ABSTRACT

SCHROETER, JOEL THOMAS. Genomic Analysis of *Lactobacillus gasseri* Strains using Suppressive Subtractive Hybridization. (Under the direction of Dr. Todd Klaenhammer).

Lactobacillus gasseri is a commensal bacterium found in the human GI tract and is considered important in the maintenance of gut health and immunomodulation. The neotype strain, *L. gasseri* ATCC 33323 was previously sequenced and its complete genome was recently published. Due to the significant genomic variability between commensal and probiotic strains, genotypic comparisons are expected to identify novel genotypes and also define the basis of genomic variability. Suppressive subtractive hybridization (SSH) is a technique used to separate DNA sequences present in an uncharacterized strain of interest that are absent in a reference strain with a sequenced genome. The goal of this study was to identify genes related to intestinal survival and activity among a group of six human *L. gasseri* strains using the SSH method. A clonal library of unique fragments was constructed for each unsequenced strain and a series of clones were sequenced to reveal a variety of genes potentially related to cellular processes important for intestinal survival including adhesion and oligosaccharide metabolism. The subtractions also yielded a significant number of plasmid encoded and phage related genes.

Genomic Analysis of *Lactobacillus gasseri* Strains using Suppressive Subtractive Hybridization

by
Joel T. Schroeter

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Master of Science

Food Science

Raleigh, North Carolina

2010

APPROVED BY:

_____	_____
Dr. Robert Kelly	Dr. Trevor Phister

_____
Dr. Todd Klaenhammer
Chair of Advisory Committee

DEDICATION

To the quick-tongued and good-hearted

BIOGRAPHY

Joel Thomas Schroeter was born January 24, 1985 in Minocqua, WI to Thomas and Susan Schroeter. In May 2007 Joel received his bachelors of Science degree from the University of Wisconsin-Madison in Food Science. In 2007 he moved to North Carolina to attend graduate school at NC State where he received the NIH Molecular Biotechnology Training Program Fellowship which supported him in his study of food microbiology within the Food, Bioprocessing, and Nutrition Sciences department under Dr. Todd R. Klaenhammer.

ACKNOWLEDGEMENTS

First and foremost I would like to extend my greatest thanks to Dr. Todd Klaenhammer for allowing me to continue my education under his guidance. I have a great appreciation for his dedication to his work and his students both inside and outside of the laboratory. I thank him most for guiding my work while still allowing me to explore and find answers for myself.

Furthermore, I would like to extend my deepest appreciation to my family and friends that have always and continue to, encourage, support, and inspire me. Nothing I have done, or ever will do, would be possible without their unwavering love.

Lastly, I would like to thank the NIH Molecular Biotechnology Training Fellowship Program for funding my research and enabling me to learn so much during my time at North Carolina State University.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter I

Literature Review

Genomics of Lactic Acid Bacteria

MINIREVIEW

Genomics of lactic acid bacteria

Joel Schroeter & Todd Klaenhammer

Department of Food Science, Bioprocessing and Nutrition Sciences, North Carolina State University, Raleigh, NC, USA

Correspondence: Todd Klaenhammer, Department of Food, Bioprocessing and Nutrition Sciences, North Carolina State University, PO Box 7624, Raleigh, NC 27695-7624, USA. Tel.: +1 919 515 2972; fax: +1 919 513 0014; e-mail: klaenhammer@ncsu.edu

Received 7 October 2008; accepted 6 November 2008.
First published online 12 December 2008.

DOI:10.1111/j.1574-6968.2008.01442.x

Editor: Rustam Aminov

Keywords
lactic acid bacteria; genomics; fermentation; probiotics.

Abstract

Lactic acid bacteria (LAB) are found to occupy a variety of ecological niches including fermented foods as well as mucosal surfaces of humans and other vertebrates. This review is based on the genomic content of LAB that is responsible for the functional and ecological diversity of these bacteria. These genomes reveal an ongoing process of reductive evolution as the LAB have specialized to different nutritionally rich environments. Species-to-species variation in the number of pseudogenes as well as genes directing nutrient uptake and metabolism reflects the adaptation of LAB to food matrices and the gastrointestinal tract. Although a general trend of genome reduction was observed, certain niche-specific genes appear to be recently acquired and appear on plasmids or adjacent to prophages. Recent work has improved our understanding of the genomic content responsible for various phenotypes that continue to be discovered, as well as those that have been exploited by man for thousands of years.

Introduction

We are exposed to a huge variety of microorganisms on a daily basis; one group of bacteria that humans have developed a particularly intimate relationship with are the lactic acid bacteria (LAB). The LAB group is composed of microaerophilic, nonsporulating rods and cocci that are functionally linked by their common capacity to produce primarily lactic acid from hexose sugars (Makarova & Koonin, 2007). The vast diversity of LAB allows them to inhabit a variety of ecological niches ranging from food matrices such as dairy products, meats, vegetables, sourdough bread, and wine to human mucosal surfaces such as the oral cavity, vagina, and gastrointestinal tract (Pfeiler & Klaenhammer, 2007). The metabolic characteristics of LAB have been exploited for the preservation of foods and have been passed down from generation to generation through food 'traditions' that continue to flourish in many cultures to this day. Foods fermented using LAB are still widely consumed, the sales of fermented foods reaching tens of billions of dollars per year, worldwide. Recently, commensal LAB have been given increased attention due to evidence suggesting their important roles in the maintenance of health and the prevention of infection (Reid *et al.*, 2003; Klaenhammer *et al.*, 2005).

Currently, 31 complete LAB genomes have been sequenced and are publicly available. Furthermore, considerable comparative and functional genomic analyses have accompanied the appearance of genomic sequence information. The goal of this review is to highlight some of the interesting outcomes from recent work relating to the genomics of LAB and to put these into context with the relationship between LAB and humans.

General genome features and history

The availability of sequenced genomes has allowed for a deeper understanding of the evolutionary divergence of the LAB, and reveals a trend of relatively recent and ongoing reduction in genome size (van de Guchte *et al.*, 2006). The last common ancestor of *Lactobacillales* appears to have lost *c.* 600–1200 genes and gained < 100 during its divergence from the *Bacilli* ancestor (Makarova & Koonin, 2007). The extent of genome reduction varies greatly among LAB with *Oenococcus oeni* having only *c.* 1700 predicted ORFs compared with the *c.* 3000 of *Lactobacillus plantarum* (Pfeiler & Klaenhammer, 2007). Analysis of the available genomes of LAB suggests that the bulk of the genes lost were due to adaptation to nutrient-rich food environments, particularly

those organisms that have adapted to milk and other food environments rich in protein and carbohydrates. The yogurt bacterium *Lactobacillus delbrueckii* ssp. *bulgaricus* shows a large difference in G–C% content (49.7%) from the closely related species *Lactobacillus acidophilus* (34.7%), a gastrointestinal commensal organism. Interestingly, the difference was primarily in the less conserved third codon position, which had a 65% G–C content implying rapid ongoing evolution to a higher G–C content. Furthermore, the number of rRNA and tRNA genes in *L. bulgaricus* is *c.* 50% higher than the average for a genome of its size. These numbers would correspond to a genome of 3–4 Mb, significantly larger than its actual size of 1.8 Mbp (van de Guchte *et al.*, 2006). The specialized adaptation to milk is particularly interesting because this fermentation environment would not exist without human intervention. The selective pressure came not only from the natural environment, but also from anthropogenic environments created by humans, which essentially domesticated these organisms over the last 5000 years through repeated transfer of LAB cultures for production of fermented dairy products.

Pseudogenes

Further evidence of the recent and ongoing genome reduction of LAB is the presence of pseudogenes, often in relatively high numbers compared with other groups of bacteria. This feature is particularly common in organisms that are associated with nutrient-rich food environments such as *L. bulgaricus*, *Lactobacillus helveticus*, *Lactobacillus lactis*, *Streptococcus thermophilus*, and *O. oeni*. The yogurt cultures *L. bulgaricus* and *S. thermophilus* are reported to have the largest number of pseudogenes, 270 and 182, respectively, showing evidence of recent specialization to the nutrient-rich milk environment (Makarova *et al.*, 2006; van de Guchte *et al.*, 2006). A study of *S. thermophilus* strains found that the most decayed genes were predicted to encode proteins for carbohydrate transport and metabolism. For example, phosphotransferase system transporting proteins responsible for the uptake of glucose, fructose, β-glucoside, and trehalose were pseudogenes in all eight strains of *S. thermophilus* studied, while these were all present in a functional form in *Streptococcus salivarius*, a closely related oral commensal organism. Alternatively, a specific lactose symporter is present in *S. thermophilus* but is missing in other streptococci species (Bolotin *et al.*, 2004).

This trend of genome reduction and specialization is also observed in *L. bulgaricus*, which has several incomplete transport systems along with pseudogenes involved in carbohydrate metabolism, amino acid/cofactor biosynthesis, and competence (van de Guchte *et al.*, 2006). Gains in lactose transport and loss of other carbohydrate transport and metabolism genes, as well as the loss of amino acid biosynthesis genes, provide evidence for the adaptation to milk, an environment rich in lactose and protein.

Transporters

Although a large number of transport systems appear to be lost in some LAB, transporters still make up 13–18% of their genomes, a number larger than what is found in many other bacteria. The diverse environments occupied by LAB require the ability to transport and utilize a variety of substrates in order to survive. The large number of transporters corresponds to the adaptation to nutrient-rich environments and subsequent loss of biosynthetic pathways. Amino acid transporters represent the largest number of uptake systems, which also include sugar, cation/anion, and peptide transporters (Lorca *et al.*, 2007). These transporters are particularly important for low G–C% LAB, which are highly auxotrophic and must scavenge nutrients from the environment (Klaenhammer *et al.*, 2005). Alternatively, a more metabolically capable organism such as *L. plantarum* retained a larger number of sugar uptake systems and complete pathways for biosynthesis of most amino acids. The maintenance of these systems reflects the ability of this organism to metabolize a wide variety of substrates in a primarily plant environment that is not as nutritionally rich (Kleerebezem *et al.*, 2003). Specific transporters have been identified in some LAB that are needed for survival in specialized environments such as the gastrointestinal tract. For example, gastrointestinal-related organisms such as *L. acidophilus* and *Lactobacillus paracasei* contain transporters able to take up fructo-oligosacharides, a group of nondigestible carbohydrates found primarily in the gastrointestinal tract (Barrangou *et al.*, 2003; Altermann *et al.*, 2005; Goh *et al.*, 2006).

Probiotic and gastrointestinal-related genes

In addition to the adaptation of LAB to various food environments, a great deal of interest is currently being paid to those LAB that have adapted to mucosal surfaces. Much of the interest in these organisms is due to relatively recent evidence indicating the important role they play in the maintenance of health. These organisms, referred to as probiotics, are defined as 'Live microorganisms which when administered in adequate amounts confer a health benefit on the host' (FAO/WHO, 2001). Some potential beneficial effects of probiotics include exclusion of pathogens, mucosal immunomodulation, and reduction of carcinogens (Reid *et al.*, 2003). A variety of LAB cultures are now being added to foods and consumed as supplements for the sole purpose of eliciting health-promoting benefits. The genome sequence of probiotic organisms has shed light on important functions needed for survival in the gastrointestinal tract.

Genes have been identified that encode for proteins involved in probiotic functions including acid/bile tolerance, surface proteins/adherence, gene transfer, and carbohydrate utilization (Klaenhammer *et al.*, 2005).

The human gut microbiota is extraordinarily complex and is receiving increased attention for its role in maintenance of health. The microbial load of the human gut reaches 10^{12}–10^{14} CFU g^{-1} of luminal content and is important for digestion, protection against pathogens, and maintenance of mucosal immunity. Mucosal surfaces such as the gastrointestinal tract are the places where the body comes in contact with the majority of antigens and infectious agents. The gastrointestinal tract produces 70–85% of the immune cells of the body and balances anti- and proinflammatory factors to respond to the presence of these antigens. Circumstances such as intestinal infection, changes in diet, and antibiotic treatment can influence the makeup of the intestinal microbiota (Candela *et al.*, in press). Although lactobacilli and other LAB make up a small portion of the total gastrointestinal microbial community, they are predominant microbiota in the small intestine and considered to play a pivotal role in its protection (Heilig *et al.*, 2002; Hayashi *et al.*, 2005).

Adhesion to intestinal cells is considered an important component of a probiotic strain because it allows the organisms to persist in the intestinal tract and potentially exclude pathogens. Many probiotic species, including *L. acidophilus*, *Lactobacillus johnsonii*, and *Lactobacillus gasseri*, encode a series of cell surface mucus-binding proteins that bind mucin glycoproteins. These species also contain a predicted fibronectin-binding protein FbpA able to bind to fibronectin, another cell surface protein (Pridmore *et al.*, 2004; Altermann *et al.*, 2005; Klaenhammer *et al.*, 2005). Interestingly, *L. helveticus* that appears to have diverged from other lactobacilli via adaptation to a milk environment does not contain any mucus-binding proteins and encodes less than half the cell wall proteins of the closely related gastrointestinal commensal *L. acidophilus* (Callanan *et al.*, 2008). Another study in *L. johnsonii* using sequencing and microarray data identified specific genes involved in increased persistence in the gastrointestinal tract. Three genes were identified that were important for a long gut persistence phenotype: two genes within an operon for a mannose-specific transporter and one gene with similarity to an immunoglobulin A protease. Reduced gastrointestinal transit time was observed when each of the three genes were deleted independently (Denou *et al.*, 2008).

In addition to competitive exclusion as one means to inhibit intestinal pathogens, some probiotic LAB also produce compounds that inhibit the growth of pathogens. Bacteriocins, for example, are small peptides produced by some bacteria that are toxic to other competing microorganisms. A recent study revealed that *Lactobacillus salivarius* UCC118 was able to protect mice from infection by *Lysteria monocytogenes*, a common food-borne pathogen. The protective effect was clearly shown to be due to production of a specific bacteriocin that was lethal for *L. monocytogenes* (Corr *et al.*, 2007). Bacteriocins, including those produced by LAB, are now being applied as food preservatives due to their ability to selectively inhibit pathogens such as *L. monocytogenes*, *Bacillus cereus*, *Clostridium botulinum*, and *Staphylococcus aureus*, common pathogens causing food-borne infections (Gálvez *et al.*, 2007; Sit & Vederas, 2008).

Resistance to bile is another trait observed in gastrointestinal-related LAB. Bile-specific hydrolases and transporters have been found in many probiotic species such as *L. johnsonii*, *L. acidophilus*, *Lactobacillus reuteri*, and *L. plantarum* (Leer *et al.*, 1993; Pridmore *et al.*, 2004; McAuliffe *et al.*, 2005; Pfeiler *et al.*, 2007; Martoni *et al.*, 2008). Microarray expression analysis of *L. acidophilus* NCFM found 289 genes that were differentially expressed in the presence of bile. Of these genes, 168 were downregulated while 78 were upregulated. The upregulated genes encoded for genes involved in carbohydrate uptake/metabolism, stress responses, and adhesion to intestinal cells (Pfeiler *et al.*, 2007). Notably, genes for lactose metabolism were induced in the presence of bile suggesting a close evolution of mammalian gastrointestinal tract and a resident microbiota adapted for utilization of milk. A study in *L. reuteri* identified an operon containing a multidrug resistance transporter and a gene of unknown function differentially expressed in the presence of bile. When these genes were deleted, the strain demonstrated a reduced ability to recover in the presence of bile, suggesting that transport of bile salts plays an important roll in bile tolerance (Whitehead *et al.*, 2008). LAB have maintained and acquired a variety of genes that allows to them to survive and interact with the human gastrointestinal tract in an amazingly complex way.

Plasmids

Plasmids are found in many LAB and vary in size and gene content. Important plasmid-encoded genes in *Lactococcus* species were discovered decades ago and include genes involved in lactose/galactose utilization, proteolysis, oligopeptide transport, bacteriophage resistance, citrate utilization, bacteriocin production, and stress response (McKay, 1983; Gasson, 1990). Many of the plasmid-encoded genes have functions not found previously in *L. lactis* and have significantly different G–C% contents. These features of plasmid-encoded genes suggest recent acquisition via horizontal gene transfer from other dairy organisms such as enterococci, streptococci, and lactobacilli. The common dairy strain *L. lactis* ssp. *cremoris* SK11 contains four plasmids that were sequenced and shown to encode a large

number of genes that reflect the organisms adaptation to the milk environment and confirm their critical roles in the industrial performance and fermentative ability of *L. lactis* in cheese and cultured dairy products (Siezen *et al.*, 2005). The probiotic strain *L. salivarius* contains the only characterized megaplasmid in the LAB, 242 kb and comprising *c.* 11% of the genome. Although the megaplasmid does not encode any essential single copy genes, it does encode important genes for additional amino acid and carbohydrate metabolism. The megaplasmid also encodes genes potentially important to the probiotic status of this organism including the ABP118 bacteriocin, a bile salt hydrolase, and a putative conjugation locus. Functions were determined for many of the genes on the megaplasmid; however, the megaplasmid also contains a relatively large number of pseudogenes (Claesson *et al.*, 2006). Plasmids are vehicles for rapid genetic transfer and typically encode genes that are important to a particular strain and its competitiveness in a specific environment.

Prophage and clustered regularly interspaced short palindromic repeats (CRISPRs)

Bacteriophages present a significant challenge in industrial fermentations using LAB. Phage and phage remnants are found in the genomes of most LAB and play a prominent role in species-to-species and strain-to-strain variability. Prophage and remnants can also encode genes directing phenotypes important for host survival or functions. Notable examples are virulence-related genes in pathogenic *Streptococcus pyogenes* and bacteriophage resistance in dairy-related organisms such as *S. thermophilus* and *L. lactis*. (Canchaya *et al.*, 2003; Ventura *et al.*, 2006; Wegmann *et al.*, 2007).

CRISPRs have been observed in the genomes of a number of LAB (Barrangou *et al.*, 2007). CRISPR sequences generally contain a series of short palindromic repeats separated by spacer sequences and adjacent to CRISPR associated, or *cas*, genes. Although the function of these structures is not fully understood, the spacer regions share significant homology with foreign DNA elements. It is suspected that these regions are involved in protecting the host from invasion by potentially harmful foreign DNA, including that from bacteriophages and plasmids. Evidence suggests that the CRISPR regions and *cas* genes provide a type of phage immunity via an RNA interference mechanism. This theory is substantiated by the presence of additional spacers in the CRISPR regions of phage-resistant strains of *S. thermophilus*. Furthermore, when *S. thermophilus* strains were challenged with phage, additional spacer regions homologous to the tested phage were observed. This evidence suggests that CRISPR regions act as a type of primitive bacterial immune system against invading DNA.

Conclusions and future issues

As genome sequence and functional genomic information continues to explode, key features of the genomes of LAB continue to be discovered. The analysis of these features leads to a greater understanding of the physiology and metabolism of organisms that are connected so intimately to humans and their food. The role of LAB in food and health continues to expand and evolve as new discoveries are made and new applications are explored. Currently, the LAB are receiving significant attention as vehicles for delivery of biotherapeutics because of their ability to reach the gastrointestinal tract and interact with the host immune system. Potential biotherapeutic applications being explored currently include using LAB for drug and vaccine delivery vehicles, where the drug is produced directly within the gastrointestinal tract in a proximal position to the immune cells present in the human gut (Delcenserie *et al.*, 2008).

Furthermore, researchers are just now beginning to scratch the surface of the complex relationship between humans and their microbiota. As discoveries in metagenomics continue to be made, the picture becomes more and more complex. The make up of the intestinal microbiota has even been linked to fat deposition and obesity in mouse and human models, further illustrating the relationship between host microbiota, diet, and energy balance (Turnbaugh *et al.*, 2006, 2008). The improved understanding of the genomics of LAB not only answers many questions but also raises many new ones, helping to expand our knowledge of their relationship with mankind.

Acknowledgements

Research at NCSU on the genomics of probiotic lactobacilli is supported by the NC Dairy Foundation, Danisco USA Inc., and Dairy Management Inc. J.S. is supported by a NIH-Molecular Biotechnology Training Fellowship.

References

Altermann E, Russell WM, Azcarate-Peril MA *et al.* (2005) Inaugural article: from the cover: complete genome sequence of the probiotic lactic acid bacterium *Lactobacillus acidophilus* NCFM. *P Natl Acad Sci USA* **102**: 3906–3912.

Barrangou R, Altermann E, Hutkins R, Cano R & Klaenhammer TR (2003) Functional and comparative genomic analyses of an operon involved in fructooligosaccharide utilization by *Lactobacillus acidophilus*. *P Natl Acad Sci USA* **100**: 8957–8962.

Barrangou R, Fremaux C, Deveau H *et al.* (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**: 1709–1712.

Bolotin A, Quinquis B, Renault P *et al.* (2004) Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. *Nat Biotechnol* **22**: 1554–1558.

Callanan M, Kaleta P, O'Callaghan J *et al.* (2008) Genome sequence of *Lactobacillus helveticus*, an organism distinguished by selective gene loss and insertion sequence element expansion. *J Bacteriol* **190**: 727–735.

Canchaya C, Proux C, Fournous G, Bruttin A & Brussow H (2003) Prophage genomics. *Microbiol Mol Biol Rev* **67**: 238–276.

Candela M, Perna F, Carnevali P *et al.* Interaction of probiotic *Lactobacillus* and *Bifidobacterium* strains with human intestinal epithelial cells: adhesion properties, competition against enteropathogens and modulation of IL-8 production. *Int J Food Microbiol* **125**: 286–292.

Claesson MJ, Li Y, Leahy S *et al.* (2006) From the cover: multireplicon genome architecture of *Lactobacillus salivarius*. *P Natl Acad Sci USA* **103**: 6718–6723.

Corr SC, Li Y, Riedel CU, O'Toole PW, Hill C & Gahan CGM (2007) From the cover: bacteriocin production as a mechanism for the antiinfective activity of *Lactobacillus salivarius* UCC118. *P Natl Acad Sci USA* **104**: 7617–7621.

Delcenserie VMD, Lamoureux M, Amiot J, Boutin Y & Roy D (2008) Immunomodulatory effects of probiotics in the intestinal tract. *Curr Issues Mol Biol* **10**: 37–54.

Denou E, Pridmore RD, Berger B, Panoff J-M, Arigoni F & Brussow H (2008) Identification of genes associated with the long-gut-persistence phenotype of the probiotic *Lactobacillus johnsonii* strain NCC533 using a combination of genomics and transcriptome analysis. *J Bacteriol* **190**: 3161–3168.

FAO/WHO (2001) Health and nutritional properties of probiotics in food including powdered milk with live lactic acid bacteria. http://www.who.int/foodsafety/publications/fs_management/en/probiotics.pdf

Gálvez A, Abriouel H, López RL & Omar NB (2007) Bacteriocin-based strategies for food biopreservation. *Int J Food Microbiol* **120**: 51–70.

Gasson MJ (1990) *In vivo* genetic systems in lactic acid bacteria. *FEMS Microbiol Rev* **7**: 43–60.

Goh YJ, Zhang C, Benson AK, Schlegel V, Lee J-H & Hutkins RW (2006) Identification of a putative operon involved in Fructooligosaccharide utilization by *Lactobacillus paracasei*. *Appl Environ Microb* **72**: 7518–7530.

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.

Heilig HGHJ, Zoetendal EG, Vaughan EE, Marteau P, Akkermans ADL & de Vos WM (2002) Molecular diversity of *Lactobacillus* spp. and other lactic acid bacteria in the human intestine as determined by specific amplification of 16S ribosomal DNA. *Appl Environ Microb* **68**: 114–123.

Klaenhammer TR, Barrangou R, Buck BL, Azcarate-Peril MA & Altermann E (2005) Genomic features of lactic acid bacteria effecting bioprocessing and health. *FEMS Microbiol Rev* **29**: 393–409.

Kleerebezem M, Boekhorst J, van Kranenburg R *et al.* (2003) Complete genome sequence of *Lactobacillus plantarum* WCFS1. *P Natl Acad Sci USA* **100**: 1990–1995.

Leer RJ, Christiaens H, Verstraete W, Peters L, Posno M & Pouwels PH (1993) Gene disruption in *Lactobacillus plantarum* strain 80 by site-specific recombination: isolation of a mutant strain deficient in conjugated bile salt hydrolase activity. *Mol Gen Genet* **239**: 269–272.

Lorca GL, Barabote RD, Zlotopolski V *et al.* (2007) Transport capabilities of eleven gram-positive bacteria: comparative genomic analyses. *BBA Biomembranes* **1768**: 1342–1366.

Makarova K, Slesarev A, Wolf Y *et al.* (2006) Comparative genomics of the lactic acid bacteria. *P Natl Acad Sci USA* **103**: 15611–15616.

Makarova KS & Koonin EV (2007) Evolutionary genomics of lactic acid bacteria. *J Bacteriol* **189**: 1199–1208.

Martoni CBJ, Urbanska AM & Prakash S (2008) Microencapsulated bile salt hydrolase producing *Lactobacillus reuteri* for oral targeted delivery in the gastrointestinal tract. *Appl Microbiol Biotechnol* **81**: 225–233.

McAuliffe O, Cano RJ & Klaenhammer TR (2005) Genetic analysis of two bile salt hydrolase activities in *Lactobacillus acidophilus* NCFM. *Appl Environ Microb* **71**: 4925–4929.

McKay LL (1983) Functional properties of plasmids in lactic streptococci. *Antonie Van Leeuwenhoek* **49**: 259–274.

Pfeiler EA & Klaenhammer TR (2007) The genomics of lactic acid bacteria. *Trends in Microbiol* **15**: 546–553.

Pfeiler EA, Azcarate-Peril MA & Klaenhammer TR (2007) Characterization of a novel bile-inducible operon encoding a two-component regulatory system in *Lactobacillus acidophilus*. *J Bacteriol* **189**: 4624–4634.

Pridmore RD, Berger B, Desiere F *et al.* (2004) The genome sequence of the probiotic intestinal bacterium *Lactobacillus johnsonii* NCC 533. *P Natl Acad Sci USA* **101**: 2512–2517.

Reid G, Jass J, Sebulsky MT & McCormick JK (2003) Potential uses of probiotics in clinical practice. *Clin Microbiol Rev* **16**: 658–672.

Siezen RJ, Renckens B, van Swam I, Peters S, van Kranenburg R, Kleerebezem M & de Vos WM (2005) Complete sequences of four plasmids of *Lactococcus lactis* subsp. *cremoris* SK11 reveal extensive adaptation to the dairy environment. *Appl Environ Microb* **71**: 8371–8382.

Sit CS & Vederas JC (2008) Approaches to the discovery of new antibacterial agents based on bacteriocins. *Biochem Cell Biol* **86**: 116–123.

Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER & Gordon JI (2006) An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**: 1027–1131.

Turnbaugh PJ, Bäckhed F, Fulton L & Gordon JI (2008) Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell Host Microbe* **3**: 213–223.

van de Guchte M, Penaud S, Grimaldi C *et al.* (2006) The complete genome sequence of *Lactobacillus bulgaricus* reveals extensive and ongoing reductive evolution. *P Natl Acad Sci USA* **103**: 9274–9279.

Ventura M, Canchaya C, Bernini V *et al.* (2006) Comparative genomics and transcriptional analysis of prophages identified in the genomes of *Lactobacillus gasseri*, *Lactobacillus salivarius*, and *Lactobacillus casei*. *Appl Environ Microb* **72**: 3130–3146.

Wegmann U, O'Connell-Motherway M, Zomer A *et al.* (2007) Complete genome sequence of the prototype lactic acid bacterium *Lactococcus lactis* subsp. *cremoris* MG1363. *J Bacteriol* **189**: 3256–3270.

Whitehead K, Versalovic J, Roos S & Britton RA (2008) Genomic and genetic characterization of the bile stress response of probiotic *Lactobacillus reuteri* ATCC 55730. *Appl Environ Microb* **74**: 1812–1819.

Chapter II

Genomic analysis of *Lactobacillus gasseri* strains using suppressive subtractive hybridization

Department of Food, Bioprocessing, and Nutrition Sciences, Southeast Dairy Foods Research Center
North Carolina State University, Raleigh , North Carolina 27695-7624

Introduction

The human body is in constant interaction with the vast diversity of microorganisms that make up an individuals microbiota. Recent studies continue to reveal that microbe-host interactions play a more crucial role in the maintenance of health than previously thought, yet many of the mechanisms remain unknown or poorly understood [22]. The study of genetic and phenotypic diversity among members of the human microbiota can help to shed light on the mechanisms used by these organisms to survive and interact in the human gut and with other mucosal surfaces.

The lactic acid bacterium *Lactobacillus gasseri* is considered one of the true autochthonous organisms associated with human mucosal surfaces and is found commonly in the oral cavity, vaginal cavity, and gastrointestinal (GI) tract of humans [4]. *L. gasseri* has also been recognized for its probiotic potential. Probiotics are defined as being "Live microorganisms which when administered in adequate amounts confer a health benefit on the host" as defined by FAO/WHO. Consumption of specific strains of *L. gasseri* has been shown to have a beneficial impact on general intestinal health as well as a suppressive effect against certain human intestinal pathogens [14,17]. The stain *L. gasseri* ATCC 33323 has recently been sequenced to reveal the genomic content of this species, including numerous genes thought to be important for survival in human mucosal environments. Incidentally, analysis of *L. gasseri* strains has shown a great deal of phenotypic variation in carbohydrate metabolism, bile tolerance, and adhesion characteristics among members of this species [4]. This variation implies that considerable genetic variability occurs among stains of *L. gasseri.*

The goal of this project was to use suppressive subtractive hybridization (SSH) as a way to screen the genomes of six unsequenced strains of *L. gasseri* for sequences absent in the sequenced neotype strain ATCC 33323. The six strains of *L. gasseri* that were the focus of this study were chosen because they are of human origin isolated from either feces or endoscopy samples that displayed widely varying phenotypes related to carbohydrate metabolism, bile tolerance, and adhesion to intestinal cells [4]. These phenotypes were targeted because they have been found to be important features associated with survival, persistence, and host interactions of probiotic microbes [6,21].

This study was conducted with the intent to expand our knowledge of the genomic content of *L. gasseri* at a species level without completely sequencing the genomes of multiple strains. The SSH procedure was used to isolate genes and gene fragments found to be present in the tester strains that were absent in the reference, neotype strain ATCC 33323. The predicted functions of these genes can then be used to develop a genomic basis for phenotypic variation among strains of *L. gasseri*, as well as reveal novel genetic features of this species.

2.1 Materials and Methods

Bacterial Strains and Growth Conditions

The strains of *L. gasseri* used in this study are listed in Table 1. Broth cultures of *L. gasseri* were grown statically at 37°C in MRS media (Difco Laboratories Inc. Detroit, MI). *Escherichia coli* cultures were grown aerobically at 37°C in Luria-Bertani (LB) media (Difco).

Genomic DNA Isolation

Genomic *L. gasseri* DNA was extracted using the method described in Azcárate-Peril and Raya [3] and cleaned using the MO-BIO Ultra Clean Microbial DNA Isolation Kit (MO-BIO Laboratories, Inc. Carlsbad, CA). Genomic DNA was stored at -20° C.

Suppressive Subtractive Hybridization

Suppressive subtractive hybridization was performed using the PCR-Select™ Bacterial Genome Subtraction Kit according to the manufacturers instructions (Clontech Laboratories, Inc, Mountain View, CA). For each each SSH experiment, *L. gasseri* ATCC 33323 was used as the driver strain while one of six *L. gasseri* strains (ML3, FR2, WD19, ADH, SD10, and FR4) served as the tester. For each SSH, 2 µg of both tester and driver DNA were independently digested for 5 hours at 37°C with *RsaI.* Digested DNA was purified using a phenol-chloroform extraction procedure according to the manufacturer's instructions. Tester DNA was then divided into two aliquots and a different adaptor (adaptor 1 and 2R) was ligated to each respective aliquot using T4 DNA ligase (Clontech) overnight

at 16°C. The first hybridization was performed by adding the *RsaI* digested driver DNA to each adaptor-ligated aliquot of tester DNA. The DNA mixture was then denatured at 98°C for 1.5 minutes and hybridized at 60°C for 1.5 hrs. The hybridized DNA aliquots were then mixed together simultaneously with additional freshly denatured driver DNA for the second hybridization reaction. The reaction was incubated at 60°C overnight. A hybridization temperature of 60°C was used instead of the suggested temperature of 63°C because of the low GC% content of *L. gasseri* (35.3%) [12]. The hybridized DNA was then used as a template for two sets of polymerase chain reactions (PCR). The first PCR reaction involves end filling and amplification of all adaptor-ligated products using primers that anneal to both adaptors 1 and 2R using Taq polymerase PCR system (Roche Molecular Biochemicals, Indianapolis, IN). The reaction consisted of 1.0 µL of 1:20 diluted DNA from the second hybridization, 2.5 µl 10x reaction buffer with $MgCl_2$ (Roche), 0.5 µL 10mM dNTP mix (Roche), 1.0 µL PCR Primer 1(Clontech), 0.5 µL Taq Polymerase(Roche), and 19.5 µL nanopure water. The thermal cycling program consisted of a initial step of 72°C for 2 min for adaptor extension followed immediately by 25cycles of (94°C for 30 sec, 66°C for 30 sec, and 72°C for 1.5 min). The secondary PCR reaction consisted of 1.0 µL of 1:40 diluted primary PCR products, 2.5µl 10x reaction buffer with $MgCl_2$ (Roche), 0.5µL 10mM dNTP mix (Roche), 1.0µL Nested Primer 1(Clontech), 1.0 µL Nested Primer 2R(Clontech), 0.5 µL Taq Polymerase (Roche), and 18.5 µL nanopure water. The nested primers used in the secondary PCR reaction are specific to the corresponding adaptor for amplification of hybridization products that consist of a DNA fragment containing one of each adaptor as opposed to two of the same adaptors. The thermal cycling program for the secondary PCR

consisted of 15 cycles of (94°C for 30 sec, 68°C for 30 sec, and 72°C for 1.5 min). The PCR products from both reactions along with unsubtracted controls were analyzed on 2% agarose gels along with 1kb Plus DNA Ladder (Invitrogen Carlsbad, CA). The subtractive hybridization procedure is also described in figure 1.

Cloning and Sequencing of SSH PCR Products

Secondary PCR products were TA cloned using Strataclone™ Cloning Kit (Stratagene La Jolla, CA) vectors and transformed into Solopack™ Competent Cells (Stragagene). *E. coli* transformants were grown on LB media supplemented with 100 µg/ml ampicilin (Sigma Chemical Co., St. Louis, MO) and 40 µg X-gal (5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside) (Sigma) for blue-white screening of transformants. White colonies were selected at random and plasmids were isolated using the Qiagen Miniprep Kit (Qiagen). The SSH generated fragments were sequenced by Genewiz, Inc. (South Plainfield, NJ).

Amplification and Sequencing of Genes Encoding Collagen Binding Proteins

The predicted collagen binding protein of *L. gasseri* MV-22 was used to design primers JS_Collflank1_F and JS_Collflank1_R (Table 1) which were used to amplify products using Expand High Fidelity Plus DNA Polymerase (Roche Diagnostics). Products were purified using the Zymoclean Cleaner Concentrator Kit (Zymo Research) and sequenced using the same primers used for amplification by Eton Biosciences (Durham, NC).

2.2 Results

Subtractive Hybridization

The amplification products from the subtracted DNA of each strain resulted in a smear of DNA fragments ranging in size from ~400-2000bp (Figure 2). Among all of the strains, the unsubtracted DNA smear showed less defined bands and larger range of fragment sizes. The primary PCR products were often visible in the unsubtracted controls compared to the subtracted samples which were often faint or invisible suggesting a lower quantity of amplifiable template as a result of the subtraction process.

The subtracted DNA from each SSH experiment was used to generate clones in pSC-A (Stratagene) that were screened for presence of an insert using blue-white screening in *E. coli* StrataClone SoloPack competent cells (Stratagene). A series of the white colonies were chosen for each of the SSH experiments and the inserts were sequenced. The sequenced fragments were screened against the genome of the reference strain *L. gasseri* ATCC 33323 to determine if the fragments were specific to the tester strain. The tester specific inserts that were sequenced ranged in size from ~200 bp – 1000 bp.

Insert sequences determined to be tester specific (e.g. absent from the ATCC 33323 genome) were searched for protein similarities in the GenBank database using blastx, which searches an amino acid database using a translated nucleotide query. The BLAST analysis revealed hits for genes related to a variety of cellular processes and genomic features. The comprehensive list of BLAST responses is shown in table 2. Significant gene identification

defined as BLAST results with expected values (E-values) $\leq 10^{-3}$ based on previous literature [12]. The proportion of tester-specific inserts varied from experiment to experiment with SD10 showing the lowest (28.6%) and ML3 showing the highest (68.6%) proportion of tester::driver DNA. The specific characteristics of each SSH experiment are listed in table 3. It is unknown whether the differences in the amount of tester-specific inserts was a result of varied levels of genetic difference between the tester and the reference driver strain or from experimental variability that may have affected the subtractive hybridization efficiency.

Identification of homologous genes

A total of 102 tester specific sequences were identified by the BLASTx search of sequenced inserts across the six strains, 76 of which were shown to be different sequences. In addition to the 76 identified sequences, a total of 14 sequences across all six SSH experiments were found to share no significant homology with any sequences in the GenBank database. The tester specific sequences shared homology with a variety of genes including genes involved in carbohydrate metabolism, transport, cell surface compounds, and transcriptional regulation as well as genes related to bacteriophages and transposons. A complete organization of protein catagories based on COG classifications is shown in table 4. The vast majority of the identified genes belonged to other lactic acid bacteria primarily *Lactobacillus* species, the most common of which were *L. gasseri, Lactobacillus casei, Lactobacillus paracasei* subsp. *paracasei,* and *Lactobacillus salivarius*. The majority of the other sequenced fragments identified, were most similar to genes found in other members of the phylum Firmicutes. Fragments also shared homology with genes found in a number of human and animal

pathogens including *Streptococcus equi* subsp. *zooepidoemicus* MGCS 10565, *Bacillus cereus* W, *Salmonella typhimurium,* and *Shigella boydii* CDC 3083-94.

Genes Related to Intestinal Survival and Host-Microbe Interaction

Among the genes identified by BLAST, a series of genes appear to be related to cell adhesion and oligosaccharide metabolism. The adhesion-related genes included an adhesion exoprotein found in strain WD19 as well as a collagen-binding A precursor-like protein found in strain ADH. Additionally, five more genes related to cell surface/membrane structure and function were identified in these two strains. Among the genes related to carbohydrate transport and metabolism, 4 genes were found to be involved in transport and metabolism of oligosaccharides. These genes were involved in transport and metabolism of oligosaccharides and polysaccharides including inulin, cellobiose, oligogalacturonide, and rhamnogalacturonate. These compounds vary significantly but are all primarily plant-base carbohydrates and non-digestible by humans.

Sources of Interstrain Variability

A trend observed among the unique genes identified was that many were related to processes commonly associated with genetic variability and horizontal gene transfer. Features such as transposons, bacteriophage and phage remnants, Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) regions, and plasmids are commonly found to vary significantly between different strains of the same species. These elements were also found in this study to be highly prevalent among the DNA fragments obtained through the SSH procedure.

Screening and Sequence Comparison of Collagen-binding Proteins

A 302 bp region obtained from strain ADH during the subtractive hybridization procedure was found to share the greatest homology with the collagen-binding A precursor like protein in *Lactobacillus casei* str. Zhang. The fragment also shared significant sequence homology with predicted collagen binding proteins in *Lactobacillus paracasei* subsp. *paracasei* and *Lactobacillus gasseri* strain MV-22. The collagen binding protein in *L. gasseri* strain MV-22 was used to generate primers designed screen for the presence of similar collagen binding proteins in the 6 *L. gasseri* strains used in the SSH experiments.

The primers specific to the collagen-binding A precursor like protein were used to amplify strong products in three of the six strains (FR2, ADH, and SD10). These products showed slight variation in size with the ADH amplicon being approximately 200 bp larger than the amplicons in FR2 and SD10. Comparison of the sequences revealed a high level of homology among the collagen binding proteins in each of the three strains with only minor sequence variations; except for a 270 bp INDEL (insertion-deletion) found to be present in the FR2 collagen binding protein and absent from the proteins in ADH and SD10. A sequence comparison of the three proteins is shown in figure 3. This region was found to contain a 224 bp repeat region (figure 3B) that was found to correspond to a Cna protein type B domain by Pfam analysis. The CnaB region is a repeat region commonly found in collagen binding proteins of *Staphylococcus aureus* and make up the stalk structure that projects the collagen binding region of the protein out away from the cell surface [5]. The collagen binding protein from strain FR2 contains two CnaB domains while those from ADH and SD10 only contain one CnaB domain.

2.3 Discussion

Adhesion related genes

Adhesion to mucosal surfaces of the gastrointestinal tract is considered an important feature of intestinal organisms for both persistence in the gut environment and interaction with the host immune system [21]. The six strains of interest used in this study varied in their ability to adhere to Caco-2 and HT-29 MTX cells, which implies differences in genomic content relating to adhesion [4]. Among the genes identified by the SSH procedure, two strains (WD19 and ADH) were found to contain sequences that shared the greatest homology with genes related to adhesion. The adhesion-related genes included an adhesion exoprotein found in strain WD19 as well as a collagen-binding A precursor protein-like protein found in strain ADH. Using primers designed to anneal to the collagen binding A precursor like protein sequence, genes of similar sequence were amplified from 3 of the 7 strains used in the study.

Additionally, four genes related to cell surface/membrane structure and function were identified in these two strains. Strain ADH was shown to have a high level of adhesion to intestinal cells in vitro in comparison to the other strains, whereas WD19 was shown to have a moderate level of adhesion. At this time it is unknown whether either of these identified genes play significant roles in the adherence characteristics of these strains.

Sources of Interstrain Variability

A large number of the sequences isolated by the SSH procedure were identified as common sources of genetic variation and horizontal gene transfer (HGT) among bacterial species. These features included transposons, bacteriophage, plasmids, and CRISPR regions. The amplification of a large number of these features was not unexpected because the SSH procedure selects for regions of DNA that differ between the two genomes. These variable regions represent the most rapidly changing regions of a genome and thus the most likely sequences to be selected for by SSH. The most common features found were bacteriophage related sequences, found in all six strains, followed by transposon related sequences, found in four strains (WD19, ADH, SD10, FR4). Sequences found to be homologous to regions assocated with CRISPR and restriction modification systems were isolated in three strains (ML3, WD19, SD10) [9]. Although these systems are known to protect the cell from invading foreign DNA, CRISPR and restriction modification systems are highly variable among species of lactic acid bacteria making these regions theoretically more likely to be selected by the SSH procedure [11,13].

Transposases

Four of the six strains (WD19, ADH, SD10, FR4) were found to contain transposases or transposase fragments. The presence of transposons is common among almost all bacterial species including lactic acid bacteria, and are sources of horizontal gene transfer and genomic variation [18]. Because transposons are sources of rapid change within a genome sequence, it was reasonable to expect such sequences to be isolated during the SSH procedure. Although the BLAST analysis revealed the presence of transposases in four of the

six strains, only one transposase was located next to other genes found to be absent from the neotype strain. This tranposase, identified in strain WD19, was found to be in close proximity to two hypothetical proteins in *Lactobacillus casei* str. Zhang (Conserved hypothetical protein/putative membrane protein and hypothetical protein LCAZH_p030. At this time it is unknown whether these genes were acquired via transposon integration or if these hypothetical genes are complete or functional. Obtaining additional sequences that flank these transposases may determine whether these transposases are associated with recent gene acquisition.

Bacteriophage and CRISPR

Bacteriophages are also common sources of genetic variability and genetic mobility between strains and species of bacteria including lactic acid bacteria [18]. The genetic variability associated with bacteriophage allowed for the isolation of a large number of bacteriophage-related sequences among these strains of *L. gasseri.* A total of 13 non-redundant bacteriophage related sequences were identified in five strains (ML3, WD19, ADH, SD10, FR4) and ranged in function from phage integrases, phage structural proteins, and CRISPR associated sequences. The isolation of these sequences shows that there is a significant level of phage influence on the genome variability of *L. gasseri* strains. A number of *L. gasseri* strains have previously been shown to contain inducible phage including strains ATCC 33323 and ADH used in this study [7,15]. Phage φadh sequences were identified in strain ADH as expected, however, a series of hypothetical proteins from phage φadh were also identified in strain FR4. At this time, it is unknown whether strain FR4 contains a phage similar to φadh or other inducible phages.

Plasmids

A selection of the sequences isolated during the SSH procedure shared homology with genes found to be present on plasmids within the organisms containing those genes. Although this finding does not indicate the presence of plasmids in any of the *L. gasseri* strains of interest, it does imply increased mobility of these plasmid associated genes. For example, strain FR4 contains a gene that shares 90% homology with a hypothetical protein on the plasmid NTP16 in *Salmonella typhimurium.* Plasmid NTP16 carries, among other features, antibiotic resistance genes as well as genes related to plasmid mobility. Furthermore, this strain also contains a sequence that shares 71% homology with a hypothetical protein from plasmid pBS512_7, a plasmid from *Shigella boydii* CDC 3083-94 that carries a gene for resistance to the bacteriocin colicin as well as mobility proteins. It is interesting that a human commensal organism contains genes that may have been acquired via conjugation of antibiotic and bacteriocin resistance determinants from known human pathogens. It is currently unknown whether strain FR4 contains plasmids or any other parts of these plasmids including the antimicrobial resistance genes or phenotypes.

In addition to plasmids from pathogens, a number of the strains contained sequences that shared the greatest homology with genes located on plasmids from other lactic acid bacteria. Four of the six strains studied (ML3, WD19, ADH, and FR4) contained sequences similar to those from the plasmid plca36 from *Lactobacillus casei* st. Zhang. This 36 bp plasmid has been shown to contain a number of features including: toxin-antitoxin, adhesion, carbohydrate metabolism, and conjugation genes. It has been suggested that the presence of this plasmid contributes significantly to the high acid and bile tolerance characteristics

observed in this strain [23]. The strains containing sequences homologous to this plasmid display varied adhesion and bile tolerance phenotypes and it is unknown how the presence of these sequences impacts these phenotypes among these *L. gasseri* strains. Strain WD-19 was also shown to contain two sequences from plasmid pUCC118-44 of *Lactobacillus salivarius* UCC118. The sequences were most homologous to a hypothetical protein as well as a resolvase located on the 44 kb plasmid. This plasmid, like the *L. casei* plasmid, carries toxin-antitoxin genes as well as genes commonly associated with resistance to bile salts and heavy metals [8]. Interestingly, strain WD-19 was found to have a relatively low bile tolerance profile compared to the other strains used in this study.

Overall, it is difficult to make comparisons between the plasmid-related sequences found in these *L. gasseri* strains to the presence or impact of the plasmids in which they are commonly found without further sequencing to determine the presence of larger fragments of plasmid sequence that may be present. However, all of the plasmids mentioned carry traits that are potentially desirable for the survival of these microbes, particularly for survival in the GI tract due to presence of genes related to adhesion, bile tolerance, and even antibiotic and bacteriocin resistance.

Polysaccharide/Oligosaccharide Metabolism Genes

Strains of *L. gasseri* have been shown to have varied capacities for the metabolism of different carbohydrates [4]. This strain-to-strain variation was reflected in the results of the SSH procedure which yielded 15 sequences that shared the greatest homology to genes related to carbohydrate metabolism. Carbohydrate metabolism related sequences represented

the largest group of identified sequences, second only to hypothetical proteins. This study focused on the metabolism of non-digestible carbohydrates that could be important for survival and persistence in the human GI tract. Some of these carbohydrates are considered "prebiotics" which are defined as "a non-digestible food ingredient that beneficially affects the host by selectively stimulating the growth and/or activity of one or a limited number of bacteria in the colon, and thus improves host health [9]. The SSH procedure was able to isolate a total of four sequences that shared the greatest homology with genes related to metabolism of four different non-digestible carbohydrates. Those carbohydrates were cellobiose, inulin, oligogalaturonide , rhamnogalacturonate. In the cases of inulin, oligogalacturonate, and rhamnogalacturonate the enzymes predicted are involved in synthesis or degradation of these compounds while the enzyme related to cellobiose metabolism was predicted to be part of PTS transport systems. Strain ML3 which contained genes related to metabolism of inulin and cellobiose was not able to ferment inulin but was partially able to ferment cellobiose. Fructosyltransferase genes similar to the inulosucrase found in ML3 have recently been characterized in *Lactobacillus johnsoneii* NCC 533, *Lactobacillus reuteri* and *Lactobacillus gasseri* [1,2,19]. These enzymes have been shown to polymerize sucrose units into levan and inulin fructan products, both of which are considered prebiotic carbohydrates. The enzymes from these organisms have also been used as recombinant enzymes to produce fructan products in situ. It is unknown whether ML3 is able to produce inulin or levan from sucrose. Further analysis of these strains and genes is necessary to determine the role of these identified genes in the metabolism of these carbohydrates.

Collagen Binding Protein DNA Sequence Comparison

The sequence analysis and comparison of genes predicted to encode collagen binding proteins in *L. gasseri* strains revealed a high level of similarity between the genes present in the three strains where the related genes were able to be amplified. The major difference being one 270 bp region found in strain FR2 was missing from strains ADH and SD10 (see figure 4). This region corresponded to a CnaB domain present in collagen binding domains of *Staphylococcus aureus*. This domain is considered to have a structural function in *S. aureus* collagen binding proteins. However, mutants lacking these repeat regions showed no reduced collagen binding capacity [10]. In *S. aureus* these structures are predicted to have a function independent from collagen binding as part of a mosaic protein. It is unknown what the role of these structures are in *L. gasseri* binding proteins [16]. Further analysis is required to determine the role of these proteins in binding to collagen or other substrates.

Conclusions

The results of the suppressive subtractive hybridization experiments reveal the presence of significant genetic variability between strains of *Lactobacillus gasseri*. A number of gene fragments were isolated and determined to be absent from the sequenced neotype strain ATCC 33323. These gene fragments were found to be related to a number of cellular processes including carbohydrate transport and metabolism, bacteriophage and bacteriophage resistance, transposition, and adherence among others. The presence of a large number of sequences associated with genetic variation and horizontal gene transfer were also isolated highlighting the regions of the *L. gasseri* genomes that may be most rapidly changing.

Although the list of genetic differences discovered using this method is far from comprehensive, these results shed light on a number of genetic differences of interest for further study. Features discovered that were related to adherence and oligosaccharide metabolism discovered may impact the ability of these organisms to adhere and survive in the GI tract and thus impact their functions as probiotics. In this regard, these strains were isolated from healthy fecal or endoscopy samples and the differences in these traits may emphasize important chances occurring in the GI vs. domesticated (neotype) stains. Future work will help to reveal the impact of these genes on probiotic and metabolic functions of these strains.

References Cited

1) Anwar, M. A., S. Kralj, et al. (2008). "The Probiotic Lactobacillus johnsonii NCC 533 Produces High-Molecular-Mass Inulin from Sucrose by Using an Inulosucrase Enzyme." Appl. Environ. Microbiol. **74**(11): 3426-3433.

2) Anwar MA, K. S., Piqué AV, Leemhuis H, van der Maarel MJ, Dijkhuizen L. (2010). "Inulin and levan synthesis by probiotic Lactobacillus gasseri strains: characterization of three novel fructansucrase enzymes and their fructan products." Microbiology.

3) Azcárate-Peril, R. R. R. a. M. A. 2001. Methods for Plasmid and Genomic DNA Isolation from Lactobacilli, p. 135-139. Food Microbiology Protocols, vol. 14. Humana Press.

4) Azcarate-Peril, M. A., E. Altermann, et al. (2008). "Analysis of the Genome Sequence of Lactobacillus gasseri ATCC 33323 Reveals the Molecular Basis of an Autochthonous Intestinal Organism." Appl. Environ. Microbiol. **74**(15): 4610-4625.

5) Deivanayagam CC, R. R., Carson M, Owens RT, Danthuluri S, Bice T, Höök M, Narayana SV. (2000). "Novel fold and assembly of the repetitive B region of the Staphylococcus aureus collagen-binding surface protein." Structure **15**: 67-78.

6) Delphine M.A. Saulnier 1, Glenn R. Gibson 1 & Sofia Kolida 1 (2008). "In vitro effects of selected synbiotics on the human faecal microbiota composition." FEMS Microbiology Ecology **Volume 66** (Issue 3): Pages 516 – 527

7) Elsayed A. Ismail1, Horst Neve1, Arnold Geis1 and Knut J. Heller1 (2009). " Characterization of Temperate Lactobacillus gasseri Phage LgaI and Its Impact as Prophage on Autolysis of Its Lysogenic Host Strains." Current Microbiology.

8) Fang, F., S. Flynn, et al. (2008). "Characterization of Endogenous Plasmids from Lactobacillus salivarius UCC118." Appl. Environ. Microbiol. **74**(10): 3216-3228.

9) Gibson, G. R. and M. B. Roberfroid (1995). "Dietary Modulation of the Human Colonic Microbiota: Introducing the Concept of Prebiotics." J. Nutr. **125**(6): 1401-1412.

10) Hartford, O., D. McDevitt, et al. (1999). "Matrix-binding proteins of Staphylococcus aureus: functional analysis of mutant and hybrid molecules." Microbiology **145**(9): 2497-2505.

11) Horvath, P., A.-C. Coûté-Monvoisin, et al. (2009). "Comparative analysis of CRISPR loci in lactic acid bacteria genomes." International Journal of Food Microbiology **131**(1): 62-70.

12) LaGier, D. S. T. (2008). "Identification of novel genes in the oral pathogen Campylobacter rectus." Oral Microbiology and Immunology **23**(5): 406-412.

13) O'Sullivan, O., J. O'Callaghan, et al. (2009). "Comparative genomics of lactic acid bacteria reveals a niche-specific gene set." BMC Microbiology **9**(1): 50.

14) Olivares, M., M. P. Díaz-Ropero, et al. (2006). "Oral administration of two probiotic strains, Lactobacillus gasseri CECT5714 and Lactobacillus coryniformis CECT5711, enhances the intestinal function of healthy adults." International Journal of Food Microbiology **107**(2): 104-111.

15) Raya, R. R., E. G. Kleeman, et al. (1989). "Characterization of the temperate bacteriophage phi adh and plasmid transduction in Lactobacillus acidophilus ADH." Appl. Environ. Microbiol. **55**(9): 2206-2213.

16) Rich, R. L., B. Demeler, et al. (1998). "Domain Structure of the Staphylococcus aureus Collagen Adhesinâ€ " Biochemistry **37**(44): 15423-15433.

17) Sakamoto, I., M. Igarashi, et al. (2001). "Suppressive effect of Lactobacillus gasseri OLL 2716 (LG21) on Helicobacter pylori infection in humans." J. Antimicrob. Chemother. **47**(5): 709-710.

18) Scott, J. R. and G. G. Churchward (1995). "Conjugative Transposition." Annual Review of Microbiology **49**(1): 367-397.

19) van Hijum, S. A. F. T., G. H. van Geel-Schutten, et al. (2002). "Characterization of a Novel Fructosyltransferase from Lactobacillus reuteri That Synthesizes High-Molecular-Weight Inulin and Inulin Oligosaccharides." Appl. Environ. Microbiol. **68**(9): 4390-4398.

20) Ventura, M., C. Canchaya, et al. (2006). "Comparative Genomics and Transcriptional Analysis of Prophages Identified in the Genomes of Lactobacillus gasseri, Lactobacillus salivarius, and Lactobacillus casei." Appl. Environ. Microbiol. **72**(5): 3130-3146.

21) Vesterlund, S., J. Paltta, et al. (2005). "Adhesion of bacteria to resected human colonic tissue: Quantitative analysis of bacterial adhesion and viability." Research in Microbiology **156**(2): 238-244.

22) Wagner R.D., (2008). "Effects of microbiota on GI health: gnotobiotic research." Adv Exp Med Biol. (635): 41-56.

23) Zhang, W., D. Yu, et al. (2008). "Complete nucleotide sequence of plasmid plca36 isolated from Lactobacillus casei Zhang." Plasmid **60**(2): 131-135.

Tables and Figures

Table 1. A list of strains, plasmids, and primers used in this study

Strain, plasmid, or primer	Relavant characteristic(s) or sequence	Source or reference
Strains		
L. gasseri ATCC 33323	ATCC (NCK 344)	(Kleeman 1982)
L. gasseri ML3	Healthy Endoscopy (NCK 1347)	(Kullen 2000)
L. gasseri FR2	Healthy Endoscopy (NCK 1344)	(Kullen 2000)
L. gasseri WD19	Patient Endoscopy (NCK 1338)	(Kullen 2000)
L. gasseri ADH	Healthy Fecal (NCK 99)	(Kullen 2000)
L. gasseri SD10	Healthy Fecal (NCK 1343)	(Kullen 2000)
L. gasseri FR4	Healthy Endoscopy (NCK 1345)	(Kullen 2000)
E. coli	StrataClone Solopack compatent cells	Stratagene
Plasmids		
pSC-A-amp/kan	StrataClone PCR Cloning Vector	Stratagene
Primers		
PCR Primer 1	5'-CTAATACGACTCACTATAGGGC-3'	Clonetech
Nested Primer 1	5'-TCGAGCGGCCGCCCGGGCAGGT-3'	Clonetech
Nested Primer 2R	5'-AGCGTGGTCGCGGCCGAGGT-3'	Clonetech
JS_Collflank1_F	5'-GACGGTTGGTGAAGTTAGTGTGAAATT-3'	IDT
JS_Collflank1_R	5'-GTATACCATTTGGTATTGGACATCAATGC-3'	IDT

Table 2. Complete list of BLAST results from SSH fragments found to be absent from the neotype strain (ATCC 33323) and significantly homologous to a protein present in the GenBank database. A significant BLAST hit was designated by a BLAST result with an E-value less than or equal to10^{-3}.

*Designates a top hit that was a hypothetical protein but also had a gene hit with a putative function that fell within the E-10^{-3} significance threshold.

a, b, c Designate a sub-classification (1a,1b etc . . .) and are given for SSH sequences that contained multiple fragments with independent BLAST results (for example if the fragment fell on a junction between two genes). These values were held to the same E-10^{-3} significance threshold.

Strain	Clone	Gene Similarity Results	# of Clones	E-value 10^	Identity	Organism
ML3	2	conserved hypothetical protein	1	-85	99%	*Lactobacillus gasseri* 224-1
	4, 9, 18, 44, 62, 63	quinone oxidoreductase	6	-139	99%	*Lactobacillus gasseri* MV-22
	5, 10	hypothetical protein LgasM_05899	2	-53	100%	*Lactobacillus gasseri* MV-22
	6, 31, 65	conserved hypothetical protein	3	-6	88%	*Lactobacillus brevis* subsp. *gravesensis* ATCC 27305
	8, 35	bacteriophage integrase	2	-4	43%	*Lactobacillus crispatus* MV-1A-US
	12, 29, 45	inulosucrase InuJ	3	-135	86%	*Lactobacillus johnsonii* FI9785
	17	resolvase	1a	-39	98%	*Lactobacillus gasseri* MV-22
		ABC transporter permease component	1b	-23	98%	*Lactobacillus gasseri* MV-22
	20, 55	CRISPR-associated Cas1 family protein	1a	-6	97%	*Lactobacillus gasseri* MV-22
		conserved hypothetical protein	1b	-7	96%	*Lactobacillus antri* DSM 16041
	49	nicotinate phosphoribosyltransferase	1	-140	100%	*Lactobacillus gasseri* MV-22
	51	conserved hypothetical protein	1	-56	92%	*Lactobacillus gasseri* 224-1
	58	putative cellobiose phosphotransferase	1	-17	58%	*Lactobacillus*

Table 2 (continued)

	61	transcription regulator (putative)	1	-51	98%	*Lactobacillus gasseri* MV-22
FR2	2, 11, 35, 46	sucrose phosphorylase	4	-46	98%	*Lactobacillus gasseri* JV-V03
	16	conserved hypothetical protein	1	-24	73%	*Lactobacillus gasseri* JV-V03
	19	fructose/mannose PTS family porter IIC component	1	-31	99%	*Lactobacillus gasseri* JV-V03
	26	phage integrase family protein	1	-4	81%	*Lactobacillus reuteri* DSM 20016
	29	phage integrase family protein	1	-4	81%	*Lactobacillus reuteri* DSM 20016
	30	xylose operon regulator	1	-21	98%	*Lactobacillus gasseri* JV-V03
	58, 60	conserved hypothetical protein	2	-19	65%	*Lactobacillus jensenii* 1153
	61	oligogalacturonide transporter	1	-55	97%	*Lactobacillus gasseri* JV-V03
WD19	3	CRISPR-associated Csn2 family protein	1	-41	97%	*Lactobacillus gasseri* JV-V03
	4	ParA-like	1	-114	93%	*Lactobacillus paracasei* subsp. *paracasei*
		ATPase for chromosome partitioning	1*	-101	67%	*Lactococcus lactis* subsp. *cremoris* SK11
	5, 14, 16, 17, 26	hypothetical protein HMPREF0539_3116	5	-18	92%	*Lactobacillus rhamnosus* LMS2-1
	13, 37	hypothetical extracellular protein	2	-3	54%	*Lactobacillus vaginalis* ATCC
	18	membrane protein with 6 predicted TMS (gaaE)	1	-13	100%	*Lactobacillus gasseri* (direct submission from Japan)
	20	conserved hypothetical protein, putative membrane protein	1a	-53	97%	*Lactobacillus casei* str. Zhang
		hypothetical protein LCAZH_p030	1b	-13	95%	*Lactobacillus casei* str. Zhang
		transposase	1c	-7	100%	*Lactobacillus fermentum* IFO 3956

Table 2 (continued)

	21	permease protein	1	-35	52%	*Streptococcus equi* subsp. *zooepidemicus*
	25	plasmid replication initiation protein	1	-32	100%	*Lactobacillus gasseri* MV-22
	28	hypothetical protein HMPREF0514_0693	1	-14	92%	*Lactobacillus gasseri* JV-V03
	29	PEP-dependent sugar PTS system EIIBC, arbutin specific	1	-19	100%	*Lactobacillus johnsonii* NCC 533
	30, 43	resolvase	2	-102	98%	*Lactobacillus salivarius* UCC118
	34	adhesion exoprotein	1	-24	95%	*Lactobacillus gasseri* JV-V03
	40	nickase	1a	-16	90%	*Lactobacillus sakei* subsp. *carnosus* DSM 15831
		putative LtrC	1b	-41	97%	*Lactobacillus paracasei* subsp. *paracasei*
	51	transposase	1a	-39	94%	*Leuconostoc mesenteroides* subsp. *cremoris* ATCC 19254
		hypothetical protein HMPREF0528_0796	1b	-21	78%	*Lactobacillus johnsonii* ATCC 33200
	61	hypothetical protein LSL_2036	1a	-40	97%	*Lactobacillus salivarius* UCC118
		conserved hypothetical protein	1b	-22	85%	*Lactobacillus gasseri* JV-V03
ADH	2	protein-N(pi)-phosphohistidine--sugar phosphotransferase	1	-54	98%	*Lactobacillusn gasseri* JV-V03
	5, 9	major tail protein	2	-48	95%	*Lactobacillus* phage phiadh
	6	ABC superfamily ATP binding cassette transporter, ABC protein	1	-21	100%	*Lactobacillus gasseri* JV-V03
	8, 13, 18, 19, 20, 25	transposase domain protein	6	-131	96%	*Lactobacillus gasseri* 224-1
	12, 38	hypothetical protein SZO_10330	2	-8	30%	*Streptococcus equi* subsp. *zooepidemicus*
	24	hypothetical protein phiadhp53	1	-38	98%	*Lactobacillus* phage phiadh

Table 2 (continued)

	41	APC family amino acid-polyamine-organocation transporter	1	-72	99%	*Lactobacillus gasseri* JV-V03
	45	accessory secretory protein Asp2	1	-47	98%	*Lactobacillus gasseri* JV-V03
	49	collagen-binding A precursor protein-like protein	1	-27	96%	*Lactobacillus casei* str. Zhang
SD10	1, 36	transketolase	2	-60	97%	*Lactobacillus gasseri* JV-V03
	3	transposase	1	-29	95%	*Lactobacillus jensenii* 1153
	9	conserved hypothetical protein	1a	-36	73%	*Lactobacillus gasseri* JV-V03
		minor tail protein gp26 family protein	1b	-5	55%	*Lactobacillus gasseri* JV-V03
	10	hypothetical protein FAEPRAM212_00540	1	-40	63%	*Faecalibacterium prausnitzii* M21/2
	11	type I restriction-modification system, subunit S	1	-10	36%	*Sphingobacterium spiritivorum* ATCC 33300
	18	hypothetical protein	1	-28	95%	*Lactobacillus johnsonii* FI9785
	20	sucrose phosphorylase	1	-146	95%	*Lactobacillus gasseri* JV-V03
	21	conserved hypothetical protein	1	-125	94%	*Lactobacillus gasseri* JV-V03
		glycosyl transferase family 8	1*	-32	48%	*Lactobacillus reuteri* 100-23
	39	DNA adenine methyltransferase	1a	-32	55%	*Enterococcus* phage phiFL3A
		conserved hypothetical protein	1b	-32	88%	*Lactobacillus gasseri* 224-1
	55	conserved hypothetical protein	1	-28	87%	*Lactobacillus johnsonii* ATCC 33200
	57, 58	hypothetical protein HMPREF0514_1104	2	-16	95%	*Lactobacillus gasseri* JV-V03
	63	terminase large subunit	1	-133	96%	*Lactobacillus prophage* Lj928
FR4	4	conjugation protein	1	-66	98%	*Lactobacillus paracasei* subsp. *paracasei* 8700:2
	6, 15, 19, 27	transposase domain protein	4	-129	96%	*Lactobacillus gasseri* 224-1

Table 2 (continued)

	12	hypothetical protein phiadhp56	1	-80	86%	*Lactobacillus* phage phiadh
	22	Rhamnogalacturonate lyase	1	-102	95%	*Lactobacillus gasseri* JV-V03
	31	hypothetical protein phiadhp57	1a	-29	79%	*Lactobacillus* phage phiadh
		hypothetical protein phiadhp58	1b	-19	71%	*Lactobacillus* phage phiadh
	37	LysR family transcriptional regulator	1	-55	97%	*Lactobacillus gasseri* JV-V03
	40	SocE	1	-15	93%	*Bacillus cereus* W
		hypothetical protein pU302Sp01	1*	-32	84%	*Salmonella enterica* subsp. *enterica* serovar Typhimurium
	48	hypothetical protein SbBS512_C0014	1	-15	64%	*Shigella boydii* CDC 3083-94
	52, 59, 62	conserved hypothetical protein	3	-17	98%	*Lactobacillus paracasei* subsp. *paracasei* 8700:2
		putative ATPase	3*	-16	98%	*Lactobacillus casei* str. Zhang
	53	PEP-dependent sugar PTS system EIIBC, arbutin specific	1	-18	98%	*Lactobacillus johnsonii* NCC 533

Table 3. COG Classifications

Protein Category	COG Classification	# Sequences
Hypothetical (Unknown function)	S	19
Carbohydrate Transport and Metabolism	G	15
Bacteriophage	n/a	13
Cell Surface/Exracellular/Adhesion	M	7
DNA Replication,recombination,repair	L	7
Transposons	L	5
Other Transporters	R	4
Regulation	K	3
Coenzyme Metabolism	H	1
Conjugation	F	1
Redox	C	1
Stress Response	n/a	1

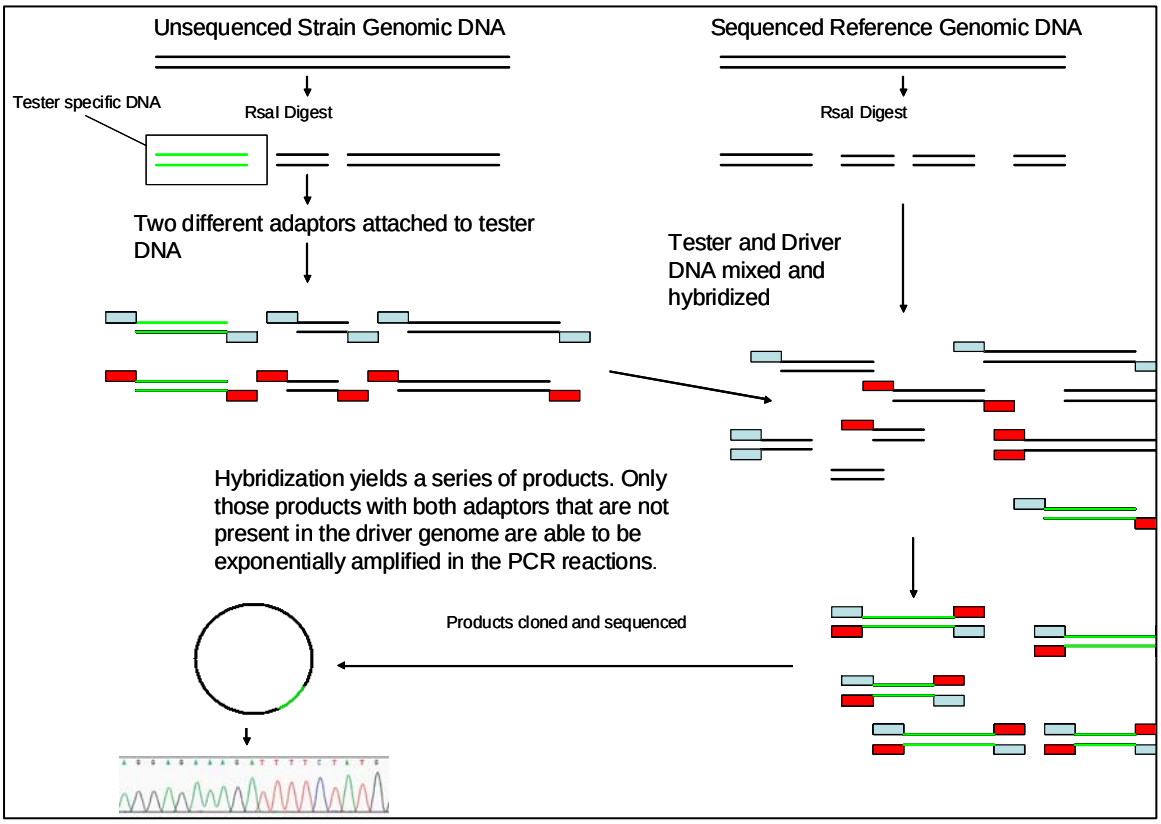


Figure 1. Graphical representation of the subtractive hybridization procedure

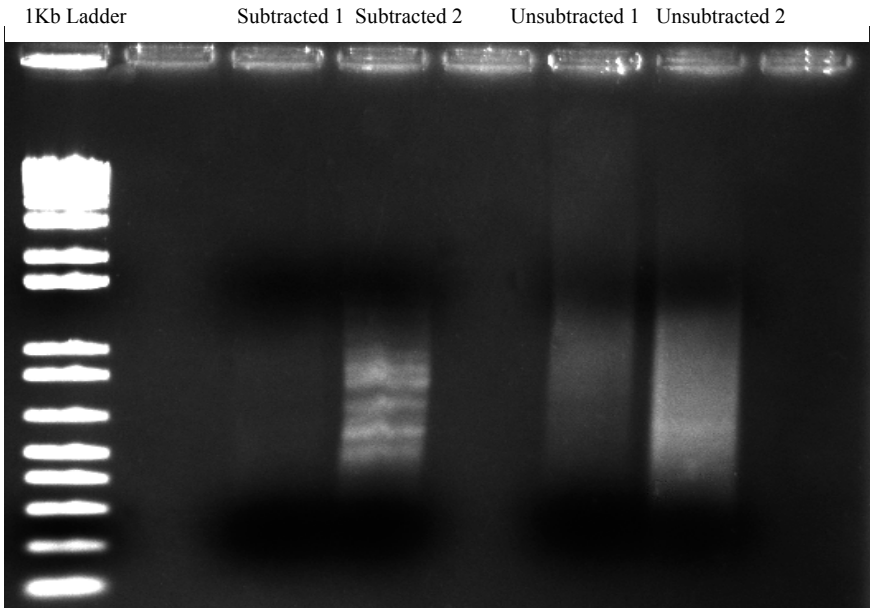


Figure 2. Electrophoresis of primary and secondary PCR products from SSH1 using *L. gasseri* ATCC 33323 as the driver and *L. gasseri* ML3 as the tester. Subtracted PCR products were amplified from the pool of adaptor-ligated and hybridized tester and driver DNA fragments. Unsubtrated PCR products were amplified from only the adaptor-ligated tester DNA fragments without hybridization with driver DNA.

1,2 Differentiates amplification products of primary and secondary PCR reactions. The Primary PCR reaction was used amplify all adaptor-ligated sequences to increase the total DNA concentration. The secondary PCR reaction was used to amplify only those fragments consisting of double stranded fragments containing adaptors 1 and 2 which was used to enrich for fragments not hybridized to driver DNA.

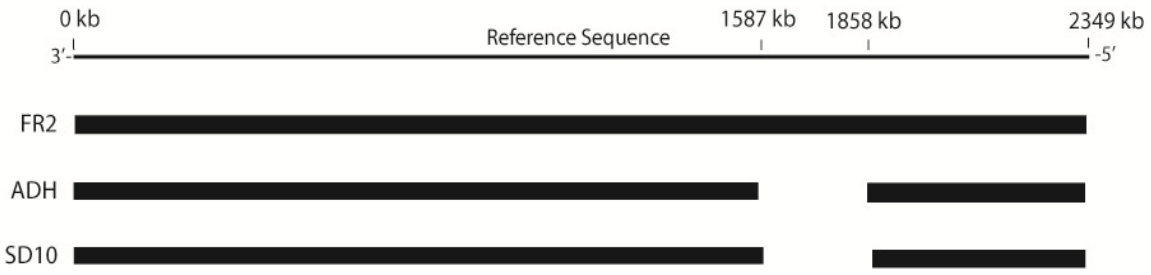


Figure 3a. Alignment of predicted genes encoding collagen binding proteins showing the INDEL repeat region absent from strains ADH and SD10.

Figure 3b. Sequence alignment of the INDEL region of the collagen binding proteins amplified from three *L. gasseri* strains. FR2 contains the 270 bp INDEL region, while this region is absent from strains ADH and SD10.

Supplementary Material

Putative Collagen Binding Protein Sequence Comparison

```
FR2 CBP Only  1
atgagtaataaatctaggcatcgtcgacttttatttatgatgggactagt
ADH CBP Only  1
atgagtaataaatctaggcatcgtcgacttttatttatgatgggactagt
SD10 CBP Only 1
atgagtaataaatctaggcatcgtcgacttttatttatgatgggactagt

FR2 CBP Only  51
cttattttcgttattagggtcaatatggatatcatctaatagtggtatga
ADH CBP Only  51
cttattttcgttattagggtcaatatggatatcatctaatagtggtatga
SD10 CBP Only 51
cttattttcgttattagggtcaatatggatatcatctaatagtggtatga

FR2 CBP Only  101
ccgtatatgctgatagcaagacaaatattactcagaacggtacggggtcc
ADH CBP Only  101
ccgtatatgctgatagcaagacaaatattactcagaacggtacggggtcc
SD10 CBP Only 101
ccgtatatgctgatagcaagacaaatattactcagaacggtacggggtcc

FR2 CBP Only  151
aggactgttattaatcaacaggaagaatcgcagataagtaccgttaatag
ADH CBP Only  151
gggactgttattaatcaacaggaagaatcgcagataagtaccgttaatag
SD10 CBP Only 151
gggactgttattaatcaacaggaagaatcgcagataagtaccgttaatag

FR2 CBP Only  201
taataccacagataatactagttcttcggacgatcagacgtctcagaagc
ADH CBP Only  201
taataccacagataatactagttcttcggacgatcagacgtctcagaagc
SD10 CBP Only 201
taataccacagataatactagttcttcggacgatcagacgtctcagaagc

FR2 CBP Only  251
aagttaccaatacagaagcacaaccaagagcaccaactgacaataatcag
ADH CBP Only  251
```

aagttaccaatacagaagcacaaccaagagcaccaactgacaataatcag
SD10 CBP Only 251
aagttaccaatacagaagcacaaccaagagcaccaactgacaataatcag

FR2 CBP Only 301
ccagttcaggaggaccgtaatcatataagtaatagtcaatactcgacttc
ADH CBP Only 301
ccagttcaggaggaccgtaatcatataagtaatagtcaatactcgacttc
SD10 CBP Only 301
ccagttcaggaggaccgtaatcatataagtaatagtcaatactcgacttc

FR2 CBP Only 351
taatggtgagactaatactaatcagggtccatcagttaatagtatttcta
ADH CBP Only 351
taatggtgagactaatactaatcagggtccatcagttaatagtatttcta
SD10 CBP Only 351
taatggtgagactaatactaatcagggtccatcagttaatagtatttcta

FR2 CBP Only 401
agaataatagttccccatctgcatataataaaggaactagcgttgatatt
ADH CBP Only 401
agaataatagttccccatctgcatataataaaggaactagcgttgatatt
SD10 CBP Only 401
agaataatagttccccatctgcatataataaaggaactagcgttgatatt

FR2 CBP Only 451
acaattaccaatcagagatcatcaactaaggagatctcaggtggaggaac
ADH CBP Only 451
acaattaccaatcagagatcatcaactaaggagatctcaggtggaggaac
SD10 CBP Only 451
acaattaccaatcagagatcatcaactaaggagatctcaggtggaggaac

FR2 CBP Only 501
tacagaattctcttttgacttttctgttcctaattcagtaagatccggag
ADH CBP Only 501
tacagaattctcttttgacttttctgttcctaattcagtaagatccggag
SD10 CBP Only 501
tacagaattctcttttgacttttctgttcctaattcagtaagatccggag

FR2 CBP Only 551
attatacgacaatatcacttcctaatgagttggattttactagagatcaa
ADH CBP Only 551
attatacgacaatatcacttcctaatgagttggattttactagagatcaa
SD10 CBP Only 551

attatacgacaatatcacttcctaatgagttggattttactagagatcaa

FR2 CBP Only 601
tcatttaatgtctacgctcctgatgggacgactacggttgctagtgcaat
ADH CBP Only 601
tcatttaatgtctacgctcctgatgggacgactacggttgctagtgcaat
SD10 CBP Only 601
tcatttaatgtctacgctcctgatgggacgactacggttgctagtgcaat

FR2 CBP Only 651
tattaataaagaaaaaaagacgttggtcttaacatacacggactatgtag
ADH CBP Only 651
tattaataaagaaaaaaagacgttggtcttaacatacacggactatgtag
SD10 CBP Only 651
tattaataaagaaaaaaagacgttggtcttaacatacacggactatgtag

FR2 CBP Only 701
atacacatgatgatgtcaccggccacattaatatgcaaataatggctgat
ADH CBP Only 701
atacacatgatgatgtcaccggccacattaatatgcaaataatggctgat
SD10 CBP Only 701
atacacatgatgatgtcaccggccacattaatatgcaaataatggctgat

FR2 CBP Only 751
acagaaaaggtgaagaaagacactatacttcctgcaacaattaaaataaa
ADH CBP Only 751
acagaaaaggtgaagaaagacactatacttcctgcaacaattaaaataaa
SD10 CBP Only 751
acagaaaaggtgaagaaagacactatacttcctgcaacaattaaaataaa

FR2 CBP Only 801
cgggaatgtagttactatcgatagtacgggtataagctataaagttaata
ADH CBP Only 801
cgggaatgtagttactatcgatagtacgggtataagctataaagttaata
SD10 CBP Only 801
cgggaatgtagttactatcgatagtacgggtataagctataaagttaata

FR2 CBP Only 851
aaggtgatagtaatattgatttttataaatatggatggatggattatgat
ADH CBP Only 851
aaggtgatagtaatattgatttttataaatatggatggatggattatgat
SD10 CBP Only 851
aaggtgatagtaatattgatttttataaatatggatggatggattatgat

FR2 CBP Only 901
aacaatgaacttatctaccgtattgttataaacacaactaattctaaacg
ADH CBP Only 901
aacaatgaacttatctaccgtattgttataaacacaactaattctaaacg
SD10 CBP Only 901
aacaatgaacttatctaccgtattgttataaacacaactaattctaaacg

FR2 CBP Only 951
gaataatgttattattaaggatattcttaattccgctggagttaattacg
ADH CBP Only 951
gaataatgttattattaaggatattcttaattctgctggagttaattacg
SD10 CBP Only 951
gaataatgttattattaaggatattcttaattctgctggagttaattacg

FR2 CBP Only 1001
aacagtcaagcataaaaattgcaactggaacttgggttaaggacgccaat
ADH CBP Only 1001
aacagtcaagcataaaaattgcaactggaacttgggttaaggacgccaat
SD10 CBP Only 1001
aacagtcaagcataaaaattgcaactggaacttgggttaaggacgccaat

FR2 CBP Only 1051
aactattggtcattgtctaactataaagaagtaactgaaaaatatacccc
ADH CBP Only 1051
aactattggtcattgtctaactataaagaagtaactgaaaaatatacccc
SD10 CBP Only 1051
aactattggtcattgtctaactataaagaagtaactgaaaaatatacccc

FR2 CBP Only 1101
tgaattcaatggctcatcttttctgtgaatcttggtaatatcaatcagg
ADH CBP Only 1101
tgaattcaatggctcatcttttctgtgaatcttggtaatatcaatcagg
SD10 CBP Only 1101
tgaattcaatggctcatcttttctgtgaatcttggtaatatcaatcagg

FR2 CBP Only 1151
gtttctctatcacctataaggtaaagccttctcataagttgattaacgga
ADH CBP Only 1151
gtttctctatcacctataaggtaaagccttctcataagttgattaacgga
SD10 CBP Only 1151
gtttctctatcacctataaggtaaagccttctcataagttgattaacgga

FR2 CBP Only 1201
gaacaaattaaaaattcatcaagttattcgtataacgaccataagatgaa

ADH CBP Only 1201
gaacaaattaaaaattcatcaagttattcgtataacgaccataagatgaa
SD10 CBP Only 1201
gaacaaattaaaaattcatcaagttattcgtataacgaccataagatgaa

FR2 CBP Only 1251
tgaatatactaatacagtatattatcaagaggcaaatggtactgcaaacg
ADH CBP Only 1251
tgaatatactaatacagtatattatcaagaggcaaatggtactgcaaacg
SD10 CBP Only 1251
tgaatatactaatacagtatattatcaagaggcaaatggtactgcaaacg

FR2 CBP Only 1301
gatataactatagtttaacaataaaaaaggtggatgaatctggcaatcca
ADH CBP Only 1301
gatataactatagtttaacaataaaaaaggtggatgaatctggcaatcca
SD10 CBP Only 1301
gatataactatagtttaacaataaaaaaggtggatgaatctggcaatcca

FR2 CBP Only 1351
ttacatggcgctatctttgatgtaattcgacagtcaactggagaagttgt
ADH CBP Only 1351
ttacatggcgctatctttgatgtaattcgacagtcaactggagaagttgt
SD10 CBP Only 1351
ttacatggcgctatctttgatgtaattcgacagtcaactggagaagttgt

FR2 CBP Only 1401
aggaactatcacaacagattcctcaggacagggcactatctatgggttgt
ADH CBP Only 1401
aggaactatcacaacagattcctcaggacagggcactatctatgggttgt
SD10 CBP Only 1401
aggaactatcacaacagattcctcaggacagggcactatctatgggttgt

FR2 CBP Only 1451
tacgtgatagttaccaattgaaagaggtcagggcacctaagggatatcaa
ADH CBP Only 1451
tacgtgatagttaccaattgaaagaggtcagggcacctaagggatatcaa
SD10 CBP Only 1451
tacgtgatagttaccaattgaaagaggtcagggcacctaagggatatcaa

FR2 CBP Only 1501
ttagcagatacggtaatcgtaacttctacaaattttgattcaaaaggggc
ADH CBP Only 1501
ttagcagatacggtaatcgtaacttctacaaattttgattcaaaaggggc

```
SD10 CBP Only 1501
ttagcagatacggtaatcgtaacttctacaaattttgattcaaaaggggc

FR2 CBP Only  1551
cgttattaaaattatcgataagatggagaaaacctcagtttcaggaacya
ADH CBP Only  1551 cgttattaaaattatcgataagatggagaaaacctca----
---------
SD10 CBP Only 1551
cgttattaaaattatcgataagatggagaaaacctcagtttca-------

FR2 CBP Only  1601
agacttggaaggataacaacaatcaagatggaatccgcccaagttcaatc
ADH CBP Only       ----------------------------------------
---------
SD10 CBP Only      ----------------------------------------
---------

FR2 CBP Only  1651
acggtgaacttgttggctaacggtcaacaagtcgcaagcaagacggttag
ADH CBP Only       ----------------------------------------
---------
SD10 CBP Only      ----------------------------------------
---------

FR2 CBP Only  1701
tgctagtgataactggcaatacagttttgacaacttagccgcttatgcta
ADH CBP Only       ----------------------------------------
---------
SD10 CBP Only      ----------------------------------------
---------

FR2 CBP Only  1751
atggtcagaagattacctacacggtaacagaaaatgcagttgctgggtac
ADH CBP Only       ----------------------------------------
---------
SD10 CBP Only      ----------------------------------------
---------

FR2 CBP Only  1801
acaagtactgttgatggttacaacgtaactaacaaccacactcctgccac
ADH CBP Only       ----------------------------------------
---------
SD10 CBP Only      ----------------------------------------
---------
```

```
FR2 CBP Only  1851
cgtcaaggtatctggaactaagacttggaaggataacaacaatcaagatg
ADH CBP Only  1588 -------
gtwtcwggaaccaagacttggaaggataacaacaatcaagatg
SD10 CBP Only 1594 -------------
ggaaccaagacttggaaggataacaacaatcaagatg

FR2 CBP Only  1901
gaatccgcccaagttcaatcacggtgaacttgttggctaacggtcaacaa
ADH CBP Only  1631
gaatccgcccaagttcaatcacrgtgaacttattggctaatggtcaacaa
SD10 CBP Only 1631
gaatccgcccaagttcaatcacrgtgaacttattggctaatggtcaacaa

FR2 CBP Only  1951
gtcgcaagcaagacggttagtgctagtgataactggcaatacagttttga
ADH CBP Only  1681
gtcgcaagcaagacggttagtgctagtgataactggcaatacagtttcga
SD10 CBP Only 1681
gtcgcaagcaagacggttagtgctagtgataactggcaatacagtttcga

FR2 CBP Only  2001
caacttagccgcttatgctaatggtcagaagattacctacacggtaacag
ADH CBP Only  1731
caacttagccgcttatgctaatggtcagaagattacctacacggtaacag
SD10 CBP Only 1731
caacttagccgcttatgctaatgrtcagaagattacctacacggtaacag

FR2 CBP Only  2051
aaaatgcagttgctgggtacacaagtactgttgatggttacaatattact
ADH CBP Only  1781
aaaatgcagttgctgggtacacaagtactatcgatggttacaatattact
SD10 CBP Only 1781
aaaatgcagttgctgggtacacaagtactrtcgatggttacaatattact

FR2 CBP Only  2101
aatacccataatccgacaacacctaagaagccccaggttcctaataaccc
ADH CBP Only  1831
aatacccataatccgacaacacctaagaagccccaggttcctaataaccc
SD10 CBP Only 1831
aatacccataatccgacaacacctaagaagccccaggttcctaataaccc

FR2 CBP Only  2151
```

gacaacgcctaagaaacctcaagttcctaataatgaaaataaggtaatac
ADH CBP Only 1881
gacaacgcctaagaaacctcaagttcctaataatgaaaataaggtaatac
SD10 CBP Only 1881
gacaacgcctaagaaacctcaagttcctaataatgaaaataaggtaatac

FR2 CBP Only 2201
caaaggcttacactcaaggtaaaacgtacgataaaacaagtagattacct
ADH CBP Only 1931
caaaggcttacactcaaggtaaaacgtacgataaaacaagtagattacct
SD10 CBP Only 1931
caaaggcttacactcaaggtaaaacgtacgataaaacaagtagattacct

FR2 CBP Only 2251
caaacaggtgatagatcatcaataggtatgatgttcgttggactagtaat
ADH CBP Only 1981
caaacaggtgatagatcatcaataggtatgatgttcgttggactagtaat
SD10 CBP Only 1981
caaacaggtgatagatcatcaataggtatgatgttcgttggactagtaat

FR2 CBP Only 2301
gttactattaagcttaggtttagtagttataaacaggtttactgtctga
ADH CBP Only 2031
gttactattaagcttaggtttagtagttataaacaggtttactgtctga
SD10 CBP Only 2031
gttactattaagcttaggtttagtagttataaacaggtttactgtctga

APPENDIX

Carbohydrate utilization of *Lactobacillus acidophilus* NCFM in a mixed carbohydrate medium containing galactooligosaccharides

Joel Schroeter, M. Andrea Azcarate-Peril, Jose M. Bruno-Barcena,
Todd R. Klaenhammer

Project Summary

In the present study we analyzed galacto-oligosaccharide (GOS) consumption and utilization by *Lactobacillus acidophilus* NCFM, a probiotic microorganism traditionally used in dairy products. We conducted pH-controlled fermentation experiments using Vivinal GOS® (Friesland Foods Domo) which contains ~50% GOS as well as traces of glucose, lactose, and galactose, as a source of carbohydrates. *L. acidophilus* NCFM was able to use the oligosaccharides under anaerobic fermentation conditions in the following preference: glucose → lactose – GOS → galactose.

Materials and Methods

Strains and culture conditions

Lactobacillus acidophilus NCFM was propagated from frozen stocks in MRS broth (Difco Laboratories Inc., Detroit, MI) and subsequently transferred in semi defined medium (SDM) [2] containing 1% Vivinal GOS, glucose, or lactose as a source of carbohydtrates. The composition of Vivinal GOS is shown in Table 1.

Growth and sample collection

Growth and time course experiments were conducted in a 1.5 L fermentor, model Bioflow 110 (New Brunswick Scientific Edison, NJ) with a 1 L working volume. The experiments were performed at 37°C under anaerobic conditions maintained by sparging with anaerobic gas mix containing 5% CO_2 and 95% N_2 at a rate of 0.1 vvm. The pH was controlled at 5.5 with addition of 5N NH_3OH and agitated at 100 rmp. A series of three

experiments were performed using SDM containing glucose, lactose, or Vivinal GOS as the carbohydrate source. Ten ml samples of the growth medium were taken every 4 hours over a 24 hour period and growth was monitored by OD_{600nm} using an off-line spectrophotometer. At each time point, 10 ml of cells were collected by centrifugation and cell pellets were flash frozen in a dry-ice ethanol bath and stored at -80°C for future use.

Analysis

Analysis of organic acids, sugars, and other fermentation products were performed by High Pressure Liquid Chromatography (HPLC) in a chromatograph (Shimadzu), separated using an Alltech IOA-1000 Organic Acid column (7.8 x 300mm), and detected with Refractive Index (RID) and/or Evaporative Light Scattering (ELSD) detectors (Shimadzu). The compounds were evaluated under isocratic conditions at 65°C using a mobile phase of 0.05 M sulfuric acid when using RID and water when using ELSD at a flow rate of 0.40 ml/min. Samples were centrifuged for 1 min at 13,000 x g, and products/substrates were determined in the supernatants.

Results and Discussion

The objective of these experiments was to determine whether, and at what point during growth, does *L. acidophilus* NCFM utilize GOS oligosaccharides. The preliminary growth curve experiments using glucose, lactose, and Vivinal GOS as carbohydrate sources yielded equivalent growth patterns and final OD's on all substrates (figure 1). HPLC analysis

of cell-free growth medium samples taken throughout the growth depicts a preferential utilization of carbohydrates present in the Vivinal GOS. NCFM utilized Vivinal GOS components in the following manner: glucose → lactose – GOS → galactose. The chromatogram shown in Figure 2 shows the preferential fermentation of glucose first followed by fermentation of lactose, and then GOS which occured between the 8-12 hr. time points.

Non-digestible oligosaccharides such as GOS have been shown to promote the growth of beneficial microorganisms within the human gastrointestinal tract [1]. Determining what mechanisms are employed to utilize these carbohydrates is an important step in understanding carbohydrate metabolism within the gut environment. These experiments identified the time points during which NCFM utilized GOS oligosaccharides within a mixed carbohydrate environment. The results showed that NCFM does metabolize GOS oligosaccharides but preferentially after glucose is exhausted. Cell pellets collected at the 8-12 hr time points during GOS utilization analyzed by whole genome microarrays for gene expression patterns of NCFM growing on glucose, lactose, and GOS. Differentially expressed genes potentially related to GOS utilization were identified and are being examined by functional genomics and confirmed using quantitative PCR. At this time this work is still ongoing.

Tables and Figures

Table 1. Chemical and physical properties of Vivinal® GOS

Table 1. Vivinal GOS composition and properties

Component	Chemical/physical Properties
Galacto-oligosaccharides	59.60%
Glucose anhydrous	21.90%
Lactose anhydrous	16.90%
Galactose	1.60%
Nitrogen	0.01%
Sulphated ash	0.20%
Nitrite	0.1ppm
Dry matter	75.70%
Viscosity (25 C)	2402 cPs
pH	3.40%

Figure 1. Growth curves of *L. acidophilus* NCFM growing on SDM with varying carbohydrate sources measured by OD_{600nm}.

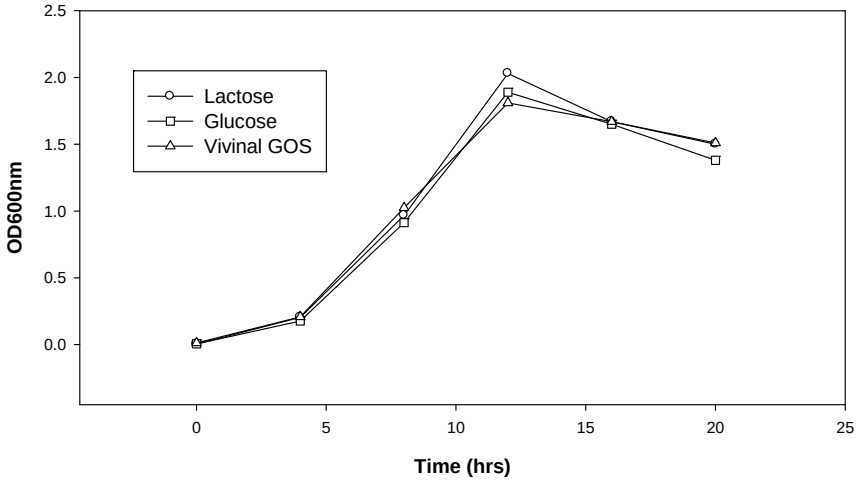
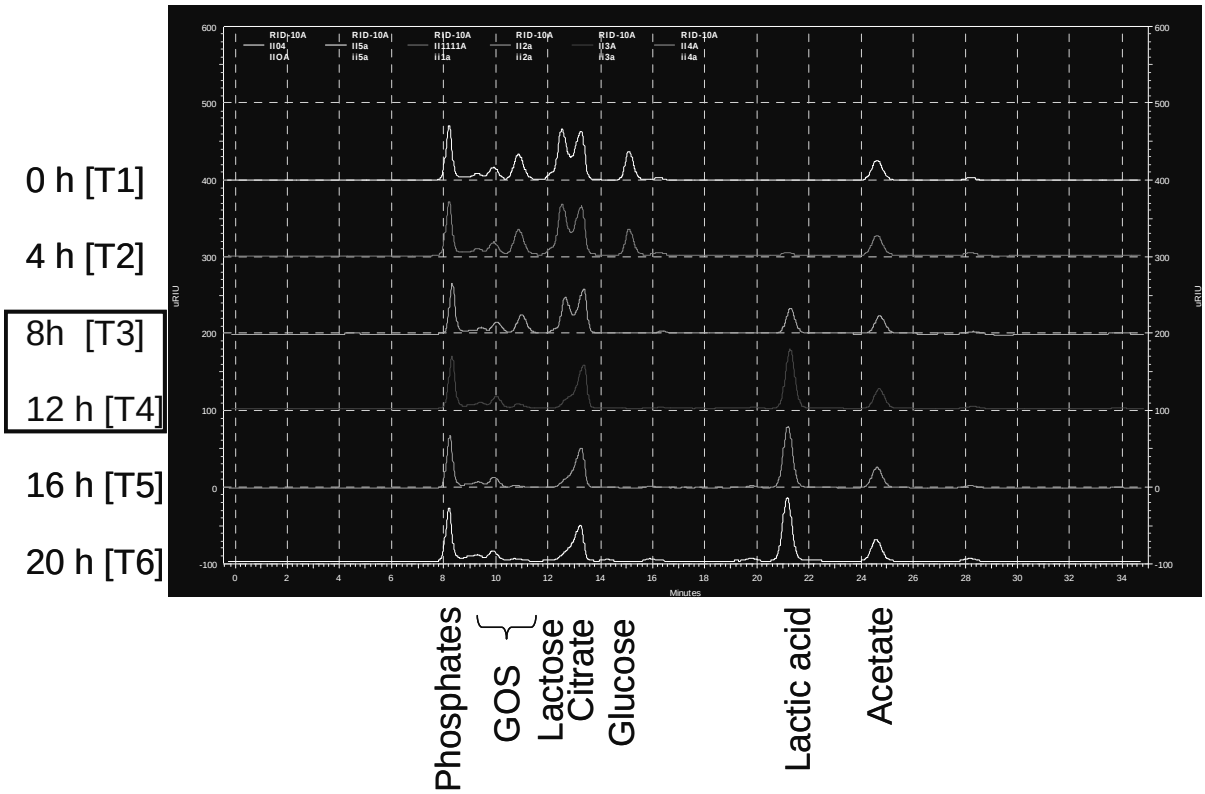


Figure 2. HPLC chromatogram showing utilization of various carbohydrates in Vivinal GOS over course of the fermentation by *L. acidophilus* NCFM. Glucose was utilized first followed by lactose, and GOS. Lactic acid is the main end product of fermentation.



References Cited

1) Boehm, G., M. Lidestri, et al. (2002). "Supplementation of a bovine milk formula with an oligosaccharide mixture increases counts of faecal bifidobacteria in preterm infants." Archives of Disease in Childhood - Fetal and Neonatal Edition **86**(3): F178-F181.

2) Kimmel, S. A. and R. F. Roberts (1998). "Development of a growth medium suitable for exopolysaccharide production by Lactobacillus delbrueckii ssp. bulgaricus RR." International Journal of Food Microbiology **40**(1-2): 87-92.