spp. present. Additionally, the percentage of lactobacilli detected from the bacterial microbiota samples can be calculated by comparing it to the total bacteria concentration.

The major limitation of qPCR using bacterial 16S rDNA is that bacterial cells usually have multiple copies of the 16S rDNA in their genome resulting in either under- or over-estimations in determing the true concentration of the target taxa. In fact, the number of 16S gene copies even varies within a single species which may require pre-calibration (Boggy *et al.*, 2010).

2.4.2.5 Conclusion of Culture-independent Methods

Culture-independent methods have greatly improved the efficiency of lactobacilli community research, but the methods outlined above still have the following challenges:

1. Many culture-independent methods are based on "community fingerprinting" which provides sequence information of individual isolates. "Sequences-based" methods are based on group specific PCR and are subject to PCR bias.
2. Culture-independent methods cannot accurately reflect the phenotype of bacteria. As seen in a previous Miller lab study (Francl *et al.*, 2008), strains with identical 16S rDNA sequence can have significant phenotypic differences.

To address the limitations of current molecular methods, more sophisticated methods have continued to be developed. In the past five years, pyrosequencing technology has emerged as a powerful tool in investigating bacterial communities. It was a breakthrough in GI microbiota study and has greatly improved our capability to analyze the structure of complex bacterial communities.

2.5 Factors Impacting Early Colonization of GI Microbiota

The neonatal period is crucial for GI microbiota colonization. The GI tract of a newborn is considered sterile; however, during the birthing process, the neonate is exposed to a wide array of bacteria from different sources (Grölund *et al.,* 1999). The development and function of the bacterial community is controlled and modulated by diverse interacting mechanisms, such as genetic endowment, intrinsic biological regulatory functions, environmental influences and the early infant diet (Long *et al.*, 1977). Due to the handicap of the tools used in the past two decades, the complex mechanism which impacts the early colonization of microbiota in the infant intestine is still poorly understood. Impacts of piglet age, route of delivery and nutrition on the early development of bacterial microbiota was studied using both traditional culture-dependent methods and latest developed bTEFAP technology in this thesis.

2.5.1 Vaginal Delivery versus Cesarean section

The initial colonization of the newborn intestine first happens during delivery when it comes in contact with vaginal and skin microbiota from the mother and the delivery environment. This process leads to an inoculation with diverse bacterial species (Mackie, *et al.,* 1999). Currently, many human babies are not exposed to vaginal microbes at birth because they are delivered by cesarean section in a sterilized environment (Dominguez-Bello *et al.,* 2010). Birth by cesarean section prevents exposure to many microorganisms and potentially influences the initial colonization of lactobacilli and other microorganisms in the gut.

The impact of the route of delivery on the short and long term development of the lactobacilli community associated with the GI is still controversial. Previous research demonstrated that cesarean section and vaginally delivered newborn infants have an obvious difference in the early colonization of the intestinal microbiota (Biasucci *et al.,* 2008; Mackie *et al.,* 1999; Penders *et al.,* 2006). However, other researchers have concluded that the bacterial species aquired at birth from the mother's vaginal fluid do not persist in the intestine of the infant in the long term (Matsumiya *et al.*, 2002).

Several investigators have hypothesized that the impact of route of delivery on intestinal microbiota diversity contributes to the functional benefits that exist from vaginal delivery as opposed to cesarean section. Some benefits include protection of infants against GI infection and reduction of issues related to dietary allergens (Matsumiya *et al.*, 2002; Orrhage *et al.*, 1999). Therefore, understanding the detailed process in microbial community development as related to the route of delivery is important in improving infant health outcomes.

2.5.2 Formula-Fed vs. Breast-fed

Human milk is considered the "gold standard" for infant nutrition. The development of the infant GI microbiota is different between breast-fed and formula-fed groups. After the first week of breast-

feeding, bifidobacteria reportedly become the dominant gastrointestinal bacteria (Harmsen *et al.*, 2000). However, formula-fed infants develop a more diverse microbiota which includes bifidobacteria, bacteroides, enterobacteria, enterococci, and clostridia (Herfel *et al.*, 2009; Stark *et al.*, 1982).

One of the biggest differences between human milk and formula milk is the oligosaccharide composition, specifically Human Milk Oligosaccharides (HMO), which constitute the third largest component in human milk after lactose and lipids (Tuohy *et al.*, 2010). HMO composition ranges in concentration from 5 to 10 g/L in mature mother's milk. HMOs potentially have some health benefits through interaction with GI microbiota (i.e. partial fermentation by GI microbiota, modification of the composition and activity of beneficial bacteria, etc.) (Tuohy *et al.*, 2010). Most of these functions are still being investigated. HMOs are virtually absent from bovine milk and most infant formula, which may account in part for the difference in GI microbiota reported between breast-fed and formula-fed infants (Kunz *et al.*, 2000; Newberg and Neubauer, 1995).

There are more than 200 different HMOs that have been discovered. The lack of HMO supplementation in infant formula is mainly due to the complexity of obtaining large quantities of these oligosaccharides. Therefore, prebiotic carbohydrates are increasingly being supplemented in infant formula which is intended to substitute for some of the functional properties of HMOs. Polydextrose (PDX) and galactooligosaccharides (GOS) are examples of these prebiotics used in current commercial infant formulas.

PDX, a randomly bonded glucose polymer with sorbitol terminal groups and citric acid attached by mono and diester bonds, was first developed as a bulking agent for foods. PDX has been proposed as a surrogate for HMOs since it contains numerous glycosidic bonds, most form a β (1–6) linkage, which limits digestion by endogenous mammalian enzymes. This limited digestion allows the complete PDX to reach the large intestine and stimulate fermentation by the commensal microbiota. PDX was designed as a selectively fermented ingredient that allows specific changes in

the GI microbiota, both in the composition and activity of the bacterial communities that confer health benefits upon the host and general well-being (Roberfroid *et al.*, 2007).

Galactooligosaccharides (GOS) are another prebiotic used commercially in foods. GOS has been used as a food ingredient in Japan and Europe for at least 30 years and their application is currently expanding rapidly (Macfarlane *et al.*, 2008). The probiotic mechanism of GOS is similar to PDX. The specificity of GOS substrates has been attributed to their selective fermentation in the intestinal tract by bifidobacteria and lactobacilli (Gibson *et al.*, 2004). The promoting effect of GOS on bifidobacteria has been demonstrated (Davis *et al.*, 2010), though the effect of GOS on the lactobacilli population is still unknown.

In order to include a similar range of molecular weights compared to most HMOs, a mixture of GOS and polydextrose PDX, and are widely use as HMO alternative in food and nutrition studies (Ashley *et al.*, 2012; Monaco *et al.*, 2011).

2.5.3 Neonatal Piglets as Animal Models and Impact of Weaning on Piglet GI Tract Microbiota

The neonatal piglet is a well-established model to study nutrient interaction of infant formula. The pig digestive tract is anatomically and functionally similar to that of the human, and is more physiologically relevant for prebiotic studies than a rodent model (Reeds and Odle, 1996). The neonatal piglet model has been used to determine the safety of single cell sources of long-chain polyunsaturated fatty acids (Mathews *et al.*, 2002; Huang *et al.*, 2002), and the effects of conjugated linoleic acids on lipid accretion and adipose tissue metabolism (Corl *et al.*, 2008).

Weaning is a stressful process for young piglets. During weaning, piglets experience abrupt separation from the sow, a change in their physical surroundings, the end of lactational immunity and exposure to a solid diet. As weaning progresses, the piglets become very susceptible to gut disorders, infections and diarrhea due to GI or respiratory diseases (Hopwood and Hampson, 2003). Weaning of piglets is associated with compromised integrity of the small intestinal mucosal surface

and an increased paracellular permeability (Kelly *et al.*, 1991; Kelly *et al.*, 1991; McCracken *et al.*, 1995; Spreeuwenberg *et al.*, 2001). Studies of the intestinal microbiota of piglets have demonstrated that weaning can cause substantial changes in the intestinal bacterial community (Franklin *et al.*, 2002; Inoue *et al.*, 2005; Konstantinov *et al.*, 2006; Castillo *et al.*, 2007), particularly, a dramatic decrease in the lactobacilli population in the ileum was observed in several studies (Su *et al.*, 2008; Franklin *et al.*, 2002; Konstantinov *et al.*, 2006; Pieper *et al.*, 2008).

2.6 Summary

In conclusion, the GI microbiota provides a wide range of essential physiological functions and is considered crucial to host health. There are various factors, including age, route of delivery and nutrition, can impact the colonization and development of GI microbiota, especially in the early age of the infants. Some unique ingredients in human milk, like HMOs, may alter the composition of bacterial community in different mechanisms, and in turn impact health status of hosts. However, study on structures of bacterial community is difficult due to its extreme complexity, as well as the inefficiency of the current approaches used for analysis.

Therefore, future research is necessary to extend our understanding of the interaction of factors and bacterial community composition. We need to focus on the the application of latest developed, high efficient technology like T-RFLP and 16S rDNA pyrosequencing on GI microbiota composition study. These new approaches are important methods to advance our knowledge regarding microbiota development, and the result will help fill the gaps currently in microbiota research.

CHAPTER 3. DEVELOPMENT OF A *LACTOBACILLUS* SPECIFIC T-RFLP METHOD TO DETERMINE LACTOBACILLI DIVERSITY IN COMPLEX SAMPLES

3.1 Abstract

Terminal restriction fragment length polymorphism (T-RFLP) analysis has been widely used for studying microbial communities. However, most T-RFLP assays use 16S rDNA as the target and are unable to accurately characterize a microbial subpopulation. In this study, we developed a novel T-RFLP protocol based on *Lactobacillus hsp60* to rapidly characterize and compare lactobacilli composition. The theoretical terminal restriction fragment (TRF) profiles were calculated from 769 *Lactobacillus hsp60* sequences from online databases. *In silico* digestion with restriction endonucleases *Alu*I and *Tac*I on *hsp60* amplicons generated 83 distinct TRF patterns, of which, 70 were species specific. To validate the assay, five previously sequenced lactobacilli were cultured independently, mixed at known concentrations and subjected to analysis by T-RFLP. All five strains generated the predicted TRFs and a qualitative consistent relationship was revealed. We performed the T-RFLP protocol on fecal samples from mice fed six different diet (n=4). Principal component analysis and agglomerative hierarchical clustering revealed that the lactobacilli community was strongly connected to dietary supplementation. Our study demonstrates the potential for using *Lactobacillus* specific T-RFLP to characterize lactobacilli communities in complex samples.

3.2 Introduction

Lactobacillus is one of the most important bacterial species in gut microbiota research. It is well established that lactobacilli play an important role in the functioning of the intestinal ecosystem (Fuller, 1989). The balance of intestinal lactobacilli closely relates to the health of the individual

(Sanders *et al.*, 2003). The composition of *Lactobacillus* spp. is greatly dependent on the diet, but currently this is poorly understood due to limitations in measuring the vast diversity of microorganisms and monitoring their changes.

To study the structure of GI bacterial community, standard plating of media agar with selective agents such as bile, esculin or antibiotics can be used for selective enrichment of certain species (Engelkirk *et al.*, 1992). Phenotypic tests such as serotyping and carbohydrate utilization can only be applied after cultivation. Traditionally, it was thought that approximately 20-40% of the microorganisms in the GI tract could be cultivated by using enrichment media (Dicksved *et al.*, 2007). Recently, Goodman *et al.* found that 56 ± 4% of species-level taxa, which were detected by 16S rDNA pyrosequencing, could be also identified by sequencing cultivable isolates (Goodman *et al.*, 2011). However, this method is extremely laborious and time consuming. Therefore, novel molecular methods have been developed over the last 15 years. Polymerase chain reaction (PCR) amplification of the specific 16S rDNA-based oligonucleotide probes has been widely used to detect different groups of bacteria directly in fecal or intestinal content samples (Singh *et al.,* 2009).

Terminal restriction fragment length polymorphism (T-RFLP) is an increasingly popular molecular approach which can determine taxa dominance and richness within samples (Engelbrektson *et al.*, 2006; Kitts *et al.*, 2001). T-RFLP allows the assessment of complex bacterial communities and rapid comparison of the community diversity between different ecosystems (Christensen *et al.*, 2004). The typical T-RFLP analysis is based on bacterial 16S rDNA specificity. Briefly, T-RFLP uses a set of fluorescently labeled primers in a PCR reaction to amplify partial 16S rDNA sequences from bacterial community DNA. The resulting mixed amplicons are then digested with restriction enzymes to create a mixture of fluorescence-tagged terminal restriction fragments (TRFs), which have specific sizes corresponding to specific microbes. The exact size of each TRF is determined by standard capillary electrophoresis and can be used in community level analysis methods such as agglomerative hierarchal clustering (AHC) and principle component analysis (PCA).

In the past 20 years, various protocols using partial 16S rDNA have been developed which were able to identify *Lactobacillus* spp. isolated from different environments (Nieminen *et al.*, 2011; Christensen *et al.*, 2004; Davis *et al.*, 2010; Collen *et al.*, 2005). However, 16S rDNA-based molecular analysis is not efficient enough to reveal significant differences between closely related species (Blaiotta *et al.*, 2008; Walter *et al.*, 2000). Additionally, the rDNA copy number in lactobacilli varies from species to species (Fogel *et al.*, 1999), which makes it difficult to quantify specific amplicons in digested products without a prior calibration. Hsp60 is a highly conserved heat shock protein found in all prokaryotic and eukaryotic cells (Kwok *et al.*, 1999). Polymorphisms within *hsp60* among different *Lactobacillus* spp. can be used to discriminate recently diverged species (Blaiotta *et al.*, 2008; Dellaglio *et al.*, 2005). Additionally, the higher variability of the *hsp60* nucleotide sequences compared to 16S rDNA offers greater opportunities to generate distinguishable TRFs which can be used to detect closely related *Lactobacillus* spp. at the strain level. Moreover, *hsp60* presents a single gene copy in each individual bacterial cell (Goh *et al.*, 1996; Kwok *et al.*, 1999), which makes it possible to develop quantitative or semi-quantitative assays.

The primary aim of this project was to develop a T-RFLP protocol using the *Lactobacillus hsp60* gene as a template. We optimized conditions for PCR and restriction enzyme digestion in order to minimize sampling, extraction and template biases. Furthermore, we used this method to evaluate the potential of dietary factors in modulating the gut lactobacilli community.

3.3 Methods and Materials

3.3.1 TRF Pattern Prediction in Silico Analysis

A set of 769 *Lactobacillus hsp60* sequences were obtained from NCBI gene and genome project databases for analysis. All sequences were imported and analyzed using the sequence analysis

software CodonCode Aligner (CodonCode Corporation, Dedham, MA). Alignment was performed using the "sequence assembly" tool. The theoretical PCR amplicons were extracted by deleting all the sequences flanking the binding sites of the two primers. The *hsp60* restriction profiles by enzymes *Alu*I and *Tac*I of each strain were obtained by using the "restriction mapping" tool. Cleavage site positions were recorded and imported into Microsoft Excel (Microsoft, Seattle, WA) for TRF size calculation.

3.3.2 Development of T-RFLP Protocol Based on *Lactobacillus hsp60*

Nested PCR was used to generate *Lactobacillus hsp60* segments for T-RFLP. In the first PCR reaction, universal bacterial *hsp60* primers H279 and H280 (**Table 3.1**) were used to amplify a 650 bp fragment internal to the *hsp60* gene (Goh *et al.*, 1996).

PCR amplification was performed with a 50 μL total volume. The reaction system includes 5 μL of target DNA used directly from the bacterial community or pure culture DNA extraction, 5 μL of Taq DNA polymerase 10X Buffer (Lucigen Corporation, Middleton, WI), 2.5 μL of 50 mM $MgCl_2$, 0.5 μL of a dNTP mix (25 mM each), 0.125 μL of each primer (0.1 mM), and 0.5 μL of Taq DNA polymerase. The PCR consisted of 40 cycles (30s at 94°C, 30s at 37°C, and 1 min at 72°C) and one final cycle at 72°C for 5 min. Two replicate PCRs in 50 μL reaction mixtures were performed for each sample and the PCR products of the replicates were pooled. PCR products were purified using Zymoclean DNA Clean & Concentrator Kit (Zymo Research Corporation, Irvine, CA) according to the manufacturer's instructions. The purified DNA was diluted 10-fold and used as the templates for a second round of PCR.

Table 3.1 The primers used in this study

Primer	Sequence (5'-3')	Reference
H279[1]	GAATTCGAIIIIGCIGGIGA(TC)GGIACIACIAC	Goh *et al.*, 1996
H280[2]	CGCGGGATCC(TC)(TG)I(TC)(TG)ITCICC(AG)AA ICCIGGIGC(TC)TT	Goh *et al.*, 1996
LB308F[1]	TGAAGAAYGTNRYNGCYGG	Blaiotta *et al.*, 2008
LB806RM[2]	AANGTNCCVCGVATCTTGTT	Blaiotta *et al.*, 2008
LB308F-D[1,3]	/56-FAM/TGAAGAAYGTNRYNGCYGG	This study
LB806RM-D[2,3]	/5HEX/AANGTNCCVCGVATCTTGTT	This study

[1] Forward primers
[2] Reverse primers
[3] "D" stands for the flurencent dyes attached to the regular primers

In the second round of PCR, lactobacilli specific *hsp60* primers, LB308F (forward, **Table 3.1**) and LB806RM (reverse), previously described by Blaiotta *et al.* (Blaiotta *et al.*, 2008), were used to amplify a 499 bp fragment internal to the first round PCR product. In order to make the TRFs visually detectable in the later steps, two fluorescent dyes, 6-FAM (6-carboxyfluorescein) and HEX (6-carboxy-2',4,4',5',7,7'-hexachlorofluorescein) were separately attached to the forward and reverse primers (LB308F-D and LB806RM-D respectively) before the PCR reaction (IDT, Coralville, IA). The reaction condition of PCR remained the same as described above.

The PCR products were subjected to electrophoresis at 100 V in a 1% (w/v) agarose gel containing 1 mg mL^{-1} ethidium bromide prior to enzyme digestion. Bands of approximately 500 bp were visualized by ultraviolet (UV) transillumination, cut from the gel, and then recovered using a Zymoclean Gel DNA Recovery Kit (Zymo Research Corporation, Irvine, CA).

Restriction enzyme digests were performed by digesting 200 ng of a gel-purified PCR product with 20 U of either restriction enzyme *Alu*I (New England Biolabs, Ipswich, MA) at 37°C for 2 h or

*Tac*I (Promega Corporation, Madison, WI) at 65°C for 1.5 h. Following the restriction digestion, the DNA was analyzed with an ABI 3730xl capillary system (Applied Biosystems Inc, Foster City, CA) according to the manufacturer's instructions.

3.3.3 Application of T-RFLP with Mixed Culture

To validate the T-RFLP protocol and determine detection limits, a mixed *Lactobacillus* spp. culture was made with five *Lactobacillus spp.*: *L. acidophilus* (NCFM), *L. jesenii* (ATCC 25258), *L. sakei* (ATCC 15521), *L. rhamnosus* (ATCC 53103), *L. plantarum* (ATCC 11146). These five strains have distinct TRF patterns (Table S1) and were selected as reference strains for this study. Working cultures were prepared in 10 mL of MRS broth (BD, Franklin Lakes, NJ) and incubated for 24 h at 37°C. 1 mL of the cultures was serially diluted and then plated on LBS agar (BD, Franklin Lakes, NJ) plates. The plates were cultivated in an anaerobic chamber at 37°C for 24 h and then enumerated. After plating, the remaining culture was stored in a -80˚C freezer immediately to end bacterial growth.

After colony forming units (CFU) numbers were calculated, the frozen media was thawed on ice, and then homogenized using a vortex. The lactobacilli cells were mixed at a ratio of 10^4: 10^3: 10^2: 10: 1, respectively. The final concentrations of cells in the mixed culture depicted in **Fig. 3.1** were: *L. rhamnosus*: 72,000 cells mL^{-1}, *L. plantarum*: 7,200 cells mL^{-1}, *L. acidophilus*: 720 cells mL^{-1}, *L. sakei*: 72 cells mL^{-1}, *L. jesenii*: 7 cells mL^{-1}. Additional trials were performed so that each species was represented at each concentration.

The total DNA of the mixed bacterial culture was extracted and purified using the commercial QIAamp DNA Stool Mini Kit (QIAGEN, Valencia, CA), following the manufacturers protocol. The DNA was finally eluted in 200 µL of Qiagen Buffer AE, and T-RFLP was applied using the protocol described above.

3.3.4 Application of T-RFLP on Lactobacilli Community Research

C57BL/6J male mice (3 weeks old) were purchased from Jackson Laboratories (Bar Harbour, ME). Mice were housed (8 per cage) in large standard shoebox cages (length 28 cm; width 17 cm; height 12.5 cm) and allowed free access to food and water. Mice were fed open source uniform-base diet purchased from Research Diets (New Brunswick, NJ). These diets varied based on amount and type of fiber as well as fat content (**Table 3.2**). Housing temperature (22°C) and humidity (45-55%) were controlled as was a 12/12 h reversed dark-light cycle (2200-1000 h). Animal use was conducted in accordance with the Institutional Animal Care and Use Committee (IACUC) approved protocols at the University of Illinois Urbana-Champaign.

Table 3.2 Animal grouping and diet information

Group	Fiber (%)[1]	Fat (%)[2]	Age (months)[3]	Diet NO.[4]
1	cellulose (10)	60	8	D07102501
2	pectin (10)	10	8	D06082202
3	pectin (10)	60	8	D08111803
4	cellulose (5)	60	8	D12492
5	cellulose (5)	10	8	D12450B
6	cellulose (10)	10	8	D06082201
7	cellulose (10)	10	3	D06082201
8	pectin (10)	10	3	D06082202

[1] Percent fiber on a weight basis
[2] Percent fat on caloric basis
[3] The animals were fed the indicated diet from weaning (21 days) to the age indicated.
[4] Product number for each diet as assigned by Research Diets

Fecal samples were collected from each mouse at the end of the treatment phase (either 3 month or 8 month). Approximately 0.5 g of fresh fecal pellets from every mouse were collected into a 1.5 mL centrifuge tube using sterile forceps, and the fecal DNA was extracted and purified using the

commercial QIAamp DNA Stool Mini Kit as described above. T-RFLP was then applied to the extracted DNA.

To specifically analyze the dominant cultivable *Lactobacillus* population in subjects from group 7 and 8, an extra 0.1 g of fresh fecal sample from each mouse in these two groups was collected. The fecal pellets were vortex-homogenized in 10 mL saline-peptone buffer for 1 hour after collection; serially diluted and then plated on LBS agar (BD, Franklin Lakes, NJ) plates. The plates were cultivated in an anaerobic chamber at 37°C for 24 h. CFU/g were calculated and twelve colonies were randomly selected from each sample and subcultured in MRS broth for 24 h. Ten colonies are believed to provide adequate representation of the major bacterial strains which were cultivated with a selective medium (Hartley *et al.*, 1977).

Genomic DNA was extracted from the *Lactobacillus* isolates using the method described by Korhonen *et al.* with slight modifications (Korhonen *et al.*, 2007). Briefly, bacterial cultures grown overnight in 3 mL MRS broth were centrifuged (10,000 x g, 2 min) and the cells were re-suspended and washed twice in 1.5 mL molecular biology grade water (MO BIO Laboratories, Inc., Carlsbad, CA). The supernatant was discarded and the pellet was re-suspended in 750 µL of TE buffer (10mM Tris-HCl-1 mM EDTA [pH 8.0]). The bacterial cells were lysed by shaking for 1 min on a mini-bead beater (Bio-Rad Laboratories, Hercules, CA) at a speed of 6m s^{-1} in 2 mL screw-cap tubes, to which 350 to 400 mg of 100-µm-diameter Silica beads (Research Products International Corp., Mt. Prospect, IL) had been added. After centrifugation at 12,000 x g for 2 min, 500 µL of supernatant was transferred to a 1.5 mL centrifuge tube and stored at -20°C.

In order to identify the *Lactobacillus* isolates, primers LB308F and LB806RM lacking florescent tags were used to generate 0.5 kb unlabeled amplicons, which were cleaned using the Zymoclean DNA Clean & Concentrator Kit (Zymo Research Corporation, Irvine, CA) and sequenced on ABI 3730xl capillary systems by either primer LB308F or LB806RM. A FASTA file containing all of the sequences was constructed and uploaded into the BLAST search engine to compare records in the

NCBI nucleotide collection database. Based on previous studies with *Bifidobacterium* (Zhu *et al.*, 2003), *Vibrio* (Tarr *et al.*, 2007), and *Other* (Sakamoto *et al.*, 2011), records with maximal similarity values greater than 97% were used to identify species in this study. The sequences of recovered in this study were deposited with GenBank under accession numbers JN998111- JN998113.

3.3.5 T-RFLP Dataset Analysis

All digitized TRF datasets were retrieved and analyzed by the Applied Biosystems Peak Scanner (Applied Biosystems, Foster City, CA) and Microsoft Excel. The word "dataset" is used to define a particular microbial community. Each dataset consisted of 4 data matrices which were derived from the same template DNA and reflected the same community, named *Alu*I-forward, *Alu*I -reverse, *Tac*I-forward, *Tac*I -reverse.

In Peak Scanner, the analysis method for the T-RFLP matrices was set to "PP" (factory default sizing algorithm for samples containing primer peaks) and size standard GS500 was used. A baseline threshold of 25 fluorescence units was used to eliminate background noise with all datasets.

Those TRFs which differed by less than 0.5 bp in different profiles were considered identical and were clustered (Korhonen *et al.*, 2007). The relative peak area was calculated by dividing each raw peak area by the cumulative peak area of the sample and was shown as a percentage value. The reproducibility of the procedure was confirmed by repeating the entire PCR, digestion and data manipulation procedure three times. While absolute peak areas varied between trials, the relative peak area varied less than 5% for each TRF from each sample across the three replicates. Incomplete digests for TRFLP profiles were also resolved by comparisons between replicate digests. (Mills *et al.*, 2003).

The relative T-RFLP matrices were entered into an Excel spreadsheet that consisted of the TRFs as variables and individuals as objects. A binary data table (presence or absence of individual peaks)

was generated, in which peaks with percentage of total area >1% were assigned as 1 (presence) and <1% were assigned as 0 (absence) at every possible TRF size in all 4 matrices. The binary table was imported into the statistics software XLStat (Addinsoft Inc., Brooklyn, NY) to generate principal component analysis (PCA) plots and an agglomerative hierarchical clustering (AHC) dendrogram.

The identification of the *Lactobacillus* spp. was done by matching the TRFs in the generated database. As described by other authors, we noted some minor difference between the predicted and observed TRF lengths (Kaplan *et al.*, 2003). So, a variation tolerance of -1 to 10 bp was applied in this study. In addition, an online tool using Google Spreadsheet (Google Inc., Mountain View, CA) was developed to rapidly conduct the pattern comparison; with all of the *Lactobacillus* spp. matching the known TRF patterns were highlighted.

3.4 Results

3.4.1 *in silico* Evaluation of the *Lactobacillus hsp60* T-RFLP Method

A total of 769 *Lactobacillus hsp60* partial sequences were obtained from NCBI databases for *in silico* analysis. Computer simulated *Lactobacillus hsp60* T-RFLP using restriction endonucleases *Alu*I and *Tac*I on fluorescently labeled PCR products was able to generate 83 distinct TRF patterns, of which, 70 were specific for one species (**Table S1**). In addition, 23 *Lactobacillus* spp. had more than one corresponding pattern.

3.4.2 Validation of Method

In order to verify the T-RFLP protocol, five lactobacilli with distinct TRF patterns were selected and mixed as reference strains. The five strains were mixed so that their concentration was 7, 72, 720, 7,200 and 72,000 CFU mL^{-1}. All species generated clear and distinct peaks at the location

predicted *in silico* (**Fig. 3.1**). Only the peak generated by *L. jensenii* was close to background noise due to the low starting cell number, indicating the high sensitivity of this assay. There were -5 to 5 bp variations between the predicted and observed TRF lengths, this variation was previously observed and described by Kaplan *et al.* (Kaplan *et al.*, 2003). When the experiment was repeated with different ratios of these five lactobacilli, we found that the minimum concentration of cells in pure culture needed to readily generate distinguishable peaks was 1×10^{3} CFU mL^{-1}. T-RFLP demonstrated a qualitative relationship between the size of peaks and bacterial cell concentration. The species' with a larger relative abundance generated the larger peak while the species' with a lesser relative abundance generated smaller peaks.

3.4.3 Use of The T-RFLP Method for Analysis of Fecal Lactobacilli Community

T-RFLP was used to obtain fingerprints of the fecal lactobacilli communities from 4 mice in each of 8 treatment groups. Each fecal lactobacilli community was evaluated by a dataset consisting of 4 data matrices: *Alu*I-forward, *Alu*I -reverse, *Tac*I -forward and *Tac*I -reverse. A total of 53 TRF patterns were generated. Each mouse had a unique lactobacilli community profile, consisting of different mixtures of TRFs. Within each treatment group, the most dominant TRFs were observed in all subjects, and the intensity for each TRF was similar. However, between groups, there were differences in both intensity and sizes of the dominant TRFs present. Agglomerative hierarchical clustering (AHC) analysis was applied to visualize the *Lactobacillus* communities (**Fig. 3.2**). Samples from different treatment groups showed much greater differences than samples within the same group, which demonstrated that both dietary fiber supplements and the duration of time on the diet significantly impact the fecal lactobacilli community.

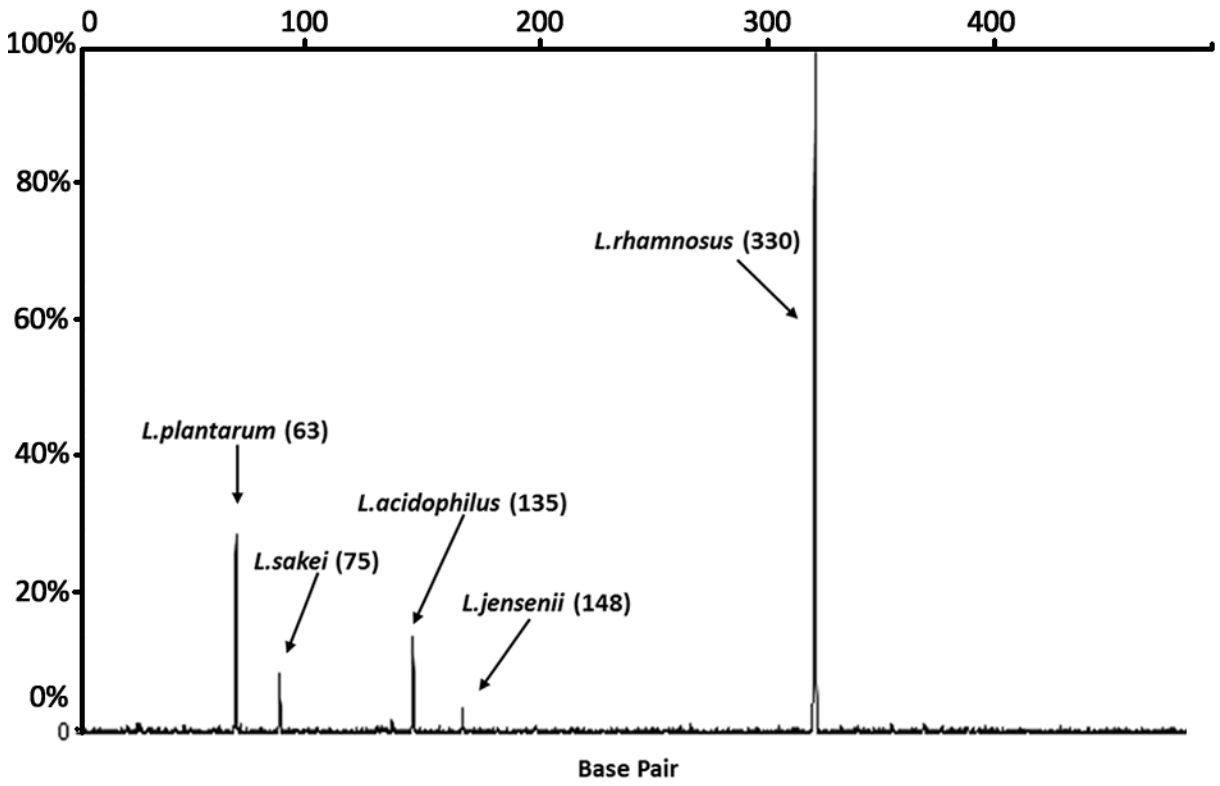


Fig. 3.1 TRF profile generated by mixed bacterial culture with *Alu*I and forward labeled *hsp60* primers. The X axis indicates the size of the TRFs, and the Y axis indicates the percentage of peaks height compared to the tallest peak. The original concentrations of the bacterial cells in 1 mL culture were: *L. rhamnosus*: 72,000 cells mL^{-1}, *L. plantarum*: 7,200 cells mL^{-1}, *L. acidophilus*: 720 cells mL^{-1}, *L. sakei*: 72 cells mL^{-1}, *L. jensenii*: 7 cells mL^{-1}. The TRF sizes (indicated in the brackets) identified corresponded with the prediction. In addition, TRF signal strength demonstrated a semi-quantitative relationship.

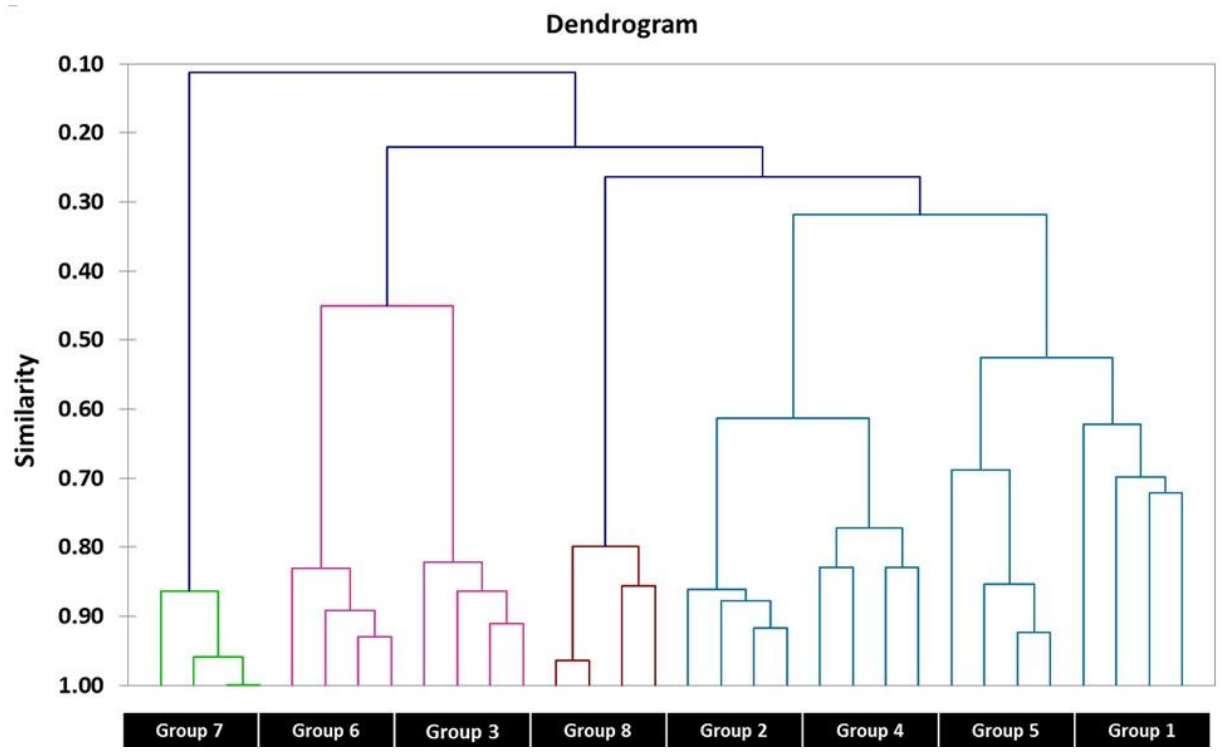


Fig. 3.2 AHC dendrogram of the T-RFLP data for the fecal lactobacilli community of 32 mice from 8 treatment groups (4 mice per group). These figures were constructed by the software XLStat (Addinsoft Inc.) based on a binary data table, which indicates presence or absence of individual peaks in all 4 matrices from the dataset of each individual samples. Samples from different groups showed much greater differences than samples within the same group.

3.4.4 Species Identification by Culture-dependent Method and T-RFLP

To compare the traditional culture-dependent method and T-RFLP in analyzing the *Lactobacillus* population, we compared the T-RFLP results from the fecal extraction (**Fig. 3.3**) with a culture-based approach. Twelve randomly selected isolates from the same feces were isolated from LBS plates for each of the four mice in group 7 and group 8. These isolates were speciated based on partial *hsp60* sequence.

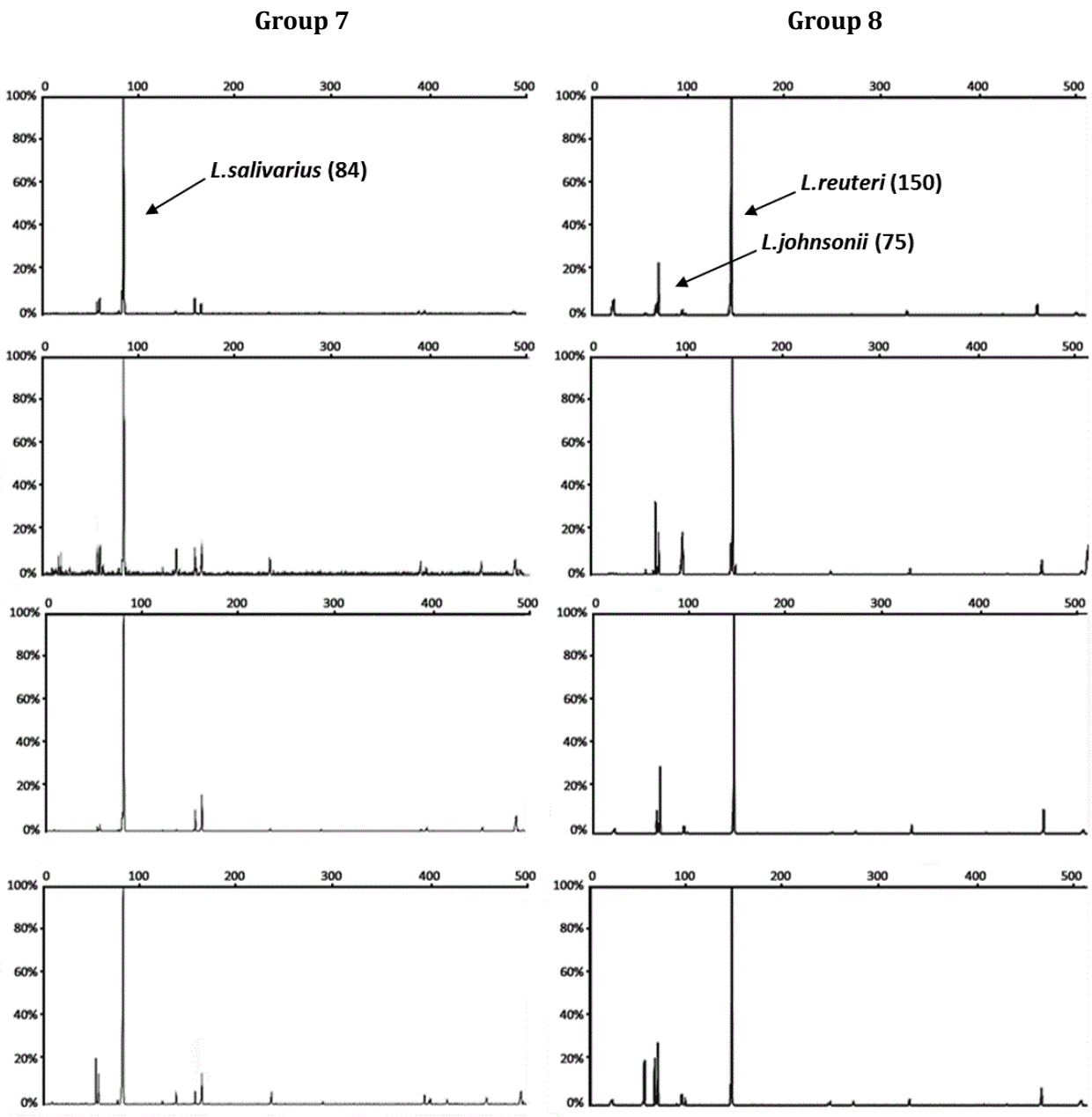


Fig. 3.3 T-RFLP raw data from the software Applied Biosystems Peak Scanner. Each panel represents a fecal extraction from a different mouse and only *Alu*I with labeled forward primer for Group 7 and 8 is depicted here. The X axis indicated the size of the TRFs, and the Y axis indicated the percentage of peaks height compared to the tallest peak. The TRF patterns within the sample dietary groups are very similar to each other, while between these two dietary groups are very different.

All 48 isolates from group 7 were identified as *L. salivarius* with a predicted TRF for *Alu*I-forward of 84 bp. In group 8, 18 isolates were identified as *L. johnsonii* (37%; predicted TRF for *Alu*I-forward of 75 bp) and 30 as *L. reuteri* (63%; predicted TRF for *Alu*I-forward of 150 bp). When compared to the T-RFLP data set for these two groups (**Fig. 3.3**), the dominance of *L. salivarius* in group 7 and *L. johnsonii* and *L. reuteri* in group 8 is confirmed.

3.5 Discussion

Selection of suitable primers and enzymes is a key step in designing T-RFLP protocols. Many investigators have used the 16S rDNA primers and enzymes *Hha*I, *Msp*I, *Rsa*I in T-RFLP analysis (Culman *et al.*, 2008) because extensive databases of sequences, primer sets and tools for analysis of 16S rDNA TRFs are well established. However, 16S rDNA gene sequence analysis is not able to reveal significant differences between recently diverged species, such as *L. plantarum, L. paraplantarum*, and *L. pentosus;* or *L. casei*, *L. rhamnosus*, and *L. zeae;* or *L. johnsonii* and *L. gasseri* (Blaiotta *et al.*, 2008; Walter *et al.,* 2000). In some cases the poor specificity of 16S rDNA-based TRF approaches make it difficult to identify bacteria at the species or genera level (Culman *et al.*, 2008). The 16S-23S rRNA internal transcribed spacer (ITS) has greater variation, which allows for discrimination of a larger number of *Lactobacillus* spp. (Moreira *et al.,* 2005). In addition, the 16S-23S rRNA ITS has been used as a template for T-RFLP in detecting 19 other species (Spasenovski *et al.*, 2009). However, Culman *et al.* suggested that the most optimal PCR primers for T-RFLP should have product length between 400 and 700 bp since this allows for the best possible estimation of diversity while avoiding the loss of data associated with long amplicons (Culman *et al.*, 2008). Song *et al.* indicated that most of the *Lactobacillus* spp. have an ITS between 200 to 300 bp (Song *et al.*, 2000), so the amplicons of ITS are usually too short to generate informative patterns in bacterial fingerprint analysis. The *hsp60*-based T-RFLP developed in this research overcomes limitations of the existing methods using either 16S rDNA or 16S-23S ITS. Seventy-nine distinct overall TRF

patterns were generated *in silico* by a set of fluorescently labeled *Lactobacillus* specific primers and endonuclease *Alu*I and *Tac*I, of which, 70 were specific for one species. The high variability and comparably long lengths of the *hsp60* nucleotide sequences allowed for the discrimination of very closely related species and even strains within these species, including *L. casei* and *L. rhamnosus*; *L. acidophilus* and *L. crispatus*; *L. acidophilus*, *L. helveticus*, and *L. amylovorus*; and *L. plantarum* and *L. pentosus* (Blaiotta *et al.*, 2008). These data confirm the usefulness of using *hsp60* as a target in *Lactobacillus* phylogenetic analyses.

Previous studies have shown that the number of rRNA genes varies significantly between different species (Culman *et al.*, 2008) and others have postulated that an average bacterial community has 3.8 16S rDNA copies per genome (Fogel *et al.*, 1999) while *L. acidophilus* has four copies in the genome (Roussel *et al.*, 1993). The TRF abundance may not describe the relative abundance of specific amplicons in a mixture, and in turn not be able to estimate relative bacterial abundance. Since the *hsp60* is present in only a single copy on the chromosome (Goh *et al.*, 1996; Kwok *et al.*, 1999), amplicon abundance after PCR is more proportional to cell abundance in the original sample without any prior calibration, which makes it possible to relate T-RFLP analysis to relative cell abundance in the original sample. Additionally, the single copy of *hsp60* avoids the potential problem associated with polymorphisms found in different copies of the same gene as has been seen with 16S rRNA (Pillidge *et al.,* 2009). This feature of *hsp60* makes is ideal for either generating a similarity index or being used in principle components analysis (PCA) (Culman *et al.*, 2008). We demonstrated that the *hsp60*-based T-RFLP could correctly reflect the quantitative relationship of the tested species in the mixed culture. This result demonstrates the potential utility of this tool for comparing the abundance of *Lactobacillus* spp. in the community.

In order to test the capability of *hsp60*-based T-RFLP in profiling the fecal *Lactobacillus* communities, fecal samples from 8 treatment groups of mice were analyzed. The graphs generated using raw data in the software Peak Scanner demonstrate that the TRF patterns from different

animals within the same treatment group are very similar, while comparison between groups are significantly different. PCA plot and AHC dendrogram were generated using a binary table. Individuals within the same dietary groups could be easily clustered together with either PCA plot or the AHC dendrogram. We also constructed PCA and AHC tests that incorporate peak area which generated very similar results with only trivial variation. However, using binary values was preferred, due to a greater resistance to errors from incomplete digestion. This test has also indicated that diet can have a great influence on the fecal *Lactobacillus* microbiota composition.

We compared the T-RFLP results with groups 7 and 8 to results from culture-dependent analysis. The major peaks found in T-RFLP can be clearly identified by *in silico* analysis of the *hsp60* sequences generated from the lactobacilli isolates. The TRF for *L. reuteri* and *L. salivarius* did not previously exist in the database and required *hsp60* sequencing of lactobacilli isolates to identify. However, *hsp60* sequencing of a few isolates will provide the necessary information to identify the major peaks in the T-RFLP data set if species identification is necessary. At a community level, our *Lactobacillus* T-RFLP could readily compare the lactobacilli community from one treatment to another.

In conclusion, we have shown that the *hsp60*-based T-RFLP method is a valid monitoring tool for *Lactobacillus* community dynamics, the high variability, suitable amplicon size and stable copy number in the chrome offers a higher usefulness than other PCR-based molecular approaches. The application of this method could considerably enhance and extend our current understanding of intestinal microbiota.

CHAPTER 4. IMPACT OF PIGLET AGE AND ROUTE OF DELIVERY ON ILEAL *LACTOBACILLUS* DIVERSITY

4.1 Abstract

Lactobacilli are inhabitants of the human and porcine GI tract. *Lactobacillus* diversity was characterized in piglets that were either Cesarean (CD, n = 11) or vaginally delivered (VD, n=9). Piglets were weaned on day 21 and ileal content samples were obtained on day 3, 14, 21 and 28. Total lactobacilli counts (CFU/g) increased from day 3 to 14, and were stable until day 21 and decreased at day 28. Ten random isolates per piglet were identified by partial (499 bp) *hsp60* gene sequencing using *Lactobacillus* specific primers. Six species were identified from 200 isolates: *L. johnsonii*, *L. mucosae*, *L. reuteri*, *L. amylovorus*, *L. delbrueckii* and *L. salivarius*. At day 3, the VD and CD piglets both had five identifiable *Lactobacillus* spp., but their identity (species) and relative abundance differed. At days 14 and 21, *L. johnsonii* was dominated in both CD and VD piglets (>85% of isolates). At day 28, *L. johnsonii* was dominant in VD (91%) piglets, whereas CD piglets had decreased *L. johnsonii* (72.5%) but had an increased abundance of *L. mucosae* (20%). *Lactobacillus*-specific terminal restriction fragment length polymorphism (T-RFLP) patterns generated from ileal content samples confirmed the *Lactobacillus* spp. composition identified by cultivation. Carbohydrate utilization profiles of 200 *Lactobacillus* isolates demonstrated intra-species diversity, confirming that several strains may be present for each *Lactobacillus* spp. identified. Thus, piglet age and route of delivery impacted the ileal lactobacilli community.

4.2 Introduction

Lactobacilli are important inhabitants of the human and animal gastrointestinal tract (GI) despite generally composing less than 1% of the human fecal bacterial population (Wall *et al.,* 2007). The *Lactobacillus* genus includes 80 recognized species and subspecies, many of which have been

isolated from human and animal fecal sources (Hammes *et al.,* 2006; Wall *et al.*, 2007). Lactobacilli have been shown to protect against harmful microorganisms and for proper intestinal immune system development of human newborns (Giraffa *et al.*, 2010; Bauer *et al.*, 2006). Additionally, lactobacilli have been widely exploited as probiotics for maintaining or improving human and animal health (Fuller R. 1989).

The GI tract of a newborn is considered sterile; however, during the birthing process, the neonate is exposed to a wide array of bacterial sources from the environment. The initial colonization of the newborn intestine first happens during delivery when contact with vaginal fluids and commensal skin microbiota from the mother occurs (Mackie, *et al.,* 1999). Additionally, the delivery environment can impact the assembly of the infant gut community. Importantly, these factors lead to the inoculation of the infant GI with numerous *Lactobacillus* spp. (Matsumiya *et al.*, 2002; Orrhage *et al.,* 1999; Deminguez-Bello *et al.*, 2010). However, many modern human babies are not exposed to vaginal microbes at birth because they are delivered by cesarean section (Deminguez-Bello *et al.*, 2010). Cesarean birth prevents exposure to many microbial sources and can potentially impact the initial colonization of lactobacilli in the infant gut. Previous research shows that cesarean section and vaginally delivered newborn infants have differences in the initial composition of the intestinal microbiota (Biasucci *et al.,* 2008; Penders *et al.,* 2006). However, there is no agreement on whether the route of delivery can impact the long-term development of intestinal the lactobacilli community. Some research shows that infants may receive lactobacilli at birth from the mother's vaginal fluid (Matsumiya *et al.*, 2002). The acquired lactobacilli do not appear to persist in the infant GI for the long term (Matsumiya *et al.*, 2002). Several investigators have speculated that *Lactobacillus* colonization due to route of delivery may contribute to the functional benefits that VD has over CD. Specifically, *Lactobacillus* colonization has been shown to be protective against GI infection and the reduction in sensitivity to dietary allergens (Murgas *et al.*,

2011). Therefore, it is important to understand what microbial community changes happen during and after birth of infants.

The difficulty in accurately describing a bacterial community in the GI tract is a major handicap in studies of microbial development and impact of route of delivery. Culture-based techniques have been considered the standard method for enumeration and species detection during the past several decades. Standard plating with selective agents such as bile, esculin or antibiotics can be used for selective enrichment of certain species (O'Sullivan, D. J. 1999.). Phenotypic tests such as serotyping and carbohydrate utilization can only be applied following cultivation. However, a majority of the microbial species are uncultivable because of environmental stress and selective agents present in the medium (Stackebrandt *et al*., 1995). Two commonly used media selective for lactobacilli are *Lactobacillus* selective (LBS) agar and de Man, Rogosa and Sharpe (MRS) agar which have showed significant differences in the degree of selectivity (Nelson, G.M. 1995). As a result, molecular methods based on 16S rDNA techniques were developed to improve microbial identification over the last 20 years such as qPCR (Ladero *et al.,* 2010) and FISH (Dominguez-Bello *et al.,* 2010). A recent modification is the use of the *hsp60* gene for the identification of lactobacilli using an RFLP approach (Bliotta *et al*., 2008). The two main advantages of using *hsp60* are [1] the greater sequence variability within the *hsp60* gene, which allows for greater discrimination, and [2] the *hsp60* gene is found as a single copy in lactobacilli genomes as opposed to the multi-copy nature of 16S rDNA, making the bias of amplification smaller in PCR reaction.

In this study, we used traditional culture-dependent methods to detect the composition of *Lactobacillus* spp. in piglet ileal content samples. Numerous lactobacilli isolates were phenotypically characterized by carbohydrate utilization assay. Additionally, the novel molecular identification tool, *Lactobacillus hsp60*-based Terminal restriction fragment length polymorphisms (T-RFLP), which we developed in previous research, was applied to compare to the cultured results and provide further information about the non-cultivable species.

4.3 Methods and Materials

4.3.1 Animal Protocols

Animals were managed throughout the study in accordance with the requirements of the Institutional Animal Care and Use Committee (IACUC) at the University of Illinois-Urbana Champaign. Piglets (n = 25) were born from six littermate sows (3 -5 piglets each sow) by either vaginal or cesarean delivery (VD: n=9; CD: n=11). The piglets from each VD- and CD sows were paired and fed by three VD sows for 21 days. The piglets were housed individually in temperature- and light-controlled rooms and weaned onto a standard weaning diet at day 21. The body weights of piglets were monitored daily.

4.3.2 Sample collection and bacterial colony counting

On day 3, 14, 21 and 28 after birth, both CD (n = 11) and VD (n = 9) piglets were sacrificed according to regulations, 4 or 5 piglets once. The animals were sanguinated, the abdomen immediately opened, and the samples of the ileal content were swabbed in 10 ml screw cap tubes which had 10 ml sterile 0.1% peptone saline suspension added beforehand. The samples were then homogenized by vortex mixing for 60 seconds at the maximum speed. Serial dilutions (10^1 to 10^6) of the samples collected were pour-plated onto *Lactobacillus* selective agar (LBS, BD). The agar plates were incubated anaerobically at 37°C for 72 hours. The bacterial concentration (CFU /g) was calculated.

To analyze the dominant *Lactobacillus* species of each piglet, 10 colonies were randomly selected from the LBS agar plates after counting and sub-cultured in MRS broth (BD) for 24to 48 hours. Ten colonies are believed to provide adequate representation of the major bacterial strains cultured on a selective medium (Hartley *et al.*, 1977).

4.3.3 DNA Extraction

Genomic DNA was extracted from the *Lactobacillus* isolates using the method described by Luchansky *et al.,* (1991) with slight modifications (Korhonen *et al*., 2007). Briefly, bacterial cultures grown overnight in 3 mL of MRS broth were centrifuged (2 minutes, 10,000 x g), followed by re-suspension and washed twice with 1.5 mL molecular biology grade water (MO BIO Laboratories, Inc. Carlsbad, CA). The supernatant was discarded and the pellet was re-suspended in 500 µL of TE buffer (10 mM Tris-HCl 1 mM EDTA [pH 8.0]). The bacterial cells were lysed by shaking for 1 minute on a minibead beater (Bio-Rad Laboratories, Hercules, CA) at a speed of 6 m/s in 2 mL screw-cap tubes with 350 to 400 mg of 100 µm diameter Silica beads (Research Products International Corp., Mt. Prospect, IL). After centrifugation at 12,000 x g for 2 minutes, 300 µL of supernatant was transferred to a 1.5 mL microcentrifuge tube and stored until purification at -20°C. The purification of DNA was conducted using the QIAamp DNA Stool Mini Kit (QIAGEN, Venlo, Netherlands) according to manufacturer's instruction. The DNA was eluted in 200 uL of Qiagen Buffer AE and was stored at -20°C.

4.3.4 *Lactobacillus hsp60* Gene Sequencing

Two *Lactobacillus* specific *hsp60* oligonucleotide primers, LB308F (TGAAGAAYGTNRYNGCYGG) and LB806RM (AANGTNCCVCGVATCTTGTT), previously described by Blaiotta (2008) (Also see **Table S2**), were used to amplify a 499 bp fragment of the *Lactobacillus hsp60* gene. PCR amplification was performed with a 50 µL total volume including 5 µL of target DNA (10 ng / µL), 5.0 µL of Taq DNA polymerase 10X Buffer (Lucigen Corporation, Middleton, WI), 2.5 µL of 50 mM $MgCl_2$, 0.5 µL of a deoxynucleoside triphosphate mix (25 mM each), 0.125 µL of each primer (0.1 mM), and 0.5 µL of Taq DNA polymerase. The PCR conditions consisted of 40 cycles (30 seconds at 94°C, 30 seconds at 37°C, and 1 minute at 72°C) and a final cycle at 72°C for 5 minutes. The samples were cleaned up using the Zymo DNA Clean & Concentrator Kit (Zymo Research Corporation, Irvine,

CA) according to the manufacturer's instruction Samples were sequenced using an ABI 3730xl capillary systems (Applied Biosystems Inc., Foster City, CA) by either primer LB308F or LB806RM. A FASTA file containing all of the sequences was constructed and uploaded onto the BLAST database to compare records in the NCBI nucleotide collection database. Based on previous studies with *Bifidobacterium* (Zhu *et al.*, 2003), *Vibrio* (Kwok *et al.*, 2002), and *Staphylococci* (Goh *et al.*, 1996), records with maximal similarity values greater than 97% were used to identify species in this study.

Phylogenetic analysis was performed using MEGA version 5.0. Multiple alignments of sequences belonging to the same species were applied and the UPGMA tree was constructed. In the tree explorer, branches shorter than 0.04 were clustered together and the species within each branch were considered as a single strain.

4.3.5 Simpson's Diversity Index (SDI)

The SDI for lactobacilli composition between CD and VD groups was calculated using the richness (number of species) and evenness (number of isolates of the same species) values of the lactobacilli community. The calculation was done using M.S. Excel 2010 (Microsoft, Seattle, WA) following an established formula $D = 1 - \left(\frac{\Sigma\, n(n-1)}{N(N-1)} \right)$. In which SDI value is calculated using n, the total number of organisms of a particular *Lactobacillus* spp., and the total number of organisms of all species N.

4.3.6 T-RFLP

Lactobacillus spp. specific T-RFLP was performed to study the diversity of lactobacilli in ileal contents. Two fluorescent dyes, the 6-FAM (6-carboxyfluorescein) and the HEX (6-carboxy-2',4,4',5',7,7'-hexachlorofluorescein) were separately attached to the forward and reverse *hsp60*

primers described above. Two replicate PCRs in 50 uL reaction mixtures were performed for each sample. The PCR products of the replicates were pooled together to reduce PCR variability. To increase the specificity of amplicons, DNA gel purification was done. The PCR products were run on a 1% (w/v) agarose electrophoresis gel at 100 Volts (Molecular Grade, Sigma, MO) in 1% (v/v) TAE buffer (40 mmol/L Tris-acetate, 1 mmol/L EDTA, pH 8.5) containing 1 mg/mL ethidium bromide. The bands of approximately 499 bp were visualized by ultraviolet (UV) transillumination, excised from the gel. The DNA was recovered using a Zymoclean Gel DNA Recovery Kit according to the manufacturer's instructions.

Two hundred ng of purified PCR product was digested by with 20 U of either restriction enzyme *Alu*I (New England Biolabs, Ipswich, MA) at 37°C for 2 hours or *Tac*I (Promega Corporation, Madison, WI) at 65°C for 1.5 hours. Following the restriction digestion, the DNA was analyzed on ABI 3730xl capillary systems.

The identification of the *Lactobacillus* spp. was achieved by matching of terminal restriction fragments (TRF). TRFs that differed by less than 0.5 bp in different profiles were considered identical and were clustered (Dicksved J. 2007). An online tool using Google spreadsheet was developed to rapidly conduct the pattern comparison (see appendix). The sizes of all the peaks from the 4 data matrices (*Alu*I-forward, *Alu*I -reverse, *Tac*I-forward, *Tac*I -reverse) can be entered in to this tool and the *Lactobacillus* spp. matching the known TRF patterns are highlighted.

4.3.7 Carbohydrate Fermentation Profile

In order to further characterize the 200 lactobacilli isolates, the fermentation capabilities of these strains were evaluated using six different carbohydrates: galactose (Gal), short-chain Fructooligosaccharide (sc-FOS), polydextrose (PDX), mannose (Man), cellobiose (Cel) and lactose (Lac). De-ionized water and dextrose solution were used as the negative and positive controls. All the strains were first cultured in MRS anaerobically for 24 hours and then inoculated into 96 well

plates with sterile pipet tips. Each well contained 250 µL modified MRS media devoid of carbohydrates (Barrangou *et al.*, 2003) with 0.01% (w/v) pH indicator (bromocresol purple) plus 10 µL of a 10% (w/v) filtration-sterilized carbohydrate solution. This special medium was used to determine the ability of bacteria to use various carbohydrates with minimal carbon resources. The 96 well plates were covered by transparent plastic films to prevent CO_2 absorption and moisture evaporation during incubation. The 96 well plates were then incubated anaerobically at 37°C overnight. Readings were taken after 24 hours and verified after 48 hours. Fermentation of carbohydrates in the modified carbohydrate free MRS was indicated by change in color from violet to yellow. Isolates which cannot change media color in 48 hours were considered as non-utilizers of the carbohydrates added in each well.

4.3.8 ERIC-PCR

To better differentiate the *Lactobacillus* strains within the same species using culture-independent approach, Enterobacterial Repetitive Intergenic Consensus Sequence-PCR (ERIC-PCR) was performed using ERIC-1R (5'-ATGTAAGCTCCTGGGGATTCAC-3') and ERIC-2 (5'-AAGTAAGTGACTGGGGTGAGCG-3') primers (Versalovic *et al.*, 1991). The concentration of components in the PCR reaction was the same as described in T-RFLP protocol. The PCR conditions are as follows: initial cycle of 95°C for 5 minutes; 30 cycles at 90°C for 30 seconds, 50°C for 30 seconds, 52°C for 1 minute, and 72°C for 1 minute; and final cycle 72°C for 8 minutes.

The ERIC-PCR products were separated using 1.5% (w/v) agarose electrophoresis gels for 1 hour. High-resolution images were obtained using a Fluor Chem 8900 fluorescence chemiluminescence and imaging system (Alpha Innotech, San Leandro, CA). A binary data table (presence or absence of individual bands) was generated, in which bands were assigned as absence or presence at every possible sizes in the picture. The binary table was imported into an Excel files and different patterns were recorded.

4.3.9 Statistical Analysis

The numbers of *Lactobacillus* spp. identified from each piglet, percentage and SDI values were analyzed using SAS (version 9.3, 2011, GraphPad Software, San Diego, CA). Normality of data was assessed. The standard ANOVA tables were constructed; variables were analyzed accordingly, using Student 2-tailed unpaired t test or nonparametric Mann-Whitney test. A value of $p \leq 0.05$ was considered significant.

4.4 Results

4.4.1 Basic Parameters of Animals

Body weight, small intestinal weight and small intestinal length of the piglets increased over time from day 3 to day 28 (from 1.90 ± 0.37 kg to 8.40 ± 1.99 kg, 495.0 ± 36.28 cm to 1000.7 ±84.2 cm, 49.8 ±13.5 g to 239.07 ±47.1 g; respectively).

4.4.2 Lactobacilli Community Analysis

Ten randomly selected isolates from ileal content samples of each piglet were cultivated for the lactobacilli community analysis. An approximately 499 bp of the partial *hsp60* gene from the randomly selected bacteria isolated from LBS was sequenced to identify the isolates to the species level. In total, 200 sequences (110 CD and 90 VD) were collected during this study (**Table S1**).

Six different *Lactobacillus* spp. were identified amongst the 200 isolates: *L. johnsonii, L. mucosae, L. reuteri, L. amylovorus, L. delbrueckii* and *L. salivarius*. An UPGMA tree based on comparison of trimmed *hsp60* gene sequences divided the two most abundant species, *L. johnsonii* and *L. reuteri,* to strain level. Branches shorter than 0.04 were clustered together as a strain and were given a name (**Fig. 4.1**). Six strains were identified within *L. johnsonii* isolates and 4 strains identified within *L. reuteri* isolates based on *hsp60* sequences. Species *L. mucosae, L. amylovorus, L. delbrueckii,* and *L. salivarius* had highly similar sequences (>99% identical), indicating that there

was only one strain within each species that could be identified with hsp60 sequencing. The percentage and number of each strain at different ages are listed in **Fig. 4.2**.

The major *Lactobacillus* spp. in samples on day 3 from both groups was *L. reuteri* (>45%). On day 14, day 21 and day 28, the major *Lactobacillus* spp. for both groups was *L. johnsonii* (>72.5%). *L. delbrueckii* was only found in CD group. *L. salivarius* was only found in day 3 samples (**Fig. 4.2**). The overall composition of species was similar for both modes of delivery. On day 3, both modes of delivery had five species, with 4 shared however the percentages of each species were different.

The Simpson's Diversity Index (SDI) values for CD and VD piglets were both highest on day 3 (CD=0.71, VD=0.78), indicating that the lactobacilli communities were most diverse shortly after the piglets were born. On day 14 and day 28, the SDI values for both groups declined. The value for the VD piglets decreased to 0, indicating only one species, *L. johnsonii* was detected. However, after weaning (day 21), the diversity value of the CD group increased to 0.44, while that of the VD group decreased to 0.13 (**Fig. 4.3**)

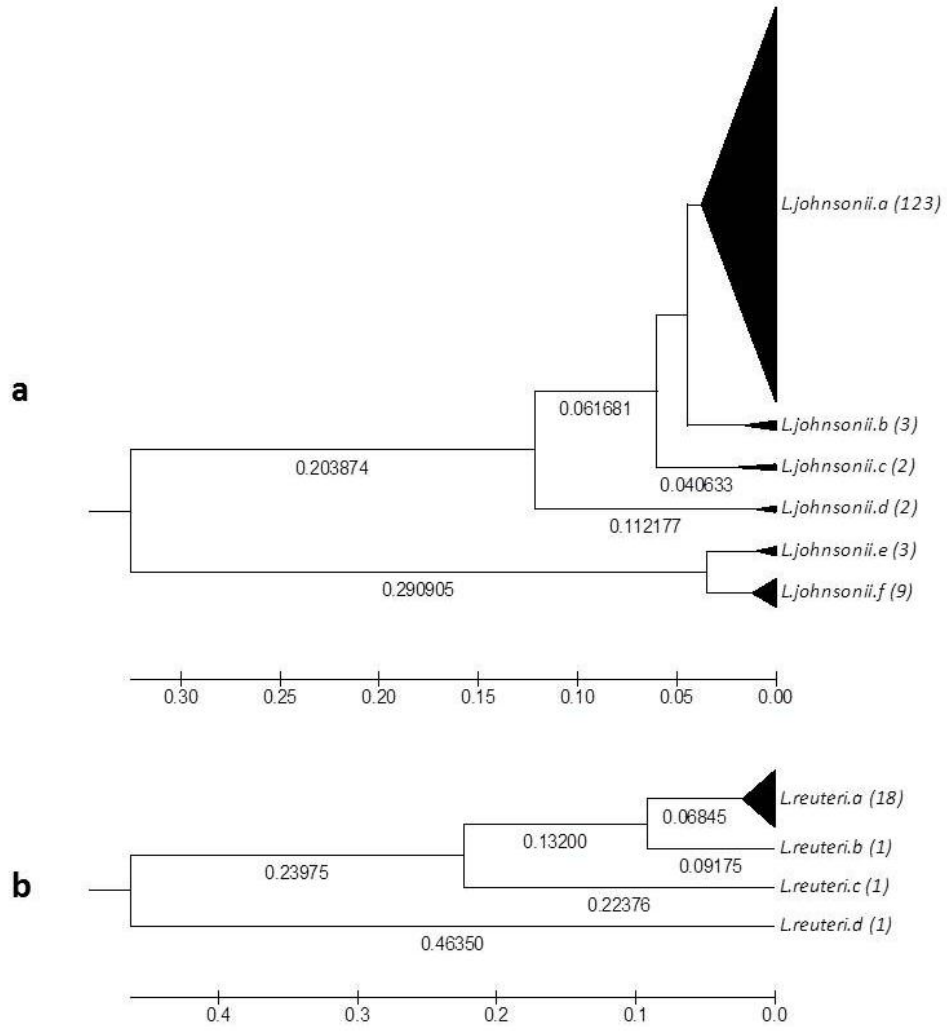


Fig. 4.1 UPGMA tree based on comparison of trimmed *hsp60* gene sequences showing the phylogenetic relationships between isolates of *L. johnsonii* (a) and *L. reuteri* (b). Branches shorter than 0.04 were clustered together, the isolates within each cluster were considered as belonging to the same strain, which was named as species name plus letter a to f. The number in the brackets shows the amount of isolates of each strain. Eight strains were defined within the species *L. johnsonii*, named as (*L. johnsonii.a-f*). Four strains were defined within the species *L. reuteri,* named as (*L. reuteri.a-d*). For the species *L. mucosae, L. amylovorus, L. delbrueckii,* and *L. salivarius,* the sequences showed a high similarity between isolates (branches shorter than 0.04) and were considered as a single strain.

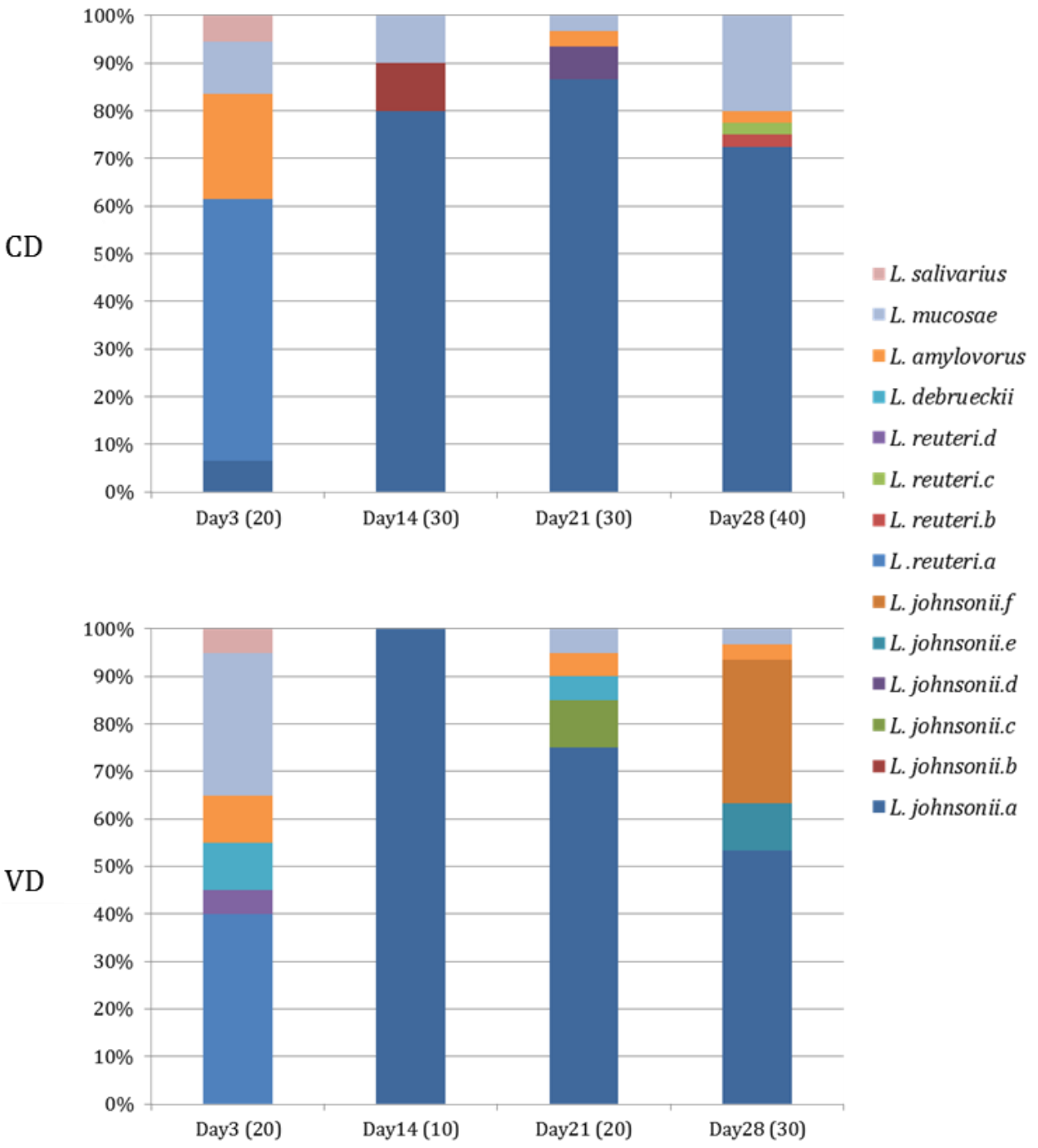


Fig. 4.2 The lactobacilli community composition of CD and VD piglets at four time points, detected by sequencing isolates randomly selected from the LBS plates. The percentages indicate the percentage of each *Lactobacillus* spp. strain and the numbers in the brackets indicate the number of isolates sequenced.

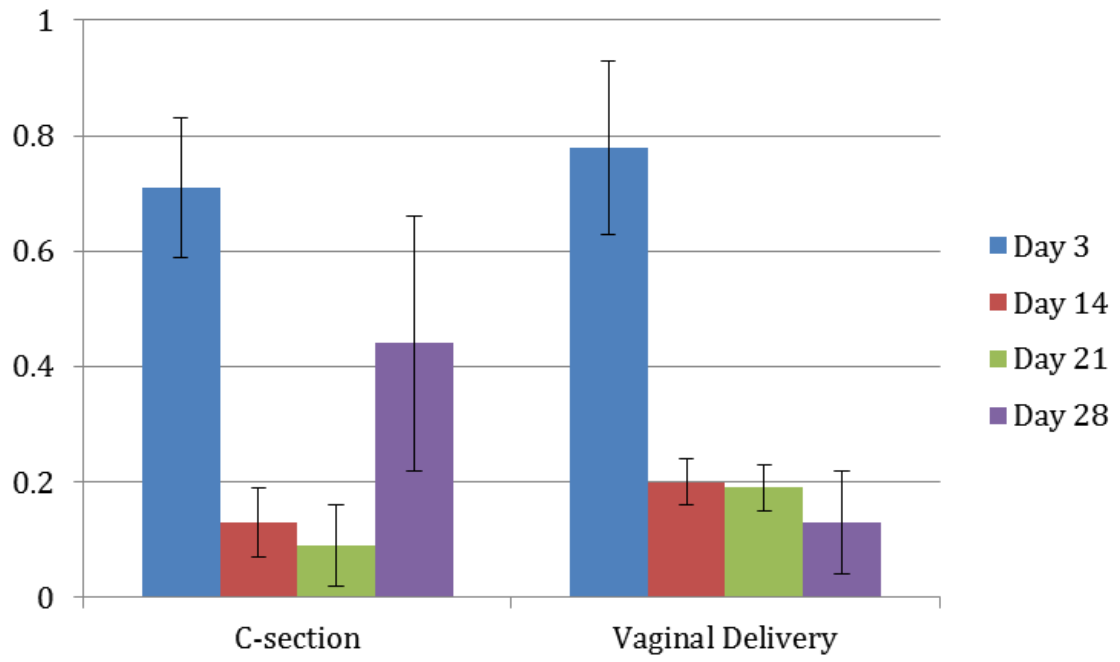


Fig. 4.3 The Simpson's Diversity Index (SDI) for lactobacilli composition between Cesarean and vaginal delivered piglets. The values were calculated using isolate sequencing results at the four time points. Statistical analysis showed that there was no significant difference between SDI of CD and VD groups ($p_{model} = 0.88$).

4.4.3 T-RFLP Result

Lactobacillus specific *hsp60* based T-RFLP was performed to verify the identification of species recovered using culture-dependent methods. Previously extracted total genomic DNA from four representative ileal content samples: piglet 308 (CD, day 21), 408 (VD, day 21), 306 (CD, day 28), 506 (VD, day 28) were used as templates. These four representative ileal content samples had the all of the *Lactobacillus* spp. detected via our culture-dependent strategy and are good for method comparison.

The TRF identified in this study is listed in **Table 4.1**, and the names of their corresponding species were given. It showed that totally 9 *Lactobacillus* spp. were identified by T-RFLP including

all seven species observed by individual isolate sequencing. T-RFLP provided strain information by comparing to our known TRF pattern database (**Table S1**). Additionally, two species, *L. acidophilus* and *L. zaea*, were only detected by T-RFLP, indicating that their abundance might be too little to be recovered using culture-based methods.

Table 4.1 TRF sizes of the species identified by *hsp60*-based T-RFLP

	***Alu*I**[1]		***Tac*I**	
	Forward	**Reverse**	**Forward**	**Reverse**
L. johnsonii (A)[3]	75[2]	34	224	221
L. mucosae	274	169	56	357
L. reuteri (A)[3]	84	225	499	0
L. amylovorus (B)[3]	72	34	278	221
L. salivarius (B)[3]	330	34	178	221
L. delbrueckii (F)[3]	330	34	449	0
L. vaginalis	84	34	278	221
L. acidophilus (A)[3,4]	135	33	217	220
L. zeae (C)[3,4]	330	168	56	281

[1] All the primers are fluorescence labeled
[2] Unit of the value: Basepairs (bp)
[3] The letters in the brackets indicate the strain of the species detected. Some species have more than one TRF patterns, which can be used as indicator of different strains within the same species.
[4] Only detected by T-RFLP.

4.4.4 Carbohydrate Utilization Profile and ERIC-PCR Result on *L. johnsonii* Interspecies Discrimination

Comparison of the carbohydrate utilization of the whole lactobacilli communities on each days showed that the CD and VD groups had similar utilization profiles (**Table 4.2**). One exception being after weaning, the lactose utilization stayed the same (100%) in CD group, while there was a slight

decrease from 100% to 83% in the VD group. This observation indicated that as solid food was introduced on day 21, abundance of species using lactose as their carbon sources decreased. The carbohydrate utilization test also revealed that there may potentially be more than one strain within each species detected, which reflected the complexity of ileal lactobacilli microbiota.

Table 4.2 Percentage of *Lactobacillus* isolates capable of utilizing different carbohydrates in the VD and CD piglets[1].

		Day3	Day14	Day21	Day28
CD	Galactose	100%	100%	83%	100%
	sc-FOS	40%	100%	100%	53%
	Polydextrose	10%	0%	0%	0%
	Mannose	30%	100%	80%	40%
	Cellobiose	20%	87%	73%	58%
	Lactose	65%	100%	100%	100%
VD	Galactose	100%	100%	100%	100%
	sc-FOS	65%	90%	95%	53%
	Polydextrose	5%	0%	0%	0%
	Mannose	45%	50%	65%	13%
	Cellobiose	15%	90%	60%	47%
	Lactose	80%	100%	100%	83%

[1]Water and dextrose were used as negative and positive controls

Carbohydrate utilization profiles of each isolate were defined by testing its capability to ferment six different carbohydrates. The strains which had carbohydrate profiles were treated as different strains within a single species. The percentages of isolates of different strains are shown in **Table 4.3**. The carbohydrate utilization profiles of strain A to J are listed in **Table 4.4.**

Table 4.3 Percentages of *Lactobacillus johnsonii* strains isolated from piglet ileal samples. The strains were defined by their carbohydrate utilization profile (A to J)[1].

	Day	A	B	C	D	E	F	G	H	I	J
C	3	33%	33%	33%							
	14		4%		74%	11%	11%				
	21		4%		61%	14%	14%	7%			
	28		34%		28%	17%	3%		3%	14%	
V	3										
	14				100%						
	21				53%	24%	24%				
	28		14%		14%	54%				4%	14%

[1]The definition of strain A to J are listed in **Table 4.4.**

Table 4.4 Carbohydrate utilization profiles of *Lactobacillus johnsonii* (A to J) strains isolated from piglet ileal samples. Six different polysaccharides were used in the carbohydrate utilization test.

	Galactose	sc-FOS	Polydextrose	Mannose	Cellobiose	Lactose
A	- [1]	-	-	-	-	+
B	+ [2]	-	-	-	-	+
C	+	+	+	+	-	+
D	+	+	-	+	+	+
E	+	+	-	-	+	+
F	+	+	-	+	-	+
G	+	+	-	-	-	+
H	+	-	-	-	+	+
I	+	-	-	-	-	-
J	+	-	+	+	+	+

[1] "-" indicated that the strains cannot utilize this saccharide as sole carbon source
[2] "+" indicated that the strains can ferment this saccharide as sole carbon source

The carbohydrate utilization test revealed that the intra-species composition of the detected *Lactobacillus* spp. was very complex. Ten strains (A to J) were found amongst the 142 *L. johnsonii* isolates in the CD or VD groups based on their different carbohydrate utilization profile (**Table 4.3**). The strains detected within CD and VD groups were also different. In the CD group, there were generally more strains detected than the VD group at the every time point. There were 9 strains found in CD group while only 6 strains in VD group.

As a genotypic method to discriminate closely related strains, ERIC-PCR confirmed that there were multiple strains defined by comparing the electrophoresis profile of different isolates (data not shown). This is consistent with our observation in the carbohydrate utilization test. However, the strains defined based on their phenotypic features (carbohydrate utilization profile) and their genotypic features (ERIC-PCR) were not the same. The differing results from the carbohydrate utilization test and ERIC-PCR shows the complexity of the lactobacilli community and the connection between the genotypes and phenotypes of *Lactobacillus* spp.

4.5 Discussion

In this study, we demonstrated the similarity and differences in lactobacilli community composition with 2 different modes of delivery at 4 time points after birth. Six different species: *L. johnsonii, L. mucosae, L. reuteri, L. amylovorus, L. delbrueckii* and *L. salivarius* were detected from 200 isolates by sequencing a 499 bp partial *Lactobacillus hsp60* gene. *L. delbrueckii* has been found in human, mouse and rat intestinal samples (Hammes *et al.,* 2006), but not previously reported in piglet intestinal or fecal samples. *L. reuteri* was the major species detected on day 3 in both CD (50%) and VD (45%) groups, and was replaced by *L. johnsonii* on day 14, 21, 28 (>72.5%). To our knowledge, this study reports *L. johnsonii* as the predominating *Lactobacillus* species for the first time in piglet GI tract. Hammes *et al.* reported that *L. amylovorus, L. salivarius* and *L. reuteri* were predominant (Hammes *et al.,* 2006). Du Toit *et al.* (2001) reported that *L. amylovorus* and *L. plantarum* were predominant in the piglet GI. The fact that one single species comes to dominate

the ileal lactobacilli community may be a result of specific growth or selection factors from breast milk (Marina Elli *et al.*, 1999) or within the GI tract environment (Bezkorovainy A., 2001).

The discrimination power of *Lactobacillus hsp60* sequencing technique was used to detect strain level identification. Previously, 16S rDNA gene sequence analysis does not allow for identification between recently diverged species, such as *L. plantarum, L. paraplantarum*, and *L. pentosus* or *L. casei*; *L. rhamnosus*, and *L. zeae* (Blaiotta, G. 2008). *L. johnsonii* and *L. gasseri* are difficult to distinguish from each other even by molecular techniques using 16S rDNA marker (Walter *et al.*, 2000). Here, most *L. johnsonii* isolates had sequences highly similar (>99%) to NCBI nucleotide collection records, and could be easily distinguished from other closely related species. In this study, *hsp60* sequencing defined strains within each species and enabled the characterization of the ileal lactobacilli composition over time (28 days). Even when the abundance of each species remains unchanged, there can be great variety on the composition of the strains within species.

Here we demonstrated that the piglet ileal *Lactobacillus* community during the first 3 days of life is influenced by the route of delivery. The lactobacilli communities of the CD and VD groups showed a noticeable difference on day 3. The CD group had five species detected while the VD group had 6. Four species were shared between the groups however the relative abundance for each species was not the same. Therefore, route of delivery may have an impact on lactobacilli community initiation at an early stage of life. This conclusion was in agreement with previous clinical and animal research (Hall *et al.*, 1990; Deminguez-Bello *et al.*, 2010; Orrhage *et al.,* 1999).

After initiation, the impact of modes of delivery had become less distinguished. Weaning is a dietary transition which has been show to change the bacterial composition in the small intestine (Castillo, M. 2007; Pieper, R. 2007). At day 28, the species detected in the CD group remained the same as day 3, whereas the relative abundance changed.

In the VD group on day 3, there was a greater number of species and diversity than the CD group. However 2 of 6 species on day 3 were detected in the ileal content samples on day 28. Additionally,

these species had a decreased abundance (~3.3%). The predominant species detected was *L. johnsonii* (>93%). This indicates that vaginal fluid may be a resource of *Lactobacillus* spp. However, growth factors present in sow's milk or nutritional factors from solid food after weaning had greater impacts on shaping bacterial composition. The Simpson's Diversity Index (SDI) helped visualize the change in diversity of the lactobacilli community. The diversity for both groups was highest on day 3, and then decreased rapidly in both groups. After weaning, diversity increased again. This result is consistent with previously published research (Matsumiya *et al.*, 2002).

Moreover, this study compared the application of a newly-developed molecular biological tool, *Lactobacillus hsp60*-based T-RFLP and traditional methods. T-RFLP is an increasingly popular molecular approach which can determine both species dominance and species richness within samples. T-RFLP allows for the assessment of a structure of complex bacterial communities and rapid comparison of the community diversity between different ecosystems (Sakamoto, M. 2003). Traditional T-RFLP analysis is based on bacterial 16S rDNA specificity. In order to discriminate closely related *Lactobacillus* spp., alternative molecular markers were used in previous studies (Berthier *et al.,* 1998; Torriani *et al.,* 2001). In this study, we used the *Lactobacillus hsp60* gene as the molecular marker. Seven species were detected by culture-dependent methods, all of which were found by TRF profile matching. Additionally, two other species, *L. acidophilus* and *L. zeae*, were only detected by T-RFLP. This indicates that T-RFLP is sensitive enough to detect the species less than 0.5% in population that could not be efficiently sampled by traditional methods.

The carbohydrates utilization profiles offered more phenotypic information and allowed for a prediction of the potential function for the overall lactobacilli community. Six carbohydrates were used in this test: Gal, sc-FOS, PDX, Man, Cel and Lac. Generally, the carbohydrate utilization profiles from CD and VD groups were very similar. One exception being after weaning, the lactose utilization stayed the same (100%) in CD group, while there was a slight decrease from 100% to 83% in the VD group. This may indicate that the number of those strains which utilize lactose as a main

carbon source decreased, as lactose was not the carbohydrate source in the solid food. The carbohydrate utilization test also revealed that there may potentially be more than one strain within each species detected. Strains defined using carbohydrate utilization profile showed that there was dynamic change within *Lactobacillus* spp. at different time, because the strains were capable of using different carbon sources, how they interacted with host and other species is considerably complex.

In conclusion, our results confirm the establishment of the gut lactobacilli community as a gradual process. The route of delivery does affect the early stage of ileal lactobacilli colonization; however, growth factors in the breast milk and in solid food after weaning had greater impact on the community composition. More importantly, we found the diversity at the species level to be much higher than at the genus level, which has been largely ignored by previous studies. Unfortunately, we did not have powerful molecular assay to characterize *Lactobacillus* spp. at strain level, which is the objective of our next study. Our results may stimulate new ideas in looking deeper into the composition of lactobacilli community at the early stages of colonization. While the relative abundance of *Lactobacillus* spp. from both modes of delivery routes might be very similar, at the species and strain level, the diversity of the lactobacilli community is different.

CHAPTER 5. ROUTE OF DELIVERY AND NUTRITION ALTER THE ILEAL MICROBIOTA

5.1 Abstract

The colonization and development of the gastrointestinal (GI) tract microbiota begins immediately at birth and is thought to be modulated by numerous factors including route of delivery and diet. In this study, 25 neonatal piglets were delivered and randomized into six treatment groups in a 2 X 3 design. They were delivered either vaginally (V) or by Cesarean section (C), and were either sow-reared (S), fed with formula milk (F), or fed with formula milk supplemented with 2g/L each galactooligosaccharides and polydextrose (FP). Following 21 days on diet, the ileal microbiome was analyzed by high-throughput pyrosequencing. Operational taxonomic unit (OTUs) based analysis revealed that the ileal bacterial diversity (Shannon Diversity Index) and richness was not significantly different amongst the groups. Phylotype analysis demonstrated that the overall bacterial communities were statistically similar at the genus level. Additionally, high-throughput pyrosequencing (culture-independent) and culture-dependent assays were done to obtain in depth information about the ileal lactobacilli community. Thirty different *Lactobacillus* spp. were detected, ten of which were present in over 1% relative abundance: *L. amylovorus, L. reuteri, L. plantarum, L. delbrueckii, L. helveticus, L. agilis, L. salivarius, L.mucosae, L.jonsonii* and *L.crispatus*. Culture-based Sanger sequencing was consistent with the high-throughput sequencing results. In conclusion, route of delivery significantly impacted the relative abundance of *Lactobacillus* spp. at genus and species level. Sow-reared piglets (S) had a different bacterial community structure compared to formula-fed piglets (F and FP) at the species level, and prebiotics could potentially impact the composition of the lactobacilli community at the species level.

5.2 Introduction

The gastrointestinal (GI) microbiome is a diverse collection of microorganisms that are assembled shortly after birth (Dethlefsen *et al.*, 2006) and is closely associated with the health status of its host (Cummings and Macfarlane, 1997). During physiological development the relationship between the host and its microbiome is very important. The GI microbiome has been shown to influence intestinal barrier function (Vaughan *et al.*, 2002), immune development (Isolauri *et al.*, 2002) and host metabolism (Martin *et al.*, 2008). However, unfavorable dysbiosis in the GI microbiome have been linked to an increased risk or development of many diseases such as, obesity and its associated syndromes (Ley *et al.*, 2006; Zhang *et al.*, 2009); necrotizing enterocolitis (Neu *et al.*, 2011); irritable bowel diseases (Kassinen *et al.*, 2007); and intestinal cancers (Marchesi *et al.*, 2011). There has been substantial research investigating the relationship between the GI microbiome and the host. However, this relationship is very complex and more research is needed.

Many investigators are interested in the contributions to the GI microbiome of individual bacterial taxa. In particular, several studies have investigated the relationship of lactobacilli and the GI microbiome (Euzeby *et al.*, 1997; Sanders *et al.*, 2003). Specific *Lactobacillus* spp. have traditionally been exploited in food fermentations however more recently they have been added to foods as probiotics (Euzeby et al., 1997; Sanders et al., 2003). Due to the importance of lactobacilli it is necessary to determine the composition of this genus in the GI tract.

Multiple factors can influence the development of the GI microbiota of neonates: route of delivery (Mackie *et al.*, 1999), exposure to the external environment during delivery (Dominguez-Bello *et al.*, 2010), exposure to maternal microbiota (faecal, vaginal and skin), and especially, the type of nutrition after birth (breast-fed vs. formula- fed) (Herfel *et al.*, 2009; Stark *et al.*, 1982). The long term impact of the route of delivery in the assembly of the GI microbiome is not well described. Several studies have shown that the composition and assembly of organisms is different between C-section and vaginally delivered newborn infants (Biasucci. *et al.*, 2008; Mackie, *et al.*,

1999; Penders *et al.*, 2006). A study investigating lactobacilli colonization of vaginally delivered babies reported that vaginal lactobacilli were detected in the feces of the majority of the infants at 5 days of age. At one month, the vaginally acquired lactobacilli were replaced by other lactobacilli not associated with their mother's vagina. The authors conclude that early lactobacilli colonizers do not persist in the infant intestine and are replaced by lactobacilli from other sources (Matsumiya *et al.*, 2002). However, early alterations in the infant microbiome may have long term health impacts (Collado *et al.*, 2012)

Human milk oligosaccharides (HMOs) are the third largest component of human milk (Kunz *et al.*, 2000) and have been shown to impact the microbiota composition by various mechanisms including prebiotic effects, anti-adhesive effects and host glycome-modifying effects (Bode *et al.*, 2009). Research directly using HMOs has been slowed, partially due to the difficulty in obtaining large quantities of the compounds economically. Therefore, research groups have been evaluating various prebiotics as an economic alternative for HMOs in infant formula.

A mixture of galactooligosaccharide (GOS) and polydextrose (PDX) is a commonly used prebiotic blend used in infant formula studies (Ashley *et al.*, 2012; Monaco *et al.*, 2011). GOS is a non-digestible oligosaccharide chain composed of galactose units with a glucose unit at the reducing end (Sako 1999). PDX is a soluble fiber synthesized from glucose; it is frequently used to increase the fiber content of food (Ranawana *et al.*, 2012). Both GOS and PDX have shown prebiotic effects in human infants based on analysis of fecal samples (Ashley *et al.*, 2012). However, since lactobacilli are usually the predominant species in ileum, information of ileal lactobacilli community is crucial for understanding the potential health benefits of GOS and PDX. As fecal bacterial community composition is significantly different than ileal microbiota (Hayashi *et al.*, 2005), study directly using ileal sample is necessary in order to obtain accurate information on bacterial community change.

Currently there are various tools available to study the effect of multiple factors on the microbial community in the GI tract. Bacterial Tag-Encoded FLX Amplicon Pyrosequencing (bTEFAP) is a pyrosequencing protocol that makes use of a novel tag priming method and an efficient bioinformatics pipeline. It allows multiplexing (Hamady *et al.*, 2008) and automatic computer analysis (Dowd *et al.*, 2008). The extremely high-throughput and comparatively cost are the most beneficial aspects of pyrosequencing technologies over other "traditional" culture-independent methods. Now, bTEFAP has been used as a novel tool in many microbial ecology studies including human microbiota research (Jones *et al.*, 2009; Costello *et al.*, 2009; Callaway *et al.*, 2009).

In this study, we analyzed the ileal bacterial community of neonatal piglets from different route of delivery and diet groups. bTEFAP was used to generate culture-independent information about the ileal microbiome. Due to the importance and relative abundance of lactobacilli in the ileum, we characterized the lactobacilli ileal community using a culture-based approach which was compared to the bTEFAP results.

5.3 Methods and Materials

5.3.1 Animal Model and Diet

Animals were cared for in accordance with the requirements of the Institutional Animal Care and Use Committee (IACUC) at the University of Illinois Urbana-Champaign (IACUC protocols 08070 and 08015) and in accordance with the NIH guidelines for animal research.

A total of 25 neonatal piglets were delivered and randomized into six groups: vaginally delivered (V) or delivered by Cesarean section (C), and were either sow-reared (S), fed formula milk (F), or fed formula milk containing mixed probiotics (see description below; FP). At 3 days of age, all piglets and sow were transported to the University of Illinois Urbana-Champaign animal facilities. The sow-reared group piglets (C:S, V:S) were co-housed with the nursing sow, whereas the formula group and prebiotic group piglets (C:F, C:FP, V:F, V:FP) were housed individually in cages. All

piglets were subjected to a 12 hour light/dark cycle. Rooms were maintained at 25°C with supplemental heat provided by radiant heaters suspended above the cages to provide a local temperature between 30 and 32°C.

The sow-reared piglets (C:S, V:S) were fed by the nursing sow multiple times per day based on the natural nursing cycle. The formula piglets (C:F, V:F) were fed with a bovine milk-based formula (Advance Baby Pig LiquiWean; Milk Specialties, Dundee, IL). The prebiotic piglets (C:FP, V:FP) were fed with the same formula supplemented with 2 g/L PDX (w/v) and 2 g/L GOS (w/v; Formulas were prepared fresh each morning (final solids of 18.3%) and were delivered 14-times daily. Formula was dispensed via a pump into a bowl at a rate of 360 mL/kg body weight. Piglets were weighed each morning and monitored three times per day for general health (Poroyko *et al.*, 2010).

5.3.2 Sample Collection and DNA Extraction

Piglets were sacrificed following 21 days on diet. Prior to sacrifice, piglets were sedated with Telazol (7 mg/kg body weight; Fort Dodge Animal Health, Fort Dodge, IA) followed by euthanasia with an intracardiac injection of sodium pentobarbital (72 mg/kg body weight; Fatal Plus, Vortech Pharmaceuticals, Ltd., Dearborn, MI). The ileum was excised and opened longitudinally, the ileal contents were collected into two separate 2 ml tubes, one tube was stored at -80°C and the other was held on ice for same day bacterial enumeration.

The bacterial genomic DNA was extracted from the ileal contents and was purified by the QIAamp DNA Stool Mini Kit (Qiagen, Valencia, CA) in combination with FastPrep-24 System (MP Biomedicals, Solon, OH). Briefly, 200 mg (wet weight) of ileal contents were weighed into a 2 ml eppendorftube containing glass beads. A 1 mL aliquot of ASL-buffer was added to the eppendorf tube and was shaken with the Fastprep-24 at 6 m/s for 30seconds. The sample was then incubated at 95°C for 5 minutes. After centrifugation at 20,800 x g for 1 minute, 0.8 mL of supernatant was collected into a 2 mL tube and 400 µL of ASL-buffer was added. The sample was mixed and treated

with one Inhibit EX tablet to remove the DNA-damaging substances and PCR inhibitors. After 3 minutes of centrifugation, 200 µL of supernatant was treated with proteinase K, AL-buffer, and precipitated with ethanol according to the manufacturer's instructions. DNA in the sample was further purified on a QIAamp spin column and eluted in 200 µL of AE-buffer and stored at −20°C. DNA concentration was determined with a NanoDrop 1000 spectrophotometer (NanoDrop Technologies, Wilmington DE).

Individual lactobacilli isolates from ileal contents were cultivated using traditional culture-dependent methods. Briefly, the ileal samples were added to a 10 mL sterile 0.1% (w/v) peptone saline suspension and homogenized by vortexing for 1 minute at maximum speed. Serial dilutions (10^1 to 10^6) were pour-plated onto *Lactobacillus* selective (LBS) agar (BD) for selective growth of lactobacilli. The agar plates were incubated anaerobically at 37°C for 72 hours. To analyze the dominant *Lactobacillus* spp. of each piglet, five colonies were randomly selected from the LBS agar plates and sub-cultured in 3 ml de Man, Rogosa and Sharpe (MRS) broth (BD) for 24 to 48 hours. After cultivation, lactobacilli isolates were frozen for long term storage at -80°C.

5.3.3 *Lactobacillus* Identification

Genomic DNA from the presumptive lactobacilli isolates was extracted using the method described by Luchansky *et al.* (1991) with slight modifications (Korhonen *et al.*, 2007). Briefly, bacterial cultures grown overnight in 3 mL of MRS broth were centrifuged (2 min, 10,000 x g), following centrifugation the cells were re-suspended and washed twice in 1.5 mL molecular biology grade water (MO BIO Laboratories, Inc. Carlsbad, CA). The supernatant was discarded and the pellet was re-suspended in 500 µL of TE buffer (10mM Tris-HCl-1 mM EDTA [pH 8.0]). The bacterial cells were lysed by shaking for one minute on a minibead beater (Bio-Rad Laboratories, Hercules, CA) at a speed of 6 m/s in 2 mL screw-cap tubes containing 350 to 400 mg of 100-µm-diameter Silica beads (Research Products International Corp., Mt. Prospect, IL). After centrifugation at 12,000 x g

for 2 minutes, 300 µL of supernatant was transferred to a 1.5 ml eppendorf tube and stored at -20°C until sequencing.

Two bacterial 16S rDNA oligonucleotide primers, Gray28F (5'-GAGTTTGATCNTGGCTCAG-3') and Gray519R (5'-GTNTTACNGCGGCKGCTG-3') (Ishak *et al.,* 2011) (Also see **Table S2**), were used to amplify a 510 bp fragment internal to the presumptive lactobacilli 16S rDNA V1-V3 region. PCR amplification was performed with a 50 µL total volume including 5 µL of target DNA (10 ng/µL), 5.0 µL of Taq DNA polymerase 10X Buffer (Lucigen Corporation, Middleton, WI), 2.5 µL of 50 mM $MgCl_2$, 0.5 µL of a deoxynucleoside triphosphate mix (25 mM each), 0.125 µL of each primer (0.1 mM), and 0.5 µL of Taq DNA polymerase. The PCR consisted of 30 cycles (30 seconds at 94°C, 30 seconds at 60°C, and 1 minute at 72°C) followed by an additional cycle at 72°C for 5 minutes. The samples were cleaned up using the Zymo DNA Clean & Concentrator Kit (Zymo Research Corporation, Irvine, CA) according to the manufacturer's instructions. The samples were sequenced on an ABI 3730xl capillary systems (Applied Biosystems Inc, Foster City, CA) using the forward primer Gray28F. The sequences were retrieved and a FASTA file containing all of the sequences was constructed and uploaded onto the BLASTn search engine to compare records in the NCBI nucleotide collection database, records with maximal similarity values greater than 97% were used to identify species.

5.3.4 Bacterial Tag-Encoded FLX 454-Pyroseqencing

Genomic DNA isolated directly from frozen ileal contents (Section 5.3.2) were analyzed by 454-pyrosequencing by the Research and Testing Laboratory (RTL, Lubbock, TX) based upon RTL protocols (Dowd *et al.,* 2008). The 16S rDNA primer set used for pyrosequencing were Gray28F (5'-GAGTTTGATCNTGGCTCAG-3') and Gray519R (5'-GTNTTACNGCGGCKGCTG-3') targeting the span of variable region V1-V3 in the bacterial 16S rDNA gene. bTEFAP was performed on the Genome Sequencer FLX instrument as previously described by Bailey using Titanium protocols and reagents

(Roche, Indianapolis, IN) (Bailey *et al.*, 2010). For a detailed description of pyrosequencing operation and protocol, refer to Ishak and colleagues (Ishak *et al.,* 2011).

5.3.5 Raw Data Processing and Sequence Preparation

Bacterial 16S rDNA sequences were retrieved and converted to a FASTA file. Sorting by tag sequence, trimming and quality control was done by using the Mothur pipeline (version 1.26.0) (Schloss *et al.*, 2009). Briefly, primers were trimmed from the sequences. Sequences were removed if they were low quality reads (Avg. Q < 25), were shorter than 200 nucleotides, contained homopolymers longer than six nucleotides (ex. AAAAAAA), or contained ambiguous bases ("N"s). In this procedure, one mismatch was allowed in the barcode, while two mismatches were allowed in the primers. Processed high quality sequences were aligned against the Silva database. Suspected chimeric sequences were detected using UCHIME (<6% sequences detected as chimeric) and removed (Edgar *et al.*, 2011). The remaining reads were pre-clustered and then clustered as previously described by Huse (Huse *et al.* 2007). Operational taxonomic units (OTUs) were defined as sharing > 97% sequence complete-linkage identity with the most abundant sequence forming the OTU seed. OTUs detected in less than three samples and fewer than three times were removed as possible artifacts. In order to accelerate this analysis procedure, an internally developed Perl script (version 5.16.0) was applied to automatize the operation (See **Code S1** in **Appendix**).

5.3.6 Phylogenetic Analysis of the Microbiome

Phylogenetic analysis was done using OTU and bacterial sequence information at the genus and species level, the main analysis procedure is demonstrated in the **Fig. 5.1.** OTUs were assigned to bacterial taxonomy according to the Ribosomal Database Project (RDP, http://rdp.cme.msu.edu/) Classifier (Cole *et al.*, 2005)., OTUs classified as the same bacterial genera were clustered together and used in statistical analysis at genus level.

To identify *Lactobacillus* sequences at species level, OTUs specifically identified as *Lactobacillus* genus in the RDP Classifier were retrieved and used to construct a new FASTA file. The new file was

uploaded into the BLASTn search engine (http://www. ncbi.nlm.nih.gov/BLAST/) and compared with records in the NCBI nucleotide collection database. Records with maximal similarity values greater than 97% were considered the same species. Sequences identified as the same species were clustered and used in statistical analysis. The reason BLASTn was used in species identification instead of RDP Classifier was that RDP Classifier can provide more precise sequence matching at genus level (Cole *et al.*, 2007); however, it doesn't provide information at species level.

To compare the composition of the OTUs, bacterial genera and species, abundance was converted from absolute to relative value. A spreadsheet was constructed using species, genera or OTU as variables and the relative abundance of OTUs or taxa as objects. The spreadsheet was imported into statistics software XLStat (Addinsoft Inc., Brooklyn, NY) to generate a principal component analysis (PCA) plots and an agglomerative hierarchical clustering (AHC) dendrogram.

SDI value is a simple estimation of the bacterial diversity in the microbiota. Shannon's Diversity Index (SDI) was calculated obeying the formula $H' = -\sum_{i=1}^{R} p_i \log p_i$, where p_i is the bacterial relative abundance belong to *i*th OTU in all the OTUs identified from the dataset.

5.3.7 Statistical Analysis

The quantities of OTUs, taxa and SDI values were analyzed using SAS (version 9.3, 2011, GraphPad Software, San Diego, CA). Descriptive statistics were calculated, and normality of sample was assessed. The data that did not fulfill the requirement of normality in distribution was converted to log scale before comparison in genus or species comparison. Standard ANOVA tables were constructed; variables were analyzed accordingly, using Student 2-tailed unpaired t-test or nonparametric Mann-Whitney test. A value of $p < 0.05$ was considered significant.

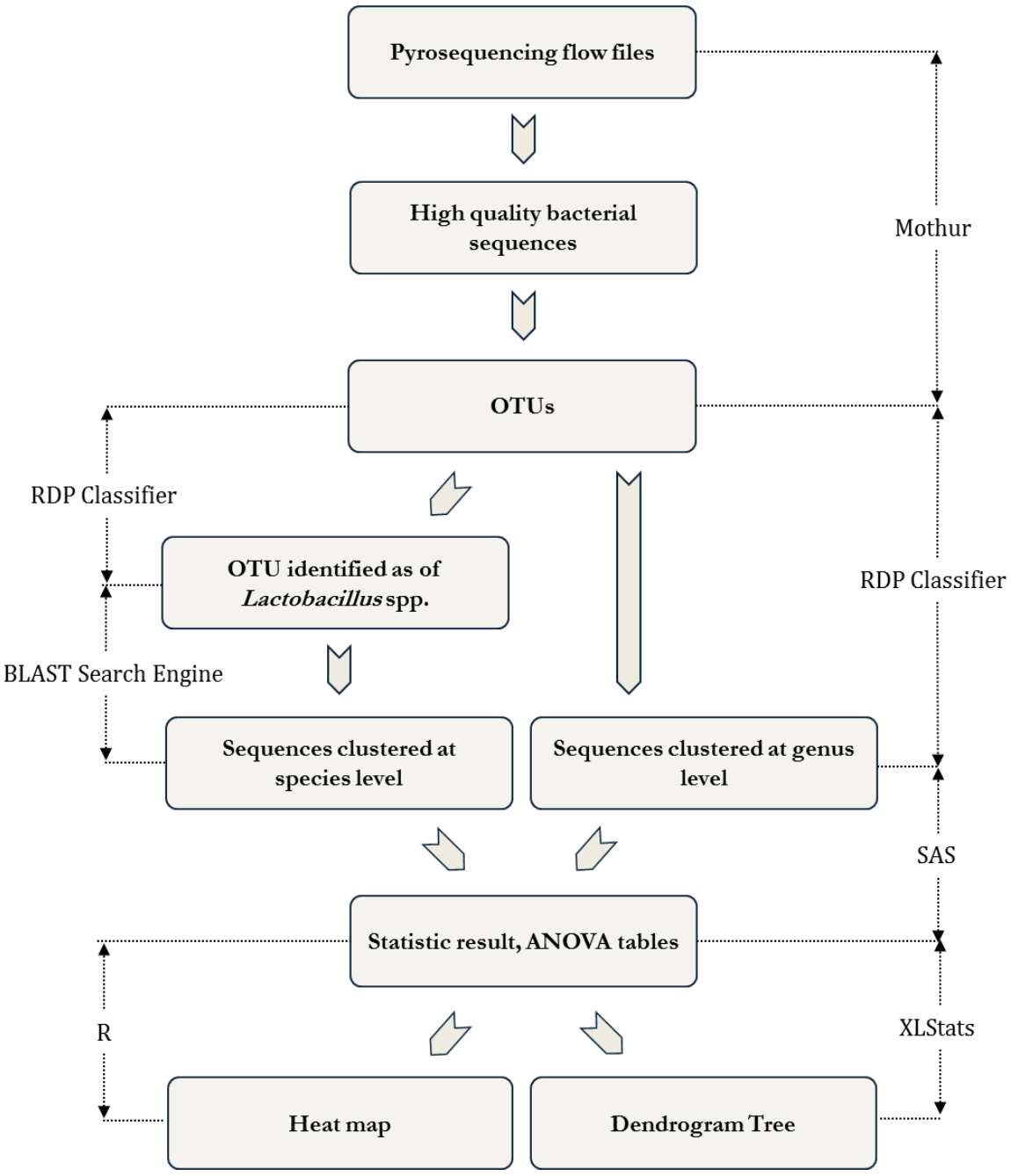


Fig. 5.1 The major steps of data processing in this study. The softwares used for sequence analysis indicated in the chart: Mothur (version 1.26.0), RDP Classifier (version 1.1), BLAST Search Engine (Database bacterial nucleotide collection), SAS (version 9.3), R (version 2.15), XLStats (version 2012). Internally developed script, "My454", was applied to control Mothur and RDP operation. Codes for analysis in different softwares are listed in **Appendix.**

5.4 Results

5.4.1 General Information of Pyrosequencing

Bacterial Tag-Encoded FLX 454-Pyroseqencing (bTEFAP) was used to analyze the bacterial community from ileal content samples collected from the 25 neonatal piglets. Following the conversion of flow files to sequences, application of trimming and screening protocols, a total of 112,002 bacterial 16S rDNA sequences with high quality were left (57% of total sequences obtained), an average of 4,500 sequences per sample. The mean length of these sequences was close to 400 bp with 75% of the sequences greater than 300 bp. A total of 3,199 OTUs were generated from bTEFAP in this study using the Mothur analysis pipeline.

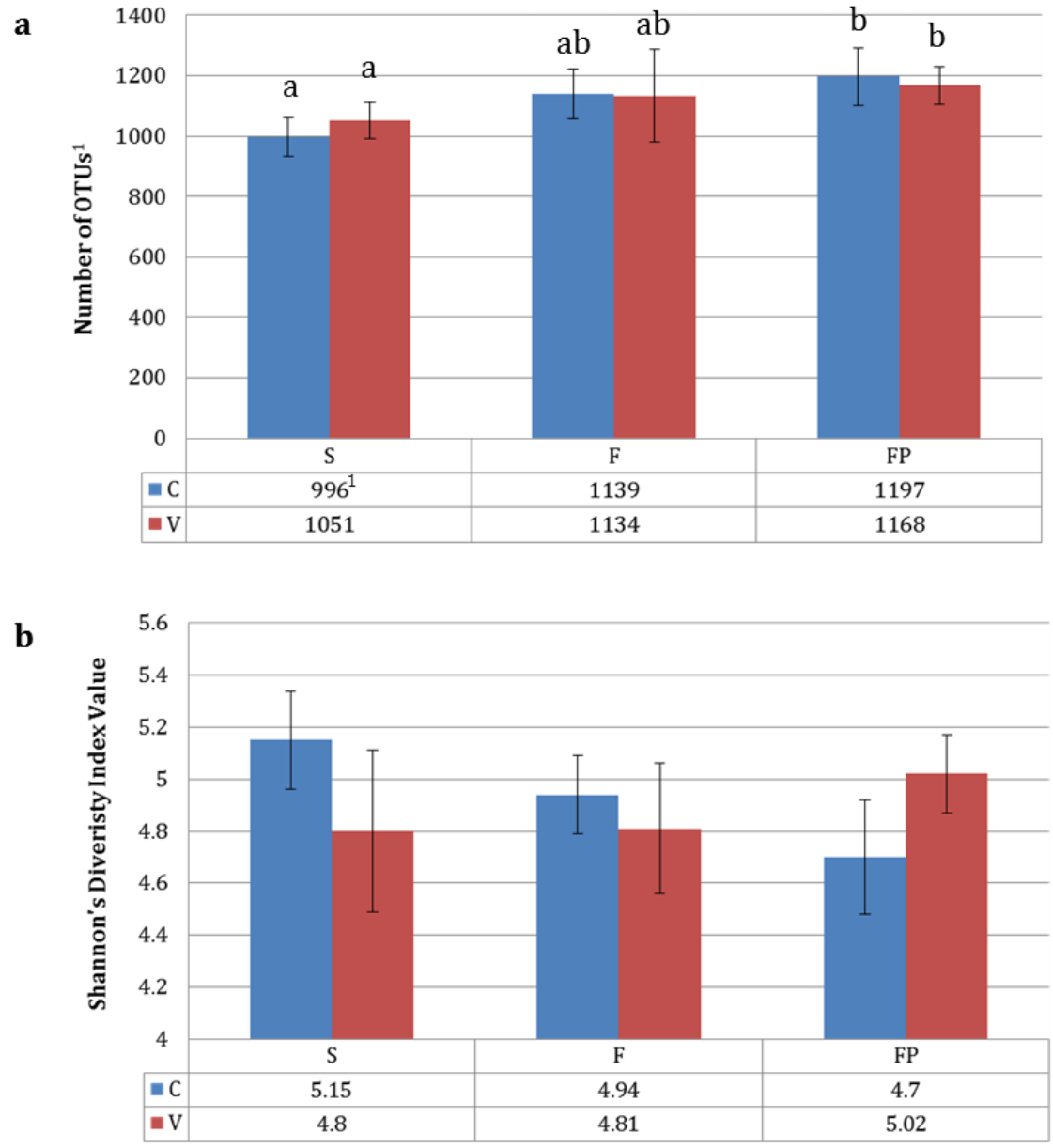
5.4.2 OTU-based Community Analysis

The number of sequences from each sample were normalized to 3177 (the smallest value of sequences from one single sample) and the bacterial richness as determined the total number of OTUs per normalized sample was compared within six groups (S:C, S:V, F:C, F:V, FP:C, FP:V), **Fig. 5.2a**. The bacterial richness from the treatment groups (animal model) was not significantly different ($p_{model} = 0.2093$). However, pair-wise comparison showed that there were significant differences between diet groups ($p_{diet} = 0.0456$), where the sow-reared (S) piglets had the lowest relative abundance of OTUs while the prebiotics (FP) piglets had the highest The relative abundance of OTUs detected in the formula (F) piglets was between that of the S piglets and the FP piglets.

Bacterial diversity (includes richness and evenness) was estimated by the Shannon Diversity Index (SDI), **Fig. 5.2b.** The SDI was calculated using the relative abundance of OTUs from each sample. Bacterial diversity was not significantly different among the six treatments ($p_{model} = 0.3986$). Pair-wise comparison of CD and VD piglets (S:C vs. S:V; F:C vs. F:V, FP:C vs. FP:V) showed no significant differences due to the route of delivery or diet ($p > 0.05$). Therefore, neither the

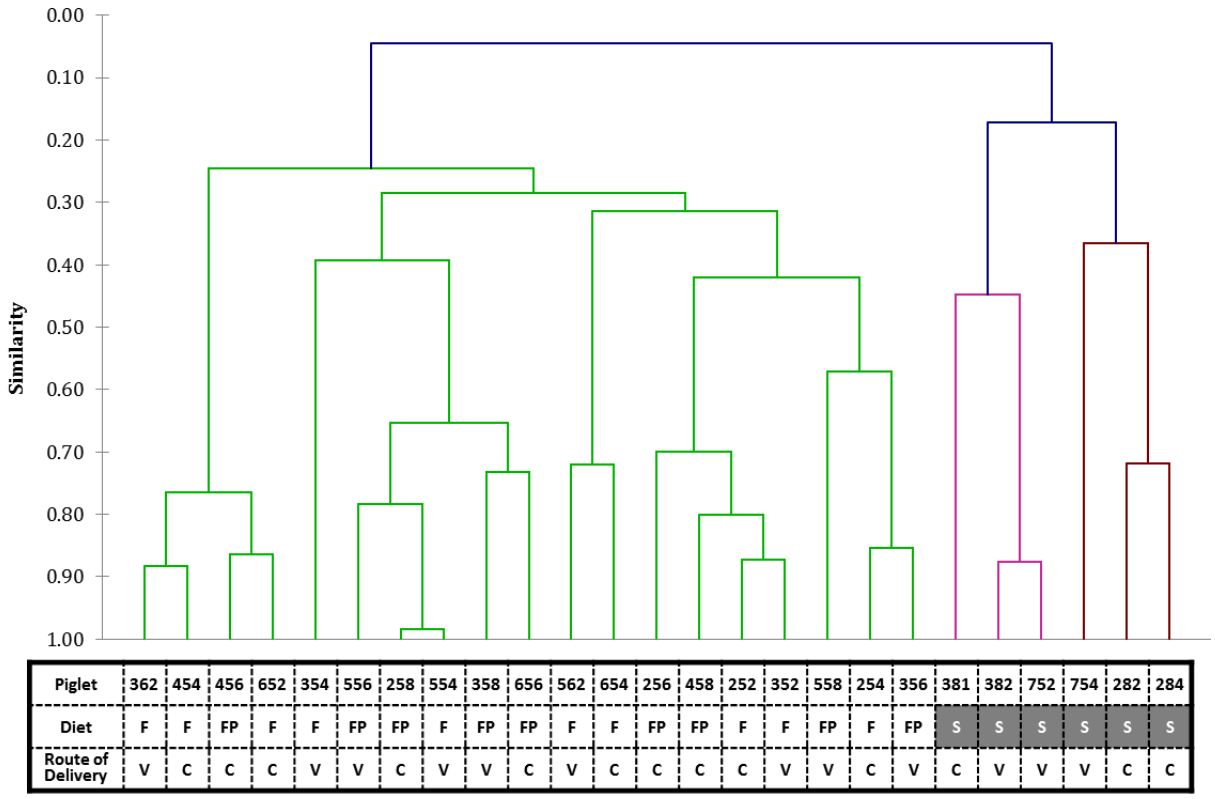
route of delivery nor diet had a significant impact on the overall bacterial diversity in piglet ileal microbiome.

To analyze and visualize the OTU composition of the piglet ileal microbiome, principal component analysis (PCA) plots (data now shown) and an agglomerative hierarchical clustering (AHC) dendrogram were generated (**Fig. 5.3**). A spreadsheet was constructed using OTU type as variables and the relative abundance of OTUs as objects, this spreadsheet represents the "fingerprints" of OTU compositions in different samples. The OTU compositions from the sow-reared piglets were clearly clustered in a single clade (similarity < 0.2) whereas the F and FP piglets were indistinguishable from each other. This shows that diet impacted the OTU composition while, route of delivery was not a shaping force of OTU composition.



[a, b, c] Values with the same letters are not different. For **Fig 5.1a** p_{model} = 0.2093, p_{diet} = 0.0456, $p_{delivery}$ = 0.8806, $p_{diet*delivery}$ = 0.7734. **Fig 5.1b** p_{model} = 0.3986, p_{diet} = 0.8567, $p_{delivery}$ = 0.0729, $p_{diet*delivery}$ = 0.1467.

Fig. 5.2 The average number of OTUs (**a**) and SDI (**b**) values generated from all groups (S:C, S:V, F:C, F:V, FP:C, FP:V). There were no significant differences in the relative OTU abundance or bacterial diversity (SDI value) between all groups. However, pair-wised comparison showed that the quantity of OTU is highest in FP groups and lowest in sow-reared groups.



Piglet	362	454	456	652	354	556	258	554	358	656	562	654	256	458	252	352	558	254	356	381	382	752	754	282	284
Diet	F	F	FP	F	F	FP	FP	F	FP	FP	F	F	FP	FP	F	F	FP	F	FP	S	S	S	S	S	S
Route of Delivery	V	C	C	C	V	V	C	V	V	C	V	C	C	C	C	V	V	C	V	C	V	V	V	C	C

Fig. 5.3 AHC dendrogram tree generated based on relative OTU abundance from the 25 piglet ileal samples. The diet and route of delivery are indicated in the table underneath. This figure was constructed using XLStat. OTU types were used as variables and their relative abundance in the microbiota were used as objects. The sow-reared piglets (S) were clearly clustered; however, the formula (F) and prebiotic supplemented formula (FP) groups were mixed together. Route of delivery had no impact on clustering.

5.4.3 Phylotype Analysis

Following OTU-based analysis methods, the sequences were classified based on a reference database (RDP classifier) to obtain genus-level identification. A total of 99 bacterial genera were identified from the 3199 total OTUs that were identified. The top ten genera in terms of relative abundance are listed in **Table 5.1**. These genera constitute more than 80% population of the microbiome. *Lactobacillus* was the most abundant genus representing an average of 25% of the total bacterial population in the ileal samples. Statistical analysis revealed that there were no significant differences among the relative abundance of the six groups of samples ($p > 0.05$). The composition of ileal microbiome is comparatively stable among different treatments. This result is consistent with previous OTU-based analysis.

However, pair-wise comparison showed that the relative abundance value of the *Lactobacillus* genus in the piglets delivered by cesarean section (CD) were significantly higher than from vaginally delivered (VD) piglets (28.00±5.90 vs. 23.44±6.39, $p_{Delivery}=0.0491$). Therefore the route of delivery had an impact on the relative abundance of overall *Lactobacillus* spp., which was generally higher in VD ileal samples as compared to CD ileal samples.

The relative abundance of *Enterobacter* spp. was different between the diet groups (S, F, FP). The prebiotic supplemented formula (FP) piglets had the highest abundance of *Enterobacter* spp., followed by the formula (F) group while the sow-reared (S) piglets had the lowest value (3.22±1.90 vs. 4.72±2.04 vs. 6.89±3.22 $p_{diet}=0.0338$).

Table 5.1 Relative abundance of top ten genera identified in ileal samples by bTEFAP

Genus	S:C	S:V	F:C	F:V	FP:C	FP:V	p_{model}	$p_{Delivery}$	p_{diet}
Lactobacillus[1]	32.32±6.86[2]	22.32±2.43	26.96±6.27	25.70±6.85	26.47±4.85	21.47±8.25	0.3226	0.0491[3]	0.5659
Streptococcus	11.49±1.98	20.71±10.3	21.23±4.25	20.77±3.98	20.00±3.93	18.04±4.55	0.1631	0.2854	0.1958
Actinobacillus	7.46±3.67	8.99±3.38	6.02±1.41	7.55±3.64	7.97±1.31	10.34±3.86	0.4013	0.1479	0.2327
Turicibacter	6.42±1.81	6.05±3.46	7.62±4.88	8.02±12.47	8.39±5.88	5.31±4.16	0.9842	0.7960	0.9002
Enterobacter	3.88±2.73	2.57±0.5a	4.69±1.23	4.75±2.8	6.01±3.13	7.99±3.43	0.1469	0.8215	0.0338[4]
Pasteurella	5.33±1.78	5.52±0.38	5.5±2.99	3.4±0.28	3.53±2.04	6.91±1.94	0.3672	0.6553	0.7285
Escherichia	2.98±0.86	4.22±2.33	3.74±1.24	4.56±2.14	4.5±1.74	4.36±0.88	0.7847	0.3576	0.6421
Veillonella	2.74±0.71	3.02±1.09	2.82±1.25	3.58±1.27	3.86±1.76	3.97±1.28	0.6554	0.4915	0.3178
Lactococcus	2.6±0.48	2.92±0.11	3.89±0.77	3.4±1.38	2.72±0.72	3.42±0.70	0.2651	0.6183	0.1378
Haemophilus	2.7±1.6	2.32±0.59	2.07±0.7	2.43±0.27	2.43±0.68	2.69±0.67	0.8493	0.8014	0.6580

[1]Genera identified using RDP classifier
[2]Data expressed as mean % ± SD of the total bacterial population.
[3]Relative abundance of *Lactobacillus* spp. was different in CD and VD groups (28.00±5.90 vs. 23.44±6.39).
[4]Relative abundance of *Enterobacter* spp. was different among diets groups (FP: 3.22±1.90 vs. F: 4.72±2.04 vs. S: 6.89±3.22).

5.4.4 *Lactobacillus* Identification at Species Level

Due to the importance and relatively high concentration of *Lactobacillus* spp., a sub-genus analysis was done to determine the impact of different factors on the *Lactobacillus* community. A total of 515 OTUs containing 29,007 *Lactobacillus* sequences were retrieved to generate a new FASTA file, the representative sequences were classified using the BLASTn tool and identified at the species level.

Thirty *Lactobacillus* spp. were identified in the 25 ileal samples, ten of which were above 1% total abundance: *L. amylovorus, L. reuteri, L. plantarum, L. delbrueckii, L. helveticus, L. agilis, L. salivarius, L.mucosae, L.jonsonii* and *L.crispatus*, **Table 5.2**. Six *Lactobacillus* spp. were identified in the sow-reared (S) piglets and the formula-fed (F, FP) piglets, and only four species were found in all groups. The relative abundance varied from less than 1% to over 35% of total lactobacilli.

Statistical analysis showed that the relative abundance of *L. plantarum, L.delbrueckii, L.helveticus, L.agilis, L. johnsonii, L. crispatus, L. mucosae* and *L. salivarius* were significantly different in the sow-reared (S) piglets than the formula-fed piglets (F, FP) ($p < 0.05$). *L. salivarius, L. mucosae, L. johnsonii, L. crispatus* and *L. reuteri* were significant different in CD and VD piglets.

Since the route of delivery was observed to have an impact on the relative abundance of *Lactobacillus* genus in previous analyses, an AHC dendrogram tree was constructed to examine whether the route of delivery also impacted the composition at the species level (**Fig 5.4**). A spreadsheet was constructed using *Lactobacillus* spp. as variables and the relative abundance of sequences as objects. There was no obvious clustering between the CD and VD groups, therefore, the way route of delivery impact the composition of overall *Lactobacillus* genus is not predictable.

Table 5.2 Relative abundance of top ten *Lactobacillus* spp. identified in ileal samples by bTEFAP

Species	S:C	S:V	F:C	F:V	FP:C	FP:V	p_{model}	$p_{Delivery}$	p_{diet}
L. amylovorus[1]	0[2]	2.03±1.21	36.35±18.11	18.71±14.40	51.44+22.31	17.45±12.14	0.2290	0.0831	0.1012
L. reuteri	14.24±7.12	20.41±15.15	10.45±7.13	11.04±4.23	25.11±12.32	4.72±2.2	0.7801	0.0392	0.4541
L. plantarum	0	0	10.56±4.3	28.71±15.51	2.31±1.17	23.26±9.79	0.3615	0.5164	<0.05[3]
L. delbrueckii	1.90±0.25	0	26.14±12.25	23.36±3.15	5.36±4.1	8.67±4.65	0.3656	0.7309	<0.05
L. helveticus	0	0	5.01±4.12	3.45±2.1	2.17±1.65	0	0.1215	0.1330	<0.05
L. agilis	0	0	0	8.75±6.63	0	17.66±15.23	0.8471	0.6060	<0.05
L. salivarius	17.10±14.44	23.13±10.4	0	3.12±2.77	0	0	0.6013	<0.05	<0.05
L. mucosae	8.01±4.56	2.15±1.53	1.05±0.56	0	0	0	0.3642	0.0127	<0.05
L. johnsonii	18.72±5.55	36.52±15.59	36±27.71	0	0.91±0.54	0	0.6445	<0.05	0.0411
L. crispatus	15.43±6.47	1.11±0.45	0	0	0	0	0.7132	0.0332	<0.05

[1]Genera identified using BLASTn with NCBI nucleotide collection database.
[2]Data expressed as mean % ± SD of the total *Lactobacillus* population. The value < 0.01% was considered as 0.
[3]p value <0.01% were shown as <0.05%.`

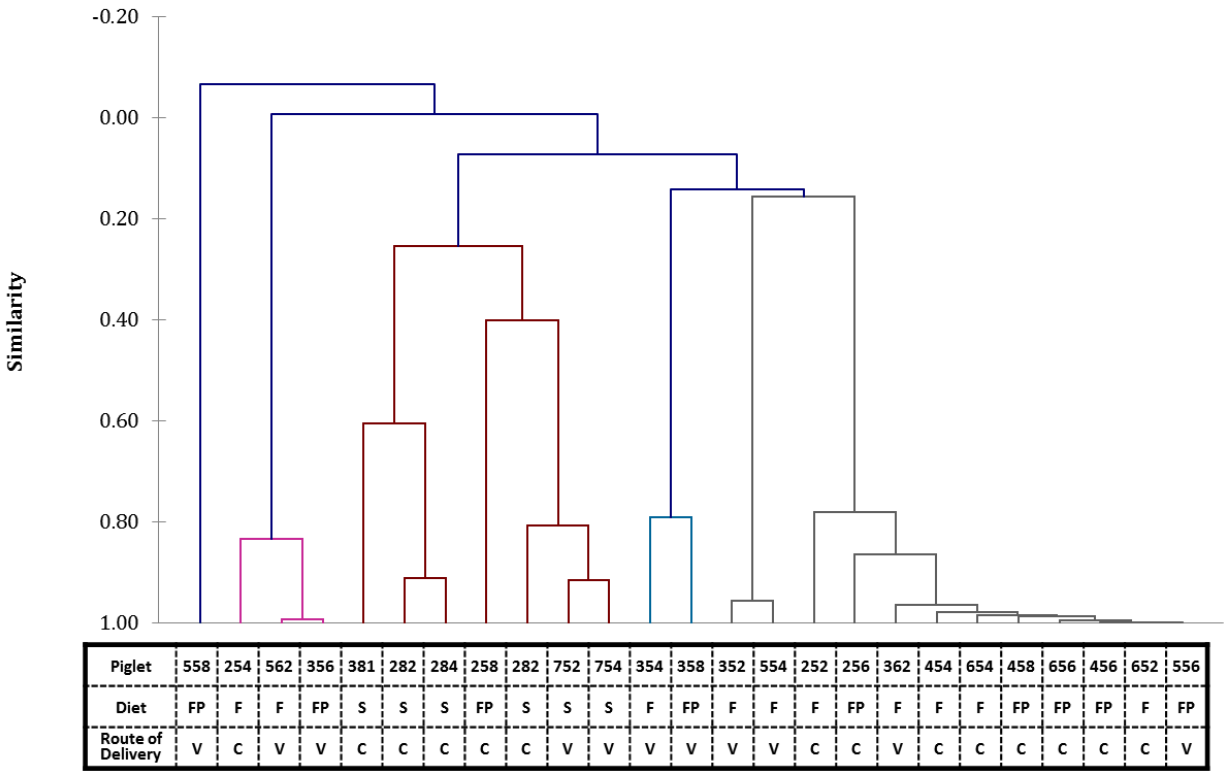


Fig. 5.4 AHC dendrogram tree generated based on the relative abundance of the top ten *Lactobacillus* spp. The diet and route of delivery are indicated in the table underneath. This figure was constructed using *Lactobacillus* spp. as variables and relative abundance of sequences as objects. There was no clustering of samples.

5.4.5 Species Identification by Culture-dependent Method

Culture-dependent methods were done to verify the major *Lactobacillus* spp. identified by pyrosequencing. PCR using primers Gray28F and Gray519R generated 510 bp amplicon covering V1 to V3 region of the *Lactobacillus* 16S rDNA was done, and Sanger sequencing was performed to identify the species. Sequences generated by Sanger sequencing are longer than the average length of the sequences from pyrosequencing (~400 bp). The extra base pairs offer a higher resolution to identify very closely related species.

A total of 226 isolates were sequenced and identified as *Lactobacillus* spp. (**Table 5.3**). The relative abundance of species which was obtained from culture-dependent methods are very close to pyrosequencing result, except for that *L. equi* and *L. pentosus* were only detected by culture-dependent methods. In addition, *L. paraplantarum* was able to be distinguished from *L. plantarum*. These two species are very closely related. Part of the sequences of "*L. plantarum* obtained from pyrosequencing may actually belong to *L. paraplantarum*, however they were not distinguishable from *L. plantarum* perhaps due to the shorter sequence length.

Table 5.3 Total *Lactobacillus* spp. identified from the ileal isolates as determined by sequencing the 510 bp 16S rDNA V1-V3 region.

	S:C	S:V	F:C	F:V	FP:C	FP:V
Total Isolates	30[1]	36	41	45	32	45
L. amylovorus	19.5%[2]	8.9%	15.6%	16.7%	6.7%	0
L. reuteri	26.8%	28.9%	34.4%	11.9%	3.3%	2.8%
L. plantarum	29.3%	22.2%	12.5%	9.5%	6.7%	11.1%
L. paraplantarum[3]	0	0	0	4.8%	0	0
L. delbrueckii	9.8%	8.9%	3.1%	21.4%	36.7%	19.4%
L. helveticus	4.9%	13.3%	18.8%	4.8%	0	0
L. equi[3]	0	0	0	0	6.7%	0
L. salivarius	2.4%	2.2%	0	0	6.7%	16.7%
L. mucosae	2.4%	2.2%	9.4%	21.4%	0	5.6%
L. johnsonii	4.9%	13.3%	6.3%	9.5%	16.7%	41.7%
L. crispatus	0	0	0	0	10.0%	2.8%
L. pentosus[3]	0	0	0	0	6.7%	0

[1]Number of isolates identified as *Lactobacillus* spp.
[2]Data is shown as % of isolates identified as *Lactobacillus* spp. Due to the small number of samples (<1 isolate per piglet sample), statistical analysis was not applied.
[3]Species only detected by culture-dependent methods

5.5 Discussion

The aim of this study is to investigate the ileal microbiome of neonatal piglets by 16S rDNA-based bTEFAP. Additionally, we aimed to determine the impact of route of delivery and nutrition (diet) on bacterial abundance and diversity, and more specifically, the lactobacilli composition at the genus and species level.

In this study, we compared the gut microbiome of at the genus and species level. Two separate approaches were applied to analyze the data from pyrosequencing: the OTU-based methods and phylotype-based (taxonomic) methods. The OTU-based methods allow researchers to compare the microbial diversity of different environments when the vast majority of microbial taxa are unknown (Bohannan *et al.*, 2003). OTU-based methods are now widely used for classification of high-throughput sequences (Cheng *et al.*, 2012). Sequences are assigned an OTU based on the similarity of the sequences and not by comparison with a reference database. However, OTU-based methods have two main limitations: [1] it does not account for differences in dissimilarity percentage cutoffs for sequences of the same species [2] cannot distinguish closely related species. Phylotype-based methods rely on classification using a reference database. Each sequence is assigned to a bacterial species or genus. Phylotype-based methods allow researchers to focus their studies on specific bacterial taxa.

The impact of diet was analyzed by both OTU and phylotype-based methods. The bacterial diversity and bacterial richness was significantly different between the sow-reared piglets and formula-fed piglets (F, FP). Additionally, a Unifrac-based dendrogram plot (ACH) showed a significant overlapping of F and FP. Interestingly, the sow-reared piglets are clustered in a unique clade separated from the F and FP piglets. This observation indicates that the life-style (fed by mother, co-housed in one cage and with the sow) is a large contributor to the development of the ileal microbiome and could be the most important factor in early lactobacilli colonization rather than the diet alone. While the total number of OTUs was lower in the sow-reared piglets the

standard deviations were larger which may mean the sow-reared piglets had a more diverse ileal microbiome. During this study, the sow-reared piglets had exposure to more diverse environmental bacteria, including the sow's skin, fecal microbiota (sow and sibling piglets) and bacteria inhabiting the pens. Additionally, the feeding schedule, "life styles" and psychological stress is not similar between the sow-reared and formula-fed piglets. The formula-fed piglets were individually housed during the study and were not exposed to the same environmental bacteria as the sow-reared piglets. In fact, care was taken to avoid the introduction of environmental bacteria to the formula-fed piglets by limiting human contact.

We also showed that *Lactobacillus* spp. were the most abundant species in all three groups which is consistent with other studies (Konstantinov *et al.*, 2004; Fuller R, 1989). We didn't observe some species which were found in other researches (Harmsen *et al.*, 2000). It was reported that formula-fed infants develop a more diverse microbiota which includes bifidobacteria, bacteroides, enterobacteria, enterococci, and clostridia (Herfel *et al.*, 2009; Stark *et al.*, 1982). This is not consistent with our result, our data showed that the composition of bacterial genera is relatively stable at the genus level and does not differ greatly due to diet. Unfortunately, more information is needed to have detailed analysis of these genera at species level and we did not study them in depth.

To reveal the difference of the ileal microbiome at the species level, we applied sub-genus analysis on *Lactobacillus* genus. Data from culture-dependent and culture-independent methods were analyzed and compared. Sanger sequencing results provided more accurate species identification, suggesting that the traditional sequence technology is still useful before the new pyrosequencing protocol is developed to generate longer sequences.

Human milk is considered the "gold standard" for infant nutrition, one of the biggest differences between human milk and formula milk is the oligosaccharide composition, specifically Human Milk Oligosaccharides (HMO). In this research, PDX and GOS mix was added into formula milk as HMO alternative. However, we did not observe huge differences between sample from formula fed and

formula-prebiotic fed piglets. This is not consistent with previous research (Kunz *et al.*, 2000; Newberg and Neubauer, 1995). We believe the result from the powerful tool, pyrosequencing, is more likely to have less bias or higher discrimination than the studies led ten years ago.

In conclusion, the bacterial community was greatly impacted by dietary and nutrition factors. Route of delivery did not impacted the overall bacterial diversity, however, it changed the relative abundance of *Lactobacillus* spp. at genus and species level, the direction of these changes were non-predictable (no patterns). Dietary factors are crucial shaping force of microbiome, it impact the bacterial diversity and relative abundance of *Lactobacillus* spp. at species level. These conclusions will help us understand the complexity of GI microbiota, and in turn, help us find the approach of impact its composition.

CHAPTER 6. OVERALL CONCLUSIONS AND FUTURE DIRECTION

The human GI microbiota is influenced by several factors. In these three studies, we have applied two traditional culture-dependent methods, including standard plating and carbohydrate utilization; and four culture-independent methods, including T-RFLP, ERIC-PCR and qPCR, and a recently developed molecular biological tool, bTEAFP. These methods were used to investigate the composition of the microbiome in the piglet GI tract, specifically the lactobacilli community. These studies resulted in several main conclusions: we observed that the age of our piglets and the route of delivery had obvious impacts on the composition of the lactobacilli community. We found that diets could be the most crucial force shaping the bacterial microbiota at both the genus and species level. Additionally, we observed other factors; including nursing style, weaning and environmental conditions could alter the composition of the lactobacilli composition for a comparatively short time. These conclusions allow for a better understanding of the complexity of the GI microbiota and to expand our knowledge of the relationship of microbial activities and host health. These efforts will in turn help us develop new approaches to alter the composition of the GI microbiota, and in turn offer health benefits to the human body.

We also encountered several issues in these studies, and propose to apply new protocols in the future research.

1. The impact of the diet made these studies challenging when analyzing the factors in this animal model. We observed that diets and nursing style (sow-reared vs. formula-fed) had tremendously huge impacts on the composition of bacterial microbiota. This single factor is overwhelmingly more influential than the other factors tested (age, route of delivery, prebiotic supplementation). Piglets were exposed to different sources of bacteria (environment, feces) during the sow nursing process, which was not considered as contributing to the dietary factors. We suggest using a sow milk collection tool instead of co-housing the piglets with the sows. Housing all the

piglets in the same environment or in a sterile pen could help to eliminate the impact from the environmental differences.

2. Other oligosaccharides which are more similar to HMOs in molecular structure could be investigated in future studies. Though GOS and PDX are widely used as HMO alternatives in animal studies, this choice is mainly due to the cost and technical difficulty of large scale synthesis of HMOs. As the development of new technologies progress experiments directly using HMO will become more feasible and more affordable.
3. Continuing improvements in current pyrosequencing analysis tools and the development of new culture-independent tools to facilitate better characterizations of microbial communities. The introduction of pyrosequencing in microbiome research has provided unprecedented depth in analyzing microbial communities. These methods provide hundreds of times more information than the traditional (culture-dependent) methods. However, the interpretation of bioinformatics data sets is computationally complex and is still challenging. Current phylogenetic analysis based on OTUs and sequences are still laborious, expensive and dependent on well trained personnel. With further development phylogenetic characterization can be improved by the application of optimized computer software programs. Computer programs could be tailored to streamline the data analysis process to limit the need for specialized bioinformaticians and pave the way for more integration of these powerful tools in research. Additionally, the simplification of data analysis could decrease the cost of these tools and increase research.

In conclusion, we made several important observations in these three studies. Future work will focus on improvement of the current animal mode (neonatal piglet model), development of a sow-reared reference group that can be used for microbial studies, inclusion of more applicable prebiotics (HMOs) and improving the bioinformatic data analysis pipeline. We believe that microbiome research will continue to be a hot topic in both food microbiology and nutritional sciences research. Most importantly, the efforts made today can have long-lasting impacts on infant nutrition study in the future.

REFERENCES

Ashley C, Johnston WH, Harris CL, Stolz SI, Wampler JL & Berseth CL (2012) Growth and tolerance of infants fed formula supplemented with polydextrose (PDX) and/or galactooligosaccharides (GOS): double-blind, randomized, controlled trial Nutr J **11:** 38.

Bailey MT, Dowd SE, Parry NM, Galley JD, Schauer DB & Lyte M (2010) Stressor exposure disrupts commensal microbial populations in the intestines and leads to increased colonization by Citrobacter rodentium Infect Immun **78:** 1509-1519.

Bartosch S, Fite A, Macfarlane GT & McMurdo ME (2004) Characterization of bacterial communities in feces from healthy elderly volunteers and hospitalized elderly patients by using real-time PCR and effects of antibiotic treatment on the fecal microbiota Appl Environ Microbiol **70:** 3575-3581.

Berthier F & Ehrlich SD (1998) Rapid species identification within two groups of closely related lactobacilli using PCR primers that target the 16S/23S rRNA spacer region FEMS Microbiol Lett **161:** 97-106.

Bezkorovainy A (2001) Probiotics: determinants of survival and growth in the gut Am J Clin Nutr **73:** 399S-405S.

Biasucci G, Benenati B, Morelli L, Bessi E & Boehm G (2008) Cesarean delivery may affect the early biodiversity of intestinal bacteria J Nutr **138:** 1796S-1800S.

Blaiotta G, Fusco V, Ercolini D, Aponte M, Pepe O &Villani F (2008) *Lactobacillus* strain diversity based on partial hsp60 gene sequences and design of PCR-restriction fragment length polymorphism assays for species identification and differentiation Appl Environ Microbiol **74:** 208-215.

Bohannan BJ & Hughes J (2003) New approaches to analyzing microbial biodiversity data Curr Opin Microbiol **6:** 282-287.

Boggy GJ & Woolf PJ (2010) A mechanistic model of PCR for accurate quantification of quantitative PCR data PLoS One **5:** e12355.

Brody JR & Kern SE (2004) History and principles of conductive media for standard DNA electrophoresis Anal Biochem **333:** 1-13.

Brugger SD, Baumberger C, Jost M, Jenni W, Brugger U & Muhlemann K (2012) Automated counting of bacterial colony forming units on agar plates PLoS One **7:** e33695.

Callaway TR, Dowd SE, Wolcott RD, Sun Y, McReynolds JL, Edrington TS, Byrd JA, Anderson RC, Krueger N &Nisbet DJ (2009) Evaluation of the bacterial diversity in cecal contents of laying hens fed various molting diets by using bacterial tag-encoded FLX ampliconpyrosequencingPoultSci **88:** 298-302.

Calmin G, Lefort F &Belbahri L (2008) Multi-loci sequence typing (MLST) for two lacto-acid bacteria (LAB) species: Pediococcusparvulus and P. damnosusMolBiotechnol **40:** 170-179.

Castillo M, Martin-Orue SM, Nofrarias M, Manzanilla EG & Gasa J (2007) Changes in caecal microbiota and mucosal morphology of weaned pigs Vet Microbiol **124:** 239-247.

Chou YK, Chen HJ, Shaio MF & Kuo YM (2004) In vitro antibiotic susceptibility of lactobacilli isolated from commercial products containing active lactobacilli Acta Paediatr Taiwan **45:** 141-144.

Christensen JE, Reynolds CE, Shukla SK & Reed KD (2004) Rapid molecular diagnosis of *Lactobacillus* bacteremia by terminal restriction fragment length polymorphism analysis of the 16S rRNA gene Clin Med Res **2:** 37-45.

Claesson MJ, Li Y, Leahy S *et al.* (2006) Multireplicon genome architecture of *Lactobacillus* salivarius. Proc Natl Acad Sci U S A **103:** 6718-6723.

Claisse O, Renouf V &Lonvaud-FunelA (2007) Differentiation of wine lactic acid bacteria species based on RFLP analysis of a partial sequence of rpoB gene J Microbiol Methods **69:** 387-390.

Clarke SC (2005) Pyrosequencing: nucleotide sequencing technology with bacterial genotyping applications Expert Rev MolDiagn **5:** 947-953.

Cole JR, Chai B, Farris RJ, Wang Q, Kulam SA, McGarrell DM, Garrity GM & Tiedje JM (2005) The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis Nucleic Acids Res **33:** D294-6.

Cole JR, Chai B, Farris RJ, Wang Q, Kulam-Syed-Mohideen AS, McGarrell DM, Bandela AM, Cardenas E, Garrity GM & Tiedje JM (2007) The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data Nucleic Acids Res **35:** D169-72.

Collado MC, Cernada M, Bauerl C, Vento M & Perez-Martinez G (2012) Microbial ecology and host-microbiota interactions during early life stages Gut Microbes **3:** 352-365.

Collado MC &Sanz Y (2007) Quantification of mucosa-adhered microbiota of lambs and calves by the use of culture methods and fluorescent in situ hybridization coupled with flow cytometry techniques Vet Microbiol **121:** 299-306.

Coolen MJ, Post E, Davis CC & Forney LJ (2005) Characterization of microbial communities found in the human vagina by analysis of terminal restriction fragment length polymorphisms of 16S rRNA genes Appl Environ Microbiol **71:** 8729-8737.

Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI & Knight R (2009) Bacterial community variation in human body habitats across space and time Science **326:** 1694-1697.

Culman SW, Gauch HG, Blackwood CB & Thies JE (2008) Analysis of T-RFLP data using analysis of variance and ordination methods: a comparative study J Microbiol Methods **75:** 55-63.

Cummings JH & Macfarlane GT (1997) Role of intestinal bacteria in nutrient metabolism Clinical Nutrition **16:** 3-11.

Davis E, Rehberger J, King M, Brown DC, Maxwell CV &Rehberger T (2010) Characterization of gastrointestinal microbial and immune populations post-weaning in conventionally-reared and segregated early weaned pigs Livestock Science **133:** 92-94.

Dellaglio F, Felis GE, Castioni A, Torriani S & Germond JE (2005) *Lactobacillus delbrueckii* subsp. indicus subsp. nov., isolated from Indian dairy products Int J Syst Evol Microbiol **55:** 401-404.

Dethlefsen L, Eckburg PB, Bik EM & Relman DA (2006) Assembly of the human intestinal microbiota Trends Ecol Evol **21:** 517-523.

de Las Rivas B, Marcobal A & Munoz R (2006) Development of a multilocus sequence typing method for analysis of *Lactobacillus* plantarum strains Microbiology **152:** 85-93.

Dicksved J, Floistrup H, Bergstrom A *et al.* (2007) Molecular fingerprinting of the fecal microbiota of children raised according to different lifestyles Appl Environ Microbiol **73:** 2284-2289.

Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N & Knight R (2010) Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns ProcNatlAcadSci U S A **107:** 11971-11975.

Dowd SE, Wolcott RD, Sun Y, McKeehan T, Smith E & Rhoads D (2008) Polymicrobial nature of chronic diabetic foot ulcer biofilm infections determined using bacterial tag encoded FLX ampliconpyrosequencing (bTEFAP) PLoS One **3:** e3326.

Dunbar J, Ticknor LO & Kuske CR (2001) Phylogenetic specificity and reproducibility and new method for analysis of terminal restriction fragment profiles of 16S rRNA genes from bacterial communities Appl Environ Microbiol **67:** 190-197.

Du Toit M, Dicks LM & Holzapfel WH (2003) Identification of heterofermentative lactobacilli isolated from pig faeces by numerical analysis of total soluble cell protein patterns and RAPD-PCR Lett Appl Microbiol **37:** 12-16.

Edgar RC, Haas BJ, Clemente JC, Quince C & Knight R (2011) UCHIME improves sensitivity and speed of chimera detection Bioinformatics **27:** 2194-2200.

Eggesbo M, Botten G, Stigum H, Nafstad P & Magnus P (2003) Is delivery by cesarean section a risk factor for food allergy? J Allergy Clin Immunol **112:** 420-426.

Elli M, Zink R, Reniero R & Morelli L (1999) Growth requirements of *Lactobacillus* johnsonii in skim and UHT milk Int Dairy J **9:** 507-513.

Endo A, Futagawa-Endo Y & Dicks LM (2009) *Lactobacillus* and Bifidobacterium diversity in horse feces, revealed by PCR-DGGE CurrMicrobiol **59:** 651-655.

Engelbrektson AL, Korzenik JR, Sanders ME, Clement BG, Leyer G, Klaenhammer TR & Kitts CL (2006) Analysis of treatment effects on the microbial ecology of the human intestine FEMS Microbiol Ecol **57:** 239-250.

Englekirk, P.G., J. Duben-Englekirk, and V.R. Dowell (1992) Processing anaerobic isolates, p. 125-147. In S. Hoffman (ed.), Clinical anaerobic bacteriology. Star Publishing Company, Belmont, CA, U.S.A.

Euzeby JP (1997) List of Bacterial Names with Standing in Nomenclature: a folder available on the Internet Int J SystBacteriol **47:** 590-592.

Fernandez-Raudales D, Hoeflinger JL, Bringe NA, Cox SB, Dowd SE, Miller MJ & Gonzalez de Mejia E (2012) Consumption of different soymilk formulations differentially affects the gut microbiomes of overweight and obese men Gut Microbes **3:** 490-500.

Fogel GB, Collins CR, Li J & Brunk CF (1999) Prokaryotic Genome Size and SSU rDNA Copy Number: Estimation of Microbial Relative Abundance from a Mixed Population Microb Ecol **38:** 93-113.

Francl AL, Hoeflinger JL & Miller MJ (2012) Identification of lactose phosphotransferase systems in Lactobacillus gasseri ATCC 33323 required for lactose utilization Microbiology **158:** 944-952.

Francl AL, Thongaram T & Miller MJ (2010) The PTS transporters of Lactobacillus gasseri ATCC 33323 BMC Microbiol **10:** 77.

Frank DN, St Amand AL, Feldman RA, Boedeker EC, Harpaz N & Pace NR (2007) Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases ProcNatlAcadSci U S A **104:** 13780-13785.

Fryklund B, Tullus K, Berglund B &Burman LG (1992) Importance of the environment and the faecal flora of infants, nursing staff and parents as sources of gram-negative bacteria colonizing newborns in three neonatal wards Infection **20:** 253-257.

Fujimoto J, Matsuki T, Sasamoto M, Tomii Y & Watanabe K (2008) Identification and quantification of *Lactobacillus* casei strain Shirota in human feces with strain-specific primers derived from randomly amplified polymorphic DNA Int J Food Microbiol **126:** 210-215.

Fuller R (1989) Probiotics in man and animals J Appl Bacteriol **66:** 365-378.

Fu-gui *et al.*, (2007) Effect of Acanthopanax Senticosus Extracts on Gut Microflora in Early-Weaned Piglets.

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

Giraffa G, De Vecchi P & Rossetti L (1998) Note: identification of *Lactobacillus* delbrueckii subspecies bulgaricus and subspecies lactis dairy isolates by amplified rDNA restriction analysis J ApplMicrobiol **85:** 918-924.

Goh SH, Potter S, Wood JO, Hemmingsen SM, Reynolds RP & Chow AW (1996) HSP60 gene sequences as universal targets for microbial species identification: studies with coagulase-negative staphylococci J Clin Microbiol **34:** 818-823.

Goldmann DA (1988) The bacterial flora of neonates in intensive care-monitoring and manipulation J Hosp Infect **11 Suppl A:** 340-351.

Goodman AL, Kallstrom G, Faith JJ, Reyes A, Moore A, Dantas G & Gordon JI (2011) Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice Proceedings of the National Academy of Sciences.

Gronlund MM, Lehtonen OP, Eerola E &Kero P (1999) Fecalmicroflora in healthy infants born by different methods of delivery: permanent changes in intestinal flora after cesarean delivery J PediatrGastroenterolNutr **28:** 19-25.

Guan LL, Hagen KE, Tannock GW, Korver DR, Fasenko GM & Allison GE (2003) Detection and identification of *Lactobacillus* species in crops of broilers of different ages by using PCR-denaturing gradient gel electrophoresis and amplified ribosomal DNA restriction analysis Appl Environ Microbiol **69:** 6750-6757.

Hall MA, Cole CB, Smith SL, Fuller R & Rolles CJ (1990) Factors influencing the presence of faecal lactobacilli in early infancy Arch Dis Child **65:** 185-188.

Hamady M, Walker JJ, Harris JK, Gold NJ & Knight R (2008) Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex Nat Methods **5:** 235-237.

Harmsen HJ, Wildeboer-Veloo AC, Raangs GC, Wagendorp AA, Klijn N, Bindels JG & Welling GW (2000) Analysis of intestinal flora development in breast-fed and formula-fed infants by using molecular identification and detection methods J PediatrGastroenterolNutr **30:** 61-67.

Harmsen HJM, Wildeboer-Veloo ACM, Raangs GC, Wagendorp AA, Klijn N, Bindels JG & Welling GW (2000) Analysis of Intestinal Flora Development in Breast-Fed and Formula-Fed Infants by Using Molecular Identification and Detection Methods J PediatrGastroenterolNutr **30:** 61-67.

Hartley CL, Clements HM & Linton KB (1977) Escherichia coli in the faecal flora of man J Appl Bacteriol **43:** 261-269.

Hayashi H, Sakamoto M &Benno Y (2002) Phylogenetic analysis of the human gut microbiota using 16S rDNA clone libraries and strictly anaerobic culture-based methods MicrobiolImmunol **46:** 535-548.

Hayashi H, Takahashi R, Nishi T, Sakamoto M & Benno Y (2005) Molecular analysis of jejunal, ileal, caecal and recto-sigmoidal human colonic microbiota using 16S rRNA gene libraries and terminal restriction fragment length polymorphism J Med Microbiol **54:** 1093-1101.

Hayford AE, Petersen A, Vogensen FK &Jakobsen M (1999) Use of conserved randomly amplified polymorphic DNA (RAPD) fragments and RAPD pattern for characterization of *Lactobacillus* fermentum in Ghanaian fermented maize dough Appl Environ Microbiol **65:** 3213-3221.

Head IM, Saunders JR & Pickup RW (1998) Microbial Evolution, Diversity, and Ecology: A Decade of Ribosomal RNA Analysis of Uncultivated Microorganisms MicrobEcol **35:** 1-21.

Herfel TM, Jacobi SK, Lin X, Walker DC, Jouni ZE &Odle J (2009) Safety evaluation of polydextrose in infant formula using a suckling piglet model Food ChemToxicol **47:** 1530-1537.

Hermie J M Harmsen, Peter Elfferich, Frits Schut, Gjalt,W.Welling (1999) A 16S rRNA-targeted Probe for Detection of Lactobacilli and Enterococci in Faecal Samples by Fluorescent In Situ Hybridization MicrobEcol Health Dis **11:** 3-12.

Hugenholtz P, Goebel BM & Pace NR (1998) Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity J Bacteriol **180:** 4765-4774.

Hungate RE *et al.* (1966) Rumen and its microbes.

Ishak HD, Plowes R, Sen R, Kellner K, Meyer E, Estrada DA, Dowd SE & Mueller UG (2011) Bacterial diversity in Solenopsis invicta and Solenopsis geminata ant colonies characterized by 16S amplicon 454 pyrosequencing Microb Ecol **61:** 821-831.

Isolauri E, Kirjavainen PV &Salminen S (2002) Probiotics: a role in the treatment of intestinal infection and inflammation? Gut **50 Suppl 3:** III54-9.

Isolauri E, Rautava S, Kalliomaki M, Kirjavainen P & Salminen S (2002) Probiotic research: learn from the evidence Allergy **57:** 1076-7; author reply 1078.

Jacobsen CN, Rosenfeldt Nielsen V, Hayford AE, Moller PL, Michaelsen KF, Paerregaard A, Sandstrom B, Tvede M &Jakobsen M (1999) Screening of probiotic activities of forty-seven strains of *Lactobacillus spp.* by *in vitro* techniques and evaluation of the colonization ability of five selected strains in humans Appl Environ Microbiol **65:** 4949-4956.

Jiang Y, Gao F, Xu XL, Su Y, Ye KP & Zhou GH (2010) Changes in the bacterial communities of vacuum-packaged pork during chilled storage analyzed by PCR-DGGE Meat Sci **86:** 889-895.

Jones RT, Robeson MS, Lauber CL, Hamady M, Knight R &Fierer N (2009) A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses ISME J **3:** 442-453.

Kaplan CW & Kitts CL (2003) Variation between observed and true Terminal Restriction Fragment length is dependent on true TRF length and purine content J Microbiol Methods **54:** 121-125.

Kassinen A, Krogius-Kurikka L, Makivuokko H, Rinttila T, Paulin L, Corander J, Malinen E, Apajalahti J &Palva A (2007) The fecal microbiota of irritable bowel syndrome patients differs significantly from that of healthy subjects Gastroenterology **133:** 24-33.

Kimura K, McCartney AL, McConnell MA &Tannock GW (1997) Analysis of fecal populations of bifidobacteria and lactobacilli and investigation of the immunological responses of their human hosts to the predominant strains Appl Environ Microbiol **63:** 3394-3398.

Kitts CL (2001) Terminal restriction fragment patterns: a tool for comparing microbial communities and assessing community dynamics Curr Issues Intest Microbiol **2:** 17-25.

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

Korhonen JM, Sclivagnotis Y & von Wright A (2007) Characterization of dominant cultivable lactobacilli and their antibiotic resistance profiles from faecal samples of weaning piglets J Appl Microbiol **103:** 2496-2503.

Konstantinov SR, Awati AA, Williams BA, Miller BG, Jones P, Stokes CR, Akkermans AD, Smidt H & de Vos WM (2006) Post-natal development of the porcine microbiota composition and activities Environ Microbiol **8:** 1191-1199.

Kunz C, Rudloff S, Baier W, Klein N &Strobel S (2000) OLIGOSACCHARIDES IN HUMAN MILK: Structural, Functional, and Metabolic Aspects Annu Rev Nutr **20:** 699-722.

Kwok AY, Su SC, Reynolds RP, Bay SJ, Av-Gay Y, Dovichi NJ & Chow AW (1999) Species identification and phylogenetic relationships based on partial HSP60 gene sequences within the genus Staphylococcus Int J Syst Bacteriol **49 Pt 3:** 1181-1192.

Ladero V, Fernandez M, Cuesta I & Alvarez MA (2010) Quantitative detection and identification of tyramine-producing enterococci and lactobacilli in cheese by multiplex qPCR Food Microbiol **27:** 933-939.

Langendijk PS, Schut F, Jansen GJ, Raangs GC, Kamphuis GR, Wilkinson MH & Welling GW (1995) Quantitative fluorescence in situ hybridization of Bifidobacterium spp. with genus-specific 16S rRNA-targeted probes and its application in fecal samples Appl Environ Microbiol **61:** 3069-3075.

Ley RE, Turnbaugh PJ, Klein S & Gordon JI (2006) Microbial ecology: Human gut microbes associated with obesity Nature **444:** 1022-1023.

Long SS & Swenson RM (1977) Development of anaerobic fecal flora in healthy newborn infants J Pediatr **91:** 298-301.

Lubbs DC, Vester BM, Fastinger ND & Swanson KS (2009) Dietary protein concentration affects intestinal microbiota of adult cats: a study using DGGE and qPCR to evaluate differences in microbial populations in the feline gastrointestinal tract J AnimPhysiolAnimNutr (Berl) **93:** 113-121.

Luchansky JB, Tennant MC & Klaenhammer TR (1991) Molecular cloning and deoxyribonucleic acid polymorphisms in Lactobacillus acidophilus and Lactobacillus gasseri J Dairy Sci **74:** 3293-3302.

Ludwig W, Strunk O, Westram R *et al.* (2004) ARB: a software environment for sequence data Nucleic Acids Res **32:** 1363-1371.

Macfarlane, G.T. Steeds, H. and Macfarlane, S. (2008) Bacterial metabolism and health related effects of galactooligosaccharides and other prebiotics. Journal of applied microbiology; Feb;104(2);305-344.

Matsumiya Y, Kato N, Watanabe K & Kato H (2002) Molecular epidemiological study of vertical transmission of vaginal *Lactobacillus* species from mothers to newborn infants in Japanese, by arbitrarily primed polymerase chain reaction J Infect Chemother **8:** 43-49.

Mackie RI, SghirA& Gaskins HR (1999) Developmental microbial ecology of the neonatal gastrointestinal tract Am J ClinNutr **69:** 1035S-1045S.

Mare L, Wolfaardt GM & Dicks LM (2006) Adhesion of *Lactobacillus* plantarum 423 and *Lactobacillus* salivarius 241 to the intestinal tract of piglets, as recorded with fluorescent in situ hybridization (FISH), and production of plantaricin 423 by cells colonized to the ileum J ApplMicrobiol **100:** 838-845.

Mackie RI, Sghir A & Gaskins HR (1999) Developmental microbial ecology of the neonatal gastrointestinal tract Am J Clin Nutr **69:** 1035S-1045S.

Marchesi JR, Dutilh BE, Hall N, Peters WH, Roelofs R, Boleij A & Tjalsma H (2011) Towards the human colorectal cancer microbiome PLoS One **6:** e20447.

Markiewicz LH, Biedrzycka E, Wasilewska E &Bielecka M (2010) Rapid molecular identification and characteristics of *Lactobacillus* strains Folia Microbiol (Praha) **55:** 481-488.

Martin FP, Wang Y, Sprenger N *et al.* (2008) Probiotic modulation of symbiotic gut microbial-host metabolic interactions in a humanized microbiome mouse model Mol Syst Biol **4:** 157.

Matsumiya Y, Kato N, Watanabe K & Kato H (2002) Molecular epidemiological study of vertical transmission of vaginal *Lactobacillus* species from mothers to newborn infants in Japanese, by arbitrarily primed polymerase chain reaction J Infect Chemother **8:** 43-49.

McIntosh RP (1967) An Index of Diversity and the Relation of Certain Concepts to Diversity Ecology **48:** 392-404.

Mills DK, Fitzgerald K, Litchfield CD & Gillevet PM (2003) A comparison of DNA profiling techniques for monitoring nutrient impact on microbial community composition during bioremediation of petroleum-contaminated soils J Microbiol Methods **54:** 57-74.

Moreira JL, Mota RM, Horta MF, Teixeira SM, Neumann E, Nicoli JR & Nunes AC (2005) Identification to the species level of *Lactobacillus* isolated in probiotic prospecting studies of human, animal or food origin by 16S-23S rRNA restriction profiling BMC Microbiol **5:** 15.

Monaco MH, Kashtanov DO, Wang M, Walker DC, Rai D, Jouni ZE, Miller MJ & Donovan SM (2011) Addition of polydextrose and galactooligosaccharide to formula does not affect bacterial translocation in the neonatal piglet J Pediatr Gastroenterol Nutr **52:** 210-216.

Nelson GM & George SE (1995) Comparison of media for selection and enumeration of mouse fecal flora populations J Microbiol Methods **22:** 293-300.

Neu J & Walker WA (2011) Necrotizing enterocolitis N Engl J Med **364:** 255-264.

Newberg and Neubauer (1995) Carbohydrates in milks, analysis, quantities, and significance. Handbook of milk composition.

Nieminen TT, Vihavainen E, Paloranta A, Lehto J, Paulin L, Auvinen P, Solismaa M &Bjorkroth KJ (2011) Characterization of psychrotrophic bacterial communities in modified atmosphere-packed meat with terminal restriction fragment length polymorphism Int J Food Microbiol **144:** 360-366.

Nigatu A, Ahrne S &Molin G (2001) Randomly amplified polymorphic DNA (RAPD) profiles for the distinction of *Lactobacillus* species Antonie Van Leeuwenhoek **79:** 1-6.

Orrhage K & Nord CE (1999) Factors controlling the bacterial colonization of the intestine in breastfed infants ActaPaediatrSuppl **88:** 47-57.

Ott SJ, Musfeldt M, Wenderoth DF, Hampe J, Brant O, Folsch UR, Timmis KN & Schreiber S (2004) Reduction in diversity of the colonic mucosa associated bacterial microflora in patients with active inflammatory bowel disease Gut **53:** 685-693.

Ouwehand A, Isolauri E &Salminen S (2002) The role of the intestinal microflora for the development of the immune system in early childhood Eur J Nutr **41 Suppl 1:** I32-7.

O'Sullivan DJ (2000) Methods for analysis of the intestinal microflora Curr Issues Intest Microbiol **1:** 39-50.

O'Sullivan DJ (2001) Screening of intestinal microflora for effective probiotic bacteria J Agric Food Chem **49:** 1751-1760.

Pang H, Kitahara M, Tan Z, Wang Y, Qin G, Ohkuma M & Cai Y (2012) Reclassification of Lactobacillus kimchii and Lactobacillus bobalius as later subjective synonyms of Lactobacillus paralimentarius Int J Syst Evol Microbiol **62:** 2383-2387.

Penders J, Thijs C, Vink C, Stelma FF, Snijders B, Kummeling I, van den Brandt PA &Stobberingh EE (2006) Factors influencing the composition of the intestinal microbiota in early infancy Pediatrics **118:** 511-521.

Petersson A, Domig KJ, Nagel P, Zollitsch W, Hagmuller W &Kneifeld W (2009) Denaturing gradient gel electrophoresis (DGGE)-based monitoring of intestinal lactobacilli and bifidobacteria of pigs during a feeding trial Arch AnimNutr **63:** 112-126.

Peu P, Brugere H, Pourcher AM, Kerouredan M, Godon JJ, Delgenes JP &Dabert P (2006) Dynamics of a pig slurry microbial community during anaerobic storage and management Appl Environ Microbiol **72:** 3578-3585.

Picozzi C, Bonacina G, Vigentini I &Foschino R (2010) Genetic diversity in Italian *Lactobacillus* sanfranciscensis strains assessed by multilocus sequence typing and pulsed-field gel electrophoresis analyses Microbiology **156:** 2035-2045.

Pieper R, Janczyk P, Zeyner A, Smidt H, Guiard V & Souffrant WB (2008) Ecophysiology of the developing total bacterial and *Lactobacillus* communities in the terminal small intestine of weaning piglets Microb Ecol **56:** 474-483.

Pillidge CJ, Sheehy LM, Shihata A, Pu Z, Dobos M & Powell IB (2009) Intragenomic 16S rRNA gene heterogeneity in Lactococcus lactis subsp. cremoris Int Dairy J **19:** 222-227.

Poroyko V, White JR, Wang M, Donovan S, Alverdy J, Liu DC & Morowitz MJ (2010) Gut microbial gene expression in mother-fed and formula-fed piglets PLoS One **5:** e12459.

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

Pozhitkov A, Noble PA, Domazet-Loso T, Nolte AW, Sonnenberg R, Staehler P, Beier M & Tautz D (2006) Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted Nucleic Acids Res **34:** e66.

Preza D, Olsen I, Willumsen T, Boches SK, Cotton SL, Grinde B &Paster BJ (2009) Microarray analysis of the microflora of root caries in elderly Eur J ClinMicrobiol Infect Dis **28:** 509-517.

Rachman C, Kabadjova P, Valcheva R, Prevost H &Dousset X (2004) Identification of Carnobacterium species by restriction fragment length polymorphism of the 16S-23S rRNA gene intergenic spacer region and species-specific PCR Appl Environ Microbiol **70:** 4468-4477.

Ranawana V, Muller A & Henry CJ (2012) Polydextrose: its impact on short-term food intake and subjective feelings of satiety in males-a randomized controlled cross-over study Eur J Nutr.

Ribeiro TC & Latronico AC (2012) Insulin-like growth factor system on adrenocortical tumorigenesis Mol Cell Endocrinol **351:** 96-100.

Roberfroid M (2007) Prebiotics: the concept revisited J Nutr **137:** 830S-7S.

Roberfroid M, Gibson GR, Hoyles L *et al.* (2010) Prebiotic effects: metabolic and health benefits Br J Nutr **104 Suppl 2:** S1-63.

Roesch LF, Lorca GL, Casella G *et al.* (2009) Culture-independent identification of gut bacteria correlated with the onset of diabetes in a rat model ISME J **3:** 536-548.

Roussel Y, Colmin C, Simonet JM & Decaris B (1993) Strain characterization, genome size and plasmid content in the *Lactobacillus acidophilus* group (Hansen and Mocquot) J Appl Bacteriol **74:** 549-556.

Roy D, Ward P, Vincent D &Mondou F (2000) Molecular identification of potentially probiotic lactobacilli CurrMicrobiol **40:** 40-46.

Saito, D., Marsh, T.L., de Souza Cannavan, F., Hofling, J.F., Goncalves, R.B. (2009) Assessment of intraradicular bacterial composition by terminal restriction fragment length polymorphism analysis. Oral microbiology and immunology. 24, 369-376.

Sakamoto M, Hayashi H & Benno Y (2003) Terminal restriction fragment length polymorphism analysis for human fecal microbiota and its application for analysis of complex bifidobacterial communities Microbiol Immunol **47:** 133-142.

Sakamoto M & Ohkuma M (2011) Identification and classification of the genus Bacteroides by multilocus sequence analysis Microbiology **157:** 3388-3397.

Salminen S, Bouley C, Boutron-Ruault MC, Cummings JH, Franck A, Gibson GR, Isolauri E, Moreau MC, Roberfroid M & Rowland I (1998) Functional food science and gastrointestinal physiology and function Br J Nutr **80 Suppl 1:** S147-71.

Samelis J, Bleicher A, Delbes-Paus C, Kakouri A, Neuhaus K & Montel MC (2011) FTIR-based polyphasic identification of lactic acid bacteria isolated from traditional Greek Graviera cheese Food Microbiol **28:** 76-83.

Sanders ME (2003) Probiotics: considerations for human health Nutr Rev **61:** 91-99.

Scalabrin DM, Mitmesser SH, Welling GW, Harris CL, Marunycz JD, Walker DC, Bos NA, Tolkko S, Salminen S & Vanderhoof JA (2012) New prebiotic blend of polydextrose and galacto-oligosaccharides has a bifidogenic effect in young infants J Pediatr Gastroenterol Nutr **54:** 343-352.

Schleifer and Ludwig (1995) Phylogenetic relationships of lactic acid bacteria. In The Genera of Lactic Acid Bacteria, pp. 7–18. Edited by Wood B. J. B., Holzapfel W. H. London: Blackie Academic & Professional

Schloss PD, Westcott SL, Ryabin T *et al.* (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities Appl Environ Microbiol **75:** 7537-7541.

Silvi, S., Nardi, M., Sulpizio, R., Orpianesi, C., Caggiano, M., Carnevali, O., Cresci,A. (2011) Effect of the addition of *Lactobacillus delbrueckii* subsp. *delbrueckii* on the gut microbiota composition and contribution to the well-being of European sea bass (*Dicentrarchuslabrax*, L.) MicrobEcol Health Dis **20:**.

Singh S, Goswami P, Singh R & Heller KJ (2009) Application of molecular identification tools for *Lactobacillus*, with a focus on discrimination between closely related species: A review LWT - Food Science and Technology **42:** 448-457.

Sjogren YM, Tomicic S, Lundberg A, Bottcher MF, Bjorksten B, Sverremark-Ekstrom E & Jenmalm MC (2009) Influence of early gut microbiota on the maturation of childhood mucosal and systemic immune responses Clin Exp Allergy **39:** 1842-1851.

Skowrońska A, Mikulski D, Mikulska M, Zapotoczny P, Zmysłowska I, Szczypiński PM (2002) Methodological Aspects and Application of Fluorescent in situ Hybridization to Identify Bacteria from the Gastrointestinal Tract of Turkeys.

Sokol H, Seksik P, Rigottier-Gois L, Lay C, Lepage P, Podglajen I, Marteau P & Dore J (2006) Specificities of the fecal microbiota in inflammatory bowel disease Inflamm Bowel Dis **12:** 106-111.

Song Y, Kato N, Liu C, Matsumiya Y, Kato H & Watanabe K (2000) Rapid identification of 11 human intestinal *Lactobacillus* species by multiplex PCR assays using group- and species-specific primers derived from the 16S-23S rRNAintergenic spacer region and its flanking 23S rRNA FEMS MicrobiolLett **187:** 167-173.

Spasenovski T, Carroll MP, Payne MS & Bruce KD (2009) Molecular analysis of diversity within the genus Pseudomonas in the lungs of cystic fibrosis patients Diagn Microbiol Infect Dis **63:** 261-267.

Spear GT, Gilbert D, Landay AL, Zariffard R, French AL, Patel P &Gillevet PM (2011) Pyrosequencing of the genital microbiotas of HIV-seropositive and -seronegative women reveals *Lactobacillus* iners as the predominant *Lactobacillus* Species Appl Environ Microbiol **77:** 378-381.

Stackebrandt E, Liesack W & Witt D (1992) Ribosomal RNA and rDNA sequence analyses Gene **115:** 255-260.

Stark PL & Lee A (1982) The microbial ecology of the large bowel of breast-fed and formula-fed infants during the first year of life J Med Microbiol **15:** 189-203.

Tannock GW, Fuller R, Smith SL & Hall MA (1990) Plasmid profiling of members of the family Enterobacteriaceae, lactobacilli, and bifidobacteria to study the transmission of bacteria from mother to infant J ClinMicrobiol **28:** 1225-1228.

Tarnberg M, Jakobsson T, Jonasson J &Forsum U (2002) Identification of randomly selected colonies of lactobacilli from normal vaginal fluid by pyrosequencing of the 16S rDNA variable V1 and V3 regions APMIS **110:** 802-810.

Tarr CL, Patel JS, Puhr ND, Sowers EG, Bopp CA & Strockbine NA (2007) Identification of Vibrio isolates by a multiplex PCR assay and rpoB sequence determination J Clin Microbiol **45:** 134-140.

Teanpaisan R &Dahlen G (2006) Use of polymerase chain reaction techniques and sodium dodecyl sulfate-polyacrylamide gel electrophoresis for differentiation of oral *Lactobacillus* species Oral MicrobiolImmunol **21:** 79-83.

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

Tost J & Gut IG (2007) DNA methylation analysis by pyrosequencing Nat Protoc **2:** 2265-2275.

Trask BJ (1991) Fluorescence in situ hybridization: applications in cytogenetics and gene mapping Trends Genet **7:** 149-154.

Tuohy KM, Rouzaud GC, Bruck WM & Gibson GR (2005) Modulation of the human gut microflora towards improved health using prebiotics--assessment of efficacy Curr Pharm Des **11:** 75-90.

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

Vaughan EE, de Vries MC, Zoetendal EG, Ben-Amor K, Akkermans AD & de Vos WM (2002) The intestinal LABs Antonie Van Leeuwenhoek **82:** 341-352.

Ventura M & Zink R (2002) Specific identification and molecular typing analysis of *Lactobacillus johnsonii* by using PCR-based methods and pulsed-field gel electrophoresis FEMS MicrobiolLett **217:** 141-154.

Wall R, Fitzgerald G, Hussey S, Ryan T, Murphy B, Ross P & Stanton C (2007) Genomic diversity of cultivable *Lactobacillus* populations residing in the neonatal and adult gastrointestinal tract FEMS MicrobiolEcol **59:** 127-137.

Walter J, Britton RA &Roos S (2011) Host-microbial symbiosis in the vertebrate gastrointestinal tract and the *Lactobacillus reuteri* paradigm ProcNatlAcadSci U S A **108 Suppl 1:** 4645-4652.

Walter J, Tannock GW, Tilsala-Timisjarvi A, Rodtong S, Loach DM, Munro K & Alatossava T (2000) Detection and identification of gastrointestinal *Lactobacillus* species by using denaturing gradient gel electrophoresis and species-specific PCR primers Appl Environ Microbiol **66:** 297-303.

Wang RF, Beggs ML, Erickson BD &Cerniglia CE (2004) DNA microarray analysis of predominant human intestinal bacteria in fecal samples Mol Cell Probes **18:** 223-234.

Weber DG, Sahm K, Polen T, Wendisch VF &Antranikian G (2008) Oligonucleotide microarrays for the detection and identification of viable beer spoilage bacteria J ApplMicrobiol **105:** 951-962.

Weckx S, Van der Meulen R, Allemeersch J, Huys G, Vandamme P, Van Hummelen P & De Vuyst L (2010) Community dynamics of bacteria in sourdough fermentations as revealed by their metatranscriptomeAppl Environ Microbiol **76:** 5402-5408.

Yilmaz LS, Bergsven LI & Noguera DR (2008) Systematic evaluation of single mismatch stability predictors for fluorescence in situ hybridization Environ Microbiol **10:** 2872-2885

Yu J, Wang WH, Menghe BL *et al.* (2011) Diversity of lactic acid bacteria associated with traditional fermented dairy products in Mongolia J Dairy Sci **94:** 3229-3241.

Zhang H, DiBaise JK, Zuccolo A *et al.* (2009) Human gut microbiota in obesity and after gastric bypass Proceedings of the National Academy of Sciences **106:** 2365-2370.

Zoetendal EG, Akkermans AD & De Vos WM (1998) Temperature gradient gel electrophoresis analysis of 16S rRNA from human fecal samples reveals stable and host-specific communities of active bacteria Appl Environ Microbiol 64: 3854-3859.

APPENDIX

Table S1 Profiles of TRFs generated by *Alu*I and *Tac*I in 499 bp *hsp60* partial products (1/3).

	***Alu*I**		***Tac*I**	
	Foward[1]	**Reverse**	**Foward**	**Reverse**
L. acetotolerans	63[2]	33	499	0
L. acidipisci	21	168	56	43
L. acidophilus (A)[3]	135	33	217	220
L. acidophilus (B)	138	33	278	220
L. alimentarius	21	168	80	220
L. amylophilus (A)	499	0	224	221
L. amylophilus (B)	63	34	278	221
L. amylophilus (C)	499	0	192	254
L. amylovorus (A)	63	54	499	0
L. amylovorus (B)	72	54	278	241
L. aviarius (A)	21	348	56	356
L. aviarius subsp. aviarius (B)	27	168	56	220
L. aviarius subsp. aviarius (C)	27	168	56	356
L. bifermentans	72	168	56	220
L. brevis (A)	250	16	278	203
L. brevis (B)	3	39	254	227
L. brevis (C)	130	39	254	227
L. brevis (D)	227	39	254	227
L. buchneri (A)	250	33	499	0
L. buchneri (B)	78	230	121	281
L. casei (A)	330	168	499	0
L. casei (B)	330	168	278	220
L. casei (C)	330	168	56	256
L. coryniformis	153	168	56	220
L. crispatus (A)	63	33	499	0
L. crispatus (B)	153	33	499	0
L. curvatus	75	168	110	220
L. cypricasei	21	168	56	43
L. delbrueckii (A)	330	33	499	0
L. delbrueckii (B)	75	33	242	220
L. delbrueckii (C)	100	33	242	220
L. delbrueckii (D)	135	33	242	220
L. delbrueckii (E)	150	33	242	220
L. delbrueckii (F)	330	33	242	220

Table S1 (Cont. 2/3)

	*Alu*I		*Tac*I	
	Foward[1]	**Reverse**	**Foward**	**Reverse**
L. dextrinicus (A)	75	169	499	0
L. dextrinicus (B)	153	169	499	0
L. dextrinicus (C)	63	34	56	257
L. durianis	150	168	56	160
L. farciminis	138	187	56	179
L. ferintoshensis	173	168	278	220
L. fermentum	274	224	499	0
L. fructivorans	330	168	499	0
L. gallinarum	63	33	56	442
L. gasseri	75	33	278	220
L. helveticus (A)	138	35	499	0
L. helveticus (B)	330	170	499	0
L. helveticus (C)	63	35	278	222
L. hilgardii	182	257	85	283
L. homohichii	75	168	224	256
L. iners	27	33	242	256
L. ingluviei	274	168	56	356
L. intesinalis	138	33	499	0
L. jensenii	138	360	499	0
L. jensenii[3]	150	360	499	0
L. johnsonii (A)	75	33	224	220
L. johnsonii (B)	75	33	278	220
L. johnsonii (C)	110	33	224	220
L. kefiranofaciens (A)	72	168	499	0
L. kefiranofaciens (B)	465	33	278	220
L. kefiri	153	345	224	122
L. kimchii	138	33	56	160
L. mali	72	426	100	398
L. mali (yamanashiensis)	72	168	499	0
L. paracasei	330	168	499	0
L. paraplantarum (A)	138	169	499	0
L. paraplantarum (B)	173	146	278	100
L. paraplantarum (C)	165	168	278	220
L. paraplantarum (D)	63	34	278	221
L. paraplantarum (E)	250	34	278	221
L. pentosus (A)	110	143	438	60
L. pentosus (B)	63	33	278	220
L. plantarum (A)[3]	63	34	278	221
L. planturum (B)	110	144	334	61

Table S1 (Cont. 3/3)

	*Alu*I		*Tac*I	
	Foward[1]	**Reverse**	**Foward**	**Reverse**
L. reuteri (A)	84	224	499	0
L. reuteri (B)	84	414	73	242
L. reuteri (C)	84	224	121	281
L. reuteri (D)	63	325	121	281
L. reuteri (E)[4]	152	224	129	281
L. rhamnosus (A)	165	168	499	0
L. rhamnosus (B)	232	168	499	0
L. rhamnosus (C)[3]	330	168	56	3
L. rhamnosus (D)	330	168	56	442
L. ruminis (A)	110	426	499	0
L. ruminis (B)	72	464	499	0
L. sakei (A)[3]	75	65	499	0
L. sakei (B)	75	168	499	0
L. sakei (C)	75	168	110	220
L. salivarius (A)	135	168	56	160
L. salivarius (B)	330	168	242	256
L. salivarius (C)[4]	78	31	57	256
L. sanfranciscensis	72	425	56	441
L. vaccinostercus	150	168	56	160
L. vaginalis	84	33	278	220
L. zeae (A)	330	168	499	0
L. zeae (B)	330	168	56	256
L. zeae (C)	330	168	56	281

[1] All the primers are fluorescence labeled

[2] Unit of the value: Basepairs (bp)

[3] Strains depicted in **Fig. 2.3**

[4] Strains found in this research

Table S2. Primers Used in this Thesis

Primer	Target	Sequence (5'-3')	Reference
H279	*Lactobacillus hsp60* gene	GAATTCGAIIIIGCIGGIGA(TC)GGIACIACIAC	Goh *et al.*, 1996
H280		CGCGGGATCC(TC)(TG)I(TC)(TG)ITCICC(AG)AAICCIGGIGC(TC)TT	
LB308F	Bacterial *hsp60* gene	TGAAGAAYGTNRYNGCYGG	Blaiotta *et al.*, 2008
LB806RM		AANGTNCCVCGVATCTTGTT	
LB308F-D[1]	Bacterial *hsp60* gene	/56-FAM/TGAAGAAYGTNRYNGCYGG	This study
LB806RM-D[2]		/5HEX/AANGTNCCVCGVATCTTGTT	
ERIC-1R	Multiple sites on bacterial genome	ATGTAAGCTCCTGGGGATTCAC	Versalovic *et al.*, 1991
ERIC-2		AAGTAAGTGACTGGGGTGAGCG	
Gray28F	Bacterial 16S rDNA V1-V3	GAGTTTGATCNTGGCTCAG	Ishak *et al.,* 2011
Gray519R		GTNTTACNGCGGCKGCTG	

[1]A fluorescent dye 6-FAM was added onto the 5' of original primer LB308F
[2]A fluorescent dye Hex was added onto the 5' of original primer LB806RM

Code S1: Internal Pearl code for pyroseuqencing data analysis developed by Dr.Carl J. Yeoman[1]

```
#! /usr/bin/perl
system("clear");
print "\n\n\n\t\t\tMY454 v3.4\n\n";
################################DETAILS#########################################################
###

#my454 - written by Carl J. Yeoman
#Is intended to optimally take raw 16s amplicon files through to clustered and analyzed material
#locate primer sequences
#separate forward and reverse reads (currently 27f and 534r)
#Reverse Complement the reverse reads and write to separate files
#parse files removing duplicates and concatenate while forming groups file

#USAGE: my454 (options) -a [association file]; -o [output prefix]
#                            -help (Will give these instructions)

#Options:
#-c [(y/n) overlook missing sample files - default = n]
#-l [(minimum length (1-1000) -space- maximum length (200-1000)) initial minimum length cutoff =
200nt and maximum 535nt]
#-h [(number) maximum homopolymers tolerated - default = 6]
#-k [(all/legible) all - keep all reads in one file
#                                allRC - keep all and expect sequences to be reverse complement
(default)
#                                legible - keep all but those with unclear primer sequence
#                                     ]
#-r [do not remove primers - default removes primers]
#-global [remove/label] - used to trigger a global search for duplicates if it is suspected that
identical read names will occur in different files
#-m [maximum number of reads to keep in any one sample]
#-ambig [keep sequences with ambiguous characters - this is my current thinking that these should
be evaluated downstream after trimming etc as they may not actually be part of the final
evaluated sequence so it seems a waste throwing them out right off the bat]
#-trimlong [Trim the 3' region of sequences that exceed the upper defined length limit (see -l)]
#NOTE:
#Association file should be formated [Sample file] -tab- [Group]
#If -q is selected their should be a third column with the quality file details
#To split reads/groups into separate files insert: Fsplit [tab] [next files name] in the
association file where you'd like this to occur.

################################PROGRAM#########################################################
###

################################Set variables#################################
my $untrimmed = 0;  my $unclear = 0;
my $linecount = 'x'; my $seqs = 0;
my $for = 0; my $rev = 0;
my $rp = 'x';
my $len_min = 200; my $len_max = 535;
my $max_homop = 6; my $keep = 'allrc'; my $pre_base = 'n';
my $ambig = 0; my $badslen = 0; my $badllen = 0;my $homo = 0; my $count = 0; my $kept = 0;
my $file = 1; my $next_file = '';
my $AJ = 0; my $pause = 'n';
my $duplicate= ''; my $found = 'n';
my $continue = 'n'; my $output = '';
my $QualProcess = 'n'; my $pooled = 'n';
my $threshold = 'x'; my $QualSoftRun = 1;
```

[1] Posted with the permission from Carl Yeoman (carl.yeoman@hotmail.com), copyright reserved.

```
my $newseq = ''; my $catfile = '';
my $prebase = ''; my $sequence ='';
my @summary = ''; my $foundQs = 'n';
my $Qscores = ''; my @Qbits = '';
my $quality = 0; my $bits = 0; my $i = 0;
my $global = 'n';
my $large = 'n';
my $DUPreadCount = 0;
my $maxReads = 'n';
my $count_to_max = 0;
my $KeepAmbig = 'n';
my $trimlong = 'n';
@now = localtime(time); $start = $now[1]; $report = ($start + 5); if ($report > 60) {$report =
($report - 60);}

#################Read and set user defined variables#######################
Options:
$left = @ARGV;

if ($left > 0) {$next = shift (@ARGV); chomp $next;

        if ($next=~ /-/) {$next=~ s/-//g;
                if ($next eq 'a') {$ass = shift (@ARGV);}
                elsif ($next eq 'c') {$continue = shift (@ARGV);}
                elsif ($next eq 'o') {$output = shift (@ARGV);}
                elsif ($next eq 'r') {$rp = 'r';}
                elsif ($next eq 'm') {$maxReads = shift (@ARGV); --$maxReads;}
                elsif ($next eq 'l') {$len_min = shift (@ARGV); if ($ARGV[0]=~ /-/) {} else
{$len_max = shift (@ARGV);}}
                elsif ($next eq 'h') {$max_homop = shift (@ARGV);}
                elsif ($next eq 'k') {$keep = shift (@ARGV); $keep = lc ($keep);}
               elsif ($next eq 'q') {$QualProcess = shift (@ARGV); if ($QualProcess=~ /Clip-Soft/)
{$QualSoftRun = shift (@ARGV);} $quality = shift (@ARGV); }
                elsif ($next eq 't') {$threshold = shift (@ARGV); }
                elsif ($next eq 'help') {goto Help;}
                elsif ($next eq 'p') {$pooled = 'y';}
                elsif ($next eq 'global') {$global = shift (@ARGV);}
                elsif ($next eq 'large') {$large = 'y';}
                elsif ($next eq 'ambig') {$KeepAmbig = 'y';}
                elsif ($next eq 'trimlong') {$trimlong = 'y';}
                else {print "Unrecognized option ", $next, "\n"; exit;}
        }
goto Options;
}

##########################Check for write to files#####################
@otpt = split (/\./, $output);
$groupsfile = ($otpt[0].".groups"); $catfile = ($otpt[0].".fasta");
@otpt = '';

######################Open Write to Files############################
open (GROUP, ">", $groupsfile); open (CAT, ">", $catfile);
#open (TEMPGROUP, ">my454tempgroup"); open (TEMPCAT, ">my454tempcat");

#########################Open Association file#########################
unless (open (ASS, $ass) ) {print "Association file not found"; exit;}
while ($line = <ASS>) { chomp $line; #1
@a = split (/\t/, $line);

        $next_file = $a[0]; #Get file name from Association file
        $group = $a[1]; #Get group
```

```
       if ($next_file eq 'Fsplit') {print"\nReads now being read to ".$group." file, ".$output."
had ".$count." reads total\n\n"; $output = $group; $groupsfile = ($output."_groups"); $catfile =
($output."_concatenated"); $count = 0; $duplicate=''; next;}
       unless (-e $next_file) {print $next_file, " file not found\n"; if ($continue eq 'n')
{exit;} next;} #Check for file existence
       unless ($QualProcess eq 'n') {$QualFile = $a[2];}
       if ($global eq 'n') {$duplicate = '';}
       $pause = 'n'; $count_to_max = 0; #Reset triggers
########################Get info associated with each file##############################
       open (NFILE, $next_file);
       while ($line = <NFILE>) { chomp $line; $pause = 'n';   #2
              if (($linecount eq 'x') and ($line =~ /^>/)  ) { ++$seqs; $line=~ s/\srank.*//;
$line=~ s/\ .*//g; $line=~ s/\t.*//g; $line=~ s/\s.*//g; $header = $line; $linecount = 'g';
$grouphead = $header; $grouphead=~ s/>//g;}
              elsif (($linecount eq 'g') and ($line =~ /^>/)) { ++$seqs; #3

######################Check for duplicate reads#################################################
                     if ($duplicate=~ m/$grouphead/g) { #4
                     #count them
                     ++$DUPreadCount;
                     #Global = label
                     if ($global eq 'label') { #5
                     print "Duplicate read ",$grouphead," was found in file ".$next_file;
                     FindNewLabel:
                     $grouphead = ($grouphead."1");
                     if ($duplicate=~ m/$grouphead/g) {goto FindNewLabel;}
                     print " it has been relabeled ".$grouphead."\n";
                     } #\5
                     #Global = remove or local remove
                     else {$pause = 'y'; print "Duplicate read ",$grouphead," was found in file
",$next_file, " it has been removed from the concatenated and group files\n";}
                     } #\4
                     unless ($pause eq 'y') {$duplicate = ($duplicate." ".$grouphead); }

############ CHECK Length#################################################
unless ($pause eq 'y') {@qualcheck = split ('', $sequence); $length = @qualcheck; #6
                  if ($length < $len_min) {$pause = 'y';++$badslen;}
              elsif ($length > $len_max) {++$badllen; #7

              if ($trimlong eq 'y') { --$len_max; #8
              $trseq = substr ($sequence, 0, $len_max); $sequence = $trseq;
              ++$len_max} #\8
              else {$pause = 'y';} #Remove if not of desired length
              } #\7
} #\6
#######Set up MaxhomoP#####
--$max_homop;

###############Check for ambiguous characters 'N's#########################
unless ($pause eq 'y') {foreach $check (@qualcheck) { #9 #10
$check = uc ($check);
              if ($check eq 'N') {++$ambig; unless ($KeepAmbig eq 'y') {$pause = 'y'; } goto
Break;}
###############Check for homopolymers#######################################
                    if ($check eq $pre_base) {++$homop; if ($homop > $max_homop) {$pause = 'y';
++$homo; $homop = 0; goto Break;}}
                     else {$pre_base = 'n'; $homop = 0; }
                     $pre_base = $check;
              } #\10
              $pre_base = 'n'; $homop = 0;

} #\9 (1,2,3 still open)
```

```
Break:
#####Return MaxhomoP######
++$max_homop;

##############################Find Primer(s)#########################################

unless ($pause eq 'y') {if ($sequence=~ /^ATTACCGCGGCTGCTGGCA/) { #If 534r is detected #11 #12
                              $sequence = reverse ($sequence); $sequence=~ tr/ACGT/TGCA/;
$rev++; } #Reverse complement #\12
                        elsif ($sequence =~ /.*ATTACCGCGGCTGCTGGCA/) {$sequence=~
s/.*ATTACCGCGGCTGCTGGCA/ATTACCGCGGCTGCTGGCA/i; $sequence = reverse ($newseq); $sequence=~
tr/ACGT/TGCA/; $untrimmed++; $rev++; }
                        elsif ($sequence =~ /^AGAGTTTGAT[CT][AC]TGGCTCAG/) { $for++; }
                        elsif ($sequence =~ /.*AGAGTTTGAT[CT][AC]TGGCTCAG/) {$sequence=~
s/.*AGAGTTTGAT/AGAGTTTGAT/i; $untrimmed ++; $for++; }
                        else {$unclear++; if ($keep eq 'allrc') {$sequence = reverse ($newseq);
$sequence=~ tr/ACGT/TGCA/;} if ($keep eq 'legible') {$pause = 'y';}}
                        unless ($rp eq 'r') { $sequence=~ s/^AGAGTTTGAT[CT][AC]TGGCTCAG//;
$sequence=~ s/TGCCAGCAGCCGCGGTAAT$//}; #Remove primer if desired
} #\11 (1,2,3 still open)

##############################Check Sequence quality#############################
#SLOW and consequently not very useful!
#unless ($QualProcess eq 'n') {
#print "Qualcheck To Be Fixed\n"; exit;}

##############################write to groups file#############################
if ($pause eq 'y') { ++$totalbad;}           #Pause occurs when read is redundant within file or
quality criteria are not met.
       else {  ++$kept;              #13
                              print GROUP "$grouphead","\t","$group","\n";
                              ++$count;
##############################write to concatenated file#########################
               $header = (">".$grouphead);

#write to cat file
                        print CAT "$header", "\n", "$sequence", "\n";
               ###Check for MaxReads###
unless ($maxReads eq 'n') { ++$count_to_max; if ($count_to_max > $maxReads) { goto FileReport;}}
               } #\13
               $line=~ s/\srank.*//; $line=~ s/\ len.*//g; $line=~ s/\ .*//g; $line=~ s/\t.*//g;
$line=~ s/\s.*//g; $grouphead = $line; $grouphead=~ s/>//g; $sequence = '';} #\3
       else {$sequence = ($sequence.$line);}

########################Report to screen at set intervals#####################################
       @now = localtime(time); $clock = ($now[2].":".$now[1].":".$now[0]);
       if ($now[1] == $report) {$report = ($report + 5); if ($report > 60) {$report = ($report -
60);}; print $clock.": ".$count." reads added\n";}
} #\2
FileReport:
############################Completed file Report#################################################
@now = localtime(time); $clock = ($now[2].":".$now[1].":".$now[0]);
print "File ",$next_file, " added at ", $clock, "\n";
print "There were ", $seqs, " reads\t(", $for, " forward reads and ", $rev, " reverse
reads):\n\n";
print $untrimmed, " reads started before the primer, that sequence was trimmed\n";
print $unclear." reads did not have an clear primer sequence and were "; if ($keep eq 'legible')
{print "removed";} elsif ($keep eq 'all') {print"kept";} elsif ($keep eq 'allrc') {print"kept and
reverse complemented";} print "\n";
print $DUPreadCount," duplicate reads have been detected and were "; if ($global eq 'label')
{print "relabeled\n";} else {print" removed\n";}
```

```
print $badslen." were shorter than the specified ".$len_min."nt ".$badllen." were longer than
".$len_max."nt\n";
if ($trimlong eq 'y') {print "Long sequences were trimmed\n";}
print $homo." contained a homopolymer greater than the specified ".$max_homop."\n";
print "and ".$ambig." sequences with ambiguous bases were detected ";if ($KeepAmbig eq 'y')
{print "but retained\n";} else {print "and removed\n";}
print "\nIn all ".$totalbad." reads were removed:\n\n";
print $kept." sequences have been added to concatenated and group files have been saved in the
workfolder\nThere are now ".$count." sequences in this file\n\n\n";
$unclear = 0; $linecount = 'x'; $seqs = 0; $for = 0; $rev = 0;  $ambig = 0; $badslen = 0;
$badllen = 0; $totalbad = 0; --$maxhomop; $kept=0;
close NFILE;
###Remove chimeras and put to main file
#system("mothur \"#set.current(fasta=my454tempcat, group=my454tempgroup); unique.seqs();
chimera.uchime(fasta=current, name=current); remove.seqs(accnos=current, name=current,
group=current)\"");
} #\1
##########Close Files##############
close GROUP; close CAT;
close ASS;
###   IN PROGRESS!!

######################OPTIMIZE SEQUENCE LENGTH / READ NUMBER#####################
################################################################################
#print "\n\n\nOptimizing processing conditions to your data\n\n";
##Evaluate summary determine what minimum length should be###
#print "Optimizing length cutoff\n"
#@all_lengths = split (/\t/, $summary[1]);
#@sorted_lengths = sort {$a <=> $b} @all_lengths;
#$PrepdReads = @sorted_lengths; $next_check_reads = int ($PrepdReads / 1.429);
#$opt_reads = $next_check_reads; $opt_length = $len_max; $opt_test = 1;
#if ($PrepdReads <= 2000) {
#print "!!!WARNING!!!:   Less than 2,000 reads were kept from preprosessing.\n
#This is much less than the reccomended number of reads required to get an accurate ecological
measure\n
#No optimization will be applied, you may wish to quit this program and restart using adapted
parameters\n";
#goto Unique;
#}
#$i = 0;
#foreach $b (@sorted_lenghts) {
#++$i;
#   if ($i < $next_check_reads) {next;}
#elsif ($opt_length == $len_max) {$opt_length = $b; ++$opt_reads;}
#elsif ($i >= $next_check_reads) {
#$loss_of_length = ($opt_length - $b); $loss_of_length = int ($loss_of_length / $opt_test);
#if ($loss_of_length > $threshold) {++$opt_test;
#  if ($opt_test > 10) {last;}
#else {++$opt_test}
#}
#else {$opt_length = $b}
#$next_check_reads += 100; ++$opt_reads;}
#else {++$opt_reads; next;}
#}
#$per_read = ($opt_reads / $PrepdReads); $per_read = int ($per_read * 100);
#print "The optimum minimum length was determined to be ".$opt_length."\n";
#print "This will retain ".$per_read." % of preprocessed reads\n\n";
#####Remove short reads#####
#print "Removing short reads\n";
#open (SUM, "tempfile_summary.txt");
#while ($line = <SUM>) {
#@a = split (/\t/, $line);
```

```
#}
#Unique:
#exit;
#$groupsfile2 = $groupsfile; $catfile2 = $catfile;
#$groupsfile2=~ s/\.txt/\.good\.txt/; $catfile2=~ s/\.txt/\.good\.txt/;
#$catfile3 = $catfile2;
#$catfile3=~ s/\.txt/\.unique\.txt/;
###Remove short sequences and long sequences### ##Eliminate redundant sequences## ##Align
sequences##
#$command2 = ('~/Mothur.source/mothur \"\#screen.seqs(fasta='.$catfile.', end='.$OptLenMin.'
maxlength=523, group='.$groupsfile.'); unique.seqs(fasta='.$catfile.');
align.seqs(candidate='.$catfile3.', template=~/Analysis/silva.bacteria/silva.bacteria.fasta)\"');
#system ($command2);
#$catfile = $catfile3; $catfile=~ s/\.txt/\.align/; $namefile=$catfile2; $namefile=~
s/\.txt/\.names/;
#Then summary to see what to trim with trimaligns
exit;
Help:
system ("clear");
print"my454 is intended to:\n*\tOptimally take raw 16s amplicon files through to clustered
material\n";
print"*\tLocate and remove primer sequences\n";
print"*\tDetect and reverse complement reverse reads (currently detects 27f and 534r)\n";
print"*\tParse files removing duplicates and concatenate while forming groups file\n\n";
print"USAGE: my454 (options) -a [association file]; -o [output prefix]\n";
print "Return for options"; <STDIN>; system ("clear");
print"Options:  \n";
print"-c \t[(y/n) overlook missing sample files - default = n] \n\n";
print"-l \t[(minimum length (1-1000) -space- maximum length (200-1000)) initial minimum length
cutoff = 200nt and maximum 535nt] \n\n";
print"-h \t[(number) maximum homopolymers tolerated - default = 6] \n\n";
print"-k \t[(all/legible) all - keep all reads in one file \n";
print"  \tallRC - keep all and expect sequences to be reverse complement (default)\n";
print"  \tlegible - keep all but those with unclear primer sequence\n";
print"   \t]\n\n";
print"-r [do not remove primers - default does remove primers] \n\n";
print"-global [remove/label] - used to trigger a global search for duplicates if it is suspected
that identical read names will occur in different files\n\n";
print"-m [maximum number of reads to keep in any one sample]\n\n";
print"-ambig [keep sequences with ambiguous characters]\n\n";
print"-trimlong [Trim the 3' region of sequences that exceed the upper defined length limit (see
-l)] \n\n";
print "Return for comming options";
<STDIN>; system ("clear");
print"-q (REMOVED - MAY FIX IN FUTURE) [(Ave/Clip-Hard/Soft  (number) number1-40) Ave - remove
reads when average qual score falls below set quality number) \n";
print"  *\tClip - clips reads when falls below set qual threshold\n";
print"  *\tHard - Clips from first instance of below threshold score(with Clip) or Removes read
(with Ave)\n";
print"  *\tSoft - Clips when a run of (number) nucleotides falls below threshold (with clip) or
Clips (with Ave)\n";
print"  *\t-default is no assessment of quality - this is currently very slow!\n";
print" Return for notes\n";
<STDIN>; system ("clear");
print"NOTE:\n";
print"Association file should be formated [Sample file] -tab- [Group]\n";
print"If -q is selected their should be a third column with the quality file details\n";
print"To split reads/groups into separate files insert: Fsplit [tab] [next files name] in the
association file where you'd like this to occur.\n\n";
exit;
```

Code S2. The R code and formatted soruce used to generate the heat map.

```
heatmap 7-30-2012 genera
> Abund<-read.table("Dexter_genera.txt", header=T)
> attach(Abund)
> names(Abund)
> row.names(Abund) <- Abund$Species
> Abund<-Abund[,2:25]
> Abund_matrix<-data.matrix(Abund)
>library(gplots) #gplots needs to be attached before using
>library(RColorBrewer)
>Heatmap<-heatmap.2(Abund_matrix, Rowv=F, Colv=T, dendrogram='column', trace='none', breaks=c(0,
0.05, 0.1, 0.2, 0.4, 0.5, 0.6, 0.8, 0.9, 1.0), density.info='none', margins=c(5,12),
col=rev(brewer.pal(9, 'YlOrRd')))

> Abund<-read.table("Dexter_Species.txt", header=T)
> attach(Abund)
> names(Abund)

> row.names(Abund) <- Abund$Species
> Abund<-Abund[,2:25]
> Abund_matrix<-data.matrix(Abund)
>library(gplots) #gplots needs to be attached before using
>library(RColorBrewer)
>Heatmap<-heatmap.2(Abund_matrix, Rowv=F, Colv=T, dendrogram='column', trace='none', breaks=c(0,
0.05, 0.1, 0.2, 0.4, 0.5, 0.6, 0.8, 0.9, 1.0), density.info='none', margins=c(5,12),
col=rev(brewer.pal(9, 'YlOrRd')))

Genera  Pig_252 Pig_254 Pig_256 Pig_258 Pig_282 Pig_284 Pig_352 Pig_354 Pig_356 Pig_358 Pig_362 Pig_381
        Pig_454 Pig_456 Pig_458 Pig_554 Pig_556 Pig_558 Pig_562 Pig_652 Pig_654 Pig_656 Pig_682 Pig_752
        Pig_754
Lactobacillus  22.42587601    21.86927306    27.60059613    32.40677966    40.20255394
        29.05617337    21.30572717    25.75092811    17.75510204    23.45415778    32.5770797
        22.51497006    36.66134185    25.87859425    18.97923111    32.10278094    31.95395146
        12.70516717    16.73828525    29.77542266    24.03622251    27.48588797    27.69272827
        24.64771913    19.81050818
Streptococcus  15.17520216    20.4642639     23.15946349    25.11864407    9.863496257
        13.68691189    15.71213892    17.65102936    12.85714286    22.60127932    21.5823153
        18.17964072    19.95207668    16.86444546    18.53733981    24.16653114    21.03298071
        15.68389058    24.73910374    24.67827403    25.87322122    16.3048198     10.93493712
        11.89395749    32.07005455
Actinobacillus 7.385444744    7.544288332    7.302533532    7.593220339    4.931748129
        5.788423154    4.408081483    13.60107999    6.693877551    10.20788913    5.293775451
        11.4491018     4.361022364    8.238247376    6.650463986    6.407545942    8.742999378
        15.71428571    8.041743401    5.021448398    5.769728331    10.07381676    11.67304538
        10.38930021    5.139247775
Turicibacter   15.87601078    7.727550397    10.01490313    3.762711864    5.019815059
        3.222127174    30.30556019    1.383732703    11.55102041    2.985074627    3.170447935
        7.592814371    3.833865815    6.891830215    17.78612461    3.415189462    3.204729309
        3.525835866    1.984857786    6.358819076    4.294954722    3.517151541    9.896118097
        7.332218772    4.335343095
Enterobacter   4.043126685    5.192425168    2.414307004    2.813559322    1.959489212
        7.014542344    2.972115545    2.868714141    4.489795918    8.02238806     6.369982548
        3.041916168    5.367412141    8.694659973    8.418029165    2.618311921    6.813939017
        12.6443769     8.921628811    5.955084532    2.871927555    7.729049066    2.679059595
        2.627179365    2.038472581
Pasteurella    2.102425876    9.865607819    1.31147541     3.355932203    6.010568032
        6.672369547    3.623309401    3.273709079    14.20408163    4.317697228    3.461314718
        5.964071856    5.52715655     6.869009585    3.203711887    3.659131566    3.422526447
        5.714285714    2.987517905    3.456977038    6.520051746    2.887537994    3.30781848
        5.278242178    5.311513063
```

```
Escherichia_Shigella   5.148247978    3.634697618    5.871833085    6.576271186    2.003522677
       3.051040776    3.523125731    4.826189673    3.632653061    5.170575693    2.065154159
       1.724550898    2.092651757    2.213601095    3.999116217    7.871198569    5.071561917
       3.556231003    4.52220176     3.103709311    4.734799483    3.864524533    3.909240022
       4.633389061    6.316393913
Veillonella    3.692722372    4.153940134    3.48733234     2.779661017    2.708058124
       2.053036784    2.6214727      3.138710766    3.102040816    4.371002132    2.675974404
       3.041916168    1.182108626    3.08078503     2.960671675    3.756708408    2.800248911
       5.592705167    5.709023941    1.867272268    3.234152652    6.969170647    3.471842537
       4.107953188    1.923629055
Lactococcus    3.58490566     3.97067807     3.666169896    2.440677966    2.531924262
       3.108069575    2.237435298    2.396220047    3.87755102     3.331556503    4.392088424
       2.80239521     3.530351438    2.647193063    1.745470614    2.650837535    2.457996266
       4.012158055    5.320237364    3.179409538    5.174644243    3.104646114    2.159650082
       2.961547647    3.01464255
Haemophilus    3.07277628     1.954795357    2.503725782    1.661016949    2.642007926
       4.334188765    1.953581566    2.53121836     3.387755102    2.50533049     2.646887725
       2.107784431    1.102236422    2.099497946    3.513035793    2.471946658    1.835718731
       3.009118541    2.537343974    2.069139541    2.147477361    2.366478506    1.120831055
       1.862909004    2.985931668
Pseudomonas    2.291105121    1.191203421    2.205663189    1.457627119    5.966534566
       5.474764756    0.617799299    1.349983125    1.387755102    2.185501066    2.414194299
       1.02994012     2.587859425    2.441807394    0.441891295    1.12213368     1.711263223
       3.860182371    3.35584203     1.665404996    2.613195343    1.411202779    0.847457627
       1.122522092    2.555268447
Sporacetigenium        1.266846361    1.099572389    0.953800298    1.152542373    1.144870101
       2.309666382    0.684588412    1.586230172    1.510204082    1.465884861    3.25770797
       2.251497006    5.495207668    2.898219991    1.723376049    1.219710522    2.240199129
       2.127659574    1.636996112    3.028009084    1.112548512    2.800694746    1.831601968
       2.268927633    3.215618719
Klebsiella     1.644204852    2.382406842    1.013412817    0.983050847    1.761338617
       1.19760479     1.068625814    0.438744516    3.87755102     1.092750533    2.792321117
       0.910179641    1.916932907    2.327704245    1.811754308    1.219710522    1.586807716
       1.975683891    3.806015961    1.993439314    3.285899094    0.890143291    2.897758338
       4.060186291    2.354292277
Weissella      2.210242588    0.946854001    1.460506706    0.440677966    4.645530603
       2.224123182    1.085323092    2.058724266    1.265306122    0.506396588    2.90866783
       2.634730539    0.591054313    1.118210863    2.209456474    0.455358595    0.622277536
       1.398176292    2.537343974    1.56447136     1.992238034    1.042118975    4.018589393
       3.678051111    0.60292851
Clostridium    1.078167116    1.405009163    0.86438152     1.118644068    1.475121092
       2.50926718     0.768074804    3.509956126    1.87755102     1.57249467     1.657940663
       1.173652695    2.332268371    2.122318576    0.972160848    1.463652626    1.275668948
       1.428571429    0.859422959    2.043906132    0.905562743    1.671732523    1.366867141
       1.528540721    0.660350273
Leuconostoc    2.587601078    1.038485034    1.371087928    1.050847458    0.946719507
       0.598802395    2.654867257    1.653729328    3.265306122    0.479744136    0.17452007
       1.748502994    0.111821086    0.616157006    3.579319487    0.520409823    0.840074673
       0.881458967    1.493758952    0.656068635    0.465717982    1.085540599    1.995626025
       2.030093145    1.80878553
Prevotella     0.754716981    0.427611484    0.149031297    0.542372881    0.572435051
       2.765896778    0.183670062    3.746203172    0.367346939    0.559701493    0.087260035
       1.053892216    0.047923323    0.410771337    0.198851083    0.390307367    0.622277536
       0.759878419    0.266011868    0.100933636    0.181112549    0.998697351    0.656096227
       0.310484834    0.746482917
Peptostreptococcus     0.539083558    0.152718387    0.208643815    0.101694915    1.210920299
       0.998003992    0.80146936     0.067499156    0.163265306    0.26652452     0.058173357
       0.359281437    0.063897764    0.251026928    0.353513036    0.032525614    0.186683261
       0.060790274    0.081849806    0.151400454    0.051746442    0.130264872    4.920721706
       3.200382135    0.488084984
Neisseria      0.754716981    0.794135614    0.715350224    0.847457627    0.506384852
       0.085543199    0.60110202     0.641241984    0.530612245    0.959488273    0.232693426
```

0.191616766 0.191693291 0.798722045 0.574458683 0.520409823 0.528935905
1.063829787 0.73664825 0.277567499 0.336351876 1.541467651 0.382722799
0.692620014 0.172265289
Enterococcus 0.700808625 0.427611484 0.655737705 0.508474576 0.814619111
0.741374394 0.283853732 0.742490719 0.244897959 0.772921109 0.319953461
1.173652695 1.006389776 0.319488818 0.287229342 0.37404456 0.373366521
0.820668693 0.409249028 0.353267726 0.724450194 0.412505428 0.464734828
0.788153809 0.746482917
Salmonella 0.269541779 0.549786194 0.655737705 0.881355932 0.154117129
0.0285144 0.40073468 0.168747891 0.734693878 0.692963753 0.785340314
0.191616766 0.159744409 1.049748973 0.41979673 0.504147016 0.342252645
0.911854103 0.900347862 0.630835226 0.879689521 1.042118975 0.300710771
0.38213518 0.976169968
Raoultella 0.916442049 0.091631032 0.149031297 0.271186441 0.374284456
0.142571999 0.751377525 2.058724266 0.530612245 0.053304904 0
0.45508982 0.111821086 0.182565039 0.463985859 0.292730525 0.280024891
0.030395137 0.020462451 0.227100681 0.155239327 0.173686496 0.601421542
0.692620014 0.229687051
Fusobacterium 0.566037736 0.335980452 0.298062593 0.271186441 0.24218406
0.370687197 0.133578227 0.438744516 0.244897959 0.186567164 0.116346713
0.071856287 0.159744409 0.159744409 0.088378259 0.341518946 0.217797138
0.486322188 0.593411091 0.328034317 0.620957309 0.499348676 0.109349371
0.167184141 0.086132644
Gemella 0.161725067 0.061087355 0.149031297 0.406779661 0 0 0.050091835
0.776240297 0.163265306 0.106609808 0.087260035 0 0.015974441
0.02282063 0.088378259 0.146365263 0.186683261 0 0.040924903
0.050466818 0.155239327 0.086843248 0.027337343 0.095533795 0.172265289
Paralactobacillus 0.134770889 0.12217471 0.178837556 0.13559322 0 0
0.083486392 0.202497469 0.040816327 0.186567164 0.087260035 0.047904192
0.095846645 0.06846189 0.022094565 0.308993332 0.062227754 0.151975684
0.225086965 0.227100681 0.103492885 0.21710812 0.082012028 0 0
Citrobacter 0.269541779 0.091631032 0.029806259 0.06779661 0 0
0.050091835 0.53999325 0 0.079957356 0.029086678 0.071856287 0
0.091282519 0.044189129 0.032525614 0.155569384 0.182370821 0.081849806
0 0.025873221 0.043421624 0.027337343 0.047766898 0.430663221
Proteocatella 0.188679245 0.030543677 0.029806259 0.033898305 0.044033465
0.0285144 0.016697278 1.181235235 0 0.106609808 0 0 0
0 0 0.016262807 0.09334163 0.060790274 0.040924903 0 0
0.130264872 0.027337343 0 0
Pigmentiphaga 0.134770889 0.030543677 0.059612519 0 0 0 0.033394557
0 0 0 0.145433392 0 0 0.04564126 0 0 0
0.030395137 0.14323716 0.075700227 0.20698577 0.021710812 0 0
0
Bibersteinia 0.134770889 0.030543677 0.149031297 0 0.022016733 0.114057599
0.016697278 0 0 0 0.029086678 0 0 0 0 0
0 0.030395137 0.122774708 0.050466818 0.129366106 0.021710812 0
0 0
Roseburia 0.269541779 0 0.149031297 0 0 0 0.016697278
0.134998313 0.040816327 0.026652452 0 0 0.031948882 0.02282063
0.022094565 0 0.031113877 0 0.020462451 0 0 0.021710812
0 0 0

Code S3: The SAS code for bacterial relative abundance analysis

```
data lactobacilli;
input diet$ delivery$ bacteria;
datalines;
f       c       1.200741385
f       c       0.888041846
f       c       0.583636909
f       c       0.803376468
f       c       0.63295859
f       v       1.481522316
f       v       0.141052205
f       v       0.501120626
f       v       0.533414802
f       v       0.297729395
P       c       1.000646753
P       c       0.575500963
P       c       0.83833457
P       c       1.250081331
P       c       0.546191081
P       v       1.062620351
P       v       0.474955193
P       v       0.505791352
P       v       0.547262091
s       c       0.700687717
s       c       0.508142678
s       c       0.995464869
s       v       0.880402782
s       v       0.865235415
s       v       0.637023473
;
run;
proc sort;
by diet delivery;
proc means n mean std stderr min max data=lactobacilli;
title 'BMSII_ileum_microbiota_LOGTuricibacter';
by diet delivery;
var bacteria;
run;
proc sort;
by diet;
proc means n mean std stderr min max data=lactobacilli;
by diet;
var bacteria;
run;
proc sort;
by delivery;
proc means n mean std stderr min max data=lactobacilli;
by delivery;
var bacteria;
run;

proc glm data=lactobacilli;
class diet delivery;
model bacteria= diet delivery diet*delivery /ss3;
lsmeans  diet delivery diet*delivery  / pdiff adj=Tukey Lines;
output out=inf p=pred r=resid student=stu student=stures rstudent=rst h=leve cookd=cook
dffits=dfit;/*check outliers, influencial*/
run;

proc print data=inf;
```

```
run;
proc univariate normal plot data =inf;
var stures;
run;
proc plot data=inf;
plot stures*pred;
run;
data inf; set inf;
absres=abs(resid);
proc corr data=inf spearman;
var absres pred;
run;
proc plot data=inf;
plot absres * pred;
run;
```