FUNCTIONAL GENOMICS OF COMMENSAL LACTOBACILLI

Candidate: Michelle M. O' Donnell (BSc)

A thesis presented in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy.

Department of Microbiology,
National University of Ireland,
Cork,
Ireland.

January 2014

Head of Department: Prof. Gerald Fitzgerald
Supervisors: Prof. R. Paul Ross and Prof. Paul W. O' Toole

Dedicated to my loving family and in memory of my father

TABLE OF CONTENTS

DECLARATION

I hereby declare that the content of this thesis is the result of my own work and has not been submitted for another degree, either at University College Cork or elsewhere.

Michelle O' Donnell

_____

PUBLICATIONS

- **O' Donnell, M. M.**, Forde, B. M., Neville, B. A., Ross, R. P. and O' Toole, P. W. Carbohydrate catabolic flexibility in the mammalian intestinal commensal *Lactobacillus ruminis* revealed by fermentation studies aligned to genome annotations. *Microbial Cell Factories*, 10, Suppl 1 2011), S12.
- **O' Donnell, M. M.**, O' Toole, P. W. and Ross, R. P. Catabolic flexibility of mammalian-associated lactobacilli. *Microbial Cell Factories*, 12, 1 2013), 48.
- **O' Donnell, M. M**., Harris, H. M. B, Jeffery, I. B., Claesson, M. J., Younge, B., O' Toole, P. W. and Ross, R. P. The core faecal bacterial microbiome of Irish Thoroughbred racehorses. *Letters in Applied Microbiology*, 57, 6, 2013), 492-501.
- **O' Donnell, M. M**., Harris, H. M. B., O' Toole P. W. and Ross, R. P. The genome of the predominant equine lactobacillus species, *Lactobacillus equi* is reflective of its lifestyle adaptations to a herbivorous host. *Genome Announcements*, 2, 1, 2014).
- **O' Donnell, M. M**., Harris, H. M. B., Ross, R. P. and O' Toole P. W. The core microbiota of domesticated herbivorous hindgut fermenters, mono-gastric and ruminant animals. In preparation as research article intended for submission to *PLOS One.*
- **O' Donnell, M. M**., Harris, H. M. B., Lynch, D., Ross, R. P. and O' Toole P. W. Genomic diversity and biochemical characterisation of *Lactobacillus ruminis* isolates of human, bovine, porcine and equine origin. In preparation as research article intended for submission to *Applied Environmental Microbiology*.
- **O' Donnell, M. M.**, Harris, H. M. B., Lakshminarayanan, B., Lawlor, P. G., O' Toole P. W. and Ross, R. P. Assessing the effect of the galactooligosaccharides and the autochthonous probiotic *Lactobacillus ruminis* on the pig intestinal microbiota. In preparation as a research article for submission to *Letters in Applied Microbiology.*

ABBREVIATIONS

AA = Amino Acid
ACT = Artemis Comparison Tool
BLAST = Basic Local Alignment Search Tool
BSH = Bile Salt Hydrolase
cDNA = Complementary DNA
CDS = Coding DNA Sequence
cfMRS = Carbohydrate-free MRS
COG = Clusters of Orthologous Groups
CFU = Colony Forming Units
DGGE = Denaturing Gel Gradient Electrophoresis
EDTA = Ethylene-diamine-tetraacetic acid
EFSA = European Food Safety Authority
EPS = Exopolysaccharide
EST = Expressed Sequence Tag
FAO/WHO = Food and Agriculture Organisation of the United Nations and the World Health Organisation
FOS = Fructooligosaccharide
GAPDH = Glyceraldehyde 3 Phosphate Dehydrogenase
GI = Gastrointestinal
GOS = Galactooligosaccharide
GRAS = Generally Regarded As Safe
HMP = Human Microbiome Project
KAAS = KEGG Automatic Annotation Server
KEGG = Kyoto Encyclopedia of Genes and Genomes
LAB = Lactic Acid Bacteria
LPS = Lipopolysaccharide
MCP = Methyl accepting chemotaxis protein
MRS = de Mann-Rogosa-Sharpe
NCBI = National Center for Biotechnology Information
NF- κB = Nuclear Factor κ B
NGS = Next-generation Sequencing
NT = Nucleotide
ORF = Open Reading Frame
OTU = Operational Taxonomic Unit
PCR = Polymerase Chain Reaction
qRT-PCR = Quantitative Real-Time PCR
QTL = Quantitative Trait Locus
RACE = Rapid Amplification of cDNA Ends
RBS = Ribosome Binding Site
RFLP = Restriction Fragment Length Polymorphism
rRNA = Ribosomal RNA
SCFA = Short Chain Fatty Acid
SOS = Soybean-oligosaccharide
S. typhimurium = Salmonella enterica serovar Typhimurium
T-RFLP = Terminal Restriction Fragment Length Polymorphism

ABSTRACT

Catabolic flexibility affords a bacterium the ability to utilise various sugars as carbon sources for energy. This ability is particularly important for commensal lactobacilli which are exposed to a variety of simple and complex carbohydrates *in vivo*. *Lactobacillus ruminis* has been identified as one of a limited number of truly autochthonous commensal lactobacilli identified in the gastrointestinal tract of humans and other mammals. However, little was known at the outset of this thesis research about the fermentation capabilities and metabolic pathways used by *L. ruminis* that allow it to survive *in vivo*. Chapter 1 provides a detailed literature perspective and context on the various catabolic flexibility mechanisms utilised by other mammalian associated lactobacilli to enable them to survive in the gastrointestinal tracts of their hosts.

A combination of *in vitro* and *in silico* techniques was used to identify the pathways, enzymes and transporters involved in the utilisation of a variety of carbohydrates by *L. ruminis*. The transport and catabolic machinery involved in the utilization of ≥50 carbohydrates including prebiotics were identified by comparison of *in vitro* fermentation profiles with the genome annotation of two *L. ruminis* strains (ATCC 25644 and ATCC 27782). Prebiotic utilisation operons and transporters were identified *in silico*. Carbohydrate symporter transport families were identified as the primary transporters of relatively complex carbohydrates in *L. ruminis*. In contrast, simpler carbohydrates like mono and di-saccharide sugars were transported via energy dependent transport systems. This suggested that *L. ruminis* has adapted to its intestinal niche which provides a steady supply of carbon sources that allows *L. ruminis* to use less energy dependent methods of carbohydrate translocation.

Microbiota dysbiosis-related diseases may be caused or aggravated by the ingestion of certain carbohydrates. Diet is a major factor that affects the bacterial diversity of the gut microbiota. To help prevent gut-related health loss in important animals, such as racehorses and other performance animals, it is important to identify the core microbiota of healthy animals consuming different feeds. Culture-independent analysis of the microbiota of six healthy racehorses revealed that the core microbiota of these hindgut animals was dominated by the *Firmicutes* and *Bacteroidetes* phyla. While not the main focus of the study, differences between the groups were noted. Active horses receiving high starch concentrate feed had lower microbiota diversity than naturally grazing horses at rest. This loss of diversity

coincided with an increase in the abundance of *Lactobacillus* and *Streptococcus* genera *in vivo*.

Diversity in the gut microbiota of humans and animals can be affected by many internal and external factors. To identify the bacterial diversity and a core microbiota of domesticated herbivores, 10 animal species from a single Irish farm were analysed. Animal gut microbiota diversity was affected primarily by host phylogeny, and by extension, the digestion physiology of the animal. The *Firmicutes* and *Bacteroidetes* phyla formed the core of the gut microbiota in the groups associated with digestion method and animal species studied.

Following on from the *in vitro* and *in silico* assessment of carbohydrate fermentation by *L. ruminis*, I supplemented the diets two groups of pigs (n=12) with a prebiotic (galactooligosaccharides) or with a synbiotic (*L. ruminis* ATCC 25644 and galactooligosaccharides). Supplementation of the porcine diet with the prebiotic alone had no effect on the diversity of the microbiota. However, the synbiotic treatment significantly reduced the microbiota diversity.

To date there is little published information describing the genomic diversity and survival characteristics of *Lactobacillus ruminis*. To expand the knowledge base of this commensal lactobacillus species, we compared 16 *L. ruminis* strains using a panel of *in vitro* growth and survival assays, molecular biology, whole genome sequencing and RNA sequencing (RNAseq). Survival assays identified *L. ruminis* S23, DPC 6832 and DPC 6835 as strains with potential use in industry. The multi locus sequence typing scheme developed in this study revealed that the strains clustered by host isolation (human, bovine, porcine and equine). Phylogenetic comparison of the four sequenced *L. ruminis* genomes (S23, ATCC 25644, ATCC 27782 and DPC 6832) revealed that both human-derived strains (S23 and ATCC 25644) were closely related, while the equine strain DPC 6832 was the most divergent strain. RNAseq of two motile strains (ATCC 27782 and DPC 6832) and three growth conditions (stationary, swimming and swarming) identified a number of carbohydrate enzymes and transporters up-regulated under swarming conditions. This suggests that carbohydrate utilisation enzymes such as beta-fructofuranosidase and 1-phosphofructokinase have unrecognised roles in *L. ruminis* swarm cell proliferation.

Lactobacillus equi, a dominant lactobacillus species in the equine hindgut, was recently isolated from a healthy Irish thoroughbred. The genome was sequenced using an Illumina HiSeq instrument and the draft genome was annotated. The

annotated genome was 2.19 Mb in size and comprised 2,263 predicted genes. When compared to other sequenced genomes in the *L. salivarius* clade, the genome of *L. equi* was most closely related to *L. ruminis* ATCC 27782. The similarity between *L. equi* and *L. ruminis* ATCC 27782 was also reflected in the proportion of shared orthologous genes between the two species. Two enzymes, tagatose 1,6 diphosphate aldolase and fructan hydrolase, previously not described in *L. salivarius* clade, were identified in the genome of *L. equi*. We surmised that these enzymes along with the other predicted glycosyl hydrolases and carbohydrate transporters may offer *L. equi* an advantage in the complex and harsh hindgut environment.

In summary, this thesis uses functional genomics to assess the effect that carbohydrates have on commensal lactobacilli but also on the microbiota as a whole. The impact of diet on the microbiota was assessed in a variety of animal hosts.

Chapter I

Catabolic flexibility of mammalian-associated lactobacilli.

INTRODUCTION

Sections 1.1-1.6 were published in full as a review article in:

O' Donnell, M. M., O' Toole, P. W. & Ross, R. P. (2013). Catabolic flexibility of mammalian-associated lactobacilli. *Microbial Cell Factories* **12**, 48**.**

Chapter I

Table of contents

Abstract

Metabolic flexibility may be generally defined as “the capacity for the organism to adapt fuel oxidation to fuel availability”. The metabolic diversification strategies used by individual bacteria vary greatly from the use of novel or acquired enzymes to the use of plasmid-localised genes and transporters. In this review, we describe the ability of lactobacilli to utilise a variety of carbon sources from their niche habitat or new environments in order to grow and survive. The genus *Lactobacillus* now includes more than 150 species, many with adaptive capabilities, broad metabolic capacity and species/strain variance. They are therefore, an informative example of a cell factory capable of adapting to new niches with differing nutritional landscapes. Indeed, lactobacilli naturally colonise and grow in a wide variety of environmental niches which include the roots and foliage of plants, silage, various fermented foods and beverages, the human vagina and the mammalian gastrointestinal tract (GIT; including the mouth, stomach, small intestine and large intestine). Here we primarily describe the metabolic flexibility of some lactobacilli isolated from the mammalian gastrointestinal tract, and we also describe some of the food-associated species with a proven ability to adapt to the GIT. As examples this review concentrates on the following species - *Lb. plantarum*, *Lb. acidophilus*, *Lb. ruminis*, *Lb. salivarius*, *Lb. reuteri* and *Lb. sakei*, to highlight the diversity and inter-relationships between the catabolic nature of species within the genus.

1.1 Introduction

The human gut is an ecological niche where bio-transformations of dietary ingredients occur, catalysed by gut bacteria including lactobacilli. With that in mind, this review describes, compares and summarises the catabolic machinery present in the mammalian-associated lactobacilli. Lactobacilli are well-characterised members of the Lactic Acid Bacteria (LAB) that are found throughout the gastrointestinal tract of humans and other mammals, and although generally sub dominant in the colon, can be present at proportionately high levels in the upper GIT (Holzapfel & Wood, 1995).

The LAB are low G+C Gram positive bacteria and have multiple uses in the food industry. Those associated with foods include the *Lactobacillus* and *Bifidobacterium* genera (Stiles & Holzapfel, 1997). Bifidobacteria are phylogenetically distant from all of the other low [G+C%]–genome LAB, but are pragmatically included in the LAB group based on their functionality and habitat (Vaughan *et al.*, 2002). In this respect, LAB are integral inhabitants of the microbiota of the gastrointestinal tract where they contribute to intestinal barrier integrity and have roles in immunomodulation and pathogen resistance (Stiles, 1996). This adds impetus to their inclusion in functional food products.

The growth of all living organisms is dependent on efficient cycling and recovery of energy from the environment. Carbohydrates are the primary source of carbon and energy for the growth of microorganisms (Gunsalus *et al.*, 1955). Glycolysis is the most important carbohydrate metabolic cycle in the majority of bacteria and constitutes the main energy generating mechanism. In many of the commensal *Lactobacillus* species, four of the main glycolytic genes along with a regulator are encoded by the *gap* operon. Such *gap* operons have previously been reported for other Gram positive bacteria including bacilli and clostridia (Ludwig *et al.*, 2001; Schreiber & Dürre, 2000). The *gap* operon in mammalian lactobacilli generally encodes the central glycolytic gene regulator (*cgg*R), glyceraldehydes-3-phosphate dehydrogenase (*gap*), phosphoglycerate kinase (*pgk*), triosephosphate isomerase (*tpi*) and an enolase (*eno*). This operon arrangement was first noted in the genomes of *Lactobacillus plantarum* NC8 and *Lactobacillus sakei* Lb790 (Naterstad *et al.*, 2007). However, this particular arrangement of the *gap* operon has also since been identified in a variety of other *Lactobacillus* species genomes (Azcarate-Peril *et al.*, 2008;

Forde *et al.*, 2011; Kankainen *et al.*, 2009; Kleerebezem *et al.*, 2003; Pridmore *et al.*, 2004), while some other genomes contain only partial operons (Altermann *et al.*, 2005; Claesson *et al.*, 2006; Jiménez *et al.*, 2010b; Kant *et al.*, 2011; Morita *et al.*, 2008). The conservation of this operon arrangement (and fragments thereof) in the genomes of a number of mammalian-associated lactobacilli has a number of implications. It suggests that, through evolution and adaptation, this glycolytic operon gene arrangement has been optimised for functionality and that there is a strong selective pressure against nucleotide, gene and operon change.

The ability of lactobacilli to efficiently utilise both of the glycolytic pathways facilitates the degradation of a wider range of carbohydrates present in a given niche, but is also information relevant for their industrial exploitation. For example, *Lactobacillus reuteri* is a commensal, facultatively hetero-fermentative species able to use both the Embden-Meyerhof pathway (EMP) and the phosphoketolase pathway (PKP) to ferment carbohydrates, exemplified by *Lb. reuteri* ATCC 55730 (Årsköld *et al.*, 2008). However, examination of the genome sequences of other heterofermentative lactobacilli has also revealed genes corresponding to both glycolytic pathways (Claesson *et al.*, 2006; Kleerebezem *et al.*, 2003). A number of genes for enzymes involved in both glycolytic cycles were identified in the genome of *Lb. reuteri* ATCC 55730; however, no recognisable *Lactobacillus*-like *pfkA* gene could be annotated. Metabolic flux analysis identified PKP as the main glycolytic pathway with EMP acting as a shunt (Årsköld *et al.*, 2008). Of the two glycolytic pathways, PKP yields less energy production overall. However, it seems that the EMP functions to provide a net gain in ATP in conjunction with the main energy production by the PKP. It is believed that the use of PKP as the main glycolytic pathway is an adaptation of *Lb. reuteri* and other heterofermentative lactobacilli to an environment rich in carbohydrates (Årsköld *et al.*, 2008). Since *Lb. reuteri* can be used as a cell factory to produce industrially exploitable metabolic intermediates or end products such as 3-hydroxypropionaldehyde for nylons and plastics , the ability to culture lactobacilli such as *Lb. reuteri* efficiently and cost-effectively will undoubtedly be informed by knowledge of its metabolism (Vollenweider *et al.*, 2003).

The structure of carbohydrates and their degrees of polymerisation determine the complexity of the sugar as well as the enzymes capable of degrading them. The building blocks of the majority of complex carbohydrates metabolised by LAB are

glucose, fructose, xylose and galactose, while the linkages between monosaccharide residues are what determine carbohydrate digestibility in the small intestine (Manning & Gibson, 2004). Related to these parameters, prebiotics are defined as "selectively fermented ingredients that allow specific changes both in the composition and/or activity in the gastrointestinal microbiota that confers benefits upon host well-being and health" (Gibson *et al.*, 2004). The lactobacilli of the mammalian microbiota are capable of fermenting a range of carbohydrates including oligosaccharides, starch, non-starch polysaccharides and many more carbohydrates (Barrangou *et al.*, 2003; Barrangou *et al.*, 2006; O' Donnell *et al.*, 2011; Saulnier *et al.*, 2007). Many different bacterial enzymes are used in the degradation of simple and complex carbohydrates; prominent among them are the glycosyl hydrolase (EC 3.2.1) family of enzymes (Henrissat, 1991; Henrissat & Bairoch, 1996). Table 1 shows a list of glycosyl hydrolases commonly identified in and utilised by lactobacilli.

Table 1.1 Common glycosyl hydrolases present in mammalian lactobacilli

Enzyme	EC number	Gene	Reaction	Associated pathways	References
Alpha-amylase	3.2.1.1	amyA	Endo-hydrolysis of (1->4)-alpha-D-glucosidic linkages in polysaccharides containing three or more (1->4)-alpha-linked D-glucose units	Starch and sucrose metabolism	(Claesson *et al.*, 2006; Forde *et al.*, 2011; Kleerebezem *et al.*, 2003; O' Donnell *et al.*, 2011)
Oligo-1,6-glucosidase	3.2.1.10	malL	Hydrolysis of (1->6)-alpha-D-glucosidic linkages in some oligosaccharides produced from starch and glycogen by EC 3.2.1.1 (alpha-amylase), and in isomaltose	Starch and sucrose metabolism	(Azcarate-Peril *et al.*, 2008; Claesson *et al.*, 2006; Forde *et al.*, 2011; Kleerebezem *et al.*, 2003; O' Donnell *et al.*, 2011; Pridmore *et al.*, 2004)
Maltose-6'-phosphate glucosidase	3.2.1.122	glvA	Hydrolysis of maltose 6'-phosphate	Starch and sucrose metabolism	(Altermann *et al.*, 2005)
Alpha-glucosidase	3.2.1.20	malZ	Hydrolysis of terminal, non-reducing (1->4)-linked alpha-D-glucose residues with release of D-glucose	Galactose, starch and sucrose metabolism	(Altermann *et al.*, 2005; Azcarate-Peril *et al.*, 2008; Forde *et al.*, 2011; Frese *et al.*, 2011; Kleerebezem *et al.*, 2003; O' Donnell *et al.*, 2011; Pridmore *et al.*, 2004)
Beta-glucosidase	3.2.1.21	bglX	Hydrolysis of terminal, non-reducing beta-D-glucosyl residues with release of beta-D-glucose	Starch and sucrose metabolism	(Altermann *et al.*, 2005; Azcarate-Peril *et al.*, 2008; Forde *et al.*, 2011; Kleerebezem *et al.*, 2003; O' Donnell *et al.*, 2011; Pridmore *et al.*, 2004)
Alpha-galactosidase	3.2.1.22	rafA	Hydrolysis of terminal, non-reducing alpha-D-galactose residues in alpha-D-galactosides, including galactose oligosaccharides, galactomannans and galactolipids	Galactose metabolism	(Altermann *et al.*, 2005; Azcarate-Peril *et al.*, 2008; Claesson *et al.*, 2006; Forde *et al.*, 2011; Frese *et al.*, 2011; Kleerebezem *et al.*, 2003; O' Donnell *et al.*, 2011; Pridmore *et al.*, 2004)
Beta-galactosidase	3.2.1.23	lacZ	Hydrolysis of terminal non-reducing beta-D-galactose residues in beta-D-galactosides	Galactose metabolism	(Altermann *et al.*, 2005; Claesson *et al.*, 2006; Forde *et al.*, 2011; Frese *et al.*, 2011; Kleerebezem *et al.*, 2003; O' Donnell *et al.*, 2011; Pridmore *et al.*, 2004)
Beta-fructofuranosidase	3.2.1.26	sacA	Hydrolysis of terminal non-reducing beta-D-fructofuranoside residues in beta-D-fructofuranosides	Galactose, starch and sucrose metabolism	(Altermann *et al.*, 2005; Azcarate-Peril *et al.*, 2008; Claesson *et al.*, 2006; Forde *et al.*, 2011; Kleerebezem *et al.*, 2003; O' Donnell *et al.*, 2011; Pridmore *et al.*, 2004)
Beta-N-acetylhexosaminidase	3.2.1.52	nagZ	Hydrolysis of terminal non-reducing N-acetyl-D-hexosamine residues in N-acetyl-beta-D-hexosaminides	Amino sugar and nucleotide sugar metabolism	(Claesson *et al.*, 2006; Forde *et al.*, 2011; Kleerebezem *et al.*, 2003; O' Donnell *et al.*, 2011)
6-phospho-beta-galactosidase	3.2.1.85	lacG	Hydrolysis of 6-phospho-beta-D-galactosides	Galactose metabolism	(Altermann *et al.*, 2005; Azcarate-Peril *et al.*, 2008; Kleerebezem *et al.*, 2003)
6-phospho-beta-glucosidase	3.2.1.86	bglA	Hydrolysis of 6-phospho-beta-D-glucosyl-(1->4)-D-glucose	Glycolysis	(Altermann *et al.*, 2005; Azcarate-Peril *et al.*, 2008; Forde *et al.*, 2011; Kleerebezem *et al.*, 2003; O' Donnell *et al.*, 2011; Pridmore *et al.*, 2004)
Trehalose-6-phosphate hydrolase	3.2.1.93	treC	Hydrolysis of alpha,alpha-trehalose 6-phosphate	Starch and sucrose metabolism	(Altermann *et al.*, 2005; Azcarate-Peril *et al.*, 2008; Claesson *et al.*, 2006; Kleerebezem *et al.*, 2003; Pridmore *et al.*, 2004)

In a more health conscious society, there has been a growing interest in recent years in the use of prebiotics as modulators of intestinal health (Gibson *et al.*, 2004), and prebiotics have become economically and industrially important as nutritional supplements for adults and as components in the burgeoning infant milk formula market. Lactose, soy oligosaccharides (stachyose and raffinose), lactulose and fructooligosaccharides are some of the carbohydrates that can be classed as prebiotics and that are commonly consumed as dairy, fruits and vegetables (Gibson & Roberfroid, 1995). The microbiota is under constant pressure to adapt to the variety of foods consumed on a daily basis, especially in omnivores like humans. Lactobacilli present in the mammalian GIT have developed an array of adaptations to facilitate their continued presence in the human intestinal microbiota, examples of which will now be discussed. These case studies illustrate how knowledge of *Lactobacillus* metabolism is useful for optimizing their growth in the laboratory or factory, or promoting their retention in the intestinal tract by functional foods.

1.2 Carbon metabolic machinery encoded by *Lactobacillus* genomes and COG assignments

In the last decade, there has been a dramatic expansion in the number of available *Lactobacillus* genome sequences from organisms isolated from a variety of environments including the mammalian GIT, dairy products and fermented foods. Based on the Integrated Microbial Genomes (IMG) website (http://img.jgi.doe.gov/cgi-bin/w/main.cgi), as of April 2013 there are 46 completed *Lactobacillus* genome sequences, comprising 18 unique species. This expansion in the number of genome sequences available has facilitated the use of comparative genomic approaches to examine the machinery involved in growth and survival of lactobacilli with unprecedented rigour.

The genome size of a *Lactobacillus* is often a determinant of the organism's capacity to metabolise a wide range of carbohydrates. Bacterial species with larger genomes are often capable of utilising a wider range of complex carbohydrates like prebiotics while those with smaller genomes are often associated with more restricted niche habitats, for example milk, and are only capable of utilising simple sugars like lactose and galactose. A comparison of the genome size and gene content for the majority of

mammalian lactobacilli is shown in Table 2. *Lb. plantarum* WCFS1 has the largest genome of any *Lactobacillus* genome sequenced to date. This organism uses the phosphoketolase pathway as a central metabolic pathway. *Lb. plantarum* has been isolated from a variety of environments including soil, vegetables, meat, dairy and from the gastrointestinal tract of humans and animals and has been used as a model *Lactobacillus* for metabolic studies (Siezen & van Hylckama Vlieg, 2011; Siezen *et al.*, 2010). Indeed, the genome of *Lb. plantarum* encodes a large contingent of PTS transporters, ABC transporters and glycosyl hydrolases associated with carbohydrate metabolic flexibility (Kleerebezem *et al.*, 2003). In contrast, *Lactobacillus gasseri* has a much smaller genome and is considered to be part of the autochthonous species present in the human gastrointestinal tract, frequently isolated from the mouth, intestines, faeces and vagina of juveniles and adults (Azcarate-Peril *et al.*, 2008; Reuter, 2001). This homofermentative organism is unable to ferment polyols (sugar alcohols), pentoses or deoxysugars, and in this respect resembles other obligate homofermenters (Felis & Dellaglio, 2007; Kandler, 1983). Its inability to ferment pentoses is because of the absence of two key enzymes of the pentose phosphate pathway namely transketolase and transaldolase. Absence of either or both of these enzymes results in the inability to utilise pentose sugars. This limitation is also clearly illustrated by two members of the *Lb. salivarius* clade; *Lb. salivarius* itself (heterofermentative) produces both enzymes and is capable of utilising pentoses while *Lactobacillus ruminis* (homofermentative) lacks a transaldolase gene in its genome and as a result is unable to utilise pentose sugars (Claesson *et al.*, 2006; O' Donnell *et al.*, 2011).

Table 1.2 Genome statistics of various mammalian *Lactobacillus* species

Genome Name	Reference	Genome Size (Mb)	Gene Count	GC (%)
Lb. acidophilus NCFM	(Altermann *et al.*, 2005)	1.99	1970	35
Lb. amylovorus GRL 1118	(Kant *et al.*, 2011)	2.07	2126	38
Lb. fermentum CECT 5716	(Jiménez *et al.*, 2010a)	2.1	1149	51
Lb. gasseri ATCC 33323	(Azcarate-Peril *et al.*, 2008)	1.9	1874	35
Lb. johnsonii FI9785	(Wegmann *et al.*, 2009)	1.8	1804	34
Lb. johnsonii NCC 533	(Pridmore *et al.*, 2004)	1.99	1941	35
Lb. plantarum JDM1	(Zhang *et al.*, 2009)	3.2	3026	45
Lb. plantarum WCFS1	(Kleerebezem *et al.*, 2003)	3.35	3230	44
Lb. reuteri F275, JCM 1112	(Morita *et al.*, 2008)	2.04	1901	39
Lb. rhamnosus GG	(Kankainen *et al.*, 2009)	3.01	3016	47
Lb. rhamnosus GG, ATCC 53103	(Morita *et al.*, 2009b)	3.00	2905	47
Lb. rhamnosus Lc 705	(Kankainen *et al.*, 2009)	3.03	3068	47
Lb. ruminis ATCC 25644	(Forde *et al.*, 2011)	2.14	1901	44
Lb. ruminis ATCC 27782	(Forde *et al.*, 2011)	2.01	2251	44
Lb. salivarius CECT 5713	(Jiménez *et al.*, 2010b)	2.13	1672	33
Lb. salivarius UCC118	(Claesson *et al.*, 2006)	2.13	2196	33

It should be emphasized, however, that examination of *Lactobacillus* genomes alone provides a limited quality of information. Functional genomics studies provide empirical experimental evidence for the functionality, mechanisms and pathways involved in carbohydrate metabolism. The fields of proteomics and transcriptomics in combination with genomics have been exploited to elucidate the mechanisms involved in carbohydrate metabolism in the host and this will be discussed in the next section.

1.3 Metabolic potential of lactobacilli – adaptation to the environment

A wide range of adaptations can potentially develop within a genus or species based on the availability of nutrients and the complexity and competition within their current environment. Adaptation to a particular environment is of great importance for survival especially in a diverse and complex milieu like the mammalian gastrointestinal tract where a wide variety of carbon sources are often present.

Lb. reuteri has previously been used as a model organism for developing and testing microbe/host symbiosis theories (Walter *et al.*, 2010). Along with other mammalian associated lactobacilli, *Lb. reuteri* is reliant on the fermentable carbohydrates and

amino acids present in the mammalian gut digesta. However, some strains of *Lb. reuteri* also have the ability to degrade 1,2-propanediol using the cobamide-enzyme-requiring propanediol dehydratase (EC 4.2.1.28), which may constitute a primary human colonisation parameter for the species. Propanediol dehydratase is a multifunctional enzyme with roles in glycerol utilisation, glycerolipid metabolism, vitamin B_{12} biosynthesis and reuterin formation (Walter *et al.*, 2010). Interestingly, an enzyme with a potentially similar function has been previously identified in *Lactobacillus brevis* ATCC 367 (Makarova *et al.*, 2006). Glycerol is used in food and beverage manufacture as a sweetener, humectant, preservative, filler, thickening agent and solvent. It has also applications in the manufacture of mono/di-glycerides and poly-glycerol for margarine production. Therefore, glycerol can form a significant part of the foods consumed daily, particularly in the western world. The capability to hydrolyse glycerol may provide lactobacilli a competitive advantage in the gastrointestinal tract.

Some *Lactobacillus* species utilise differentially present or differentially expressed features of their carbohydrate metabolic machinery in order to facilitate their colonisation and persistence in the mammalian gut. For example, *Lactobacillus johnsonii* and *Lb. reuteri* do not compete in the mouse fore-stomach because the former utilizes glucose and the latter maltose, even though both species have the genes for metabolizing both substrates (Tannock *et al.*, 2012). This is an example of niche sharing by way of resource partitioning. Using a mouse model Denou *et al.* showed that *Lb. johnsonii* strains use a number of genes (carbohydrate utilisation genes included) for long-term gut persistence. Correlating the datasets from the genomic hybridisation of two strains (ATCC 33200 and NCC533) and the *in vivo* microarray transcription data from strain NCC533 identified six genes, forming three loci that are *Lb. johnsonii* NCC533 strain specific. Two of the loci are involved in carbohydrate metabolism namely exo-polysaccharide biosynthesis (glycosyltransferases) and a mannose phosphoenolpyruvate phosphotransferase system PTS (transporter) (Denou *et al.*, 2008).

A similar transcriptomic study, focusing on the adaptations of *Lb. plantarum,* demonstrated the capacity of a *Lactobacillus* to alter its metabolism in response to the human or murine intestine (Marco *et al.*, 2009; Marco *et al.*, 2010). In those studies, a number of genes required for carbohydrate metabolism were identified as differentially transcribed in the human and mouse gastrointestinal tract under

different dietary conditions. The genes up-regulated included those encoding glycosyl hydrolases, glycolytic enzymes and various carbohydrate transporter classes (Marco *et al.*, 2009; Marco *et al.*, 2010). An overlap in the enzymes induced in the mammalian GIT included those involved in the degradation and transport of lactose and the plant derived-disaccharides melibiose, cellobiose and maltose. In animals fed a Western diet there was also a noteworthy up-regulation of glycerol metabolism-related enzymes, which relates to the presence of glycerol in many foods discussed above. The induction of carbohydrate metabolism genes highlights the importance of metabolic flexibility in the adaptation of *Lactobacillus* and other bacteria to the human and mammalian intestine (Marco *et al.*, 2009; Marco *et al.*, 2010).

Metabonomic studies using Nuclear Magnetic Resonance (NMR) spectroscopy have identified the metabolites most affected by supplementation of the human diet with fructooligosaccharides (FOS) and *Lactobacillus acidophilus* and *Bifidobacterium longum* based synbiotics (Ndagijimana *et al.*, 2009). Beneficial short chain fatty acids (SCFA) namely propionate and butyrate were identified in faeces of individuals receiving the synbiotic treatments. There was also a marked decrease in the recoverable amino acids in the samples. The increase in *Lactobacillus* numbers over the month-long period as well as the increase in SCFA levels and decrease in amino acid concentrations indicate that the feeding of a synbiotic resulted in a shift of the intestinal metabolome from an overall proteolytic pattern to a saccharolytic one. The presence of FOS in the diet, which is indigestible in the upper GIT, had the ability to affect the SCFA profile of the lower GIT when fermented by bacterial species like lactobacilli and bifidobacteria (Ndagijimana *et al.*, 2009).

Another recent study focussed on the adaptation by *Lb. reuteri* to the GIT of mice (Frese *et al.*, 2011). *In vivo* studies using *Lactobacillus*-free (LF) mice and different vertebrate-derived *Lb. reuteri* isolates established that only the rodent isolates were capable of reaching colonising numbers in the LF mice, supporting the theory of host specialisation. Using comparative genome hybridisation, the genome of an *Lb. reuteri* mouse isolate was compared to that of 24 other *Lb. reuteri* strains from various sources. A xylose utilisation operon was conserved in the strains of rodent and porcine origin (Frese *et al.*, 2011) but absent in the others. Xylose forms a large percentage of the hemi-cellulose found in some plants and so is consumed as part of animal diet.

Other examples of niche-specific genes or host specialisation genes between dairy and gastrointestinal lactobacilli have also been revealed using comparative genomic approaches. For example, mannose-6-phosphate glucosidase (EC 3.2.1.122), a mannose catabolic enzyme, was identified as a solely gut-specific gene in the genome sequences of a number of frequently present mammalian lactobacilli (Altermann *et al.*, 2005; Kankainen *et al.*, 2009; Kant *et al.*, 2011; O' Sullivan *et al.*, 2009; Pridmore *et al.*, 2004; Zhang *et al.*, 2009). This enzyme works in conjunction with a maltose phosphotransferase system to import phosphorylated maltose into the cell. Once internalised the enzyme converts maltose-6-phosphate into glucose and glucose-6-phosphate, and it is this method of transport and degradation that is thought to be specific to strains of gut origin. However, this mechanism of maltose utilisation is not ubiquitous among the gut lactobacilli (Azcarate-Peril *et al.*, 2008; Claesson *et al.*, 2006; Forde *et al.*, 2011; Kleerebezem *et al.*, 2003; O' Sullivan *et al.*, 2009). Genome decay, due to gene loss, seems to operate in the dairy lactobacilli that have higher numbers of pseudogenes in their genomes than other lactobacilli. The majority of the pseudo-genes present are related to carbon catabolism, amino acid metabolism and transport, reflecting the fact that these organisms (for example *Lactobacillus helveticus* (Callanan *et al.*, 2008) have less need for these processes in a milk environment. However, it must be noted that even for an organism like *Lb. plantarum* with a diverse range of habitats, continual passage in a nutrient rich medium can lead to genome contraction and loss of certain types of carbohydrate transporters and enzymes (Zhang *et al.*, 2009). A genome level comparison of *Lb. plantarum* JDM1 with *Lb. plantarum* WCFS1 revealed that certain saccharolytic genes and transporters present in strain WCFS1 were absent in the closely related strain JDM1 (Kleerebezem *et al.*, 2003; Zhang *et al.*, 2009). Examples of the absent enzymes include alpha-amylase, alpha-L-rhamnosidase, beta-N-acetylhexosaminidase, mannosyl-glycoprotein, endo-beta-N-acetylglucosaminidase and glucan 1,4-alpha-maltohydrolase (Zhang *et al.*, 2009). This variability of saccharolytic capability within a species is also clearly illustrated by the work of Molenaar *et al*., 2005 who compared over 20 *Lb. plantarum* species using microarray genotyping technology (Molenaar *et al.*, 2005). These were clear examples of a species adapting to their environment and altering their metabolic profile to suit the new environment either by gene acquisition or in this case gene loss.

Recent studies have also focussed on the cellular response of certain lactobacilli to complex carbohydrates. For example, Majumder and colleagues identified a number of proteins involved in the adaptation of *Lactobacillus acidophilus* NCFM to growth in the presence of the prebiotic lactitol (a synthetic sugar alcohol derived from lactose, used in the food industry and in some medications) (Majumder *et al.*, 2011). Examination of the late exponential phase whole-cell extract proteome revealed a number of proteins present which may be involved in utilization of lactitol including a β-galactosidase subunit, galactokinase and other galactose utilisation proteins. The majority of enzymes identified in lactitol utilisation were the same enzymes involved in the Leloir pathway (the lactose utilisation pathway) and transportation of lactitol into the cell was facilitated by LacS (a glycoside-pentoside-hexuronide cation symporter). While transport of lactitol is facilitated by a permease, it is the phosphotransferase system that transports and metabolises sorbitol (Majumder *et al.*, 2011). *Lb. reuteri* (as well as the other mammalian lactobacilli) also possess the genetic determinants for enzymes associated with the utilisation of raffinose family oligosaccharides (RFO). RFOs are present in many vegetables namely legumes and are associated with flatulence and gastrointestinal upset (Rackis Joseph, 1975). Alpha galactosidase (EC 3.2.1.20) and to a lesser extent levansucrase (EC 2.4.1.10) are the main enzymes commonly encoded in the genome sequences of mammalian derived lactobacilli, which are responsible for the hydrolysis and partial hydrolysis of RFO, respectively (Altermann *et al.*, 2005; Azcarate-Peril *et al.*, 2008; Claesson *et al.*, 2006; Jiménez *et al.*, 2010a; Kankainen *et al.*, 2009; Kant *et al.*, 2011; Kleerebezem *et al.*, 2003; O' Donnell *et al.*, 2011; Pridmore *et al.*, 2004; Teixeira *et al.*, 2012). Interestingly, the genome sequences of dairy lactobacilli such as *Lactobacillus bulgaricus* and *Lb. helveticus* (Callanan *et al.*, 2008; Hao *et al.*, 2011) are devoid of RFO degradation associated enzymes, consistent with the fact that milk generally contains negligible amounts of RFO.

Dairy derived lactobacilli, however, can possess considerable and demonstrable metabolic flexibility. Burns *et al.* investigated the "progressive adaptation" of dairy *Lactobacillus delbrueckii* strains to bile (a bio-surfactant produced in the liver for emulsifying fats in the diet). The proteomes of *Lb. delbrueckii* and an enhanced bile resistant derivative were examined using cells grown in the presence and absence of bile. A total of 35 proteins were affected by the inclusion of bile. Three of the proteins were found to be part of the glycolytic cycle with phosphoglycerate mutase

(*pgm*) and glyceraldehyde-3P-dehydrogenase genes up-regulated, while fructose-bisphosphate aldolase was down-regulated at the protein level (Burns *et al.*, 2010). *Lactobacillus casei*, a predominantly dairy associated isolate, is frequently isolated from a range of other niches, including plants, and the human GIT (Cai *et al.*, 2007b; Kandler & Weiss, 1986). Examination of the *Lb. casei* strain fermentation profiles from these various niches identified several trends, for example the increased utilisation of polyols by strains of plant and human origin. Not surprisingly, strains of cheese origin also were found to have an increased capacity for lactose utilisation when compared to non-dairy isolates. The data suggest that *Lb. casei* can adjust its metabolic capabilities in order to adapt to the carbon sources available in a particular niche.

Lactobacilli also have the capacity to alter their metabolism to adapt to a new environment. This is clearly exemplified by a study of *Lb. sakei* where Chiaramonte and colleagues (2010) showed that the meat-borne *Lactobacillus sakei* is capable of colonizing the GIT of mice (Chiaramonte *et al.*, 2010). Analysis of *Lb. sakei* wild-type and morphological mutants revealed an increased capacity for the utilisation of some carbon sources (fructose, ribose and galactose) when compared to the original meat-borne parent strain. Up-regulation of the genes encoding 6-phosphofructokinase, L-lactate dehydrogenase and fructose-bisphosphate aldolase was considered to be the likely cause of this capacity to colonize the mouse GIT. Two genes involved in nucleotide metabolism, CTP synthase and xanthine phosphoribosyltransferase were also up-regulated in the mutants derived from the passage of meat-borne *Lb. sakei* strain through the GIT of axenic mice (Chiaramonte *et al.*, 2010).

1.4 Transporters and their importance in metabolic flexibility and regulation of metabolism

Carbohydrate transporters or permeases are an essential component in carbohydrate metabolism to facilitate permeability of the cell to carbon metabolites, and may be the rate limiting step in their utilization (Lengeler, 1993). Transporters involved in carbohydrate metabolism include proton coupled active transport and group

translocators (Dills *et al.*, 1980). A summary of those systems most commonly found in lactobacilli is presented in Table 3.

Table 1.3 Common carbohydrate transporters utilised by mammalian lactobacilli

Superfamily	Transport family	Transporter class	Transporter subclass	Transport Classification system	Trans-membrane domain range
MFS	Major Facilitator Superfamily (MFS)	Electrochemical Potential-driven Transporters	Porters (uniporters, symporters, antiporters)	TC 2.A.1	12-24
GPH	Glycoside-Pentoside-Hexuronide (GPH):Cation Symporter Family	Electrochemical Potential-driven Transporters	Porters (uniporters, symporters, antiporters)	TC 2.A.2	12
ATP Binding Cassette	ATP-binding Cassette (ABC)	Primary Active Transporters	P-P-bond-hydrolysis-driven transporters	TC 3.A.1	5-6
PTS-GFL	PTS Glucose-Glucoside (Glc) Family	Group Translocators	Phosphotransfer-driven Group Translocators	TC 4.A.1	8
PTS-GFL	PTS Fructose-Mannitol (Fru) Family	Group Translocators	Phosphotransfer-driven Group Translocators	TC 4.A.2	8
PTS-GFL	PTS Lactose-N,N'-Diacetylchitobiose-β-glucoside (Lac) Family	Group Translocators	Phosphotransfer-driven Group Translocators	TC 4.A.3	8
PTS-GFL	PTS Glucitol (Gut) Family	Group Translocators	Phosphotransfer-driven Group Translocators	TC 4.A.4	8
PTS-GFL	PTS Galactitol (Gat) Family	Group Translocators	Phosphotransfer-driven Group Translocators	TC 4.A.5	8
PTS-GFL	PTS Mannose-Fructose-Sorbose (Man) Family	Group Translocators	Phosphotransfer-driven Group Translocators	TC 4.A.6	8
PTS-GFL	PTS L-Ascorbate (L-Asc) Family	Group Translocators	Phosphotransfer-driven Group Translocators	TC 4.A.7	8

Within the LAB, the ATP binding cassette (ABC) transporters form the largest group (Poolman, 2002), whereby a metabolite or macromolecule is transported using energy derived from ATP hydrolysis (Jojima *et al.*, 2010). ABC transporters are capable of transporting mono, di, tri, poly and oligosaccharide as well as polyols (Saier *et al.*, 2006). ABC transporters encoded by the genome sequences of mammalian lactobacilli include those for maltose, lactose, arabinose, sorbitol, mannitol, glucose, N-acetylglucosamine and cellobiose transport together with ribose, xylose, fructose

and rhamnose, all of which are commonly found in the mammalian digesta, especially of omnivores (Altermann *et al.*, 2005; Azcarate-Peril *et al.*, 2008; Claesson *et al.*, 2006; Forde *et al.*, 2011; Kleerebezem *et al.*, 2003). However, genomes from strains of dairy and meat origin so far examined harbour only gene fragments of carbon-transport-related ABC transporters and do not therefore encode a complete transporter protein (Chaillou *et al.*, 2005; Hao *et al.*, 2011).

Transporters that use chemo-osmosis in order to import carbohydrates are called secondary active transporters and are categorised as either uni-porters, symporters or anti-porters (Konings, 2006). The majority of uni/sym/anti-porters are part of a large group called the Major Facilitator Superfamily (MFS) with over 40 recognised MFS families (Chang *et al.*, 2004). MFS transporters are capable of transporting the majority of micro-molecules (like low DP carbohydrates) but are unable to transport macromolecules. Glycoside-pentoside-hexuronide (GPH) transporters are a class of sodium ion symporters that are used by both homo and heterofermentative lactobacilli to transport carbohydrates (Andersen *et al.*, 2011; Barrangou *et al.*, 2006; Kleerebezem *et al.*, 2003; Marco *et al.*, 2010; O' Donnell *et al.*, 2011). Lactobacilli found exclusively in the gastrointestinal tract, for instance *Lb. ruminis*, have been found to harbour a lower number of complete PTS transporters but a higher number of symporters otherwise known as secondary active transporters (O' Donnell *et al.*, 2011). In contrast, *Lb. gasseri*, another autochthonous species in the human gut, encodes two glucose permeases but does not encode a lactose/galactose permease (Azcarate-Peril *et al.*, 2008). The reliance of some lactobacilli on symporters may be due in part to the fact that the gastrointestinal tract is a nutrient-rich, complex environment. Thus the cells do not have to expend as much energy in order to internalize carbohydrates; instead a carbohydrate is transported into the cell using simultaneous sodium ion exchange. Often the sugars found in the GIT are of a high degree of polymerisation like inulin and starches which require alternate transportation methods to the PTS system.

The majority of carbohydrate transport in lactobacilli isolated from a variety of environments, for example *Lb. plantarum* and *Lb. acidophilus*, is done using PTS systems (Altermann *et al.*, 2005; Kleerebezem *et al.*, 2003). This method of transport involves the coupling of energy molecules with phosphorylation, to bring the phosphorylated carbohydrates into the cell, and is of particular importance in the transport of low complexity hexose sugars (Postma *et al.*, 1993). PTS transporters are

characterised by a phosphate transfer cascade involving phosphoenolpyruvate (PEP), enzyme I (EI), histidine protein (HPr) and various EIIABC's. HPr is phosphorylated at site serine 46 by HPrK/P which is only present in the low [G+C%] Gram positives (Saier Jr *et al.*, 1996). PEP-dependent phosphorylation of HPr by EI yields HPr-His-P, which is required for PTS-mediated transport of carbon sources (Titgemeyer & Hillen, 2002).

Many mammalian lactobacilli rely on the PEP-PTS to facilitate nutrient uptake in the gastrointestinal tract and contain a number of PTS classes. This is best exemplified by *Lb. plantarum* and members of the acidophilus complex (Altermann *et al.*, 2005; Azcarate-Peril *et al.*, 2008; Pridmore *et al.*, 2004). The *Lb. plantarum* WCFS1 genome encodes 25 predicted complete PTS EII complexes; it also encodes some incomplete complexes (Kleerebezem *et al.*, 2003). This high number of PTS genes is one of the largest counts in a sequenced microbial genome and currently comes second only to *Listeria monocytogenes* (Glaser *et al.*, 2001). The genome of *Lb. acidophilus* NCFM encodes 20 PEP-PTS; the transporters have predicted specificity for trehalose, fructose, sucrose, glucose, mannose, melibiose, gentiobiose, cellobiose, salicin, arbutin and N-acetylglucosamine PTS (Altermann *et al.*, 2005). The genome of *Lb. gasseri* ATCC 33323, another acidophilus complex bacterium, encodes 21 PEP-PTS transporters including those for predicted transport of fructose, mannose, glucose, cellobiose, lactose, sucrose, trehalose, β-glucosides and N-acetylglucosamine (Azcarate-Peril *et al.*, 2008). The genome of *Lb. johnsonii* NCC 533 encodes 16 PEP-PTS which is a large number for a genome of its size; allowing the predicted transport of sugars such as mannose, melibiose, cellobiose, raffinose, N-acetylglucosamine, trehalose and sucrose, which is supported experimentally by physiological (API CH50, Biomerieux, France) data (Pridmore *et al.*, 2004).

As mentioned above, bacterial species will often preferentially utilise one carbohydrate prior to utilising another by means of the phosphotransferase system. This system requires strict regulation to ensure the ability to preferentially utilise the particular carbohydrate, for example glucose, before any other carbon source. This type of control is called carbon catabolite repression (CCR). CCR is defined as "a regulatory phenomenon by which the expression of functions for the use of secondary carbon sources and the activities of the corresponding enzymes are reduced in the presence of a preferred carbon source" (Deutscher, 2008). Various methods of CCR are present in nearly all free living microorganisms. In phylum Firmicutes, the main

components are catabolite control protein A (CcpA), HPr, HPr kinase/phosphorylase (HPrK) and the glycolytic enzymes fructose 1,6-bisphosphate and glucose-6-phosphate. In *Enterobacteriaceae* the phosphorylation state of EIIA is crucial for CCR, whereas in Firmicutes the phosphorylation state of HPr is essential (Brückner & Titgemeyer, 2002). HPr phosphorylation can occur at two sites, at Histidine-15 by EI and at Serine-46 by HPrK. In the presence of glucose, there is an increase in the level of fructose 1,6-bisphosphate which indicates a high level of glycolytic activity. HPrK kinase activity is triggered by this increase causing phosphorylated HPr to bind to CcpA, which then binds to the cre site on the DNA thereby repressing transcription of the catabolic genes. When glucose levels are low there is a decreased level of Fructose 1,6-bisphosphate, which dephosphorylates HPrK/P at Ser-46 (Gorke & Stulke, 2008; Stulke & Hillen, 1999). The outcome from CCR is the same with the preferential use of a carbon source.

Regulation of carbohydrate metabolism (especially lactose) has also been identified in *Lb. acidophilus* NCFM (Majumder *et al.*, 2011). In the presence of lactose there was an increase in the abundance of pyruvate kinase, a noted indicator of regulation via carbon catabolite repression, and the down regulation of genes for nucleotide metabolism proteins (Majumder *et al.*, 2011). A similar phenomenon was noted in the proteome of *Lactococcus lactis* when grown in the presence of lactose as a carbon source (Guillot *et al.*, 2003). Similarly, in *Lb. plantarum* CCR has been shown to control the expression of phospho-β-glucosidase (Marasco *et al.*, 1998). Lactobacilli like *Lb. brevis* and *Lb. pentosus* which have relaxed control of their carbon catabolite machinery are being investigated for their carbon degradation potential for industry (Kim *et al.*, 2009; Kim *et al.*, 2010). This alternative or relaxed mechanism of carbon catabolite control is being used in industrial fermentations of cellulolytic and ligno-celluloytic materials to form lactic acid and ethanol, respectively (Kim *et al.*, 2009; Kim *et al.*, 2010). The use of lactobacilli that are capable of using mixed carbohydrate sources for growth is of great importance for industries utilising lignocellulose hydrolysate-like biomass containing hexose and pentose sugars like glucose, arabinose and xylose.

1.5 Horizontal gene transfer and plasmid-encoded carbon metabolism genes

Horizontal gene transfer (HGT) has long been recognised as a method by which bacteria receive genes and other genetic elements conferring new abilities from another species, for example *Escherichia coli* transferring ampicillin resistance to *Shigella flexneri* (Tauxe *et al.*, 1989). Mobile genetic elements include transposons, bacteriophages and plasmids (Rankin *et al.*, 2011). While examining the genomes of two species of GIT-associated lactobacilli and a dairy isolate in particular (*Lb. delbrueckii* ssp. *bulgaricus*), it was noted that extensive horizontal gene transfer (HGT) had occurred between the three species (Nicolas *et al.*, 2007). Comparison of phylogenetic trees for over four hundred proteins highlighted the variance between the members of the acidophilus complex. In many cases, the acquisition of new genetic capabilities can include a new method of solute transportation. Mannose PTS transporters are a class of PTS transporters (TC 4.A.6) affliliated with the mammalian-associated *Lactobacillus* species with the exception of *Lb. reuteri* (Morita *et al.*, 2008). Comparison of phylogenetic trees created from the ClustalW alignment of mannose PTS transporters from twenty five bacteria including *Lb. plantarum*, highlighted the likelihood of HGT having occurred (Zúñiga *et al.*, 2005). The study identified the lack of concordance between evolutionary data from 16S ribosomal RNA gene sequences and the evolutionary data generated from the mannose PTS sequences. The analysis also noted that within the mannose transporters in particular, there was a high level of sequence variation among the bacteria studied. Sequence analysis and comparison of the 58 mannose PTS proteins identified the varying patterns caused by HGT and allowed organising the species into six groups (Zúñiga *et al.*, 2005).

A plasmid is defined as "a linear or circular double-stranded DNA that is capable of replicating independently of the chromosomal DNA". Plasmids are very common within the *Lactobacillus* genus with approximately 38% of all species containing one or more plasmids of varying sizes (Wang & Lee, 1997), including most of the species routinely used for industrial applications. Regions of homology have been identified in plasmids from the same species, genus and from other genera (Vogel *et al.*, 1991). Plasmids contribute to horizontal gene transfer, with plasmids often containing genes for carbohydrate, citrate and amino acid utilisation, production of bacteriocins or

other biosynthetic genes (Wang & Lee, 1997). This is best exemplified by *Lb. salivarius* UCC118 which contains 2 cryptic plasmids and one megaplasmid (Li *et al.*, 2007). The megaplasmid (pMP118) harbours genes for the utilisation of pentoses and polyols. It also carries genes involved in glycolysis (FBP) and genes for two pentose pathway essential enzymes, transketolase and transaldolase. The plasmid pMP118 encodes an additional copy of the enzyme ribose-5-phosphate isomerase which may contribute to its metabolic flexibility and adaptive capabilities. Thus, for *Lb. salivarius* to survive in an environment dominated by pentose sugars these plasmid acquired genes would be essential (Claesson *et al.*, 2006; Li *et al.*, 2007). However, the most striking example in the mammalian derived lactobacilli of the importance of plasmids in carbohydrate metabolism is the case of the *Lactobacillus rhamnosus* Lc705 plasmid pLC1 (Kankainen *et al.*, 2009). This 64 Kbp plasmid sequence encodes proteins predicted for the fructose PTS, glucose uptake proteins, a glycosyl hydrolase and a number of genes involved in alpha and beta-galactoside utilisation and transport (Kankainen *et al.*, 2009). It is obvious that without the presence of these plasmid-borne genes, *Lb. rhamnosus* Lc705 would be at a severe competitive disadvantage in the mammalian GIT compared to other *Lactobacillus* species that have these genes integrated in the chromosome. The presence of these genes in the plasmid presumably allows *Lb. rhamnosus* to compete for the alpha galactosides and fructose from plant sources and also for the beta-galactosides from dairy products. It is clear from the available plasmid sequences that, while not always present, carbohydrate genes carried by plasmids are important mobile genetic elements for lactobacilli.

The presence of carbohydrate metabolic genes located on plasmids is also common in food, plant and dairy lactobacilli. Another example of plasmid encoded pentose sugar utilisation genes is the xylose utilisation cluster present in plasmids isolated from *Lactobacillus pentosus* (Posno *et al.*, 1991), a plant derived *Lactobacillus*. A study comparing 34 sequenced Lactobacillus plasmids revealed that the carbohydrate and amino acid transport category was that most frequently encoded among the plasmids analysed (Zhang *et al.*, 2008). The presence of a larger cohort of carbohydrate and amino acid transporters is possibly a niche adaptation. *Lb. casei* 64H lacking the plasmid pLZ64, which contains a lactose PEP-PTS and phospho-β-galactosidase, is unable to utilize lactose. There is limited knowledge on the true extent of plasmids from mammalian derived lactobacilli and their impact on gut health. However, there

is detailed knowledge on the presence and function of plasmids in dairy-derived lactobacilli for example *Lb. casei* (Lee *et al.*, 1982).

1.6 Conclusions

Carbon metabolism is essential for life and the survival of many bacterial species depends on their ability to exert some degree of metabolic flexibility. *Lactobacillus* as a genus, has a broad range of environmental niches and is equipped with an intricate array of enzymatic systems and adaptive responses to cope with differing carbohydrate sources. This poses challenges for examining the effect of lactobacilli on the gut microbiota but also opportunities for their efficient industrial exploitation. Although there is an extensive amount of information on the *in vitro* and *in silico* catabolic flexibility of mammalian lactobacilli, additional studies and investigations are required to elucidate all the factors and systems that are involved in carbohydrate degradation mechanisms *in vivo* in the mammalian GIT. Further metabolomic, metabonomic and metatranscriptomic studies along with concerted effort are needed to fully elucidate all of the effects that carbohydrate metabolism has on strain phenotypes. With advances in sequencing technologies it is now possible and "affordable" to use RNA-seq (whole transcriptome shotgun sequencing) rather than using microarrays. Microarrays have shortcomings that including for example requiring prior sequence information of a strain, and the need to use pure cultures which makes it difficult to assess the effect of species or carbohydrate on the microbiome as a system of interconnected genera and species. Metatranscriptomics can identify the gene expression of mixed communities of organisms *in vivo* under a wide range of parameters including diet, stresses, disease state and other environmental and health factors. The use of metatranscriptomics in conjunction with animal model feeding studies would allow a more accurate measurement of the effect diet has on the *Lactobacillus* component of the microbiota. For *in vivo* studies the use of a "standard" mammalian GIT model, for example the pig, whose physiology is similar to that of humans would be advantageous in allowing more rigorous comparisons of *in vivo* feeding studies. The use of mouse models, while convenient and relatively inexpensive, should be viewed as a "small-scale" step before transitioning the research into a larger human GIT analogue model like the pig. Further investigations using some of the techniques outline above on a wider number

of mammalian derived lactobacilli will provide information that will lead to a greater understanding of *in vivo* carbohydrate metabolism of mammalian derived lactobacilli and the implications for human and animal health. The industrial usage of lactobacilli for production of metabolites and process ingredients will benefit from progress in metabolic modelling, exemplified to date by *Lb. plantarum* WCFS1 (Teusink *et al.*, 2009), but not yet applied to many relevant lactobacillus species. Success of these modelling experiments will be aided by empirical data provided by complementary "omics" analyses, generating greater precision in establishing and fine-tuning models for lactobacillus growth in the laboratory and in the factory.

1.7 The mammalian intestinal microbiota

1.7.1 Health relevancy and methods for studying

The human and animal intestinal microbiota has been implicated as an important factor in many diseases and health states including Irritable Bowel Syndrome (IBS), Inflammatory Bowel Disease (IBD), obesity and Type 2 Diabetes Mellitus (T2DM) and laminitis (Jeffery *et al.*, 2012; Larsen *et al.*, 2010; Payne *et al.*, 2012; Pollitt, 2004; Wu *et al.*, 2013). Thus it is extremely important to identify the diversity and composition of the microbiota to understand its role in maintaining host health and intestinal homeostasis.

The description of the gastrointestinal microbiota composition in humans and animals has been accomplished using a variety of methods including culture-based and molecular techniques (Fraher *et al.*, 2012). But, the use of culture independent methods has revolutionised our view of the microbiota (Riesenfeld *et al.*, 2004). The evolution of molecular technologies has allowed researchers to examine the complex intestinal microbiota environment using DNA microarrays (Tottey *et al.*, 2013) and amplicon next generation sequencing (Andersson *et al.*, 2008). High throughput next generation sequencing allows researchers to identify a large proportion of the microbiota from a relatively small sample input. There are however, limitations to the PCR-based next generation sequencing technologies, particularly pyrosequencing. The inability to phylogenetically assign bacterial identities from short sequence reads of amplicons derived from the intestinal microbiota reduces the efficiency and output of sequencing studies, particularly at the genus level with large proportions of

unclassified reads present (Claesson *et al.*, 2010b). Similarly, the choice of the variable region of the 16S rRNA gene used and the sequencing technology used can influence the proportions of the taxa identified (Claesson *et al.*, 2010b). Metatranscriptomics is an RNA-based analysis technique being used to assess the functional and metabolic diversity of intestinal microbial communities (Gosalbes *et al.*, 2011). However, this technique is limited because it cannot differentiate between genes expressed or repressed by the microbiota at the sampling times and also by the inherent difficulties associated with working with RNA and its half-life (Simon & Daniel, 2011).

Despite the limitations to both metagenomics and metatranscriptomic technologies, they provide a platform for further studies and future technological breakthroughs for the analysis of the complex gastrointestinal microbiota.

1.7.2 The microbiota of humans and other animals

1.7.2.1 Microbiota function

The commensal intestinal microbiota of both humans and animals serves many functions (homeostasis, immunomodulatory); primary among these functions is the digestion of food components that are indigestible by human enzymes. The short chain fatty acids (SCFA), by-products of microbial digestion, are absorbed by the hosts colonocytes as a source of energy (Wong *et al.*, 2006). The commensal microbiota of humans and animals also acts a barrier to colonisation by pathogens (O'Hara & Shanahan, 2006). The host gastrointestinal microbiota is vitally important in regulating host health and efficient digestion of nutrients.

1.7.2.2 Microbiota composition

Colonisation of the gastrointestinal tract begins at birth and with an estimated 10^{10} – 10^{14} bacteria CFU/ml present in the human microbiota. The intestinal microbiota is colonised also by archaea, fungi and yeasts. Up to 800 bacterial species are thought to comprise the human microbiota (Bäckhed *et al.*, 2005). As mentioned previously, culture independent techniques have afforded researchers the capability to examine the microbiota of large numbers of humans and animals in-depth and with relative ease (Ley *et al.*, 2008; Mitreva, 2012). Analysis of the microbial composition of

multiple regions within the human body revealed that no single taxon was conserved across each region (Mitreva, 2012). This is in agreement with an earlier study which failed to identify a universal core microbiota (Tap *et al.*, 2009). The dominance of the *Firmicutes* and *Bacteroidetes* phyla present in the human and animal intestinal microbiota (Ley *et al.*, 2008) is not consistently observed (Andersson *et al.*, 2008; Shepherd *et al.*, 2012). The dominant genera common in the microbiota of humans and animals include *Ruminococcus*, *Bacteroides*, *Alistipes* and Akkermansia (Dowd *et al.*, 2008; Rajilić-Stojanović *et al.*, 2013; Shepherd *et al.*, 2012). The microbiota composition is also subject to temporal variation throughout the lifetime of the individual from infants to adults to the elderly (O'Toole & Claesson, 2010). The intestinal tract encompasses the second largest set of nerve cells outside the brain and therefore the brain-gut axis is very important and has been associated with behavioural and mental function issues (Cryan & O'Mahony, 2011). Thus it is vital to identify the composition of the intestinal microbiota as a way of ameliorating various diseases and health issues.

1.7.3 Alteration of the microbiota and diseases

Disturbances in the microbiota or dysbiosis is a common hypothetical aetiology for gastrointestinal-associated diseases. To date, the role of the microbiota in some diseases has not been clearly defined.

Two of the main non-genetic factors linked to the development of IBD are the gastrointestinal microbiota and diet. Ulcerative colitis (UC) and Crohn's disease (CD) are forms of IBD characterised by ulcers/lesions in the colon and chronic abdominal pain, respectively. Reduction in proportional abundance of Cluster IV *Clostridia*, for example *Faecalibacterium prausnitzii*, is associated with the inflammation caused by CD (Sokol *et al.*, 2009; Willing *et al.*, 2009a). Consumption of diets high in fats and meat were also associated with an increased risk of developing UC and CD. Diets high in fibre, fruits and vegetables were associated with lowering the risk of developing these conditions (Hou *et al.*, 2011). This has led many to hypothesise that the development of IBD is linked with low fibre, high fat "Western" style diets (Wu *et al.*, 2013).

IBS is characterised by severe abdominal pain and discomfort, as well as bloating. There is currently no cure for IBS but symptoms can be alleviated by modulating the

diet of the affected individual (Gibson *et al.*, 2013; Grundmann & Yoon, 2010). This indicates there is an interplay between diet and host intestinal microbiota in the development of IBS. The intestinal microbiota of individuals with IBS differs from that of the healthy controls (Rajilić-Stojanović *et al.*, 2011). The differences in the microbiota resulted from an increase in the *Firmicutes*:*Bacteroidetes* ratio stemming from an increase in genera like *Dorea*, *Ruminococcus* and *Clostridium*. The *Bacteroidetes* phylum decreased in proportion in IBS patients along with other genera including *Bifidobacterium* and *Faecalibacterium* (Rajilić-Stojanović *et al.*, 2011). Decreasing proportions of the *Faecalibacterium* genus has been associated with other health conditions like obesity (Balamurugana *et al.*, 2010).

Obesity is now considered a worldwide epidemic and is a major concern for researchers and healthcare professionals, alike. The increase in weight gain leading to an obese phenotype is as a result of an energy imbalance from the food consumed. The intestinal microbiota of the host is responsible for converting excess energy to fat storage (Bäckhed *et al.*, 2005). An increase in food intake has correlated with a reduction in the diversity of the microbiota (Turnbaugh *et al.*, 2009). Studies have implicated a reduction in the proportion of *Bacteroidetes* in the microbiota and therefore, an increase in the *Firmicutes*:*Bacteroidetes* ratio as a factor in weight gain (Ley *et al.*, 2005). The reduction in the *Bacteroidetes* phylum abundance in the microbiota is contentious issue with other studies noting a significant increase in this phylum in obese individuals (Schwiertz *et al.*, 2010). Similarly, the link between the *Firmicutes Bacteroidetes* ratio and weight gain is also controversial with some studies failing to identify a difference in the proportions of these phyla in lean and obese individuals (Duncan *et al.*, 2008). Although considered to be a beneficial commensal genus, an increase in the proportions of *Lactobacillus* in the microbiota may also be a contributing factor in obesity (Armougom *et al.*, 2009; Million *et al.*, 2012a; Million *et al.*, 2012b). A decrease in the proportions of *Faecalibacterium prausnitzii* was found to be associated with obesity in children (Balamurugana *et al.*, 2010). However, further studies on lean and obese humans and animals are needed to verify the effect that *Firmicutes Bacteroidetes* ratio and the proportions of *Lactobacillus* or *Faecalibacterium* have on weight gain.

Obesity is commonly associated with an increased risk of Type 2 diabetes mellitus (T2DM), a metabolic disorder correlated with a high blood glucose level. Differences between the microbiota of healthy individuals and T2DM patients resulted from a

reduction in the proportions of the *Firmicutes* phylum in T2DM patients, particularly the *Clostridia* class and an increase in the *Betaproteobacteria* class (Larsen *et al.*, 2010). Individuals with T2DM were also found to have a moderate dysbiosis in their microbiota which correlated with a reduced population of butyrate-producers and an increase in the proportion of pathogens (Qin *et al.*, 2012).

Animals, including horses, are also prone to gastrointestinal disease and there is a correlation between dietary intake and disease proliferation in these animals. Laminitis is the "failure of the distal phalanx to maintain its attachment to the lamellae of the inner hoof wall, causing unrelenting pain and a characteristic lameness" (Pollitt, 2004). It is hypothesised that there is a causal link between dietary fructans and the disease laminitis (Milinovich *et al.*, 2006; Pollitt, 2004). The most common hypothesis is that of pasture-induced laminitis, whereby carbohydrate overload may occur when non-structural carbohydrate (NSC) is present at a level greater than 0.4% of the animal's body weight (Potter *et al.*, 1992). *Streptococcus* spp. have also been highlighted as being possible participants in the cycle leading to laminitis (Milinovich *et al.*, 2008a; Milinovich *et al.*, 2010).

It is clear that while further studies are needed, the microbiota of the host is important for maintaining health and that dysbiosis can lead to inflammatory and metabolic disorders. Particular attention should be paid to the *Firmicutes* and *Bacteroidetes* phyla which appear to be the regulators of homeostasis *in vivo*.

1.7.4 Effects of diet on the microbiota

The composition of the diet of the host effects the composition and diversity of the microbiota. Notwithstanding the great inter-individual variability of the microbiome, the composition of a given microbiota can be classified into one of just three "enterotypes" (Arumugam *et al.*, 2011). Each enterotype is characterised as having very high proportions of a single genus (*Bacteroides*, *Prevotella* or *Ruminococcus*) (Arumugam *et al.*, 2011). Wu and colleagues determined that habitual or long-term dietary intake assembled the faecal microbiota into two primary enterotypes (Wu *et al.*, 2011). Diets high in proteins and saturated fats associated with the *Bacteroides* enterotype and diets high in carbohydrates were associated with the *Prevotella* enterotype (Wu *et al.*, 2011). Analysis of the elderly microbiome identified four dietary groupings with the high fat and low fibre group associated with low microbial

diversity and poor health (Claesson *et al.*, 2012). Ley and colleagues showed that diet is a key factor in determining the diversity of the microbiota in ruminant, hindgut fermenter and mono-gastric animals (Ley *et al.*, 2008). Animals consuming a polysaccharide diet had a more diverse microbiota compared to those consuming a meat-based diet (Ley *et al.*, 2008). Similar differences were seen between the microbiota of children from different geographic regions consuming Western diets or an agrarian diet (high fibre, low fat) (De Filippo *et al.*, 2010). Consumption of dietary fibre has an effect on the microbiota of animals altering the dominant phyla (Middelbos *et al.*, 2010). The data would suggest that a more agrarian-based diet rich in fibre and low in saturated fats would promote a higher diversity in the microbiome and may improve gut health.

Additional research indicates that the microbiota of the host can be beneficially modulated using dietary supplements like prebiotics. In the short-term, prebiotics have increased the proportions of beneficial bacteria in the microbiota of infants (Wainwright, 2006). Galactooligosaccharides consumed by healthy adults were bifidogenic. However, the effect was dose dependent (Davis *et al.*, 2011). A bifidogenic response to galactooligosaccharides consumption was also noted in the elderly and animals (Biagi *et al.*, 2013; Walton *et al.*, 2011). However, prebiotics can have potential negative effects on the microbiome and host health as well (Firkins *et al.*, 2008; Milinovich *et al.*, 2006; Milinovich *et al.*, 2007; Milinovich *et al.*, 2008b; Pollitt, 2004; Rada *et al.*, 2008). Similarly, some prebiotics were ineffective on the host microbiota studied and this may suggest that prebiotics need to be paired with host animals with a particular baseline microbiota (Mountzouris *et al.*, 2006). In human trials, the effect of prebiotics on the faecal microbiota of 14 obese males depended in part on their starting microbiota (Duncan *et al.*, 2007). Further studies are required before prebiotics can be deemed reliable and effective as modulators of human and animal health. But it is clear that diet, host health and the microbiota are closely linked and that disturbance or alteration to one of these affects the others.

1.8 References

Andersen, J. M., Barrangou, R., Abou Hachem, M., Lahtinen, S., Goh, Y. J., Svensson, B. & Klaenhammer, T. R. (2011). Transcriptional and functional analysis of galactooligosaccharide uptake by lacS in *Lactobacillus acidophilus*. *Proc Natl Acad Sci USA* **108**, 17785-17790.

Andersson, A. F., Lindberg, M., Jakobsson, H., Bäckhed, F., Nyrén, P. & Engstrand, L. (2008). Comparative analysis of human gut microbiota by barcoded pyrosequencing. *PLOS One* **3**, e2836.

Armougom, F., Henry, M., Vialettes, B., Raccah, D. & Raoult, D. (2009). Monitoring bacterial community of human gut microbiota reveals an increase in *Lactobacillus* in obese patients and Methanogens in anorexic patients. *PLOS One* **4**, e7125.

Årsköld, E., Lohmeier-Vogel, E., Cao, R., Roos, S., Rådström, P. & van Niel, E. W. J. (2008). Phosphoketolase pathway dominates in *Lactobacillus reuteri* ATCC 55730 containing dual pathways for glycolysis. *J Bacteriol* **190**, 206-212.

Arumugam, M., Raes, J., Pelletier, E. & other authors (2011). Enterotypes of the human gut microbiome. *Nature* **473**, 174-180.

Azcarate-Peril, M. A., Altermann, E., Goh, Y. J. & other authors (2008). Analysis of the genome sequence of *Lactobacillus gasseri* ATCC 33323 reveals the molecular basis of an autochthonous intestinal organism. *Appl Environ Microbiol* **74**, 4610-4625.

Bäckhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A. & Gordon, J. I. (2005). Host-bacterial mutualism in the human intestine. *Science* **307**, 1915-1920.

Balamurugana, R., Georgea, G., Kabeerdossa, J., Hepsibaa, J., Chandragunasekarana, A. & Ramakrishnaa, B. S. (2010). Quantitative

differences in intestinal *Faecalibacterium prausnitzii* in obese Indian children. *Brit J Nutr* **103**, 335-338.

Barrangou, R., Altermann, E., Hutkins, R., Cano, R. & Klaenhammer, T. R. (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., Azcarate-Peril, M. A., Duong, T., Conners, S. B., Kelly, R. M. & Klaenhammer, T. R. (2006). Global analysis of carbohydrate utilization by *Lactobacillus acidophilus* using cDNA microarrays. *P Natl Acad Sci USA* **103**, 3816-3821.

Biagi, G., Cipollini, I., Bonaldo, A., Grandi, M., Pompei, A., Stefanelli, C. & Zaghini, G. (2013). Effect of feeding a selected combination of galacto-oligosaccharides and a strain of *Bifidobacterium pseudocatenulatum* on the intestinal microbiota of cats. *Am J Vet Res* **74**, 90-95.

Brückner, R. & Titgemeyer, F. (2002). Carbon catabolite repression in bacteria: choice of the carbon source and autoregulatory limitation of sugar utilization. *FEMS Microbiol Lett* **209**, 141-148.

Burns, P., Sánchez, B., Vinderola, G., Ruas-Madiedo, P., Ruiz, L., Margolles, A., Reinheimer, J. & de los Reyes-Gavilán, C. G. (2010). Inside the adaptation process of *Lactobacillus delbrueckii* subsp. *lactis* to bile. *Int J Food Microbiol* **142**, 132-141.

Cai, H., Rodríguez, B. T., Zhang, W., Broadbent, J. R. & Steele, J. L. (2007). Genotypic and phenotypic characterization of *Lactobacillus casei* strains isolated from different ecological niches suggests frequent recombination and niche specificity. *Microbiology* **153**, 2655-2665.

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

Chaillou, S., Champomier-Verges, M.-C., Cornet, M. & other authors (2005). The complete genome sequence of the meat-borne lactic acid bacterium *Lactobacillus sakei* 23K. *Nat Biotech* **23**, 1527-1533.

Chang, A. B., Lin, R., Studley, W. K., Tran, C. V. & Saier, J. M. H. (2004). Phylogeny as a guide to structure and function of membrane transport proteins (Review). *Molecular Membrane Biology* **21**, 171-181.

Chiaramonte, F., Anglade, P., Baraige, F., Gratadoux, J.-J., Langella, P., Champomier-Vergès, M.-C. & Zagorec, M. (2010). Analysis of *Lactobacillus sakei* mutants selected after adaptation to the gastrointestinal tracts of axenic mice. *Appl Environ Microbiol* **76**, 2932-2939.

Claesson, M. J., Li, Y., Leahy, S. & other authors (2006). Multireplicon genome architecture of *Lactobacillus salivarius*. *Proc Natl Acad Sci USA* **103**, 6718-6723.

Claesson, M. J., Wang, Q., O'Sullivan, O., Greene-Diniz, R., Cole, J. R., Ross, R. P. & O'Toole, P. W. (2010). Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Res* **38**, e200-e200.

Claesson, M. J., Jeffery, I. B., Conde, S. & other authors (2012). Gut microbiota composition correlates with diet and health in the elderly. *Nature* **488**, 178-184.

Cryan, J. F. & O'Mahony, S. M. (2011). The microbiome-gut-brain axis: from bowel to behavior. *Neurogastroent Motil* **23**, 187-192.

Davis, L. M. G., Martínez, I., Walter, J., Goin, C. & Hutkins, R. W. (2011). Barcoded pyrosequencing reveals that consumption of galactooligosaccharides results in a highly specific bifidogenic response in humans. *PLOS One* **6**, e25200.

De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J. B., Massart, S., Collini, S., Pieraccini, G. & Lionetti, P. (2010). Impact of diet in

shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *P Natl Acad Sci USA* **107**, 14691-14696.

Denou, E., Pridmore, R. D., Berger, B., Panoff, J.-M., Arigoni, F. & Brüssow, 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.

Deutscher, J. (2008). The mechanisms of carbon catabolite repression in bacteria. *Curr Opin Microbiol* **11**, 87-93.

Dills, S. S., Apperson, A., Schmidt, M. R. & Saier, M. H., Jr. (1980). Carbohydrate transport in bacteria. *Microbiol Mol Biol Rev* **44**, 385-418.

Dowd, S., Callaway, T., Wolcott, R., Sun, Y., McKeehan, T., Hagevoort, R. & Edrington, T. (2008). Evaluation of the bacterial diversity in the feces of cattle using 16S rDNA bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP). *BMC Microbiol* **8**, 125.

Duncan, S. H., Belenguer, A., Holtrop, G., Johnstone, A. M., Flint, H. J. & Lobley, G. E. (2007). Reduced dietary intake of carbohydrates by obese subjects results in decreased concentrations of butyrate and butyrate-producing bacteria in feces. *Appl Environ Microbiol* **73**, 1073-1078.

Duncan, S. H., Lobley, G. E., Holtrop, G., Ince, J., Johnstone, A. M., Louis, P. & Flint, H. J. (2008). Human colonic microbiota associated with diet, obesity and weight loss. *Int J Obes* **32**, 1720-1724.

Felis, G. E. & Dellaglio, F. (2007). Taxonomy of Lactobacilli and Bifidobacteria. *Curr Issues Intest Microbiol* **8**, 44-61.

Firkins, J., Karnati, S. & Yu, Z. (2008). Linking rumen function to animal response by application of metagenomics techniques. *Aust J Exp Agric* **48**, 711-721.

Forde, B. M., Neville, B. A., O' Donnell, M. M., Riboulet-Bisson, E., Claesson, M. J., Coghlan, A., Ross, R. P. & O' Toole, P. W. (2011). Genome sequences and comparative genomics of two *Lactobacillus ruminis* strains from the bovine and human intestinal tracts. *Microb Cell Fact* **10**, S13.

Fraher, M. H., O'Toole, P. W. & Quigley, E. M. M. (2012). Techniques used to characterize the gut microbiota: a guide for the clinician. *Nat Rev Gastroenterol Hepatol* **9**, 312-322.

Frese, S. A., Benson, A. K., Tannock, G. W. & other authors (2011). The evolution of host specialization in the vertebrate gut symbiont *Lactobacillus reuteri*. *PLOS Genet* **7**, e1001314.

Gibson, G. R. & Roberfroid, M. B. (1995). Dietary modulation of the human colonic microbiota: Introducing the concept of prebiotics. *J Nutr* **125**, 1401-1412.

Gibson, G. R., Probert, H. M., Loo, J. V., Rastall, R. A. & Roberfroid, M. B. (2004). Dietary modulation of the human colonic microbiota: updating the concept of prebiotics. *Nutr Res Rev* **17**, 259-275.

Gibson, P. R., Barrett, J. S. & Muir, J. G. (2013). Functional bowel symptoms and diet. *Intern Med J* **43**, 1067-1074.

Glaser, P., Frangeul, L., Buchrieser, C. & other authors (2001). Comparative Genomics of Listeria Species. *Science* **294**, 849-852.

Gorke, B. & Stulke, J. (2008). Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat Rev Microbiol* **6**, 613-624.

Gosalbes, M. J., Durbán, A., Pignatelli, M., Abellan, J. J., Jimenez-Hernandez, N., Perez-Cobas, A. E., Latorre, A. & Moya, A. (2011). Metatranscriptomic approach to analyze the functional human gut microbiota. *PLOS One* **6**, e17447.

Grundmann, O. & Yoon, S. L. (2010). Irritable bowel syndrome: Epidemiology, diagnosis and treatment: An update for health-care practitioners. *J Gastroen Hepatol* **25**, 691-699.

Guillot, A., Gitton, C., Anglade, P. & Mistou, M.-Y. (2003). Proteomic analysis of *Lactococcus lactis*, a lactic acid bacterium. *Proteomics* **3**, 337-354.

Gunsalus, I. C., Horecker, B. L. & Wood, W. A. (1955). Pathways of carbohydrate metabolism in microorganisms. *Microbiol Mol Biol Rev* **19**, 79-128.

Hao, P., Zheng, H., Yu, Y. & other authors (2011). Complete sequencing and pan-genomic analysis of *Lactobacillus delbrueckii* subsp. *bulgaricus* reveal its genetic basis for industrial yogurt production. *PLOS one* **6**, e15964.

Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J* **280**, 309-316.

Henrissat, B. & Bairoch, A. (1996). Updating the sequence-based classification of glycosyl hydrolases. *Biochem J* **316 (Pt 2)**, 695-696.

Holzapfel, W. H. N. & Wood, B. J. (1995). The Lactic Acid Bacteria: The genera of lactic acid bacteria. In *The Lactic Acid Bacteria: The genera of lactic acid bacteria*, pp. 420: Springer.

Hou, J. K., Abraham, B. & El-Serag, H. (2011). Dietary intake and risk of developing inflammatory bowel disease: A systematic review of the literature. *Am J Gastroenterol* **106**, 563-573.

Jeffery, I. B., O'Toole, P. W., Öhman, L., Claesson, M. J., Deane, J., Quigley, E. M. M. & Simrén, M. (2012). An irritable bowel syndrome subtype defined by species-specific alterations in faecal microbiota. *Gut* **61**, 997-1006.

Jiménez, E., Langa, S., Martín, V., Arroyo, R., Martín, R., Fernández, L. & Rodríguez, J. M. (2010a). Complete genome sequence of *Lactobacillus fermentum* CECT 5716, a probiotic strain isolated from human milk. *J Bacteriol* **192**, 4800.

Jiménez, E., Martín, R., Maldonado, A., Martín, V., Gómez de Segura, A., Fernández, L. & Rodríguez, J. M. (2010b). Complete genome sequence of *Lactobacillus salivarius* CECT 5713, a probiotic strain isolated from human milk and infant feces. *J Bacteriol* **192**, 5266-5267.

Jojima, T., Omumasaba, C., Inui, M. & Yukawa, H. (2010). Sugar transporters in efficient utilization of mixed sugar substrates: current knowledge and outlook. *Appl Microbiol Biot* **85**, 471-480.

Kandler, O. (1983). Carbohydrate metabolism in lactic acid bacteria. *A Van Leeuw* **49**, 209-224.

Kandler, O. & Weiss, N. (1986). Genus *Lactobacillus*. In *Bergey's manual of systematic bacteriology*, pp. 1209–1234.

Kankainen, M., Paulin, L., Tynkkynen, S. & other authors (2009). Comparative genomic analysis of *Lactobacillus rhamnosus* GG reveals pili containing a human-mucus binding protein. *Proc Natl Acad Sci USA* **106**, 17193-17198.

Kant, R., Paulin, L., Alatalo, E., de Vos, W. M. & Palva, A. (2011). Genome sequence of *Lactobacillus amylovorus* GRL1112. *J Bacteriol* **193**, 789-790.

Kim, J.-H., Shoemaker, S. P. & Mills, D. A. (2009). Relaxed control of sugar utilization in *Lactobacillus brevis*. *Microbiology* **155**, 1351-1359.

Kim, J.-H., Block, D. E., Shoemaker, S. P. & Mills, D. A. (2010). Atypical ethanol production by carbon catabolite derepressed lactobacilli. *Bioresource Technol* **101**, 8790-8797.

Kleerebezem, M., Boekhorst, J., van Kranenburg, R. & other authors (2003). Complete genome sequence of *Lactobacillus plantarum* WCFS1. *Proc Natl Acad Sci USA* **100**, 1990-1995.

Konings, W. N. (2006). Microbial transport: adaptations to natural environments. *A Van Leeuw* **90**, 325-342.

Larsen, N., Vogensen, F. K., van den Berg, F. W. J. & other authors (2010). Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. *PLOS One* **5**, e9085.

Lee, L. J., Hansen, J. B., Jagusztyn-Krynicka, E. K. & Chassy, B. M. (1982). Cloning and expression of the beta-D-phosphogalactoside galactohydrolase gene of Lactobacillus casei in Escherichia coli K-12. *J Bacteriol* **152**, 1138-1146.

Lengeler, J. W. (1993). Carbohydrate transport in bacteria under environmental conditions, a black box? *Antonie Van Leeuwenhoek* **63**, 275-288.

Ley, R. E., Bäckhed, F., Turnbaugh, P., Lozupone, C. A., Knight, R. D. & Gordon, J. I. (2005). Obesity alters gut microbial ecology. *PNAS* **102**, 11070-11075.

Ley, R. E., Hamady, M., Lozupone, C. & other authors (2008). Evolution of mammals and their gut microbes. *Science* **320**, 1647.

Li, Y., Canchaya, C., Fang, F., Raftis, E., Ryan, K. A., van Pijkeren, J. P., van Sinderen, D. & O'Toole, P. W. (2007). Distribution of megaplasmids in *Lactobacillus salivarius* and other lactobacilli. *J Bacteriol* **189**, 6128-6139.

Ludwig, H., Homuth, G., Schmalisch, M., Dyka, F. M., Hecker, M. & Stülke, J. (2001). Transcription of glycolytic genes and operons in *Bacillus subtilis*: evidence for the presence of multiple levels of control of the gapA operon. *Mol Microbiol* **41**, 409-422.

Majumder, A., Sultan, A., Jersie-Christensen, R. R., Ejby, M., Schmidt, B. G., Lahtinen, S. J., Jacobsen, S. & Svensson, B. (2011). Proteome reference map of Lactobacillus acidophilus NCFM and quantitative proteomics towards understanding the prebiotic action of lactitol. *Proteomics* **11**, 3470-3481.

Makarova, K., Slesarev, A., Wolf, Y. & other authors (2006). Comparative genomics of the lactic acid bacteria. *Proc Natl Acad Sci USA* **103**, 15611-15616.

Manning, T. S. & Gibson, G. R. (2004). Microbial-gut interactions in health and disease. Prebiotics. *Best Pract Res Clin Gastroenterol* **18**, 287-298.

Marasco, R., Muscariello, L., Varcamonti, M., De Felice, M. & Sacco, M. (1998). Expression of the bglH gene of *Lactobacillus plantarum* is controlled by carbon catabolite repression. *J Bacteriol* **180**, 3400-3404.

Marco, M. L., Peters, T. H. F., Bongers, R. S., Molenaar, D., Van Hemert, S., Sonnenburg, J. L., Gordon, J. I. & Kleerebezem, M. (2009). Lifestyle of Lactobacillus plantarum in the mouse caecum. *Environ Microbiol* **11**, 2747-2757.

Marco, M. L., de Vries, M. C., Wels, M., Molenaar, D., Mangell, P., Ahrne, S., de Vos, W. M., Vaughan, E. E. & Kleerebezem, M. (2010). Convergence in probiotic *Lactobacillus* gut-adaptive responses in humans and mice. *ISME J* **4**, 1481-1484.

Middelbos, I. S., Boler, B. M. V., Qu, A., White, B. A., Swanson, K. S. & Fahey Jr, G. C. (2010). Phylogenetic characterization of fecal microbial communities of dogs fed diets with or without supplemental dietary fiber using 454 pyrosequencing. *PLOS One* **5**, e9768.

Milinovich, G., Trott, D., Burrell, P., Van Eps, A., Thoefner, M., Blackall, L., Al Jassim, R., Morton, J. & Pollitt, C. (2006). Changes in equine hindgut bacterial populations during oligofructose induced laminitis. *Environ Microbiol* **8**, 885-898.

Milinovich, G. J., Trott, D. J., Burrell, P. C., Croser, E. L., Al Jassim, R. A. M., Morton, J. M., Van Eps, A. W. & Pollitt, C. C. (2007). Fluorescence *in situ*

hybridization analysis of hindgut bacteria associated with the development of equine laminitis. *Environ Microbiol* **9**, 2090-2100.

Milinovich, G. J., Burrell, P. C., Pollitt, C. C., Bouvet, A. & Trott, D. J. (2008a). *Streptococcus henryi* sp. nov. and *Streptococcus caballi* sp. nov., isolated from the hindgut of horses with oligofructose-induced laminitis. *Int J Syst Evol Microbiol* **58**, 262-266.

Milinovich, G. J., Burrell, P. C., Pollitt, C. C., Klieve, A. V., Blackall, L. L., Ouwerkerk, D., Woodland, E. & Trott, D. J. (2008b). Microbial ecology of the equine hindgut during oligofructose-induced laminitis. *ISME J* **2**, 1089-1100.

Milinovich, G. J., Klieve, A. V., Pollitt, C. C. & Trott, D. J. (2010). Microbial events in the hindgut during carbohydrate-induced equine laminitis. *Vet Clin North Am-Equine* **26**, 79-94.

Million, M., Angelakis, E., Paul, M., Armougom, F., Leibovici, L. & Raoult, D. (2012a). Comparative meta-analysis of the effect of *Lactobacillus* species on weight gain in humans and animals. *Microb Pathogenesis* **53**, 100-108.

Million, M., Maraninchi, M., Henry, M. & other authors (2012b). Obesity-associated gut microbiota is enriched in *Lactobacillus reuteri* and depleted in *Bifidobacterium animalis* and *Methanobrevibacter smithii*. *Int J Obes* **36**, 817-825.

Mitreva, M. (2012). Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207-214.

Molenaar, D., Bringel, F., Schuren, F. H., de Vos, W. M., Siezen, R. J. & Kleerebezem, M. (2005). Exploring *Lactobacillus plantarum* genome diversity by using microarrays. *J Bacteriol* **187**, 6119-6127.

Morita, H., Toh, H., Fukuda, S. & other authors (2008). Comparative genome analysis of *Lactobacillus reuteri* and *Lactobacillus fermentum* reveal a genomic island for reuterin and cobalamin production. *DNA Res* **15**, 151-161.

Morita, H., Toh, H., Oshima, K., Murakami, M., Taylor, T. D., Igimi, S. & Hattori, M. (2009). Complete genome sequence of the probiotic *Lactobacillus rhamnosus* ATCC 53103. *J Bacteriol* **191**, 7630-7631.

Mountzouris, K. C., Balaskas, C., Fava, F., Tuohy, K. M., Gibson, G. R. & Fegeros, K. (2006). Profiling of composition and metabolic activities of the colonic microflora of growing pigs fed diets supplemented with prebiotic oligosaccharides. *Anaerobe* **12**, 178-185.

Naterstad, K., Rud, I., Kvam, I. & Axelsson, L. (2007). Characterisation of the gap operon from *Lactobacillus plantarum* and *Lactobacillus sakei*. *Curr Microbiol* **54**, 180-185.

Ndagijimana, M., Laghi, L., Vitali, B., Placucci, G., Brigidi, P. & Guerzoni, M. E. (2009). Effect of a synbiotic food consumption on human gut metabolic profiles evaluated by 1H Nuclear Magnetic Resonance spectroscopy. *Int J Food Microbiol* **134**, 147-153.

Nicolas, P., Bessières, P., Ehrlich, S. D., Maguin, E. & Van De Guchte, M. (2007). Extensive horizontal transfer of core genome genes between two Lactobacillus species found in the gastrointestinal tract. *BMC Evol Biol* **7**, 141.

O' Sullivan, O., O' Callaghan, J., Sangrador-Vegas, A. & other authors (2009). Comparative genomics of lactic acid bacteria reveals a niche-specific gene set. *BMC Microbiol* **9**, 50.

O'Hara, A. M. & Shanahan, F. (2006). The gut flora as a forgotten organ. *EMBO Rep* **7**, 688-693.

O'Toole, P. W. & Claesson, M. J. (2010). Gut microbiota: Changes throughout the lifespan from infancy to elderly. *Int Dairy J* **20**, 281-291.

O'Donnell, M. M., Forde, B. M., Neville, B. A., Ross, R. P. & O'Toole, P. W. (2011). Carbohydrate catabolic flexibility in the mammalian intestinal commensal *Lactobacillus ruminis* revealed by fermentation studies aligned to genome annotations. *Microb Cell Fact* **10**, S12.

Payne, A. N., Chassard, C. & Lacroix, C. (2012). Gut microbial adaptation to dietary consumption of fructose, artificial sweeteners and sugar alcohols: implications for host-microbe interactions contributing to obesity. *Obes Rev* **13**, 799-809.

Pollitt, C. C. (2004). Equine laminitis. *Clin Tech Equine Pract* **3**, 34-44.

Poolman, B. (2002). Transporters and their roles in LAB cell physiology. *A Van Leeuw* **82**, 147-164.

Posno, M., Heuvelmans, P. T., van Giezen, M. J., Lokman, B. C., Leer, R. J. & Pouwels, P. H. (1991). Complementation of the inability of *Lactobacillus* strains to utilize D-xylose with D-xylose catabolism-encoding genes of *Lactobacillus pentosus*. *Appl Environ Microbiol* **57**, 2764-2766.

Postma, P. W., Lengeler, J. W. & Jacobson, G. R. (1993). Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. *Microbiol Mol Biol Rev* **57**, 543-594.

Potter, G., Arnold, F., Householder, D., Hansen, D. & Brown, K. (1992). Digestion of starch in the small or large intestine of the equine. *Pferdeheilkunde* **1**.

Pridmore, R. D., Berger, B., Desiere, F. & other authors (2004). The genome sequence of the probiotic intestinal bacterium *Lactobacillus johnsonii* NCC 533. *Proc Natl Acad Sci USA* **101**, 2512-2517.

Qin, J., Li, Y., Cai, Z. & other authors (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490**, 55-60.

Rackis Joseph, J. (1975). Oligosaccharides of food legumes: Alpha-galactosidase activity and the flatus problem. In *Physiological Effects of Food Carbohydrates*, pp. 207-222. Edited by Allene Jeanes & J. Hodge: American Chemical Society.

Rada, V., Nevoral, J., Trojanová, I., Tománková, E., Šmehilová, M. & Killer, J. (2008). Growth of infant faecal bifidobacteria and clostridia on prebiotic oligosaccharides in *in vitro* conditions. *Anaerobe* **14**, 205-208.

Rajilić-Stojanović, M., Biagi, E., Heilig, H. G. H. J., Kajander, K., Kekkonen, R. A., Tims, S. & de Vos, W. M. (2011). Global and deep molecular analysis of microbiota signatures in fecal samples from patients with irritable bowel syndrome. *Gastroenterology* **141**, 1792-1801.

Rajilić-Stojanović, M., Heilig, H. G. H. J., Tims, S., Zoetendal, E. G. & de Vos, W. M. (2013). Long-term monitoring of the human intestinal microbiota composition. *Environ Microbiol* **15**, 1146-1159.

Rankin, D. J., Rocha, E. P. C. & Brown, S. P. (2011). What traits are carried on mobile genetic elements, and why? *Heredity* **106**, 1-10.

Reuter, G. (2001). The *Lactobacillus* and *Bifidobacterium* microflora of the human intestine: composition and succession. *Curr Issues Intest Microbiol* **2**, 43-53.

Riesenfeld, C. S., Schloss, P. D. & Handelsman, J. (2004). Metagenomics: genomic analysis of microbial communities. *Annu Rev Genet* **38**, 525-552.

Saier Jr, M. H., Chauvaux, S., Cook, G. M., Deutscher, J., Paulsen, I. T., Reizer, J. & Ye, J. J. (1996). Catabolite repression and inducer control in Gram-positive bacteria. *Microbiology* **142**, 217.

Saier, M. H., Tran, C. V. & Barabote, R. D. (2006). TCDB: the Transporter Classification Database for membrane transport protein analyses and information. *Nucleic Acids Res* **34**, D181-D186.

Saulnier, D. M., Molenaar, D., de Vos, W. M., Gibson, G. R. & Kolida, S. (2007). Identification of prebiotic fructooligosaccharide metabolism in *Lactobacillus plantarum* WCFS1 through microarrays. *Appl Environ Microbiol* **73**, 1753-1765.

Schreiber, W. & Dürre, P. (2000). Differential expression of genes within the gap operon of *Clostridium acetobutylicum*. *Anaerobe* **6**, 291-297.

Schwiertz, A., Taras, D., Schäfer, K., Beijer, S., Bos, N. A., Donus, C. & Hardt, P. D. (2010). Microbiota and SCFA in lean and overweight healthy subjects. *Obesity* **18**, 190-195.

Shepherd, M. L., Swecker Jr, W. S., Jensen, R. V. & Ponder, M. A. (2012). Characterization of the fecal bacteria communities of forage fed horses by pyrosequencing of 16S rRNA V4 gene amplicons. *FEMS Microbiol Lett* **326**, 62-68.

Siezen, R. & van Hylckama Vlieg, J. (2011). Genomic diversity and versatility of *Lactobacillus plantarum*, a natural metabolic engineer. *Microb Cell Fact* **10**, S3.

Siezen, R. J., Tzeneva, V. A., Castioni, A. & other authors (2010). Phenotypic and genomic diversity of *Lactobacillus plantarum* strains isolated from various environmental niches. *Environ Microbiol* **12**, 758-773.

Simon, C. & Daniel, R. (2011). Metagenomic Analyses: Past and Future Trends. *Appl Environ Microbiol* **77**, 1153-1161.

Sokol, H., Seksik, P., Furet, J. P. & other authors (2009). Low counts of *Faecalibacterium prausnitzii* in colitis microbiota. *Inflamm Bowel Dis* **15**, 1183-1189.

Stiles, M. E. (1996). Biopreservation by lactic acid bacteria. *A Van Leeuw* **70**, 331-345.

Stiles, M. E. & Holzapfel, W. H. (1997). Lactic acid bacteria of foods and their current taxonomy. *Int J Food Microbiol* **36**, 1-29.

Stulke, J. & Hillen, W. (1999). Carbon catabolite repression in bacteria. *Curr Opin Microbiol* **2**, 195-201.

Tannock, G. W., Wilson, C. M., Loach, D., Cook, G. M., Eason, J., O'Toole, P. W., Holtrop, G. & Lawley, B. (2012). Resource partitioning in relation to cohabitation of *Lactobacillus* species in the mouse forestomach. *ISME J* **6**, 927-938.

Tap, J., Mondot, S., Levenez, F. & other authors (2009). Towards the human intestinal microbiota phylogenetic core. *Environ Microbiol* **11**, 2574-2584.

Tauxe, R. V., Cavanagh, T. R. & Cohen, M. L. (1989). Interspecies gene transfer *in vivo* producing an outbreak of multiply resistant Shigellosis. *J Infect Diseases* **160**, 1067-1070.

Teixeira, J. S., McNeill, V. & Gänzle, M. G. (2012). Levansucrase and sucrose phoshorylase contribute to raffinose, stachyose, and verbascose metabolism by lactobacilli. *Food Microbiology* **31**, 278-284.

Teusink, B., Wiersma, A., Jacobs, L., Notebaart, R. A. & Smid, E. J. (2009). Understanding the adaptive growth strategy of *Lactobacillus plantarum* by *in silico* optimisation. *PLoS Comput Biol* **5**, e1000410.

Titgemeyer, F. & Hillen, W. (2002). Global control of sugar metabolism: a gram-positive solution. *A Van Leeuw* **82**, 59-71.

Tottey, W., Denonfoux, J., Jaziri, F. & other authors (2013). The Human Gut Chip "HuGChip", an explorative phylogenetic microarray for determining gut microbiome diversity at family level. *PLOS One* **8**, e62544.

Turnbaugh, P. J., Ridaura, V. K., Faith, J. J., Rey, F. E., Knight, R. & Gordon, J. I. (2009). The effect of diet on the human gut microbiome: A metagenomic analysis in humanized gnotobiotic mice. *Sci Transl Med* **1**, 6ra14.

Vaughan, E. E., de Vries, M. C., Zoetendal, E. G., Ben-Amor, K., Akkermans, A. D. & de Vos, W. M. (2002). The intestinal LABs. *Antonie Van Leeuwenhoek* **82**, 341-352.

Vogel, R. F., Lohmann, M., Weller, A. N., Hugas, M. & Hammes, W. P. (1991). Structural similarity and distribution of small cryptic plasmids of *Lactobacillus curvatus* and *L. sake*. *FEMS Microbiol Lett* **84**, 183-190.

Vollenweider, S., Grassi, G., König, I. & Puhan, Z. (2003). Purification and structural characterization of 3-Hydroxypropionaldehyde and its derivatives. *J Agr Food Chem* **51**, 3287-3293.

Wainwright, L. (2006). Does the addition of prebiotics to infant formula have beneficial effects for the baby? *J Neonatal Nurs* **12**, 130-137.

Walter, J., Britton, R. A. & Roos, S. (2010). Host-microbial symbiosis in the vertebrate gastrointestinal tract and the *Lactobacillus reuteri* paradigm. *Proc Natl Acad Sci USA* **108**, 4645-4652.

Walton, G. E., van den Heuvel, E. G., Kosters, M. H., Rastall, R. A., Tuohy, K. M. & Gibson, G. R. (2011). A randomised crossover study investigating the effects of galacto-oligosaccharides on the faecal microbiota in men and women over 50 years of age. *Br J Nutr* **13**, 1-10.

Wang, T. T. & Lee, B. H. (1997). Plasmids in *Lactobacillus*. *Crit Rev Biotechnol* **17**, 227-272.

Wegmann, U., Overweg, K., Horn, N., Goesmann, A., Narbad, A., Gasson, M. J. & Shearman, C. (2009). Complete genome sequence of *Lactobacillus johnsonii* FI9785, a competitive exclusion agent against pathogens in poultry. *J Bacteriol* **191**, 7142-7143.

Willing, B., Halfvarson, J., Dicksved, J., Rosenquist, M., JÃ¤rnerot, G., Engstrand, L., Tysk, C. & Jansson, J. K. (2009). Twin studies reveal specific

imbalances in the mucosa-associated microbiota of patients with ileal Crohn's disease. *Inflamm Bowel Dis* **15**, 653-660.

Wong, J. M. W., de Souza, R., Kendall, C. W. C., Emam, A. & Jenkins, D. J. A. (2006). Colonic health: fermentation and short chain fatty acids. *J Clin Gastroenterol* **40**, 235-243.

Wu, G. D., Chen, J., Hoffmann, C. & other authors (2011). Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334**, 105-108.

Wu, G. D., Bushmanc, F. D. & Lewis, J. D. (2013). Diet, the human gut microbiota, and IBD. *Anaerobe* **24**, 117-120.

Zhang, W., Yu, D., Sun, Z., Chen, X., Bao, Q., Meng, H., Hu, S. & Zhang, H. (2008). Complete nucleotide sequence of plasmid plca36 isolated from *Lactobacillus casei* Zhang. *Plasmid* **60**, 131-135.

Zhang, Z., Liu, C., Zhu, Y., Zhong, Y., Zhu, Y., Zheng, H., Zhao, G.-P., Wang, S.-Y. & Guo, X. (2009). Complete genome sequence of *Lactobacillus plantarum* JDM1. *J Bacteriol* **191**, 5020-5021.

Zúñiga, M., Comas, I., Linaje, R., Monedero, V., Yebra, M. J., Esteban, C. D., Deutscher, J., Pérez-Martínez, G. & González-Candelas, F. (2005). Horizontal gene transfer in the molecular evolution of mannose PTS transporters. *Mol biol evol* **22**, 1673-1685.

Chapter II

Carbohydrate catabolic flexibility in the mammalian intestinal commensal *Lactobacillus ruminis* revealed by fermentation studies aligned to genome annotations.

This chapter was published in full in
O' Donnell, M. M., Forde, B. M., Neville, B. A., Ross, R. P. & O' Toole, P. W. (2011). Carbohydrate catabolic flexibility in the mammalian intestinal commensal *Lactobacillus ruminis* revealed by fermentation studies aligned to genome annotations. *Microb Cell Fact* 10, S12.

Notes:

In vitro growth profiling, manual curation of the genomes, manual curation of the carbohydrate operons using Artemis, KEGG and KAAS and Transmembrane domain prediction was carried out by M.M. O' Donnell (author of this thesis)

Genome sequences were generated for both *L. ruminis* strains and performed other bioinformatics analyses on the data by B. M. Forde.

Additional sequence annotation was carried out by B. A. Neville

Chapter II

Table of contents

Abstract

Background

Lactobacillus ruminis is a poorly characterized member of the *Lactobacillus salivarius* clade that is part of the intestinal microbiota of pigs, humans and other mammals. Its variable abundance in human and animals may be linked to historical changes over time and geographical differences in dietary intake of complex carbohydrates.

Results

In this study, we investigated the ability of nine *L. ruminis* strains of human and bovine origin to utilize fifty carbohydrates including simple sugars, oligosaccharides, and prebiotic polysaccharides. The growth patterns were compared with metabolic pathways predicted by annotation of a high quality draft genome sequence of ATCC 25644 (human isolate) and the complete genome of ATCC 27782 (bovine isolate). All of the strains tested utilized prebiotics including fructooligosaccharides (FOS), soybean-oligosaccharides (SOS) and 1,3:1,4-β-D-gluco-oligosaccharides to varying degrees. Six strains isolated from humans utilized FOS-enriched inulin, as well as FOS. In contrast, three strains isolated from cows grew poorly in FOS-supplemented medium. In general, carbohydrate utilisation patterns were strain-dependent and also varied depending on the degree of polymerisation or complexity of structure. Six putative operons were identified in the genome of the human isolate ATCC 25644 for the transport and utilisation of the prebiotics FOS, galacto-oligosaccharides (GOS), SOS, and 1,3:1,4-β-D-Gluco-oligosaccharides. One of these comprised a novel FOS utilisation operon with predicted capacity to degrade chicory-derived FOS. However, only three of these operons were identified in the ATCC 27782 genome that might account for the utilisation of only SOS and 1,3:1,4-β-D-Gluco-oligosaccharides.

Conclusions

This study has provided definitive genome-based evidence to support the fermentation patterns of nine strains of *Lactobacillus ruminis*, and has linked it to gene distribution patterns in strains from different sources. Furthermore, the study has identified prebiotic carbohydrates with the potential to promote *L. ruminis* growth *in vivo*.

2.1 Background

Immediately following birth, humans are colonised by a variety of bacteria which form the gastrointestinal tract microbiota (Qin *et al.*, 2010). Lactic acid bacteria (LAB), which include *Lactobacillus* spp., are a subdominant element of the microbiota of humans and animals (O'Toole & Claesson, 2010).

Lactobacillus ruminis is a LAB which is part of the autochthonous microbiota in the intestines of both humans (Reuter, 2001), and pigs (Al Jassim, 2003) and it has also been isolated from the bovine rumen (Sharpe *et al.*, 1973). *L. ruminis* is a low G+C Gram positive bacillus (Krieg & Holt, 1984). It is a candidate probiotic organism (see below), since it has been reported to have immunomodulatory characteristics (Taweechotipatr *et al.*, 2009), specifically the ability to induce Nuclear Factor Kappa B (NF-κB) in the absence of lipopolysaccharide production and to activate Tumour Necrosis Factor alpha (TNFα) production in THP-1 monocytes (Taweechotipatr *et al.*, 2009). Unusually, some strains of *L. ruminis* are motile (Sharpe *et al.*, 1973). Limited studies have identified some of the carbohydrates utilised by *L. ruminis* which include cellobiose and raffinose (Krieg & Holt, 1984; Sharpe *et al.*, 1973; Yin & Zheng, 2005). However, little information is available about the fermentation of oligosaccharides/prebiotics by *Lactobacillus ruminis*.

There is growing interest in modulating the human microbiota using dietary supplements including probiotics and prebiotics. Probiotics are defined as "live microorganisms which when administered in adequate amounts confer a health benefit on the host" (FAO/WHO, 2001). However, maintained ingestion of probiotic cultures is generally required to sustain the probiotic effect, with only some of the inoculum surviving gastrointestinal transit, and the vast majority of surviving bacteria shed days after ingestion (Bezkorovainy, 2001). For this reason there has been an increasing research effort expended in the area of prebiotics in order to extend the persistence of particular bacteria (mainly bifidobacteria) in the intestine. Prebiotics are "selectively fermented ingredients that result in specific changes in the composition and/or activity of the gastrointestinal microbiota, thus conferring benefit(s) upon host health" (Gibson *et al.*, 2010). To be considered a prebiotic, the compound has to resist hydrolysis by gastrointestinal tract enzymes and pass into the large intestine, where ideally it promotes the growth of commensal bacteria (Gibson & Roberfroid, 1995). The fermentation of prebiotics in the colon is largely influenced

by the type of sugar monomer, the degree of polymerisation and the nature of the glycosidic bonds between the sugar moieties (Swennen *et al.*, 2006). The constituent sugars of the majority of prebiotics are monosaccharides such as glucose, fructose, galactose and xylose (Manning & Gibson, 2004). The degree of polymerisation (DP) of prebiotics can vary from as low as two for lactulose and in excess of 23 for chicory-derived inulin (Gibson *et al.*, 2004). Humans lack the gastrointestinal enzymes necessary to degrade many of the glycosidic bonds between the sugar units of compounds that are prebiotics, which accounts for their resistance to hydrolysis (Manning & Gibson, 2004). A number of enzymes produced by colonic commensal bacteria may hydrolyse these bonds. These glycosyl hydrolase (GH) enzymes include β-Glucosidases, α-Glucosidases, β-Fructofuranosidases, β-Galactosidases and α-Galactosidases (Henrissat, 1991; Henrissat & Bairoch, 1993; Henrissat & Bairoch, 1996).

Studies of other *Lactobacillus* species have identified a variety of genetic systems that encode the ability to utilize carbohydrates of varying complexity. β-fructofuranosidase is responsible for the hydrolysis of FOS, and this activity was identified in *L. plantarum* WCFS1 (Saulnier *et al.*, 2007), *L. acidophilus* NCFM (Barrangou *et al.*, 2003), and *L. paracasei* 1195 (Goh *et al.*, 2006). β-galactosidases involved in lactose degradation were characterised in *L. sakei* (Stentz *et al.*, 2000), *L. bulgaricus* (Schmidt *et al.*, 1989), *L. coryniformis* (Corral *et al.*, 2006) and *L. reuteri* (Nguyen *et al.*, 2006). β-glucosidase activity (which is responsible for the hydrolysis of 1,4-β-D-Glucans like cellobiose) has been identified in *L. plantarum* (Spano *et al.*, 2005). α-galactosidases, which hydrolyse α-galactosides like raffinose, stachyose and melibiose, were identified in *L. plantarum* ATCC 8014 (Silvestroni *et al.*, 2002) and *L. reuteri* (Tzortzis *et al.*, 2003). Moreover, several α-glucosidases have been characterised in *L. brevis* (De Cort *et al.*, 1994), *L. acidophilus* (Li & Chan, 1983) and *L. pentosus* (Chaillou *et al.*, 1998).

In this study, we describe the fermentation profiles of nine strains of *Lactobacillus ruminis*. The interpretation of the carbohydrate utilisation profiles generated was complemented by the annotation of carbohydrate utilisation genes in the genomes of *L. ruminis* ATCC 25644 and ATCC 27782.

2.2 Methods

2.2.1 Bacterial strains and culture conditions.

Nine *Lactobacillus ruminis* strains were used in this study, and were obtained courtesy of Prof. Gerald Tannock, University of Otago, New Zealand. Four of these are American Type Culture Collection strains: ATCC 25644 (human isolate), ATCC 27780T, ATCC 27781 and ATCC 27782 (bovine isolates). Five human-derived *L. ruminis* strains, L5, S21, S23, S36 and S38 were also studied. All strains were stored at -80°C in de Man-Rogosa-Sharpe (MRS) broth (Difco, BD, Ireland), supplemented with 25% (vol/vol) glycerol as a cryoprotectant. *Lactobacillus* strains were grown anaerobically on MRS agar plates at 37°C for two days. Growth tests were initiated by growing *Lactobacillus* strains anaerobically in MRS-glucose broth at 37°C overnight and unless otherwise stated, all further incubations were also performed under anaerobic conditions at 37°C.

2.2.2 Growth medium.

Modifications were made to the de Man-Rogosa-Sharpe (MRS) (De Man *et al.*, 1960) medium by omitting the carbohydrate source (glucose) and meat extract. Carbohydrate-free MRS (cfMRS) was used as a basal growth medium to study the ability of *Lactobacillus ruminis* strains to utilise various carbohydrates, because it contains no additional carbohydrates and lacks Lab Lembco as a source of carbohydrates. The cfMRS medium contained the following components (gL^{-1}): bacteriological peptone (Oxoid) 10.0, yeast extract (Fluka) 5.0, sodium acetate (Sigma) 5.0, ammonium citrate (Sigma) 2.0, potassium phosphate (Sigma) 2.0, magnesium sulphate (BDH Chemical) 0.2, Manganese sulphate (BDH Chemical) 0.05. The medium also includes Tween 80 (Sigma) 1 ml $litre^{-1}$. The pH was adjusted to between 6.2 and 6.5 and the medium was sterilised at 121°C for 15 minutes. Carbohydrate-free MRS was unable to support bacterial growth above an OD_{600nm} of 0.1 for any of the strains tested.

2.2.3 Carbohydrates and prebiotics.

Fifty-two carbohydrates were used in this study (Table S2.1). Stock solutions of the 50 carbohydrates were filter-sterilized (0.45μm) (Sarstedt) into the cfMRS basal medium to yield a concentration of 0.5% (v/v) for use in the fermentation tests.

2.2.4 Growth measurements.

The fermentation profiles of the various strains were determined using optical density (OD) measurements. The sterile carbohydrate supplemented MRS media was added to the wells of 96 well microtiter plates. The medium in the wells was inoculated with 1% (v/v) of the overnight bacterial culture in MRS-glucose. The OD values of the 96 well microtiter plate wells were read using a Synergy 2 plate reader (BioTek Instruments, Inc., Vermont, US). The inoculated microtiter plates were incubated anaerobically at 37°C and OD readings were taken before and after a 48 hour period (Brewster, 2003). The mean OD readings, standard deviations and standard errors were calculated using technical triplicate data from biological duplicate experiments.

2.2.5 *Lactobacillus ruminis* genome sequencing and assembly.

The genome sequencing, assembly and detailed annotation of the *L. ruminis* ATCC2772 and 25644 genomes will be described elsewhere in this volume (Forde *et al*, manuscript in preparation). In brief, a hybrid next-generation strategy generated 28-fold coverage of the ATCC27782 genome by 454 pyrosequencing, complimented by 217-fold coverage with Illumina paired-end sequences. The assembly of *L. ruminis* ATCC 27782 is a finished genome; the genome assembly of *L. ruminis* ATCC 25644 a high-quality draft (Chain *et al.*, 2009).

2.2.6 Bioinformatic analysis and gene annotation.

The Artemis program (Rutherford *et al.*, 2000) was used to visualise and identify carbohydrate metabolism genes in the genome of *Lactobacillus ruminis* ATCC 25644 and ATCC 27782 (Mural, 2000). Open reading frames were predicted using Glimmer 3 (Delcher *et al.*, 2007). Each carbohydrate utilisation enzyme, predicted from opening reading frames (ORF), was assigned a KEGG orthology (KO) identifier by KAAS and graphical representations for each metabolic pathway were generated (Moriya *et al.*, 2007). The TMHMM 2.0 server was used to predict the

transmembrane helices of proteins, which were identified from annotation as putative carbohydrate transporters. THHMM 2.0 uses Hidden Markov models to predict the proteins topology with a high degree of accuracy (Krogh *et al.*, 2001). TransTermHP (Kingsford *et al.*, 2007) was used to predict rho-independent transcriptional terminators. Comparisons to other Lactobacillus genomes were made using data available from both NCBI and KEGG Organisms .

2.2.7 Sequence data availability and accession numbers.

The finished genome of ATCC 27782 is available under accession number XXYYZZ123. The draft genome of ATCC 25644 is available under accession number CCGGHHIIUU.

2.3 Results

2.3.1 Growth of *L. ruminis* in media containing diverse carbon sources

A carbohydrate utilisation profile for each of nine strains of *L. ruminis* on fifty carbohydrates was established as described in Methods. Table S2.2 summarizes the data, with individual strain data in Figure S2.1-2.9. In summary, there was significant variation with respect to carbohydrate fermentation profiles at the strain level. Moderate growth was observed for strains L5 and S21 when grown on α-galactosides (melibiose, raffinose, stachyose) and β-glucosides (β-glucotriose B, cellobiose) (Table S2.2). The majority of bovine isolates could poorly utilize fructooligosaccharides, except for ATCC 27781 with Beneo P95 and Raftilose P95. Moderate growth was observed for the majority of isolates with galactooligosaccharides (GOS, GOS-inulin, lactose, lactulose). All strains were able to ferment β-Glucotriose B, cellobiose, galactose, glucose, maltose, mannose, melibiose, raffinose, stachyose and sucrose (Table S2). Some strains showed a distinctly higher ability to utilize specific carbo+hydrates e.g. fructose by strains L5 and S21, (Figure S2.1 and 2.2); lactose by strains S23, ATCC 25644 and ATCC 27780T (Figure S2.3, 2.6 and 2.7); raffinose by ATCC 27781 (Figure S2.8); and Raftilose P95 by strain S36 (Figure S2.4).

2.3.2 Growth and fermentation analysis of human and bovine-derived *L. ruminis* type strains.

Table 2.1 shows the final cell numbers and culture-medium pH values reached for the two strains ATCC 25644 (human isolate) and ATCC 27782 (bovine isolate), in the presence of various carbohydrates and prebiotics for 24 h. *L. ruminis* ATCC 25644 reached the highest cell density (8.9 x 10^8 cfu/ml) when grown on Raftilose Synergy 1 which coincided with the lowest culture medium pH value of 4.86. ATCC 27782 reached the highest cell density values (2.7x10^8 cfu/ml) when grown on Beta Glucotriose B, and fermentation resulted in a culture medium pH value of 5.19 following 24 hours incubation. This was far higher than cellobiose, the other beta-glucoside tested, although the final pH of both cultures was very similar, and the medium was buffered in the same way as MRS.

Table 2.1 - Growth and fermentation analysis of *L. ruminis* strains ATCC 25644 (human isolate) and ATCC 27782 (bovine isolate).

Carbohydrate type	Carbohydrate	ATCC 25644		ATCC 27782	
		Cfu/ml	pH*	Cfu/ml	pH*
Disaccharide	Cellobiose	2.40 x 10^8	5.21	7.00 x 10^6	5.13
	Lactulose	3.20 x 10^8	4.99	0	6.53
	Lactose	2.76 x 10^8	4.76	0	6.57
Monosaccharide	Glucose	4.39 x 10^8	4.86	1.53 x 10^8	4.85
Oligosaccharide	Beta Glucotriose B	4.05 x 10^8	5.17	2.66 x 10^8	5.19
	Raftilose Synergy 1	8.90 x 10^8	5.01	1.35 x 10^7	6.04
	Raftilose P95	2.91 x 10^8	5.28	2.51 x 10^6	5.42
Tetrasaccharide	Stachyose	3.94 x 10^8	5.13	2.37 x 10^8	5.11
Trisaccharide	Raffinose	3.24 x 10^8	5.2	1.40 x 10^8	5.2

***** pH value of culture medium after 24 h growth in indicated carbon source. Values tabulated are the average of two replicates carried out on separate days.

2.3.3 Annotation of carbohydrate pathways in the *L. ruminis* genome.

A high-quality draft genome sequence was generated for *L. ruminis* ATCC 25644 and a finished genome sequence was generated for ATCC 27782, as described in Methods. The complete functional and comparative analysis of these genomes will be

described elsewhere (Forde *et al.*, 2011; Neville *et al*., 2012). A draft sequence of ATCC 25644 has also been generated by the Human Microbiome Project; however it has a different scaffold structure and assembly statistics to that which we generated for ATCC 25644, and for that reason was not used in the current study. The carbohydrate utilisation genes of ATCC 25644 and ATCC 27782 were annotated by manual curation in conjunction with KEGG Automatic Annotation Server (KAAS). *L. ruminis*-specific Kyoto Encyclopaedia of Genes and Genomes (KEGG) maps were generated based upon our annotated genome sequences that we analyzed with KAAS. As a representative example, the galactose metabolic pathway (for both sequenced *L. ruminis* genomes) is presented in Figure 2.1. It demonstrates the predicted reliance on glycosyl hydrolases to ferment carbohydrates in *L. ruminis* as well as highlighting the fermentable α and β-galactosides.

Sixteen major pathways or systems involved in carbohydrate utilization were annotated in both genomes, and are shown in Figures S2.10-2.25. These include those for glycolysis, pentose and glucuronate interconversions, fructose and mannose utilization, starch and sucrose. Of the sixteen pathways identified, eight are considered partial pathways (Figures S2.10-2.25).

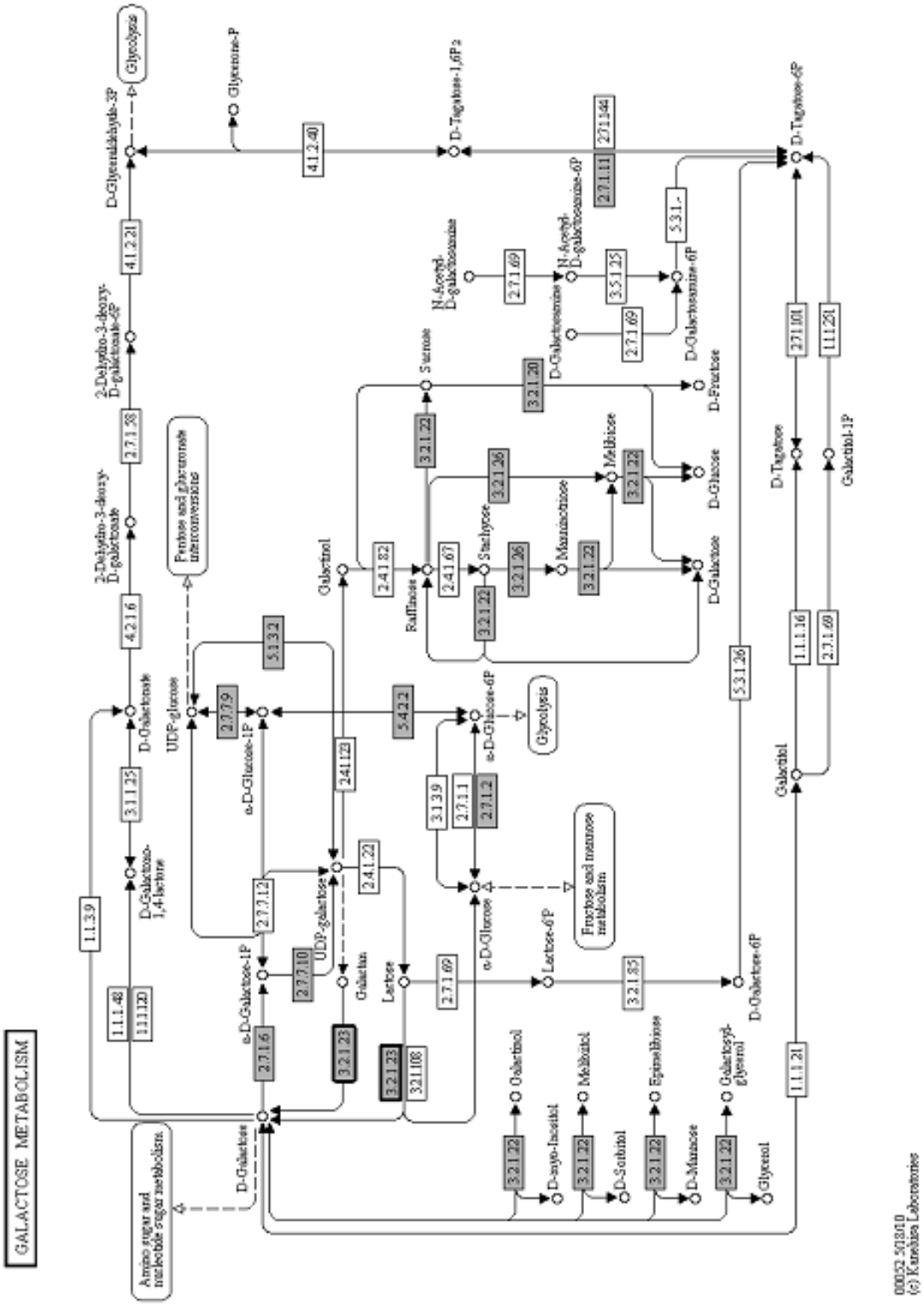


Figure 2.1 - Galactoside utilisation metabolic map for *L. ruminis* ATCC 25644 and ATCC 27782. Grey boxes, enzymes present in both ATCC 25644 and ATCC 27782; Grey boxes with emphasised black border, enzymes present in ATCC 25644 and absent from ATCC 27782.

2.3.4 Identification of Glycosyl Hydrolases.

Glycosyl hydrolases are key to prebiotic utilization, and can also be manipulated to synthesize prebiotics. Twenty glycosyl hydrolases were annotated in the genome of ATCC 25644, and fourteen were annotated in the genome of ATCC 27782. The glycosyl hydrolases include α-amylase (EC 3.2.1.1), endo-1,4-β-xylanase (EC 3.2.1.8), oligo-1,6-glucosidase (EC 3.2.1.10), lysozyme (EC 3.2.1.17), α-glucosidase (EC 3.2.1.20), β-glucosidase (EC 3.2.1.21), α-galactosidase (EC 3.2.1.22), β-galactosidase (EC 3.2.1.23), β-fructofuranosidase (EC 3.2.1.26), β-N-acetylhexosaminidase (EC 3.2.1.52), glucan 1,6-α-glucosidase (EC 3.2.1.70), 6-phospho-β--glucosidase (EC 3.2.1.86) and neopullalanase (EC 3.2.1.135). The majority of these enzymes are present in ATCC 27782 with the exceptions of α-amylase, oligo-1,6-glucosidase and β-galactosidase.

2.3.5 Identification of putative genes and operons involved in prebiotic utilisation.

The sequenced *L. ruminis* genomes were extensively scrutinized to identify putative operons involved in carbohydrate transport and utilisation. Specificity of substrate was based upon manual curation of the annotated region, including reference to BLAST identity to functionally characterized homologues, genetic neighbourhood analysis, and protein motif matching. Six putative prebiotic utilisation operons were annotated in the *L. ruminis* ATCC 25644 genome (human isolate; Figure 2.2), only three of which were identified in the bovine isolate ATCC 27782 (Figure S2.26). Most of the operons are flanked by predicted rho-independent transcriptional terminators (Figure 2.2), and these operons constitute one to two transcriptional units, with a gene for a LacI-type transcriptional regulator in four of six cases.

We annotated a predicted FOS utilization operon only in the human isolate *L. ruminis* ATCC 25644. β-fructofuranosidase, a Glycosyl hydrolase (GH) family 32 enzyme (Henrissat, 1991), has been identified as the key enzyme in operons involved in FOS utilisation in other *Lactobacillus* species (Barrangou *et al.*, 2003; Goh *et al.*, 2006; Saulnier *et al.*, 2007). This activity is predicted to be encoded by the *L. ruminis bfrA* gene, which is linked to a presumptive oligosaccharide symporter gene. The ATCC 25644 genome was also distinguished by having two additional operons for lactose/galactose utilization (Figure 2.2). The genomes of both strains harboured

operons predicted to confer utilization of sucrose, cellobiose and raffinose. As well as the β-fructofuranosidase (*sacA*) in the sucrose operon, genes for an amylopullalanase (*amyB*) and an α-glucosidase (*malZ*) are also contiguous and are potentially co-transcribed with the sucrose operon, but do not have a predicted function in the hydrolysis of sucrose or FOS (Figure 2.2 B).

The cellobiose operon is predicted to be responsible for the transport and hydrolysis of both cellobiose and 1,3:1,4-β-D-Glucan hydrolysates, and in *L. ruminis* it appears to involve two β glucosidases (Figure 2.2) that belong to the GH1 family of glycosyl hydrolases (Henrissat, 1991). The amino acid sequence of BglB and $BglB_2$ showed 70% and 77% identity to the β-glucosidases identified in the genomes of *L. helveticus* DPC 4571 and *L. ultunensis* DSM 16047, respectively. The products of the raffinose operon (Fig. 2.2D; also present in ATCC 27782) are predicted to have the additional ability to breakdown melibiose and stachyose. All of the glycolytic enzymes discussed above lack predicted transmembrane domains (TMD) and therefore most likely require import of their respective substrates.

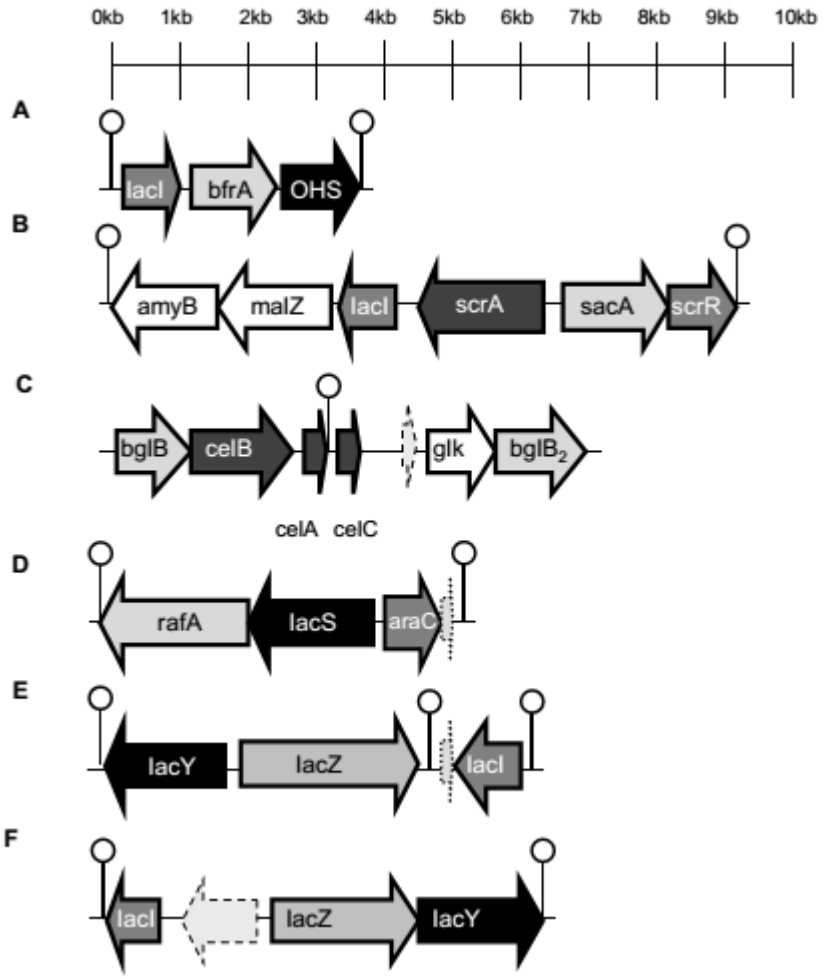


Figure 2.2 - Putative operons for the predicted utilisation of carbohydrates in *L. ruminis* ATCC 25644. Predicted substrates are A, FOS; B, Sucrose; C, Cellobiose; D, raffinose; E, lactose/galactose; F, lactose/galactose operon. Light grey arrows with thick black border, glycosyl hydrolase family enzyme; Black arrows, major facilitator superfamily transporters; Medium grey arrows, transcriptional regulators; Dark grey arrows with thick grey border, phosphotransferase system transporters; Lollipops, rho-dependent transcriptional terminators; White arrows with dashed surround, transposases; white arrows with dotted surround, hypothetical proteins; White arrows with black continuous surround, potentially co-transcribed enzymes. Operons B, C and D were also annotated in the ATCC27782 genome (Figure S2.26).

2.3.6 Predicted carbohydrate transporters.

A relationship exists between the genomic association of genes and the functional interaction of the proteins they encode (Snel *et al.*, 2002). To refine our annotation of the carbohydrate utilisation operons, we therefore performed a detailed analysis of the predicted transporter proteins encoded by the contiguous genes. As for hydrolases, specificity of substrate was predicted based upon an integrated analysis of the annotated region, including reference to BLAST identity to functionally characterized homologues, linked genes, and protein motif matching. Putative carbohydrate transporters were analysed with transmembrane prediction software, with 14 and 10 transporters identified in the genome sequences of *L. ruminis* ATCC 25644 and 27782, respectively (Table 2.2). The predicted carbohydrate transporters belong to the ATP-binding Cassette family (ABC), the Glycoside-Pentoside-Hexuronide cation symporter family (GPH), the Oligosaccharide H^+ Symporter (OHS) and the Phosphotransferase System (PTS). Transmembrane domain (TMD) numbers are generally indicative of the type of carbohydrate transporter, with some exceptions (Saier, 2000). ABC transporters have on average 10-12 TMD but this can be highly variable. PTS transporters have been identified with up to 10 TMD (this study). GPH and OHS transporters (both being Major Facilitator Superfamily transporters) generally have 12 TMD (Saier, 2000). In ATCC 25644, three GPH transporters were identified (Table 2.2) and these are predicted to transport the β-galactosides (lactose, galactose, lactulose and GOS) and the α-galactosides (raffinose, melibiose and stachyose). However, in ATCC 27782 only one GPH transporter was identified, which was predicted to transport α-galactosides. The OHS identified in the genome of

ATCC 25644 is adjacent to a β-fructofuranosidase and may be involved in transporting FOS. Both genomes encode six predicted PTS transporters, which potentially transport mannose, sucrose, fructose, cellobiose and glucose. In both *L. ruminis* genome sequences, four ABC transporters were identified, with the putative substrates identified as mannose and glycerol-3-phosphate. All of the transporters identified in each genome had associated metabolic genes located either upstream or downstream in the genome, and the majority were arranged in operons. Both genomes also encoded proteins for glucose uptake (with TMD counts of 5 and 9 in ATCC 25644 and ATCC 27782, respectively), and a simple sugar transport system permease protein which was predicted to transport monosaccharides like galactose.

Table 2.2 - Transmembrane domains (TMD) of the predicted carbohydrate transport proteins in *Lactobacillus ruminis*

Family	Gene	Locus number[a]		Predicted substrate	TMD[b]	
		ATCC 25644	ATCC 27782		ATCC 25644	ATCC 27782
OHS	*lacY*	ANHS_218	-	FOS	12	-
GPH	*lacY*	ANHS_744c ANHS_924	-	Lactose, galactose, galactan	12	-
	lacS	ANHS_783	LRU_18250	Raffinose, stachyose, melibiose	12	12
ABC	*ugpE*	ANHS_648	LRU_16940	Glycerol	6	6
	ugpA	ANHS_649c	LRU_16950	Glycerol	6	6
	malG	ANHS_839c	LRU_18720	Maltose	6	6
	malF	ANHS_840c	LRU_18730	Maltose	8	8
PTS	*manY*	ANHS_242	LRU_18860	Mannose	7	7
	manZ	ANHS_243	LRU_18850	Mannose	5	4
	scrA	ANHS_846c	LRU_18780	Sucrose, FOS	8	8
	fruA	ANHS_1075	LRU_00800	Fructose	9	9
	celB	ANHS_1218	LRU_02240	Cellobiose	10	10
	gluA	ANHS_851c	LRU_18820	Glucose	9	9

a. Locus number in draft genome sequences

TMD: predicted trans-membrane domains, as described in Materials & Methods

2.4 Discussion

We consider *L. ruminis* as a candidate probiotic, which we are also investigating as a potential responder for prebiotic/symbiotic supplementation in humans and animals. Several studies have identified *L. ruminis* in the gastrointestinal tract of humans (Antonio *et al.*, 1999; Delgado *et al.*, 2004; Kimura *et al.*, 2010). *L. ruminis* was isolated from the bovine rumen (Sharpe *et al.*, 1973), from the pig (Al Jassim, 2003; Yin & Zheng, 2005), chickens (Kovalenko *et al.*, 1989), sheep (Mueller *et al.*, 1984), Svalbard reindeer (Mathiesen *et al.*, 1987), horses (Hidetoshi *et al.*, 2009; Vörös, 2008; Willing *et al.*, 2009c), cats (Desai *et al.*, 2009; Ritchie *et al.*, 2009), dogs (Greetham *et al.*, 2002) and parrots (Xenoulis *et al.*). *L. ruminis* thus appears to be variably present in the microbiota of humans and many domesticated animals.

L. ruminis was previously described as a homofermentative bacterium, with the ability to ferment amygdalin, cellobiose, galactose, maltose, mannose, melibiose, raffinose, salicin, sorbitol and sucrose (Kandler & Weiss, 1986). In the current study, the nine strains of *L. ruminis* were unable to utilise sorbitol as a carbon source. *L. ruminis* has also been reported to have the ability to ferment D-ribose (Tanasupawat *et al.*, 2000). However, we observed no growth for any of the nine *L. ruminis* strains when cultured in cfMRS supplemented with ribose. ATCC 27782 lacks a transaldolase gene (and the draft genome sequence suggests ATCC 25644 also lacks this gene), which would account for inability to utilise any of the pentose sugars tested. All of the *L. ruminis* strains tested (with the exception of ATCC 27782 which lacks a *lacZ* gene) had strong growth in lactose. This contrasts with a previous study, where moderate growth was recorded on lactose (Kandler & Weiss, 1986). It has also been reported that *L. ruminis* showed a strain dependent fermentation of starch (Kato *et al.*, 2000), and very little growth was recorded for any of the strains tested here.

As a species, *L. ruminis* is generally able to ferment prebiotic compounds including FOS, GOS, lactulose, 1,3:1,4 β-D-Glucooligosaccharides, raffinose and stachyose. Only one strain, S36 was capable of (weakly) fermenting the prebiotic disaccharide palatinose. Palatinose is made by enzymatic rearrangement of the glycosidic linkages present in sucrose from an α-1,2-fructoside to an α-1,6-fructoside (Lina *et al.*, 2002). This suggests that the catalytic enzymes involved in sucrose utilisation may no longer be able to degrade the α-1,6-fructoside linkage in this disaccharide. The majority of

L. ruminis strains achieved higher cell densities when grown on the prebiotic carbohydrates raffinose, lactulose, FOS, GOS and stachyose than when grown in other mono- and disaccharide carbohydrates tested. This growth pattern may be attributed to a niche for *L. ruminis* in the lower gastrointestinal tract (GIT). Mono and disaccharides are often unable to resist the hydrolytic action of the upper GIT, unlike prebiotics, and would not therefore be as freely available as carbon sources for *L. ruminis* in the large intestine. Lactulose, a disaccharide derivative of lactose, has previously been shown to support high level growth of other lactobacilli namely *L. rhamnosus*, *L. paracasei* and *L. salivarius* (Saarela *et al.*, 2003). Lactulose also supported a high level of growth for the majority of *L. ruminis* strains. The β-galactosides lactulose and GOS are predicted to be transported and hydrolysed in ATCC 25644 by LacY and LacZ as part of the lactose operon. Two operons for β-galactoside utilisation were identified in the genome of ATCC 25644; however neither of these operons or any potential genetic determinants could be identified for lactose utilisation in ATCC 27782. The absence of a lactose operon in the genome may suggest an ecological niche adaptation by ATCC 27782 to an environment devoid of milk sugars.

β-glucooligosaccharides such as cellobiose are generally transported and hydrolysed using the cellobiose PTS and β-glucosidase enzymes. Both cellobiose and β-glucotriose B are 1,4-β-D-glucooligosaccharides with a similar structure which allows the transport and utilisation of these carbohydrates by the products of the cellobiose operon. The bovine *L. ruminis* isolates, ATCC 27780T, 27781 and 27782 were previously reported to utilise β-glucan hydrolysates as a carbohydrate source (Snart *et al.*, 2006), and in that study, all bovine isolates utilised β-glucan hydrolysates of DP3, and only ATCC 27780T was unable to utilise DP4 oligosaccharide. ATCC 27781 was distinguished by being able to utilise the highest percentage of both DP3 and DP4 β glucan. We have shown that all the strains tested in this study were able to utilise the DP3 β-glucan hydrolysates to a moderate degree. The bovine isolate ATCC 27780T achieved the highest growth (data not shown) when utilizing β glucan hydrolysate, in contrast to a previous study which identified ATCC 27781 as having the highest percentage utilisation of β-glucan oligosaccharide (Snart *et al.*, 2006).

In previous analysis of sixteen *Lactobacillus* species, only *L. acidophilus* L3, *L. acidophilus* 74-2 and *L. casei* CRL431 were able to utilise Raftilose P95, an

oligofructose (Kneifel *et al.*, 2000). In the current study, eight strains of *L. ruminis* were capable of utilizing Raftilose P95. In addition, *L. ruminis* was capable of moderate to strong fermentation of Raftilose Synergy 1, an oligofructose-enriched inulin. *L. paracasei* subsp. *paracasei* 8700:2 was previously shown to be the only strain, out of ten strains tested, that was capable of strong growth on Raftilose Synergy 1, while three other species were capable of moderate growth (Makras *et al.*, 2005). Based on these comparisons, *L. ruminis* may have a growth advantage over other lactobacilli in the presence of fructooligosaccharides.

A novel β-fructofuranosidase was identified in the genome of *L. ruminis* ATCC 25644 that potentially hydrolyses the linkages present in chicory derived fructooligosaccharides. The cognate transporter OHS was identified only in the strains isolated from humans. FOS may be transported using the sucrose PTS transporter in the bovine strains ATCC 27780 and 27781. The human isolates of *L. ruminis* apparently use an OHS to transport FOS into the cell. Both sequenced strains likely use the ABC transport system to transport simple carbohydrates like maltose and glycerol. The most populated class of transporter identified was the phosphotransferase system transporter, with six such systems present. However, in *L. ruminis* many of the fermentable carbohydrates including α-galactosides and β-galactosides are predicted to be transported by GPH symporters. GPH transporters contain a C-terminal hydrophilic domain which interacts with the PTS system (Saier, 2000), which may thus be an important regulatory mechanism in *L. ruminis*.

2.5 Conclusions

Lactobacillus ruminis is a saccharolytic member of the intestinal microbiota capable of degrading a variety of prebiotics. Genes and operons were identified in the genomes of two sequenced strains for the hydrolysis and transport of the utilisable prebiotics. This work is the first step in the characterisation of carbohydrate metabolism, transportation and regulation in *L. ruminis*. Further studies will focus on the functional characterisation of the putative operons identified in this study and also *in vivo* studies with dietary supplementation by selected carbohydrates. Characterisation of the novel FOS degrading enzyme *bfrA* may facilitate applications including reverse engineering of the FOS degradation pathway to allow the biosynthesis of a potentially novel fructooligosaccharide.

2.6 References

KEGG Organisms: Complete Genomes. In *http://keggjp/kegg/catalog/org_listhtml*.

Human Microbiome Project. In *http://nihroadmapnihgov/hmp/*.

NCBI Genome Homepage. In *http://wwwncbinlmnihgov/sites/genome*.

Al Jassim, R. A. (2003). *Lactobacillus ruminis* is a predominant lactic acid producing bacterium in the caecum and rectum of the pig. *Lett Appl Microbiol* **37**, 213-217.

Antonio, M. A., Hawes, S. E. & Hillier, S. L. (1999). The identification of vaginal *Lactobacillus* species and the demographic and microbiologic characteristics of women colonized by these species. *J Infect Dis* **180**, 1950-1956.

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

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

Brewster, J. D. (2003). A simple micro-growth assay for enumerating bacteria. *Journal of Microbiological Methods* **53**, 77-86.

Chaillou, S., Lokman, B. C., Leer, R. J., Posthuma, C., Postma, P. W. & Pouwels, P. H. (1998). Cloning, sequence analysis, and characterization of the genes involved in isoprimeverose metabolism in *Lactobacillus pentosus*. *J Bacteriol* **180**, 2312-2320.

Chain, P. S. G., Grafham, D. V., Fulton, R. S. & other authors (2009). Genome Project Standards in a New Era of Sequencing. *Science* **326**, 236-237.

Corral, J., Bañuelos, O., Adrio, J. & Velasco, J. (2006). Cloning and characterization of a β-galactosidase encoding region in *Lactobacillus coryniformis* CECT 5711. *Applied microbiology and biotechnology* **73**, 640-646.

De Cort, S., Kumara, H. M. C. S. & Verachtert, H. (1994). Localization and characterization of alpha-glucosidase activity in *Lactobacillus brevis*. *Appl Environ Microbiol* **60**, 3074-3078.

De Man, J. C., M. Rogosa & Sharpe, M. E. (1960). A medium for the cultivation of lactobacilli. *Journal of Applied Microbiology* **23**, 130-135.

Delcher, A. L., Bratke, K. A., Powers, E. C. & Salzberg, S. L. (2007). Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* **23**, 673-679.

Delgado, S., Suarez, A., Otero, L. & Mayo, B. (2004). Variation of microbiological and biochemical parameters in the faeces of two healthy people over a 15 day period. *Eur J Nutr* **43**, 375-380.

Desai, A. R., Musil, K. M., Carr, A. P. & Hill, J. E. (2009). Characterization and quantification of feline fecal microbiota using cpn60 sequence-based methods and investigation of animal-to-animal variation in microbial population structure. *Veterinary Microbiology* **137**, 120-128.

FAO/WHO (2001). Report on Joint FAO/WHO expert consultation on evaluation of health and nutritional properties of probiotics in food Including powder milk with live lactic acid bacteria. In *ftp://ftpfaoorg/es/esn/food/probio_report_enpdf*.

Gibson, G. R. & Roberfroid, M. B. (1995). Dietary modulation of the human colonic microbiota: Introducing the concept of prebiotics. *J Nutr* **125**, 1401-1412.

Gibson, G. R., Probert, H. M., Loo, J. V., Rastall, R. A. & Roberfroid, M. B. (2004). Dietary modulation of the human colonic microbiota: updating the concept of prebiotics. *Nutr Res Rev* **17**, 259-275.

Gibson, G. R., Scott, K. P., Rastall, R. A. & other authors (2010). Dietary prebiotics: current status and new definition. *Food Science & Technology Bulletin: Functional Foods* **7**, 1-19.

Goh, Y. J., Zhang, C., Benson, A. K., Schlegel, V., Lee, J.-H. & Hutkins, R. W. (2006). Identification of a putative operon involved in fructooligosaccharide utilization by *Lactobacillus paracasei*. *Appl Environ Microbiol* **72**, 7518-7530.

Greetham, H. L., Giffard, C., Hutson, R. A., Collins, M. D. & Gibson, G. R. (2002). Bacteriology of the Labrador dog gut: a cultural and genotypic approach. *Journal of Applied Microbiology* **93**, 640-646.

Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J* **280**, 309-316.

Henrissat, B. & Bairoch, A. (1993). New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J* **293 (Pt 3)**, 781-788.

Henrissat, B. & Bairoch, A. (1996). Updating the sequence-based classification of glycosyl hydrolases. *Biochem J* **316 (Pt 2)**, 695-696.

Hidetoshi, M., Akiyo, N., Mitsuharu, S. & other authors (2009). *Lactobacillus hayakitensis*, *L. equigenerosi* and *L. equi*, predominant lactobacilli in the intestinal flora of healthy thoroughbreds. *Anim Sci J* **80**, 339-346.

Kandler, O. & Weiss, N. (1986). Genus Lactobacillus. In *Bergey's manual of systematic bacteriology*, pp. 1209–1234.

Kato, Y., Sakala, R. M., Hayashidani, H., Kiuchi, A., Kaneuchi, C. & Ogawa, M. (2000). *Lactobacillus algidus* sp. nov., a psychrophilic lactic acid bacterium isolated from vacuum-packaged refrigerated beef. *Int J Syst Evol Microbiol* **50**, 1143-1149.

Kimura, K., Nishio, T., Mizoguchi, C. & Koizumi, A. (2010). Analysis of the composition of lactobacilli in humans. *Bioscience and Microflora* **29**, 47-50.

Kingsford, C., Ayanbule, K. & Salzberg, S. (2007). Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biology* **8**, R22.

Kneifel, W., Rajal, A. & Kulbe, K. D. (2000). *In vitro* growth behaviour of probiotic bacteria in culture media with carbohydrates of prebiotic importance. *Microbial Ecology in Health & Disease* **12**, 27-34.

Kovalenko, N. K., Golovach, T. N. & Kvasnikov, E. I. (1989). Lactic bacteria in the digestive tract of poultry. *Mikrobiologiia* **58**, 137-143.

Krieg, N. R. & Holt, J. G. (1984). *Bergey's manual of systematic bacteriology*.

Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *Journal of Molecular Biology* **305**, 567-580.

Li, K. B. & Chan, K. Y. (1983). Production and properties of alpha-glucosidase from *Lactobacillus acidophilus*. *Appl Environ Microbiol* **46**, 1380-1387.

Lina, B. A. R., Jonker, D. & Kozianowski, G. (2002). Isomaltulose (Palatinose®): a review of biological and toxicological studies. *Food and Chemical Toxicology* **40**, 1375-1381.

Makras, L., Van Acker, G. & De Vuyst, L. (2005). *Lactobacillus paracase*i subsp. *paracasei* 8700:2 degrades inulin-type fructans exhibiting different degrees of polymerization. *Appl Environ Microbiol* **71**, 6531-6537.

Manning, T. S. & Gibson, G. R. (2004). Microbial-gut interactions in health and disease. Prebiotics. *Best Pract Res Clin Gastroenterol* **18**, 287-298.

Mathiesen, S. D., Orpin, C. G., Greenwood, Y. & Blix, A. S. (1987). Seasonal changes in the cecal microflora of the high-arctic Svalbard reindeer (*Rangifer tarandus platyrhynchus*). *Appl Environ Microbiol* **53**, 114-118.

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C. & Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* **35**, W182-185.

Mueller, R. E., Iannotti, E. L. & Asplund, J. M. (1984). Isolation and identification of adherent epimural bacteria during succession in young lambs. *Appl Environ Microbiol* **47**, 724-730.

Mural, R. J. (2000). ARTEMIS: a tool for displaying and annotating DNA sequence. *Brief Bioinform* **1**, 199-200.

Nguyen, T.-h., Splechtna, B., Steinböck, M., Kneifel, W., Lettner, H. P., Kulbe, K. D. & Haltrich, D. (2006). Purification and characterization of two novel beta-galactosidases from *Lactobacillus reuteri*. *Journal of Agricultural and Food Chemistry* **54**, 4989-4998.

O'Toole, P. W. & Claesson, M. J. (2010). Gut microbiota: Changes throughout the lifespan from infancy to elderly. *Int Dairy J* **20**, 281-291.

Qin, J., Li, R., Raes, J. & other authors (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59-65.

Reuter, G. (2001). The *Lactobacillus* and *Bifidobacterium* microflora of the human intestine: composition and succession. *Curr Issues Intest Microbiol* **2**, 43-53.

Ritchie, L. E., Burke, K. F., Garcia-Mazcorro, J. F., Steiner, J. M. & Suchodolski, J. S. (2009). Characterization of fecal microbiota in cats using universal 16S rRNA gene and group-specific primers for *Lactobacillus* and *Bifidobacterium* spp. *Veterinary Microbiology* **In Press, Corrected Proof**.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.-A. l. & Barrell, B. (2000). Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944-945.

Saarela, M., Hallamaa, K., Mattila-Sandholm, T. & Mättö, J. (2003). The effect of lactose derivatives lactulose, lactitol and lactobionic acid on the functional and technological properties of potentially probiotic *Lactobacillus* strains. *International Dairy Journal* **13**, 291-302.

Saier, M. H., Jr. (2000). A functional-phylogenetic classification system for transmembrane solute transporters. *Microbiol Mol Biol Rev* **64**, 354-411.

Saulnier, D. M., Molenaar, D., de Vos, W. M., Gibson, G. R. & Kolida, S. (2007). Identification of prebiotic fructooligosaccharide metabolism in *Lactobacillus plantarum* WCFS1 through microarrays. *Appl Environ Microbiol* **73**, 1753-1765.

Schmidt, B. F., Adams, R. M., Requadt, C., Power, S. & Mainzer, S. E. (1989). Expression and nucleotide sequence of the *Lactobacillus bulgaricus* beta-galactosidase gene cloned in Escherichia coli. *J Bacteriol* **171**, 625-635.

Sharpe, M. E., Latham, M. J., Garvie, E. I., Zirngibl, J. & Kandler, O. (1973). Two new species of *Lactobacillus* isolated from the bovine rumen, *Lactobacillus ruminis* sp.nov. and *Lactobacillus vitulinus* sp.nov. *J Gen Microbiol* **77**, 37-49.

Silvestroni, A., Connes, C., Sesma, F., de Giori, G. S. & Piard, J.-C. (2002). Characterization of the melA locus for α-galactosidase in *Lactobacillus plantarum*. *Appl Environ Microbiol* **68**, 5464-5471.

Snart, J., Bibiloni, R., Grayson, T. & other authors (2006). Supplementation of the diet with high-viscosity beta-glucan results in enrichment for lactobacilli in the rat cecum. *Appl Environ Microbiol* **72**, 1925-1931.

Snel, B., Bork, P. & Huynen, M. A. (2002). The identification of functional modules from the genomic association of genes. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 5890-5895.

Spano, G., Rinaldi, A., Ugliano, M., Moio, L., Beneduce, L. & Massa, S. (2005). A beta-glucosidase gene isolated from wine *Lactobacillus plantarum* is regulated by abiotic stresses. *Journal of Applied Microbiology* **98**, 855-861.

Stentz, R., Loizel, C., Malleret, C. & Zagorec, M. (2000). Development of genetic tools for *Lactobacillus sakei*: Disruption of the beta-galactosidase gene and use of lacZ as a reporter gene to study regulation of the putative copper ATPase, AtkB. *Appl Environ Microbiol* **66**, 4272-4278.

Swennen, K., Courtin, C. & Delcour, J. (2006). Non-digestible oligosaccharides with prebiotic properties. *Critical Reviews in Food Science & Nutrition* **46**, 459-471.

Tanasupawat, S., Shida, O., Okada, S. & Komagata, K. (2000). *Lactobacillus acidipiscis* sp. nov. and *Weissella thailandensis* sp. nov., isolated from fermented fish in Thailand. *Int J Syst Evol Microbiol* **50**, 1479-1485.

Taweechotipatr, M., Iyer, C., Spinler, J. K., Versalovic, J. & Tumwasorn, S. (2009). *Lactobacillus saerimneri* and *Lactobacillus ruminis*: novel human-derived probiotic strains with immunomodulatory activities. *FEMS Microbiology Letters* **293**, 65-72.

Tzortzis, G., Jay, A. J., Baillon, M. L. A., Gibson, G. R. & Rastall, R. A. (2003). Synthesis of α-galactooligosaccharides with α-galactosidase from *Lactobacillus reuteri* of canine origin. *Applied microbiology and biotechnology* **63**, 286-292.

Vörös, A. (2008). Diet related changes in the gastrointestinal microbiota of horses. In *Dept of Animal Nutrition and Management*, pp. 36. Uppsala: Swedish University of Agricultural Sciences.

Willing, B., Vörös, A., Roos, S., Jones, C., Jansson, A. & Lindberg, J. E. (2009). Changes in faecal bacteria associated with concentrate and forage-only diets fed to horses in training. *Equine Vet J* **41**, 908-914.

Xenoulis, P. G., Gray, P. L., Brightsmith, D., Palculict, B., Hoppes, S., Steiner, J. M., Tizard, I. & Suchodolski, J. S. Molecular characterization of the cloacal microbiota of wild and captive parrots. *Veterinary Microbiology* **In Press, Accepted Manuscript**.

Yin, Q. & Zheng, Q. (2005). Isolation and identification of the dominant *Lactobacillus* in gut and faeces of pigs using carbohydrate fermentation and 16S rDNA analysis. *J Biosci Bioeng* **99**, 68-71.

2.7 Supplementary information

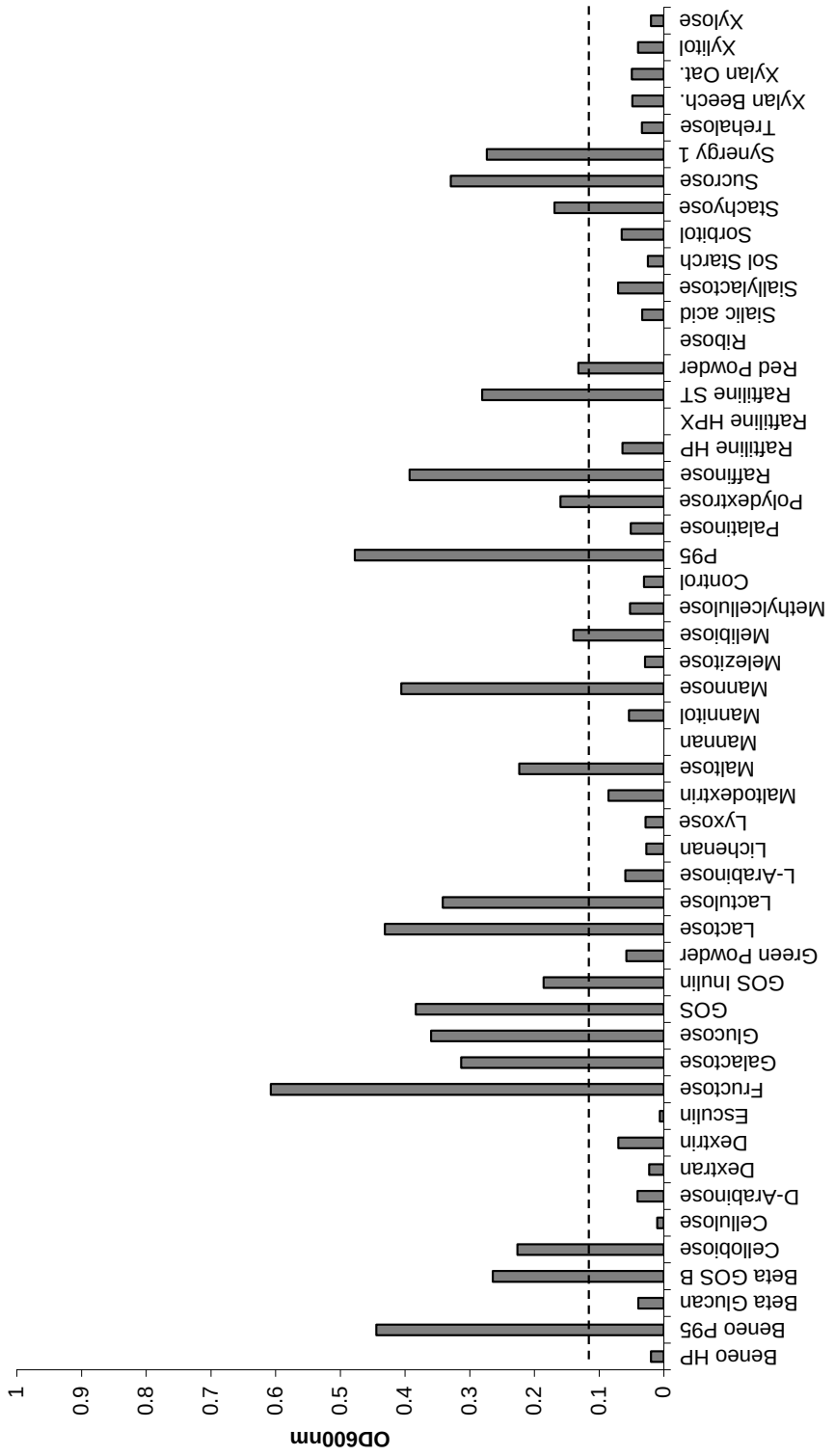


Figure S2.1 Fermentation profile for *L. ruminis* L5. Dashed line, cut-off point

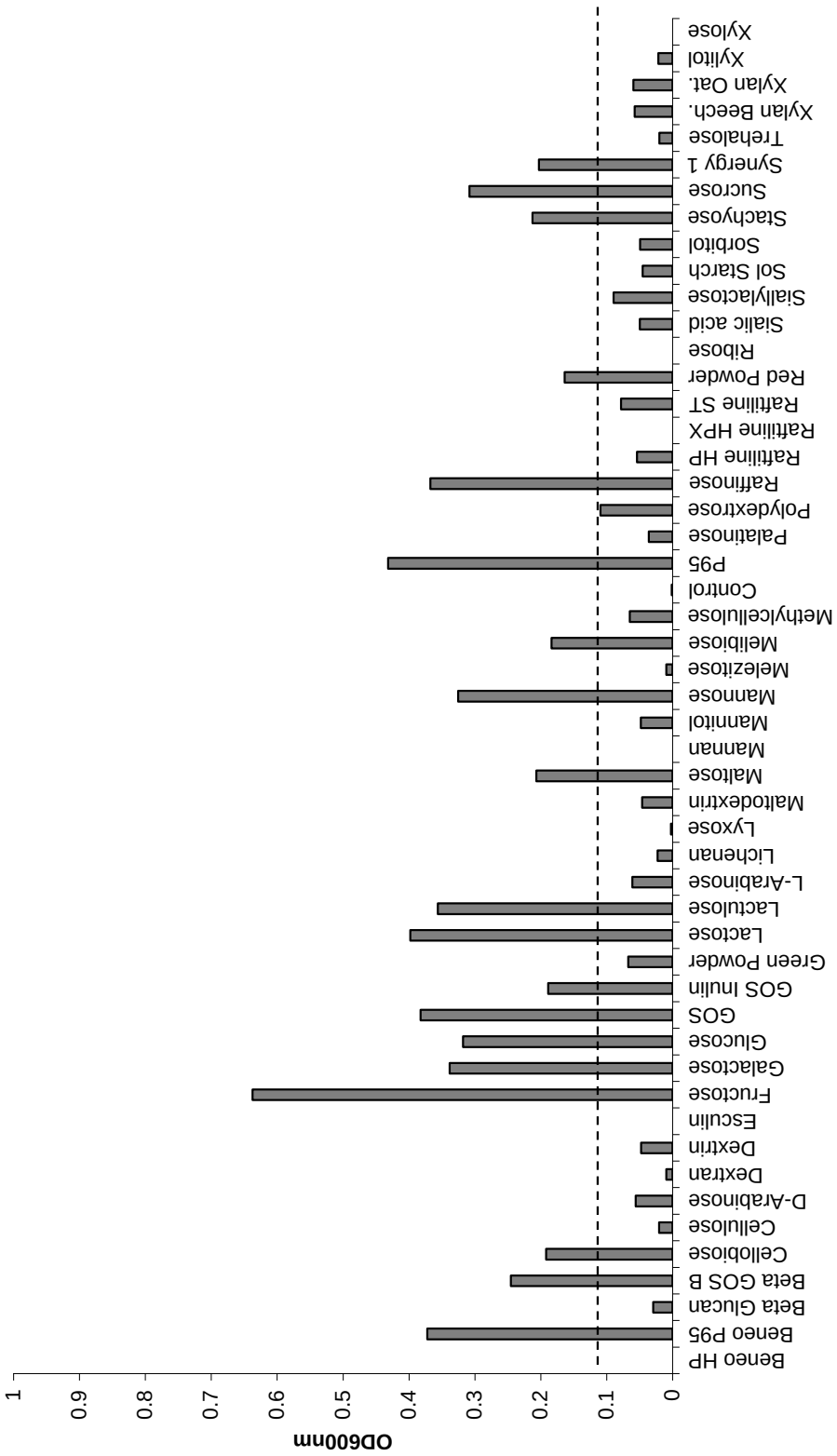


Figure S2.2 Fermentation profile for *L. ruminis* S21. Dashed line, cut-off point

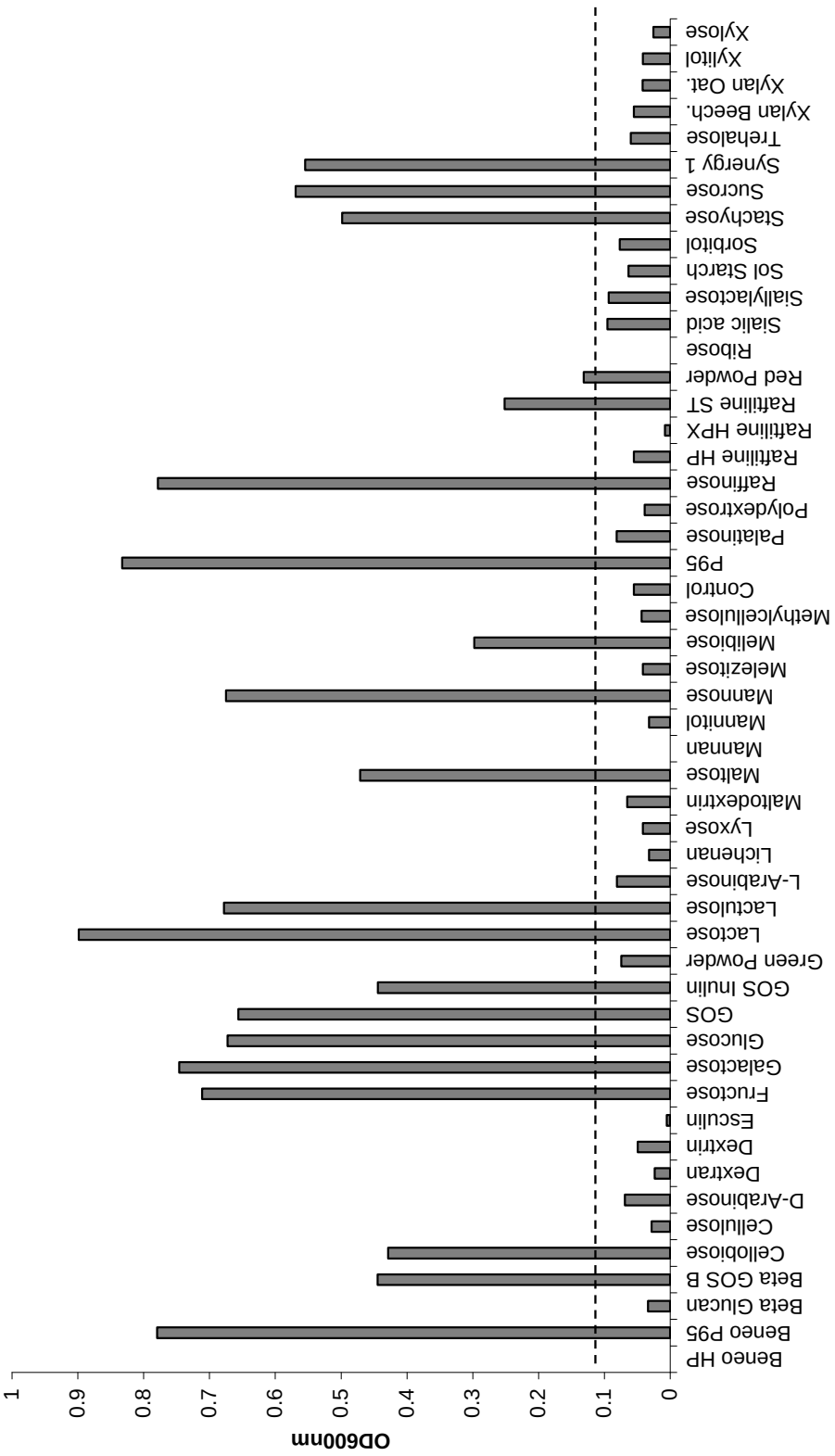


Figure S2.3 Fermentation profile for *L. ruminis* S23. Dashed line, cut-off point

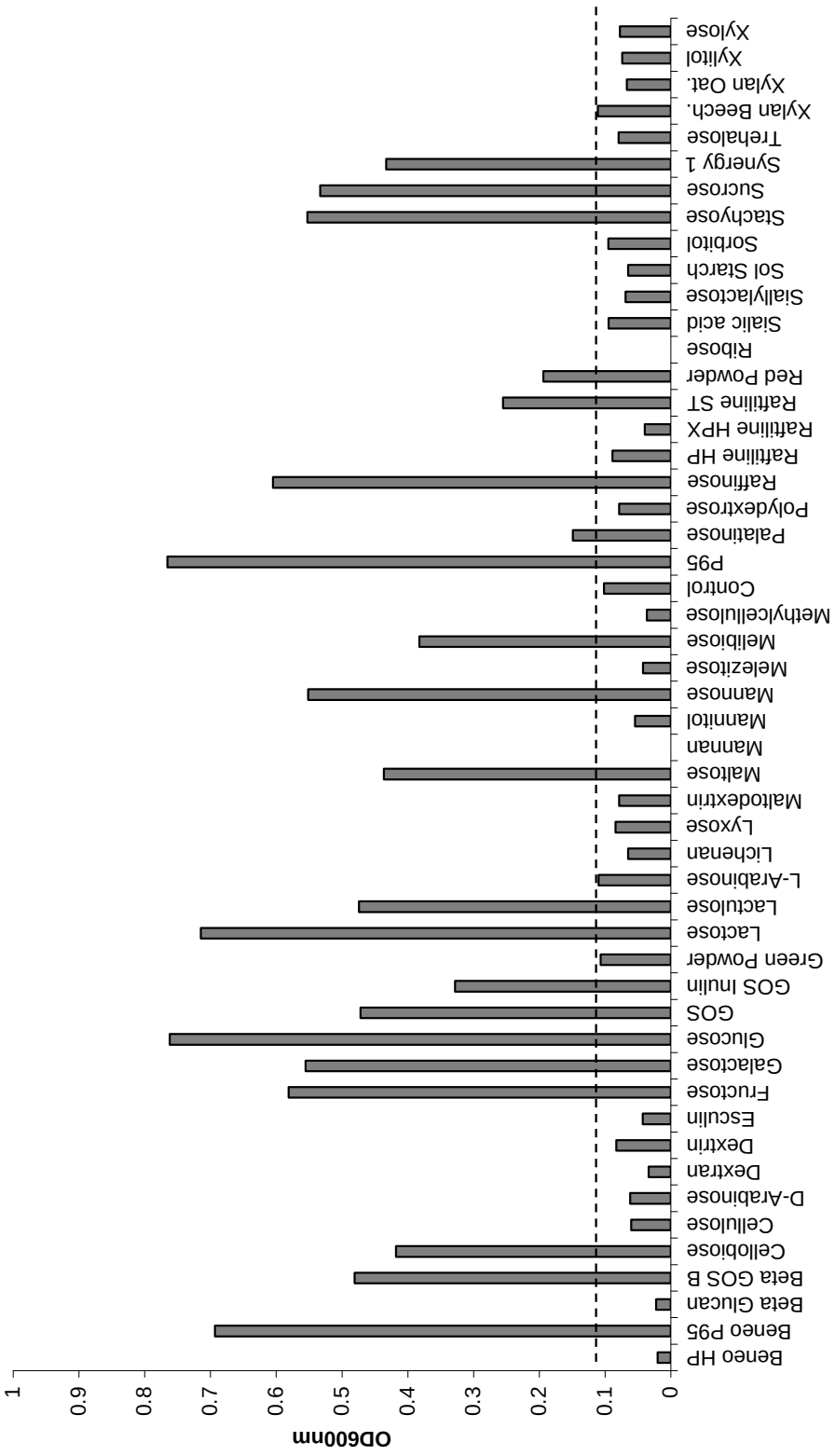


Figure S2.4 Fermentation profile for *L. ruminis* S36. Dashed line, cut-off point

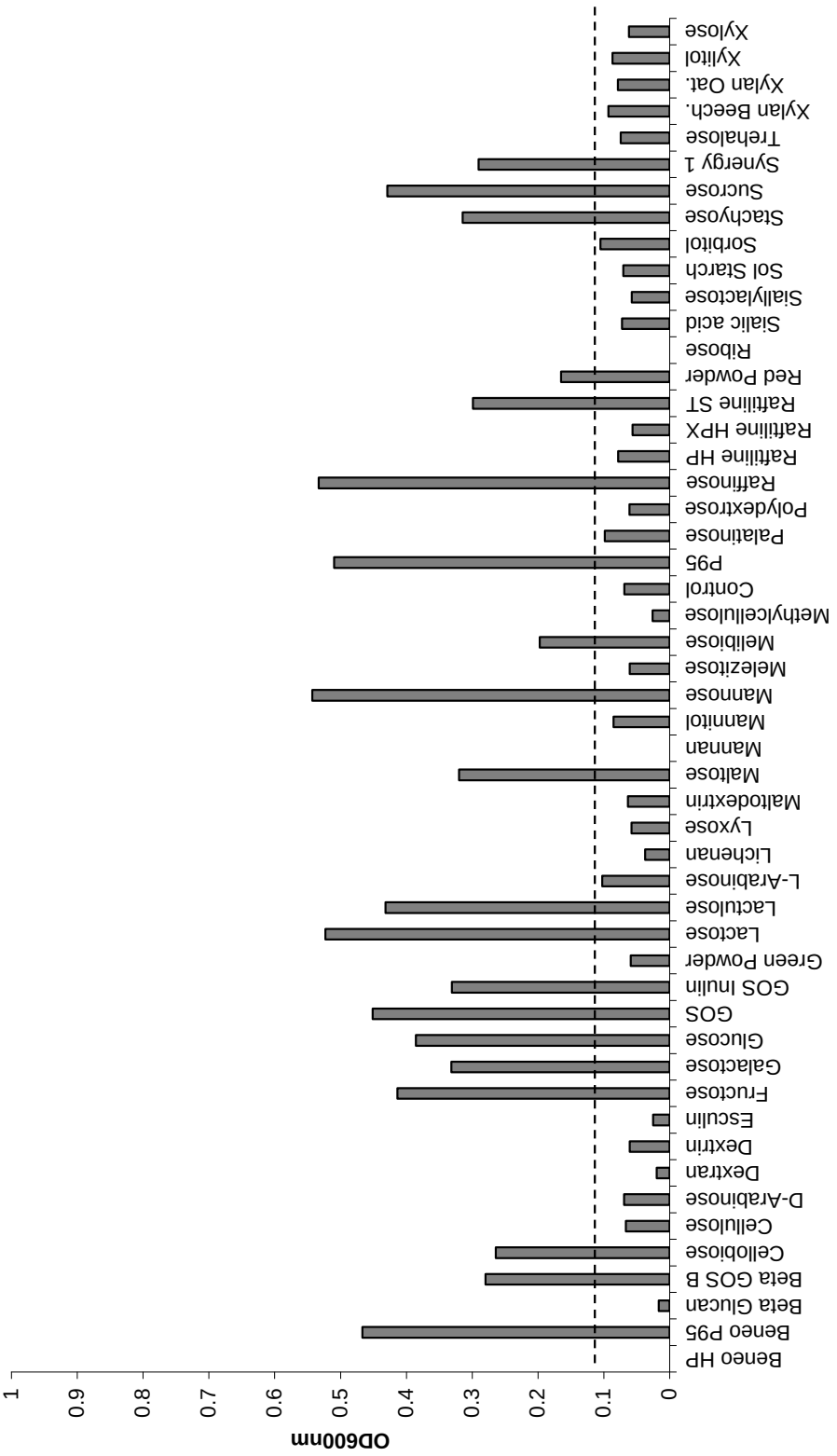


Figure S2.5 Fermentation profile for *L. ruminis* S38. Dashed line, cut-off point

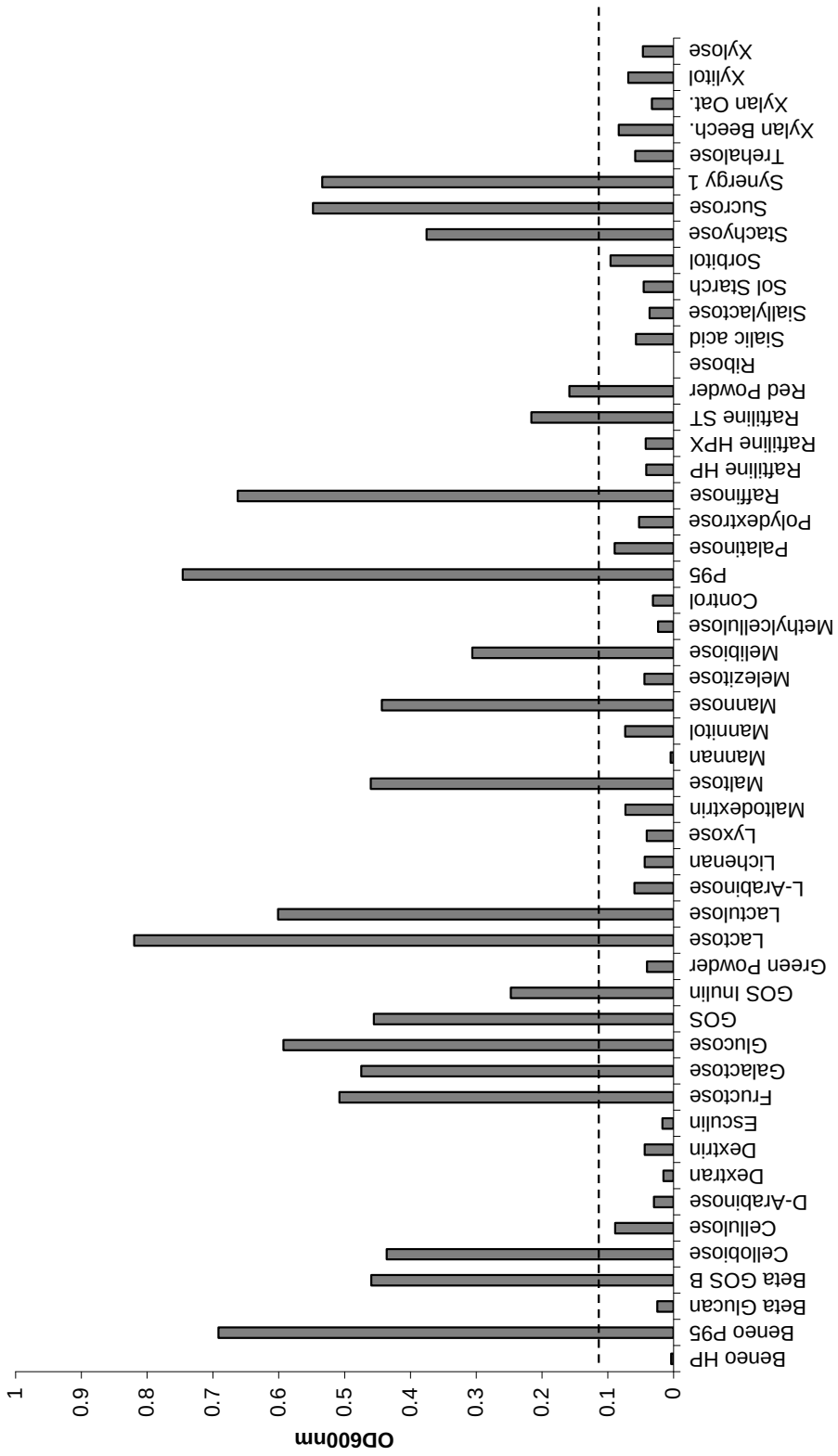


Figure S2.6 Fermentation profile for *L. ruminis* ATCC 25644. Dashed line, cut-off point

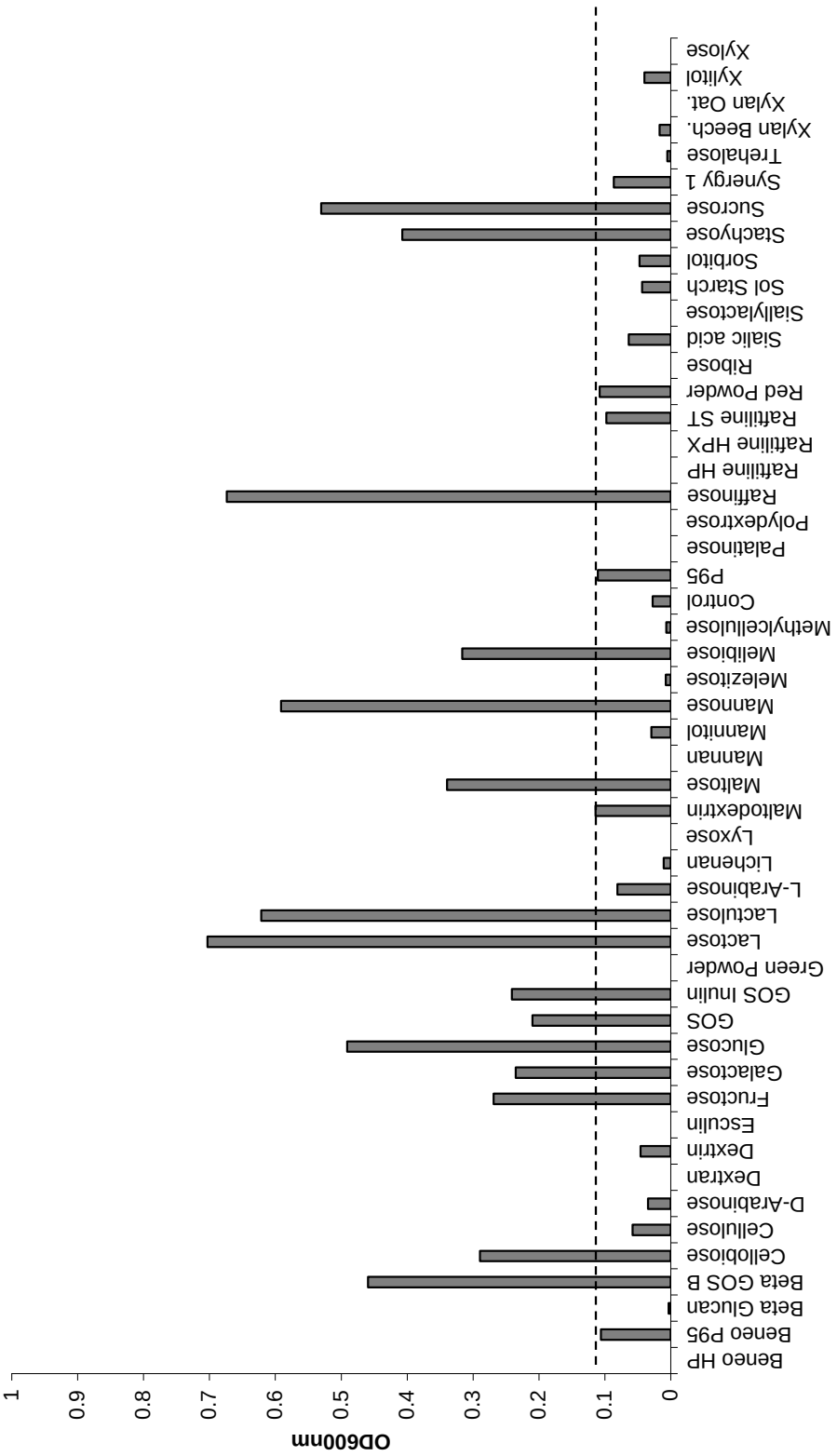


Figure S2.7 Fermentation profile for *L. ruminis* ATCC 27780T. Dashed line, cut-off point

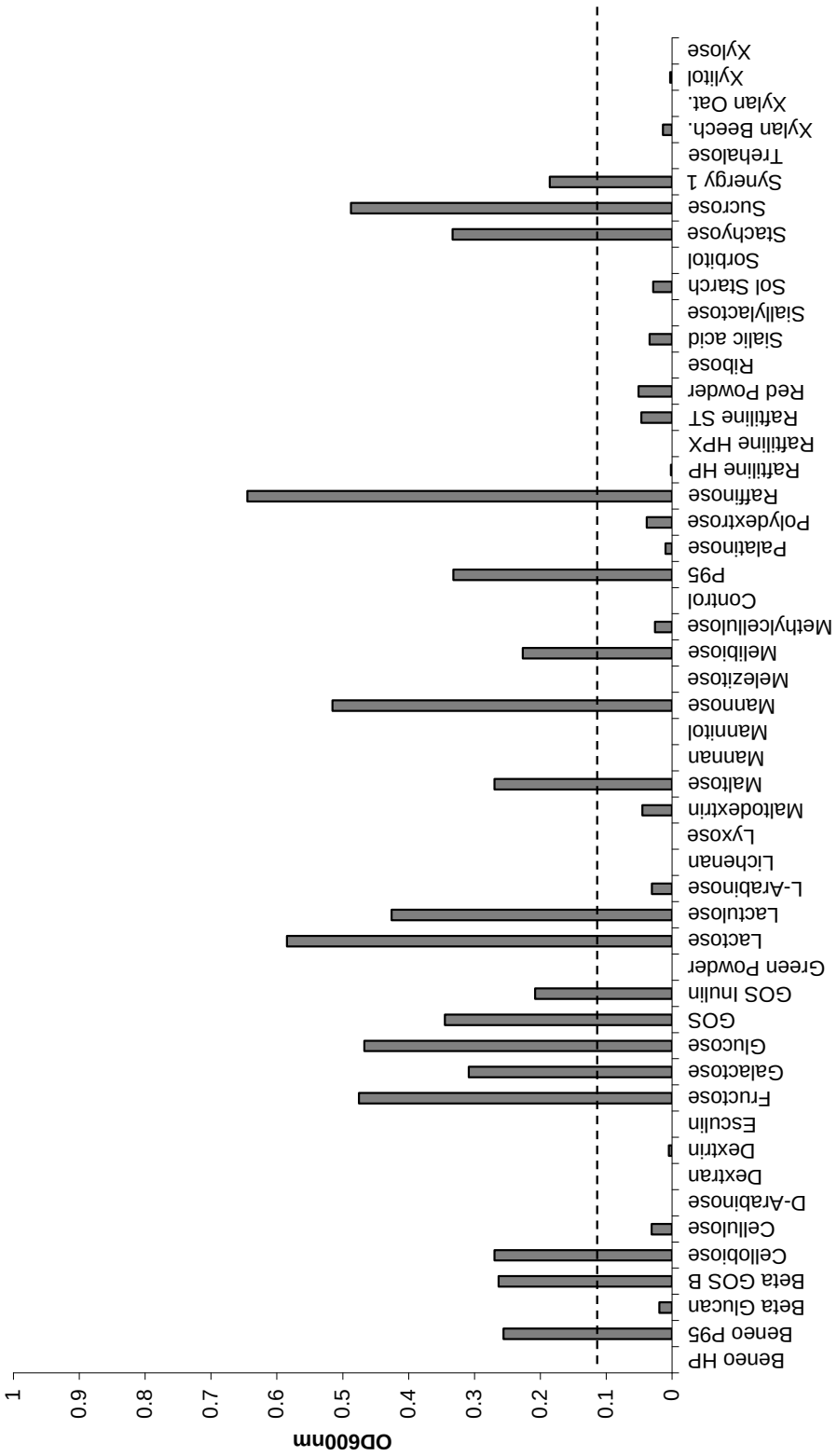


Figure S2.8 Fermentation profile for *L. ruminis* ATCC 27781. Dashed line, cut-off point

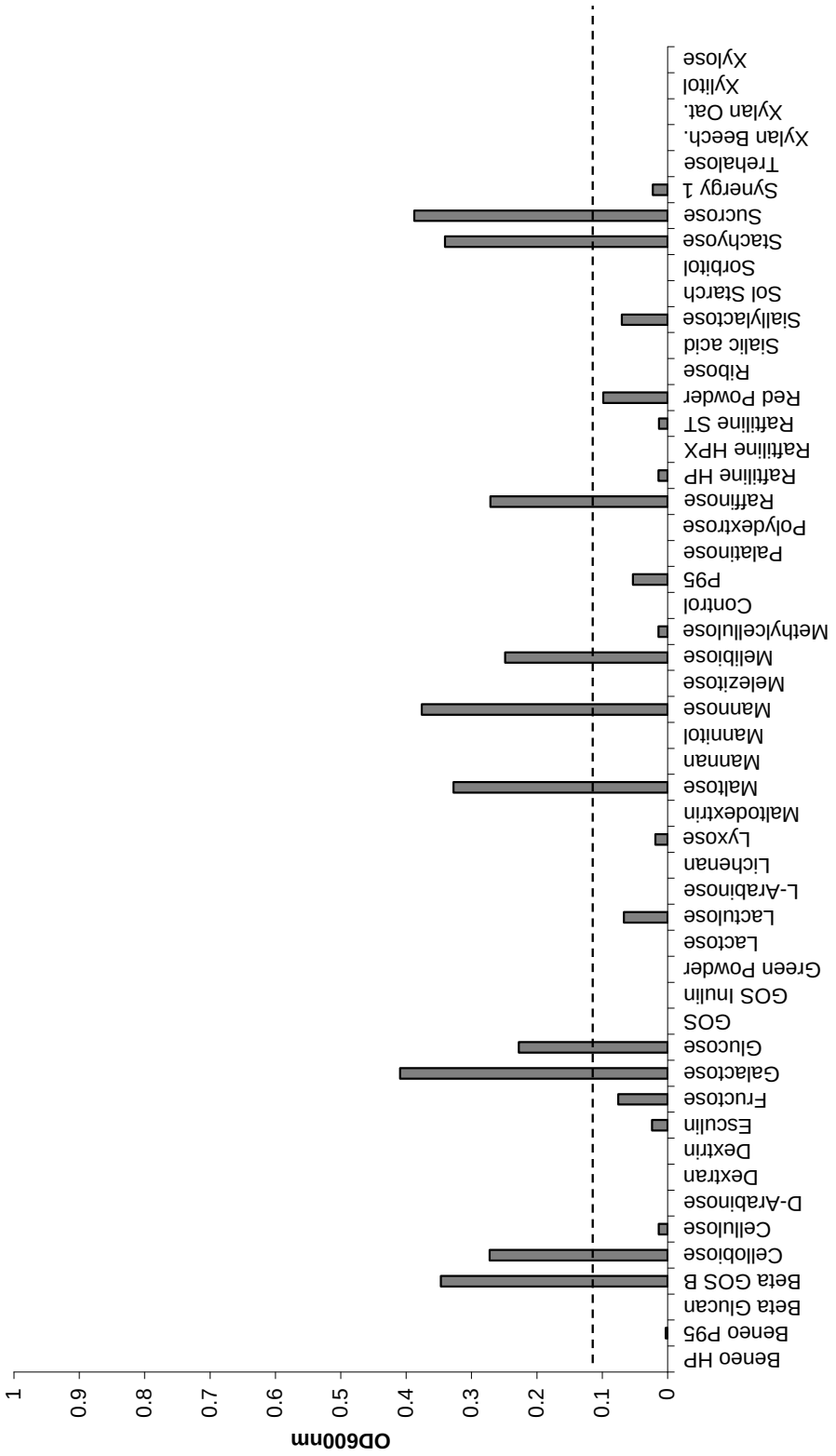


Figure S2.9 Fermentation profile for *L. ruminis* ATCC 27782. Dashed line, cut-off point

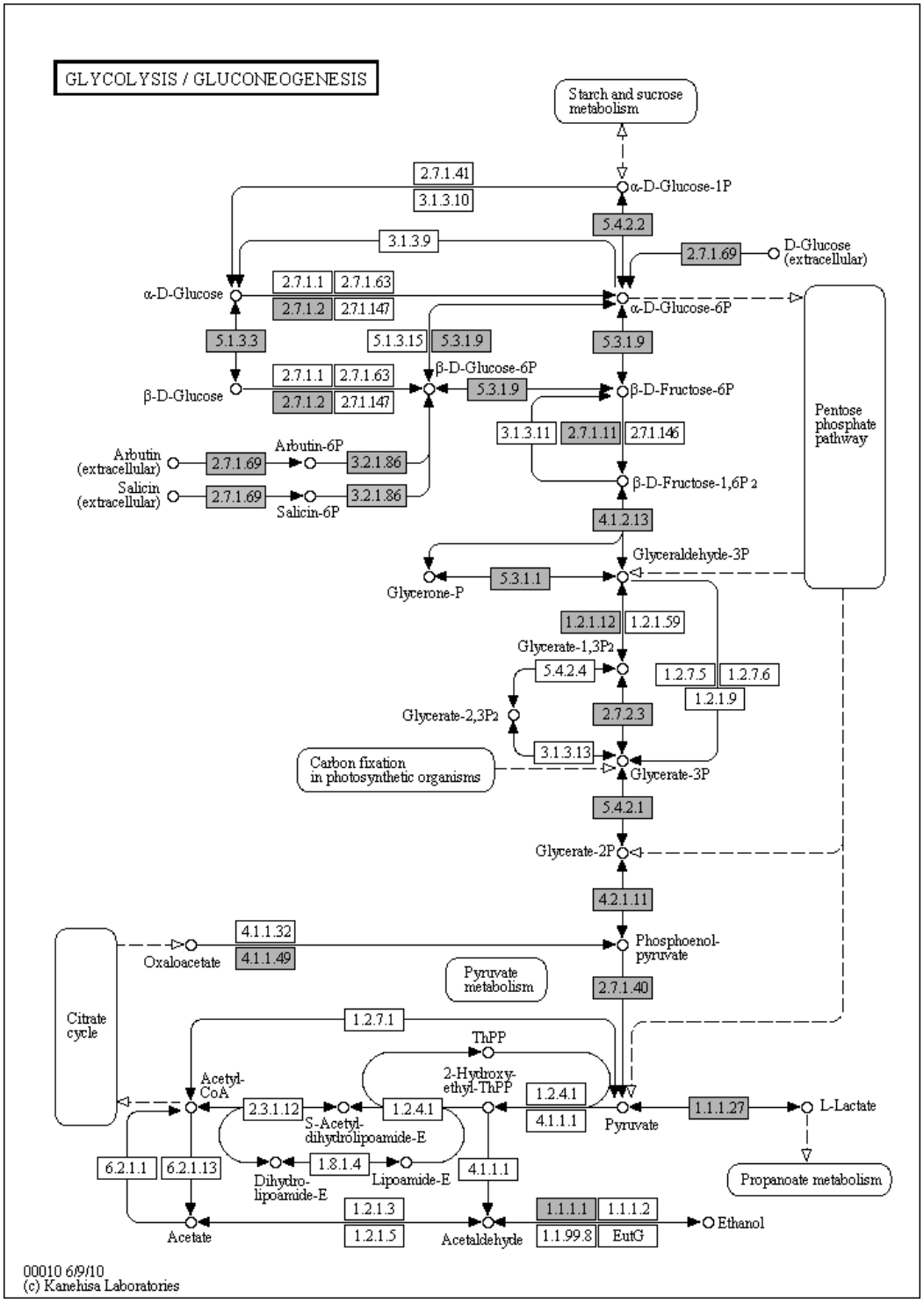


Figure S2.10 Glycolysis map representing enzymes present in both *L. ruminis* ATCC 25644 and ATCC 27782. Grey boxes, enzymes present in both sequenced strains

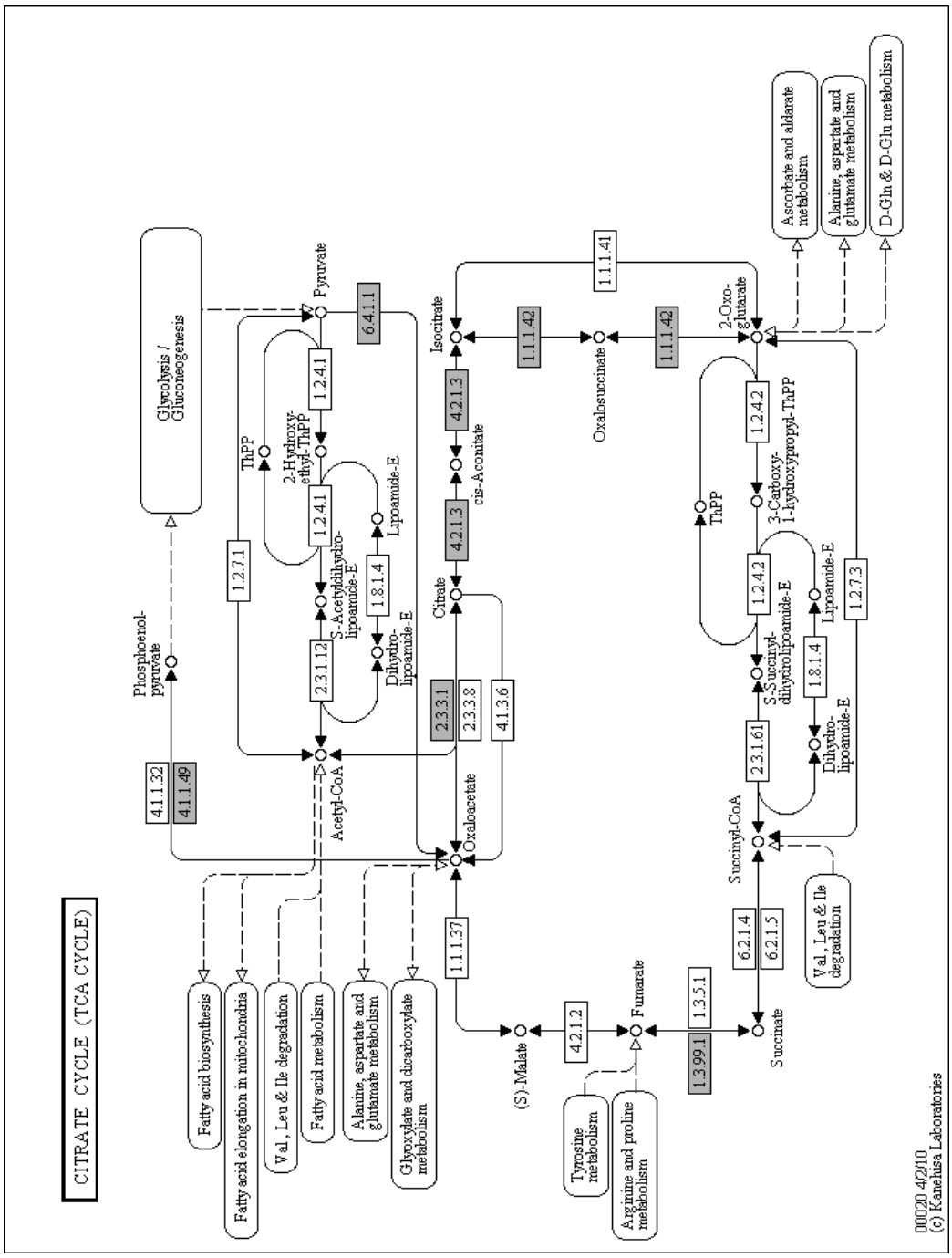


Figure S2.11 Citrate cycle map representing enzymes present in both *L. ruminis* ATCC 25644 and ATCC 27782. Grey boxes, enzymes present in both sequenced strains

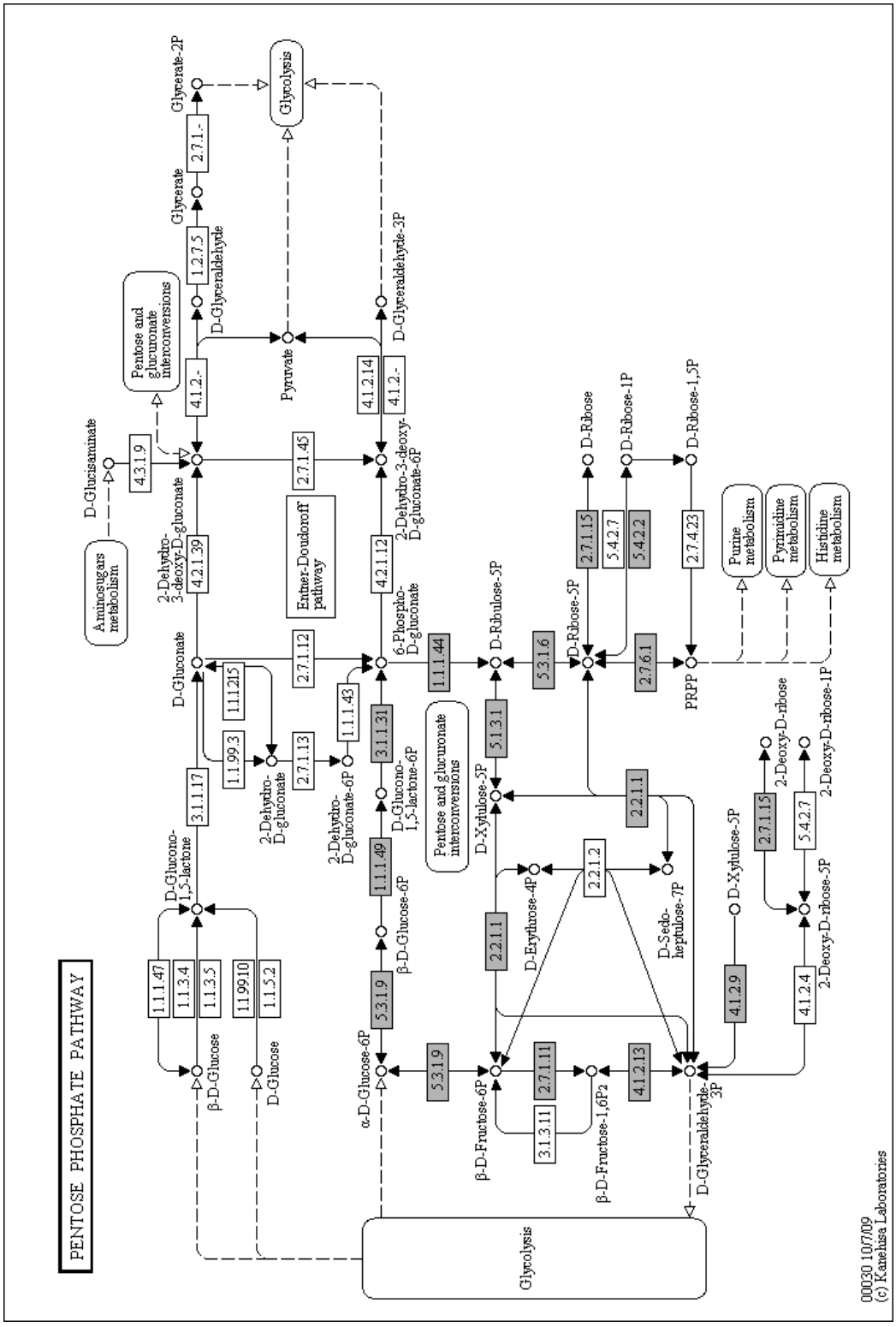


Figure S2.12 Pentose phosphate pathway map representing enzymes present in both *L. ruminis* ATCC 25644 and ATCC 27782. Grey boxes, enzymes present in both sequenced strains

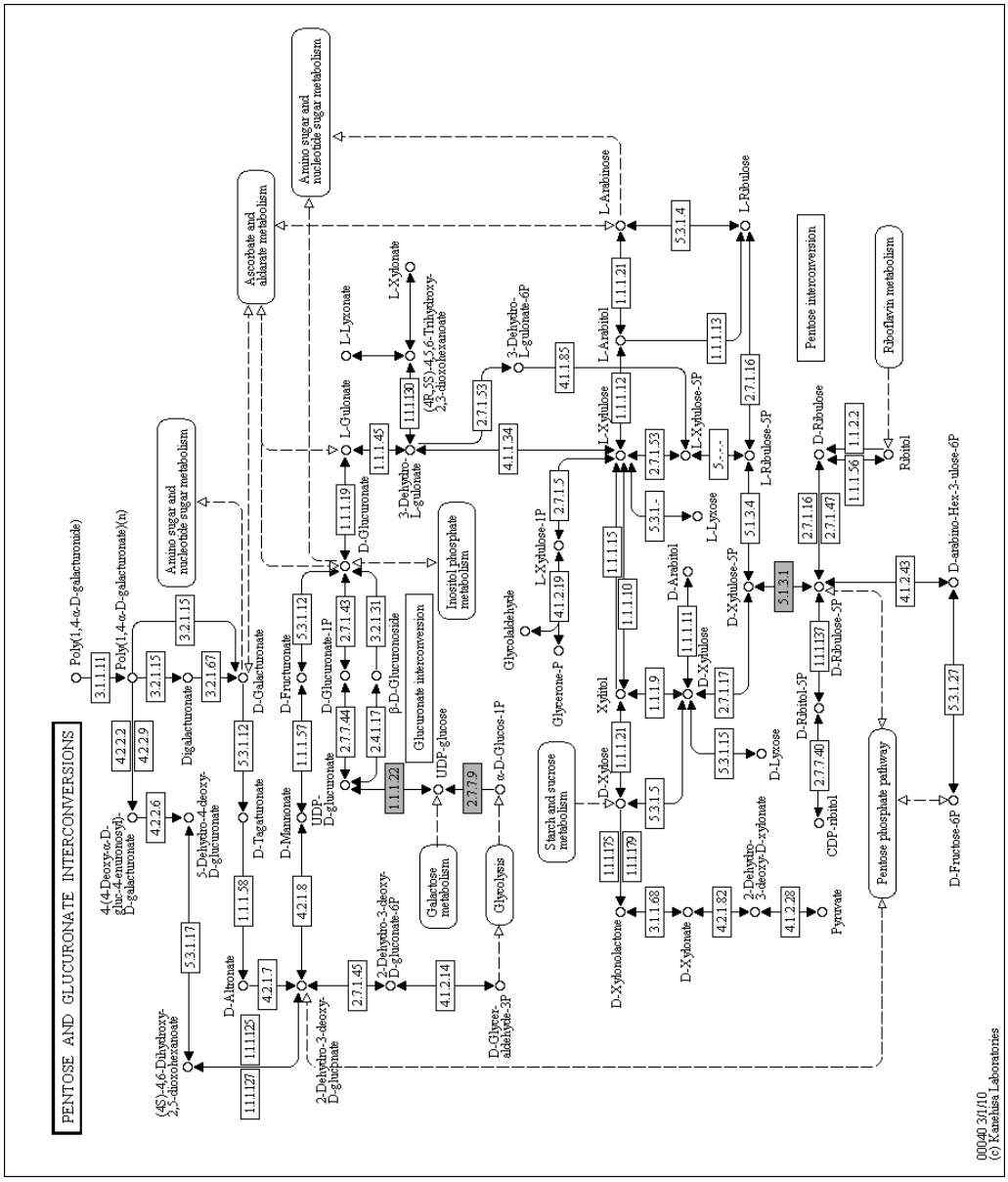


Figure S2.13 Pentose and glucuronate interconversions map representing enzymes present in both *L. ruminis* ATCC 25644 and ATCC 27782. Grey boxes, enzymes present in both sequenced strains

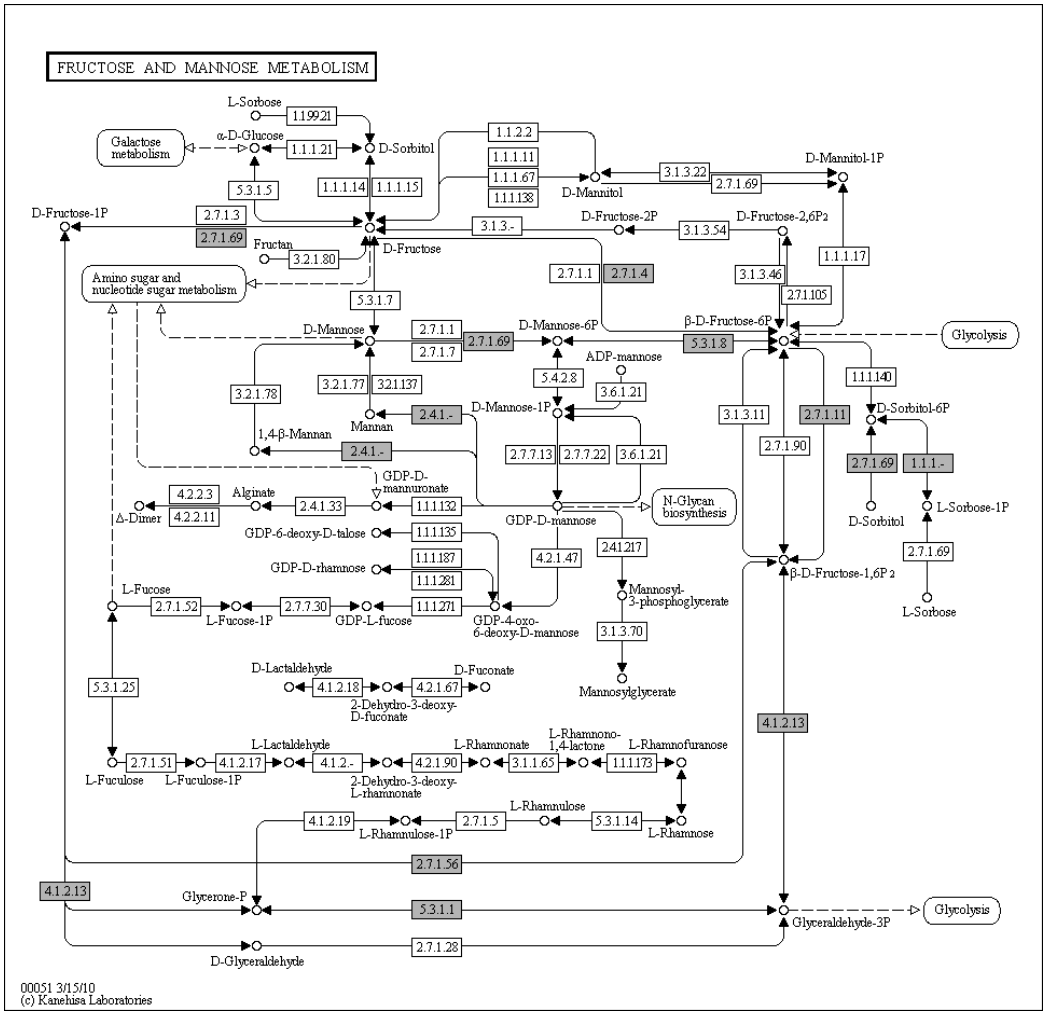


Figure S2.14 Fructose and Mannose metabolism map representing enzymes present in both *L. ruminis* ATCC 25644 and ATCC 27782. Grey boxes, enzymes present in both sequenced strains

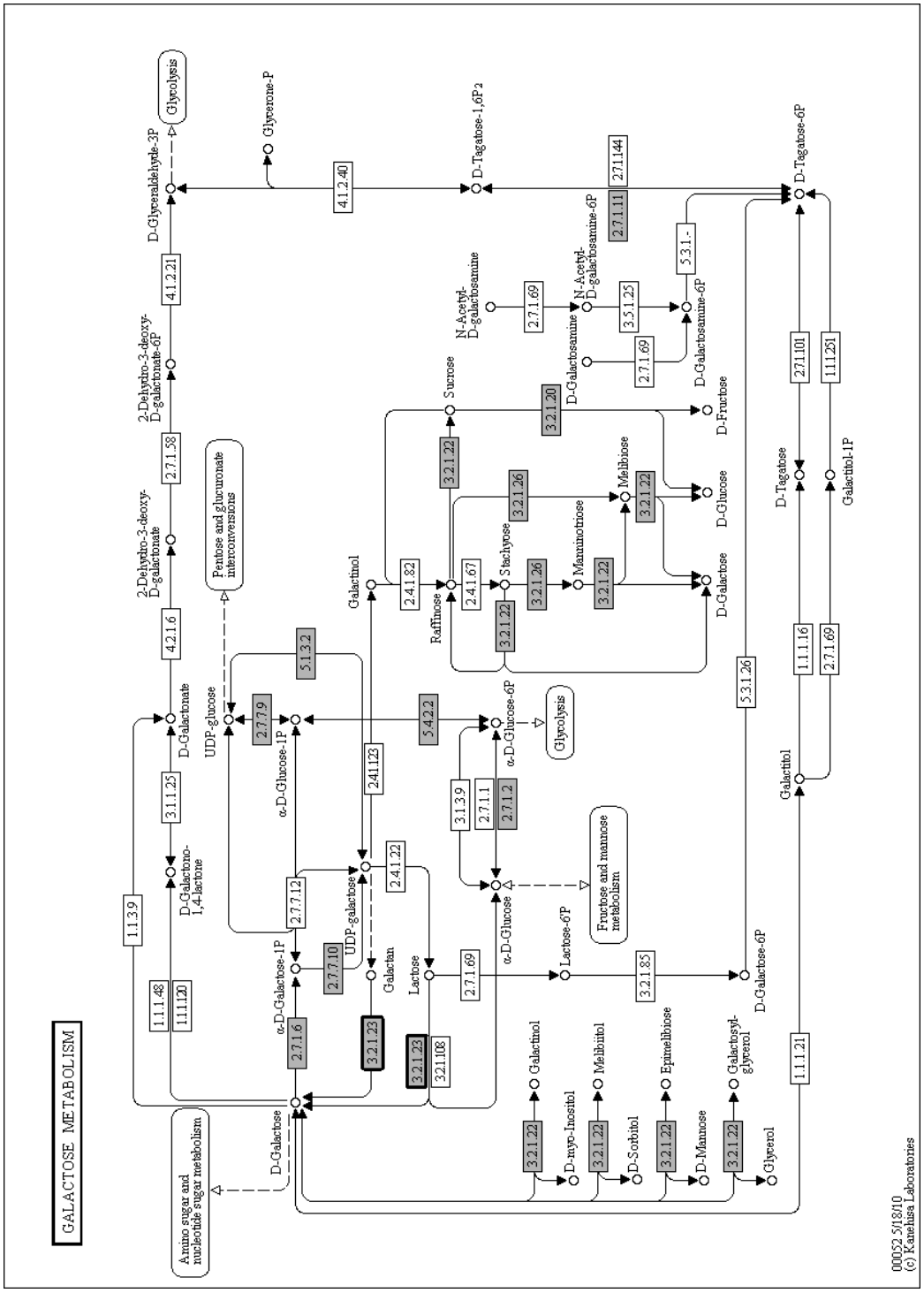


Figure S2.15 Galactose metabolism map representing enzymes present in both *L. ruminis* ATCC 25644 and ATCC 27782. Grey boxes, enzymes present in both sequenced strains; Grey boxes with bold border, enzymes present in ATCC 25644 only.

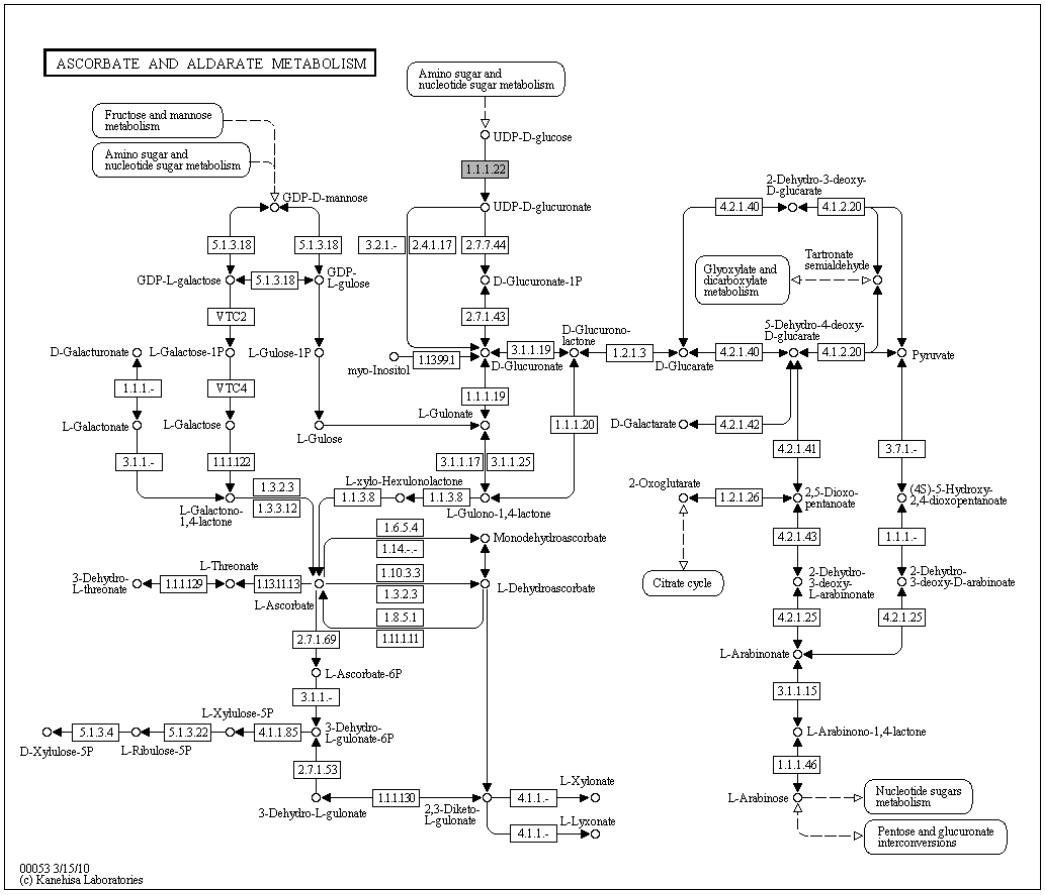


Figure S2.16 Ascorbate and aldarate metabolism map representing enzymes present in both *L. ruminis* ATCC 25644 and ATCC 27782. Grey boxes, enzymes present in both sequenced strains

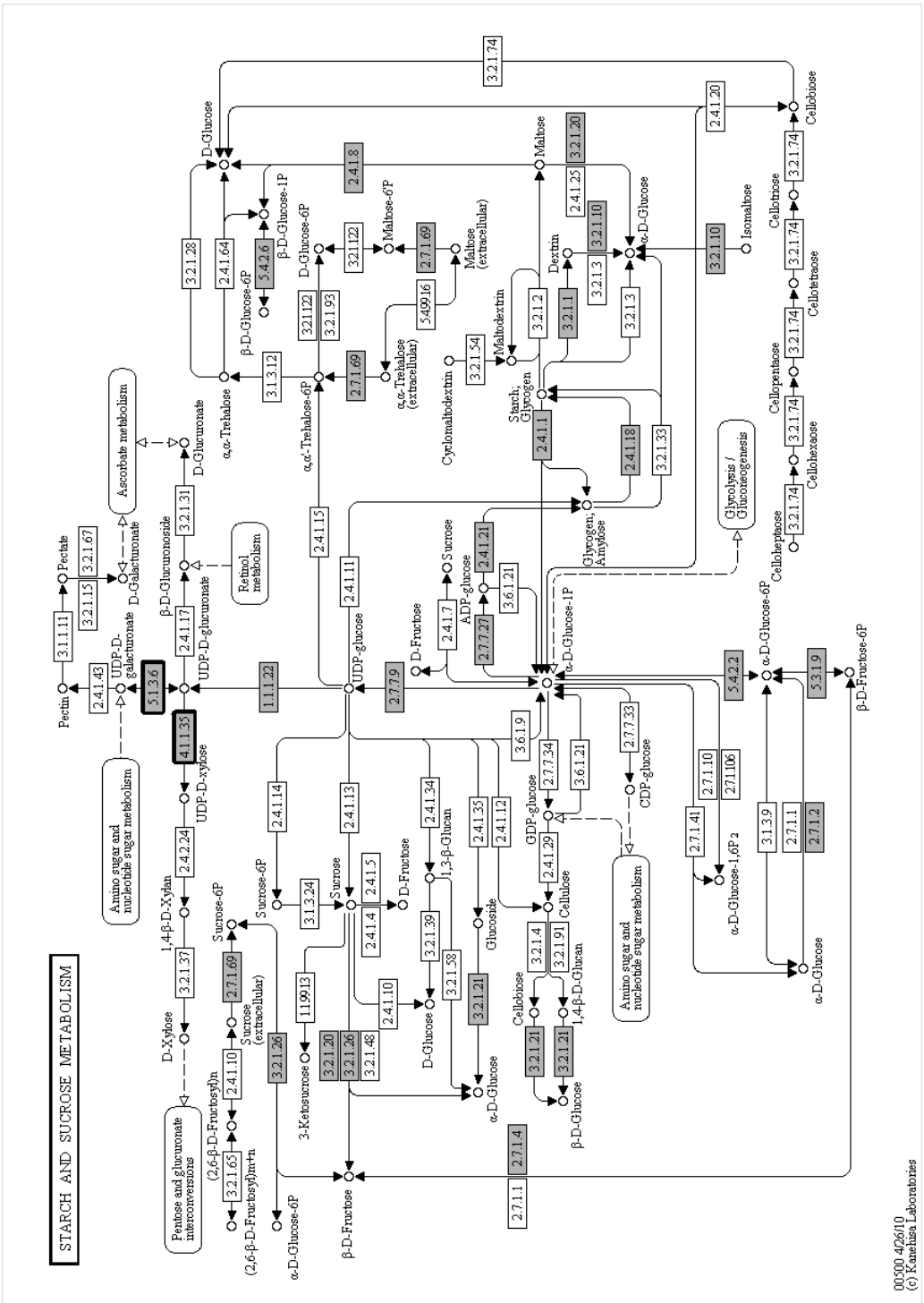


Figure S2.17 Starch and sucrose metabolism map representing enzymes present in both *L. ruminis* ATCC 25644 and ATCC 27782. Grey boxes, enzymes present in both sequenced strains

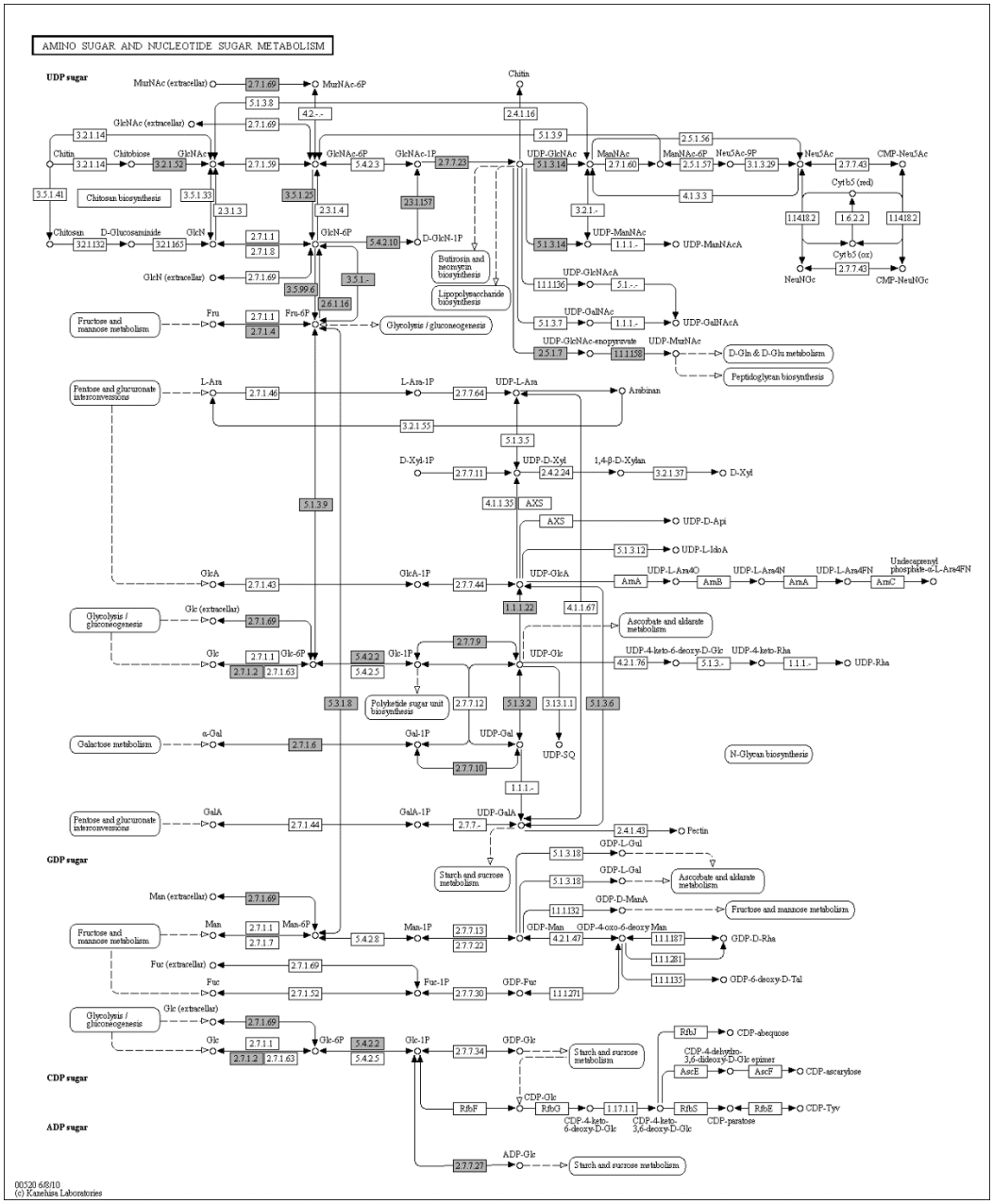


Figure S2.18 Amino and nucleotide sugar metabolism map representing enzymes present in both *L. ruminis* ATCC 25644 and ATCC 27782. Grey boxes, enzymes present in both sequenced strains

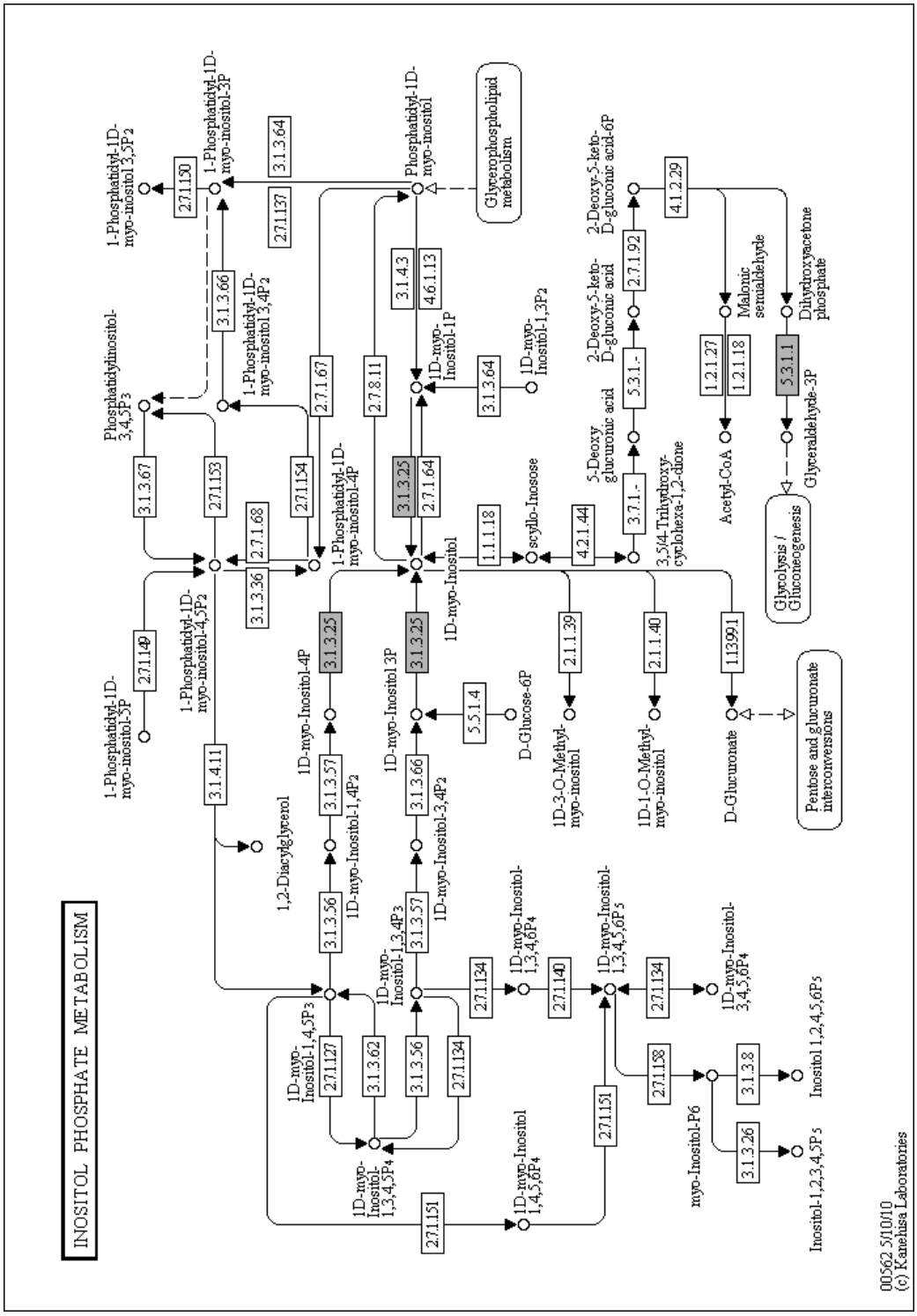


Figure S2.19 Inositol Phosphate metabolism map representing enzymes present in both *L. ruminis* ATCC 25644 and ATCC 27782. Grey boxes, enzymes present in both sequenced strains

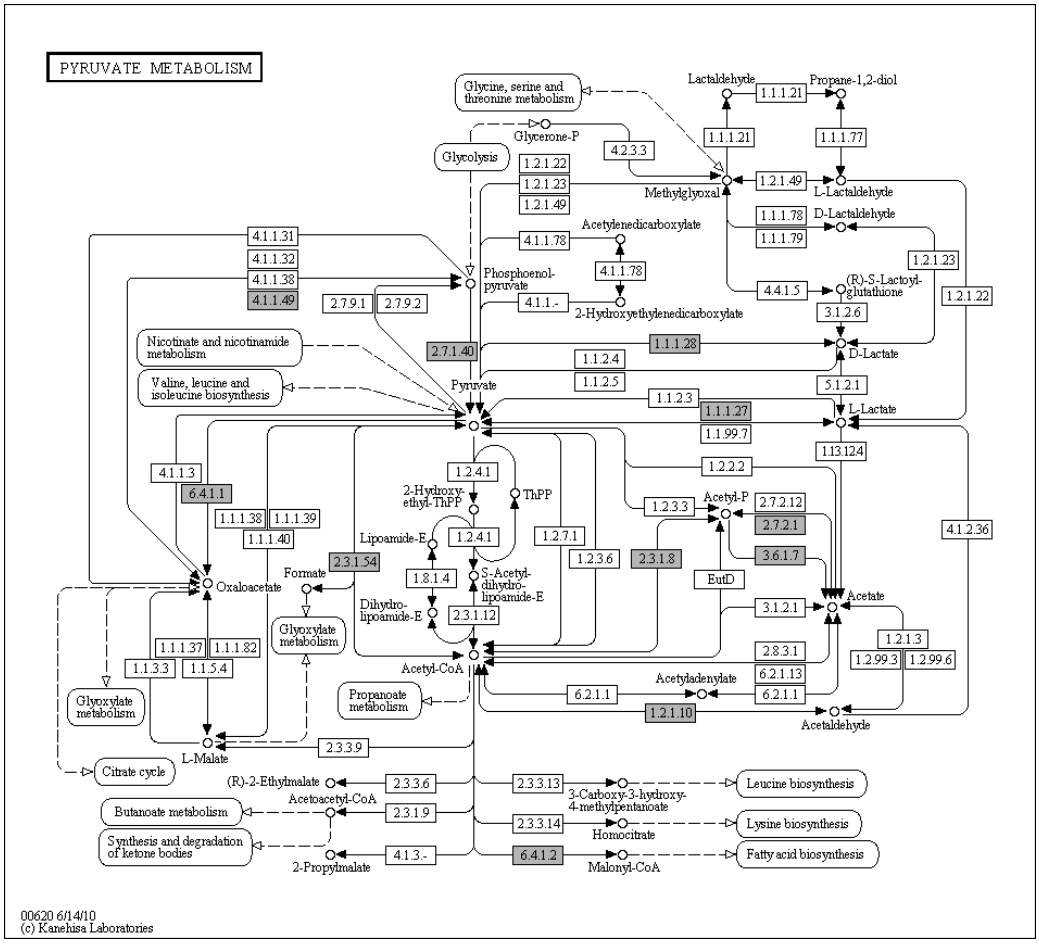


Figure S2.20 Pyruvate metabolism map representing enzymes present in both *L. ruminis* ATCC 25644 and ATCC 27782. Grey boxes, enzymes present in both sequenced strains

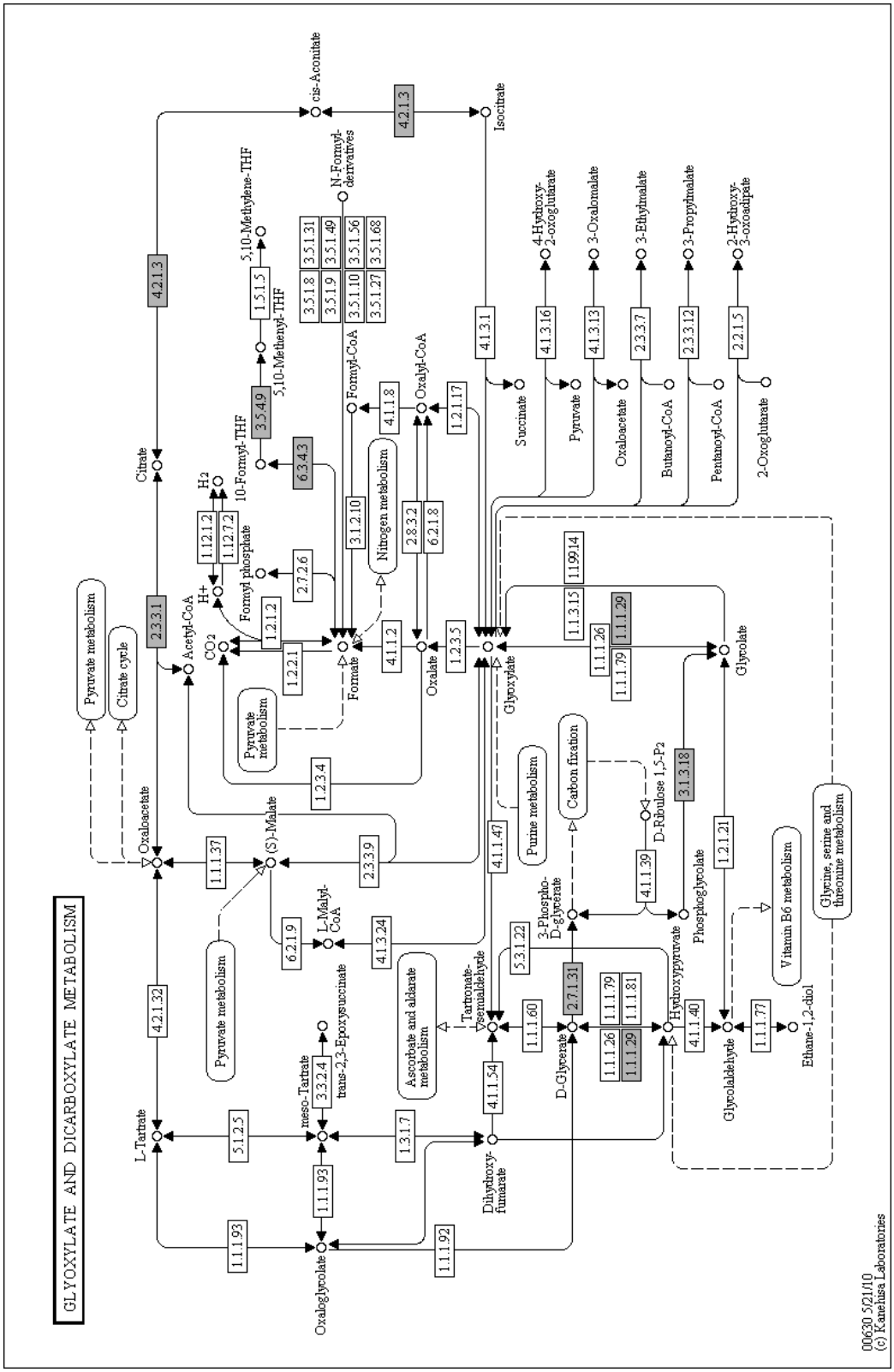


Figure S2.21 Glyoxylate and Dicarboxylate metabolism map representing enzymes present in both *L. ruminis* ATCC 25644 and ATCC 27782. Grey boxes, enzymes present in both sequenced strains

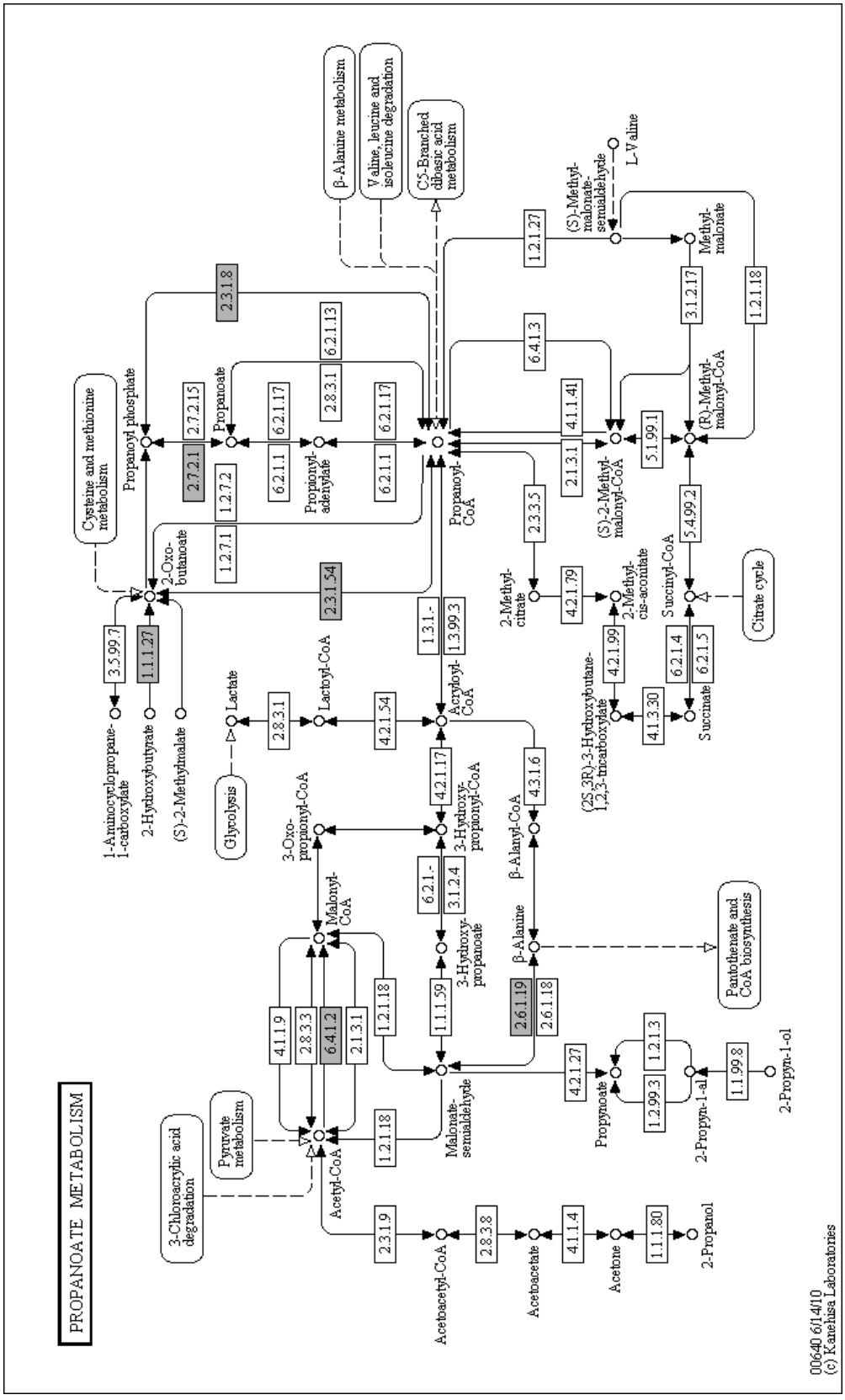


Figure S2.22 Propanoate metabolic map representing enzymes present in both *L. ruminis* ATCC 25644 and ATCC 27782. Grey boxes, enzymes present in both sequenced strains

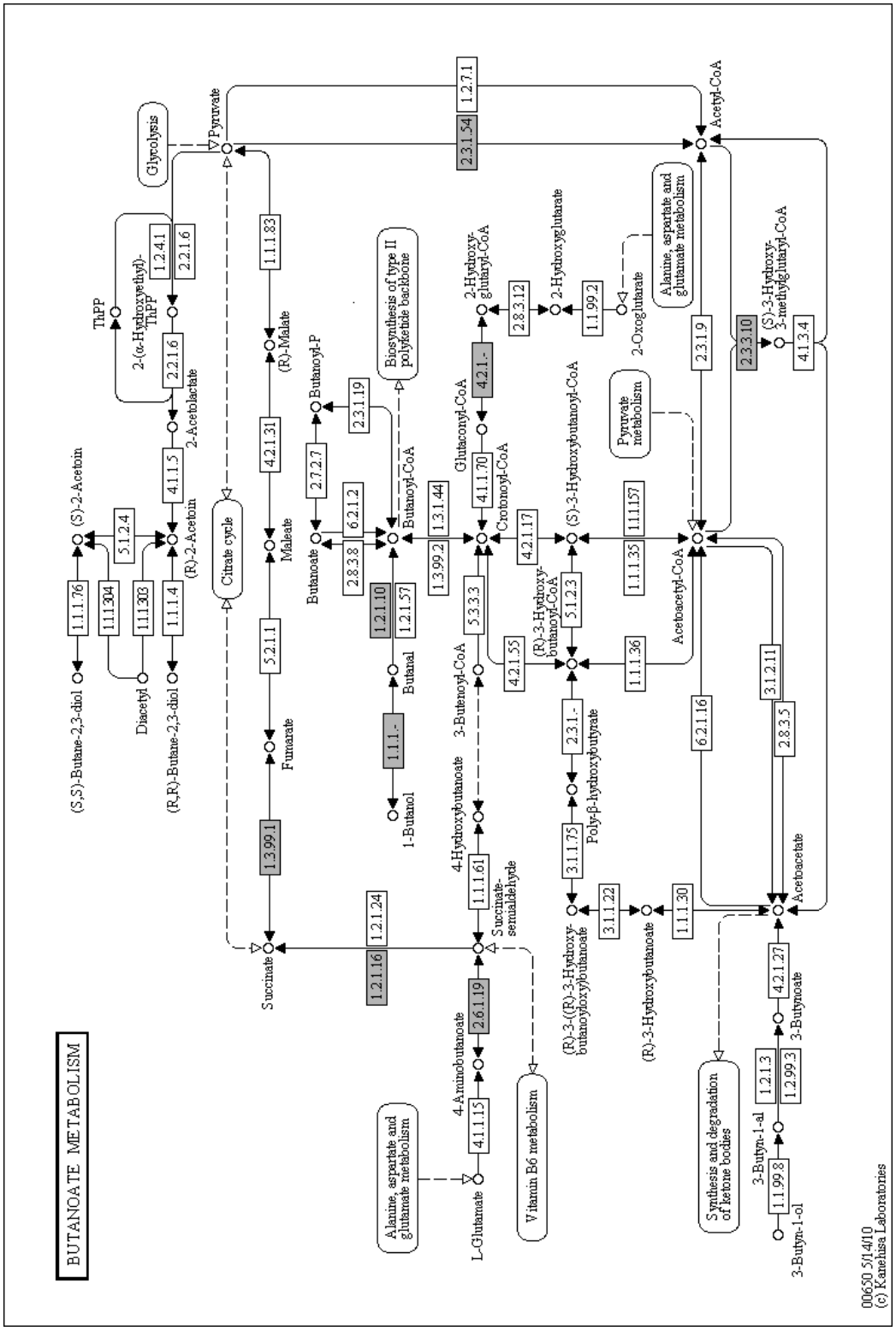


Figure S2.23 Butanoate metabolic map representing enzymes present in both *L. ruminis* ATCC 25644 and ATCC 27782. Grey boxes, enzymes present in both sequenced strains

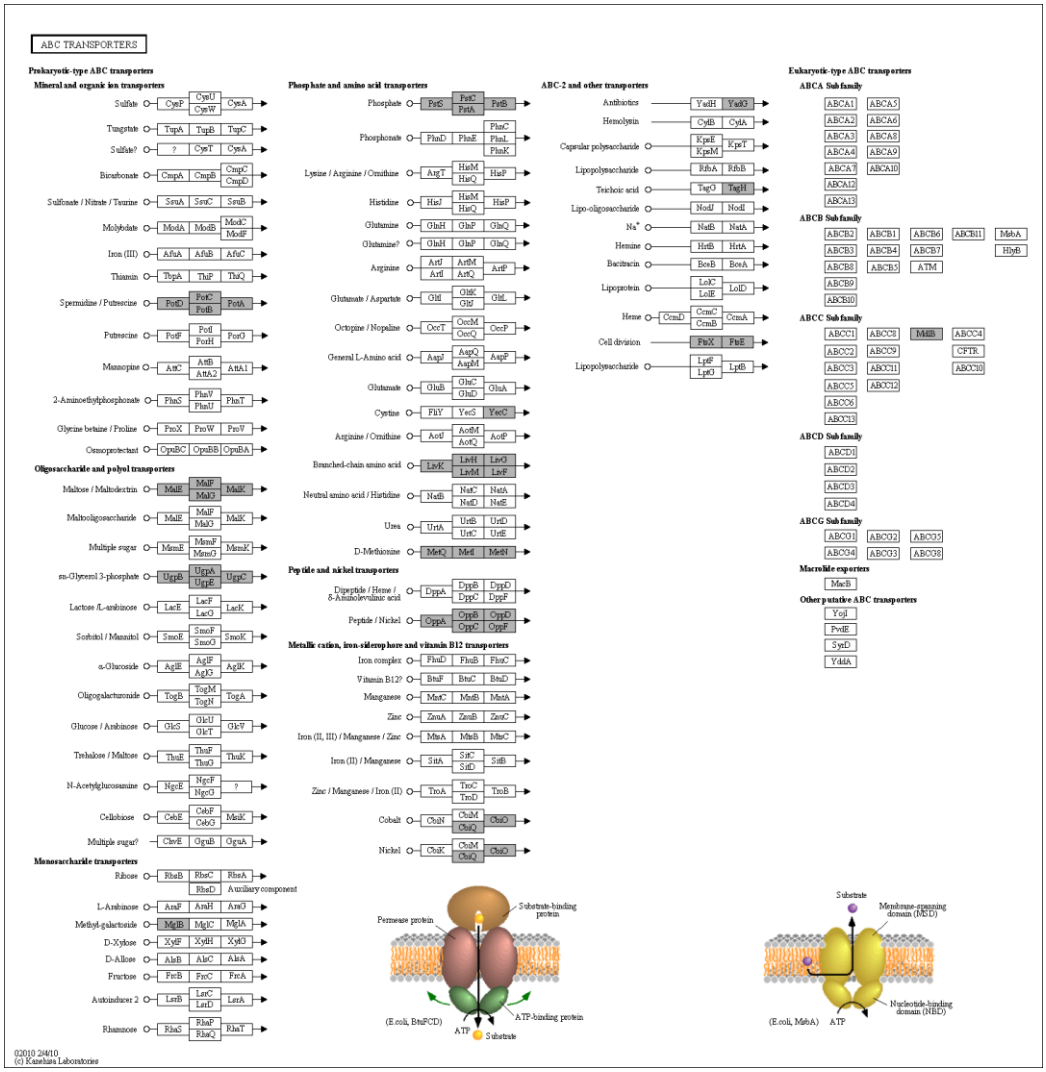


Figure S2.24 ABC transporters map representing enzymes present in both *L. ruminis* ATCC 25644 and ATCC 27782. Grey boxes, enzymes present in both sequenced strains

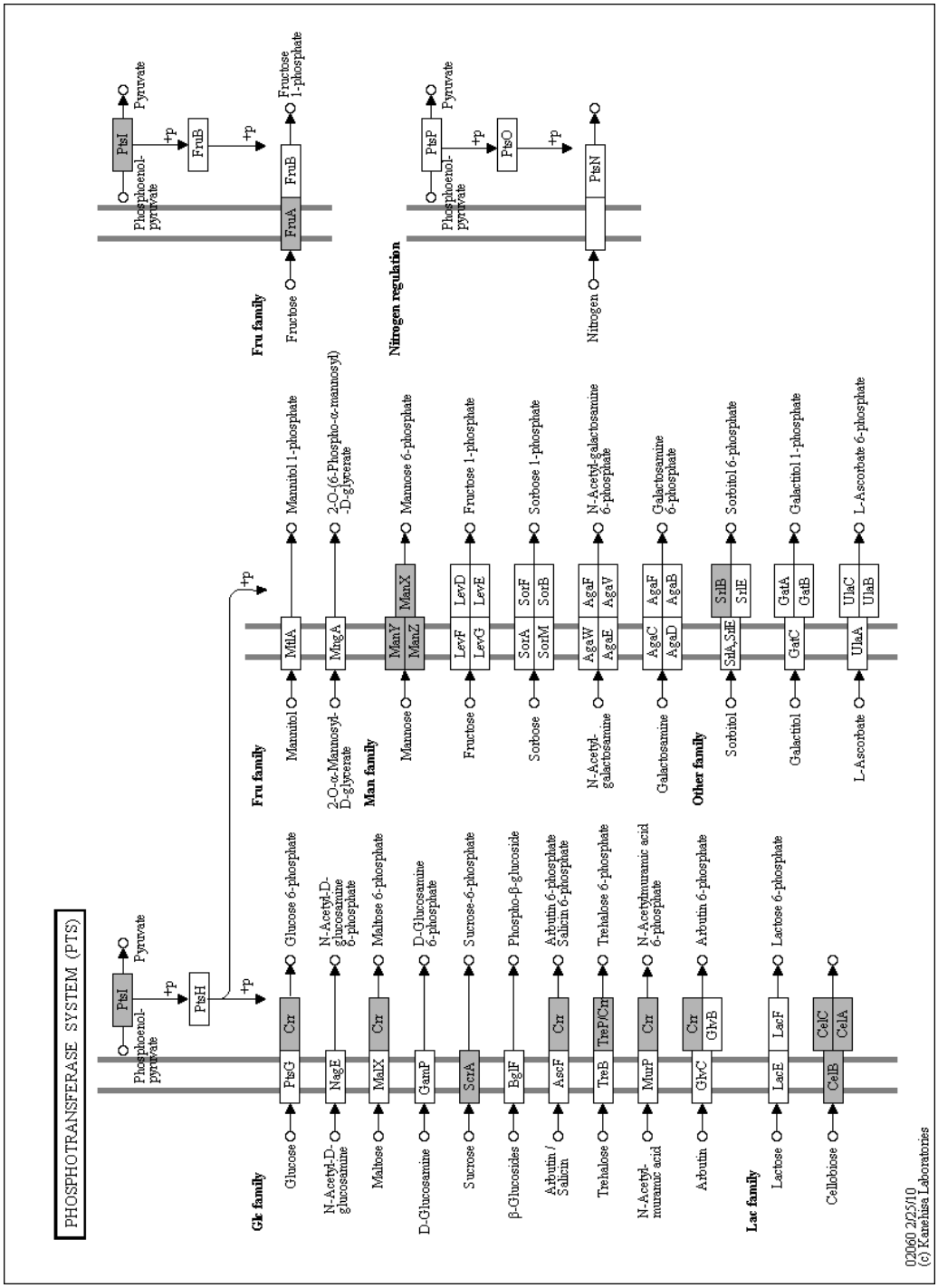


Figure S2.25 Phosphotransferase system map representing enzymes present in both *L. ruminis* ATCC 25644 and ATCC 27782. Grey boxes, enzymes present in both sequenced strains

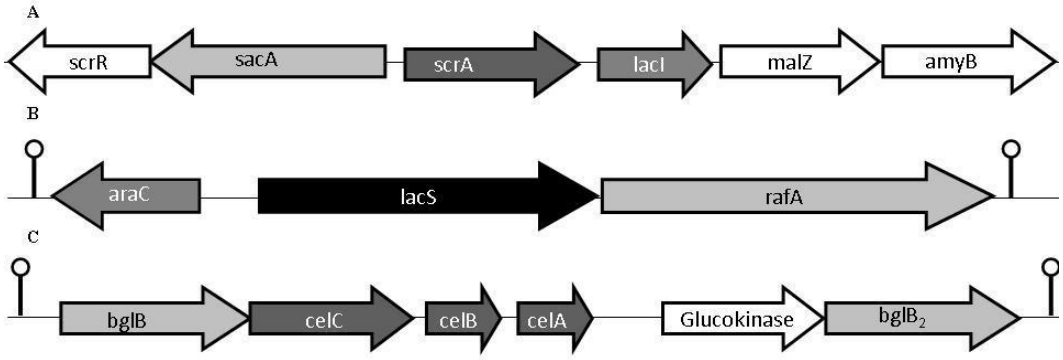


Figure S2.26. Putative operons predicted to be involved in the utilisation of carbohydrates in ATCC 27782. A, Sucrose operon; B, Raffinose operon; C, Cellobiose operon. Light grey arrows, glycosyl hydrolase family enzyme; Black arrows, major facilitator superfamily transporters; Medium grey arrows transcriptional regulators; Dark grey arrows, phosphotransferase system transporters; Lollipops, rho-independent transcriptional regulators.

Table S2.1 - Carbohydrates used in this study

Carbohydrate type	Name	Source	Degree of Polymerisation
Monosaccharide	Glucose	Fisher Scientific	n/a
	Fructose	Sigma Aldrich, Poole, UK	n/a
	Galactose	Sigma Aldrich, Poole, UK	n/a
	D – Arabinose	Sigma Aldrich, Poole, UK	n/a
	L - Arabinose	Sigma Aldrich, Poole, UK	n/a
	Mannose	Sigma Aldrich, Poole, UK	n/a
	Ribose	Sigma Aldrich, Poole, UK	n/a
	Lyxose	Sigma Aldrich, Poole, UK	n/a
	Xylose	Sigma Aldrich, Poole, UK	n/a
	Sialic acid	Friesland Foods, Zwolle, Netherlands	n/a
Disaccharide	Cellobiose	Sigma Aldrich, Poole, UK	2
	Trehalose	Sigma Aldrich, Poole, UK	2
	Sucrose	Sigma Aldrich, Poole, UK	2
	Maltose	Sigma Aldrich, Poole, UK	2
	Lactose	Sigma Aldrich, Poole, UK	2
	Lactulose	Sigma Aldrich, Poole, UK	2
	Melibiose	Sigma Aldrich, Poole, UK	2
	Palatinose	Sigma Aldrich, Poole, UK	2
Trisaccharide	Melezitose	Sigma Aldrich, Poole, UK	3
	Raffinose	Sigma Aldrich, Poole, UK	3
Tetrasaccharide	Stachyose	Sigma Aldrich, Poole, UK	4
Oligosaccharide	Soluble Starch	BDH Analar	-
	Maltodextrin	Cargill-Cerestar	Avg. 7
	Polydextrose	Danisco	Avg. 12
	Galactooligosaccharide (GOS)	Friesland Foods ,Zwolle, Netherlands	2 to 8
	GOS inulin	Friesland Foods, Zwolle, Netherlands	Unknown
	β-Glucotriose B (β-glucan hydrolysate)	Megazyme, Co. Wicklow, Ireland	3
	Raftilose P95	Orafti, Tienen, Belgium	2 to 8
	Raftilose Synergy 1 (oligofructose enriched inulin)	Orafti, Tienen, Belgium	2 to 8
	Beneo P95	Orafti, Tienen, Belgium	2 to 8
	Dextran	Sigma Aldrich, Poole, UK	- -
	Dextrin	Sigma Aldrich, Poole, UK	-
Polysaccharide	β Glucan	Megazyme, Co. Wicklow, Ireland	>100
	Mannan	Megazyme, Co. Wicklow, Ireland	15
	Lichenan	Megazyme, Co. Wicklow, Ireland	80-400
	Beneo HP	Orafti, Tienen, Belgium	>23
	Raftiline ST	Orafti, Tienen, Belgium	≥10
	Raftiline HPX	Orafti, Tienen, Belgium	≥23
	Raftiline HP	Orafti, Tienen, Belgium	>23
	Xylan from Beechwood	Sigma Aldrich, Poole, UK	100-200
	Xylan from Oatspelts	Sigma Aldrich, Poole, UK	100-200
	Cellulose	Sigma Aldrich, Poole, UK	300-1700
	Methylcellulose	Sigma Aldrich, Poole, UK	-
Polyol	Mannitol	Sigma Aldrich, Poole, UK	n/a
	Sorbitol	Sigma Aldrich, Poole, UK	n/a
	Xylitol	Sigma Aldrich, Poole, UK	n/a
Algal source	Green Powder	Algae derived powder	Unknown
	Red Powder	Algae derived powder	Unknown
Unknown	Esculin	Sigma Aldrich, Poole, UK	-
	Sialyllactose	Friesland Foods, Zwolle, Netherlands	-

(?), Unknown carbohydrate type; (-), unknown degree of polymerisation; n/a, monosaccharides

Table S2.2 Fermentation profiles for nine *Lactobacillus ruminis* strains

		Lactobacillus ruminis strains								
		Human strains						Bovine strains		
Carbohydrate type	**Carbohydrate**	**L5**	**S21**	**S23**	**S36**	**S38**	**ATCC 25644**	**ATCC 27780T**	**ATCC 27781**	**ATCC 27782**
Monosaccharides	D-Arabinose	-	-	-	-	-	-	-	-	-
	L-Arabinose	-	-	-	+	+	-	-	-	-
	Fructose	+++	+++	+++	+++	++	+++	++	++	-
	Galactose	++	++	+++	+++	++	++	++	++	++
	Glucose	++	++	+++	+++	++	+++	++	++	++
	Lyxose	-	-	-	-	-	-	-	-	-
	Mannose	++	++	+++	+++	+++	++	+++	+++	++
	Melibiose	+	+	++	++	+	++	++	++	++
	Ribose	-	-	-	-	-	-	-	-	-
	Xylose	-	-	-	-	-	-	-	-	-
Disaccharides	Cellobiose	++	+	++	++	++	++	++	++	++
	Lactose	++	++	++++	+++	+++	++++	+++	+++	-
	Lactulose	++	++	+++	++	++	+++	+++	++	-
	Maltose	++	++	++	++	++	++	++	++	++
	Sucrose	++	++	+++	+++	++	+++	+++	++	++
	Trehalose	-	-	-	-	-	-	-	-	ND
Trisaccharides	Raffinose	++	++	+++	+++	+++	+++	+++	+++	++
	Melezitose	-	-	-	-	-	-	-	-	ND
Tetrasaccharide	Stachyose	+	++	++	+++	++	++	++	++	++
Oligosaccharides	Beneo P95	++	++	+++	+++	++	+++	+	++	-
	B-Glucotriose (B)[a]	++	++	++	++	++	++	++	++	++
	Raftilose P95	++	++	++++	+++	+++	+++	+	++	-
	Raftilose Synergy 1	++	++	+++	++	++	+++	-	+	-
	GOS	++	++	+++	++	++	++	++	++	ND
	GOS Inulin	+	+	++	++	++	++	++	++	-
	Palatinose	-	-	-	+	-	-	-	-	ND
	Polydextrose	+	+	-	-	-	-	-	-	-
Polyols	Mannitol	-	-	-	-	-	-	-	-	ND
	Sorbitol	-	-	-	-	-	-	-	-	ND
	Xylitol	-	-	-	-	-	-	-	-	-
Polysaccharides	Sialic acid	-	-	-	-	-	-	-	-	ND
	Siallylactose	-	-	-	-	-	-	-	-	-
	Soluble Starch	-	-	-	-	-	-	-	-	ND
	Xylan Beechwood	-	-	-	+	-	-	-	-	ND
	Xylan Oatspelts	-	-	-	-	-	-	-	-	ND
	Cellulose	-	-	-	-	-	-	-	-	-
	B- Glucan	-	-	-	-	-	-	-	-	ND
	Dextran	-	-	-	-	-	-	-	-	ND
	Dextrin	-	-	-	-	-	-	-	-	ND
	Esculin	-	-	-	-	-	-	-	-	-
	Beneo HP	-	-	-	-	-	-	-	-	ND
	Lichenan	-	-	-	-	-	-	-	-	ND
	Maltodextrin	-	-	-	-	-	-	+	-	ND
	Mannan	-	-	-	-	-	-	-	-	ND
	Methylcellulose	-	-	-	-	-	-	-	-	-
	Raftiline HP	-	-	-	-	-	-	-	-	-
	Raftiline HPX	-	-	-	-	-	-	-	-	ND
	Raftiline ST	++	-	++	++	++	++	-	-	-
Algal sources	Red Powder	+	+	+	+	+	+	+	+	-
	Green Powder	-	-	-	-	-	-	-	-	ND

(-) no growth (OD ≤ 0.1); (+) weak growth (OD 0.1 – 023); (++) moderate growth (OD 0.2 – 0.5); (+++) strong growth (OD 0.5 – 0.8); (++++) very strong growth (OD 0.8 – 1.0); (ND) not determined. a: β-glucan hydrolysate.

Chapter III

The core faecal bacterial microbiome of Irish Thoroughbred racehorses

This chapter has been published in full as a research article in

O' Donnell, M. M., Harris, H., Jeffery, I. B., Claesson, M. J., Younge, B., O' Toole, P. W. & Ross, R. P. (2013). The core faecal bacterial microbiome of Irish Thoroughbred racehorses. *Letters in Applied Microbiology*.

Note:

Sample collection, pyrosequencing PCR and purification andpyrosequencing data analysis were carried out by M.M. O' Donnell (author of this thesis)

Qiime analysis, statistical analysis and species level assignments was carried out by H. Harris

Clostridium clusters data was generated by M.J. Claesson

Access to the animals and details about their health was provided by B. Younge

Chapter III

Table of contents

Abstract

In this study, we characterised the gut microbiota in six healthy Irish thoroughbred racehorses and showed it to be dominated by the phyla *Firmicutes*, *Bacteroidetes*, *Proteobacteria*, *Verrucomicrobia*, *Actinobacteria*, *Euryarchaeota*, *Fibrobacteres* and *Spirochaetes*. Moreover, all the horses harboured *Clostridium*, *Fibrobacter*, *Faecalibacterium*, *Ruminococcus*, *Eubacterium*, *Oscillospira*, *Blautia Anaerotruncus*, *Coprococcus*, *Treponema*, and *Lactobacillus* spp. Notwithstanding the sample size, it was noteworthy that the core microbiota species assignments identified *Fibrobacter succinogenes*, *Eubacterium coprostanoligenes*, *Eubacterium hallii*, *Eubacterium ruminantium*, *Oscillospira guillermondii*, *Sporobacter termiditis*, *Lactobacillus equicursoris*, *Treponema parvum* and *Treponema porcinum* in all the horses. This is the first study of the faecal microbiota in the Irish Thoroughbred racehorse, a significant competitor in the global bloodstock industry. The information gathered in this pilot study provides a foundation for veterinarians and other equine health associated professionals to begin to analyse the microbiome of performance racehorses. This study and subsequent work may lead to alternate dietary approaches aimed at minimizing the risk of microbiota-related dysbiosis in these performance animals.

Significance and Impact of the Study. Although Irish Thoroughbreds are used nationally and internationally as performance animals very little is known about the core faecal microbiota of these animals. This is the first study to characterise the bacterial microbiota present in the Irish Thoroughbred racehorse faeces and elucidate a core microbiome irrespective of diet, animal management and geographic location.

3.1 Introduction

The horse is a member of the family *Equidae* and is "a mono-gastric" or non-ruminant herbivore whose physiology is suited to digesting and utilising high fibre diets as a result of continual microbial fermentation within the hindgut. Ireland is now the third largest producer of Thoroughbreds in the world after the USA and Australia, with approximately 40% of European Thoroughbreds originating from Ireland (Leadon & Herholz, 2009) and the equine sector is worth an estimated €100 billion a year to the European economy.

Until recently the equine hindgut microbiota had remained relatively poorly characterised. Previous studies have used culture and molecular methods to identify the bacterial genera present in the equine gastrointestinal microbiota affected by laminitis and colic (Milinovich *et al.*, 2006; Pollitt, 2004; Respondek *et al.*, 2008; Shirazi-Beechey, 2008). Recent studies have used next generation sequencing to investigate the faecal microbiota of two Arabian Geldings (Shepherd *et al.*, 2012). Comparison of the microbiota of healthy and unhealthy horses suffering from colitis (Costa *et al.*, 2012) revealed a shift in the predominant phyla. The *Firmicutes* phylum predominated in healthy horses while in colitis-affected horses, *Bacteroidetes* predominated. A similar investigation comparing the microbiota of healthy horses and those with laminitis revealed an increase in the *Verrucomicrobia* phylum for those horses with the disease (Steelman *et al.*, 2012).

The link between altered gastrointestinal microbiota and disease risk is becoming a well-established concept in both humans and animals (Yatsunenko *et al.*, 2012). Identification of the core microbiota present in the faeces of horses would allow for a better understanding of the dietary requirements needed to prevent or to inhibit microbiota-related diseases and to promote gut health. Fructans and starches are present at varying levels in grasses depending on the growing season and the cultivar (Hoffman *et al.*, 2001; Superchi *et al.*, 2010) and thus can have seasonal effects on the composition of the microbiota of the grazing horse. Knowledge of the effect that different grasses, types of forage, concentrates and supplements can have on the horse microbiota is therefore very valuable especially to the bloodstock industry.

The objective of this study was to characterise the microbiota of Irish thoroughbred horses fed various commonly consumed diets to elucidate the core microbiome of the

Irish Thoroughbred racehorse independent of diet, management regime, geographic location or age.

3.2 Materials and Methods

3.2.1 Animals and diets

Faecal samples were collected from six mature Irish Thoroughbreds horses that were housed in two stables; horse weights, ages and genders are shown in Table S3.1. All faecal sample collection and analysis was consistent with the current animal welfare legislation in Ireland. The horses were each assigned the abbreviation TCM (Thoroughbred core microbiome) and a numbered from one to six. Faecal samples were collected and all the faecal samples were held anaerobically at 4°C prior to DNA extraction within 24 hours. Grass and haylage were chosen as diets to represent racehorses at rest; while haylage supplemented with starch concentrate represents those performance horses in active training. Each horse had been receiving their respective feed for a month.

3.2.2 DNA extraction, PCR amplification and 454 pyrosequencing

Total genomic faecal DNA was isolated from the six faecal samples using the Isolate faecal DNA kit (myBio, Ireland). The V4 region PCR reaction conditions were outlined previously by Claesson *et al*., 2009 (Claesson *et al.*, 2009). Table S3.2 contains a full list of the primers used in the study. PCR products were purified and quantified using the Agencourt AMPure XP PCR (Beckman Coulter, High Wycombe, UK) purification beads and the Quant-It Picogreen dsDNA kit (Invitrogen, Amhersham, US), respectively. The 16S rRNA V4 amplicons were sequenced on a 454 Genome Sequencer FLX platform (Teagasc Food Research Centre, Moorepark).

3.2.3 DNA Sequence processing and statistical analysis

Raw sequencing reads were quality trimmed using a locally installed version of the RDP Pyrosequencing Pipeline (Claesson *et al.*, 2009). The following analysis of the pyrosequencing data was performed in Qiime . All of the sequences from the six samples were clustered into OTUs (operational taxonomic units) of 97% sequence

identity using uclust . The representative sequences for each OTU were aligned using PyNAST , using the best match from the Greengenes (DeSantis *et al.*, 2006) core set (http://greengenes.lbl.gov/). Taxonomy was assigned to the unaligned representative set using the RDP classifier (Cole *et al.*, 2005) with a minimum confidence value of 0.8. Chimeras were identified in the aligned representative set using ChimeraSlayer (Haas *et al.*, 2011) and the same core set of Greengenes aligned sequences used to align the representative set. A phylogenetic tree was constructed from the aligned, filtered representative set using Fasttree (Price *et al.*, 2009). Before rarefaction, the OTU table was filtered for OTUs represented by a single read in a single sample. If an OTU represented by a single read was identified in more than 1 sample it was included in the study. The OTU Table was rarefied to account for variations in sequencing depth among the samples and a subsample of 17,000 sequences was taken from each sample. Weighted and un-weighted Uni-frac (Lozupone & Knight, 2005) distance matrices were constructed from the rarefied OTU Table. Single rarefaction was carried out on the OTU table and the rarefied samples from this table were subjected to Unifrac, principle coordinates analysis and statistics. Multiple rarefaction was used on the OTU table to generate the rarefaction plots. 2D and 3D PCoA plots were constructed from the weighted and un-weighted distance matrices. The 2D plots were generated in R (version 2.13.1) from collated alpha diversity values imported from Qiime (Caporaso *et al.*, 2010). Due to the small sample sizes, statistical analysis of the data was carried out using Fisher's exact test (Clayton *et al.*, 2012; Hynes *et al.*, 2002; Ruijter *et al.*, 2002). The method used to assign reads to *Clostridium* clusters is outlined in (Claesson *et al.*, 2011).

For species level assignments, all the sequences in the RDP database were blasted against themselves in an all-against-all blast (Altschul *et al.*, 1990). Since multiple strains of the same species are present in the database, the blast score varied slightly for the multiple within-species blast alignments. Any sequence from our analysis blasted against the RDP database that had a score ≥ the lowest within-species blast score was assigned to that species as a "strict" species assignment (Jeffery *et al.*, 2012). If the blast score was lower than the lowest within-species blast score but higher than the next highest blast score to another species, it was assigned to the species as a "relaxed" species assignment. Otherwise, no species was assigned to the read. A core genus or species was assigned if it was present in the microbiota of ≥4

horses at 0.1% of the total read assignments. Relaxed species assignments were primarily used in this study.

3.2.4 Alpha and Beta Diversity metrics

Four alpha diversity metrics were calculated to measure the microbial diversity in each of the six horses. These metrics are Observed OTUs, Phylogenetic Diversity, the Shannon index ($H' = -\sum p_i .\log p_{i)}$ and the Species or Pielou's Evenness ($E = H'/H'_{max)}$. Each metric was calculated from a rarefied OTU table consisting of sub-samples of 17,000 reads per sample. The last index used was Phylogenetic Diversity using a phylogenetic tree created from all the reads in the six samples. The phylogenetic diversity for any one sample was then the sum of the branch lengths that lead to every read in the tree that belongs to that sample. Rarefaction curves for each sample was based upon the calculated alpha diversity metric for sub-samples ranging from 100 to 17,000 reads at increments of 100 reads.

Beta diversity was calculated using weighted and un-weighted Unifrac distance in Qiime and displayed graphically using principle coordinates analysis in R. Unifrac distance is calculated by constructing a phylogenetic tree from all the OTUs and, for each pair of samples, calculating a distance measure using the equation (sum of unshared branch lengths) / (sum of total branch lengths).

3.3 Results and Discussion

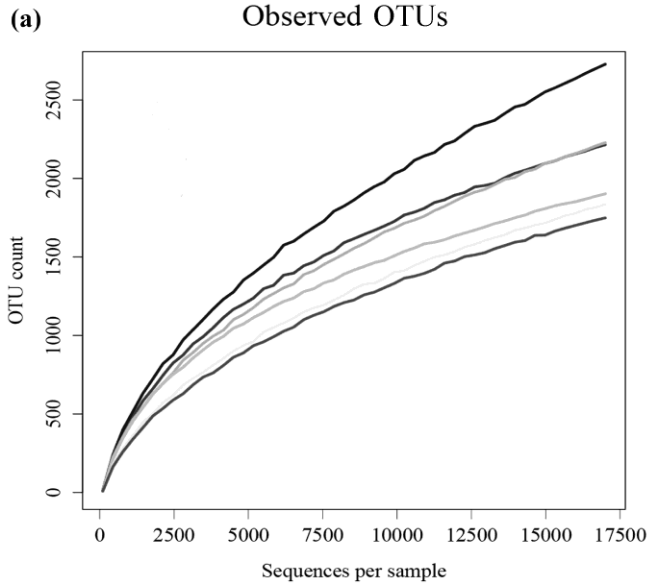
Horses feed naturally by what is termed "trickle feeding" (Hill, 2002); however modern practices have necessarily altered this feeding pattern. Thoroughbreds and other performance racehorses are often fed a high energy, carbohydrate-enriched feed twice a day. The microbiological impact of this alteration to the natural grazing-based feeding pattern of the horse has yet to be fully elucidated. Starch concentrate was chosen as a representative of "high sugar" feeds which are often detrimental to equine health. Feeding excessive carbohydrates to horses in the form of either starch or fructooligosaccharides may result in laminitis (Milinovich *et al.*, 2010).

We applied 16S rRNA gene (V4 region) amplicon pyrosequencing to determine the faecal microbiota composition in six Thoroughbred racehorses. Following the removal of low quality reads, a total of 178,975 sequences were obtained from the six samples. Read numbers ranged from 17,757 to 38,378 (SD = 8,002; Table S3.3). The

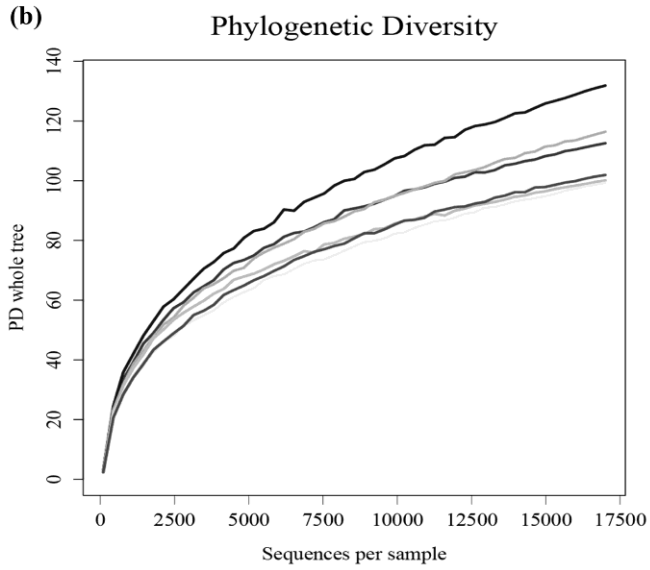
average read length following quality trimming was 224.8bp (SD = 3.23). A total of 19 phyla, 229 genera and 143 bacterial species were identified across the six horses. At the phylum level an average of 93% of the reads from the trial animals were classified as bacterial phyla with 6% of the reads remaining unclassified and <0.3% Archea. An average 43% of the reads identified were assigned to bacteria at the genus level while 57% (average) remain unclassified and a small proportion was assigned as Archaea (less than 0.3% on average). The high level of unclassified read assignments may suggest that the equine faecal samples contain many genera that are distinct and novel from those isolated from other mammals and the wider environment. The Archaea present in all horses were identified as *Methanobrevibacter* and *Methanocorpusculum. Methanobrevibacter woesei* was identified at the species level in all of the horses. However, it must be noted that a previous study has shown that the V4 region of the 16S rRNA gene may underestimate the true population levels of Archaea present in a faecal sample (Yu *et al.*, 2008).

Four different measures of alpha diversity (microbiota diversity within a subject) were calculated to assess the diversity of faecal microbiota in the 6 racehorses (Table S3.3). In each metric, the average diversity of samples TCM 1-2 is the lowest, followed by samples TCM 3-4, with samples TCM 5-6 having the highest alpha diversity in all four metrics. This may be due to the different feeding regimes and diets however, a larger sample size would be needed for a conclusive analysis. Rarefaction curves were also generated for three of the alpha diversity metrics: observed species, phylogenetic diversity and the Shannon index (Fig. 3.1). The observed OTU's and the phylogenetic diversity metric curves have not reached a plateau at 17,000 reads which suggests that the equine faecal microbiota is more diverse than that measured in this pilot study. Curves for the Shannon index plateaux at relatively low read numbers. However, the saturation of microbial diversity at these read numbers is unlikely since the addition of low-abundance OTUs has a minor effect on the value of the Shannon index.

(a)


Observed OTUs
OTU count
0
500
1000
1500
2000
2500
0
2500
5000
7500
10000
12500
15000
17500
Sequences per sample

(b)


Phylogenetic Diversity
PD whole tree
0
20
40
60
80
100
120
140
0
2500
5000
7500
10000
12500
15000
17500
Sequences per sample

(c)

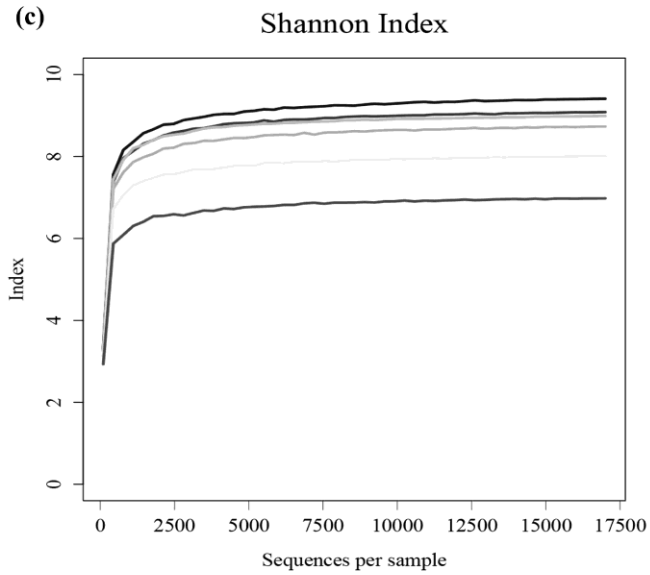

Shannon Index
Index
0
2
4
6
8
10
0
2500
5000
7500
10000
12500
15000
17500
Sequences per sample

Fig. 3.1. (a) Observed species (OTUs); (b) Phylogenetic diversity; (c) Shannon index identified from the faecal samples of each horse used in this study. Each of the plots was generated by multiple rarefaction where sub-samples of different depths (read number) were taken from each sample in increments of 100 reads (x-axis) and the appropriate diversity metric at each sub-sampling was calculated (y-axis). The colour scheme is the same for all three plots. ( ) TCM 1**,** ( ) TCM 2, ( ) TCM 3, ( ) TCM 4, ( ) TCM 5, ( ) TCM 6.

The beta diversity (i.e. between animals) of the six faecal microbiota samples was measured by generating PCoA plots based on the rarefied OTU table (Fig. S3.1). Figure S3.1 (a) shows an un-weighted PCoA plot of the six samples, coloured by feed received. The first two principle axes, which explain 50% of the variation in the samples, show a grouping of the samples according to diet group. This suggests that the two samples from each diet group are more similar to each other in terms of the presence/absence of microbial taxa than they are to the samples from the other diet groups. Un-weighted PCoA plots therefore may be affected by the inclusion of low abundance reads however, it is difficult to quantify the severity of this effect. While the weighted PCoA plot includes proportional data and therefore the inclusion of low abundance reads will probably have a negligible effect. Figure S3.1 (b) shows a weighted PCoA plot of the six samples, coloured by feed received. The first two principle axes, which explain 68% of the variation in the samples, do not group the samples according to diet. When relative abundance of taxa is taken into account, TCM 4 is grouped closer to the grass-fed samples and TCM 3 and TCM 2 are grouped together, while TCM 1 lies a considerable distance away from both groups. A possible reason for this "outlier" status may be due to the significantly younger age of the animal TCM 1 but it is more likely due to this horse only being housed at the sample collection stable for approximately a month. The previous feeding regime and management style experienced by this horse may have greatly influenced its microbiota and thus our beta diversity indices. This outlier status can also be seen in the large number of reads assigned to the *Streptococcus* genus in this horse compared to the other animals in the study.

We measured greater phylotype diversity in the equine faecal microbiota compared to data from the distal bowel microbiota of other animals (Pitta *et al.*, 2010)

(Lamendella *et al.*, 2011). Our phylotype number estimations for the equine faecal microbiota (1,755 - 2,736) are higher than those estimated for the human microbiota (Claesson *et al.*, 2009; Nam *et al.*, 2011) and other horses (Shepherd *et al.*, 2012; Steelman *et al.*, 2012). However, this difference in observed phylotypes might be influenced by the metric used to generate the phylotype numbers (Kemp & Aller, 2004). Additionally, we opted to not to use a de-noise step in the Qiime pipeline and this too can have an influence on the alpha diversity matrices (Reeder & Knight, 2010). The diversity indices indicated that consumption of the starch concentrate in conjunction with haylage reduced the faecal microbiota diversity, where the forage fed horses harboured the most diverse microbiota. However, further study is needed to confirm this trend.

The relative phylum abundance in the faecal microbiota of the six racehorses is shown Fig. 3.2. A total of 19 phyla were identified, twelve of which were present in all horses. In addition to the phyla shown in Fig. 3.2, these 12 phyla include [*Chlamydiae*, *Chloroflexi*, *Deferribacteres*, *Cyanobacteria* and *Synergistetes* present at low abundance levels]. Phyla *Firmicutes* and *Bacteroidetes* were dominant in all the horses, with all microbiota displaying a *Firmicutes* to *Bacteroidetes* ratio of greater than 2:1. Collectively these two phyla accounted for 73-85% (SD = 5.7%) of the sequences. Although the current analysis of the Thoroughbred microbiota identified 19 phyla, only five were present in all horses above a 0.5% cut-off. The dominance of *Firmicutes* and *Bacterioidetes* phyla in the faecal microbiome is similar to that measured in humans and cows (Jami & Mizrahi, 2012; van den Bogert *et al.*, 2011). The *Firmicutes* range (47-74%) is consistent with other equine studies, which attributed from 15-83% (Costa *et al.*, 2012; Daly *et al.*, 2001; Perkins *et al.*, 2012; Shepherd *et al.*, 2012; Steelman *et al.*, 2012; Willing *et al.*, 2009b) of the total reads to the *Firmicutes* phylum. The relative abundance of *Bacteroidetes* (3.65-9.94%) identified by other equine microbiota studies (Shepherd *et al.*, 2012; Steelman *et al.*, 2012) is far lower than the average relative abundance identified by this study but is similar to the levels identified by Willing *et al*. (Willing *et al.*, 2009b). The relative abundances of the *Proteobacteria*, *Verrucomicrobia*, *Spirochaetes*, *Tenericutes* and *Fibrobacteres* phyla identified in this study were statistically significantly higher ($P<0.001$) in the horses fed the forage based diets. To our knowledge, this is also the first time the *Euryarchaeota* phylum has been identified in horses, albeit at low levels; however, it is as of yet unknown what function members of this phylum have

in the microbiome of horses. However, as stated earlier this may be an underestimation of the true extent of the presence of the Archaea in the horse faecal samples due to the use of the V4 region primer pair.

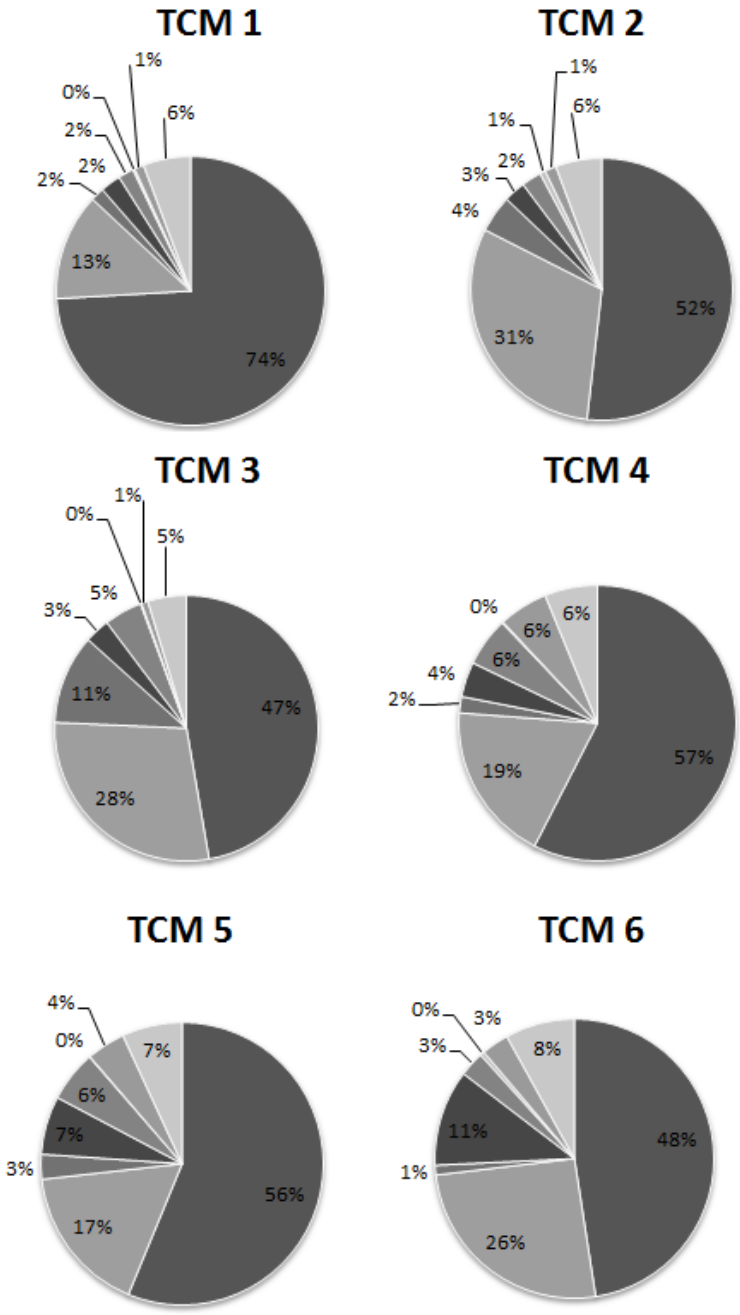


Fig. 3.2. Phylum-level assignment of V4 16S rRNA sequences from individual horses, according to the RDP classifier (CI ≥ 97%). Reading clockwise: ( ) *Firmicutes*, ( ) *Bacteroidetes*, ( ) *Proteobacteria*, ( ) *Verrucomicrobia*, ( ) *Spirochaetes*, ( ) *Euryarchaeota*, ( ) Other, ( ) Unclassified.

Consideration of the assignment of sequences to phylogenetic orders within the *Firmicutes* phylum (Fig. 3.2) revealed that the increased *Firmicutes* abundance in the faeces of samples TCM 1 & 2 was due to increased abundance of the *Lactobacillales*, especially in sample TCM 1. At the order level the microbiota of the horses was dominated by *Lactobacillales*, *Clostridia* and *Erysipelotrichi.* In our study, *Streptococcus* and *Lactobacillus* were identified at significantly elevated levels ($P<0.001$) in the horses from the Limerick stable, samples TCM 1-2. However, at the species level, only *Streptococcus caballi* (Milinovich *et al.*, 2008a), was identified at less than 0.1% of the total reads from samples TCM 1-2 only. This loss of resolution at the species level is probably firstly due to the fact that some sequences are too short to accurately identify to species level. Secondly, the limited size of the RDP database (i.e. more sequences would lead to a greater representation of bacterial diversity and more sequences would be assigned to species level).

Two hundred and twenty-nine genera were identified across the six horse samples, Table 3.2 lists the genus level diversity of the faecal microbiota between the horses used in this study; 93 were found in ≥4 of the datasets and 64 of those were present in all the horses. We can thus consider these genera as being part of the core faecal microbiota of Thoroughbreds. This means that approximately 41% of the genera identified were consistently found in the majority of the horses sampled and 28% of the genera are present at varying levels in all the horses. The genera most commonly found at relatively high levels (≥0.2 % of reads) in the majority of the samples include *Prevotella, Fibrobacter, Clostridium, Ruminococcus, Sporobacter, Acinetobacter* and *Trepomena*. Further scrutiny of the genus data revealed 34 genera that were present in the faecal microbiota of ≥4 racehorses, the identities of which are listed in Table 3.1. Therefore in this study 15% of the faecal microbiota of the majority of horses tested was consistently found irrespective of feed or geographic locale.

Table 3.1 The core genera[a] and relative abundance identified in the hindgut microbiota of Irish Thoroughbred racehorses

	% of total reads per animal						
Genus	TCM 1	TCM 2	TCM 3	TCM 4	TCM 5	TCM 6	Order › Family
Methanocorpusculum	0.0%	0.6%	0.1%	0.1%	0.1%	0.5%	*Methanomicrobiales › Methanocorpusculaceae*
Anaerophaga	0.4%	0.3%	1.2%	0.8%	0.2%	1.3%	*Bacteroidales › Marinilabiaceae*
Paludibacter	0.3%	0.4%	0.2%	0.5%	1.5%	1.5%	*Bacteroidales › Porphyromonadaceae*
Paraprevotella	0.1%	0.1%	0.9%	0.3%	0.2%	0.3%	*Bacteroidales › Prevotellaceae*
Prevotella	0.3%	1.1%	1.4%	0.5%	0.3%	1.2%	*Bacteroidales › Prevotellaceae*
Galbibacter	0.3%	0.2%	0.3%	0.6%	0.1%	0.3%	*Flavobacteriales › Flavobacteriaceae*
Fibrobacter	0.3%	0.3%	0.5%	5.2%	3.7%	2.3%	*Fibrobacterales › Fibrobacteraceae*
Anaerosporobacter	0.1%	0.1%	0.2%	0.3%	0.3%	0.2%	*Clostridiales › Clostridiaceae*
Clostridium	0.6%	0.7%	0.4%	0.1%	1.3%	1.0%	*Clostridiales › Clostridiaceae*
Lactonifactor	0.2%	0.4%	0.1%	0.9%	0.4%	0.2%	*Clostridiales › Clostridiaceae*
Eubacterium	0.0%	0.5%	0.0%	0.1%	0.5%	0.1%	*Clostridiales › Eubacteriaceae*
Acetitomaculum	0.1%	0.5%	0.0%	0.5%	1.3%	0.4%	*Clostridiales › Lachnospiraceae*
Blautia	0.3%	0.8%	0.1%	1.9%	1.8%	1.1%	*Clostridiales › Lachnospiraceae*
Coprococcus	0.5%	0.5%	0.6%	1.3%	1.4%	0.7%	*Clostridiales › Lachnospiraceae*
Dorea	0.1%	0.3%	0.1%	0.3%	0.4%	0.4%	*Clostridiales › Lachnospiraceae*
Pseudobutyrivibrio	0.1%	0.3%	0.2%	1.0%	0.5%	0.5%	*Clostridiales › Lachnospiraceae*
Robinsoniella	0.1%	0.0%	0.1%	0.5%	0.2%	0.5%	*Clostridiales › Lachnospiraceae*
Roseburia	0.1%	0.2%	0.1%	0.5%	0.8%	0.5%	*Clostridiales › Lachnospiraceae*
Oscillibacter	1.7%	5.3%	2.4%	0.6%	1.0%	2.0%	*Clostridiales › Oscillospiracea*
Acetivibrio	0.4%	1.1%	0.7%	1.2%	1.0%	1.0%	*Clostridiales › Ruminococcaceae*
Anaerotruncus	0.0%	2.2%	1.2%	0.4%	0.2%	0.5%	*Clostridiales › Ruminococcaceae*
Faecalibacterium	0.2%	0.1%	0.2%	0.5%	0.4%	1.1%	*Clostridiales › Ruminococcaceae*
Papillibacter	0.3%	0.4%	0.9%	0.3%	0.3%	1.0%	*Clostridiales › Ruminococcaceae*
Ruminococcus	0.7%	0.9%	1.3%	3.2%	2.3%	1.2%	*Clostridiales › Ruminococcaceae*
Sporobacter	1.3%	6.3%	3.5%	3.6%	3.1%	3.3%	*Clostridiales › Ruminococcaceae*
Holdemania	0.1%	0.1%	0.3%	0.2%	0.2%	0.3%	*Erysipelotrichales › Erysipelotrichaceae*
Lactobacillus	2.6%	0.8%	0.0%	0.4%	0.4%	0.0%	*Lactobacillales › Lactobacillaceae*
Acidaminococcus	0.2%	0.3%	0.7%	0.2%	0.2%	0.4%	*Selenomonadales › Acidaminococcaceae*
Phascolarctobacterium	0.1%	0.1%	0.1%	0.2%	0.1%	0.0%	*Selenomonadales › Acidaminococcaceae*
Acinetobacter	0.2%	0.0%	9.1%	0.4%	1.0%	0.0%	*Pseudomonadales › Moraxellaceae*
Treponema	1.8%	2.3%	4.7%	5.7%	5.8%	2.9%	*Spirochaetales › Spirochaetaceae*
Anaeroplasma	0.0%	0.1%	0.0%	0.2%	0.2%	0.5%	*Anaeroplasmatales › Anaeroplasmataceae*
Akkermansia	0.2%	0.2%	0.0%	0.1%	0.1%	0.0%	*Verrucomicrobiales › Verrucomicrobiaceae*
Subdivision5_*incertae_sedis*	2.1%	2.4%	3.0%	3.9%	6.4%	10.8%	-

[a]Core as defined by presence in the microbiota of 4 or more of the 6 racehorses at ≥0.1% of the total reads

Table 3.2 Genus level diversity of the faecal microbiota between the horses used in this study.

Genus	% of total reads per animal					
	TCM 1	TCM 2	TCM 3	TCM 4	TCM 5	TCM 6
Streptococcus	26.85%	6.46%	0.02%	0.15%	0.00%	0.00%
Subdivision5_incertae_sedis	2.13%	2.40%	2.99%	3.91%	6.39%	10.83%
Treponema	1.82%	2.28%	4.65%	5.72%	5.84%	2.86%
Sporobacter	1.31%	6.31%	3.52%	3.64%	3.12%	3.28%
Oscillibacter	1.69%	5.26%	2.43%	0.57%	1.02%	1.99%
Acinetobacter	0.19%	0.01%	9.13%	0.35%	0.99%	0.00%
Fibrobacter	0.30%	0.28%	0.45%	5.21%	3.69%	2.31%
Ruminococcus	0.70%	0.91%	1.34%	3.20%	2.30%	1.22%
Allobaculum	7.93%	0.02%	0.01%	0.08%	0.01%	0.00%
Lysinibacillus	0.04%	0.00%	4.62%	0.09%	2.05%	0.00%
Blautia	0.28%	0.80%	0.09%	1.85%	1.77%	1.10%
Acetivibrio	0.40%	1.09%	0.68%	1.17%	1.03%	1.03%
Prevotella	0.34%	1.11%	1.36%	0.52%	0.33%	1.21%
Coprococcus	0.46%	0.46%	0.62%	1.26%	1.39%	0.66%
Clostridium	0.60%	0.66%	0.42%	0.10%	1.32%	0.99%
Anaerophaga	0.44%	0.26%	1.19%	0.78%	0.18%	1.28%
Anaerotruncus	0.02%	2.17%	1.21%	0.39%	0.17%	0.51%
Lactobacillus	2.58%	0.76%	0.03%	0.40%	0.38%	0.02%
Other	6.08%	7.41%	8.12%	10.02%	11.31%	10.83%
Unclassified	45.86%	61.37%	57.12%	60.60%	56.71%	59.88%

Thirty-five species were present as core microbiota in four or more racehorses present at ≥ 0.1% which are listed in Table S3.5. Of these 35 species, 19 belonged to the *Clostridiales* order. The majority of reads were assigned to the *Eubacteriaceae* and *Ruminococcaceae* families. However, this core species is calculated from the reads assigned to the species level and not the total reads for each horse. There are 13 species of bacteria present as a core microbiome when we calculated the species of bacteria present at ≥0.1% of the total read assignments for each horse. The majority of reads are assigned to the *Clostridiales* and *Spirochaetales* order. The read assignments for the total reads assigned can be seen in Table 3.3. *Sporobacter*

termiditis was the most abundant species identified (2.9% average aggregate proportion across the six samples. When the species level assignments generated from the total reads are compared to the genera level assignments *Sporobacter termiditis* accounts for between 52-100% of those reads assigned to the *Sporobacter* genus. The cellulolytic species *Fibrobacter succinogenes* was the second most abundant species identified in this study and accounted for 88-100% of the reads assigned to the Fibrobacter genus. *Lactobacillus equicursoris,* a predominant equine lactobacillus, accounts for between 13-94% of the reads assigned to the *Lactobacillus* genus. *Lactobacillus equicursoris* was present in all samples, but statistically higher levels were present in samples TCM 1-2. On average the percentage of the total reads in the study we could identify to the species level in each horse was approximately 13%, which is in line with previous studies from our lab (Claesson *et al.*, 2009).

Table 3.3 The 13 species that form the core microbiome accounting for ≥0.1% of the total reads for 4 or more animals used in this study.

	% of the total reads per animal						
Species	**TCM 1**	**TCM 2**	**TCM 3**	**TCM 4**	**TCM 5**	**TCM 6**	**Order › Family › Genus**
Paludibacter propionicigenes	0.2%	0.1%	0.0%	0.1%	0.9%	0.4%	*Bacteroidales › Porphyromonadaceae › Paludibacter*
Fibrobacter succinogenes	0.3%	0.3%	0.4%	5.2%	3.7%	2.3%	*Fibrobacterales › Fibrobacteraceae › Fibrobacter*
Eubacterium coprostanoligenes	0.2%	0.1%	0.3%	0.3%	0.3%	0.1%	*Clostridiales › Eubacteriaceae › Eubacterium*
Eubacterium hallii	0.2%	0.6%	0.0%	0.9%	1.2%	0.3%	*Clostridiales › Eubacteriaceae › Eubacterium*
Eubacterium ruminantium	0.2%	0.2%	0.2%	0.8%	0.3%	0.1%	*Clostridiales › Eubacteriaceae › Eubacterium*
Oscillospira guilliermondii	0.8%	3.9%	1.4%	0.3%	1.0%	1.1%	*Clostridiales › Ruminococcaceae › Oscillospira*
Sporobacter termitidis	1.3%	5.6%	3.2%	2.2%	1.6%	3.4%	*Clostridiales › Ruminococcaceae › Sporobacter*
Lactobacillus equicursoris	1.8%	0.5%	0.0%	0.4%	0.1%	0.0%	*Lactobacillales › Lactobacillaceae › Lactobacillus*
Phascolarctobacterium faecium	0.1%	0.0%	0.1%	0.1%	0.1%	0.0%	*Selenomonadales › Acidaminococcaceae › Phascolarctobacterium*
Treponema brennaborense	0.1%	0.0%	0.0%	0.2%	0.1%	0.1%	*Spirochaetales › Spirochaetaceae › Treponema*
Treponema parvum	0.1%	0.1%	0.2%	0.4%	0.6%	0.3%	*Spirochaetales › Spirochaetaceae › Treponema*
Treponema porcinum	0.5%	0.6%	0.5%	1.8%	0.6%	0.7%	*Spirochaetales › Spirochaetaceae › Treponema*
Treponema saccharophilum	0.1%	0.1%	0.0%	0.0%	0.3%	0.3%	*Spirochaetales › Spirochaetaceae › Treponema*

Major *Clostridium* clusters in humans have been linked to changes in diet, short-chain fatty acid production, and anti-inflammatory effects (O'Toole & Claesson, 2010). *Clostridium* clusters IV and XIVa dominate in all horses in this study (Figure S3.3)

similar to studies on the human faecal microbiota (Claesson *et al.*, 2011). The microbiota of samples TCM 5-6 had a higher proportion of Cluster I (1.76% average) clostridia. All racehorses examined had a similar proportion of *Clostridium* Cluster III. Sixteen *Clostridium* species were identified across the six racehorse microbiota datasets, though the efficiency of assignment using this approach is not high (Claesson *et al.*, 2009; Claesson *et al.*, 2010a). *Clostridium butyricum, Cl. caenicola, Cl. hathewayi, Cl. hylemonae, Cl. lactatifermentans, Cl. leptum, Cl. methylpentosum* were present in the microbiota independent of diet.

We investigated the microbiota at a single time-point, from a single breed of horse housed in two stables close to Limerick City, Ireland. Seasonal and geographical influences on the forages consumed may also prevail; for example Yamano *et al*. monitored the faecal bacteria from two horse breeds, the Hokkaido native horse dominated by cellulolytic species and a light horse breed dominated by soluble sugar utilisers grazing on hilly winter woodland pasture (Yamano *et al.*, 2008). Costa et al. also noted that the two Thoroughbreds that were housed similarly on the same farm with the same feeding regime had a similar microbiota and that feeding regimes, location and other management factors may influence the microbiota (Costa *et al.*, 2012).

This study clearly outlines that the horse faecal microbiome is a diverse and practically unknown habitat and as such further large scale studies are required to identify "unclassified" genera and species present in the core microbiome. Although not the primary focus of this work we noted that the feed consumed by the horses did have an effect on the levels of certain genera in the faecal microbiome. As a multi-million euro industry and given the high monitory value of performance horses future work should also concentrate on identifying the effect that diet has on the microbiome. A practical future application of this study and corroborated by future work might be, for example, that horses transferred to starch concentrate feed for performance enhancement might be supplemented with a microbiota cocktail corresponding to that typical for forage-animals, or with forage extracts to maintain levels of associated genera. This might off-set or preclude the observed increases in *Streptococcus* or *Lactobacillus* abundance.

3.4 References

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

Caporaso, J. G., Bittinger, K., Bushman, F. D., DeSantis, T. Z., Andersen, G. L. & Knight, R. (2010a). PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26**, 266-267.

Caporaso, J. G., Kuczynski, J., Stombaugh, J. & other authors (2010b). QIIME allows analysis of high-throughput community sequencing data. *Nat methods* **7**, 335-336.

Claesson, M. J., O'Sullivan, O., Wang, Q., Nikkilä, J., Marchesi, J. R., Smidt, H., De Vos, W. M., Ross, R. P. & O'Toole, P. W. (2009). Comparative analysis of pyrosequencing and a phylogenetic microarray for exploring microbial community structures in the human distal intestine. *PLoS ONE* **4**, e6669.

Claesson, M. J., Wang, Q., O'Sullivan, O., Greene-Diniz, R., Cole, J. R., Ross, R. P. & O'Toole, P. W. (2010). Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Res* **38**, e200.

Claesson, M. J., Cusack, S., O'Sullivan, O. & other authors (2011). Composition, variability, and temporal stability of the intestinal microbiota of the elderly. *Proc Natl Acad Sci USA* **108**, 4586.

Clayton, E. M., Rea, M. C., Shanahan, F., Quigley, E. M. M., Kiely, B., Ross, R. P. & Hill, C. (2012). Carriage of *Clostridium difficile* in outpatients with irritable bowel syndrome. *J Med Microbiol* **61**, 1290-1294.

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

Costa, M. C., Arroyo, L. G., Allen-Vercoe, E., StÃ¤mpfli, H. R., Kim, P. T., Sturgeon, A. & Weese, J. S. (2012). Comparison of the Fecal Microbiota of Healthy Horses and Horses with Colitis by High Throughput Sequencing of the V3-V5 Region of the 16S rRNA Gene. *PLoS ONE* **7**, e41484.

Daly, K., Stewart, C. S., Flint, H. J. & Shirazi Beechey, S. P. (2001). Bacterial diversity within the equine large intestine as revealed by molecular analysis of cloned 16S rRNA genes. *FEMS Microbiol Ecol* **38**, 141-151.

DeSantis, T. Z., Hugenholtz, P., Larsen, N. & other authors (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72**, 5069-5072.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460-2461.

Haas, B. J., Gevers, D., Earl, A. M. & other authors (2011). Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res* **21**, 494-504.

Hill, J. (2002). Effect of level of inclusion and method of presentation of a single distillery by-product on the processes of ingestion of concentrate feeds by horses. *Livest Prod Sci* **75**, 209-218.

Hoffman, R., Wilson, J., Kronfeld, D., Cooper, W., Lawrence, L., Sklan, D. & Harris, P. (2001). Hydrolyzable carbohydrates in pasture, hay, and horse feeds: direct assay and seasonal variation. *J Anim Sci* **79**, 500.

Hynes, S. O., Broutet, N., Eurohepygast Study Group & other authors (2002). Phenotypic Variation of *Helicobacter pylori* Isolates from Geographically Distinct Regions Detected by Lectin Typing. *J Clin Microbiol* **40**, 227-232.

Jami, E. & Mizrahi, I. (2012). Composition and Similarity of Bovine Rumen Microbiota across Individual Animals. *PLoS ONE* **7**, e33306.

Jeffery, I. B., O'Toole, P. W., Öhman, L., Claesson, M. J., Deane, J., Quigley, E. M. M. & Simrén, M. (2012). An irritable bowel syndrome subtype defined by species-specific alterations in faecal microbiota. *Gut* **61**, 997-1006.

Kemp, P. F. & Aller, J. Y. (2004). Bacterial diversity in aquatic and other environments: what 16S rDNA libraries can tell us. *FEMS Microbiol Ecol* **47**, 161-177.

Lamendella, R., Santo Domingo, J., Ghosh, S., Martinson, J. & Oerther, D. (2011). Comparative fecal metagenomics unveils unique functional capacity of the swine gut. *BMC Microbiol* **11**, 103.

Leadon, D. & Herholz, C. (2009). Globalisation of trade and the spread of infectious disease. *Equine Vet Educ Man 8*.

Lozupone, C. & Knight, R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* **71**, 8228-8235.

Milinovich, G., Trott, D., Burrell, P., Van Eps, A., Thoefner, M., Blackall, L., Al Jassim, R., Morton, J. & Pollitt, C. (2006). Changes in equine hindgut bacterial populations during oligofructose induced laminitis. *Environ Microbiol* **8**, 885-898.

Milinovich, G. J., Burrell, P. C., Pollitt, C. C., Bouvet, A. & Trott, D. J. (2008). *Streptococcus henryi* sp. nov. and *Streptococcus caballi* sp. nov., isolated from the hindgut of horses with oligofructose-induced laminitis. *Int J Syst Evol Microbiol* **58**, 262-266.

Milinovich, G. J., Klieve, A. V., Pollitt, C. C. & Trott, D. J. (2010). Microbial events in the hindgut during carbohydrate-induced equine laminitis. *Vet Clin North Am-Equine* **26**, 79-94.

Nam, Y.-D., Jung, M.-J., Roh, S. W., Kim, M.-S. & Bae, J.-W. (2011). Comparative Analysis of Korean Human Gut Microbiota by Barcoded Pyrosequencing. *PLoS ONE* **6**, e22109.

O'Toole, P. W. & Claesson, M. J. (2010). Gut microbiota: Changes throughout the lifespan from infancy to elderly. *Int Dairy J* **20**, 281-291.

Perkins, G. A., den Bakker, H. C., Burton, A. J. & other authors (2012). The equine stomach harbors an abundant and diverse mucosal microbiota. *Appl Environ Microbiol* **78**, 2522-2532.

Pitta, D., Pinchak, W., Dowd, S. & other authors (2010). Rumen Bacterial Diversity Dynamics Associated with Changing from Bermudagrass Hay to Grazed Winter Wheat Diets. *Microbiol Ecol* **59**, 511-522.

Pollitt, C. C. (2004). Equine laminitis. *Clin Tech Equine Pract* **3**, 34-44.

Price, M. N., Dehal, P. S. & Arkin, A. P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* **26**, 1641-1650.

Reeder, J. & Knight, R. (2010). Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. *Nat Meth* **7**, 668-669.

Respondek, F., Goachet, A. G. & Julliand, V. (2008). Effects of dietary short-chain fructooligosaccharides on the intestinal microflora of horses subjected to a sudden change in diet. *J Anim Sci* **86**, 316-323.

Ruijter, J. M., Van Kampen, A. H. C. & Baas, F. (2002). Statistical evaluation of SAGE libraries: consequences for experimental design. *Physiol Genomics* **11**, 37-44.

Shepherd, M. L., Swecker Jr, W. S., Jensen, R. V. & Ponder, M. A. (2012). Characterization of the fecal bacteria communities of forage fed horses by pyrosequencing of 16S rRNA V4 gene amplicons. *FEMS Microbiol Lett* **326**, 62-68.

Shirazi-Beechey, S. P. (2008). Molecular insights into dietary induced colic in the horse. *Equine Vet J* **40**, 414-421.

Steelman, S. M., Chowdhary, B. P., Dowd, S., Suchodolski, J. & JaneÄ • ka, J. E. (2012). Pyrosequencing of 16S rRNA genes in fecal samples reveals high diversity of hindgut microflora in horses and potential links to chronic laminitis. *BMC Vet Res* **8**, 231.

Superchi, P., Sabbioni, A., Beretti, V., Vecchi, I. & Vaccari Simonini, F. (2010). Hydrolyzable and fermentable carbohydrates in North Italian pastures for horses. *Ital J Anim Sci* **9**, e66.

van den Bogert, B., de Vos, W. M., Zoetendal, E. G. & Kleerebezem, M. (2011). Microarray Analysis and Barcoded Pyrosequencing Provide Consistent Microbial Profiles Depending on the Source of Human Intestinal Samples. *Appl Environ Microbiol* **77**, 2071-2080.

Willing, B., Vörös, A., Roos, S., Jones, C., Jansson, A. & Lindberg, J. (2009). Changes in faecal bacteria associated with concentrate and forage only diets fed to horses in training. *Equine Vet J* **41**, 908-914.

Yamano, H., Koike, S., Kobayashi, Y. & Hata, H. (2008). Phylogenetic analysis of hindgut microbiota in Hokkaido native horses compared to light horses. *Anim Sci J* **79**, 234-242.

Yatsunenko, T., Rey, F. E., Manary, M. J. & other authors (2012). Human gut microbiome viewed across age and geography. *Nature* **486**, 222–227.

Yu, Z., García-González, R., Schanbacher, F. L. & Morrison, M. (2008). Evaluations of Different Hypervariable Regions of Archaeal 16S rRNA Genes in Profiling of Methanogens by Archaea-Specific PCR and Denaturing Gradient Gel Electrophoresis. *Appl Environ Microbiol* **74**, 889-893.

3.5 Supplementary information

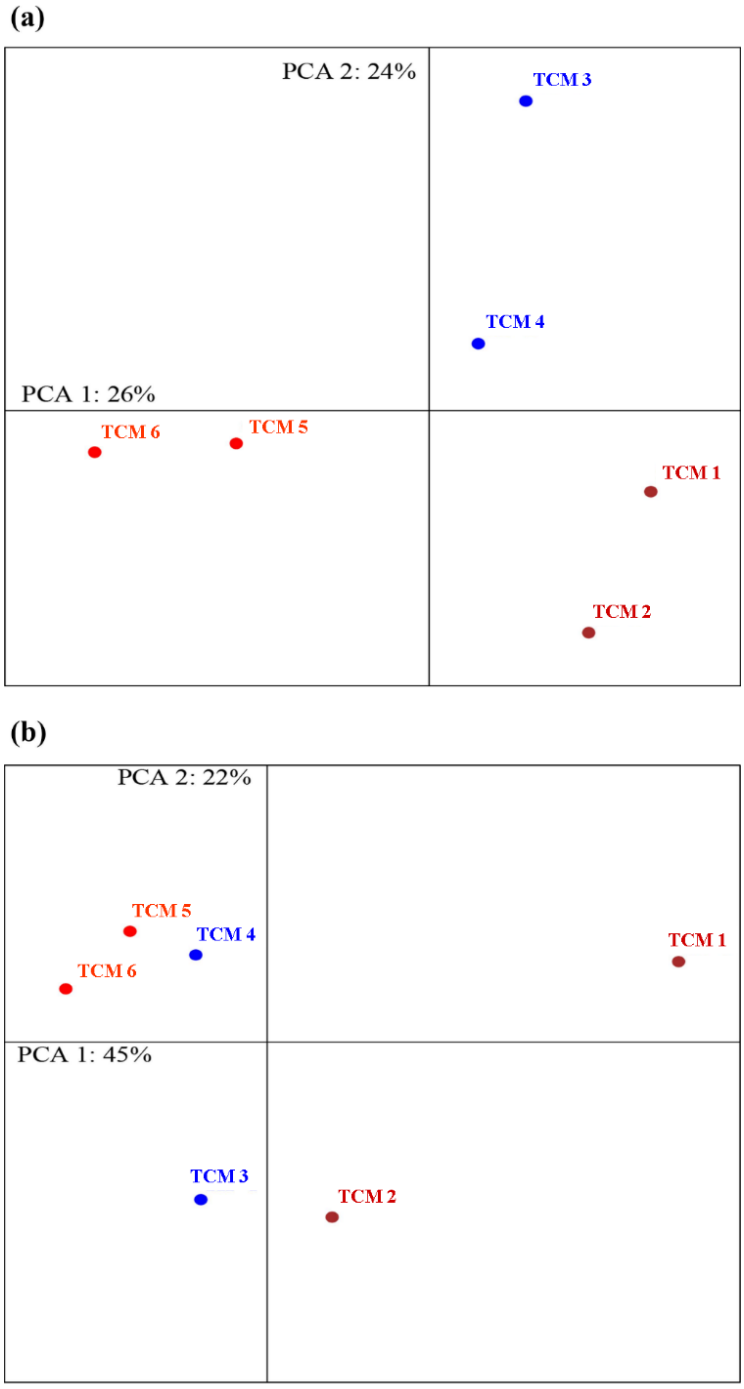


Fig. S3.1 – (a) un-weighted, and panel (b) weighted, PCoA plots, of the six microbiota samples. Each sample is coloured according to the feed received. The closer two samples are in the plots, the more similar their microbiota. Un-weighted PCoA considers presence/absence of OTU's while weighted PCoA also takes relative abundance into account. The percentage values on each axis show the proportion of variation in the microbiota samples that is explained by that axis.

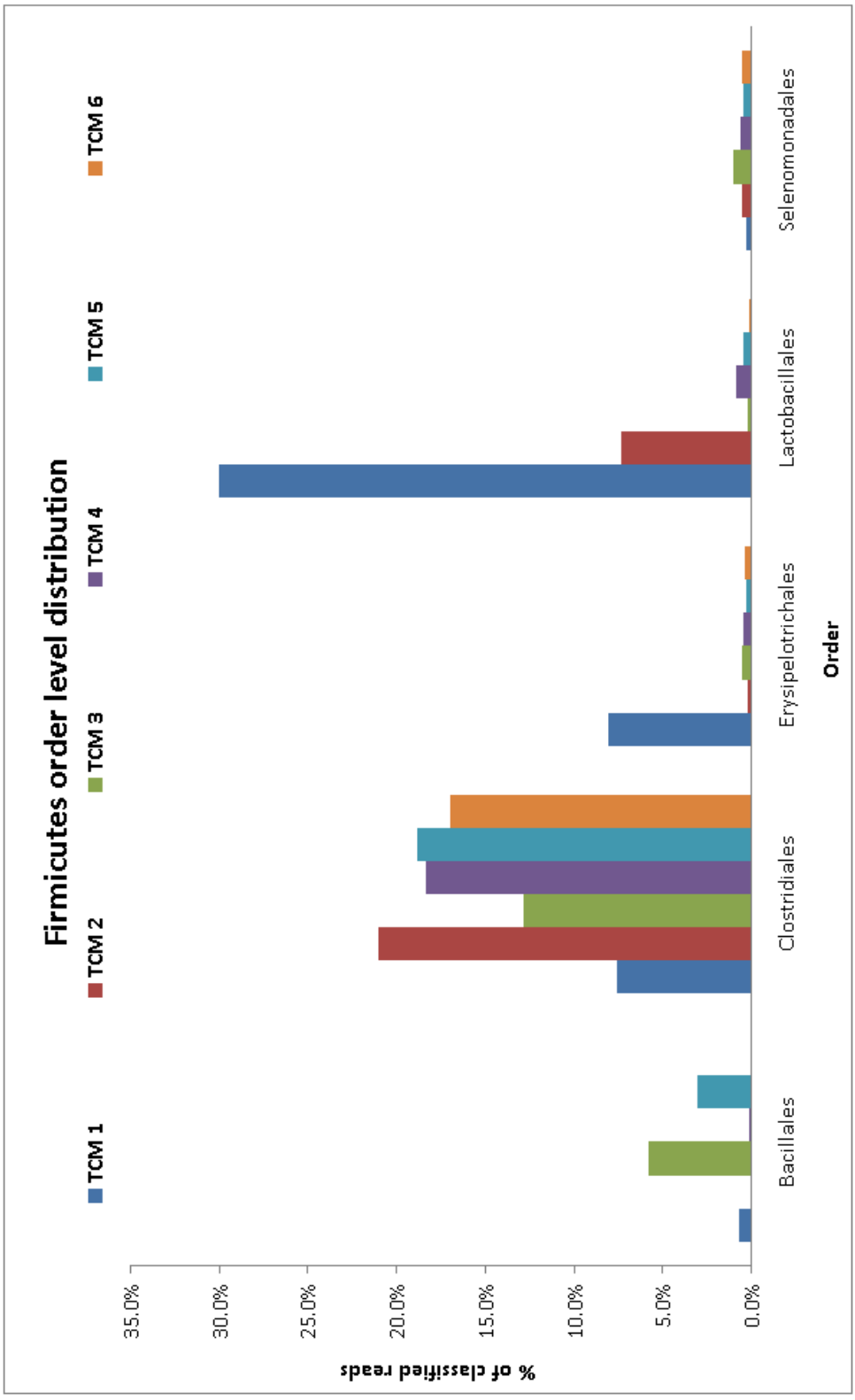


Fig. S3.2 - *Firmicutes* order level read distribution in the hindgut microbiota between the six horses (TCM 1-6) used in this study.

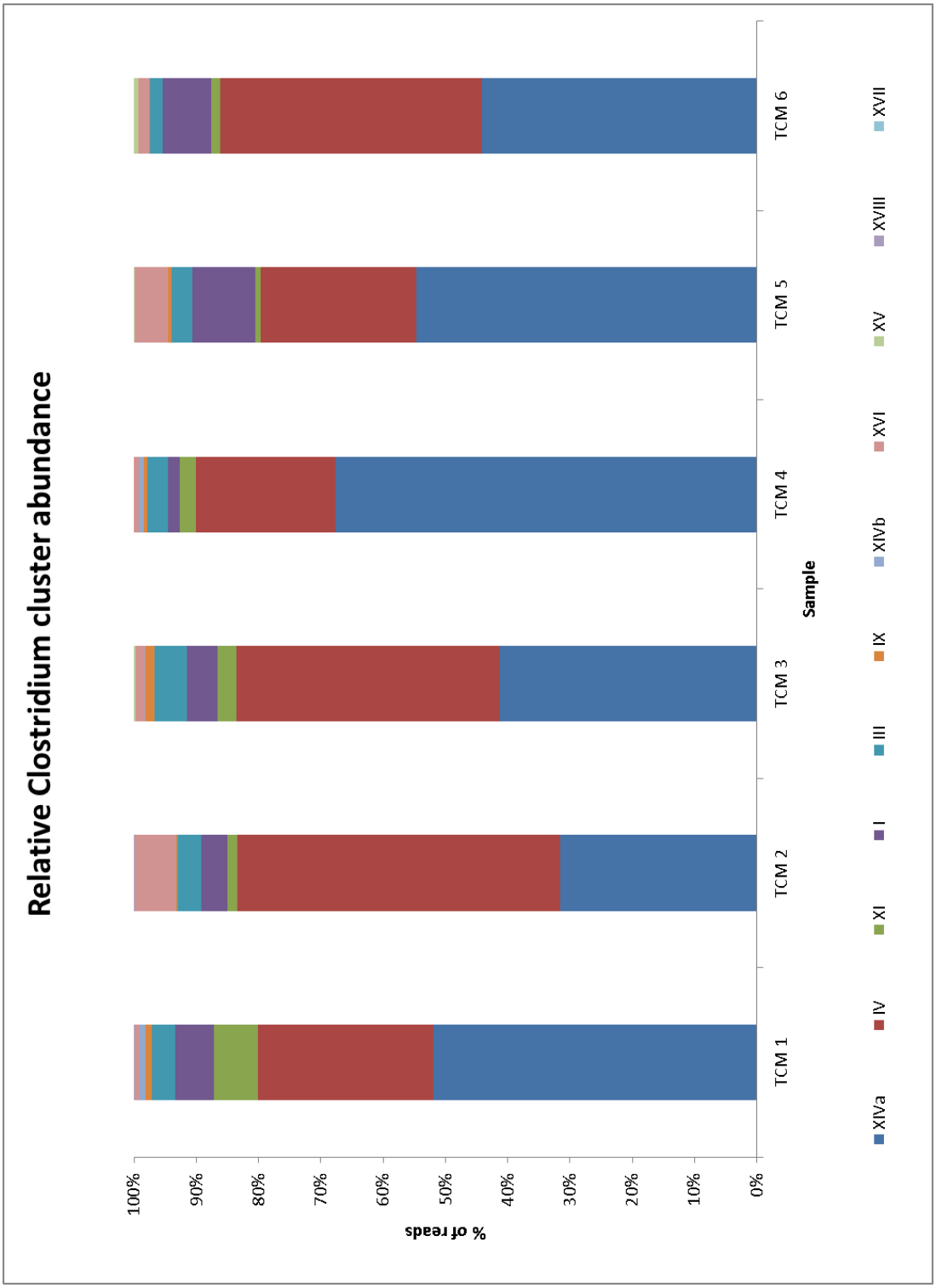


Fig. S3.3 - *Clostridium* cluster assignment in the hindgut microbiota between the Thoroughbred racehorses used in this study.

Table S3.1 Animals and Diets used in this study

Diet	Age	Approximate Weight (kg)	Sex	Time at trial stable (yrs)	Lead-in diet[d]	Stables used
TCM 1[a]	3-4	450-500	Gelding	< 1 month	Mixture of race horse cubes and race horse mix	Stable X, Co. Limerick
TCM 2[a]	7	500	Mare	2-3		
TCM 3[b]	8	400	Filly	4	Oats & Haylage	Stable Y, Co. Clare
TCM 4[b]	7	425	Filly	4		
TCM 5[c]	7	390	Gelding	4	Oats, Nuts & Haylage	
TCM 6[c]	6	420	Gelding	3		

[a] SF, starch-fed;

[b]GF, grass-fed;

[c]HF, haylage-fed

[d] Diet fed on a regular basis before the study began.

Table S3.2 Barcode primers used in this study

Name	Adaptor	Barcode	V4 primer region
EM_01	CGTATCGCCTCCCTCGCGCCATCAG	ACGAGTGCGT	AYTGGGYDTAAAGNG
EM_02	CGTATCGCCTCCCTCGCGCCATCAG	ACGCTCGACA	AYTGGGYDTAAAGNG
EM_03	CGTATCGCCTCCCTCGCGCCATCAG	AGACGCACTC	AYTGGGYDTAAAGNG
EM_04	CGTATCGCCTCCCTCGCGCCATCAG	AGCACTGTAG	AYTGGGYDTAAAGNG
EM_20	CGTATCGCCTCCCTCGCGCCATCAG	TACGAGTATG	AYTGGGYDTAAAGNG
EM_21	CGTATCGCCTCCCTCGCGCCATCAG	TACTCTCGTG	AYTGGGYDTAAAGNG
EM_R	GCCTTGCCAGCCCGCTCAG		TACNVGGGTATCTAATCC

Table S3.3 The number of sequences obtained from faecal samples from racehorses and species richness estimates (using 97% CI).

Sample	Total sequences	OTU count	Phylogenetic diversity	Shannon index (H')	Species Evenness (E)
TCM 1	31,052	1755	102.1	6.99	0.6
TCM 2	22,448	1838	99.5	8.04	0.7
TCM 3	33,694	2234	116.6	8.78	0.8
TCM 4	17,757	1908	100.3	8.99	0.8
TCM 5	35,403	2219	112.7	9.16	0.8
TCM 6	38,378	2736	132.1	9.51	0.8

Table S3.4 Genera and read assignments for each horse used in the study

	Read assignments for each sample						
Genera	**TCM 1**	**TCM 2**	**TCM 3**	**TCM 4**	**TCM 5**	**TCM 6**	**Superkingdom › Phylum › Class › Order › Family › Genus**
Methanobrevibacter	62	21	2	2	3	18	*Archaea › Euryarchaeota › Methanobacteria › Methanobacteriales › Methanobacteriaceae*
Methanocorpusculum	10	137	42	25	25	175	*Archaea › Euryarchaeota › Methanomicrobia › Methanomicrobiales › Methanocorpusculaceae*
Corynebacterium	27	1	0	1	0	0	*Bacteria › Actinobacteria › Actinobacteridae › Actinomycetales › Corynebacterineae › Corynebacteriaceae*
Dietzia	1	0	1	0	0	0	*Bacteria › Actinobacteria › Actinobacteridae › Actinomycetales › Corynebacterineae › Dietziaceae*
Gordonia	4	0	0	0	0	0	*Bacteria › Actinobacteria › Actinobacteridae › Actinomycetales › Corynebacterineae › Gordoniaceae*
Mycobacterium	2	0	0	0	2	1	*Bacteria › Actinobacteria › Actinobacteridae › Actinomycetales › Corynebacterineae › Mycobacteriaceae*
Rhodococcus	6	0	0	2	0	0	*Bacteria › Actinobacteria › Actinobacteridae › Actinomycetales › Corynebacterineae › Nocardiaceae*
Blastococcus	0	0	1	0	0	0	*Bacteria › Actinobacteria › Actinobacteridae › Actinomycetales › Frankineae › Geodermatophilaceae*
Stackebrandtia	1	0	2	0	0	0	*Bacteria › Actinobacteria › Actinobacteridae › Actinomycetales › Glycomycineae › Glycomycetaceae*
Brachybacterium	10	0	0	0	0	0	*Bacteria › Actinobacteria › Actinobacteridae › Actinomycetales › Micrococcineae › Dermabacteraceae*
Dermacoccus	1	0	0	0	0	0	*Bacteria › Actinobacteria › Actinobacteridae › Actinomycetales › Micrococcineae › Dermacoccaceae*
Janibacter	11	0	0	0	0	0	*Bacteria › Actinobacteria › Actinobacteridae › Actinomycetales › Micrococcineae › Intrasporangiaceae*
Ornithinicoccus	2	0	0	0	0	0	*Bacteria › Actinobacteria › Actinobacteridae › Actinomycetales › Micrococcineae › Intrasporangiaceae*
Ornithinimicrobium	2	0	0	0	0	0	*Bacteria › Actinobacteria › Actinobacteridae › Actinomycetales › Micrococcineae › Intrasporangiaceae*
Phycicoccus	0	2	0	0	0	0	*Bacteria › Actinobacteria › Actinobacteridae › Actinomycetales › Micrococcineae › Intrasporangiaceae*
Agrococcus	0	0	0	0	1	0	*Bacteria › Actinobacteria › Actinobacteridae › Actinomycetales › Micrococcineae › Microbacteriaceae*
Leucobacter	0	1	1	0	0	0	*Bacteria › Actinobacteria › Actinobacteridae › Actinomycetales › Micrococcineae › Microbacteriaceae*
Arthrobacter	4	0	10	3	2	0	*Bacteria › Actinobacteria › Actinobacteridae › Actinomycetales › Micrococcineae › Micrococcaceae*
Kocuria	0	0	0	0	1	0	*Bacteria › Actinobacteria › Actinobacteridae › Actinomycetales › Micrococcineae › Micrococcaceae*
Rothia	1	0	0	0	0	0	*Bacteria › Actinobacteria › Actinobacteridae › Actinomycetales › Micrococcineae › Micrococcaceae*
Sinomonas	0	0	0	0	1	0	*Bacteria › Actinobacteria › Actinobacteridae › Actinomycetales › Micrococcineae › Micrococcaceae*
Acaricomes	0	0	1	0	0	0	*Bacteria › Actinobacteria › Actinobacteridae › Actinomycetales › Micrococcineae › Micrococcaceae*
Isoptericola	11	1	1	0	0	0	*Bacteria › Actinobacteria › Actinobacteridae › Actinomycetales › Micrococcineae*
Promicromonospora	4	0	0	0	0	0	*Bacteria › Actinobacteria › Actinobacteridae › Actinomycetales › Micrococcineae*
Yaniella	0	1	0	0	0	0	*Bacteria › Actinobacteria › Actinobacteridae › Actinomycetales › Micrococcineae › Yaniellaceae*
Aeromicrobium	1	0	0	2	1	0	*Bacteria › Actinobacteria › Actinobacteridae › Actinomycetales › Propionibacterineae › Nocardioidaceae*

Nocardioides	8	3	1	0	0	0	*Bacteria › Actinobacteria › Actinobacteridae › Actinomycetales › Propionibacterineae › Nocardioidaceae*
Ponticoccus	1	0	0	0	0	0	*Bacteria › Actinobacteria › Actinobacteridae › Actinomycetales › Propionibacterineae › Propionibacteriaceae*
Actinokineospora	1	0	0	0	0	0	*Bacteria › Actinobacteria › Actinobacteridae › Actinomycetales › Pseudonocardineae › Actinosynnemataceae*
Streptomyces	3	0	0	0	0	0	*Bacteria › Actinobacteria › Actinobacteridae › Actinomycetales › Streptomycineae › Streptomycetaceae*
Bifidobacterium	10	1	2	0	0	0	*Bacteria › Actinobacteria › Actinobacteridae › Bifidobacteriales › Bifidobacteriaceae*
Asaccharobacter	0	0	0	0	6	0	*Bacteria › Actinobacteria › Coriobacteridae › Coriobacteriales › Coriobacterineae › Coriobacteriaceae*
Denitrobacterium	0	0	0	0	2	1	*Bacteria › Actinobacteria › Coriobacteridae › Coriobacteriales › Coriobacterineae › Coriobacteriaceae*
Enterorhabdus	0	0	0	0	1	0	*Bacteria › Actinobacteria › Coriobacteridae › Coriobacteriales › Coriobacterineae › Coriobacteriaceae*
Olsenella	0	0	0	0	2	0	*Bacteria › Actinobacteria › Coriobacteridae › Coriobacteriales › Coriobacterineae › Coriobacteriaceae*
Paraeggerthella	0	0	0	0	0	3	*Bacteria › Actinobacteria › Coriobacteridae › Coriobacteriales › Coriobacterineae › Coriobacteriaceae*
Phocaeicola	2	1	5	19	54	66	*Bacteria › Bacteroidetes › Bacteroidia › Bacteroidales*
Bacteroides	30	5	10	3	17	6	*Bacteria › Bacteroidetes › Bacteroidia › Bacteroidales › Bacteroidaceae*
Anaerophaga	138	58	402	139	65	493	*Bacteria › Bacteroidetes › Bacteroidia › Bacteroidales › Marinilabiaceae*
Barnesiella	6	12	6	0	19	2	*Bacteria › Bacteroidetes › Bacteroidia › Bacteroidales › Porphyromonadaceae*
Butyricimonas	1	0	0	0	0	0	*Bacteria › Bacteroidetes › Bacteroidia › Bacteroidales › Porphyromonadaceae*
Paludibacter	107	90	60	92	514	565	*Bacteria › Bacteroidetes › Bacteroidia › Bacteroidales › Porphyromonadaceae*
Parabacteroides	4	11	6	12	0	0	*Bacteria › Bacteroidetes › Bacteroidia › Bacteroidales › Porphyromonadaceae*
Hallella	0	2	4	8	2	8	*Bacteria › Bacteroidetes › Bacteroidia › Bacteroidales › Prevotellaceae*
Paraprevotella	17	16	302	54	59	127	*Bacteria › Bacteroidetes › Bacteroidia › Bacteroidales › Prevotellaceae*
Prevotella	105	250	459	92	117	465	*Bacteria › Bacteroidetes › Bacteroidia › Bacteroidales › Prevotellaceae*
Xylanibacter	2	0	15	1	0	0	*Bacteria › Bacteroidetes › Bacteroidia › Bacteroidales › Prevotellaceae*
Alistipes	6	0	0	12	0	7	*Bacteria › Bacteroidetes › Bacteroidia › Bacteroidales › Rikenellaceae*
Algoriphagus	0	0	4	0	0	0	*Bacteria › Bacteroidetes › Cytophagia › Cytophagales › Cyclobacteriaceae*
Brumimicrobium	1	0	53	1	0	43	*Bacteria › Bacteroidetes › Flavobacteria › Flavobacteriales › Cryomorphaceae*
Lishizhenia	1	2	1	1	0	0	*Bacteria › Bacteroidetes › Flavobacteria › Flavobacteriales › Cryomorphaceae*
Aequorivita	16	0	10	0	0	0	*Bacteria › Bacteroidetes › Flavobacteria › Flavobacteriales › Flavobacteriaceae*
Galbibacter	83	41	104	109	31	105	*Bacteria › Bacteroidetes › Flavobacteria › Flavobacteriales › Flavobacteriaceae*
Fluviicola	0	0	1	0	0	0	*Bacteria › Bacteroidetes › Flavobacteriia › Flavobacteriales › Cryomorphaceae*
Chryseobacterium	0	0	0	0	1	0	*Bacteria › Bacteroidetes › Flavobacteriia › Flavobacteriales › Flavobacteriaceae*
Coenonia	0	0	0	0	1	4	*Bacteria › Bacteroidetes › Flavobacteriia › Flavobacteriales › Flavobacteriaceae*
Croceibacter	0	0	1	1	0	1	*Bacteria › Bacteroidetes › Flavobacteriia › Flavobacteriales › Flavobacteriaceae*

Empedobacter	0	0	10	4	136	0	*Bacteria › Bacteroidetes › Flavobacteriia › Flavobacteriales › Flavobacteriaceae*
Flavobacterium	0	0	2	0	0	0	*Bacteria › Bacteroidetes › Flavobacteriia › Flavobacteriales › Flavobacteriaceae*
Gelidibacter	0	0	1	0	0	0	*Bacteria › Bacteroidetes › Flavobacteriia › Flavobacteriales › Flavobacteriaceae*
Marixanthomonas	0	0	4	0	0	0	*Bacteria › Bacteroidetes › Flavobacteriia › Flavobacteriales › Flavobacteriaceae*
Subsaxibacter	0	0	0	1	0	0	*Bacteria › Bacteroidetes › Flavobacteriia › Flavobacteriales › Flavobacteriaceae*
Vitellibacter	0	0	1	0	0	0	*Bacteria › Bacteroidetes › Flavobacteriia › Flavobacteriales › Flavobacteriaceae*
Haliscomenobacter	2	0	4	0	0	0	*Bacteria › Bacteroidetes › Sphingobacteria › Sphingobacteriales › Saprospiraceae*
Parapedobacter	12	0	29	7	0	2	*Bacteria › Bacteroidetes › Sphingobacteria › Sphingobacteriales › Sphingobacteriaceae*
Pedobacter	4	0	22	0	0	0	*Bacteria › Bacteroidetes › Sphingobacteria › Sphingobacteriales › Sphingobacteriaceae*
Pseudosphingobacterium	0	1	1	0	0	0	*Bacteria › Bacteroidetes › Sphingobacteria › Sphingobacteriales › Sphingobacteriaceae*
Solitalea	2	0	0	0	0	0	*Bacteria › Bacteroidetes › Sphingobacteria › Sphingobacteriales › Sphingobacteriaceae*
Ferruginibacter	0	0	0	0	1	1	*Bacteria › Bacteroidetes › Sphingobacteriia › Sphingobacteriales › Chitinophagaceae*
Lewinella	0	0	1	0	0	0	*Bacteria › Bacteroidetes › Sphingobacteriia › Sphingobacteriales › Saprospiraceae*
Mucilaginibacter	0	0	4	0	0	0	*Bacteria › Bacteroidetes › Sphingobacteriia › Sphingobacteriales › Sphingobacteriaceae*
Nubsella	0	0	0	0	2	0	*Bacteria › Bacteroidetes › Sphingobacteriia › Sphingobacteriales › Sphingobacteriaceae*
Sphingobacterium	0	0	16	0	0	0	*Bacteria › Bacteroidetes › Sphingobacteriia › Sphingobacteriales › Sphingobacteriaceae*
TM7_genera	69	34	38	4	10	13	*Bacteria › candidate division TM7*
Neochlamydia	10	0	0	0	0	0	*Bacteria › Chlamydiae › Chlamydiales › Parachlamydiaceae*
Parachlamydia	0	0	2	1	0	0	*Bacteria › Chlamydiae › Chlamydiales › Parachlamydiaceae*
Sphaerobacter	1	0	0	0	0	0	*Bacteria › Chloroflexi › Sphaerobacteridae › Sphaerobacterales › Sphaerobacterineae › Sphaerobacteraceae*
GpXIII	0	0	0	0	0	1	*Bacteria › Cyanobacteria ›*
Mucispirillum	0	2	0	0	0	0	*Bacteria › Deferribacteres › Deferribacterales › Deferribacteraceae*
Fibrobacter	94	63	151	925	1307	886	*Bacteria › Fibrobacteres › Fibrobacterales › Fibrobacteraceae*
Bacillus	1	1	2	0	20	6	*Bacteria › Firmicutes › Bacillales › Bacillaceae*
Geobacillus	1	0	0	0	0	0	*Bacteria › Firmicutes › Bacillales › Bacillaceae*
Lysinibacillus	12	0	1556	16	724	1	*Bacteria › Firmicutes › Bacillales › Bacillaceae*
Paraliobacillus	1	0	0	0	0	0	*Bacteria › Firmicutes › Bacillales › Bacillaceae*
Paenibacillus	1	0	4	0	21	1	*Bacteria › Firmicutes › Bacillales › Paenibacillaceae*
Caryophanon	2	0	180	0	199	0	*Bacteria › Firmicutes › Bacillales › Planococcaceae*
Kurthia	2	1	74	0	16	0	*Bacteria › Firmicutes › Bacillales › Planococcaceae*
Paenisporosarcina	0	0	1	0	3	0	*Bacteria › Firmicutes › Bacillales › Planococcaceae*
Rummeliibacillus	161	3	102	1	0	0	*Bacteria › Firmicutes › Bacillales › Planococcaceae*

Solibacillus	0	0	6	0	9	0	*Bacteria › Firmicutes › Bacillales › Planococcaceae*
Sporosarcina	0	0	0	0	2	1	*Bacteria › Firmicutes › Bacillales › Planococcaceae*
Viridibacillus	20	0	1	0	5	0	*Bacteria › Firmicutes › Bacillales › Planococcaceae*
Marinibacillus	0	0	10	0	71	2	*Bacteria › Firmicutes › Bacillales › Planococcaceae › Jeotgalibacillus*
Anaerovirgula	0	0	1	0	0	0	*Bacteria › Firmicutes › Clostridia › Clostridiales*
Blautia	86	180	30	328	625	421	*Bacteria › Firmicutes › Clostridia › Clostridiales*
Proteiniborus	0	0	0	1	0	0	*Bacteria › Firmicutes › Clostridia › Clostridiales*
Anaerobacter	0	1	0	0	1	1	*Bacteria › Firmicutes › Clostridia › Clostridiales › Clostridiaceae*
Anaerosporobacter	32	25	62	52	94	87	*Bacteria › Firmicutes › Clostridia › Clostridiales › Clostridiaceae*
Butyricicoccus	8	5	29	10	14	50	*Bacteria › Firmicutes › Clostridia › Clostridiales › Clostridiaceae*
Clostridium	186	149	143	18	469	380	*Bacteria › Firmicutes › Clostridia › Clostridiales › Clostridiaceae*
Lactonifactor	53	96	34	161	136	71	*Bacteria › Firmicutes › Clostridia › Clostridiales › Clostridiaceae*
Lutispora	4	3	24	1	25	10	*Bacteria › Firmicutes › Clostridia › Clostridiales › Clostridiaceae*
Natronincola	0	0	0	1	0	0	*Bacteria › Firmicutes › Clostridia › Clostridiales › Clostridiaceae*
Oxobacter	0	0	2	1	0	1	*Bacteria › Firmicutes › Clostridia › Clostridiales › Clostridiaceae*
Sarcina	6	10	2	68	293	141	*Bacteria › Firmicutes › Clostridia › Clostridiales › Clostridiaceae*
Thermobrachium	3	0	0	0	0	0	*Bacteria › Firmicutes › Clostridia › Clostridiales › Clostridiaceae*
Dethiosulfatibacter	0	1	0	0	0	0	*Bacteria › Firmicutes › Clostridia › Clostridiales › Clostridiales Family XI. Incertae Sedis*
Sedimentibacter	0	0	0	0	0	1	*Bacteria › Firmicutes › Clostridia › Clostridiales › Clostridiales Family XI. Incertae Sedis*
Acidaminobacter	0	0	0	0	1	0	*Bacteria › Firmicutes › Clostridia › Clostridiales › Clostridiales Family XII. Incertae Sedis*
Guggenheimella	0	0	1	0	0	0	*Bacteria › Firmicutes › Clostridia › Clostridiales › Clostridiales Family XII. Incertae Sedis*
Anaerovorax	4	3	4	1	5	7	*Bacteria › Firmicutes › Clostridia › Clostridiales › Clostridiales Family XIII. Incertae Sedis*
Mogibacterium	159	27	11	29	14	17	*Bacteria › Firmicutes › Clostridia › Clostridiales › Clostridiales Family XIII. Incertae Sedis*
Acetobacterium	0	1	0	1	1	2	*Bacteria › Firmicutes › Clostridia › Clostridiales › Eubacteriaceae*
Alkalibacter	0	5	0	0	0	1	*Bacteria › Firmicutes › Clostridia › Clostridiales › Eubacteriaceae*
Anaerofustis	4	0	0	8	0	0	*Bacteria › Firmicutes › Clostridia › Clostridiales › Eubacteriaceae*
Eubacterium	5	106	5	24	192	53	*Bacteria › Firmicutes › Clostridia › Clostridiales › Eubacteriaceae*
Pseudoramibacter	3	0	0	0	0	5	*Bacteria › Firmicutes › Clostridia › Clostridiales › Eubacteriaceae*
Gracilibacter	0	1	0	0	0	2	*Bacteria › Firmicutes › Clostridia › Clostridiales › Graciibacteraceae*
Acetitomaculum	30	112	13	92	450	160	*Bacteria › Firmicutes › Clostridia › Clostridiales › Lachnospiraceae*
Anaerostipes	6	14	3	3	3	2	*Bacteria › Firmicutes › Clostridia › Clostridiales › Lachnospiraceae*
Butyrivibrio	0	1	0	1	4	1	*Bacteria › Firmicutes › Clostridia › Clostridiales › Lachnospiraceae*

Coprococcus	143	103	210	223	493	252	*Bacteria › Firmicutes › Clostridia › Clostridiales › Lachnospiraceae*
Dorea	30	57	21	48	142	135	*Bacteria › Firmicutes › Clostridia › Clostridiales › Lachnospiraceae*
Hespellia	4	4	4	9	96	10	*Bacteria › Firmicutes › Clostridia › Clostridiales › Lachnospiraceae*
Johnsonella	1	0	1	4	2	4	*Bacteria › Firmicutes › Clostridia › Clostridiales › Lachnospiraceae*
Lachnobacterium	4	0	2	2	6	6	*Bacteria › Firmicutes › Clostridia › Clostridiales › Lachnospiraceae*
Marvinbryantia	0	1	0	0	0	0	*Bacteria › Firmicutes › Clostridia › Clostridiales › Lachnospiraceae*
Moryella	2	0	5	2	5	6	*Bacteria › Firmicutes › Clostridia › Clostridiales › Lachnospiraceae*
Oribacterium	10	5	17	7	33	77	*Bacteria › Firmicutes › Clostridia › Clostridiales › Lachnospiraceae*
Parasporobacterium	0	16	1	24	9	5	*Bacteria › Firmicutes › Clostridia › Clostridiales › Lachnospiraceae*
Pseudobutyrivibrio	30	71	49	173	170	200	*Bacteria › Firmicutes › Clostridia › Clostridiales › Lachnospiraceae*
Robinsoniella	37	7	30	81	65	177	*Bacteria › Firmicutes › Clostridia › Clostridiales › Lachnospiraceae*
Roseburia	34	39	39	92	289	202	*Bacteria › Firmicutes › Clostridia › Clostridiales › Lachnospiraceae*
Syntrophococcus	5	5	9	8	23	12	*Bacteria › Firmicutes › Clostridia › Clostridiales › Lachnospiraceae*
Oscillibacter	527	1185	819	101	361	764	*Bacteria › Firmicutes › Clostridia › Clostridiales › Oscillospiracea*
Peptococcus	2	0	0	1	0	4	*Bacteria › Firmicutes › Clostridia › Clostridiales › Peptococcaceae*
Acetanaerobacterium	2	2	24	8	6	26	*Bacteria › Firmicutes › Clostridia › Clostridiales › Ruminococcaceae*
Acetivibrio	124	246	230	208	363	395	*Bacteria › Firmicutes › Clostridia › Clostridiales › Ruminococcaceae*
Anaerofilum	1	1	4	1	3	3	*Bacteria › Firmicutes › Clostridia › Clostridiales › Ruminococcaceae*
Anaerotruncus	6	489	407	69	61	194	*Bacteria › Firmicutes › Clostridia › Clostridiales › Ruminococcaceae*
Ethanoligenens	10	23	48	5	6	25	*Bacteria › Firmicutes › Clostridia › Clostridiales › Ruminococcaceae*
Faecalibacterium	47	24	81	86	141	438	*Bacteria › Firmicutes › Clostridia › Clostridiales › Ruminococcaceae*
Hydrogenoanaerobacterium	13	5	19	21	26	13	*Bacteria › Firmicutes › Clostridia › Clostridiales › Ruminococcaceae*
Papillibacter	95	79	285	59	101	385	*Bacteria › Firmicutes › Clostridia › Clostridiales › Ruminococcaceae*
Ruminococcus	217	204	450	568	815	470	*Bacteria › Firmicutes › Clostridia › Clostridiales › Ruminococcaceae*
Sporobacter	410	1420	1187	647	1104	1260	*Bacteria › Firmicutes › Clostridia › Clostridiales › Ruminococcaceae*
Subdoligranulum	0	1	3	0	3	12	*Bacteria › Firmicutes › Clostridia › Clostridiales › Ruminococcaceae*
Pelospora	1	0	11	0	2	0	*Bacteria › Firmicutes › Clostridia › Clostridiales › Syntrophomonadaceae*
Allobaculum	2477	4	2	14	3	0	*Bacteria › Firmicutes › Erysipelotrichi › Erysipelotrichales › Erysipelotrichaceae*
Bulleidia	4	0	60	1	9	13	*Bacteria › Firmicutes › Erysipelotrichi › Erysipelotrichales › Erysipelotrichaceae*
Catenibacterium	2	1	0	0	1	1	*Bacteria › Firmicutes › Erysipelotrichi › Erysipelotrichales › Erysipelotrichaceae*
Coprobacillus	8	13	0	1	10	5	*Bacteria › Firmicutes › Erysipelotrichi › Erysipelotrichales › Erysipelotrichaceae*
Erysipelothrix	0	0	1	0	0	0	*Bacteria › Firmicutes › Erysipelotrichi › Erysipelotrichales › Erysipelotrichaceae*

Holdemania	21	12	97	38	53	103	*Bacteria › Firmicutes › Erysipelotrichi › Erysipelotrichales › Erysipelotrichaceae*
Solobacterium	0	1	0	11	9	6	*Bacteria › Firmicutes › Erysipelotrichi › Erysipelotrichales › Erysipelotrichaceae*
Weissella	112	24	3	30	0	0	*Bacteria › Firmicutes › Lactobacillales*
Facklamia	4	0	0	0	0	0	*Bacteria › Firmicutes › Lactobacillales › Aerococcaceae*
Atopostipes	0	0	1	0	0	0	*Bacteria › Firmicutes › Lactobacillales › Carnobacteriaceae*
Carnobacterium	0	0	13	0	0	0	*Bacteria › Firmicutes › Lactobacillales › Carnobacteriaceae*
Desemzia	0	0	19	0	0	0	*Bacteria › Firmicutes › Lactobacillales › Carnobacteriaceae*
Enterococcus	58	0	1	20	2	0	*Bacteria › Firmicutes › Lactobacillales › Enterococcaceae*
Melissococcus	0	0	0	0	1	0	*Bacteria › Firmicutes › Lactobacillales › Enterococcaceae*
Lactobacillus	805	171	9	71	134	6	*Bacteria › Firmicutes › Lactobacillales › Lactobacillaceae*
Sharpea	1	1	4	0	2	17	*Bacteria › Firmicutes › Lactobacillales › Lactobacillaceae*
Streptococcus	8383	1455	7	27	0	0	*Bacteria › Firmicutes › Lactobacillales › Streptococcaceae*
Acidaminococcus	47	74	233	29	76	152	*Bacteria › Firmicutes › Negativicutes › Selenomonadales › Acidaminococcaceae*
Phascolarctobacterium	31	13	47	26	42	5	*Bacteria › Firmicutes › Negativicutes › Selenomonadales › Acidaminococcaceae*
Anaerovibrio	6	16	4	4	24	15	*Bacteria › Firmicutes › Negativicutes › Selenomonadales › Veillonellaceae*
Centipeda	0	0	0	3	0	4	*Bacteria › Firmicutes › Negativicutes › Selenomonadales › Veillonellaceae*
Propionispira	0	0	0	0	1	0	*Bacteria › Firmicutes › Negativicutes › Selenomonadales › Veillonellaceae*
Schwartzia	3	4	30	10	4	0	*Bacteria › Firmicutes › Negativicutes › Selenomonadales › Veillonellaceae*
Selenomonas	1	0	0	32	2	7	*Bacteria › Firmicutes › Negativicutes › Selenomonadales › Veillonellaceae*
Veillonella	0	1	0	0	0	0	*Bacteria › Firmicutes › Negativicutes › Selenomonadales › Veillonellaceae*
Victivallis	10	8	6	6	8	47	*Bacteria › Lentisphaerae › Victivallales › Victivallaceae*
Pirellula	5	9	6	11	1	6	*Bacteria › Planctomycetes › Planctomycetacia › Planctomycetales › Planctomycetaceae*
Phenylobacterium	1	0	0	0	0	0	*Bacteria › Proteobacteria › Alphaproteobacteria › Caulobacterales › Caulobacteraceae*
Rhodopseudomonas	0	0	0	0	2	0	*Bacteria › Proteobacteria › Alphaproteobacteria › Rhizobiales › Bradyrhizobiaceae*
Hyphomicrobium	0	0	0	1	0	0	*Bacteria › Proteobacteria › Alphaproteobacteria › Rhizobiales › Hyphomicrobiaceae*
Methylobacterium	0	0	0	0	1	0	*Bacteria › Proteobacteria › Alphaproteobacteria › Rhizobiales › Methylobacteriaceae*
Rhizobium	2	0	1	0	5	1	*Bacteria › Proteobacteria › Alphaproteobacteria › Rhizobiales › Rhizobiaceae › Rhizobium/Agrobacterium*
Ensifer	0	0	0	0	1	0	*Bacteria › Proteobacteria › Alphaproteobacteria › Rhizobiales › Rhizobiaceae › Sinorhizobium/Ensifer*
Paracoccus	1	0	0	0	0	0	*Bacteria › Proteobacteria › Alphaproteobacteria › Rhodobacterales › Rhodobacteraceae*
Acetobacter	0	0	0	0	0	1	*Bacteria › Proteobacteria › Alphaproteobacteria › Rhodospirillales › Acetobacteraceae*
Orientia	3	2	13	17	9	8	*Bacteria › Proteobacteria › Alphaproteobacteria › Rickettsiales › Rickettsiaceae › Rickettsieae*
Pelagibacter	0	2	2	5	2	22	*Bacteria › Proteobacteria › Alphaproteobacteria › Rickettsiales › SAR11 cluster*

Sphingopyxis	0	0	1	0	0	0	*Bacteria › Proteobacteria › Alphaproteobacteria › Sphingomonadales › Sphingomonadaceae*
Achromobacter	0	0	1	0	0	0	*Bacteria › Proteobacteria › Betaproteobacteria › Burkholderiales › Alcaligenaceae*
Castellaniella	1	0	1	0	0	0	*Bacteria › Proteobacteria › Betaproteobacteria › Burkholderiales › Alcaligenaceae*
Oligella	0	0	1	0	0	0	*Bacteria › Proteobacteria › Betaproteobacteria › Burkholderiales › Alcaligenaceae*
Comamonas	0	0	1	0	0	0	*Bacteria › Proteobacteria › Betaproteobacteria › Burkholderiales › Comamonadaceae*
Parasutterella	1	26	0	0	0	0	*Bacteria › Proteobacteria › Betaproteobacteria › Burkholderiales › Sutterellaceae*
Sutterella	0	1	1	0	6	5	*Bacteria › Proteobacteria › Betaproteobacteria › Burkholderiales › Sutterellaceae*
Kingella	0	2	0	0	0	0	*Bacteria › Proteobacteria › Betaproteobacteria › Neisseriales › Neisseriaceae*
Azoarcus	0	0	2	0	0	0	*Bacteria › Proteobacteria › Betaproteobacteria › Rhodocyclales › Rhodocyclaceae*
Desulfobulbus	0	0	0	0	0	1	*Bacteria › Proteobacteria › Deltaproteobacteria › Desulfobacterales › Desulfobulbaceae*
Desulfovibrio	0	1	8	3	4	6	*Bacteria › Proteobacteria › Deltaproteobacteria › Desulfovibrionales › Desulfovibrionaceae*
Desulfovirga	3	0	0	1	0	0	*Bacteria › Proteobacteria › Deltaproteobacteria › Syntrophobacterales › Syntrophobacteraceae*
Campylobacter	4	0	0	0	0	0	*Bacteria › Proteobacteria › Epsilonproteobacteria › Campylobacterales › Campylobacteraceae*
Helicobacter	2	1	1	2	10	0	*Bacteria › Proteobacteria › Epsilonproteobacteria › Campylobacterales › Helicobacteraceae*
Anaerobiospirillum	0	0	2	0	0	0	*Bacteria › Proteobacteria › Gammaproteobacteria › Aeromonadales › Succinivibrionaceae*
Ruminobacter	1	2	3	1	0	0	*Bacteria › Proteobacteria › Gammaproteobacteria › Aeromonadales › Succinivibrionaceae*
Succinivibrio	68	246	0	0	0	0	*Bacteria › Proteobacteria › Gammaproteobacteria › Aeromonadales › Succinivibrionaceae*
Nitrosococcus	1	0	0	0	0	0	*Bacteria › Proteobacteria › Gammaproteobacteria › Chromatiales › Chromatiaceae*
Escherichia/Shigella	8	1	130	33	2	0	*Bacteria › Proteobacteria › Gammaproteobacteria › Enterobacteriales › Enterobacteriaceae*
Pectobacterium	0	0	0	0	2	0	*Bacteria › Proteobacteria › Gammaproteobacteria › Enterobacteriales › Enterobacteriaceae*
Halomonas	0	0	1	0	0	0	*Bacteria › Proteobacteria › Gammaproteobacteria › Oceanospirillales › Halomonadaceae*
Oleiphilus	1	0	0	0	0	0	*Bacteria › Proteobacteria › Gammaproteobacteria › Oceanospirillales › Oleiphilaceae*
Actinobacillus	0	1	0	0	0	0	*Bacteria › Proteobacteria › Gammaproteobacteria › Pasteurellales › Pasteurellaceae*
Alkanindiges	0	0	0	1	0	0	*Bacteria › Proteobacteria › Gammaproteobacteria › Pseudomonadales › Moraxellaceae*
Psychrobacter	0	0	2	0	0	0	*Bacteria › Proteobacteria › Gammaproteobacteria › Pseudomonadales › Moraxellaceae*
Acinetobacter	59	2	3075	62	352	1	*Bacteria › Proteobacteria › Gammaproteobacteria › Pseudomonadales › Moraxellaceae › Acinetobacter*
Pseudomonas	0	0	18	0	1	0	*Bacteria › Proteobacteria › Gammaproteobacteria › Pseudomonadales › Pseudomonadaceae*
Methylophaga	0	0	2	0	0	0	*Bacteria › Proteobacteria › Gammaproteobacteria › Thiotrichales › Piscirickettsiaceae*
Luteibacter	4	0	1	0	0	0	*Bacteria › Proteobacteria › Gammaproteobacteria › Xanthomonadales › Xanthomonadaceae*
Luteimonas	11	0	0	0	0	0	*Bacteria › Proteobacteria › Gammaproteobacteria › Xanthomonadales › Xanthomonadaceae*
Lysobacter	1	0	0	0	0	0	*Bacteria › Proteobacteria › Gammaproteobacteria › Xanthomonadales › Xanthomonadaceae*
Treponema	568	513	1566	1015	2067	1098	*Bacteria › Spirochaetes › Spirochaetales › Spirochaetaceae*

Pyramidobacter	0	0	0	0	5	2	*Bacteria › Synergistetes › Synergistia › Synergistales › Synergistaceae*
Synergistes	1	0	5	2	1	4	*Bacteria › Synergistetes › Synergistia › Synergistales › Synergistaceae*
Acholeplasma	0	0	0	1	1	1	*Bacteria › Tenericutes › Mollicutes › Acholeplasmatales › Acholeplasmataceae*
Anaeroplasma	10	16	2	30	68	199	*Bacteria › Tenericutes › Mollicutes › Anaeroplasmatales › Anaeroplasmataceae*
Cerasicoccus	12	1	0	0	2	2	*Bacteria › Verrucomicrobia › Opitutae › Puniceicoccales › Puniceicoccaceae*
Coraliomargarita	0	0	0	0	0	1	*Bacteria › Verrucomicrobia › Opitutae › Puniceicoccales › Puniceicoccaceae*
Spartobacteria_incertae_sedis	0	0	0	0	0	1	*Bacteria › Verrucomicrobia › Spartobacteria*
Subdivision3	0	0	0	0	0	1	*Bacteria › Verrucomicrobia › Verrucomicrobiae › Verrucomicrobiales › Verrucomicrobia subdivision 3*
Subdivision5	665	540	1009	695	2263	4155	*Bacteria › Verrucomicrobia › Verrucomicrobiae › Verrucomicrobiales › Verrucomicrobia subdivision 5*
Akkermansia	55	47	6	9	34	7	*Bacteria › Verrucomicrobia › Verrucomicrobiae › Verrucomicrobiales › Verrucomicrobiaceae*
Persicirhabdus	2	1	1	0	0	1	*Bacteria › Verrucomicrobia › Verrucomicrobiae › Verrucomicrobiales › Verrucomicrobiaceae*
Streptophyta	0	0	0	0	1	0	*Eukaryota › Viridiplantae*
Unclassified	14318	13821	19246	10760	20077	22982	Unclassified
Total reads	31222	22521	33694	17757	35403	38378	

Table S3.5 The 13 species that form the core microbiome accounting for ≥0.1% of the total reads for 4 or more animals used in this study.

	% of the total reads per animal						
Species	**TCM 1**	**TCM 2**	**TCM 3**	**TCM 4**	**TCM 5**	**TCM 6**	**Order › Family › Genus**
Paludibacter propionicigenes	0.2%	0.1%	0.0%	0.1%	0.9%	0.4%	*Bacteroidales › Porphyromonadaceae › Paludibacter*
Fibrobacter succinogenes	0.3%	0.3%	0.4%	5.2%	3.7%	2.3%	*Fibrobacterales › Fibrobacteraceae › Fibrobacter*
Eubacterium coprostanoligenes	0.2%	0.1%	0.3%	0.3%	0.3%	0.1%	*Clostridiales › Eubacteriaceae › Eubacterium*
Eubacterium hallii	0.2%	0.6%	0.0%	0.9%	1.2%	0.3%	*Clostridiales › Eubacteriaceae › Eubacterium*
Eubacterium ruminantium	0.2%	0.2%	0.2%	0.8%	0.3%	0.1%	*Clostridiales › Eubacteriaceae › Eubacterium*
Oscillospira guilliermondii	0.8%	3.9%	1.4%	0.3%	1.0%	1.1%	*Clostridiales › Ruminococcaceae › Oscillospira*
Sporobacter termitidis	1.3%	5.6%	3.2%	2.2%	1.6%	3.4%	*Clostridiales › Ruminococcaceae › Sporobacter*
Lactobacillus equicursoris	1.8%	0.5%	0.0%	0.4%	0.1%	0.0%	*Lactobacillales › Lactobacillaceae › Lactobacillus*
Phascolarctobacterium faecium	0.1%	0.0%	0.1%	0.1%	0.1%	0.0%	*Selenomonadales › Acidaminococcaceae › Phascolarctobacterium*
Treponema brennaborense	0.1%	0.0%	0.0%	0.2%	0.1%	0.1%	*Spirochaetales › Spirochaetaceae › Treponema*
Treponema parvum	0.1%	0.1%	0.2%	0.4%	0.6%	0.3%	*Spirochaetales › Spirochaetaceae › Treponema*
Treponema porcinum	0.5%	0.6%	0.5%	1.8%	0.6%	0.7%	*Spirochaetales › Spirochaetaceae › Treponema*
Treponema saccharophilum	0.1%	0.1%	0.0%	0.0%	0.3%	0.3%	*Spirochaetales › Spirochaetaceae › Treponema*

Chapter IV

The core microbiota of domesticated herbivorous hindgut fermenters, mono-gastric and ruminant animals

Michelle M. O' Donnell[1], Hugh M.B. Harris[2], R. Paul Ross[1], Paul W. O'Toole[2]

[1]Food Biosciences Department, Teagasc Food Research Centre, Moorepark, Ireland.
[2]Department of Microbiology, University College Cork, Ireland

This chapter is intended for submission as a research article to *Applied Environmental Microbiology*.

Notes:
Sample collection, pyrosequencing PCR and purifications, pyrosequencing read data analysis was carried out by M.M. O' Donnell (author of this thesis)

Pyrosequencing read analysis using Qiime and statistical analysis was run by H. M. B. Harris

Chapter IV

Table of Contents

Abstract

In this study, we aimed to characterise the core faecal microbiota composition and diversity in domesticated herbivorous animals that use three different digestion methods (hindgut fermenters, ruminant and monogastric) to harvest energy from food. The 42 animals, spanning 10 animal species were housed on a single farm in the south of Ireland allowing us to assess these domesticated herbivores as they consume similar feeds while under the same management regime, thereby eliminating some of the factors that influence the microbiota. This study is also, to our knowledge the first to examine in depth the faecal microbiota of the donkey, chinchilla, rabbit, alpaca and llama. The microbiota of all animals tested was dominated by the *Firmicutes* and *Bacteroidetes* phyla. The core microbiota of the each digestion type comprised 18% of the genera identified. The large proportion of unclassified reads (36-72%) identified in the animal species at the genus level, suggests that further studies are required to elucidate the true microbiota of the domesticated herbivores. Fifty-nine species were identified between the different animal faecal samples. *Lactobacillus ruminis* (0.03-0.17%), *Clostridium septicum* (0.01-0.53%) and *Clostridium bifermentans* (0.003-0.04%) were identified in the majority of animal species in this study irrespective of digestion method. We also determined that host phylogeny and to a lesser extent digestion method affect the bacterial diversity in the domesticated herbivore. This study forms a platform for future studies into the microbiota of non-bovine and non-equine domesticated herbivorous animals. It also suggests that the microbiota of domesticated herbivores (equids especially) is an important niche for the mammalian-associated commensal bacterium *Lactobacillus ruminis*.

4.1 Introduction

Animals including humans consume food in order to fulfil their nutritional and energy requirements. Foregut fermenters, otherwise known as ruminants each have a specialised chambered digestive system which has evolved to support a symbiotic relationship with the microorganisms in the microbiota (Warner *et al.*, 1956). The microorganisms execute the breakdown of complex polysaccharides and produce short chain fatty acids (SCFA), carbon dioxide, hydrogen and ammonia (Playne & Kennedy, 1976). The rumen of the animal then absorbs the SCFA (acetic, propionic and butyric acids) and uses them for energy (Bergman, 1990). Camelids (llamas and alpacas) are considered to be pseudo-ruminants as they lack a reticulum and therefore, cannot be considered true ruminants like bovids (cows, goats and sheep) (Abdel-Magied & Taha, 2003). The large intestine of horses and other hindgut fermenters is a fermentation system analagous to the rumen. The hindgut fermenters are able to digest some of the cellulose in their diet by way of symbiotic bacteria in the microbiota by employing longer gut retention times. However, their ability to extract energy from cellulose digestion is less efficient than that of ruminants (Stevens & Hume, 2004). Fermentation by the hindgut microbiota also generates SCFA with approximately 75% of them absorbed by the host animal's intestinal epithelium (Duncan *et al.*, 1990). However, unlike the ruminants, the hindgut fermenters excrete the vast majority of the amino acids generated by the gut microbiota (Demeyer, 1991). Lysine is the primary rate limiting amino acid in horses and therefore equine diets are often supplemented with this amino acid (Hintz & Cymbaluk, 1994).

Many studies on the ruminant microbiota have focused on the bovine microbiota because of their importance in the beef and dairy industry, and also they have primarily analysed solid and liquid fractions taken directly from the rumen (Brulc *et al.*, 2009; Callaway *et al.*, 2010; Jami & Mizrahi, 2012; Welkie *et al.*, 2010). Similarly hindgut fermenter microbiota research has focused on the horse microbiota because of their importance as work and performance animals (Costa *et al.*, 2012; Daly *et al.*, 2001; O' Donnell *et al.*, 2013; Shepherd *et al.*, 2012; Steelman *et al.*, 2012). No study to date has used next-generation sequencing techniques to compare the faecal microbiota of a variety of common domesticated ruminants and hindgut fermenters.

The inter-play and symbiotic relationship between the intestinal microbiota and the host are essential for life. A recent review has summarised the effect and influence that the gut microbiota can have on animal behaviour and highlighted that bacteria either as a total microbiota or as single species can affect host behaviour (Ezenwa *et al.*, 2012). Ley *et al.* (2008) compared the gut microbiota of over 100 animals to that of the humans to assess the composition of the vertebrate microbiota. The study concluded that gut microbiota diversity is influenced by diet (herbivorous, carnivorous or omnivorous) and host phylogeny with herbivorous animals having the most diverse microbiota (Ley *et al.*, 2008). However, a follow-up study examined the faecal microbiota of carnivores, herbivores and omnivores to assess whether diet or phylogeny of host determined the genera and species present in the microbiota (Muegge *et al.*, 2011). Using Principle Coordinate analysis plots to illustrate the differences between the samples, there was a clear separation of carnivores, omnivores and herbivores. This was clear evidence that diet and not phylogeny of the host had the greatest influence on the taxa present (Muegge *et al.*, 2011). Other studies have shown that geographic location as well as diet can influence the microbiota of humans and animals, particularly domesticated/farmed animals (De Filippo *et al.*, 2010; Shanks *et al.*, 2011; Yamano *et al.*, 2008).

The gut microbiota of humans and animals contain between 10^{10}-10^{14} bacteria; however, it is extremely difficult to culture *in vitro* the majority of bacteria present, especially fastidious anaerobic bacteria. In depth examination and comparison of the gut microbiota from humans and animals has been possible through the use of DNA microarrays (Human gut chip and Phylochip) and next generation sequencing (Petrosino *et al.*, 2009; Tottey *et al.*, 2013; Wagner *et al.*, 2007).

This study aimed to identify the bacterial diversity and the core microbiota of 10 species of herbivorous domesticated animal that span three different digestion physiologies. We also aimed to elucidate the taxa specific to hindgut fermenters, ruminants and monogastric animal species.

4.2 Materials and Methods

4.2.1 Animals and diet

All of the animals were housed in the mini farm in the south east of Ireland. None of the animals used in the study had received antibiotic treatments in the 12 months prior to sampling. Similarly none of the animals tested had any health issues prior to sampling and are thus considered to be healthy animals. A list of each animal (and the sample number of each) and the feed consumed by each is given in Table 4.1. The Kingdom, Phylum, Class, Order and Family for each animal species is listed in Table S4.1. Twenty-five hindgut fermenting, sixteen ruminant and four monogastric animals were used in this study, spanning 10 animal species. Animals that were housed indoors (rabbits, chinchillas and pigs) were fed twice daily and had access to water *ab libitum*. The other animals were kept on separate pasture paddocks and therefore fed naturally by grazing also with access to water *ab libitum*.

4.2.2 Faecal sample collection, DNA extraction and 454 pyrosequencing

Fresh faecal samples were collected from each animal placed in sterile 100mL pots and frozen at -80°C. Total bacterial genomic DNA was isolated from the faeces according to the Repeat Bead Beating plus column method (RBB+C) (Yu & Morrison, 2004). The extracted DNA was then used as a template in the V4 region PCR amplifications using a method outlined previously (O' Donnell *et al.*, 2013). Samples were sequenced with 454 Titanium technologies (Teagasc Food Research Centre, Moorepark, Ireland).

4.2.3 Sequence processing and OTU clustering

Raw sequencing reads were quality trimmed and following analysis of the pyrosequencing data was performed in Qiime as outlined previously (O' Donnell *et al*., 2013). For the classification of the reads to the species level, the most common sequences (100% identity) were chosen from each OTU cluster as a representative sequence. Only unique species classifications were accepted if the following criteria were met with (a) if a representative sequence aligned with equal percentage identity (b) length and (c) had a blast score to a single species. If a representative sequence had a blast score to 2 or more species the sequence remained unclassified. To define a

core taxa the following criteria were used (a) present at ≥0.1% of total reads and (b) present in > two digestion types or 5 animal species.

The median read proportions at each taxon level for each individual animal species were pooled to form the animal species datasets. The median proportions of each animal species were then pooled to generate the three digestion type datasets. VENNY, an online Venn diagram tool was used to create a figure representing the core genera (Oliveros, 2007).

4.2.4 Alpha and beta diversity matrices

Rarefaction was performed to remove any bias in diversity estimation that might have been present due to uneven sample sizes. Five alpha diversity metrics were calculated to measure the microbial diversity in the three digestion types and in each animal species. Three of these metrics (Shannon index, OTU count and Phylogenetic diversiy) were previously described (O' Donnell *et al*., 2013). Each metric was calculated from a rarefied OTU table consisting of sub-samples of 2,440 reads per sample. Simpson's Index (D) measures the probability that two individuals randomly selected from a sample will belong to different OTUs. Good's coverage (ESC) was estimated using the formula ESC = 1 - n/N, where n= number of singleton OTUs and N=number of assigned reads. Rarefaction plots were generated in R (version 2.13.1) using a collated alpha diversity table imported from Qiime. The curve for each sample was based upon the calculated alpha diversity metric for sub-samples ranging from 100 to 2,440 reads at increments of 100 reads. A second sub-set of 10,000 reads was also used to generate rarefaction curves, to plot the alpha diversity in the hindgut fermenters (n=8) and ruminants (n=6). Each animal chosen as a representative of its digestion type had read assignments greater than 10,000 reads.

Beta diversity is the difference in diversity between one community and another and the method used in this study was described previously (O' Donnell *et al.*, 2013a). Beta diversity was calculated using weighted and un-weighted Unifrac distances in Qiime.

4.2.5 Statistics

The Mann-Whitney test (Siegel, 1956) was used for all pair-wise comparisons in this study and, in cases where multiple correction of p-values were necessary, Benjamini-

Hochberg (Benjamini and Hochberg, 1995) was used. Before statistics were carried out on the data, each group of taxa from phylum to species was filtered for those that were present in 50% of samples or greater; this ensured that the number of zero values was not heavily biased in one group over the other, which would lead to inaccurate p-values. Statistics were only performed on groups where the sample size was >=4; this was true for comparison of the 3 digestion groups (mono-gastric animals were omitted for low sample size) and also for comparison of the 10 animal groups.

4.3 Results

We used 16S rRNA gene (V4 region) amplicon pyrosequencing to determine the faecal microbiota composition of ten animal species totalling 42 animals (having removed two of the porcine datasets due to low read counts). The total number of reads identified following filtering and chimeric identification was 560,957 bp. The read numbers for each animal species ranged from 10,837 - 220,774 and a full list of the read counts for each sample animal species is presented in Table 4.1. The average read length calculated from the total reads identified was 207 bp. Assignments to the Bacterial kingdom accounted for a median 96% of the total reads in each animal with a median 0.01% of the reads assigned to the Archaea . At each taxon level only the *Equidae* animals (donkeys and miniature ponies) and sheep had *Archaea,* consisting of *Methanocorpusculum* and *Methanobrevibacter*. The remaining phylum level reads were uncharacterised read assignments (between 3-4% for the three digestion types).

Table 4.1. The animals, diets and total reads used in this study

Animals	Binomial nomenclature	Abbrev.	n	Digestion	Feed consumed	Total reads
Chinchillas	*Chinchilla lanigera*	Ch	3	Hindgut fermenter	Commercial feed[a]	37,013
Rabbits	*Oryctolagus cuniculus*	Ra	8	Hindgut fermenter	Commercial feed[b]	74,963
Donkeys	*Equus africanus asinus*	Do	7	Hindgut fermenter	Grass	220,774
Miniature ponies	*Equus ferus caballus*	MP	7	Hindgut fermenter	Grass	46,884
Deer	*Cervus nippon*	De	4	Ruminant	Grass	32,635
Goats	*Capra aegagrus hircus*	Go	5	Ruminant	Grass	27,791
Sheep	*Ovis aries*	Sh	4	Ruminant	Grass	43,559
Llamas	*Lama glama*	Ll	2	Ruminant	Grass	52,461
Alpacas	*Vicugna pacos*	Al	1	Ruminant	Grass	10,837
Pigs	*Sus scrofa scrofa kunekune*	Pi	2	Monogastric	Sow pellets and bread	14,040

[a] Dehydrated grass pellets, alfalfa pellets, chopped alfalfa hay, flaked field Peas, flaked corn, vitamins and minerals.

[b] Dry grass flaked maize, carrots, corn and oat grains supplemented with additional carrots

4.3.1 Dominant taxa in the animal species intestinal microbiota

The predominant phyla identified in the three digestion types and 10 animal species were *Firmicutes* and *Bacteroidetes*. The abundance of the *Firmicutes* phyla was significantly higher ($P \leq 0.05$) in the ruminants compared to the hindgut fermenters. The dominant taxa level assignments for each digestion type are listed in Table 4.2**.** The dominance of the *Firmicutes* phylum in the microbiota of domesticated herbivores was reflected in the other predominant taxa identified (*Clostridia* > *Clostridiales* > *Ruminococcaceae* > *Sporobacter*). *Actinobacteria* was identified as a dominant phylum in the microbiota of rabbits. The predominance of this phylum in the rabbit microbiota was seen throughout other taxa level data (*Actinobacteria* > *Bifidobacteriales* > *Bifidobacteriaceae* > *Bifidobacterium*). The dominance of *Betaproteobacteria* in the chinchilla microbiota was the single host animal-specific class identified in this study. Host animal-specific dominant orders included *Burkholderiales* (chinchillas) and *Verrucomicrobiales* (rabbits and sheep). Host animal-specific families identified included *Marinilabiaceae* (donkeys and miniature ponies), *Chitinophagaceae* (deer) and *Moraxellaceae* (llamas). The predominant

genus in the faecal microbiota of the mono-gastric animal was Treponema. Host animal-associated dominant genera were identified in the chinchillas (*Parabacteroides* and *Barnesiella*), rabbits (*Persichirhabdus* and *Subdoligranulum*), donkeys (*Anerophaga*), llamas (*Hydrogenoanaerobacterium* and *Acinetobacter*) and alpacas (*Roseburia*). *Galbibacter* and *Clostridium* were identified as dominant genera in the equids and camelids, respectively. Statistically significant differences in the taxa proportions between ruminants and hindgut fermenter microbiota are given in Table S4.2.

4.3.2

Table 4.2. Dominant taxa percentage proportions identified from the three digestion types microbiota calculated from total reads

	Digestion		
Taxa	Hindgut	Ruminant	Mono-gastric
Phylum			
Firmicutes	53.11	65.35	52.27
Bacteroidetes	31.36	20.95	26.95
Verrucomicrobia	2.90	1.24	0.54
Spirochaetes	1.93	0.91	10.34
Proteobacteria	1.68	1.52	3.44
Class			
Alphaproteobacteria	0.23	0.45	0.12
Bacilli	0.37	0.12	1.08
Bacteroidia	8.26	10.67	7.37
Clostridia	45.91	62.65	48.83
Deltaproteobacteria	0.18	0.37	0.47
Erysipelotrichi	1.17	0.86	1.38
Flavobacteria	4.60	0.75	2.26
Sphingobacteria	2.15	4.96	3.33
Spirochaetes	1.93	0.91	10.34
Subdivision5	1.07	0.10	0.31
Order			
Bacteroidales	8.26	10.67	7.37
Clostridiales	44.09	60.73	48.31
Erysipelotrichales	1.17	0.86	1.38
Flavobacteriales	4.60	0.75	2.26
Sphingobacteriales	2.15	4.96	3.33
Spirochaetales	1.93	0.91	10.34
Subdivision5	1.07	0.10	0.31
Family			
Bacteroidaceae	0.36	1.85	0.32
Clostridiaceae	0.27	0.44	0.43
Erysipelotrichaceae	1.17	0.86	1.38
Eubacteriaceae	0.28	0.23	0.65
Flavobacteriaceae	3.40	0.64	1.69
Incertae Sedis XIV	0.50	0.20	0.78
Lachnospiraceae	6.84	5.26	3.30
Porphyromonadaceae	2.10	3.73	3.06
Prevotellaceae	2.09	1.41	2.93
Ruminococcaceae	20.48	33.46	23.97
Sphingobacteriaceae	1.97	0.55	2.44
Spirochaetaceae	1.87	0.82	10.34
Veillonellaceae	0.82	0.76	2.88
Genus			
Anaerosporobacter	0.15	0.11	0.11
Acidaminococcus	0.33	0.10	0.30
Hydrogenoanaerobacterium	0.18	0.34	0.31
Bacteroides	0.36	1.85	0.32
Clostridium	0.16	0.28	0.33
Anaerotruncus	0.35	0.37	0.46
Acetivibrio	0.93	1.25	0.60
Eubacterium	0.18	0.19	0.63
Blautia	0.50	0.20	0.78
Butyricicoccus	0.13	0.24	0.80
Coprococcus	0.42	0.89	0.82
Papillibacter	0.45	1.65	0.93
Oscillibacter	0.71	1.55	1.74
Prevotella	0.91	0.36	2.38
Faecalibacterium	1.10	0.34	2.92
Ruminococcus	2.29	1.78	2.98
Sporobacter	3.63	5.05	4.34
Treponema	1.87	0.82	10.33

Core microbiota of domesticated herbivores

The *Firmicutes*, *Bacteroidetes*, *Verrucomicrobia*, *Spirochaetes* and *Proteobacteria* were identified as the core phyla in the faecal microbiota of the domesticated herbivores. These five phyla were also noted as the dominant phyla in each animal species (Table 4.2). Eighteen core genera were identified between the three digestion types (Figure S4.1). *Acidaminobacter*, *Anaerophaga*, *Dorea*, *Fibrobacter*, *Lactobacillus*, *Subdoligranulum* and *Parabacteroides* were recognised as core hindgut fermenter-associated genera. *Acetanaerobacterium*, *Acetitomaculum*, *Croceibacter*, Holdemania, *Lutispora*, *Persicirhabdus* and *Victivallis* were identified as core ruminant microbiota-associated genera. Mono-gastric microbiota-associated core genera identified in this study were *Bulleidia*, *Catenibacterium*, *Herspellia*, *Lysinibacillus*, *Megasphaera*, *Parasporobacterium*, *Petrimonas* and *Pseudomonas*. *Akkermansia*, *Alistipes*, *Paludibacter*, *Paraprevotella*, *Robinsoniella* and *Roseburia* were recognised as the six additional genera forming the core microbiota of the hindgut and ruminants only.

Thirty-three genera were identified as forming the core microbiota of domesticated herbivores and are given in Table 4.3. The majority of the genera forming the core microbiota were identified as members of the *Clostridia* class.

Table 4.3. Core genera percentage proportions from the domesticated herbivore animal species identified from the total reads

	Animals									
Genus	Chinchillas	Rabbits	Donkeys	Miniature ponies	Deer	Goats	Sheep	Llama	Alpaca	Pigs
*Acetivibrio**	0.89	0.32	1.05	0.99	1.17	1.06	1.70	1.65	1.37	0.60
Acidaminobacter	0.05	0.22	0.18	0.10	0.36	0.00	0.17	0.51	0.00	0.05
*Acidaminococcus**	0.19	0.10	0.36	0.42	0.10	0.11	0.08	0.24	0.55	0.30
Akkermansia	0.00	0.25	0.83	0.02	0.48	0.37	0.28	0.41	0.04	0.00
Alistipes	0.52	0.40	0.07	0.02	4.51	5.43	2.24	0.25	0.06	0.02
*Anaerosporobacter**	0.04	0.10	0.17	0.23	0.05	0.10	0.11	0.13	0.35	0.11
*Anaerotruncus**	0.17	0.23	0.61	0.47	0.52	0.39	0.51	0.16	0.19	0.46
*Bacteroides**	3.38	2.29	0.11	0.16	2.39	3.14	2.00	1.37	0.92	0.32
*Blautia**	0.61	0.99	0.26	0.43	0.13	0.22	0.20	0.38	0.84	0.78
*Butyricicoccus**	0.29	0.22	0.07	0.12	0.18	0.29	0.18	0.28	0.21	0.80
*Clostridium**	0.00	0.01	0.38	0.16	0.32	0.16	0.20	1.80	1.67	0.33
*Coprococcus**	0.58	0.35	0.41	0.44	0.95	0.96	0.80	0.97	1.05	0.82
Dorea	0.24	0.42	0.04	0.08	0.12	0.11	0.07	0.09	0.29	0.01
*Eubacterium**	0.19	0.62	0.06	0.12	0.43	0.22	0.14	0.12	0.27	0.63
*Faecalibacterium**	1.21	2.77	0.50	1.04	0.21	0.80	0.56	0.42	0.51	2.92
Galbibacter	0.00	0.00	3.51	5.56	0.00	0.02	0.14	0.60	0.83	1.02
Holdemania	0.01	0.02	0.04	0.13	0.16	0.29	0.10	0.20	0.11	0.01
*Hydrogenoanaerobacterium**	0.75	0.09	0.18	0.26	0.59	0.27	0.18	1.24	0.48	0.31
Lactonifactor	0.02	0.00	0.14	0.29	0.04	0.30	0.51	0.12	0.07	0.01
Lutispora	0.00	0.00	0.00	0.01	0.36	0.31	0.32	0.01	0.04	0.01
Oribacterium	0.05	0.03	0.05	0.18	0.00	0.12	0.10	0.26	0.23	0.04
*Oscillibacter**	0.58	0.30	1.38	0.98	1.30	1.74	1.95	1.57	1.42	1.74
*Paludibacter**	0.00	0.00	0.96	0.68	3.44	1.53	0.27	1.36	1.45	0.06
Papillibacter	1.10	0.22	0.65	0.52	1.60	2.11	2.32	1.27	1.14	0.93
Parabacteroides	2.37	0.49	0.24	0.00	0.06	0.11	0.10	0.19	0.27	0.10
Paraprevotella	0.37	0.09	0.23	2.34	0.14	0.63	0.38	0.64	0.96	0.04
Persicirhabdus	0.00	2.96	0.03	0.00	0.21	0.18	0.77	0.14	0.00	0.00
*Prevotella**	1.28	0.53	0.89	1.49	0.10	1.42	0.44	0.54	0.39	2.38
Robinsoniella	0.43	0.01	0.29	0.23	0.35	0.30	0.15	0.16	0.08	0.09
Roseburia	0.27	0.30	0.25	0.89	0.33	0.26	0.14	0.14	1.03	0.00
*Ruminococcus**	5.65	14.23	1.17	1.67	1.63	1.64	1.79	2.34	1.61	2.98
*Sporobacter**	0.63	4.29	3.42	5.02	5.15	5.09	4.67	4.47	2.88	4.34
*Treponema**	0.03	0.14	6.55	2.02	1.24	0.85	0.52	2.78	6.51	10.33

* The 18 core genera identified from the three digestion types.

4.3.3 Digestion type and animal host-specific species identified in the animals studied

Fifty-nine species were identified between the 10 animal species used in this study and are listed in Table 4.4. The dominant bacterial species were recognised as being digestion type-specific. The hindgut fermenters dominant faecal bacterial species were *Lactobacillus ruminis*, *Ruminococcus albus* and *Clostridium bifermentans*. The dominant ruminant bacterial species recognised were *Clostridium septicum*, *L. ruminis* and *Butyvibrio hungatei*. The dominant species present in the faecal samples

of the mono-gastric animal were *Cl. septicum*, *Butyricoccus pullicaecorum* and *Tisserella praecuta*. However, only *L. ruminis* was identified in all three digestion types at greater than 0.03% of the total reads.

Host digestion type-specific bacterial species identified included *Butyvibrio hungatei* present in the microbiota of the ruminant animals only. Similarly, *Bacteroides intestinalis*, *Bifidobacterium breve*, *Ruminococcus albus* and *Ruminococcus flavefaciens* were identified in the faecal microbiota of the hindgut fermenters only. The majority of digestion type-specific bacterial species were associated with the faecal samples from the mono-gastric kune-kune pigs. These species include *Aequorivita capsosiphonis*, *Aerosphaera taetra*, *Brumimicrobium mesophilum*, *Clostridium neonatale*, *Erysipelothrix rhusiopathiae*, *Escherichia coli*, *Fibrobacter intestinalis* and *Paracoccus alcaliphilus*.

Table 4.4. Species level assignment percentages calculated from the total reads for the animals used in this study

	Animals									
Taxa	Chinchilla	Rabbits	Donkeys	Miniature ponies	Deer	Goats	Sheep	Llama	Alpaca	Pigs
Acidaminococcus fermentans				0.013	0.013			0.001	0.009	
Actinobacillus capsulatus								0.006		
Actinomyces hyovaginalis								0.044		
Aequorivita capsosiphonis										0.021
Aeromicrobium kwangyangensis								0.006	0.028	
Aerosphaera taetra		0.017						**0.219**		0.012
Ahrensia kielensis								0.001		
Anaerovibrio lipolyticus						0.014		0.025		
Antarctic bacterium									0.009	
Bacillus insolitus								0.006		
Bacteroides intestinalis	0.064	**0.407**	0.001							
Bifidobacterium breve	0.008	**1.157**	0.002					0.008	0.009	
Bifidobacterium magnum		**0.358**								
Brachybacterium arcticum		0.007								
Brumimicrobium mesophilum										0.008
Butyricicoccus pullicaecorum				0.017	0.023		0.007	0.060	0.083	**0.268**
Butyricimonas virosa	0.013							0.014		
Butyrivibrio fibrisolvens								0.006		
Butyrivibrio hungatei						0.073	**0.164**	**0.175**		
Campylobacter cuniculorum		0.086								
Clostridium aldenense	**1.961**	0.010						0.001		
Clostridium bifermentans	0.013		0.026	0.035	0.035	0.016	0.003	0.017	0.055	0.025
Clostridium lavalense								0.003		
Clostridium neonatale					0.003			0.007		0.008
Clostridium septicum			**0.188**	0.010	**0.311**	**0.125**	**0.148**	0.060	**0.138**	**0.525**
Clostridium tetani									0.009	
Coprococcus catus			0.001	0.070						0.008
Erysipelothrix rhusiopathiae		0.028						**0.494**		0.006
Escherichia coli			0.018					0.010	0.055	**0.230**
Eubacterium cellulosolvens						0.011				
Fibrobacter intestinalis								0.015	0.009	0.019
Fibrobacter succinogenes			0.006			0.037	**0.136**			
Glaciibacter superstes								0.001		
Lactobacillus plantarum										0.008
Lactobacillus ruminis			**0.166**	**0.130**	0.025	0.025	0.016	0.040	0.065	0.066
Leucobacter chironomi									0.009	
Leuconostoc pseudomesenteroides										0.006
Mesorhizobium thiogangeticum								0.002		
Methylobacterium komagatae								0.001		
Oerskovia turbata								0.007		0.015
Oscillibacter valericigenes								0.001		
Paenibacillus contaminans										0.006
Paracoccus alcaliphilus								0.002	0.009	0.017
Paraprevotella xylaniphila		0.027								
Pasteurella caballi							0.003			
Pseudaminobacter salicylatoxidans		0.003						0.007	0.074	0.046
Rhizobium leguminosarum									0.028	
Rhizobium radiobacter								0.002	0.046	
Ruminococcus albus		**7.717**	0.044	0.013						
Ruminococcus bromii				**0.402**						
Ruminococcus flavefaciens	**0.171**	0.010	0.007			0.007		0.003		
Ruminococcus gauvreauii										
Selenomonas ruminantium						0.007	0.018	0.025		
Sharpea azabuensis	0.004		0.007							
Sphingobacterium anhuiense	0.013	0.017								0.006
Sphingobacterium mizutaii		0.016								
Sphingoterrabacterium composti								0.006		
Tissierella praeacuta										**0.260**

Table 4.4 Note: Values highlighted in bold were those present in the animal's microbiota at greater than 0.1% of the total reads. Blank spaces denote values equal to zero or values equal to zero in at least 50% of the individuals in an animal species.

Some bacterial species identified were noted as having an affiliation with a particular host animal. *Bifidobacterium magnum* and *Campylobacter cuniculorum* were identified in the rabbit faecal samples only. *Clostridium aldenense* was not identified as a host-specific bacterial species however, it did account for 1.96% of the total reads from the chinchilla faecal samples. *Pasteurella caballi* and *Actinomyces hyovaginalis* were identified from the faecal microbiota of sheep and llamas, respectively. *Ruminococcus bromii* and *Eubacterium cellulosolvens* were identified from the faecal samples of the miniature ponies and goats, respectively. *Aeromicrobium kwangyangensis* and *Rhizobium radiobacter* were recognised as camelid/pseudoruminant-associated bacterial species. *Aequorivita capsosiphonis* was identified as a unique species in the microbiota of the kune-kune pigs.

4.3.4 Bacterial diversity estimations between digestion types and animals

Comparisons of the alpha bacterial diversities of the hindgut fermenting and ruminant digestion types revealed that the diversity of the ruminant microbiota was larger than the hindgut fermenters. The microbiota diversity estimated in the ruminants was significantly higher than the hindgut fermenters using the Shannon diversity and OTU counts ($P<0.01$ and $P<0.05$, respectively) indices. The rarefaction curves generated from the 2,440bp subset of the populations are shown in Figure S4.2. The phylogenetic diversity and OTU count curves failed to reach a saturation plateau for any of the digestion types/animals which suggests that the sampling in this study failed to encompass the true microbial diversity of each animal. However, both the Shannon diversity and Simpson diversity indices plots did reach a plateau, suggesting that further sampling would not yield additional phylotypes. The Goods coverage metric was used to estimate the completeness of sampling with median coverage percentages of 90 to 96%. The Goods coverage percentages for each sample also indicate that, like the Shannon and Simpson diversity indices, further microbiota sampling would result in a small number of additional phylotypes.

A summary of the alpha diversity indices results generated for the individual animal species are given in Table 4.5. The alpha bacterial diversity indices from individual animal species revealed that the donkey microbiota was the most diverse of the animal species studied and that the rabbit microbiota was the least diverse. This difference between animals with a similar digestion type may be due to the relative size of the animals and the longer gut retention times of the equids.

Table 4.5. Alpha diversity metrics in the different animals groups

Digestion type	Animal species	Diversity metrics					
		Phylogenetic Diversity	Shannon Weaver	Simpson index	Chao1 score	Observed species	Goods coverage (RSD%)
Hindgut fermenters	Chinchillas	38.97	6.78	0.973	1027	445	95% (4.01)
	Rabbits	41.03	6.43	0.962	865	415	95% (4.26)
	Donkeys	62.66	7.82	0.988	1278	606	92% (5.08)
	Miniature ponies	49.08	7.37	0.987	910	472	91% (2.88)
Ruminants	Deer	52.01	7.69	0.979	1112	614	90% (2.38)
	Goats	57.78	8.00	0.987	1262	660	87% (3.48)
	Sheep	58.22	7.95	0.986	1144	645	92% (2.82)
	Llamas	54.78	7.43	0.982	1028	542	95% (4.42)
	Alpaca	48.87	7.22	0.979	844	472	97% (N/A)
Monogastric	Pigs	48.22	6.98	0.976	739	434	94% (0.74)

N/A – single animal therefore we were unable to calculate relative standard deviation

4.3.5 Clustering of the intestinal microbiota by digestion type and host phylogeny

Unifrac un-weighted and weighted principle coordinate analysis (PCoA) plots were used to visualise and examine the beta-diversity of both the digestion types and the animal species, the plots of each are displayed in Figure 4.1. The low variance explained by the first two axes (27.7%) in the un-weighted plots is common when explaining the variance when many diverse factors may affect the samples. The first two axes in the weighted plots accounted for 48.7% of the variance.

The un-weighted and weighted PCoA plots showed a clustering of bacteria within each microbiota by the digestion type (Figure 4.1 (a) and (c)). However, there was an overlap noted with the samples from the mono-gastric animal species (pig) with those from the hindgut fermenters in the weighted PCoA plot (Figure 4.1 (c)). The

weighted PCoA animal species microbiota plots (Figure 4.1 (d)) showed a clustering of the microbiota of each animal species based on their Family as well as digestion type. Groupings include the *equidae* (donkeys & miniature ponies; hindgut fermenters), *camelidae* (llama & alpaca; ruminants/pseudo-ruminants) and *bovidae* (sheep & goats; ruminants). The remaining animal species microbiota appear to cluster based on the digestion type and Order (*Artiodactyla*). This suggests that host phylogeny and therefore, digestion type may predetermine the microbiota of the herbivorous domesticated animals studied. However, it should be noted that digestion type and host phylogeny are not independent of each other and closely related animal species are more likely to share the same digestive strategy (for example, goats and sheep).

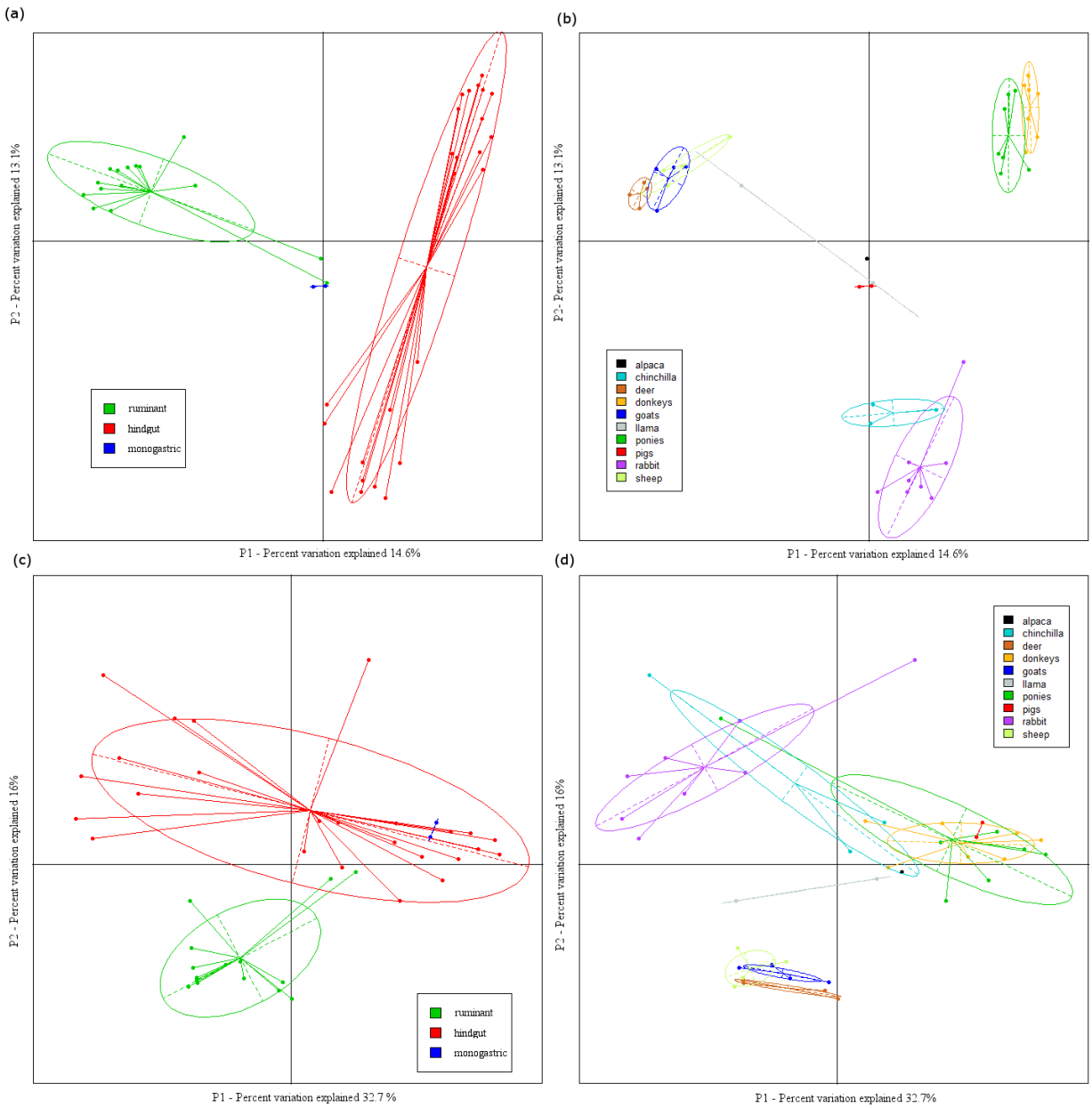


Figure 4.1. Unifrac beta diversity measures (a) un-weighted plot for the microbiota of three digestion types (b) un-weighted plot for the microbiota of the 10 animal species (c) weighted plot for the microbiota of the three digestion types (d) weighted plot for the microbiota of the 10 animal species.

4.4 Discussion

The objective of this study was to identify the bacterial taxa present in hindgut fermenters and ruminant animals dwelling on a single Irish farm. In the current study, the composition and diversity of the faecal microbiota of the chinchilla, rabbit, donkey, llama and alpaca was elucidated for the first time. This is the first study also to investigate the microbiota of a number of animals with different types located on the same farm. The co-localisation of each test animal removes the geographic, management regime and diet differences noted in other studies (O' Donnell *et al.*, 2013; Shanks *et al.*, 2011; Yamano *et al.*, 2008). In the current study, we showed that the domesticated herbivorous animals shared a core microbiota but that some genera were associated with particular digestion types only.

In this study, *Firmicutes* was identified as the predominant phyla in the microbiota of hindgut fermenters and ruminants. However, the examination of the rumen microbiota of North American moose (*Alces alces*) revealed that the *Bacteroidetes* (27%) was the predominant phyla present and the *Firmicutes* phylum was only present at low levels (4%) in the moose microbiota (Ishaq & Wright, 2012). Sika deer (*Cervus nippon*) is a native of East Asia, introduced to European countries 150 years ago, and in the wild favours foraging in forested areas (Bartoš, 2009). Previous analysis of four domesticated sika deer consuming two different diets (oak leaves and corn stalks) revealed that in both diet treatments *Bacteroidetes* predominated the rumen samples (Li *et al.*, 2013). In our study however, the *Firmicutes* phyla was predominant in the faecal samples of the sika deer. The differences in the dominant phyla in both the moose and deer populations may be a result of the different diets consumed, PCR amplification bias or due to the DNA extraction methods employed (Henderson *et al.*, 2013).

The potential effects of the different diets is also reflected in the genera with *Prevotella* identified as the dominant genus in the sika deer population (Li *et al.*, 2013). However, in this study *Sporobacter* was the dominant genus identified in the faecal microbiota of the sika deer. Both miniature ponies/horses (*Equus ferus caballus*) and donkeys (*Equus africanus asinus*) are members of the *Equidae* family. We recently characterised the faecal microbiota of the healthy Irish Thoroughbred racehorse and identified *Firmicutes* as the predominant phylum in the faecal

microbiota irrespective of feed consumed (O' Donnell *et al.*, 2013). In this study, similar proportions of this phylum was identified in both equid animal species. However, the proportions of the *Bacteroidetes* phylum was higher in this study than our previous work on grass fed horses (O' Donnell *et al.*, 2013). The difference between our two studies may be due to the different extraction methods used to generate the data or primer bias (Berry *et al.*, 2011; Henderson *et al.*, 2013). When examining the effects that domestication can have on an animal species De Jesus-Laboy *et al.*, 2012 noted that the *Actinobacteria* phylum was present in all of the domesticated goats (De Jesús-Laboy *et al.*, 2012). However, in our study, the *Actinobacteria* phylum was absent from the domesticated pygmy goat dataset. Instead, the *Actinobacteria* phylum was associated with the hindgut-fermenting animals and in particular the rabbits. The *Prevotellaceae* were also found to be associated with the feral goat microbiota (De Jesús-Laboy *et al.*, 2012) and were also found as a dominant family in the pygmy goats in this study. This may be due in part to both animals grazing naturally on plant matter.

The dominant families identified in the hindgut fermenter and ruminant animals are consistent with those identified in other large ruminants and hindgut fermenters (Bhatt *et al.*, 2013; Bian *et al.*, 2013; Thoetkiattikul *et al.*, 2013). Many genera have been identified as important rumen-associated bacteria involved in the digestion of plant polysaccharides. Important rumen-associated polysaccharide degrading bacteria include *Ruminococcus, Prevotella, Butyvibrio, Alistipes* and *Succniniclasticum* (Dowd *et al.*, 2008; Kim *et al.*, 2011) and were identified in this study. However, in our study only the *Succiniclasticum* and *Butyvibrio* genera was associated with the microbiota of ruminants only, but at very low proportions (<0.2%). We previously identified *Ruminococcus, Sporobacter* and *Treponema* as dominant genera in the microbiota of grass-fed Thoroughbred racehorses, while other studies have identified *Prevotella, Oscillibacter, Faecalibacterium, Coprococcus* and *Butyricicoccus* and *Blautia* as important hindgut fermenter-associated genera (O' Donnell *et al.*, 2013a; Yildirim *et al.*, 2010). *Fibrobacter* was also identified as an important genus particularly for the hindgut-fermenting equids; this is consistent with previous studies (Shepherd *et al.*, 2012). Alpacas (*Lama pacos*) are camelids native to South America and like other ruminants rely on bacteria to aid in the digestion of their food (Van Saun, 2006). *Eubacterium* spp. was identified as predominant genus

in the alpaca forestomach microbiota. However, in our study the predominant genera in the Alpaca faecal samples were *Treponema*, *Sporobacter* and *Clostridium*. The difference in the sample regions examined in both studies may be an explanation for the differences in the predominant genera identified in both studies (Dougal *et al.*, 2013; Pei *et al.*, 2010).

The proportion of unclassified reads identified at the genus level in this study is consistent with other studies carried out on humans, hindgut fermenters and less commonly studied ruminants (Claesson *et al.*, 2009; Janssen & Kirs, 2008; O' Donnell *et al.*, 2013a). The high percentage of unclassified read proportions in this and other studies is due to the lack of culturing and sequence identification work on the more obscure hindgut fermenters and ruminants (Pei *et al.*, 2010).

Lactobacillus ruminis, a potentially probiotic autochthonous commensal bacteria in the microbiota of humans and animals, was identified in eight of the animal species examined (Reuter, 2001). The proportion of the *L. ruminis* species in the equids was higher than those seen in any of the ruminant animals. We have identified approximately 7% of lactic acid bacteria in the faecal samples of a horse as *L. ruminis* (data not shown). *Clostridium septicum*, an opportunistic pathogen, was also identified in eight of the animal species (Koransky *et al.*, 1979; Songer, 1996). *Clostridium bifermentans* a species previously identified in other ruminant animals (Princewell & Agba, 1982) but not hindgut fermenters (O' Donnell *et al.*, 2013) was identified in both digestion types in this study. To our knowledge the function of both clostridial species in the microbiota of animals is unknown. Some of the species identified were associated with primarily one specific animal species. *Bifidobacterium magnum* and *Campylobacter cuniculorum* were both identified in the faeces of rabbits. Both of these species have previously been identified and cultured from rabbit faeces (Scardovi & Zani, 1974; Zanoni *et al.*, 2009). However, in this study we also identified both species in two faecal samples from donkeys. This is the first study to identify these species in the donkey microbiota. In this study, *Butyrivibrio hungatei*, a butyrate-producing rumen bacteria was identified only in the true ruminant animals (Kopečný *et al.*, 2003). This suggests a strong association between this *B. hungatei* and the rumen's fibrous digesta.

Methanogenic archaea are important members of the rumen microbiota and facilitate the removal of hydrogen, generated by the fermentation of plant material, from the GIT. However, only a small proportion of the prominent archaeal genera, *Methanobrevibacter* and *Methanocorpusculum* were identified in this study (Bian *et al.*, 2013). A recent study, showed that the DNA extraction method used can have an effect on the archaea identified in a sample with *Methanobrevibacter* (identified here in the pigs at ≤0.01% of the total reads) as the only genus not effected by the extraction method (Henderson *et al.*, 2013).

Less phylotype diversity was measured in the hindgut fermenter, ruminant and monogastric animal microbiota compared to data from the distal bowel microbiota of other animals (Lamendella *et al.*, 2011; Pitta *et al.*, 2010). Our phylotype estimations for the animal species (415 - 660) were within the ranges estimated for the human microbiota (Claesson *et al.*, 2009; Nam *et al.*, 2011) but lower than our previous equine microbiota estimates (O' Donnell *et al.*, 2013). The Chao1 diversity ranged from 844-1278 and was similar to those seen for humans, cows and other hindgut fermenters (Bian *et al.*, 2013; Claesson *et al.*, 2009; Pitta *et al.*, 2010). The failure of the OTU count and phylogenetic diversity rarefaction curves to plateau indicated that complete sampling of the domesticated herbivore faecal microbiota has not yet been achieved. This result indicates that further sampling is required to truly reflect the diversity present in the microbiota of hindgut fermenters and ruminants. The Shannon diversity index results for all animals studied was higher than those seen in other animals and humans (Claesson *et al.*, 2009; Lamendella *et al.*, 2011; Nam *et al.*, 2011; Pitta *et al.*, 2010) but consistent with those seen in our previous study on horses (O' Donnell *et al.*, 2013). The diversity indices used in this study indicated that while the ruminant bacteria are more diverse than their hindgut fermenting counterparts but compared to other animals sequencing studies they are less diverse. Good's coverage ranged from 90-96% for the animal species, indicating that 10-28 additional reads would need to be sequenced to detect a new phylotype. This level of coverage indicates that the 16S rDNA V4 sequences identified in these samples represent the majority of bacterial sequences present in the domesticated herbivore microbiota. The Good's coverage estimates are consistent with those seen for humans, hindgut-fermenting mammals and larger than for some ruminants (Berry *et al.*, 2011; Janssen & Kirs, 2008; Nam *et al.*, 2011). However, there are caveats to

bear in mind when comparing and interpreting the differences in the diversity present in a particular microbiota or study. Each study may be affected by the method used to generate the data and assignments (Kemp & Aller, 2004).

The Kune-kune pigs used in this study are considered to be primarily herbivores. It is possible that the Kune-kune pig faecal microbiota may be closer to those of other mono-gastric herbivores like the hindgut fermenters. This may account for the overlap of the monogastric animals with the hindgut fermenters in the weighted PCoA beta diversity plot. The weighted PCoA plots (which include proportional data) displayed the animal species microbiota clustered by their families and digestion type, with the true ruminant animals (deer, goats and sheep) clustered by digestion type and Order. Ley and colleagues identified the herbivorous microbiota as the most diverse when compared to omnivores and carnivores (Ley *et al.*, 2008). Our study, expanded on this by focusing only on herbivorous animals and within these parameters we noted that the ruminant faecal microbiota is more diverse than the hindgut microbiota.

In conclusion, we have shown that the hindgut fermenting, ruminant and monogastric microbiota share 50% of their phyla and over 15% of their genera forming a core domesticated herbivore microbiota. This degree of overlap between the microbiota of the 10 animal species may suggest that these genera are essential for all herbivorous fibrous polysaccharide-consuming animals. Host phylogeny and digestion method were shown to be key determinants of bacterial diversity in the domesticated herbivores. Further studies in larger multi-animal farms in other countries would further allow us to identify other determinants shaping the diversity in the animal microbiota.

4.5 References

Abdel-Magied, E. M. & Taha, A. A. M. (2003). Morphological, morphometric and histochemical characterization of the gastric mucosa of the camel (*Camelus dromedarius*). *Anat Histol Embryol* **32**, 42-47.

Bergman, E. N. (1990). Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. *Physiol Rev* **70**, 567-590.

Berry, D., Ben Mahfoudh, K., Wagner, M. & Loy, A. (2011). Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. *Appl Environ Microbiol* **77**, 7846-7849.

Bhatt, V., Dande, S., Patil, N. & Joshi, C. (2013). Molecular analysis of the bacterial microbiome in the forestomach fluid from the dromedary camel (*Camelus dromedarius*). *Mol Biol Rep* **40**, 3363-3371.

Bian, G., Ma, L., Su, Y. & Zhu, W. (2013). The Microbial Community in the Feces of the White Rhinoceros (*Ceratotherium simum*) as Determined by Barcoded Pyrosequencing Analysis. *PloS one* **8**, e70103.

Brulc, J. M., Antonopoulos, D. A., Berg Miller, M. E. & other authors (2009). Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proc Natl Acad Sci* **106**, 1948-1953.

Callaway, T. R., Dowd, S. E., Edrington, T. S., Anderson, R. C., Krueger, N., Bauer, N., Kononoff, P. J. & Nisbet, D. J. (2010). Evaluation of bacterial diversity in the rumen and feces of cattle fed different levels of dried distillers grains plus solubles using bacterial tag-encoded FLX amplicon pyrosequencing. *J Anim Sci* **88**, 3977-3983.

Claesson, M. J., O'Sullivan, O., Wang, Q., Nikkilä, J., Marchesi, J. R., Smidt, H., De Vos, W. M., Ross, R. P. & O'Toole, P. W. (2009). Comparative analysis of pyrosequencing and a phylogenetic microarray for exploring microbial community structures in the human distal intestine. *PLOS One* **4**, e6669.

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

Costa, M. C., Arroyo, L. G., Allen-Vercoe, E., StÃ¤mpfli, H. R., Kim, P. T., Sturgeon, A. & Weese, J. S. (2012). Comparison of the fecal microbiota of healthy horses and horses with colitis by high throughput sequencing of the V3-V5 region of the 16S rRNA gene. *PLOS One* **7**, e41484.

Daly, K., Stewart, C. S., Flint, H. J. & Shirazi Beechey, S. P. (2001). Bacterial diversity within the equine large intestine as revealed by molecular analysis of cloned 16S rRNA genes. *FEMS Microbiol Ecol* **38**, 141-151.

De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J. B., Massart, S., Collini, S., Pieraccini, G. & Lionetti, P. (2010). Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *P Natl Acad Sci USA* **107**, 14691-14696.

De Jesús-Laboy, K. M., Godoy-Vitorino, F., Piceno, Y. M., Tom, L. M., Pantoja-Feliciano, I. G., Rivera-Rivera, M. J., Andersen, G. L. & Domínguez-Bello, M. G. (2012). Comparison of the Fecal Microbiota in Feral and Domestic Goats. *Genes* **3**, 1-18.

Demeyer, D. I. (1991). Quantitative aspects of microbial metabolism in the rumen and hindgut. *Rumen microbial metabolism and ruminant digestion INRA, Paris*, 217-237.

Dougal, K., Harris, P. A., Edwards, A., Pachebat, J. A., Blackmore, T. M., Worgan, H. J. & Newbold, C. J. (2013). A comparison of the microbiome and the metabolome of different regions of the equine hindgut. *FEMS Microbiol Ecol* **82**, 642-652.

Dowd, S., Callaway, T., Wolcott, R., Sun, Y., McKeehan, T., Hagevoort, R. & Edrington, T. (2008). Evaluation of the bacterial diversity in the feces of cattle using 16S rDNA bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP). *BMC Microbiol* **8**, 125.

Duncan, P., Foose, T. J., Gordon, I. J., Gakahu, C. G. & Lloyd, M. (1990). Comparative nutrient extraction from forages by grazing bovids and equids: a test of the nutritional model of equid/bovid competition and coexistence. *Oecologia* **84**, 411-418.

Ezenwa, V. O., Gerardo, N. M., Inouye, D. W., Medina, M. n. & Xavier, J. B. (2012). Animal behavior and the microbiome. *Science* **338**, 198-199.

Henderson, G., Cox, F., Kittelmann, S., Miri, V. H., Zethof, M., Noel, S. J., Waghorn, G. C. & Janssen, P. H. (2013). Effect of DNA extraction methods and sampling techniques on the apparent structure of cow and sheep rumen microbial communities. *PLOS One* **8**, e74787.

Hintz, H. F. & Cymbaluk, N. F. (1994). Nutrition of the horse. *Ann Rev Nutr* **14**, 243-267.

Ishaq, S. L. & Wright, A.-D. G. (2012). Insight into the bacterial gut microbiome of the North American moose (*Alces alces*). *BMC Microbiol* **12**, 212.

Jami, E. & Mizrahi, I. (2012). Composition and Similarity of Bovine Rumen Microbiota across Individual Animals. *PLoS ONE* **7**, e33306.

Janssen, P. H. & Kirs, M. (2008). Structure of the archaeal community of the rumen. *Applied and environmental microbiology* **74**, 3619-3625.

Kemp, P. F. & Aller, J. Y. (2004). Bacterial diversity in aquatic and other environments: what 16S rDNA libraries can tell us. *FEMS Microbiol Ecol* **47**, 161-177.

Kim, M., Morrison, M. & Yu, Z. (2011). Phylogenetic diversity of bacterial communities in bovine rumen as affected by diets and microenvironments. *Folia Microbiol* **56**, 453-458.

Kopečný, J., Zorec, M., Mrázek, J., Kobayashi, Y. & Marinšek-Logar, R. (2003). *Butyrivibrio hungatei* sp. nov. and *Pseudobutyrivibrio xylanivorans* sp. nov., butyrate-producing bacteria from the rumen. *Int J Sys Evol Microbiol* **53**, 201-209.

Koransky, J. R., Stargel, M. D. & Dowell Jr, V. R. (1979). *Clostridium septicum* bacteremia its clinical significance. *Amer J Med* **66**, 63-66.

Lamendella, R., Santo Domingo, J., Ghosh, S., Martinson, J. & Oerther, D. (2011). Comparative fecal metagenomics unveils unique functional capacity of the swine gut. *BMC Microbiol* **11**, 103.

Ley, R. E., Hamady, M., Lozupone, C. & other authors (2008). Evolution of mammals and their gut microbes. *Science* **320**, 1647.

Li, Z. P., Liu, H. L., Li, G. Y., Bao, K., Wang, K. Y., Xu, C., Yang, Y. F., Yang, F. H. & Wright, A.-D. G. (2013). Molecular diversity of rumen bacterial communities from tannin-rich and fiber-rich forage fed domestic Sika deer (*Cervus nippon*) in China. *BMC Microbiol* **13**, 151.

McCullough, D., Takatsuki, S., Kaji, K. & BartoÅ¡, L. k. (2009). Sika Deer in Continental Europe. In *Sika Deer*, pp. 573-594: Springer Japan.

Muegge, B. D., Kuczynski, J., Knights, D., Clemente, J. C., GonzÃ¡lez, A., Fontana, L., Henrissat, B., Knight, R. & Gordon, J. I. (2011). Diet drives

convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science* **332**, 970-974.

Nam, Y.-D., Jung, M.-J., Roh, S. W., Kim, M.-S. & Bae, J.-W. (2011). Comparative analysis of Korean human gut microbiota by barcoded pyrosequencing. *PLOS One* **6**, e22109.

O' Donnell, M. M., Harris, H. M. B., Jeffery, I. B., Claesson, M. J., Younge, B., O' Toole, P. W. & Ross, R. P. (2013). The core faecal bacterial microbiome of Irish Thoroughbred racehorses. *Lett Appl Microbiol* **57**, 467–560.

Oliveros, J. C. (2007).VENNY. An interactive tool for comparing lists with Venn Diagrams. In *BioinfoGP, CNB-CSIC*.

Pei, C.-X., Liu, Q., Dong, C.-S., Li, H., Jiang, J.-B. & Gao, W.-J. (2010). Diversity and abundance of the bacterial 16S rRNA gene sequences in forestomach of alpacas (*Lama pacos*) and sheep (*Ovis aries*). *Anaerobe* **16**, 426-432.

Petrosino, J. F., Highlander, S., Luna, R. A., Gibbs, R. A. & Versalovic, J. (2009). Metagenomic pyrosequencing and microbial identification. *Clin Chem* **55**, 856-866.

Pitta, D., Pinchak, W., Dowd, S. & other authors (2010). Rumen bacterial diversity dynamics associated with changing from bermudagrass hay to grazed winter wheat diets. *Microbiol Ecol* **59**, 511-522.

Playne, M. J. & Kennedy, P. M. (1976). Ruminal volatile fatty acids and ammonia in cattle grazing dry tropical pastures. *J Agric Sci* **86**, 367-372.

Princewell, T. J. T. & Agba, M. I. (1982). Examination of bovine faeces for the isolation and identification of *Clostridium* species. *J Appl Bacteriol* **52**, 97-102.

Reuter, G. (2001). The *Lactobacillus* and *Bifidobacterium* microflora of the human intestine: composition and succession. *Curr Issues Intest Microbiol* **2**, 43-53.

Scardovi, V. & Zani, G. (1974). *Bifidobacterium magnum* sp. nov., a large, acidophilic *bifidobacterium* isolated from rabbit feces. *Int J Syst Bacteriol* **24**, 29-34.

Shanks, O. C., Kelty, C. A., Archibeque, S., Jenkins, M., Newton, R. J., McLellan, S. L., Huse, S. M. & Sogin, M. L. (2011). Community structures of fecal bacteria in cattle from different animal feeding operations. *Appl Environ Microbiol* **77**, 2992-3001.

Shepherd, M. L., Swecker Jr, W. S., Jensen, R. V. & Ponder, M. A. (2012). Characterization of the fecal bacteria communities of forage fed horses by pyrosequencing of 16S rRNA V4 gene amplicons. *FEMS Microbiol Lett* **326**, 62-68.

Songer, J. G. (1996). Clostridial enteric diseases of domestic animals. *Clin Microbiol Rev* **9**, 216.

Steelman, S. M., Chowdhary, B. P., Dowd, S., Suchodolski, J. & JaneÄ • ka, J. E. (2012). Pyrosequencing of 16S rRNA genes in fecal samples reveals high diversity of hindgut microflora in horses and potential links to chronic laminitis. *BMC Vet Res* **8**, 231.

Stevens, C. E. & Hume, I. D. (2004). *Comparative physiology of the vertebrate digestive system*: Cambridge University Press.

Thoetkiattikul, H., Mhuantong, W., Laothanachareon, T., Tangphatsornruang, S., Pattarajinda, V., Eurwilaichitr, L. & Champreda, V. (2013a). Comparative analysis of microbial profiles in cow rumen fed with different dietary fiber by tagged 16S rRNA gene pyrosequencing. *Curr Microbiol* **67**, 130-137.

Thoetkiattikul, H., Mhuantong, W., Laothanachareon, T., Tangphatsornruang, S., Pattarajinda, V., Eurwilaichitr, L. & Champreda, V. (2013b). Comparative Analysis of Microbial Profiles in Cow Rumen Fed with Different Dietary Fiber by Tagged 16S rRNA Gene Pyrosequencing. *Current Microbiology* **67**, 130-137.

Tottey, W., Denonfoux, J., Jaziri, F. & other authors (2013). The Human Gut Chip "HuGChip", an explorative phylogenetic microarray for determining gut microbiome diversity at family level. *PLOS One* **8**, e62544.

Van Saun, R. J. (2006). Nutrient requirements of South American camelids: A factorial approach. *Small Ruminant Research* **61**, 165-186.

Wagner, M., Smidt, H., Loy, A. & Zhou, J. (2007). Unravelling microbial communities with DNA-microarrays: challenges and future directions. *Microbial Ecology* **53**, 498-506.

Warner, R. G., Flatt, W. P. & Loosli, J. K. (1956). Ruminant nutrition, dietary factors influencing development of ruminant stomach. *J Agric Food Chem* **4**, 788-792.

Welkie, D. G., Stevenson, D. M. & Weimer, P. J. (2010). ARISA analysis of ruminal bacterial community dynamics in lactating dairy cows during the feeding cycle. *Anaerobe* **16**, 94-100.

Yamano, H., Koike, S., Kobayashi, Y. & Hata, H. (2008). Phylogenetic analysis of hindgut microbiota in Hokkaido native horses compared to light horses. *Anim Sci J* **79**, 234-242.

Yildirim, S., Yeoman, C. J., Sipos, M. & other authors (2010). Characterization of the fecal microbiome from non-human wild primates reveals species specific microbial communities. *PLOS One* **5**, e13963.

Yu, Z. & Morrison, M. (2004). Improved extraction of PCR-quality community DNA from digesta and fecal samples. *Biotechniques* **36**, 808-813.

Zanoni, R. G., Debruyne, L., Rossi, M., Revez, J. & Vandamme, P. (2009). *Campylobacter cuniculorum* sp. nov., from rabbits. *Int J Sys Evol Microbiol* **59**, 1666-1671.

4.6 Supplementary information

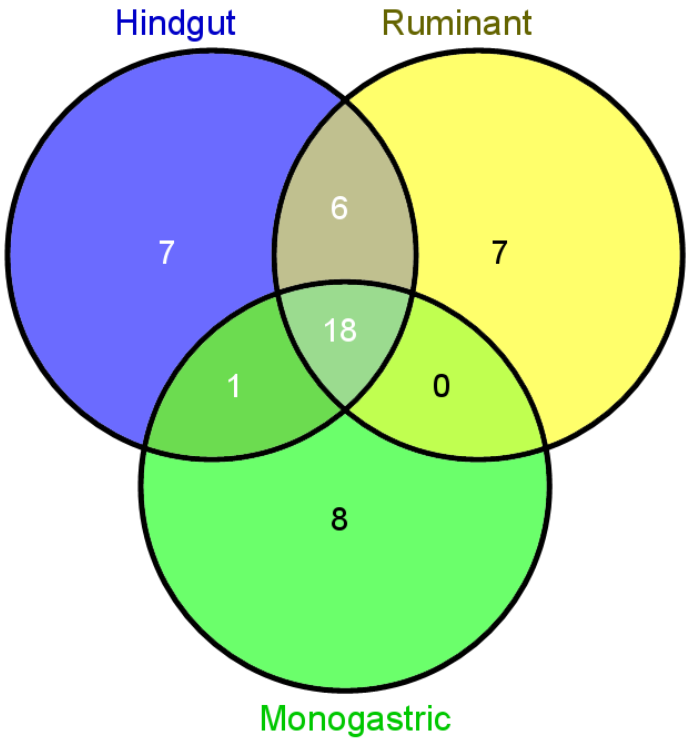


Figure S4.1. Venn diagram representation of the number of shared, core and unique genera in the microbiota of the ruminant, hindgut fermenters and monogastric animals.

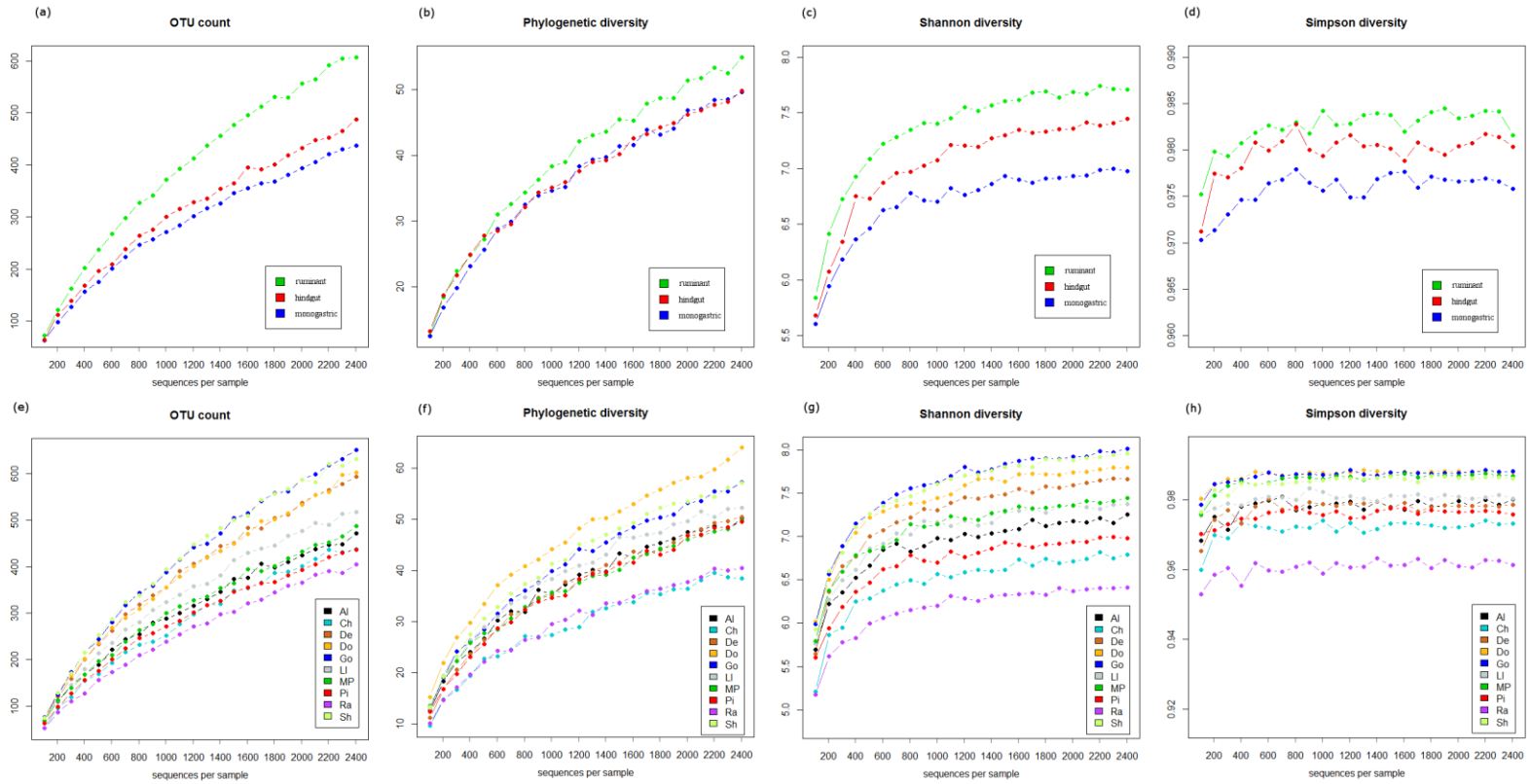


Figure S4.2. Alpha diversity rarefaction curves. (a) OTU count for each digestion type (b) Phylogenetic diversity indices for each digestion type (c) Shannon diversity indices for each digestion type (d) Simpson diversity indices for each digestion type (e) OTU count for each animal species (f) Phylogenetic diversity indices for each animal species (g) Shannon diversity indices for each animal species (h) Simpson diversity indices for each animal species.

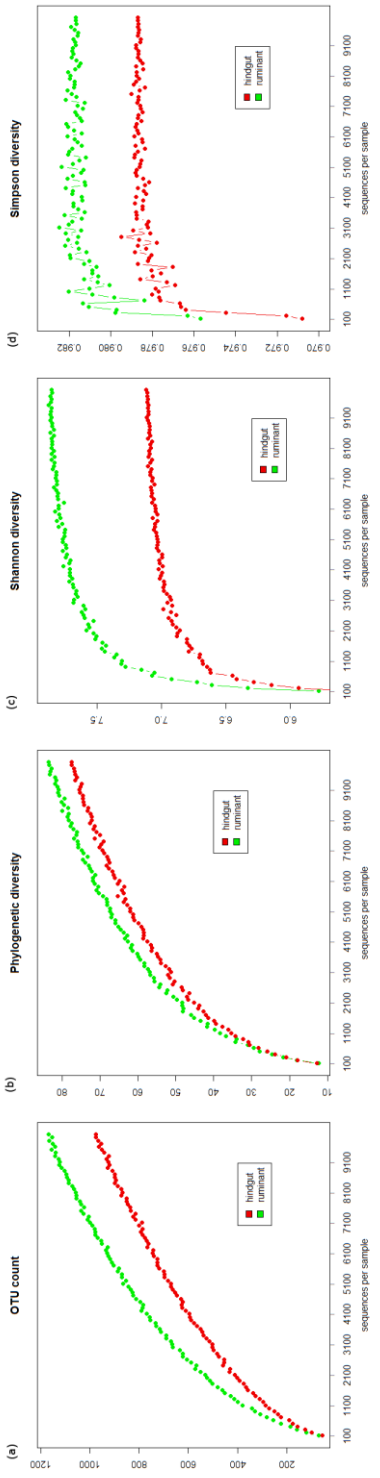


Figure S4.3. Rarefaction curves for two digestion types; hindgut fermenters and ruminants, using a 10,000 subset sample.

Table S4.1. Taxa of the animals used in this study

	Taxa levels				
Animal	Kingdom	Phylum	Class	Order	Family
Chinchilla	*Animalia*	*Chordata*	*Mammalia*	*Rodentia*	*Chinchillidae*
Rabbit	*Animalia*	*Chordata*	*Mammalia*	*Lagomorpha*	*Leporidae*
Donkey	*Animalia*	*Chordata*	*Mammalia*	*Perissodactyla*	*Equidae*
Miniature ponies	*Animalia*	*Chordata*	*Mammalia*		
Deer	*Animalia*	*Chordata*	*Mammalia*	*Artiodactyla*	*Cervidae*
Goats	*Animalia*	*Chordata*	*Mammalia*		*Bovidae*
Sheep	*Animalia*	*Chordata*	*Mammalia*		
Llama	*Animalia*	*Chordata*	*Mammalia*		*Camelidae*
Alpaca	*Animalia*	*Chordata*	*Mammalia*		
Pigs	*Animalia*	*Chordata*	*Mammalia*		*Suidae*

Table S4.2. Statistically significant differences in the taxa abundance in the Hindgut fermenter and Ruminant microbiota.

Taxa	Median read percentages (%) Hindgut	Ruminant	P-value
Phylum			
Firmicutes	53.111	65.346	*
Actinobacteria	0.176	0.018	**
Class			
Clostridia	45.912	62.651	**
Bacilli	0.373	0.117	*
Gammaproteobacteria	0.336	0.056	**
Alphaproteobacteria	0.235	0.451	*
Actinobacteria	0.176	0.018	**
Deltaproteobacteria	0.175	0.372	*
Betaproteobacteria	0.133	0.013	*
Epsilonproteobacteria	0.018	0.000	*
Order			
Clostridiales	44.088	60.725	*
Lactobacillales	0.306	0.065	*
Pseudomonadales	0.117	0.011	*
Burkholderiales	0.077	0.013	*
Desulfovibrionales	0.037	0.012	*
Aeromonadales	0.035	0.013	*
Bifidobacteriales	0.030	0.000	*
Campylobacterales	0.018	0.000	*
Rickettsiales	0.000	0.062	*
Family			
Ruminococcaceae	20.480	33.461	**
Marinilabiaceae	0.402	0.002	**
Bacteroidaceae	0.363	1.853	*
Clostridiaceae	0.274	0.445	*
Lactobacillaceae	0.172	0.041	**
Rikenellaceae	0.107	3.083	***
Moraxellaceae	0.083	0.007	*
Incertae Sedis XIII	0.050	0.142	**
Desulfovibrionaceae	0.030	0.007	*
Bifidobacteriaceae	0.030	0.000	*
Peptostreptococcaceae	0.000	0.104	***
Gracilibacteraceae	0.000	0.312	***
Rickettsiaceae	0.000	0.052	*
Genus			
Faecalibacterium[a]	1.102	0.344	**
Acetivibrio[a]	0.927	1.247	**
Prevotella[a]	0.915	0.365	*
Oscillibacter[a]	0.709	1.554	***
Papillibacter[a]	0.449	1.651	***
Paludibacter	0.444	1.424	*
Coprococcus[a]	0.418	0.894	***
Anaerophaga	0.402	0.002	**
Bacteroides[a]	0.363	1.853	*
Acidaminococcus	0.329	0.104	*
Lactobacillus	0.172	0.041	**
Subdoligranulum	0.134	0.015	**
Alistipes	0.107	3.083	***
Acinetobacter	0.083	0.007	*
Parasporobacterium	0.078	0.020	*
Catenibacterium	0.078	0.025	*
Anaerostipes	0.069	0.000	***
Holdemania	0.050	0.198	*
Desulfovibrio	0.030	0.007	*
Acetitomaculum	0.023	0.104	*
Acetanaerobacterium	0.015	0.199	**
Lutispora	0.000	0.286	***
Orientia	0.000	0.052	*

a – genera present in the core microbiota both at animal level and digestion method level. * = <0.05, ** = <0.01, *** = <0.001. Note: Some of the values listed in the table as zero have at least 50% of the values are equal to zero.

Chapter V

Assessing the effect of the galactooligosaccharides and the autochthonous probiotic *Lactobacillus ruminis* on the pig intestinal microbiota

Michelle M. O' Donnell[1,2], Hugh M.B. Harris[2], Bhuvaneswari Lakshminarayanan[1,2], Peadar G. Lawlor[3], Paul W. O' Toole[2,4] and R. Paul Ross[1,4*].

[1] Teagasc Food Research Centre, Moorepark, Fermoy, Co. Cork

[2] School of Microbiology, University College Cork, Cork City, Co. Cork.

[3] Teagasc, Pig Development Department, Animal and Grassland Research and Innovation Centre, Moorepark, Fermoy, Co. Cork, Ireland

[4] Alimentary Pharmabiotic Centre, University College Cork, Cork City, Co. Cork

This chapter is intended for submission as a research article to *Letters in Applied Microbiology*

Note:

Animal handling, treatment preparation, sample collection, culturing, DNA extraction, pyrosequencing PCR and purifications, pyrosequencing data analysis was carried out by M.M. O' Donnell (author of this thesis)

Pyrosequencing read analysis using Qiime and statistical analysis was run by H. M. B. Harris

Sample collection and culturing was assisted by B. Lakshminarayanan

Statistical analysis on pig weights and feed indices was run by P. G. Lawlor

Chapter V

Table of Contents

Abstract

Ingestion of either galactooligosaccharides or lactobacilli has been associated with health benefits, but the precise mechanisms are unclear. In this study, we assessed the effect that feeding galactooligosaccharides, or *Lactobacillus ruminis* ATCC 25644/galactooligosaccharides as a synbiotic, had on the porcine gastrointestinal microbiota. Using a combination of traditional, molecular and next-generation sequencing technologies. ***L. ruminis*** was shown to survive gastric transit and was recovered by culture in the faecal samples of the synbiotic treated group on day 14. Both prebiotic and synbiotic treatment decreased the enumerated *Enterobacteriaceae* by 1-2 logs. Consumption of the prebiotic or synbiotic did not affect the quantifiable levels of total bacteria as studied by qPCR. However, synbiotic treatment increased the proportion of total lactobacilli by over 0.5 log. Both treatment groups increased the proportion of quantifiable *Lactobacillus salivarius* clade bacteria. In contrast, both treatments caused a reduction in the proportion of the *Firmicutes* phylum. The 10% reduction in the *Firmicutes* phylum abundance in the synbiotic treated animals was accompanied by a reduction in the bacterial diversity present in the porcine microbiota. Both treatments increased the proportion of *Lactobacillus* genus assignments and induced a reduction in the *Clostridium* genus assignments in the pigs. At the class, order, family and genus levels the synbiotic treatment group displayed a decrease in the proportions of *Clostridium*–related levels assignment. There was a gender effect in the microbiota in the *Firmicutes*, *Bacteroidetes*, *Spirochaetes* and *Proteobacteria* phyla. This study also showed that consumption of the synbiotic reduced microbiota diversity. The mechanism responsible for this microbial shift by the synbiotic treatment remains unclear.

5.1 Introduction

In a non-commercial environment pigs are weaned at between 14 and 17 weeks of age (Cox & Cooper, 2001). Weaning is an important time in raising pigs commercially (3-4 weeks) in which the piglets are separated from their mother and weaned from milk feed to solid feed (Richards *et al.*, 2005). EU Directive 2001/93/EC states that "Piglets should not be weaned from the sow at less than three weeks of age unless the welfare of the sow or piglets would be adversely affected" (European, 2001). This is because a reduced weaning period and change in environment often results in many problems including a decreased feed intake, post-weaning diarrhoea and mortality (Richards *et al.*, 2005). The porcine microbiota at this time is in a state of flux, has not fully established, and the animal is prone to colonisation by potentially pathogenic genera like *Escherichia* and *Salmonella* (Farzan *et al.*, 2010; Salajka *et al.*, 1992).

Antimicrobials incorporated in feed are used extensively through in the pig industry to prevent disease, and to improve growth rates and feed efficiencies (Looft *et al.*, 2012). However, in recent years there has been increased effort to use alternatives to antimicrobials in animal feeds due to increased risk of antibiotic resistant strains of pathogenic bacteria and concerns regarding the use of antimicrobials in foods consumed by humans (Roselli *et al.*, 2005). Studies on the effect of feed incorporating antibiotics noted an increase in the proportions of *Proteobacteria* (*Shigella* sp. and *Escherichia coli*) in pigs receiving antibiotics (Looft *et al.*, 2012). At the meta-transcriptomic level there was a resultant increase in expression of antibiotic resistance genes in the samples from corresponding pigs (Looft *et al.*, 2012). Additional concerns are the effect that antibiotic remnants, from the poor absorption of the antibiotics in the animal gut, can have on the environment (Sarmah *et al.*, 2006) .

An example of a feed antibacterial alterative to antibiotics is the heavy inorganic mineral zinc oxide (ZnO) (Smith *et al.*, 1997). However, the consumption of feeds incorporating ZnO can also have a variety of effects on the porcine microbiota (Bednorz *et al.*, 2013; Bratz *et al.*, 2013; Vahjen *et al.*, 2010). Similar undesired environmental effects on the water ways and soil have also been identified by monitoring the heavy metals excreted in the waste products of the pigs (Atieno *et al.*, 2013; Seiler & Berendonk, 2012). Alternative feed additives with the ability to

improve host health and having no adverseeffect on the environment are therefore needed for the future of commercial pig production.

Such alternatives potentially include the use of prebiotics and there has been a growing interest in recent years in the use of prebiotics as modulators of intestinal health. Prebiotics are defined as "selectively fermented ingredients that allow specific changes both in the composition and/or activity in the gastrointestinal microbiota that confers benefits upon host well-being and health" (Gibson *et al.*, 2004). Galactooligosaccharides (GOS), lactose-derived oligosaccharides, soy oligosaccharides, lactulose and fructooligosaccharides are some of the carbohydrates that can be classified as prebiotics and which are commonly consumed as dairy, fruits and vegetables (Gibson & Roberfroid, 1995). The *Lactobacillus* and *Bifidobacterium* genera which are both found in the mammalian microbiota are both the targets of prebiotics and they are capable of fermenting a range of carbohydrates including oligosaccharides, starch, non-starch polysaccharides among many more (Barrangou *et al.*, 2003; Barrangou *et al.*, 2006; O' Donnell *et al.*, 2011; Saulnier *et al.*, 2007). However, studies have shown that *in vitro* other genera are also able to degrade the prebiotic sugars undermining the goal of specificity (Van der Meulen *et al.*, 2006).

In this study, we aimed to investigate the effect on the pig microbiota of introducing galactooligosaccharides and *Lactobacillus ruminis* ATCC 25644 into the diet. A number of techniques including traditional culturing methods, quantitative PCR and 16S rRNA amplicon pyrosequencing were used to assess the effect of both treatments following a 14 day feeding period.

5.2 Materials and Methods

5.2.1 Animals and housing

Faecal samples were collected from 24 crossbred (Large White x Landrace) weaned pigs. Pigs within each block were assigned at random to one of two treatment groups (n=12), as follows: prebiotic (GOS) only or synbiotic (GOS and *L. ruminis* ATCC 25644). The pigs were monitored daily for any sign of illness throughout the trial. In addition to the treatments fed throughout the trial, all pigs had unrestricted access to water. Body weights of each animal were recorded on Day 0 before feeding of the treatments began and a final body weight was taken on Day 14 following the feeding of the last treatment.

Each animal was penned individually to prevent cross-contamination and all animals were housed in a single room to avoid inter-room variation. Each pen was fully slatted (1.07m × 0.0.6m) with plastic slats (Faroex, Manitoba, Canada). Each pen had a door mounted stainless steel trough (410 mm long) with a divider in the middle. The compartment of each trough was used for feeding the feed/probiotic/prebiotic. Heat was provided by an electric bar heater and thermostatically controlled. The rooms were naturally ventilated with an air inlet in the door and exhaust by way of a roof mounted chimney. Temperature was maintained at 28-30^{o}C in the first week and reduced by 2^{o}C per week to 26^{o}C in the second week. Maximum and minimum temperatures for the previous 24 hours were recorded daily at 0900h. Lighting was provided by tubular fluorescent lights from 0830h to 1630h daily.

5.2.2 Diet

The feed consumed by each pig did not contain antibiotic feed additives or therapeutic levels of zinc oxide. It contained wheat, maize, full fat soya, milk powders, soya oil and minerals and vitamins and provided 15.5 MJ digestible energy/kg, 15g/kg lysine, 225g/kg crude protein, 80g/kg oil, 30g/kg crude fibre and 62g/kg ash. Despite animals of this age being prone to infections which are limited by zinc oxide it was decided that this antimicrobial with previously published effects on the viability of members of the *Lactobacillus* genus would compromise the study objectives and was therefore omitted (Starke *et al.*, 2013).

5.2.3 Preparation of probiotic strain used for pig feeding trial

The rifampicin-resistant variant of *L. ruminis* ATCC 25644 was generated by selection on MRS agar plates supplemented with increasing concentration of rifampicin (5 to 150 µg/ml). The rifampicin tagged (resistant) strain *L. ruminis* ATCC 25644 was grown overnight in 2 L volumes of de Man Rogosa Sharpe broth (MRS) [Merck, Darmstadt, Germany] supplemented with 150 µg/ml rifampicin at 37°C. This volume provides enough cells to ensure that each animal would receive $1x10^{10}$ cells each day. The culture was then centrifuged at 2,704 x g for 20 minutes at 4°C. The supernatant was decanted and the cells washed with sterile water and re-centrifuged using the same parameters. This step was repeated twice. The resulting cell pellet was then resuspended in 120 ml sterile water and 12 ml aliquots were made into sterile containers. The containers were then freeze-dried overnight (Virtis AdVantage 2.0 BenchTop Freeze Dryer, Suffolk, UK). Bacterial numbers were enumerated for each batch of freeze dried culture to ensure it maintained viability.

5.2.4 Preparation of the prebiotic and synbiotic

The galactooligosaccharide (GOS) used in this trial was Vivinal GOS syrup (Friesland Campina Domo, Amersfoort, The Netherlands). This is a GOS-rich whey product with a thick consistency and sweet taste. The GOS syrup typically contains 59% GOS and 41% mono/di-saccharides (lactose, glucose and galactose). For week 1 of the trial 350 ml of GOS syrup was aliquoted into beakers and for week 2 of the trial 530 ml of GOS was aliquoted. These volumes were calculated to ensure that each pig would receive the GOS at 4% of their average body weight. One container of freeze-dried *L. ruminis* cells was mixed well with the beaker of GOS syrup and the synbiotic mixture syringed directly onto the feed of the synbiotic treatment group. The prebiotic-only group received the prebiotic GOS syrup syringed directly onto their feed. The troughs of each pen were inspected 30 minutes after feeding to ensure that the prebiotic and synbiotic treatments were consumed entirely.

5.2.5 Enumeration of bacteria in pig faeces

In order to assess the effect of the treatments on the pig microbiota we plated the diluted faecal samples from each pig taken at Day 0, Day 7 and Day 14 on Violet Red Bile Glucose agar (VRBG) (Hampshire, England), MRS with 150µg/mL rifampicin and Lactobacillus selective agar (LBS) (BD BBL, Heidelberg, Germany). All

incubations were performed at 37°C unless stated otherwise. Faecal samples were homogenized in maximum recovery diluent (Lab M, United Kingdom) as 10-fold dilutions prepared using stomacher bags.

5.2.6 Faecal DNA extraction and pyrosequencing

Total bacterial genomic DNA was isolated from the faeces according to the Repeat Bead Beating plus column method (RBB+C) (Yu & Morrison, 2004). The extracted DNA was then used as a template in PCR amplifications as described previously (O' Donnell *et al.*, 2013). Samples were sequenced with 454 Titanium technologies (Teagasc Food Research Centre, Moorepark, Ireland).

5.2.7 Quantitative PCR

Absolute quantitative PCR using the standard curve method was carried out to estimate and quantify the effect of the two treatments on a) Total bacterial numbers, b) Total Lactobacilli and c) relative abundance of the *L. salivarius* clade bacteria. The primers used in the study are shown in Table S5.1. Absolute quantitative PCR was carried out using the SensiMix SYBR No-ROX Kit (myBio, Kilkenny, Ireland) and the manufacturers recommended protocol. Colony forming units (CFU) were calculated from the copy number results from each qPCR's using the following formula: [(C/µl)(TV) x (T cfu/ml)]/TCN = [cfu/ml(S)]/1.

C/µl = Copy number/µl, TV = Template volume, TCN = Total copy number of the standard used, T cfu/ml = Total cfu/ml of standard used and cfu/ml(S) = cfu/ml of test sample. The estimated $\log_{10}$CFU/ml from each test group (Total Lactobacilli and *L. salivarius* clade) was normalized to the $\log_{10}$CFU/mL of the total bacteria for statistical comparisons.

5.2.8 Pyrosequencing

Raw sequencing reads were processed as previously outlined (O' Donnell *et al.*, 2013). The species level assignments were subject to strict criteria. The representative sequence of each OTU was blasted against the RDP database. For the classification of the reads to the species level, the most common sequences (100% identity) were chosen from each OTU cluster as a representative sequence. Only unique species classifications were accepted if the following criteria were met with (a) if a representative sequence aligned with equal percentage identity (b) length and (c) had

a blast score to a single species. If a representative sequence had a blast score to 2 or more species the sequence remained unclassified. Core taxa were defined as being present at ≥0.1% of total reads.

5.2.9 Statistics

Statistical analysis on the pigs and the effect of treatment was carried out by the mixed models procedure in SAS (Institute, 1990). The fixed effects analysed were Treatment and sex. Block was included as a random effect (pigs had been blocked on sex litter origin and initial weight). Initial pig weight was included as a covariate for the analysis of all growth performance parameters. Least squares means were computed and significance was reported for $P < 0.05$ and tendencies towards significance were reported for $0.05 < P < 0.10$. For all response criteria, the individual pig was considered the experimental unit.

Colony count standard deviations were calculated using the STDEVP function in Microsoft Excel and the t-test with paired two samples with means (Data analysis tool kit) was utilised to identify statistical significance within the plate counts. For the rifampicin-MRS colony count a single sample t-test was utilised using 0 as the null hypothesis for the 12 pigs in the synbiotic treatment group at Day 14.

Statistical significance of pyrosequencing read assignment proportionalities was assessed using the Mann Whitney U test (Siegel, 1956). The p-values were corrected for multiple testing using the Benjamini and Hochberg multiple correction method (Benjamini & Hochberg, 1995).

5.2.10 Alpha and beta diversity indices

The alpha diversity metrics and beta diversity principle coordinate plots were generated from a rarefied OTU table with 4,250 sequences per sample, which excluded 4 of the 48 samples (with < 4,250 sequences).

Five alpha diversity metrics were calculated to measure the microbial diversity in the porcine microbiota of the animals consuming the two feeds. Rarefaction analysis and three of the metrics used in this study (Phylogenetic Diversity, Observed species and Shannon index) were previously described (O' Donnell *et al.*, 2013). The Simpson's Index (D) measures the probability that two individuals randomly selected from a sample will belong to the same species (or some category other than species). Good’s

coverage (ESC) was estimated using the formula ESC = 1-n/N, where n= number of singleton OTU's and N=number of assigned reads.

Beta diversity analysis was carried out as previously outlined (O' Donnell *et al.*, 2013).

5.3 Results

5.3.1 Effect on pig weights and growth characteristics

The weights of all trial animals were recorded on day 0 and day 14. The weight measurements were used to calculate the average daily gain, average daily feed intake and the feed conversion efficiency. No statistically significant effect was observed for either treatment on any of the metrics. The results of the statistical analysis are shown in Table S5.2. No animal used in the trial showed any signs of illness (diarrhoea) or other aberrant behaviour over the 14 day trial.

5.3.2 Microbiological analysis

No colony counts were feasible for the pigs No. 7, 9 and 16 as there was no faecal sample remaining following its utilisation for pyrosequencing. Therefore, these samples were omitted in the colony count analysis and the plate count analysis was limited to nine pigs for statistical analysis. Rifampicin resistant bacteria were only identified in the faecal samples taken from the synbiotic treatment group pigs on Day 14. The colonies that were identified were typical of *L. ruminis* ATCC 25644 i.e. circular, flat non-mucoid, moderately sized colonies (dull, opaque colonies).

Between 4.5 and 8.3 logs of rifampicin-resistant bacteria were recovered from the faecal samples of the synbiotic group on day 14. Some of the colonies identified on the rifampicin plates were harvested and used for a colony PCR screen to confirm that the colonies were *L. ruminis* ATCC 25644 (data not shown). The data for the plate counts determined for each group are shown in Table 5.1. The increases in the numbers of total lactobacilli and rifampicin-resistant bacteria in the synbiotic group were statistically significant ($P<0.001$ and $P<0.01$, respectively. Both treatments had a noticeable increase in the number of total lactobacilli at the median level. With a few exceptions (n=4) the total number of *Enterobacteriaceae* decreased in both treatment groups. However, this effect was only significant ($P<0.05$) in those animals receiving the prebiotic feed.

Table 5.1. Plate colony count results and estimated cell numbers from absolute qPCR analysis of both treatment groups

	Prebiotic			Synbiotic		
Test	**Day 0**	**Day 14**	**P value**	**Day 0**	**Day 14**	**P value**
Colony Count[a]						
Total *Enterobacteriaceae*	7.97 (1.09)	6.88 (0.97)	0.02	8.45 (1.37)	6.56 (1.12)	0.08
Total Lactobacilli	9.04 (0.60)	9.48 (0.73)	0.48	9.11 (0.40)	9.51 (0.53)	0.002
Rifampicin-resistant bacteria	0.00	0.00	-	0.00	6.96 (0.94)	0.000
Absolute qPCR[b]						
Total Bacteria	9.07 (0.04)	9.10 (0.04)	0.44	9.08 (0.05)	9.18 (0.08)	0.52
Total Lactobacilli	7.37 (0.08)	7.22 (0.07)	0.49	6.71 (0.08)	7.42 (0.04)	0.22
Lactobacillus salivarius clade	2.97 (0.05)	3.77 (0.04)	0.14	3.19 (0.02)	3.70 (0.02)	0.10

Standard deviation values in parentheses.

a log10 CFU/ml by viable counts on plates

b log10 CFU/ml by qPCR

5.3.3 Absolute quantitative PCR

Absolute quantitation is the process that determines the absolute quantity of a single nucleic acid target sequence within an unknown sample. The results of the absolute quantitative PCR are shown in Table 5.1. In both treatment groups there was a negligible difference in the day 0 and day 14 total bacterial numbers. However, examination of each animal by gender revealed that 67% of female animals and 58% of male animals showed an increase or maintenance of Total bacterial cell numbers over the 14 day trial (data not shown).

A moderate increase in the estimated total lactobacilli CFU/ml was observed in the synbiotic treatment group. However, this increase was not statistically significant. Gender differences of the animals also had an influence on the results of the quantitation analysis. In the prebiotic treatment group there was an increase in the *Lactobacillus* numbers of four pigs. Two of the pigs, both of which were male, displayed a significant increase in total *Lactobacillus* proportions. There was an

increase in the log_{10}CFU/ml numbers in the synbiotic group in 7 animals of which, n=5 were female

Both treatment groups showed a moderate increase (0.5 log) in the levels of the *L. salivarius* clade bacteria as measured by qPCR. This apparent increase in the prebiotic group (not receiving *L. ruminis* supplementation) may be due to lack of specifity in the primers and amplification of the other members of the *L. salivarius clade*.

5.3.4 Amplicon sequencing

Further microbiota analysis was provided by sequencing the 16S rRNA gene V4 amplicon from bacterial DNA extracted from faeces. The total read count generated by this study was 711,977 with 462,592 reads in the prebiotic group and 249,385 reads attributed to the synbiotic group. Following chimera removal and trimming the total number of sequence reads was 597,543 with an average read length of 225 bp (124/325 bp min/max). Within each treatment group there was 187,389 (Prebiotic Day 0); 197,774 (Prebiotic Day 14); 130,890 (Synbiotic Day 0) and 81,490 (Synbiotic Day 14) reads.

5.3.5 Bacterial diversity estimations

Alpha diversity is defined as the bacterial diversity identified within a sample. We used six different alpha diversity measures to assess the effect of treatment on the porcine microbiota. A summary of the metrics is provided in Table 5.2 and the plots for each are shown in Figure 5.1.

Table 5.2. Alpha diversity indices of treated animal groups

Alpha diversity metric	Prebiotic			Synbiotic		
	Day 0	Day 14	P value[a]	Day 0	Day 14	P value[a]
Phylogenetic diversity	63.12	63.35	0.20	65.7	48.4	0.08
Chao1 score	1050.3	1072.9	0.42	1017.5	603.1	0.06
OTU count	540	536	0.31	520	336	**0.04**
Shannon Weaver index	6.69	6.6	1.00	6.73	6.24	0.08
Simpson index	0.98	0.98	0.85	0.97	0.97	0.38
Goods coverage	0.97	0.97	-	0.96	0.97	-

- No statistical significance test carried out.

a Statistically significance values for Day 0-Day 14 comparison of indices

Analysis of the alpha diversity metrics from the prebiotic group indicated that consumption of galactooligosaccharides alone had little effect on the bacterial diversity in the microbiota. However, the synbiotic treatment group experienced a reduction in microbiota alpha diversity. The reduction in OTU count diversity by day 14 in this group compared with day 0 was statistically significant. The Chao1, Phylogenetic diversity and OTU count rarefaction plots failed to plateau in the prebiotic group (day 0 & day 14) and synbiotic group (day 0) samples. These results indicated that true numbers of phylotypes in the porcine microbiota were not reliably measured. The curves for these metrics in the synbiotic group samples from day 14 almost reached a plateau. However this is most likely due to the reduced diversity when compared to the other samples.

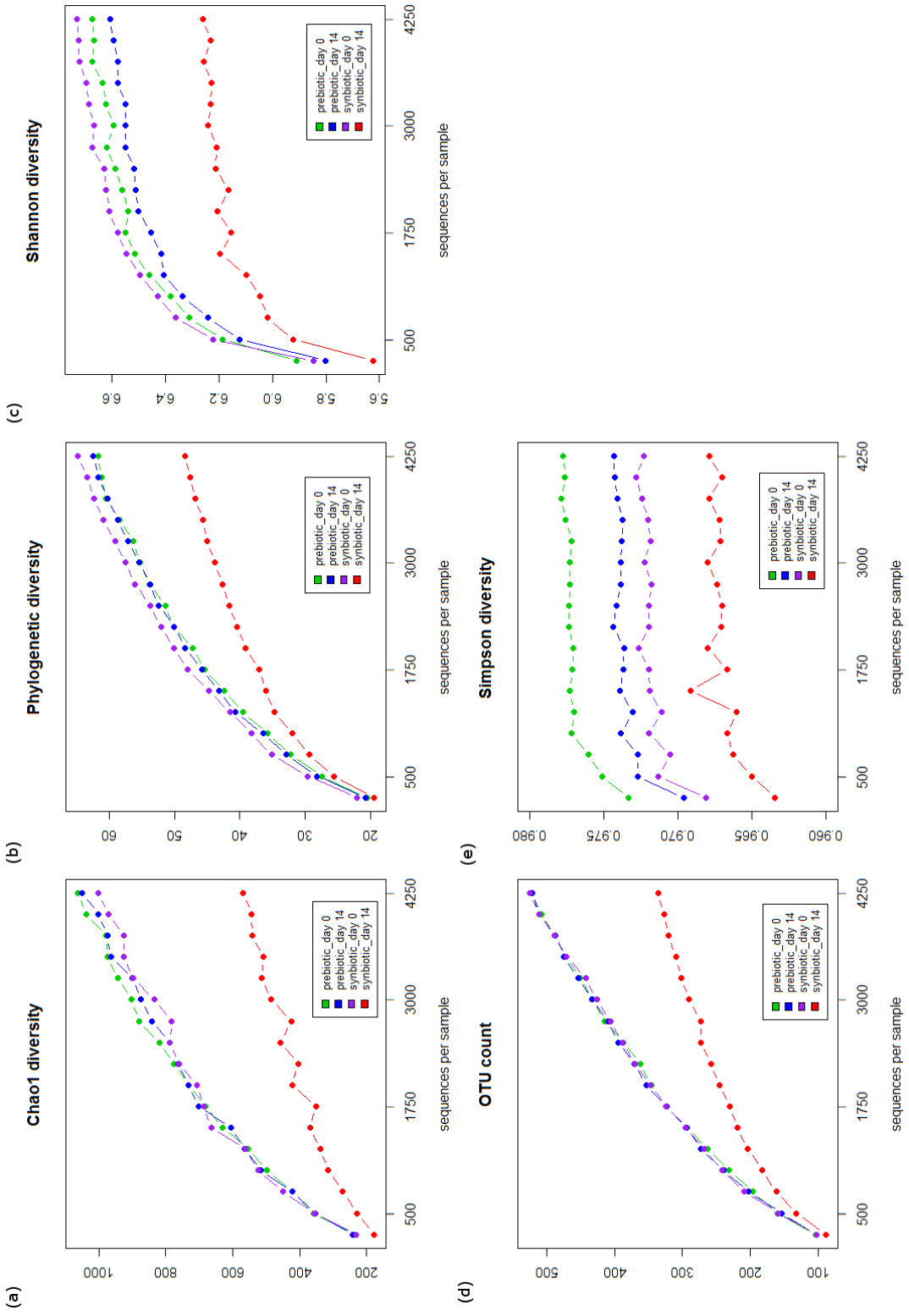


Figure 5.1. Alpha diversity measures in treatment groups (a) Chao1 diversity (b) Phylogenetic diversity (c) Shannon diversity (d) OTU count (e) Simpson diversity.

5.3.6 Dietary treatments did not affect bacterial diversity

Beta diversity is the diversity identified between a collection of samples. Visualisation of the un-weighted Unifrac PCoA plots (Figure 5.2 a) revealed an overlap between the samples in the prebiotic treated group from day 0 and the synbiotic treated group from day 14. The remaining samples separated by treatment and sample time. The first two axes in the un-weighted plot explain 17% of the variation. However, the weighted Unifrac PCoA plots (Figure 5.2 b) showed no clear separation of samples from either group or sample time. Therefore, it is very difficult to discriminate the treatment groups based on their microbiota composition, suggesting that the microbiota composition of these animals regardless of treatment or sampling time is similar.

5.3.7 Taxonomic shifts between diet groups

A summary of the taxa that showed altered abundance in both treatment groups is shown in Table 5.3 while the taxa that showed an altered abundance between treatment groups are shown in Table 5.4. Figure 5.3 illustrated the dominant taxa and the changes in abundance over the 14 day trial period.

Sixteen phyla were identified with varying levels of abundance between the two groups used in this study. The dominant phyla were *Firmicutes* and *Bacteroidetes* (Figure 5.3 (a)) however, by Day 14 both treatment groups showed a reduction in abundance of *Firmicutes* phylum. The reduction in *Firmicutes* abundance was statistically significantly ($P < 0.01$, unadjusted) in the synbiotic treatment group. This loss of read proportions coincided with an increase in the *Bacteroidetes* abundance (1.7-4.1% increase) and *Spirochaetes* abundance (0.4-4% increase) phylum in both treatment groups.

Twenty-four class assignments were identified from the microbiota analysis of two treatment groups. The *Clostridia* and *Bacteroidia* were the dominant class in both treatment groups (Figure 5.3 (b)). Only the pigs in the synbiotic treatment group showed a reduction in the faecal microbiota abundance of *Clostridia, Deltaproteobacteria* and *Sphingobacteria* class assignments.

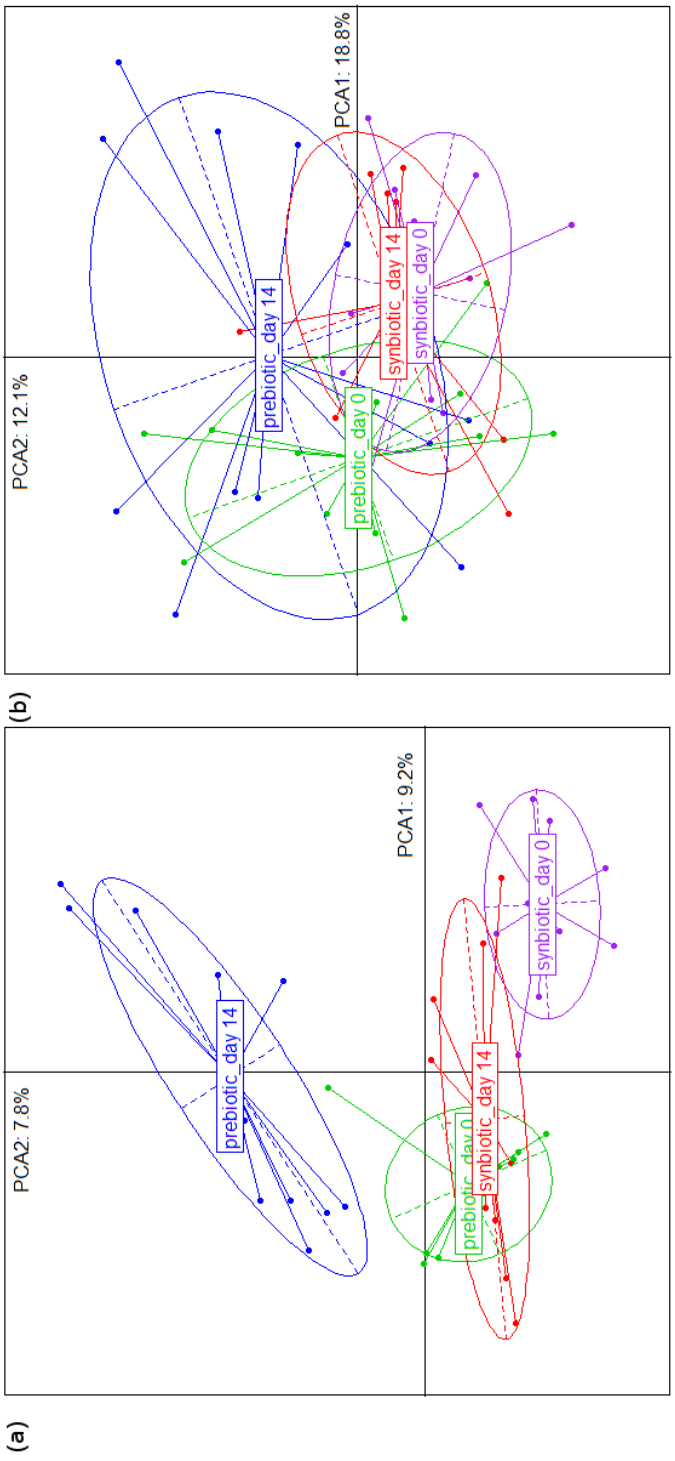


Figure 5.2. Beta diversity Principle coordinate plots for both treatment groups and sampling timepoints. (a) Un-weighted Uni-frac PCoA plot (b) Weighted Uni-frac PCoA plot.

Table 5.3. Altered abundance of taxa observed in both treatments groups

Taxa	Prebiotic			Synbiotic		
	Day 0	**Day 14**	**P value[#]**	**Day 0**	**Day 14**	**P value[#]**
Phylum						
Bacteroidetes	22.34	24.02		18.06	22.14	
Spirochaetes	0.68	1.08		2.47	6.44	
Firmicutes	70.35	68.15		73.23	62.93	**
Class						
Alphaproteobacteria	0.48	0.82		0.40	0.55	*
Negativicutes	3.82	6.14		1.45	1.97	
Spirochaetes	0.68	1.08		2.47	6.44	
Subdivision5	0.00	0.32		0.01	0.08	
Bacteroidia	16.65	9.74		9.73	9.61	
Erysipelotrichia	6.58	1.94	**	3.42	1.79	
Betaproteobacteria	0.09	0.01	***	0.01	0.00	
Order						
Selenomonadales	3.82	6.14		1.45	1.97	
Spirochaetales	0.68	1.08		2.47	6.44	
Subdivision5	0.00	0.32		0.01	0.08	
Aeromonadales	0.51	0.14	*	0.17	0.06	
Bacteroidales	16.65	9.74		9.73	9.61	
Erysipelotrichales	6.58	1.94	**	3.42	1.79	
Verrucomicrobiales	0.10	0.04		0.11	0.00	
Desulfovibrionales	0.09	0.04		0.06	0.03	
Family						
Spirochaetaceae	0.67	1.08		2.45	6.39	
Veillonellaceae	2.73	3.17		0.85	1.65	
Acidaminococcaceae	0.99	1.48		0.53	0.56	
Eubacteriaceae	0.03	0.12	*	0.04	0.05	
Ruminococcaceae	29.47	22.92		36.04	31.13	
Bacteroidaceae	0.47	0.02	**	0.63	0.31	
Erysipelotrichaceae	6.58	1.94		3.42	1.79	
Lachnospiraceae	15.41	11.10		12.80	7.83	
Rikenellaceae	0.08	0.00	***	0.61	0.04	**
Succinivibrionaceae	0.51	0.14	*	0.17	0.06	
Verrucomicrobiaceae	0.10	0.04		0.11	0.00	
Desulfovibrionaceae	0.09	0.04		0.05	0.02	
Peptostreptococcaceae	0.09	0.05		0.12	0.04	
Genus						
Mitsuokella	1.02	1.59		0.12	0.43	
Papillibacter	0.04	0.12	*	0.11	0.17	
Subdivision5_genus_*incertae_sedis*	0.00	0.32		0.01	0.08	
Treponema	0.66	1.07		2.41	6.34	
Acetivibrio	0.09	0.19	**	0.17	0.21	
Acidaminococcus	0.99	1.48		0.49	0.50	
Megasphaera	0.18	0.20		0.04	0.16	
Oribacterium	0.54	0.63		0.12	0.13	
Alistipes	0.08	0.00	***	0.60	0.04	**
Bacteroides	0.47	0.02	**	0.63	0.31	
Barnesiella	0.39	0.29		0.34	0.20	
Blautia	1.26	0.51	**	1.05	0.55	
Clostridium	1.70	1.46		2.38	1.02	
Pseudoflavonifractor	0.32	0.01	***	0.07	0.02	**
Ruminococcus	0.88	0.32		1.20	0.65	
Succinivibrio	0.50	0.13	*	0.16	0.06	
Bulleidia	0.16	0.12		0.36	0.18	
Desulfovibrio	0.09	0.03		0.05	0.02	
Dorea	0.16	0.07		0.22	0.17	
Flavonifractor	0.12	0.02		0.07	0.05	
Subdoligranulum	0.09	0.00	**	0.02	0.00	*

#P values are unadjusted. *P<0.05; **P<0.01; ***P<0.001

Thirty-three order assignments were observed between the two groups. The *Clostridales* and *Bacteroidales* were the dominant orders present in both treatment groups (Figure 5.3 (c)). Following the 14 day trial period *Bacteroidales* abundance decreased in the both treatment groups.

Fifty-four family assignments were identified from the two treatment groups. The *Ruminococcaceae* and *Lachnospiraceae* families were dominant in both treatment groups (Figure 5.3 (d)). The reduction of *Enterobacteriaceae* abundance in both treatment groups, confirmed the reduction in the culturable species in the microbiological analysis. The *Lactobacillaceae* increased in abundance in both treatment groups. None of the increases or decreases at the family level for the synbiotic group were statistically significant upon multiple sample correction.

Ninety-four genera were identified in this study between the two groups. The dominant genera in the prebiotic group after the 14 day feeding regime were *Roseburia*, *Sporobacter* and *Faecalibacterium* (Figure 5.3 (e)). The dominant genera in the synbiotic group were *Treponema*, *Sporobacter* and *Oscillibacter*. Treatment-associated changes in abundance were observed for the *Roseburia*, *Sporobacter*, *Galbibacter*, *Anaerostipes* and *Paludibacter* genera, which only increased in abundance in the prebiotic treatment group. While the *Faecalibacterium*, *Parabacteroides*, *Paraprevotella*, *Prevotella*, *Butyricicoccus*, *Oscillibacter* and *Catenibacterium* genera increased in abundance in the synbiotic treatment group only. No increase or decrease in the synbiotic genera proportions were significant following multiple assignment correction.

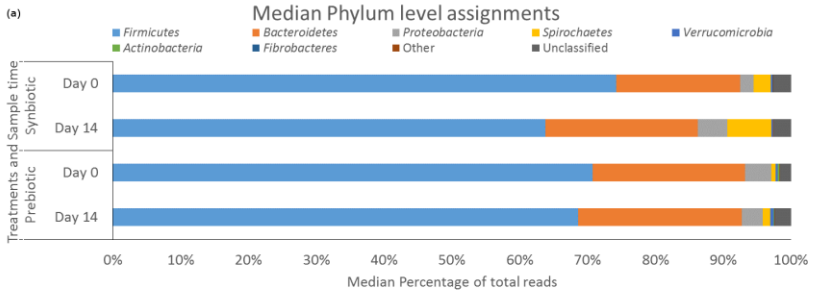
Table 5.4. Altered abundance of taxa observed between treatment groups

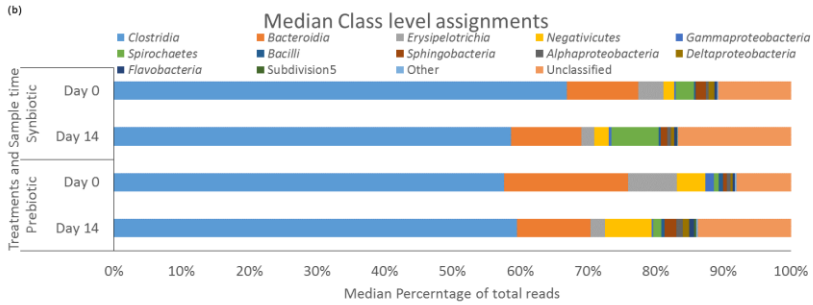
	Prebiotic			Synbiotic		
Taxa	**Day 0**	**Day 14**	**P value[#]**	**Day 0**	**Day 14**	**P value[#]**
Phylum						
Actinobacteria	0.24	0.05		0.01	0.02	
Fibrobacteres	0.02	0.11	**	0.05	0.03	
Proteobacteria	3.85	3.09		1.98	4.25	*
Verrucomicrobia	0.27	0.46		0.23	0.17	
Class						
Bacilli	0.53	0.44		0.23	0.35	
Clostridia	52.46	53.41		61.73	54.39	**
Deltaproteobacteria	0.32	0.87		0.78	0.44	
Fibrobacteria	0.02	0.11	**	0.05	0.03	
Flavobacteria	0.29	0.63		0.33	0.31	
Gammaproteobacteria	1.13	0.22	***	0.23	0.34	
Sphingobacteria	0.52	1.57		1.40	0.85	
Order						
Clostridiales	52.30	53.16		61.21	52.80	**
Fibrobacterales	0.02	0.11	**	0.05	0.03	
Flavobacteriales	0.29	0.63		0.33	0.31	
Lactobacillales	0.52	0.44		0.22	0.27	
Sphingobacteriales	0.52	1.57		1.40	0.85	
Family						
Clostridiaceae	0.08	0.20		0.26	0.06	
Fibrobacteraceae	0.02	0.11	**	0.05	0.03	
Flavobacteriaceae	0.22	0.60		0.21	0.26	
Marinilabiaceae	0.26	0.33		0.06	0.05	
Porphyromonadaceae	3.32	3.00		2.78	3.39	
Prevotellaceae	9.50	3.45		2.59	3.72	
Sphingobacteriaceae	0.04	0.12		0.08	0.00	
Streptococcaceae	0.00	0.09		0.01	0.00	
Genus						
Akkermansia	0.01	0.03		0.05	0.00	
Anaerophaga	0.20	0.29		0.05	0.01	
Anaerostipes	0.12	0.22		0.05	0.02	
Anaerotruncus	0.04	0.12	*	0.09	0.07	
Butyricicoccus	1.34	0.21	***	1.60	1.83	
Catenibacterium	1.19	0.27		0.19	0.21	
Coprococcus	0.16	0.24		0.47	0.47	
Faecalibacterium	3.00	2.96		1.66	2.49	
Fibrobacter	0.02	0.11	**	0.05	0.03	
Galbibacter	0.00	0.41	***	0.02	0.00	
Hydrogenoanaerobacterium	0.08	0.16		0.18	0.15	
Oscillibacter	2.77	2.07		2.37	2.93	
Paludibacter	0.00	0.05	**	0.26	0.06	*
Parabacteroides	0.41	0.24		0.10	0.18	
Paraprevotella	0.61	0.48		0.34	0.43	
Prevotella	8.66	2.78		1.43	2.81	
Roseburia	3.02	4.10		1.32	0.40	
Saccharofermentans	0.00	0.07	**	0.03	0.00	*
Sporobacter	2.99	3.35		6.24	3.45	*
Streptococcus	0.00	0.09		0.01	0.00	

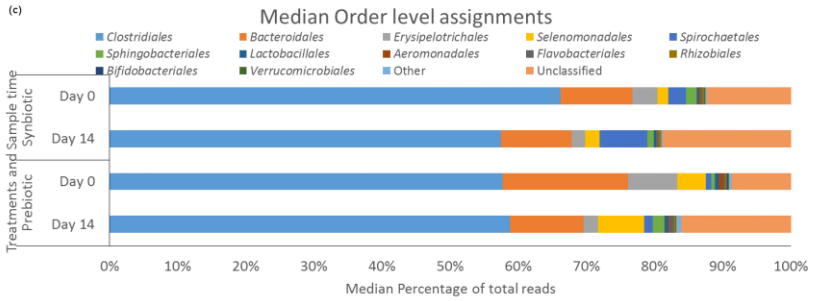
#P values are unadjusted. *P<0.05; **P<0.01; ***P<0.001

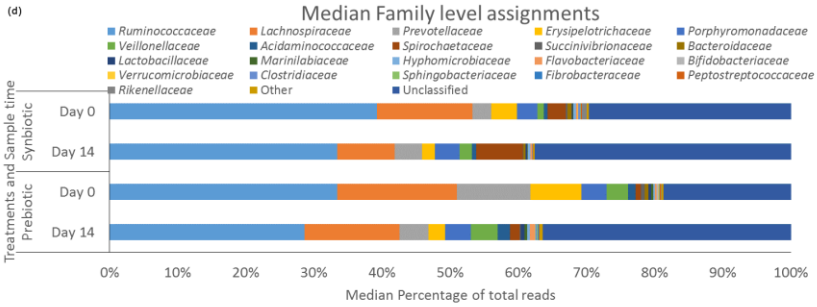
Fifteen species were identified between each group and the relative abundance of each can be seen in Figure 5.3 (f). There was an increase in abundance of the *Lactobacillus salivarius* species and a reduction in *Lactobacillus reuteri* in both treatments. This data suggested that prebiotic has the potential to promote *Lactobacillus in vivo* but this phenomenon is not universal for the whole genus. There was a reduction in abundance of potentially pathogenic species in both treatment groups, for example *Escherichia coli* and *Clostridium bifermentans*. None of the increases or decreases in species abundance were statistically significant.

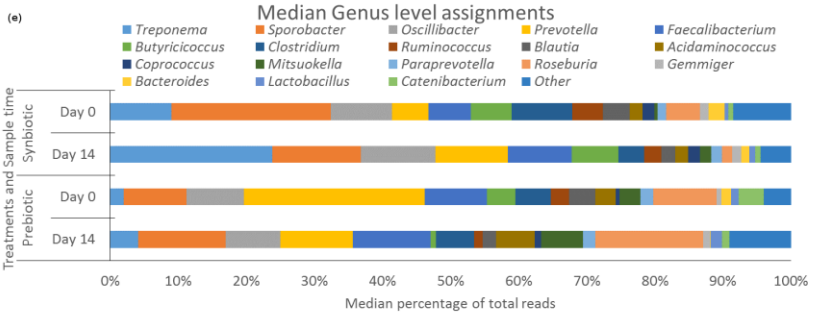
We were able to recover culturable rifampicin-tagged *L. ruminis* ATCC 25644 from the faecal samples on Day 14 in the synbiotic treatment group. But, no *L. ruminis* species level assignments were identified from the sequencing dataset. This may be as a result of the amplification or processing of the reads and the stringent criteria used to assign the species. However, the presence of *L. salivarius* clade bacteria identified in the both groups using qPCR was confirmed by the species level read assignments.


(a)
Median Phylum level assignments
Firmicutes
Bacteroidetes
Proteobacteria
Spirochaetes
Verrucomicrobia
Actinobacteria
Fibrobacteres
Other
Unclassified
Treatments and Sample time
Synbiotic
Prebiotic
Day 0
Day 14
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Median Percentage of total reads


(b)
Median Class level assignments
Clostridia
Bacteroidia
Erysipelotrichia
Negativicutes
Gammaproteobacteria
Spirochaetes
Bacilli
Sphingobacteria
Alphaproteobacteria
Deltaproteobacteria
Flavobacteria
Subdivision5
Other
Unclassified
Treatments and Sample time
Synbiotic
Prebiotic
Day 0
Day 14
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Median Percerntage of total reads


(c)
Median Order level assignments
Clostridiales
Bacteroidales
Erysipelotrichales
Selenomonadales
Spirochaetales
Sphingobacteriales
Lactobacillales
Aeromonadales
Flavobacteriales
Rhizobiales
Bifidobacteriales
Verrucomicrobiales
Other
Unclassified
Treatments and Sample time
Synbiotic
Prebiotic
Day 0
Day 14
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Median Percentage of total reads


(d)
Median Family level assignments
Ruminococcaceae
Lachnospiraceae
Prevotellaceae
Erysipelotrichaceae
Porphyromonadaceae
Veillonellaceae
Acidaminococcaceae
Spirochaetaceae
Succinivibrionaceae
Bacteroidaceae
Lactobacillaceae
Marinilabiaceae
Hyphomicrobiaceae
Flavobacteriaceae
Bifidobacteriaceae
Verrucomicrobiaceae
Clostridiaceae
Sphingobacteriaceae
Fibrobacteraceae
Peptostreptococcaceae
Rikenellaceae
Other
Unclassified
Treatments and Sample time
Synbiotic
Prebiotic
Day 0
Day 14
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Median Percentage of total reads


(e)
Median Genus level assignments
Treponema
Sporobacter
Oscillibacter
Prevotella
Faecalibacterium
Butyricicoccus
Clostridium
Ruminococcus
Blautia
Acidaminococcus
Coprococcus
Mitsuokella
Paraprevotella
Roseburia
Gemmiger
Bacteroides
Lactobacillus
Catenibacterium
Other
Treatments and Sample time
Synbiotic
Prebiotic
Day 0
Day 14
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Median percentage of total reads

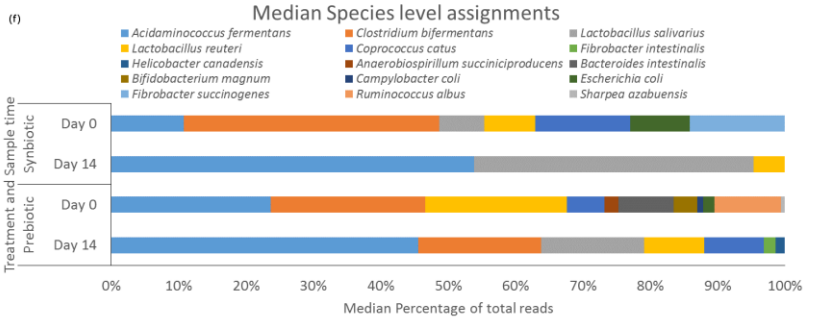

(f)
Median Species level assignments
Acidaminococcus fermentans
Clostridium bifermentans
Lactobacillus salivarius
Lactobacillus reuteri
Coprococcus catus
Fibrobacter intestinalis
Helicobacter canadensis
Anaerobiospirillum succiniciproducens
Bacteroides intestinalis
Bifidobacterium magnum
Campylobacter coli
Escherichia coli
Fibrobacter succinogenes
Ruminococcus albus
Sharpea azabuensis
Treatment and Sample time
Synbiotic
Prebiotic
Day 0
Day 14
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Median Percentage of total reads

Figure 5.3 Dominant Taxa present in the faecal microbiota of the two treatment groups (a) Phylum, (b) Class (c) Order (d) Family (e) Species level assignments

5.3.8 Influence of gender on microbiota composition and development

The read proportions over the two week period in both groups were also affected by the gender of the animal. The males in the prebiotic treatment group showed an increase in read proportions for both the *Firmicutes* and *Bacteroidetes* phyla. The *Spirochaetes*, *Verrucomicrobia* and *Actinobacteria* phyla showed different trends in the female pigs in the synbiotic groups when compared to the male pigs within each group. The differences for the taxon levels between the genders within each group are shown in Table S5.3. Boxplots depicting the influence of the host gender on the proportions of the dominant phyla *in vivo* are shown in Figure 5.4.

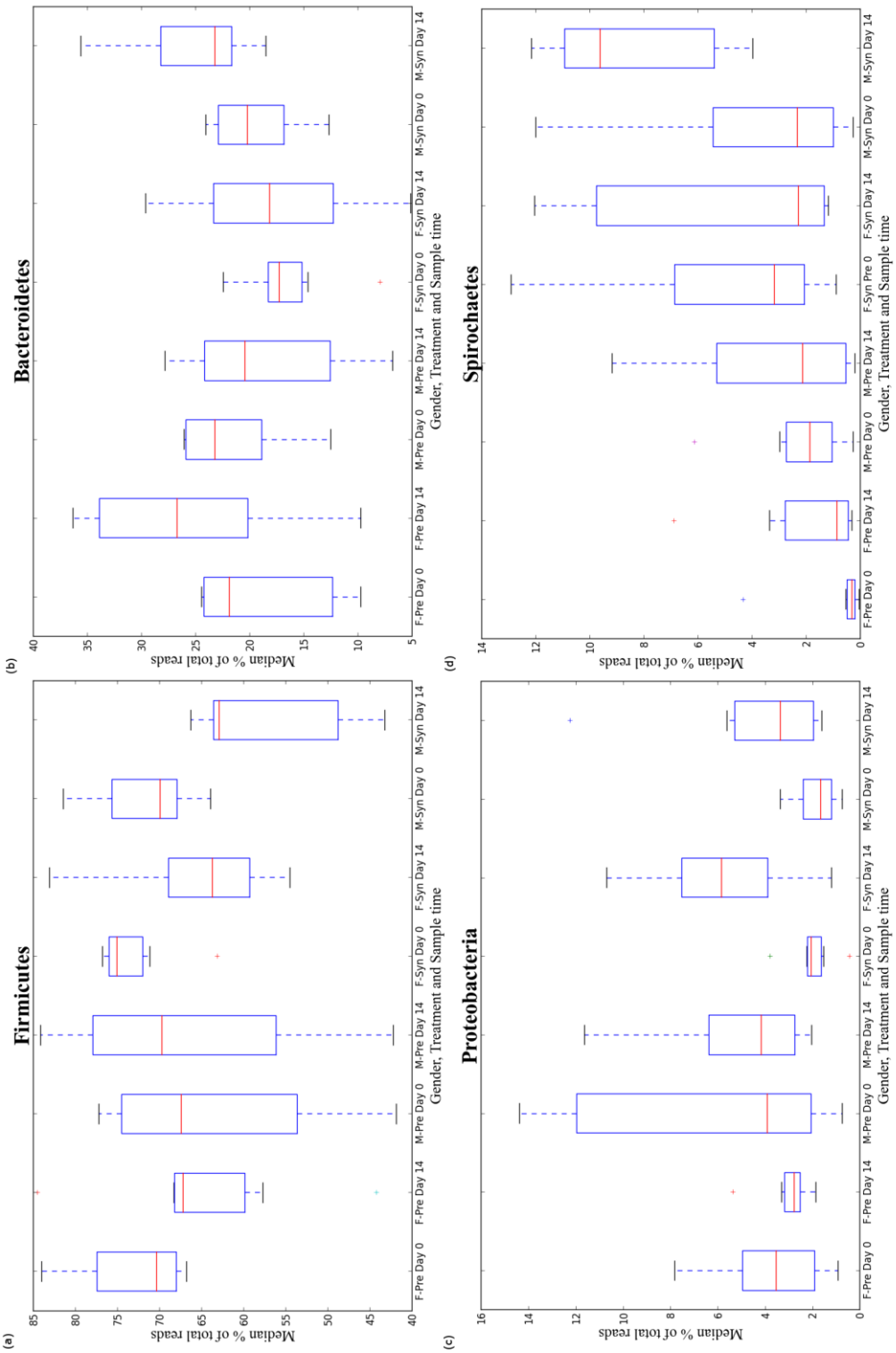


Figure 5.4. Host gender associations with the *Firmicutes*, *Bacteroidetes*, *Proteobacteria* and *Spirochaetes* phyla in each treatment group. F – Females, M - males, Pre – Prebiotic treatment, Syn – Synbiotic treatment.

Only the male pigs in the prebiotic treatment group showed an increase in the *Clostridia* class assignments. This increase in the male pigs was large enough to influence the trend of this class in the prebiotic treatment group negating the decrease in the female animals. Similar trends were observed for the groups and genders for the *Bacilli*, *Gammaproteobacteria*, *Actinobacteria* and *Alphaproteobacteria* classes. There was an increase and decrease in the proportions of the *Lactobacillales* and *Clostridiales* orders, respectively in the female pigs of synbiotic treatment group only. The reduction in *Ruminococcaceae* family within the prebiotic group was present in the female pigs only. The noted increase in the *Prevotellaceae* family in the synbiotic group however, was identified only in the female pigs. The increase in the *Treponema* genus observed in the synbiotic treatment group was as a result of the large increase in the abundance of this genus in the male pigs only. The reduction in the *Clostridium* genus proportional abundance was more pronounced in the females for both treatments. The noted increase in the *Prevotella* genus in the synbiotic treatment was restricted to the females of this group.

5.4 Discussion

Probiotics and prebiotics have various applications in both human and animal health. However, to date little is known about the microbiota-wide effects of these treatments in pigs. Despite the need for alternatives to antibiotic and heavy metal (ZnO) very few studies have assessed the effect of probiotics and or prebiotics in pigs. To the best of our knowledge, no study has investigated the effects of a *Lactobacillus*-based synbiotic treatment on pigs. Previous studies investigating the effects of probiotics in pigs have utilised culture-based and molecular techniques to ascertain the outcome of various feeding interventions (Casey *et al.*, 2007; Siggers *et al.*, 2008). It is recognised that culture-based techniques and molecular techniques have limitations (Maurer, 2011), allowing only the quantification of the culturable bacteria or the amplification/analysis of targeted genera and species. Only recently, have next generation sequencing techniques been applied to monitoring the effects of probiotics on the pig microbiota (Dobson *et al.*, 2011; Riboulet-Bisson *et al.*, 2012).

In a previous study, we analysed the fermentation capabilities of *L. ruminis* ATCC 25644 in media supplemented with various carbohydrates (O' Donnell *et al.*, 2011). The majority of strains tested, including ATCC 25644, could ferment GOS.

We cannot with absolute certainty say that the colonies identified on the MRS agar supplemented with rifampicin were the probiotic *L. ruminis* strain administered in the feed. However, the use of a rifampicin-tagged strain and the microscopic examination of the colonies suggests that the increases in rifampicin resistant bacilli in the synbiotic-consuming pigs was due to the probiotic *L. ruminis* ATCC 25644.

Consumption of GOS alone did not have an effect on the faecal microbiota diversity of the pigs. The findings of our study for the prebiotic group are consistent with those of Davis *et al*. who observed no alteration in microbiota diversity in human faecal samples following the consumption of GOS (Davis *et al.*, 2011). The reduction in microbiota diversity identified in the synbiotic group following the 14 day feeding regime may be a result of secondary metabolites produced by the *L. ruminis* ATCC 25644 strain *in vivo*. Previous studies have shown an antagonistic effect by other Lactobacilli on intestinal pathogens *in vitro*. Production of the short chain fatty acids (SCFA) acetate and lactate were proposed as the mechanism for combatting the pathogenic species (Fooks & Gibson, 2002). However, it is unknown if such SCFAs were responsible for the noted reduction particularly in the *Firmicutes* phylum.

Prebiotics by definition are resistant to gastric enzymatic action and reach the colon to promote the growth of particular genera and species in order to benefit host health (Gibson *et al.*, 2004). Typical probiotic bacteria include lactic acid bacteria such as *Lactobacillus* and *Bifidobacterium* (Collins & Gibson, 1999). Unlike previous studies (Davis *et al.*, 2011), GOS supplementation in this study did not result in a positive bifidogenic response. This difference in outcomes could be due to the inherent differences in the microbiota of humans and pigs, but may also be a feature of the concentration of prebiotic consumed. Davis *et al*. noted a dose-dependent specific bifidogenic response during the feeding of GOS. As we used the same concentration of GOS throughout the study, this may explain why we did not observe such a bifidogenic response in the pigs in either treatment group in this study. Data generated in this study suggest that bacteria other than *Lactobacillus* and *Bifidobacterium* are able to ferment galactooligosaccharides. However, it is difficult to say if the changes in the microbiota are as a result of the fermentation of the prebiotic and synbiotic treatments directly or by other metabolites produced *in vivo* by the microbiota. Future metatranscriptomic studies would be needed to monitor if

the glycosyl hydrolases needed to degrade galactooligosaccharides increased in expression levels in the microbiota during consumption of the prebiotic.

Ruminococcaceae and *Lachnospiraceae* were identified as the dominant families in pigs consuming galactooligosaccharides. This is also consistent with data generated from a human feeding trial (Davis *et al.*, 2011). At the genus level the pig microbiota irrespective of treatment, were dominated by *Roseburia, Sporobacter, Faecalibacterium, Prevotella, Oscillibacter, Mitsuokella, Acidominococcus, Clostridium, Treponema, Lachnospiracea, Oribacterium, Blautia, Paraprevotella* and *Lactobacillus*. The most common genera identified in the microbiota of humans consuming GOS were *Bacteriodes, Faecalibacterium, Blautia, Ruminococcus, Roseburia, Bifidobacterium* and *Dorea* (Davis *et al.*, 2011). The overlap in genera between the studies would suggest that GOS consumption favours the promotion of members of the *Clostridiales* order, but as noted above, this may not be a direct effect.

The gender of the host can affect the gut microbiota (Zhao *et al.*, 2013). In this study, the association of gender with different microbiota compositions was also observed across the assignments for the *Firmicutes, Bacteroidetes, Spirochaetes* and *Proteobacteria*. In previous studies, the influence of gender was identified for the *Bacteroides-Prevotella* group with males having greater proportions of this group (de Carcer *et al.*, 2011; Mueller *et al.*, 2006). However, in this study the female pigs had a greater abundance of *Prevotella* compared to the male pigs in the same group. The differences between the studies may be due to the differences in the hosts, diets and other external environmental and geographic factors. The effects of gender on the microbiota are important to allow optimal design of prebiotic or synbiotic interventions that can be used with a population of mixed gender. Future studies will need to take gender into account when deciding upon the subjects for feeding trials, and trial groups consisting of a single gender may not truly reflect the potential effects of a prebiotic or synbiotic.

In conclusion, *L. ruminis* ATCC 25644 was able to survive gastric transit and was recovered at high levels (7 logs) from the porcine faecal samples. Consumption of the prebiotic galactooligosaccharide alone did not affect the porcine microbiota diversity nor result in a bifidogenic response. However, the synbiotic treatment of

galactooligosaccharides and *L. ruminis* ATCC 25644 significantly reduced the bacterial diversity. The host animal gender was also identified as a factor when assessing the effects of both treatments on the porcine microbiota. Future studies are needed to elucidate the events that occur *in vivo* that result in the loss of diversity in the synbiotic treated animals and if there are any consequences for this reduction in microbiota diversity

5.5 References

Atieno, N. R., Owuor, O. P. & Omwoyo, O. (2013). Heavy metal and associated antibiotic resistance of fecal coliforms, fecal streptococci and pathogens isolated from wastewaters of abattoirs in Nairobi, Kenya. *J Appl Biosci* **64**, 4858-4866.

Barrangou, R., Altermann, E., Hutkins, R., Cano, R. & Klaenhammer, T. R. (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., Azcarate-Peril, M. A., Duong, T., Conners, S. B., Kelly, R. M. & Klaenhammer, T. R. (2006). Global analysis of carbohydrate utilization by *Lactobacillus acidophilus* using cDNA microarrays. *P Natl Acad Sci USA* **103**, 3816-3821.

Bednorz, C., Oelgeschläger, K., Kinnemann, B. & other authors (2013). The broader context of antibiotic resistance: Zinc feed supplementation of piglets increases the proportion of multi-resistant *Escherichia coli in vivo*. *Int J Med Microbiol* **303**, 396-403.

Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc B Met*, 289-300.

Bratz, K., Gölz, G., Riedel, C., Janczyk, P., Nöckler, K. & Alter, T. (2013). Inhibitory effect of high-dosage zinc oxide dietary supplementation on *Campylobacter coli* excretion in weaned piglets. *J Appl Micro* **115**, 1194–1202.

Casey, P. G., Gardiner, G. E., Casey, G. & other authors (2007). A five-strain probiotic combination reduces pathogen shedding and alleviates disease signs in pigs challenged with *Salmonella enterica* Serovar Typhimurium. *Appl Environ Microbiol* **73**, 1858-1863.

Collins, M. D. & Gibson, G. R. (1999). Probiotics, prebiotics, and synbiotics: approaches for modulating the microbial ecology of the gut. *Am J Clin Nutr* **69**, 1052S-1057.

Cox, L. & Cooper, J. (2001). Observations on the pre-and post-weaning behaviour of piglets reared in commercial indoor and outdoor environments. *Anim Sci* **72**, 75-86.

Davis, L. M. G., Martínez, I., Walter, J., Goin, C. & Hutkins, R. W. (2011). Barcoded pyrosequencing reveals that consumption of galactooligosaccharides results in a highly specific bifidogenic response in humans. *PLOS One* **6**, e25200.

de Carcer, D. A., Cuiv, P. O., Wang, T. & other authors (2011). Numerical ecology validates a biogeographical distribution and gender-based effect on mucosa-associated bacteria along the human colon. *ISME J* **5**, 801-809.

Dobson, A., Crispie, F., Rea, M. C. & other authors (2011). Fate and efficacy of lacticin 3147-producing *Lactococcus lactis* in the mammalian gastrointestinal tract. *FEMS Micro Ecol* **76**, 602-614.

European, C. (2001). Commission Directive 2001/93/EC of 9 November 2001 amending Directive 91/630/EEC laying down minimum standards for the protection of pigs. *Off J Eur Commun L* **316**, 36.

Farzan, A., Friendship, R. M., Cook, A. & Pollari, F. (2010). Occurrence of *Salmonella*, *Campylobacter*, *Yersinia enterocolitica*, *Escherichia coli* O157 and *Listeria monocytogenes* in swine. *Zoonoses Public Hlth* **57**, 388-396.

Fooks, L. J. & Gibson, G. R. (2002). *In vitro* investigations of the effect of probiotics and prebiotics on selected human intestinal pathogens. *FEMS Micro Ecol* **39**, 67-75.

Gibson, G. R. & Roberfroid, M. B. (1995). Dietary modulation of the human colonic microbiota: Introducing the concept of prebiotics. *J Nutr* **125**, 1401-1412.

Gibson, G. R., Probert, H. M., Loo, J. V., Rastall, R. A. & Roberfroid, M. B. (2004). Dietary modulation of the human colonic microbiota: updating the concept of prebiotics. *Nutr Res Rev* **17**, 259-275.

Heilig, H. G., Zoetendal, E. G., Vaughan, E. E., Marteau, P., Akkermans, A. D. & de Vos, W. M. (2002). Molecular diversity of *Lactobacillus* spp. and other lactic acid bacteria in the human intestine as determined by specific amplification of 16S ribosomal DNA. *Appl Environ Microbiol* **68**, 114-123.

Institute, S. (1990). *SAS/STAT user's guide: version 6*: Sas Inst.

Lane, D. (1991). 16s/23s rRNA sequencing. In *Nucleic acid techniques in bacterial systematics*. Edited by E. Stackebrandt & M. Goodfellow: John Wiley & Sons, West Sussex, United Kingdom.

Looft, T., Johnson, T. A., Allen, H. K. & other authors (2012). In-feed antibiotic effects on the swine intestinal microbiome. *P Natl Acad Sci USA*.

Maurer, J. J. (2011). Rapid detection and limitations of molecular techniques. *Annu Rev Food Sci Technol* **2**, 259-279.

Mueller, S., Saunier, K., Hanisch, C. & other authors (2006). Differences in fecal microbiota in different European study populations in relation to age, gender, and country: a cross-sectional study. *Appl Environ Microbiol* **72**, 1027-1033.

Muyzer, G., De Waal, E. C. & Uitterlinden, A. G. (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl Environ Microbiol* **59**, 695-700.

O' Donnell, M. M., Harris, H. M. B., Jeffery, I. B., Claesson, M. J., Younge, B., O' Toole, P. W. & Ross, R. P. (2013). The core faecal bacterial microbiome of Irish Thoroughbred racehorses. *Lett Appl Microbiol* **57**, 467–560.

O' Donnell, M. M., Forde, B. M., Neville, B. A., Ross, R. P. & O' Toole, P. W. (2011). Carbohydrate catabolic flexibility in the mammalian intestinal commensal *Lactobacillus ruminis* revealed by fermentation studies aligned to genome annotations. *Microb Cell Fact* **10**, S12.

Riboulet-Bisson, E., Sturme, M. H. J., Jeffery, I. B. & other authors (2012). Effect of *Lactobacillus salivarius* bacteriocin Abp118 on the mouse and pig intestinal microbiota. *PLOS One* **7**, e31113.

Richards, J. D., Gong, J. & de Lange, C. F. M. (2005). The gastrointestinal microbiota and its role in monogastric nutrition and health with an emphasis on pigs: Current understanding, possible modulations, and new technologies for ecological studies. *Can J Anim Sci* **85**, 421-435.

Roselli, M., Finamore, A., Britti, M. S., Bosi, P., Oswald, I. & Mengheri, E. (2005). Alternatives to in-feed antibiotics in pigs: Evaluation of probiotics, zinc or organic acids as protective agents for the intestinal mucosa. A comparison of *in vitro* and *in vivo* results. *Anim Res* **54**, 203.

Salajka, E., Salajkova, Z., Alexa, P. & Hornich, M. (1992). Colonization factor different from K88, K99, F41 and 987P in enterotoxigenic *Escherichia coli* strains isolated from postweaning diarrhoea in pigs. *Vet Microbiol* **32**, 163-175.

Sarmah, A. K., Meyer, M. T. & Boxall, A. B. A. (2006). A global perspective on the use, sales, exposure pathways, occurrence, fate and effects of veterinary antibiotics (VAs) in the environment. *Chemosphere* **65**, 725-759.

Saulnier, D. M., Molenaar, D., de Vos, W. M., Gibson, G. R. & Kolida, S. (2007). Identification of prebiotic fructooligosaccharide metabolism in *Lactobacillus plantarum* WCFS1 through microarrays. *Appl Environ Microbiol* **73**, 1753-1765.

Seiler, C. & Berendonk, T. U. (2012). Heavy metal driven co-selection of antibiotic resistance in soil and water bodies impacted by agriculture and aquaculture. *Front Microbiol* **3**, 1-10.

Siegel, S. (1956). The Mann-Whitney U test. In *Nonparametric Statistics for the Behavioral Sciences*, pp. 116-127. Edited by S. Siegel: McGraw-Hill.

Siggers, R. H., Siggers, J., Boye, M., Thymann, T., Mølbak, L., Leser, T., Jensen, B. B. & Sangild, P. T. (2008). Early administration of probiotics alters bacterial colonization and limits diet-induced gut dysfunction and severity of necrotizing enterocolitis in preterm pigs. *J Nutr* **138**, 1437-1444.

Smith, J. W., Tokach, M. D., Goodband, R. D., Nelssen, J. L. & Richert, B. T. (1997). Effects of the interrelationship between zinc oxide and copper sulfate on growth performance of early-weaned pigs. *J Anim Sci* **75**, 1861-1866.

Starke, I. C., Zentek, J. r. & Vahjen, W. (2013). *Ex Vivo* - Growth Response of Porcine Small Intestinal Bacterial Communities to Pharmacological Doses of Dietary Zinc Oxide. *PloS one* **8**, e56405.

Vahjen, W., Pieper, R. & Zentek, J. (2010). Bar-coded pyrosequencing of 16S rRNA gene amplicons reveals changes in ileal porcine bacterial communities due to high dietary zinc intake. *Appl Environ Microbiol* **76**, 6689-6691.

Van der Meulen, R., Makras, L., Verbrugghe, K., Adriany, T. & De Vuyst, L. (2006). *In vitro* kinetic analysis of oligofructose consumption by *Bacteroides* and *Bifidobacterium* spp. indicates different degradation mechanisms. *Appl Environ Microbiol* **72**, 1006-1012.

Walter, J., Hertel, C., Tannock, G. W., Lis, C. M., Munro, K. & Hammes, W. P. (2001). Detection of *Lactobacillus, Pediococcus, Leuconostoc,* and *Weissella* species in human feces by using group-specific PCR primers and denaturing gradient gel electrophoresis. *Appl Environ Microbiol* **67**, 2578-2585.

Yu, Z. & Morrison, M. (2004). Improved extraction of PCR-quality community DNA from digesta and fecal samples. *Biotechniques* **36**, 808-813.

Zhao, L., Wang, G., Siegel, P. & other authors (2013). Quantitative genetic background of the host influences gut microbiomes in chickens. *Sci rep* **3**.

5.6 Supplementary information

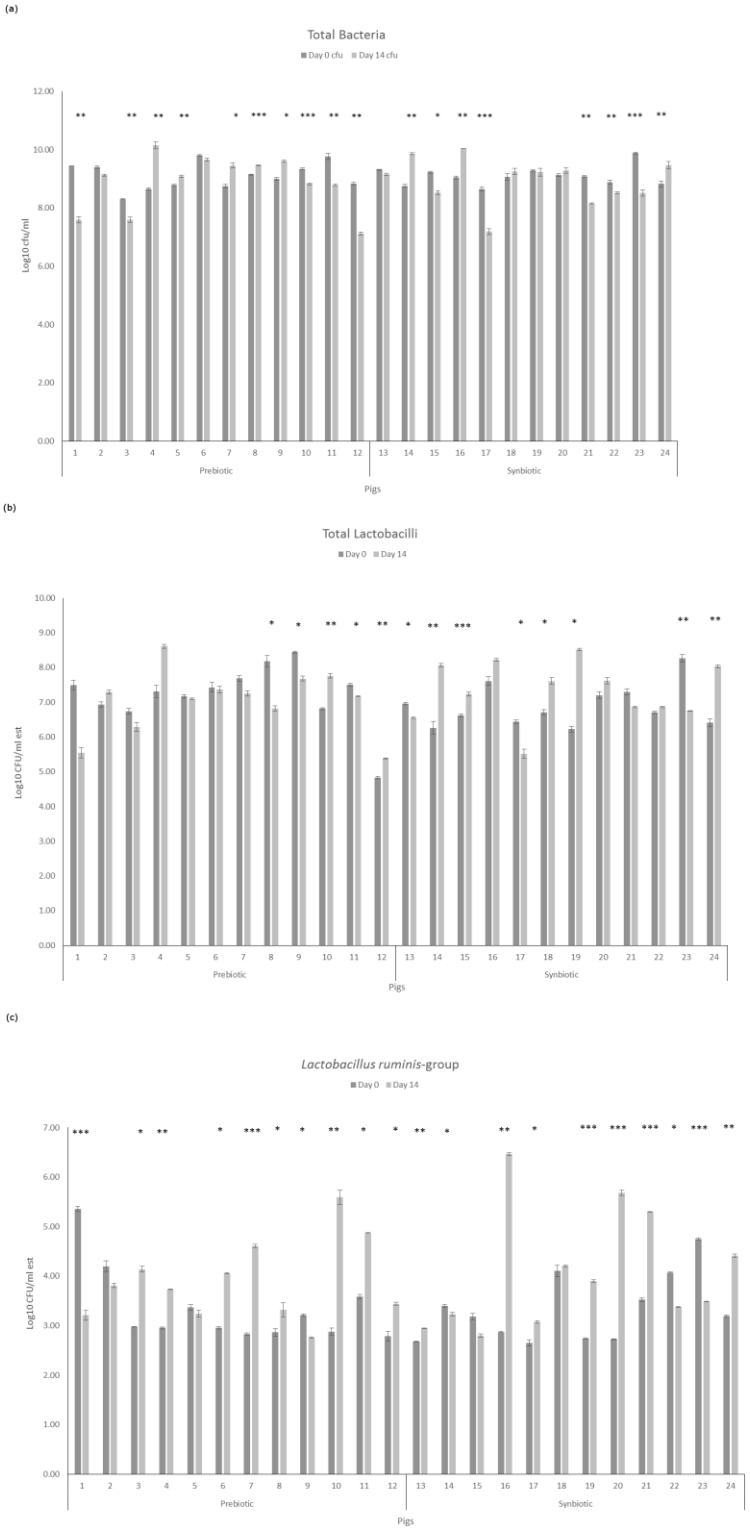


Figure S5.1 The absolute quantitative PCR results for (a) Total Bacteria, (b) Total lactobacilli and (c) *L. salivarius clade* bacteria. Note: the star values indicate p values calculated for each animal comparing the values on Day 0 to those on Day 14; *P<0.05; **P<0.01;***P<0.001

Table S5.1. Primers used in this study

Name	Sequence (5'-3')	Fragment Size (bp)	Reference
Lsal clade qPCR for	GCGGCGTATTAACTTGTTG	162	This study
Lsal clade qPCR rev	TTGCTCCATCAGACTTTCG		
Lb all qPCR for	AGCAGTAGGGAATCTTCCA	341	(Heilig *et al.*,
Lb all qPCR rev	CACCGCTACACATGGAG		2002; Walter *et al.*,
			2001)
Total Bacteria qPCR for	ACTCCTACGGGAGGCAGCAG	195	(Lane, 1991;
Total Bacteria qPCR rev	ATTACCGCGGCTGCTGG		Muyzer *et al.*,
			1993)

Table S5.2. Effect of treatment on the growth performance of pigs over the 14 day experimental period

	Prebiotic	**Synbiotic**	**S.E.#**	**P value**
No. of pigs/treatment	12	12		
Pig weight (kg)				
Day 0	9.9	9.9	0.30	0.80
Day 14	12.6	12.5	0.1	0.56
Average Daily Feed Intake (g)	232	225	4.6	0.28
Average Daily Gain (g)	190	183	7.3	0.49
Feed Conversion Efficiency (g/g)	1.23	1.25	0.045	0.850

Standard error, $^{*}P<0.05$, $^{**}P<0.01$, $^{***}P<0.001$

Table S5.3. Gender influences on the taxa identified in the two treatment groups

	Prebiotic				Synbiotic			
	Females		Males		Females		Males	
Taxa	**Day 0**	**Day 14**	**Day 0**	**Day 14**	**Day 0**	**Day 14**	**Day 0**	**Day 14**
Phylum								
Actinobacteria	0.24	0.05	0.43	0.02	0.00	0.02	0.01	0.00
Bacteroidetes	21.89	26.75	23.21	20.49	17.31	18.20	20.23	23.24
Firmicutes	70.35	67.24	67.41	69.76	75.06	63.74	69.96	62.93
Proteobacteria	3.55	2.77	3.92	4.16	2.06	5.84	1.65	3.35
Spirochaetes	0.32	0.88	1.86	2.14	3.19	2.28	2.35	9.63
Verrucomicrobia	0.18	0.40	0.37	0.53	0.17	0.06	0.26	0.27
Class								
Flavobacteria	0.22	0.58	0.34	1.43	0.33	0.16	0.29	1.03
Actinobacteria	0.24	0.05	0.43	0.02	0.00	0.02	0.01	0.00
Alphaproteobacteria	1.02	0.73	0.36	0.84	0.41	1.01	0.27	0.40
Bacilli	0.53	0.44	0.86	0.93	0.25	0.35	0.22	0.41
Clostridia	58.40	48.70	42.89	55.81	63.94	53.72	60.88	54.39
Gammaproteobacteria	0.88	0.15	1.32	0.27	0.20	0.39	0.29	0.21
Negativicutes	2.36	7.29	6.07	3.83	0.79	4.01	1.87	1.32
Spirochaetes	0.32	0.88	1.86	2.14	3.19	2.28	2.35	9.63
Order								
Clostridiales	58.26	48.51	42.76	55.60	63.46	53.48	60.66	52.80
Flavobacteriales	0.22	0.58	0.34	1.43	0.33	0.16	0.29	1.03
Lactobacillales	0.52	0.44	0.85	0.89	0.25	0.35	0.20	0.15
Selenomonadales	2.36	7.29	6.07	3.83	0.79	4.01	1.87	1.32
Spirochaetales	0.32	0.88	1.86	2.14	3.19	2.28	2.35	9.63
Family								
Ruminococcaceae	33.52	22.62	20.93	22.92	33.41	31.54	36.93	31.13
Acidaminococcaceae	0.98	1.54	1.98	1.13	0.31	0.58	0.70	0.54
Clostridiaceae	0.08	0.17	0.09	0.22	0.16	0.03	0.34	0.29
Flavobacteriaceae	0.17	0.54	0.29	1.32	0.24	0.15	0.19	0.93
Hyphomicrobiaceae	0.89	0.41	0.18	0.20	0.33	0.53	0.20	0.26
Lactobacillaceae	0.46	0.43	0.28	0.37	0.20	0.31	0.12	0.11
Porphyromonadaceae	2.70	3.21	3.61	3.00	2.97	2.93	2.48	4.10
Prevotellaceae	9.50	7.85	9.39	2.98	1.85	4.58	2.90	2.88
Spirochaetaceae	0.29	0.87	1.86	2.12	3.17	2.23	2.34	9.60
Veillonellaceae	1.69	4.42	4.31	2.55	0.58	3.03	1.21	0.85
Genus								
Anaerophaga	0.09	0.42	0.31	0.22	0.06	0.03	0.03	0.01
Barnesiella	0.23	0.29	0.57	0.28	0.25	0.21	0.34	0.21
Paraprevotella	0.49	0.89	0.71	0.25	0.12	0.35	0.73	0.55
Prevotella	8.66	6.02	8.18	2.39	1.06	3.89	1.37	1.34
Lactobacillus	0.46	0.43	0.28	0.37	0.20	0.31	0.08	0.14
Blautia	1.83	0.64	0.68	0.44	1.83	0.47	0.90	1.01
Butyricicoccus	1.83	0.19	1.15	0.22	1.76	1.28	0.97	1.76
Oribacterium	0.07	0.94	0.94	0.21	0.02	0.88	0.20	0.02
Roseburia	4.81	4.10	2.33	4.10	0.99	1.55	1.63	0.35
Faecalibacterium	7.49	3.49	1.84	2.38	0.93	3.95	2.04	0.83
Sporobacter	2.99	3.07	3.61	3.35	7.89	3.38	5.52	3.34
Catenibacterium	0.55	0.21	1.98	0.54	0.09	0.37	0.46	0.09
Acidaminococcus	0.94	1.53	1.96	1.10	0.29	0.57	0.56	0.49
Mitsuokella	0.42	3.15	2.09	0.98	0.06	1.55	0.16	0.27
Treponema	0.29	0.87	1.84	2.10	3.13	2.20	2.57	8.48

Chapter VI

Genomic diversity and biochemical characterisation of *Lactobacillus ruminis* isolates of human, bovine, porcine and equine origin.

Michelle M. O' Donnell[1,2], Hugh M.B. Harris[2], Denise Lynch[2], R. Paul Ross[1], Paul W. O' Toole[2*]

Teagasc Food Research Centre, Moorepark, Fermoy, Co. Cork, Ireland[1].
Department Microbiology, University College Cork, Ireland[2].

This chapter is intended for submission as a research article to *Applied Environmental Microbiology*

Note:

Isolation of new *L. ruminis* strains, biochemical and technological assessment of each strain, motility assays, manual curation and annotation of the two *L. ruminis* genomes, MLST analysis, BRIG and ACT analysis, RNA sample prep and purification, DESeq data output analysis, RT-PCR were carried out by M.M. O' Donnell (author of this thesis).

RNAseq analysis using Bowtie2, Trimmomatic and DESeq was run by D. Lynch

Genome assembly, ACT files and Whole genome phylogenetic trees were generated by H.M.B Harris

Chapter VI

Table of Contents

Abstract

In this study, a mixture of classical microbiological techniques and modern molecular techniques were used to identify and characterise *Lactobacillus ruminis*. Seven newly identified (porcine and equine) and nine (human and bovine) previously identified mammalian-associated *L. ruminis* strains were characterized. The survival, biochemical and metabolic characteristics of *L. ruminis* isolated from various mammalian microbiomes were determined. Three *L. ruminis* strains (S23, DPC 6832 and DPC 6835) were identified as candidate strains for use as probiotics. In this study, we describe the development and use of a multilocus sequence typing (MLST) scheme for *L. ruminis*. The MLST method developed had good discriminatory ability: the 16 isolates of *L. ruminis* examined were divided into three clades in the phylogenetic trees. These groups were based on the host origin of the isolates. Whole genome comparisons also revealed that gaps in the sequences when compared to ATCC 25644 were caused by hypothetical, CRISPR, phage, restriction modification and in some cases carbohydrate-related proteins. From the genome phylogenetic comparisons of the core gene set from the four sequenced strains we observed that *L. ruminis* DPC 6832 was the most divergent strain examined. The novel ability of some of the motile *L. ruminis* isolates to swarm on MRS plates containing up to 1.8% agar was investigated. All the porcine and equine strains had the ability to swarm on agar plates with the standard 1% to 1.5% (w/v) agar concentration, while the motile bovine strain was only able to swarm on MRS agar plates with 0.5% (w/v) agar. Transcriptional studies revealed that fructose, sucrose and fructooligosaccharide enzymes and transporters as well as the flagellar biosynthesis gene *fliC* were important genes transcribed by motile *Lactobacillus ruminis* cells. Swarming *L. ruminis* cells may have an altered metabolism and novel metabolic pathways which are distinct between swimming and stationary cells.

6.1 Introduction

Lactobacillus ruminis is a commensal bacterial species found in the intestines of humans (Heilig *et al.*, 2002; Makivuokko *et al.*, 2010; Reuter, 2001; Wall *et al.*, 2007) that is also present in the gastrointestinal tracts of many mammals including ruminants (Sharpe *et al.*, 1973; Stewart *et al.*, 1988), mono-gastric fermenters (Al Jassim, 2003; Desai *et al.*, 2009; Greetham *et al.*, 2002; Mathiesen *et al.*, 1987; Ritchie *et al.*, 2009), hindgut fermenters (Vörös, 2008; Willing *et al.*, 2009c) and other mammals (Endo *et al.*, 2010). In some studies it has also been identified in birds (Kovalenko *et al.*, 1989; Xenoulis *et al.*). *L. ruminis* was first identified in 1961 and originally classified as *Catenabacterium catenaforme* (Lerche & Reuter, 1961). It was not formally recognised under its current taxonomic classification until 1973 when Sharpe *et al*. characterised 3 isolates from the steer rumen (Sharpe *et al.*, 1973). *L. ruminis* has been described as an autochthonous species in the GIT of humans (Reuter, 2001; Tannock *et al.*, 2000). Previous studies have noted that *L. ruminis* has potential immunomodulatory properties (Neville *et al.*, 2012; Taweechotipatr *et al.*, 2009) as well as a possible use in combating antibiotic resistant bacteria (Yun *et al.*, 2005).

Previously, we characterised the fermentation capabilities of six human and three bovine *L. ruminis* isolates (O' Donnell *et al.*, 2011). Comparison of the fermentation profiles and genome annotations of ATCC 25644 and ATCC 27782 allowed us to identify the enzymes and pathways that *L. ruminis* uses to ferment carbohydrates. The pathways identified include those for the degradation of α-galactosides, β-galactosides, α-glucosides, β-glucosides and β-fructofuranosides (Forde *et al.*, 2011; O' Donnell *et al.*, 2011). The degree of polymerisation (DP) was identified as an important factor is the fermentability of the carbohydrates tested, with high DP carbohydrates not being fermented, and carbohydrates with DP of ≤ 10 being readily fermented. The prebiotic fructooligosaccharide (FOS) was fermented by all of the humans strains tested. However, the bovine strain ATCC 27782 failed to ferment this carbohydrate and this was attributed to a lack of the enzyme beta-fructofuranosidase (Forde *et al.*, 2011; O' Donnell *et al.*, 2011).

Mammalian-associated lactobacilli and those consumed as components of foods and beverages encounter many stresses and variable conditions in the human gastrointestinal tract. Common intestinal stresses include gastric acidity and bile salts.

Bile salts are formed by the conversion of cholesterol in the liver and concentrations fluctuate between 0.3-0.5% *in vivo* (Dunne *et al.*, 1999). Bile salts are known to exert an antimicrobial effect on microorganisms *in vivo* (Hänninen, 1991). Persistence to such action is therefore essential for viable intestinal transit and survival of a mammalian-associated *Lactobacillus*. The lower intestine (caecum and colon) is a nutrient rich environment containing polysaccharides and non-digestible oligosaccharides (NDO) like prebiotics. The catabolic flexibility of a bacterium to utilise these NDO is a factor in its ability to survive or colonise these gastrointestinal regions. Antibiotic resistance is a global problem and studies have shown that horizontal transfer of antibiotic resistance genes between Gram-positive and Gram-negative genera can take place *in vivo* (Salyers *et al.*, 2004). Similarly, exchange of resistance genes has be shown between *Lactobacillus* spp. and other intestinal bacteria (Jacobsen *et al.*, 2007). Therefore determining a mammalian-associated *Lactobacillus* species or strain with increased resistance to certain antibiotics is important for host health and wellbeing. Developing a strain as a candidate probiotic also includes investigating the strains response to technological stresses such as high salt concentration, aerobic environment and temperatures (Champagne *et al.*, 2005).

Motility has previously been noted in the bovine isolates of *L. ruminis* (Neville *et al.*, 2012; Sharpe *et al.*, 1973). Motility has also been identified in other lactobacilli. However, the motility of these isolates was poorly characterised (Chao *et al.*, 2008; Deibel & Niven Jr, 1958; Harrison Jr & Hansen, 1950; Nielsen *et al.*, 2007). Neville *et al*. (2012) assessed the reportedly motile lactobacilli and noted that *L. ruminis* was the only mammalian-derived species with a motile phenotype. Transcriptomic analysis of a non-motile human isolate and a motile bovine isolate revealed there was a significant up-regulation of genes in the motility locus (Neville *et al.*, 2012). The flagellar components of bacteria have immunomodulatory properties whereby flagellin is recognised by toll-like receptor 5 (TLR5) and nuclear factor κB (NF- κB) (Hayashi *et al.*, 2001). Potential benefits of a cell maintaining its flagellar apparatus include offering a cell competitive advantage over other aflagellate species allowing better access to nutrients and adaptation to its niche. Flagellate *L. ruminis* cells induced a greater IL-8 secretory response than aflagellate cells (Neville *et al.*, 2012). Motility in bacterial cells can be classified as swimming or swarming. Swarming is a flagellar-driven movement of bacteria over a solidified agar surface (Harshey, 2003; Rather, 2005; Verstraeten *et al.*, 2008). Each swarming organism

appears to have its own "unique" mechanism for facilitating swarming (Partridge & Harshey, 2013). Examination of the microbiota of starch fed horses identified the presence of swarming *L. ruminis* on agar plates (Vörös, 2008; Willing *et al.*, 2009c). The swarming ability of bacteria is often cell density dependent and involves hyper-flagellation, cell differentiation and the possible involvement of polysaccharides and bio-surfactants (Sharma & Anand, 2002; Verstraeten *et al.*, 2008). The addition of bio-surfactants like Tween 80 has been shown to facilitate swarming and aid in the ease of measurement of a swarm halo (Niu *et al.*, 2005). FliL, a part of the type III flagellar export system and the switch complex has been shown to be a key component for swarming behaviours in Salmonella (Attmannspacher *et al.*, 2008).

There is a paucity of information on the genomic diversity of *L. ruminis*. A multilocus sequence typing (MLST) approach was therefore used to rapidly evaluate the diversity of a culture bank of *L. ruminis* isolates. MLST schemes involve the examination of the nucleotide variation in housekeeping genes which slowly accumulate over time (Roumagnac *et al.*, 2006). As the housekeeping genes encode essential and functional gene products they are not affected by rapid evolution which makes them an ideal target for assessing the genomic diversity of isolates. MLST has been employed to analyse the genomic diversity in other *Lactobacillus* species including *Lactobacillus salivarius* (Raftis *et al.*, 2011), *Lactobacillus plantarum* (de las Rivas *et al.*, 2006), *Lactobacillus casei* (Cai *et al.*, 2007a; Diancourt *et al.*, 2007) and *Lactobacillus sanfranciscensis* (Picozzi *et al.*, 2010). We focused on housekeeping genes routinely used and validated for MLST (Raftis *et al.*, 2011).

In this study, the survival characteristics and genomic diversity of the culture bank of *Lactobacillus ruminis* isolates was assessed and compared. The presence of *L. ruminis* in the majority of domesticated animal species and in humans highlights the need to characterise the species. This study also aimed to perform fermentation profiling and genomic identification of the pathways involved in carbohydrate utilisation for the newly identified porcine and equine strains. The study also aimed to characterise the swarming phenotype of two *L. ruminis* species using *in vitro*, molecular and next generation sequencing techniques.

6.2 Materials and Methods

6.2.1 Bacterial strains and culture conditions

Nine *Lactobacillus ruminis* strains previously investigated for motility and catabolic flexibility (Neville *et al.*, 2012; O' Donnell *et al.*, 2011) were further examined in this study. Six strains had been isolated from human faeces and three strains isolated from the bovine rumen. All strains were stored at -80°C in de Man-Rogosa-Sharpe (MRS) broth (Difco, BD, Ireland), supplemented with 25% (v/v) glycerol as a cryo-protectant. *Lactobacillus* strains were grown anaerobically on MRS agar plates at 37°C for two days. Growth tests were initiated by growing *Lactobacillus* strains anaerobically in MRS broth at 37°C overnight and unless otherwise stated, all further incubations were also performed under anaerobic conditions at 37°C (O' Donnell *et al.*, 2011).

6.2.2 Animals and diets

Faecal samples were collected from four Large White x Landrace cross weanlings and sows. The animals are housed in the pig production unit of Teagasc Moorepark, Fermoy, Co. Cork, Ireland. The age of the weanlings was approximately 10-12 weeks old. The diets mainly consisted of Barley, Wheat, Maize, Soya Full Fat, Soya Hi Pro, Fat, Amino Acids, Vitamins and Minerals.

Faecal samples were also collected from six mature racehorses. The horses used in this study were housed in a stable in Co. Limerick, Ireland. The horses were fed on diets containing forage and a high starch concentrate. All samples were collected in accordance with the current Irish legislation on animal handling.

6.2.3 Media and solutions

MRS and Raffinose-MRS were used as the plating media for the isolation of *L. ruminis* from porcine and equine faecal matter. Modifications were made to the MRS (De Man *et al.*, 1960) medium by the omission of dextrose and the addition of 0.5% (w/v) raffinose.

Carbohydrate-free MRS (cfMRS) (O' Donnell *et al.*, 2011) with added bromocresol purple was used as a basal screening medium to study the ability of the potential

Lactobacillus ruminis strains to utilise various carbohydrates. These carbohydrates were then used as a selective method to isolate *L. ruminis* based on its carbohydrate fermentation profile (O' Donnell *et al.*, 2011). The carbohydrate free MRS was supplemented with 0.5% (v/v) of cellobiose, Raftilose P95 (Beneo-Orafti, Mannheim, Germany), mannitol or ribose for screening the porcine faecal isolates while the additional carbohydrates glucose, lactose, raffinose, Raftiline HP (Beneo-Orafti, Mannheim, Germany) and sucrose were used in the screening of the equine faecal isolates. Mannitol and ribose were used as negative controls i.e. carbohydrates that *L. ruminis* is unable to metabolise.

6.2.4 Simulated gastric juice

To simulate the *in vivo* gastric environment, a sterile electrolyte solution (de Palencia *et al.*, 2008) containing NaCl 6.2 gL^{-1}, KCl 2.2 gL^{-1}, $CaCl_2$ 0.22 gL^{-1} and $NaHCO_3$ 1.2 gL^{-1} was used. Lysozyme and pepsin (Sigma-Aldrich) were added at a concentration of 0.01% and 0.3%, respectively. The pH of the solution was reduced to pH 2.0 using 1 M HCl. Five millilitre volumes of overnight cultures were centrifuged at 4000 x g for 10min. The cell pellets were then re-suspended in the simulated gastric juice and incubated for 24 h. Viable plate counts were performed after 0 h, 3 h and 24 h incubation.

6.2.5 Swarming behaviour assays

MRS was modified and prepared to characterise the swarming behaviour of the *L. ruminis* isolates: (i) containing increasing percentage of agar from 0.5% up to 3%; (ii) containing increasing concentrations of Tween 80 from 0.2% up to 1%; (iii) minimal MRS was prepared containing 0.5% (w/v) of four different carbohydrates – glucose, lactose, cellobiose and Raftilose P95.

6.2.6 Carbohydrate fermentation profiling

The porcine and equine *L. ruminis* strains were tested for their ability to utilise twenty-eight carbohydrates and compared to previously established carbohydrate utilisation profiles for the other nine strains (O' Donnell *et al.*, 2011). Each carbohydrate solution was filter sterilised into cfMRS at a concentration of 0.5% (v/v). A Synergy 2 plate reader (BioTek Instruments Inc., Vermont, US) with Gen5 software was used to measure absorbance at 0 hrs and a second reading at 48 hrs. The

carbohydrates tested include cellulose, dextran, esculin, lichenan, lyxose, Raftiline HP, Raftiline ST (Beneo-Orafti, Mannheim, Germany), ribose, sialic acid, sialyllactose, soluble starch, trehalose, melibiose, raffinose, GOS, GOS inulin, lactose, lactulose, beta-glucotriose B, cellobiose, Beneo P95 (Beneo-Orafti, Mannheim, Germany), Raftilose P95, Raftilose Synergy 1 (Beneo-Orafti, Mannheim, Germany), fructose, galactose, glucose, maltose, mannose, sucrose.

6.2.7 Bile salt tolerance, pH tolerance and EPS production

To assess the effect of increasing concentrations of porcine bile salts (Sigma Aldrich, Wicklow) and a range of pH values on *L. ruminis* isolates modifications were made to MRS. In the bile salt assay MRS was supplemented with 0.25-5% (w/v) porcine bile salts. In the pH assay the pH was reduced using acetic acid from pH 5.5 to 3.0 in pH 0.5 unit increments.

Exopolysaccharide (EPS) production was analysed using modified MRS supplemented with 70% (v/v) of filter sterilised glucose, sucrose and lactose. A strain was marked as a potential EPS producer if a mucoid or ropy colony formation could be identified (Wang *et al.*, 2008).

6.2.8 Antibiotic resistance

Rifampicin and chloramphenicol were chosen as exemplars of broad spectrum antibiotics. Each antibiotic was tested using sterile disks (Sigma Aldrich, Wicklow, Ireland) on MRS agar plates supplemented with each test strain. The disks were saturated with rifampicin (0.1-1µg/ml) and chloramphenicol (1-4µg/ml). The test plates containing the disks were the grown at 37°C for 48hrs. A strain was considered resistant if no zone of clearing was present surrounding the antibiotic disk.

6.2.9 DNA extraction and PCR amplification and identification of 16S rRNA genes

DNA was extracted from potential isolates using the Sigma Genelute Bacterial genomic DNA kit (Sigma Aldrich, Wicklow, Ireland). Universal primers 27F and 1492R (O' Donnell *et al.*, 2011) were used to amplify the 16S rRNA gene from isolated bacterial genomic DNA. 16S rRNA genes were amplified in a 50µl reaction mixture consisting of 45µl Platinum High Fidelity Supermix (Invitrogen, USA), each

primer at 25µM, 20ng of template DNA and water to make the reaction up to 50µl. Amplification conditions for the PCR included an initial denaturation step of 94°C for 2 min, followed by 35 cycles of 94°C for 20 s, 52°C for 30 s and 68°C for 2 min and a final extension step of 68°C for 10min. PCR products were checked for size and purity on a 1% (w/v) agarose gel using gel electrophoresis. PCR products were purified with the QIAquick PCR purification kit (Qiagen, USA). DNA sequencing of the amplified 16S rRNA region was carried out by Beckmann Coulter Genomics (Takely, UK). The primers used in this study are listed in Table S6.1.

6.2.10 API-ZYM and OPNG assays

The API-ZYM kit (bio-Merieux, France) was used to characterise enzyme activity in newly isolated *L. ruminis* strains. The tests were carried out as per the manufacturer's instructions with all tests were carried out in duplicate.

Beta galactosidase activity, in particular, was assayed using OPNG disks (Sigma Aldrich, Co. Wicklow, Ireland) as per the manufacturer's instructions. Each test was carried out in duplicate.

6.2.11 Growth in reconstituted skimmed milk

Reconstituted skimmed (RSM) was prepared as a 10% (w/v) solution and autoclaved at 121°C for 10 minutes. All strains were inoculated into the RSM at 1% (v/v) and incubated for 72 hours at 37°C. Following the incubation period the pH of the growth medium was recorded and adjusted by that of the negative control to identify the pH change.

6.2.12 Aerobic growth

Each strain was inoculated as a 1% (v/v) inoculum in 5ml of MRS overnight aerobically at 37°C. Optical density (OD) readings were recorded at time 0 and time 24.

To assess the effect of carbohydrates on aerobic growth in the porcine and equine strains, they were grown in glucose, Beneo P95 and raffinose at 0.5% (w/v). Growth was measured in the Gen5 plate reader at 37°C aerobically for 20 hours.

6.2.12 Assessment of strain motility

The sixteen *L. ruminis* strains were stained with a crystal violet based flagellar stain (BD Diagnostics). The procedure was carried out as outlined by the manufacturer.

The stain is used to demonstrate the presence and arrangement of flagella on a bacterial cell. Stained cells were then examined on an oil-immersion microscope using the 1000x lens and images captured using the Olympus DP50 camera attached to the microscope.

6.2.13 Phylogenetic analysis of 16S rRNA gene sequences

Sequence alignments were performed using the ClustalW application in BioEdit (Hall, 1999). MEGA (version 5) (Tamura *et al.*, 2011) was used to construct trees by using the neighbour-joining algorithm and the Kimura two-parameter substitution model. Branch support wasted by 1,000 replicate bootstrap tests in each analysis.

6.2.14 Multi Locus Sequence Typing

The nucleotide sequences of the following genes were used for MLST analysis: *ftsQ, nrdB, parB, pheS, pstB* and *rpoA*. Primers for each locus were designed using BioEdit (Hall, 1999). An approximately 800-bp internal fragment of each gene was amplified which allowed the accurate sequencing of a 600 - 760-bp fragment within each amplicon, using the primers specified in Table S6.1. Each PCR product was sequenced (Beckman Coulter genomics, Takely, UK) and trimmed using Bioedit. Different allelic sequences, with at least one nucleotide difference per allele, were assigned arbitrary numbers. A combination of seven alleles defined the allelic profile of each strain, and a unique allelic profile was designated with a sequence type (ST). Split decomposition analysis of the allelic profile data and individual alleles was performed using SplitsTree 4.8 (Huson & Bryant, 2006). Concatenated sequences (4,103bp) of the loci (ordered as *ftsQ, nrdB, parB, pheS, pstB, rpoA*) were generated using the Sequence type Analysis and Recombinatorial Tests (START2) software (Jolley *et al.*, 2001). One thousand replicate neighbour-joining bootstrap trees, using the Kimura 2-parameter method (Kimura, 1980) in MEGA version 5 (Tamura *et al.*, 2011), were constructed to determine phylogeny. The relatedness of the isolates was assessed using START2. Related STs were clustered in groups or lineages using BURST analysis. START2 was also used to determine the ratio of non-synonymous to synonymous polymorphisms (dN/dS ratio) for each locus (Jolley *et al.*, 2001). Statistical comparisons of the loci were carried out using the maximum chi-square analysis application in the START2 package.

6.2.15 Genome sequencing and genome comparisons

Genome sequences of human (S23) and equine (DPC 6832) isolates were generated (Macrogen, Seoul, Korea). The sequence data was obtained using the Illumina HiSeq 2000 reversible dye terminator system (Macrogen, Seoul, Korea) with average read lengths of 101bp. The functional assignment of predicted genes was performed using Metagene (Noguchi *et al.*, 2006) to predict open reading frames (ORFs) and BLASTP to annotate them using the NCBI database (Altschul *et al.*, 1990). Whole genome comparisons were made between the *L. ruminis* isolates using the Artemis Comparison Tool (ACT) (Carver *et al.*, 2005). The Blast ring image generator (BRIG) (Alikhan *et al.*, 2011) was used to create an image of the whole genome comparison of ATCC 25644, S23, DPC 6836 and ATCC 27782. The threshold levels used for the comparison were 99% and 90% sequence similarity.

To generate the whole genome phylogenetic trees the core genome of each of the four *L. ruminis* strains - ATCC27782, ATCC25644, S23 and DPC6832 – was predicted using the ortholog prediction software QuartetS (Yu *et al.*, 2011). The size of the core genome was 1,388 genes.

For each core gene, an out-group was chosen by blasting a representative gene (from ATCC25644) against a protein database (unpublished) of predicted genes for 33 *L. salivarius* strains taken from various environments (including human blood, intestines, faeces, gallbladder and saliva and also from animals and food). To be confident that the top blast hit was a homolog of the core gene, the following thresholds were used: e-value <= 1e-05, % ID >= 30 and alignment length of query gene >= 45%. This left 1,154 of the original 1,388 *L. ruminis* core genes to be used in the building of the phylogenetic tree.

ClustalW was used to align the five sequences from each core gene (4 *L. ruminis* plus the *L. salivarius* out-group). A similarity matrix (from Fitch distances) was generated for each alignment where the distance between two sequences was represented by the square root of the dissimilarity (i.e. 80%/0.8 similar, so 20%/0.2 dissimilar; 0.2^0.5 = 0.447). The distances for all core genes were summed and the neighbouring-joining algorithm was used to build the consensus tree with the summed *L. salivarius* distances specified as the out-group.

6.2.16 RNA isolation and transcriptome sequencing

L. ruminis ATCC 27782 and DPC 6832 were cultured anaerobically at 37°C for 18 hours in 5 ml aliquots of MRS media (swimming cells) and also on MRS agar plates containing 0.5% (w/v) agar (swarming cells) and 2% (w/v) agar (stationary cells) for 48 hours. The broth cultures were centrifuged at 4°C to harvest the cells that were immediately resuspended in 10ml of RNAprotect Bacteria Reagent (Qiagen, Germany). To each agar plate 10ml of RNAprotect Bacteria Reagent was added and the cells gently harvested using sterile spreaders and removed from the plate using a wide bore pipette tip into a fresh 50mL falcon tube. Subsequently each tube was centrifuged at 4000 x g for 15mins at 4°C. Total RNA was isolated according to the protocol for Gram positive bacteria outlined by the Roche High Pure Isolation kit (Roche, Indiana, USA), but with minor modifications. The lysozyme concentration used was increased to 100mg/ml. Additionally, this step was also merged with a bead beating step to ensure complete cell lysis, whereby the cells were incubated for 60 mins at 37°C shaking at 1400 rpm in a 2ml stock tube containing 0.1mm zirconia beads in an Eppendorf thermomixer. DNA was removed with the Turbo DNA-free kit (Invitrogen, Dun Laoghaire, Ireland). The total RNA was ribo-depleted using the Gram-Positive Bacteria Ribo-Zero™ Magnetic Kit (Cambio Ltd., Cambridge, UK) and cleaned using the RNA Clean & Concentrator™-5 (Cambridge Biosciences, Cambridge, UK).

6.2.16 RNA-seq pipeline analysis

Six tagged strand specific cDNA libraries were prepared. Each sample was sequenced on an Illumina HiSeq sequencer (GATC Biotech, Konstanz, Germany) to generate 101bp in length reads using the pair-end sequencing. Sample coverage ranged from 1,774 to 1,936-fold for the three ATCC 27782 samples and from 1748 to 2377-fold for the three DPC 6832 samples. FastaQC was used to identify the quality of the RNA-seq reads from each treatment (www.bioinformatics.babraham.ac.uk). The Trimmomatic program was used to trim low quality section of reads (Bolger & Giorgi; Lohse *et al.*, 2012). Alignment of the reads to the complete genome of ATCC 27782 and the draft genome of DPC 6832 was carried out using Bowtie2 (Langmead & Salzberg, 2012). HTSeq-count and DESeq were utilised to assess differential gene expression between stationary, swimming and swarming *L. ruminis* cells (Anders, 2010a; Anders, 2010b; Anders & Huber, 2010) .

6.2.17 RT-PCR

RT-PCR was used to confirm a selection of the differentially expressed genes identified by the RNA-seq data. The SensiFAST™ SYBR® No-ROX One-Step Kit (Bioline, myBio, Ireland) was used to generate the cDNA and carry out the RT-PCR analysis according to the manufacturer's specifications. The amplification temperature for all reactions was 55°C. The gene expression data generated for each condition (stationary, swimming and swarming) for the *L. ruminis* strains ATCC 27782 and DPC 6832 were normalised using the housekeeping gene r*ecA*. Following the normalisation of the data using the *recA* gene, the fold changes between the swimming cells vs. the stationary cells and the swarming cells vs. the stationary cells for both *L. ruminis* strains was calculated using the following formula: fold change = $2\wedge^{(\Delta\Delta Ct)}$. The standard deviation of the ΔCT was calculated from the standard deviations of the target and reference values using the formula: S.D. = $(S_1^2 + S_2^2)\wedge^{0.5}$. The resulting value was then added or subtracted (+/-) to the ΔΔCT values to generate a range for the $2\wedge^{(\Delta\Delta Ct)}$ values.

6.2.18 Nucleotide sequences

This Whole Genome Shotgun projects for *L. ruminis* DPC 6832 and S23 have been deposited at DDBJ/EMBL/GenBank under the accession AWYA00000000 and AWYB00000000, respectively. The version described in this paper is version AWYA01000000 and AWYB01000000, respectively.

6.3 Results

6.3.1 *L. ruminis* isolation

To expand the host-range and metabolic diversity of a strain panel for molecular characterisation, faecal samples from 4 sows, 4 weanlings and 10 horses were serially diluted (10^{-8}) and plated to identify new *L. ruminis* isolates. Two hundred and fifty-nine colonies from the sows and weanlings and 77 from horses were sub-cultured into MRS broth and grown anaerobically at 37°C for further phenotypic screening. Seventy percent (63/90) of the plates had swarming colonies. Isolation of single colonies from the equine faecal samples was particularly difficult due to the

abundance of swarming bacteria covering the plates. A similar level of swarming colony abundance was noted from faecal culture of Swedish racehorses (Willing *et al.*, 2009c).

6.3.2 Phenotypic screening

In a previous study we established the carbohydrate fermentation profile for nine *L. ruminis* strains of human and bovine origin (O' Donnell *et al.*, 2011). This established *L. ruminis* profile was used to screen the potential *L. ruminis* isolates from the stocked isolates of porcine and equine origin. From the 259 porcine isolates, 57 were identified as having a fermentation profile similar to that of *L. ruminis*. 25 of the 57 isolates were Gram Positive, catalase negative rods. A similar method was used for the 77 equine strains, whereby morphological and phenotypic screening reduced the number of isolates to 24.

6.3.3 16S rRNA gene sequencing and isolate identification

Genomic DNA of the 59 potential *L. ruminis* isolates was extracted and the 16S rRNA gene was amplified using an *L. ruminis*-targeting primer pair. Non-amplification reduced the number of isolates to 14 (6 porcine and 8 equine), from which the 16S rRNA gene was sequenced for six porcine isolates and eight equine isolates. Two porcine and 5 equine isolates were identified as *L. ruminis*. The 16S rRNA sequences of the 14 isolates were compared to the type strain ATCC 27780 and the results are shown in Table S6.2. 16S rRNA phylogenetic trees were created from the 16S rRNA gene sequences of the *L. ruminis* isolates (Figure S6.1). The *L. ruminis* isolates were arranged into 3 clades. The human and porcine isolates clustered together and formed Clade 1. The bovine and equine isolates formed Clades 2 and 3, respectively.

6.3.4 Carbohydrate fermentation profiling

The growth profiles of the seven confirmed *L. ruminis* isolates are summarised in Table 6.1. Similar to the human and bovine isolate fermentation profiles, the isolates were able to utilise mono/di/tri and tetra-saccharides. The porcine isolates were able to utilise lactose and lactulose for growth but were unable to utilise GOS and GOS inulin. The strain-dependent β-galactoside utilisation capabilities of the porcine and equine *L. ruminis* strains is consistent with the similar strain variability of the human

and bovine isolates (O' Donnell *et al.*, 2011). A particularly heterogeneous fermentation pattern was identified for the porcine and equine strains when grown on beta-fructofuranosides. The porcine isolates were weakly able to ferment sialic acid for growth. With the exception of sialic acid this fermentation profile is similar to the human and bovine isolates. The majority of strains were unable to ferment polysaccharides and inulins. However, DPC 6831 was able to weakly ferment cellulose. DPC 6831 and DPC 6835 were also able to ferment dextran and Raftiline HP. This would suggest that that the majority of *L. ruminis* isolates are unable to ferment carbohydrates with a DP greater than 10 (O' Donnell *et al.*, 2011). No demonstrable amylase activity was identified in any isolate which is considered a desirable trait for potential probiotics.

Table 6.1. Growth profiles for newly isolated *L. ruminis* strains on diverse carbohydrates

		***Lactobacillus ruminis* strains**						
		Porcine			Equine			
Carbohydrate class	**Carbohydrate**	DPC 6830	DPC 6831	DPC 6832	DPC 6833	DPC 6834	DPC 6835	DPC 6836
Mono and Di-saccharides	Fructose	+	+	++	+	++	++	+
	Galactose	++	++	++	++	++	++	++
	Glucose	-	+	+++	++	++	++	++
	Lyxose	-	+	-	-	-	-	-
	Maltose	+	+	+	++	++	++	-
	Mannose	+++	+++	+++	++	++	++	+++
	Ribose	-	-	-	-	-	-	-
	Sucrose	++	+++	+++	++	++	++	++
α-galactosides	Melibiose	++	++	+++	++	++	++	++
	Raffinose	++	++	++	++	+	++	-
	Stachyose	++	++	++	++	+++	+++	+++
β-galactosides	GOS	+	+	-	-	-	-	-
	GOS Inulin	+	++	-	-	-	-	-
	Lactose	+++	++	-	-	-	-	-
	Lactulose	++	++	-	-	-	-	-
β-glucosides	β-Glucotriose B	++	++	+++	++	++	++	++
	Cellobiose	+	+	++	+	+	-	-
β-fructofuranosides & Inulins	Raftiline HP	-	+	-	-	-	+	-
	Raftiline ST	+	++	+	+	-	-	++
	Raftilose P95	-	-	++	+	+	++	++
	Raftilose Synergy 1	-	++	+++	++	++	++	++
	Dextran	-	+	-	-	-	++	-
Polysaccharides	Esculin	-	+	-	-	-	-	-
	Lichenan	-		-	-	-	-	-
	Sialic acid	++	+	-	-	-	+	+
	Siallylactose	+	+	ND	++	-	-	-
	Soluble Starch	-	-	-	-	-	-	-
	Cellulose	-	+	-	-	-	-	-

- = no growth, + = poor growth, ++ = moderate growth, +++ = strong growth, ND = Not done.

6.3.5 Biochemical and metabolic characterisation

One of the overall aims was to determine if the extended panel of *L. ruminis* strains included isolates with biochemical/metabolic traits that might allow their further development as probiotics.

Exopolysaccharides produced by a bacterium have potential uses in the food and pharmaceutical industries. The results of the characterisation of potential EPS production in the *L. ruminis* isolates is shown in Table 6.2. Forty percent of the isolates had a positive "ropy" phenotype with all of the media (glucose, sucrose and

lactose) used. Thirty percent of the isolates were negative for any discernable EPS production. No growth and therefore no EPS production was identified from the equine isolates and bovine strain ATCC 27782 on lactose-MRS plates. Future studies will be needed to confirm these initial findings.

Antibiotic resistance is a global issue and a major health concern, therefore identification of resistance or susceptibility to various antibiotics is important when characterising new bacterial isolates. All isolates were susceptible to the broad spectrum antibiotic rifampicin. Seven isolates were resistant to up to 4μg/ml of chloramphenicol. The resistant strains included both ATCC 25644 and ATCC 27782 therefore they may not be suitable as probiotic strains.

Biochemical characterisation is an important tool in identifying potential nutrients and pathways used by a bacterium. API-ZYM is a semi-quantitative method that can be used to identify enzymatic activity from the 16 *L. ruminis* isolates. Table 6.2 shows the enzymatic profiling data generated using the API-ZYM strips. All of the strains tested were positive for leucine arylamidase, valine arylamidase, α-galactosidase, Napthol-AS-BI-phosphohydrolase, N-acetyl-β-glucoaminidase and acid phosphatase. β-glucosidase activity was identified in all of the human isolate strains, in DPC 6833 (equine) and ATCC 27782 (bovine). Weak β-glucuronidase activity was noted in some of the strains tested (L5, S36, 27781 and DPC 6831). No enzymatic activities were detected for the majority of strains for alkaline phosphatase, esterase (C4), esterase lipase (C8), lipase, trypsin, α-chymotrypsin, α-mannosidase or α-fucosidase. ONPG disks were used to detect the presence of β-galactosidase, an enzyme found in lactose-fermenting organisms. A yellow colour change indicative of the presence of beta galactosidase was obtained for all of the human, porcine and two bovine (ATCC 27780 and 27781) isolates. No colour change was seen for ATCC 27782 and all of the equine isolates suggesting the absence of β-galactosidase activity in these strains. This is concordant with the results obtained from the carbohydrate fermentation profiling and from the API-ZYM assays.

6.3.6 Resistance profiling

All of the strains were able to grow in porcine bile salts at a concentration of ≤0.5% (w/v). The equine and porcine strains had the highest resistance to the action of the bile salts *in vitro,* as shown in Table 6.2. All of the human isolate strains were

unable to grow in MRS with a pH below 5.5. This may indicate that all of these strains would be unable to survive the pH stress of gastric transit. Only the equine strains were able to tolerate the lower pH levels (3.5-3.0). The ability to tolerate and survive the enzymatic and pH stresses is essential for characterisation of potential probiotics. The results of the simulated gastric juice survival assay are summarised in Table 6.2 and in Figure S6.2. Variable strain-dependent reductions in cell numbers followed 3 hours incubation. After 24 hours all of the strains showed a complete loss of viability (data not shown). Isolates S23, DPC 6833 and DPC 6836 showed the best survival in simulated gastric juice with just over a 1 log reduction in cell numbers. Isolates L5, S21, S36 and ATCC 27780 showed the largest reduction in cell numbers with a 4-5 log reduction in cell numbers after 3 hours. All the reductions were statistically significant.

Table. 6.2 Resistance and biochemical characteristics of the human, bovine, porcine and equine *L. ruminis* isolates

Tests	Conc./Variable	Human					Bovine				Porcine	Equine					
		L5	S21	S23	S36	S38	25644	27780	27781	27782	DPC 6830	DPC 6831	DPC 6832	DPC 6833	DPC 6834	DPC 6835	DPC 6836
Resistance assays																	
Bile Salts	0.25%	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	0.50%	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	0.75%	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	1%	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+
	2%	-	-	-	+	+	+	-	-	-	-	+	+	+	+	-	-
	5%	-	-	-	-	+	-	-	-	-	-	-	+	-	-	-	-
pH	5.5	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	4.5	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+
	4	-	-	-	-	-	-	-	+	-	+	+	+	+	+	+	+
	3.5	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+
	3	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+
Chloroamphenicol	≤4µg/ml	S	S	S	S	R	R	S	S	R	R	S	S	R	R	S	R
Rifampicin	≤1µg/ml	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
Simulated gastric juice	Survival	54%	53%	86%	36%	60%	63%	44%	50%	45%	66%	67%	78%	85%	82%	71%	83%
Biochemical assays																	
OPNG		+	+	+	+	+	+	+	+	-	+	+	-	-	-	-	-
Leucine arylamidase (EC. 3.4.11.1)	API-ZYM	++	++	++	+	++	++	++	++	++	++	++	++	++	++	++	++
Valine arylamidase		++	++	++	+	++	++	++	++	+	++	++	+/-	+	++	++	++
Cystine arylamidase (EC. 3.4.11.3)		+	+/-	+/-	-	+/-	+	+	+/-	-	-	+/-	-	-	-	+/-	-
Acid phosphatase (EC. 3.1.3.2)		+/-	+/-	+	+/-	+/-	+/-	++	+	+/-	+	+	++	+	+	+/-	++
Naphthol-AS-BI-phosphohydrolase		+	+	+	+	+/-	+/-	+	+	+	+	+	+	+	+	+	+
α-galactosidase (EC. 3.2.1.22)		+	+	+	+/-	+/-	+	+	+	+	++	++	++	+	+/-	+/-	+
β-galactosidase (EC. 3.2.1.23)		++	++	++	++	++	++	++	++	-	++	++	-	-	-	-	-
β-glucuronidase (EC. 3.2.1.31)		+/-	-	-	+/-	-	+/-	-	+/-	+/-	-	+/-	-	-	-	-	-
α-glucosidase (EC. 3.2.1.20)		+/-	-	-	-	-	+	+/-	+/-	+	++	+	+	-	-	-	+
β-glucosidase (EC. 3.2.1.21)		+/-	+	+	+/-	+/-	+	-	-	+	-	-	-	+/-	-	-	-
N-acetyl-β-glucosaminidase (EC. 3.2.1.52)		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
α-mannosidase (EC. 3.2.1.24)		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
α-fucosidase (EC. 3.2.1.51)		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
α-chymotrypsin (EC. 3.4.21.1)		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Trypsin (EC. 3.4.21.4)		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Alkaline phosphatase (EC. 3.1.3.1)		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Esterase (C4)		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Esterase lipase (C8)		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Lipase (C14)		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Glucose-MRS	EPS	+	+	-	+	-	+	+	+	-	+	-	+	-	+	-	+
Sucrose-MRS		+	+	-	+	-	-	+	+	-	+	+	+	-	+	+	-
Lactose-MRS		+	+	-	+	-	-	+	+	-	+	+	-	-	-	-	-

++ strong positive; + positive; - negative; +/- weak; S susceptible; R resistant.

Technological stresses are a common occurrence during the processing of a candidate probiotic strain. The ability to tolerate technological stresses like an oxygen-rich environment and high saline conditions are therefore important first stage characteristics to identify, when screening a culture bank for strains of potential use and further testing as candidate probiotics. The technological traits of each isolate can be seen in Table 6.3. All of the isolates were able to grow in media supplemented with up to 3% NaCl. With the exception of the equine strains DPC 6835 and 6836 concentrations of NaCl above 4% was inhibitory to growth. The ability of a strain to grow in milk is a benefit for use in a dairy based delivery vector. Milk acidifying capacity was examined by growing each strain in milk over 72h. Acidification was monitored using pH levels and comparing each strain to the negative control (pH 6.3). DPC 6834 was unable to grow and acidify the milk. All of the other strains tested were able ferment milk with final pH ranging from pH 4.1 to pH 5.1. Future studies will be needed to assess the organoleptic characteristics of the *L. ruminis* fermented milk.

Oxygen tolerance is an advantageous trait for a strain as it allows the bacteria to survive in a variety of niches. The majority of the human and bovine strains were negatively affected by the aerobic environment. A median 81% reduction in final culture absorbance was noted for the human strains and a 73% of a reduction was noted for the bovine strains. Between 4-13% reduction in final culture absorbance was noted for the porcine and equine strains, respectively. This suggests that the porcine and equine *L. ruminis* strains are aero-tolerant and as such are suitable candidates for probiotic processing. However, reducing the concentration of the carbohydrate in the media (from 2% to 0.5%) resulted in a decrease in the porcine and equine isolates ability to grow in the aerobic environment (Table 6.3). Cells grown aerobically in glucose reduced final culture absorbance from between 51-92%. The porcine and equine strains grown in raffinose supplemented MRS resulted in a decrease in absorbance readings of 66-89%. The isolates were most affected by the aerobic environment when grown in Beneo P95 as a carbohydrate source with reduction in growth of between 59 and 98%.

Table 6.3 Technologically related phenotypic traits of the human, bovine, porcine and equine *L. ruminis* isolates

Tests	Conc./Variable	Human					Bovine				Porcine		Equine				
		L5	S21	S23	S36	S38	25644	27780	27781	27782	DPC 6830	DPC 6831	DPC 6832	DPC 6833	DPC 6834	DPC 6835	DPC 6836
NaCl	2%	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	3%	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	4%	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-
	6%	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Temp.	4°C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	30°C	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	37°C	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	45°C	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	55°C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Anaerobic	OD	1.7	2	2.1	2.1	2.2	2.2	2.2	2.2	2.2	2.3	2.3	2.3	2.3	2.3	2.3	2.4
Aerobic		0	1.7	0.4	0.4	0.3	0.5	0.5	0.6	2.2	2.2	2.2	2	2.2	1.4	2.2	2.1
Aerobic growth & Carbohydrate growth reduction %	Glucose	ND	ND	ND	ND	ND	ND	ND	ND	ND	64	59	51	92	63	59	72
	Raffinose	ND	ND	ND	ND	ND	ND	ND	ND	ND	73	66	80	89	93	72	94
	Beneo P95	ND	ND	ND	ND	ND	ND	ND	ND	ND	95	95	59	66	80	90	98
Milk acidification*	pH reduction	1.6	1.4	2	1.6	2	1.2	2	2.2	2.2	1.9	2.2	1.4	1.5	0.2	1.8	1.3

+ positive for growth; - negative for growth; * pH difference between the negative control and test strain. ND – not done

6.3.7 Assessment of motility

Motility of *L. ruminis* is also a strain-variable trait. Microscopic examination of the stained flagellar organelle revealed that, as noted previously (Neville *et al.*, 2012), all of the human isolate strains lacked any flagella or remnants of flagella (Figure 1 a-p). All of the bovine, porcine and equine isolates produced flagella. The bovine strains had one to two flagella attached to each cell. The porcine and equine strains had between 4 and 16 peritrichous flagella. The average number of flagella attached to the porcine and equine isolates was 6.

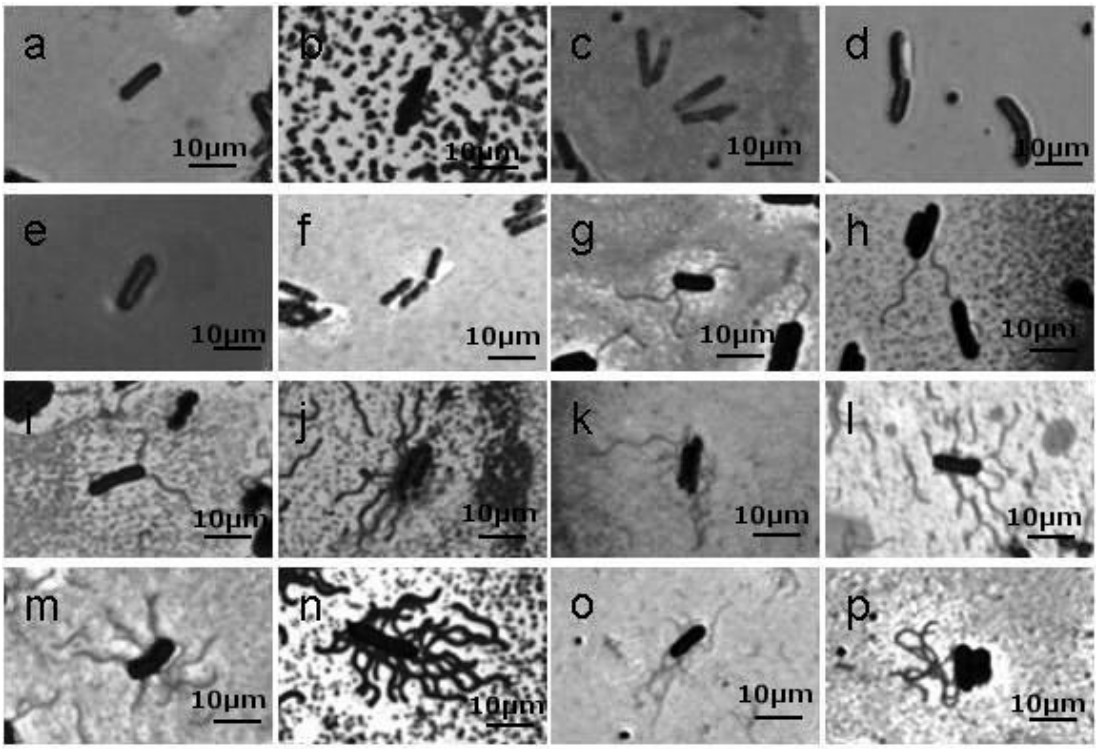


Figure 6.1. Flagella staining of 16 strains of *Lactobacillus ruminis* using light microscopy. (a) L5, (b) S21, (c) S23, (d) S36, (e) S38, (f) ATCC 25644, (g) ATCC 27780, (h) ATCC 27781, (i) ATCC 27782, (j) DPC 6830, (k) DPC 6831, (l) DPC 6832, (m) DPC 6833, (n) DPC 6834, (o) DPC 6835, (p) DPC 6836. Note: images (a-f) are strains which are non-motile and therefore lack a flagella apparatus.

Swarming is recognised as form of solid surface motility. Figure S6.3 shows representative data from the swarm assays for the porcine and equine strains. All of the porcine and equine strains were able to swarm on MRS with an agar concentrations ranging from 0.5% to 1.8% (w/v). These strains are therefore classified as hard swarmers (Butler *et al.*, 2010). ATCC 27782 was only able to swarm on MRS with 0.5% (w/v) agar which classifies it as a soft swarmer (Butler *et al.*, 2010). None of the human isolate strains or the other bovine isolate strains (ATCC 27780 and ATCC 27781) had the ability to swarm. The presence of increasing concentrations of the biosurfactant, Tween 80, had no stimulatory effect on swarming. All of the porcine and equine isolates were able to swarm at the lowest concentration of Tween 80 (0.1% v/v) present in MRS media as standard. Altering the carbohydrate and reducing the concentration from 2% to 0.5% (w/v) negatively impacted the ability to swarm. ATCC 27782 was unable to swarm under any of the Tween 80 or carbohydrate conditions tested. The porcine and equine isolates were

only able to swarm on the plates containing 0.5% (w/v) of glucose. Therefore, carbohydrate and agar concentrations are key factors in the ability of a strain to swarm.

6.3.8 Transcriptome analysis by RNAseq

RNA sequencing was carried out to generate molecular data to understand motility differences in strains. Lawley and colleagues recently used the same technology to identify the genes differentially expressed in the aflagellate human strain (L5) when grown in MRS and grown in MRS supplemented with cellobiose. The cellobiose supplemented MRS media restores the swimming phenotype to these cells (Lawley *et al.*, 2013). In our study, we aimed to examine the expression of genes in two strains both of which are naturally motile and are also able to swarm on a solid agar surface. The swimming and swarming motility phenotypes are important for bacterial survival and allow a cell to gain access to nutrients or move away from a repellent. Three conditions (swimming, swarming and stationary) were analysed for the two *L. ruminis* strains (ATCC 27782 and DPC 6832). The six samples analysed each mapped with a high percentage score to their respective genomes, an average of 99.67% and 97.13% for *L. ruminis* ATCC 27782 and DPC 6832, respectively. The lower percentage mapping of the DPC 6832 genome may be due to the draft quality of the genome. The average total number of aligned sequences for *L. ruminis* ATCC 27782 and DPC 6832 was 30,227,006 and 30,577,156, respectively. We performed a non-replicate based RNA-seq method as a high throughput screening method to identify significantly differentially expressed swimming or swarming-associated genes. The results were then used to identify a select number of genes for further examination with qRT-PCR. Seventy-four genes and 83 genes were identified as being statistically significantly differentially expressed from the RNA-seq data in motile (swimming or swarming) cells of *L. ruminis* ATCC 27782 and *L. ruminis* DPC 6832, respectively when compared to the control (stationary growth on agar plates). These statistically significant differentially expressed genes in Table S6.3 and Table S6.4 for *L. ruminis* ATCC 27782 and DPC 6832, respectively. From the RNA-seq data we selected 15 genes for further studies whose functions were divided between flagella biosynthesis, carbohydrate utilisation and uncharacterised hypothetical proteins. These 15 genes were examined with qRT-PCR in triplicate to quantify and confirm the differential expression identified in the RNA-seq data. The data generated

from the 15 genes for both RNA-seq experiment and RT-PCR can be seen in Table 6.4 and Table 6.5 for ATCC 27782 and DPC 6832, respectively. The majority of the differentially expressed genes identified in *L. ruminis* ATCC 27782 were identified as ribosomal proteins (Table S6.3) and essential for growth and proliferation of cells in general and were therefore excluded from further analysis.

Table 6.4. Genes differentially regulated in *Lactobacillus ruminis* ATCC 27782

Primer pair	ATCC 27782								Function
	ID	RNA-seq					RT-PCR		
		Swimming vs. Stationary log2 fold change[a]	pval	Swarming vs. Stationary log2 fold change[b]	pval	Fold change	2ΔΔCT Fold Change Swimming vs. Stationary[c]	2ΔΔCT Fold Change Swarming vs. Stationary[d]	
MMOD 1	LRC_00640	-2.97	*	-0.03	>0.05	8.00			hypothetical protein
MMOD 2	LRC_00780	-5.54	***	-0.94	>0.05	24.00	0.02 (0.02-0.02)	7.09 (6.35-7.93)	DeoR family transcriptional regulator
MMOD 3	pfkB	-4.10	**	-1.54	>0.05	6.00	0.01 (0.01-0.01)	0.08 (0.06-0.10)	1-phosphofructokinase
MMOD 4	LRC_00800	-3.18	*	-0.74	>0.05	5.00	0.2 (0.19-0.21)	0.19 (0.16-0.23)	PTS system fructose-specific
MMOD 5	LRC_03250	2.62	>0.05	-0.60	>0.05	9.00	17.92 (17.39-18.46)	0.22 (0.18-0.27)	hypothetical protein
MMOD 6	LRC_04370	0.96	>0.05	1.00	>0.05	1.03	1.06 (1.01-1.10)	0.22 (0.18-0.26)	hypothetical_protein
MMOD 9	LRC_05780	4.04	**	0.89	>0.05	9.00	28.54 (28.01-29.08)	0.27 (0.23-0.32)	hypothetical protein
MMOD 10	LRC_06170	0.83	>0.05	-1.16	>0.05	3.99	0.74 (0.70-0.77)	0.41 (0.33-0.50)	flagellin
MMOD 11	iD=LRC_04600	1.35	>0.05	2.42	>0.05	2.10	8.31 (7.94-8.70)	0.62 (0.51-0.75)	hypothetical_protein
MMOD 12	fliC	1.27	>0.05	0.16	>0.05	2.15	1.83 (1.83-1.83)	0.44 (0.36-0.54)	flagellin
MMOD 13	LRC_15700	1.07	>0.05	0.20	>0.05	1.82	1.48 (1.41-1.55)	0.31 (0.25-0.38)	flagellin
MMOD 14	LRC_18780	5.11	**	-0.48	>0.05	48.00	94.03 (89.54-98.74)	0.19 (0.16-0.23)	PTS system sucrose-specific transporter subunit IIABC
MMOD 15	LRC_16260	3.70	*	1.16	>0.05	6.00	2.13 (2.06-2.21)	0.05 (0.04-0.07)	hypothetical protein

a – negative values indicate a down-regulation in the swimming cells; b – negative values indicate a down-regulation in the swarming cells; c – values below 1 indicate a down-regulation of swimming cells; d – values below 1 indicate a down-regulation of swarming cells

Table 6.5. Genes differentially regulated in *Lactobacillus ruminis* DPC 6832

Primer pair	DPC 6832								Function
	ID	RNA-seq					RT-PCR		
		Swimming vs. Stationary log2 fold change[a]	pval	Swarming vs. Stationary log2 fold change[b]	pval	Fold change	2ΔΔCT Fold Change Swimming vs. Stationary[c]	2ΔΔCT Fold Change Swarming vs. Stationary[d]	
MMOD 1	LRN_87	-3.36	**	1.81	>0.05	36	12.92 (9.72-17.18)	105.18 (95.45-115.89)	hypothetical protein
MMOD 2	LRN_108	-5.11	***	2.66	*	218	2.80 (2.40-3.24)	4.34 (3.21-5.87)	DeoR family transcriptional regulator
MMOD 3	LRN_109	-3.94	**	3.62	**	189	0.06 (0.06-0.07	7.80 (6.01-10.13)	1-phosphofructokinase
MMOD 4	LRN_110	-2.93	*	4.29	**	149	0.02 (0.02-0.03)	6.06 (4.40-8.35)	PTS system fructose-specific
MMOD 5	LRN_324	2.50	*	-0.86	>0.05	10	0.97 (0.74-1.27)	0.11 (0.08-0.14)	hypothetical protein
MMOD 6	LRN_409	-4.62	***	2.28	>0.05	120	0.03 (0.02-0.04)	1.03 (0.81-1.29)	hypothetical_protein
MMOD 7	LRN_520	-0.74	>0.05	3.39	**	18	0.49 (0.36-0.68)	7.35 (5.57-9.71)	beta-fructofuranosidase
MMOD 8	LRN_521	0.06	>0.05	4.49	**	22	0.16 (0.12-0.21)	6.07 (4.61-7.98)	MFS Transporter Beta fructofuranosidase
MMOD 9	LRN_561	1.83	>0.05	-1.15	>0.05	8	0.58 (0.43-0.77)	0.12 (0.09-0.15)	hypothetical protein
MMOD 10	LRN_598	1.24	>0.05	0.67	>0.05	1.48	0.24 (0.18-0.33)	0.46 (0.35-0.62)	flagellin
MMOD 11	LRN_933	5.15	***	3.24	*	4	1.42 (1.20-1.70)	0.72 (0.53-0.98)	hypothetical_protein
MMOD 12	LRN_1405/1777	1.70	>0.05	2.89	*	2.28	0.46 (0.33-0.62)	2.09 (1.54-2.84)	flagellin
MMOD 13	LRN_1410	1.81	>0.05	3.04	*	2.36	0.3 (0.27-0.32)	1.78 (1.31-2.42)	flagellin
MMOD 14	LRN_1655	0.94	>0.05	1.84	>0.05	1.87	0.23 (0.18-0.29)	0.27 (0.20-0.36)	PTS system sucrose-specific transporter subunit IIABC

a – negative values indicate a down-regulation in the swimming cells; b – negative values indicate a down-regulation in the swarming cells; c – values below 1 indicate a down-regulation of swimming cells; d – values below 1 indicate a down-regulation of swarming cells

From the analysis of the RNA-seq data, no statistically significant differential gene expression was observed for the flagellar locus of ATCC 27782. However, nineteen flagellar locus genes were statistically significantly differentially expressed in DPC 6832 (Table S6.4**)**. Of particular interest were the two gene copies of flagellin that were up-regulated in both the swimming and swarming cells in DPC 6832. However, the up-regulated expression of the flagellin genes was not observed for the swimming cells examining the RT-PCR 2ΔΔCT fold change results. The RT-PCR expression data would suggest that the flagellin genes are extremely important for swarming DPC 6832 cells but not for swimming. While not significantly expressed in the RNA-seq dataset the two copies of flagellin were up-regulated in the swimming cells of ATCC 27782 and this trend was also reflected in the RT-PCR 2ΔΔCT fold change results. Examination of the statistically significantly expressed genes in both strains revealed that the fructose utilisation operon (LRN_108-110) was down-regulated in both test conditions in ATCC 27782 and significantly up-regulated in the swarming cells of DPC 6832 in the RNA-seq dataset. However, examination of the RT-PCR data showed that the DeoR fructose transcriptional regulator (LRN_108 & LRC_00780) was also up-regulated in swarming ATCC 27782 cells and in the swimming DPC 6832 cells. Other carbohydrate metabolism genes were significantly differentially expressed in both strains. The sucrose PTS transporter (LRC_18780) was significantly up-regulated in swimming ATCC 27782 cells in both datasets. This suggests that this transporter plays an unrecognised but important role in motility in *L. ruminis* ATCC 27782. Two genes that form part of the fructooligosaccharide utilisation operon (LRN_520-521)) were up-regulated in swarming cells in DPC 6832 in both the RNA-seq and RT-PCR datasets. A number of hypothetical proteins were also identified as being significantly up or down-regulated. The hypothetical proteins (LRC_03250/LRN_324 and LRC_05780/LRN_561) were up-regulated in the swimming cells of ATCC 27782 and DPC 6832. However, in DPC 6832 only LRN_324 was up-regulated in the RT-PCR and RNAseq dataset. This suggests that both hypothetical proteins are important for swimming cells in ATCC 27782, while only LRN_324 is important for swimming in *L. ruminis* DPC 6832. The hypothetical protein (LRN_87) may be important for swimming and swarming cells in DPC 6832 with a large up-regulation of this gene noted in the data generated from RT-PCR. Hypothetical proteins unique to a particular strain may also play a part in swimming or swarming in their respective strains; for example LRC_16260 may be essential for

motility in ATCC 27782. This hypothetical gene was up-regulated in swimming cells in both datasets. However, while these hypothetical proteins appear to be important for swimming and swarming motility in *L. ruminis* future work will need to be carried out to identify the function of each of these proteins and verify their importance in the different motility phenotypes.

6.3.9 MLST

Genomic DNA was isolated from 16 isolates, the nine strains we previously examined (O' Donnell *et al.*, 2011) and the 7 newly isolated strains from pigs and horses. The sequences of six loci were determined for each isolate and allelic profiles were assigned. The alleles defined for the MLST scheme were based on gene regions with sequence lengths ranging from 616 bp to 765 bp. The 16 isolates were assigned into 9 STs, 4 of which only occurred once (only one member in each). The strain with a complete genome sequence, ATCC 27782 (Forde *et al.*, 2011), was assigned as ST-1 and was found to be unique in this data set. In this study, the small number of isolates and loci did not allow the identification of the most prevalent ST.

The Neighbor-joining trees can be considered as robust due to the high bootstrapping values (Figure 6.2). Three major clades were identified from the concatenated sequence tree (Figure 6.2 (a)), Clade A contained all of the human derived isolates; Clade B contained the bovine and porcine isolates and Clade C contained the equine isolates. When examining each locus individually (Figure 6.2 (b)) analysis of 4 loci (*ftsQ, nrdB, pheS, pstB*) produced the same 3 clades as the concatenated tree. However, *rpoA* showed 2 clades, Clade AB combining all of the human, bovine and porcine isolates into a single clade and Clade C containing the equine isolates. Examination of *parB*-based trees showed 2 clades, Clade BC combining all of the bovine, porcine and equine isolates. For all loci tested, the bovine and porcine strains clustered with each other indicating that while they are from different hosts these strains are closely genetically related.

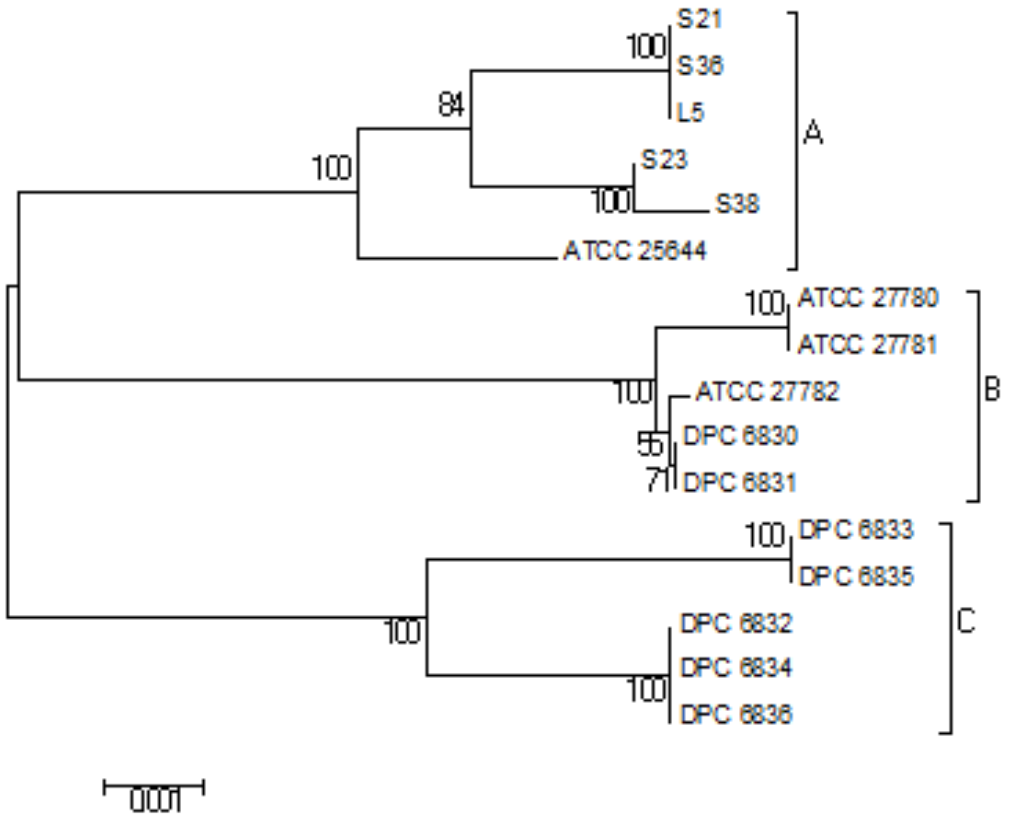


Figure 6.2 (a) Neighbor-joining tree for the concatenated sequences for all loci

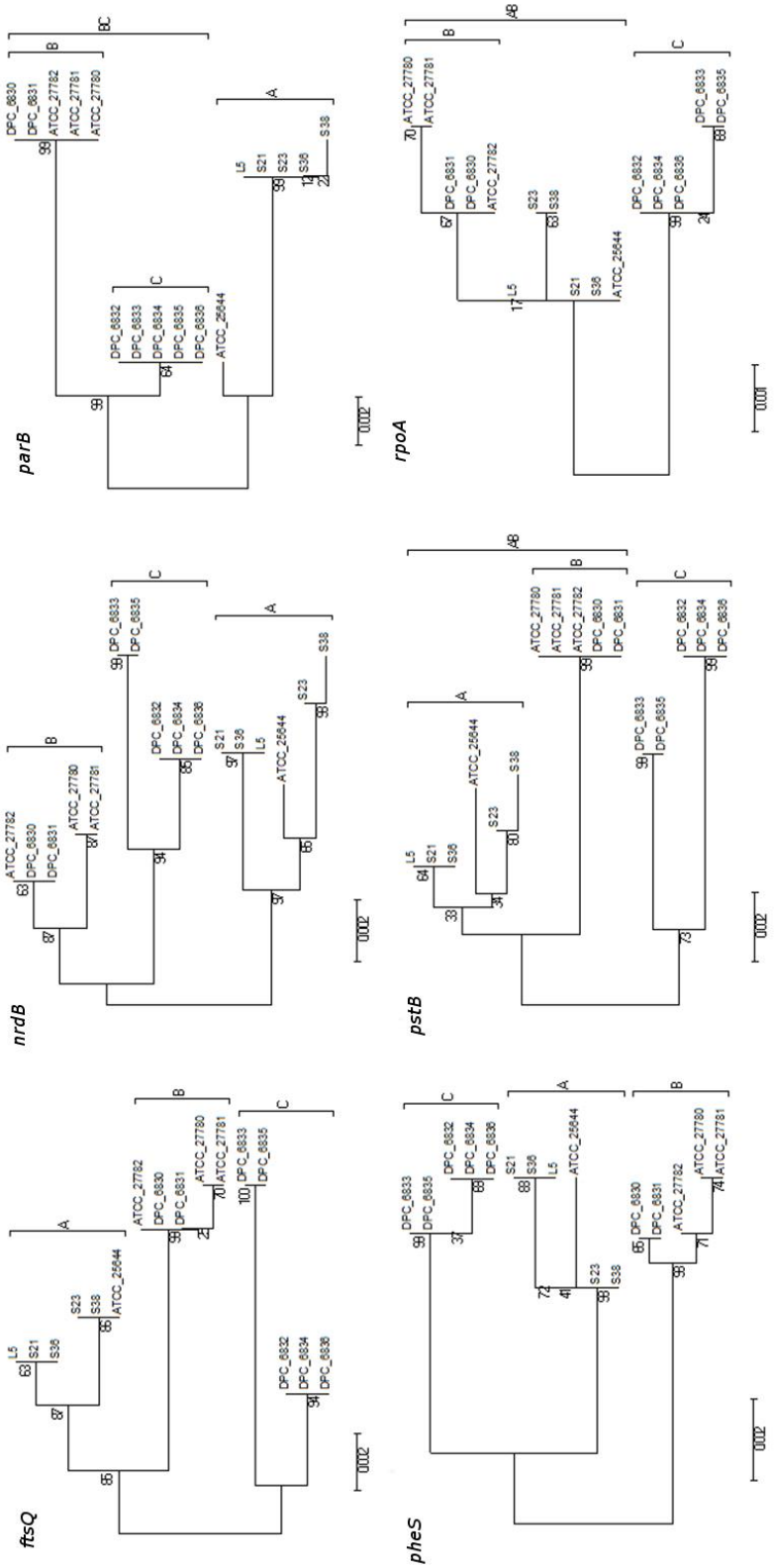


Figure 6.2 (b) Neighbour-joining phylogenetic trees for the MLST housekeeping genes

Polymorphisms were noted for all six loci tested. The number of polymorphisms varied between 9 (*rpoA*) and 23 (*nrdB*), which suggests that each locus has a different rate of evolution. Between 6 and 8 alleles were observed for each locus (Table S6.5). The average number of alleles at each locus was 3.8. All six loci are considered to be under a stabilising selective pressure as most of the noted substitutions were synonymous. The dN/dS ratios for each locus are listed in Table S6.6. In-frame concatenated gene fragments from all loci were analysed using Splits decomposition and the results suggest that intragenic recombination has occurred (Figure S6.4). Similarly, three of the splits decomposition trees generated for *ftsQ*, *pheS* and *pstB* also suggest intragenic recombination for these loci. The remaining loci had a tree like structure indicative of a clonal structure. No statistically significant recombination event was identified using the Sawyer's Run test. The linkage disequilibrium between alleles was calculated from the I_A value (3.4541). This value is significantly (P=0.000) higher than 0 which is also indicative of clade/clonal population identification. The standardised I_A value (I^S_A), was 0.6908 and indicates a low level of recombination within the loci.

6.3.10 Genome sequencing and comparisons

To complement the genome sequences already generated (Forde *et al.*, 2011), two other *L. ruminis* sequences were selected for sequencing. These strains were chosen based on the carbohydrate flexibility and in the case of DPC 6832 motility they exhibited *in vitro*. The genomes of *L. ruminis* S23 and DPC 6832 were 1,905,680 bp and 1,953,752 bp in length, respectively. The GC% content of the genomes of *L. ruminis* S23 and DPC 6832 was 42.96% and 42.87%, respectively. There were 1907 CDS and 46 tRNA's present in the genome of *L. ruminis* S23. There were 1806 CDS and 20 tRNA's present in the genome of *L. ruminis* DPC 6832.

The Blast ring image generator (BRIG) was used to visually compare the sequenced *L. ruminis* strains to the complete reference genome of ATCC 25644 (Figure 6.3). The comparison revealed large regions of similarity (99%) interspersed with small regions of dissimilarity and gaps. Examination of the BRIG image and manual curation of genomes aligned with the Artemis Comparison Tool (ACT) revealed that gaps and regions of dissimilarity in the sequence alignments were due to phage-related, hypothetical, CRISPR and restriction modification proteins. To complete the carbohydrate catabolic flexibility assessment of the *L. ruminis* isolates,

the carbohydrate operons in each genome were compared using ACT and the percentage identities between each operon are shown in Figure S6.5. A mannose PTS (mannose PTS1) operon present in ATCC 25644 was also identified in S23, which suggests that this operon is a "human" isolate only operon. However, a second mannose PTS (mannose PTS2) operon and one of the lactose operons (*lacZ*2) (O' Donnell *et al.*, 2011) were only present in ATCC 25644. A high level of conservation (95-99% both at the nucleotide and amino acid level) was noted for the raffinose, glycogen, sucrose, fructose operons and the third mannose PTS operon (mannose PTS3). A fragment of the lactose operon (lacZ1) was also identified in the genome of strain S23. It consisted of the β-galactosidase enzyme and GPH transporter but lacked the *lacI* regulator. A fragment of the maltose ABC operon was identified in the genome of strain S23. The genome of strain S23 contains only one of the two operons for lactose and maltose utilisation that are present in *L. ruminis* ATCC 25644. The fragmented operons may be as a result of gaps in the draft genome of S23 as the carbohydrate fermentation profiles of revealed the ability to ferment both lactose and maltose. A similar level of similarity was observed when using the complete genome of ATCC 27782 as the reference genome in the BRIG analysis (Figure S6.6).

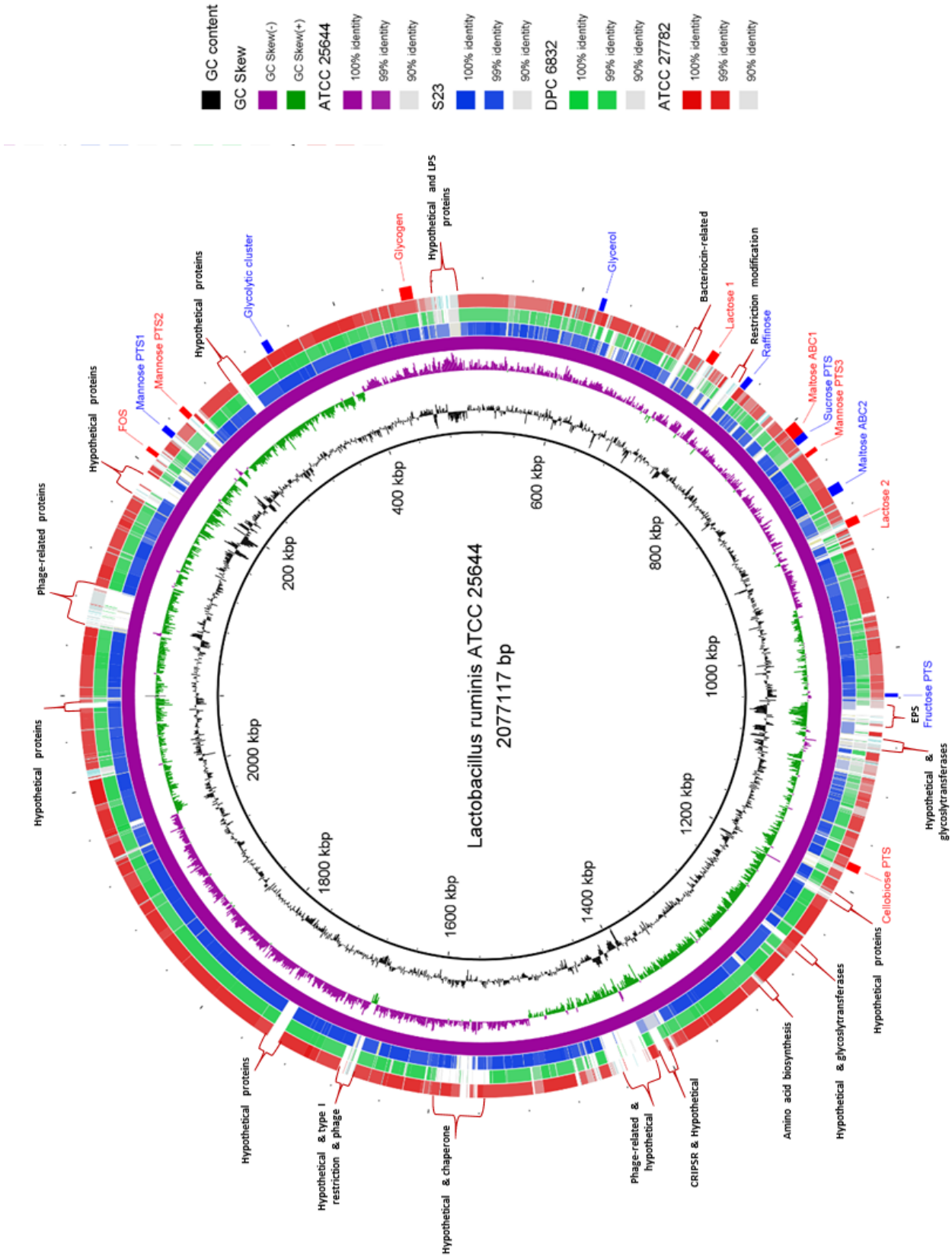

GC content
GC Skew
GC Skew(-)
GC Skew(+)
ATCC 25644
100% identity
99% identity
90% identity
S23
100% identity
99% identity
90% identity
DPC 6832
100% identity
99% identity
90% identity
ATCC 27782
100% identity
99% identity
90% identity
Lactobacillus ruminis ATCC 25644
2077117 bp
200 kbp
400 kbp
600 kbp
800 kbp
1000 kbp
1200 kbp
1400 kbp
1600 kbp
1800 kbp
2000 kbp
Hypothetical and LPS proteins
Glycogen
Glycerol
Bacteriocin-related
Lactose 1
Restriction modification
Raffinose
Maltose ABC1
Sucrose PTS
Mannose PTS3
Maltose ABC2
Lactose 2
EPS
Fructose PTS
Hypothetical & glycosylytransferases
Cellobiose PTS
Hypothetical proteins
Hypothetical & glycosylytransferases
Amino acid biosynthesis
CRIPSR & Hypothetical
Phage-related & hypothetical
Hypothetical & chaperone
Hypothetical & type I restriction & phage
Hypothetical proteins
Hypothetical proteins
Phage-related proteins
Hypothetical proteins
FOS
Mannose PTS1
Mannose PTS2
Hypothetical proteins
Glycolytic cluster

Figure 6.3. Blast ring image generator comparison of the sequenced *L. ruminis* genomes ATCC 25644, S23, DPC 6832 and ATCC 27782 using a 90-99% similarity threshold. The rings from the central nucleotide scale marker outwards, are GC% content, GC skew, ATCC 25644 (reference genome), S23 genome, DPC 6832 genome and ATCC 27782 with % identity to the reference genome colour coded as per the legend to the right. Carbohydrate genes and operons are marked in alternating red and blue colours; hypotheticals, phage-related proteins and other gaps in the sequences are marked in black.

6.3.11 Whole genome phylogeny

To complement the carbohydrate utilisation comparison, and strain relatedness analysis by MLST, we performed whole-genome phylogeny for the four *L. ruminis* genomes available. Comparisons were made using a core gene set present in the four *L. ruminis* genomes and also in the *L. salivarius* out-group. The results of the whole genome phylogenetic tree generated using the 33 *L. salivarius* strains as an out-group in Figure S6.7. The results showed a clustering of the genome sequences from the human-derived strains suggesting the core genes of these strains have independently adapted to life as human microbiota commensals. From the data and the tree generated it is clear that DPC 6832 is the most divergent of *L. ruminis* strains and that ATCC 25644 and S23 (both human isolates) are more closely related.

6.4 Discussion

Lactobacillus ruminis is an autochthonous species present in the mammalian microbiome (Reuter, 2001). In this study, we aimed to determine the genomic diversity, biochemical and metabolic characteristics of the known *L. ruminis* isolates. To date *L. ruminis* has only been isolated and identified in the lower intestines and has therefore, been overlooked as a potential probiotic with the ability to maintain cell viability under upper gastrointestinal tract conditions. A battery of tests were carried out to simulate the conditions faced by a strain as it migrates through the gastrointestinal tract (Dunne *et al.*, 1999). In this study, 63% of the strains (n=10) showed an ability to survive the simulated gastric juice (at greater than 60% of their original population numbers) *in vitro*. The survival rates for *L. ruminis* in SGJ (36-

85%) were similar to those of human isolates of *L. plantarum* (Başyiğit Kılıç *et al.*, 2013) . This data indicates that *L. ruminis* has the potential to survive gastric transit at as high cell numbers as other robust lactobacilli. All of the strains showed resistance and the ability to grow in media containing up to 0.75% (w/v) bile salts. This is greater than the levels estimated to be found in the intestines (Dunne *et al.*, 1999). Testing with increasing concentrations of bile salts greater than those found *in vivo* revealed that 44% of the isolates were able to grow in the presence of up to 2% bile salts. This is consistent with similar tests carried out on other human-derived *Lactobacillus* spp. including *L. ruminis* isolated from the faecal samples of healthy Spanish volunteers (Delgado *et al.*, 2007a; Delgado *et al.*, 2007b; Karasu *et al.*, 2010). Low pH was identified as a major growth limiting factor for *L. ruminis* strains with less than half of the isolates tested able to survive pH of 4.5 and 4. Similar levels of survival at pH 4.5 was noted by Delgado et al. using other *L. ruminis* strains (Delgado *et al.*, 2007a). The data generated here and by Delgado and colleagues (Delgado *et al.*, 2007a) suggests that the *L. ruminis* species has a high tolerance to bile salts and that human-derived strains are susceptible to acidic pH. But, all equine isolate strains tested here were able to survive and maintain minimal growth at pH 3.0. This suggests that the equine strains have evolved a greater tolerance to low pH and this was also reflected in the response of these strains to SGJ. Antibiotic resistance is a global problem for healthcare providers and human and animal health. The possibility of horizontal transfer of resistance genes *in vivo* means it is important to assess a strains resistance to a variety antibiotics (Salyers *et al.*, 2004). Due to the high level of aminoglycoside antibiotic resistance among the *Lactobacillus* species (Danielsen & Wind, 2003) and some initial tests carried out in this study (data not shown) they were omitted.Using the EFSA guidelines (EFSA, 2012) 44% of the *L. ruminis* strains were de-selected from the probiotic assessment based on their resistance to up to 4ug/ml of chloramphenicol. However, due to noted resistance and safety concerns chloramphenicol is no longer used as common antibiotic in medicine (Yunis, 1989) and these resistant strains may be revisited in the future for further probiotic assessment. All of the isolates were susceptible to rifampicin.

The catabolic flexibility of mammalian-derived lactobacilli is important for their survival in the gastrointestinal tract (O' Donnell *et al.*, 2013b). Assessment of the prebiotic utilisation of each individual strain has the potential to allow for the creation of targeted synbiotic products. The ability of each *L. ruminis* strain tested to

ferment at least one class of prebiotic carbohydrate is indicative of its adaptation to the lower gastrointestinal tract rich in NDO. The combination of a prebiotic with a *L. ruminis* strain could be used to modulate the microbiota of human and animals. Further testing would be required to assess the efficacy of the treatment on the microbiota. The additional copies of the lactose and maltose operons identified in the genome of ATCC 25644 when compared to the other human-derived strain S23, is indicative of horizontal transfer from another species present in the human microbiome.

Technological assessment of the potential probiotics was assessed by monitoring growth in a high saline environment. All strains were able to grow up to 3% NaCl and the majority of the isolates were able to tolerate and grow in 4% NaCl. No growth was identified for any isolate in media supplemented with 6% NaCl, indicating that NaCl concentrations between 4% and 6% exert an inhibitory effect on the *L. ruminis* strains. The ability to grow and survive in an aerobic environment is also a positive technological attribute for a potential probiotic. Aerobic conditions negatively impacted the growth of the majority of human and bovine isolates. The porcine and equine isolates showed very little inhibition in their growth when exposed to the aerobic environment. The ability to survive in the aerobic and saline environments suggests that the equine isolates of *L. ruminis* should be considered as probiotic candidates.

The β-galactosidase activity of potential probiotics may be a positive attribute in individuals suffering from lactose intolerance (de Vrese *et al.*, 2001). Six isolates had no β-galactosidase activity but the remaining isolates (n=10) were able to ferment β-galactosides. The presence of β-galactosidase enzymatic activity in the human, bovine and porcine strains is most likely a niche adaptation. Humans, steers and weanlings (from which both porcine strains were identified) are more likely to have consumed milk and other lactose products. An exception to this was ATCC 27782, a bovine isolate strain, which lacks the ability to utilise lactose. The horses used in this study were mature racehorses and had not received any lactose-related feed in many years. All of the equine isolates like ATCC 27782 were unable to utilise lactose.

The MLST scheme described here showed high discriminating powers since it was able to differentiate between highly similar isolates. Unlike other MLST schemes (de las Rivas *et al.*, 2006) and studies we found an association between ST, clades and the isolation source of each strain. The clade groupings identified by MLST were

divergent from those identified from sequencing of the 16S rRNA gene (Figure 6.1 (b)). This highlights the need to use a multi-testing approach for the identification of strains and species. The efficacy of MLST for the comparison of the genetic structure of bacterial populations is based on the ability of housekeeping genes to have selectively unbiased variability (de las Rivas *et al.*, 2006). The dN/dS ratios for each locus were less than 1, which indicates that they are not subject to positive selection and have neutral variability and were therefore suitable for use in the MLST scheme. Comparing the housekeeping gene nucleotide diversity estimated values to other lactobacilli (de las Rivas *et al.*, 2006; Diancourt *et al.*, 2007) revealed that the *L. ruminis* values were higher. This data suggests a higher level of polymorphisms present in the housekeeping genes examined in *L. ruminis*. This higher value may be related to different housekeeping genes used by the MLST scheme to generate the nucleotide diversity estimates. The application of this MLST analysis scheme on larger numbers of *L. ruminis* isolates could improve our knowledge of *L. ruminis* population structure.

Some of the phenotypic analyses corroborated the groupings identified in the MLST scheme. The equine isolates and human isolates cluster together when analysed for their bile salts, pH, salt and gastric juice tolerance. This behaviour is inconsistent and highlights the issue of relying on phenotypic diversity alone to differentiate between strains and species. Similar results were noted in *Lactobacillus delbrueckii* subspecies when grown in media supplemented with lactose (Tanigawa & Watanabe, 2011). N-acetyl-β-glucosaminidase production was not observed in the API-ZYM assay but the gene was present in each genome sequence. Despite half of strains in the API-ZYM test lacking β-glucosidase activity, all of the strains were able to ferment β-glucosides *in vitro*. Both leucine and cysteine arylamidase activity was identified in each isolate. However, examination of the available *L. ruminis* genome sequences (S23, ATCC 25644, ATCC 27782 and DPC 6832) failed to identify any enzymes consistent with either leucine arylamidase or cysteine arylamidase. False negative and positive results identified using the API-ZYM assay reflect the problem in using chromogenic assays only for assessing the presence of enzymes in bacteria.

Swarming is a type of flagella-mediated translocation in the presence of an extracellular slime matrix. This slime matrix has been identified in many Gram negative species and is often composed of bio-surfactants, carbohydrates and proteins (Daniels *et al.*, 2004). The increased number of hyper-flagellate, elongated cells noted

in this study may also be a factor in the *L. ruminis* strains ability to swarm on harder concentrations of agar (1-1.8%). To elucidate the genes transcribed during swarming and swimming in *L. ruminis* a combination of molecular and high throughput sequencing techniques were used. RNA-sequencing has previously been used to study the swimming motility in *L. ruminis* L5 in response to a medium supplemented with cellobiose (Lawley *et al.*, 2013). In the present study we focussed on the motility of cells grown in un-supplemented MRS media. We thus identified 14 genes in both motile *L. ruminis* strains ATCC 27782 and DPC 6832 which were differentially expressed between the two motility phenotypes. Unlike other studies where flagellar locus genes were significantly up-regulated when examining swimming motility (Attmannspacher *et al.*, 2008; Lawley *et al.*, 2013; Neville *et al.*, 2012), few flagellar locus-associated genes were significantly up-regulated here.

Swarming assays in the pathogen *Salmonella* have revealed that swarming cells have a different metabolism compared to swimming cells grown in the same nutrient medium (Kim & Surette, 2004). This difference is metabolism is reflected in the use of metabolic pathways in novel ways. Kim and Surette (2004) identified an up-regulation in expression of flagellin when comparing swimming and swarming *Salmonella Typhimurium* cells (Kim & Surette, 2004), a similar up-regulation in flagellin gene expression for both motile phenotypes was identified in this study. In our study, the expression of a number of carbohydrate metabolism and transport genes were significantly up-regulated. This suggests that carbohydrate metabolic components especially PTS transporters play a heretofore unrecognised role in swimming and more specifically swarming in *Lactobacillus ruminis*, perhaps for generating extracellular slime to promote swarming. Studies in other bacteria have noted a relationship between chemotaxis and the phosphotransferase transport system (Lux *et al.*, 1999). However, in these studies the swimming or swarming response was restricted to the PTS specific carbohydrate present in the test medium (Lux *et al.*, 1999). In our study, glucose was present in the medium for each condition, but there was an up-regulation in expression of genes related to fructose, FOS and sucrose metabolism. Further characterisation studies are needed to identify the role of the carbohydrate metabolism genes and transporters in the motile phenotypes of *L. ruminis*.

The expression of a number of uncharacterised hypothetical proteins was also identified as being up-regulated in the motile cells. It is possible that these

hypothetical proteins may be some form of novel glycolipid or lipo-peptide which may act as bio-surfactant facilitating swarm proliferation. However, until further characterisation work is carried out it is impossible to say what function these proteins have in the motile phenotypes of both strains. The data generated here on the differences between swimming and swarming cells suggests that swarming cells are a distinct cell type with novel pathways which need to be investigated further.

In conclusion, *L. ruminis* S23, DPC 6832 and DPC 6835 were identified as the best candidates for further testing and potential use in the future as probiotics based on their ability to survive gastric stresses, processing stresses and lack of antibiotic resistance genes. The MLST scheme designed and used in the study was sufficient to identify isolates and their original hosts. *In vitro* analysis of *L. ruminis* noted that agar concentration, carbohydrate type, carbohydrate concentration and hydration of the agar surface are important factors in swarming phenotype development. The transcriptional studies carried out identified carbohydrate metabolism as an important factor for swarming cells in both motile *L. ruminis* cells. This behaviour differs from that seen in swimming cells and suggests that swarming cells may have evolved novel metabolic pathways to facilitate agar surface translocation. However, further studies are needed to elucidate the function of these metabolic genes and pathways in motile *L. ruminis* cells.

6.5 References

Al Jassim, R. A. (2003). *Lactobacillus ruminis* is a predominant lactic acid producing bacterium in the caecum and rectum of the pig. *Lett Appl Microbiol* **37**, 213-217.

Alikhan, N.-F., Petty, N. K., Zakour, N. L. B. & Beatson, S. A. (2011). BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC genomics* **12**, 402.

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

Anders, S. (2010a). Analysing RNA-Seq data with the "DESeq" package. *Mol Biol*, 1-17.

Anders, S. (2010b).HTSeq: Analysing high-throughput sequencing data with Python. In *http://www-huberemblde/users/anders/HTSeq/doc/overviewhtml*.

Anders, S. & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol* **11**, R106.

Attmannspacher, U., Scharf, B. E. & Harshey, R. M. (2008). FliL is essential for swarming: motor rotation in absence of FliL fractures the flagellar rod in swarmer cells of *Salmonella enterica*. *Mol Microbiol* **68**, 328-341.

Başyiğit Kılıç, G., Kuleaşan, H., Sömer, V. F. & Akpınar, D. (2013). Determining potential probiotic properties of human originated *Lactobacillus plantarum* strains. *Biotechnol Bioprocess Eng* **18**, 479-485.

Bolger, A. & Giorgi, F. Trimmomatic: A Flexible Read Trimming Tool for Illumina NGS Data. In *http://wwwusadellaborg/cms/indexphp*.

Butler, M. T., Wang, Q. & Harshey, R. M. (2010). Cell density and mobility protect swarming bacteria against antibiotics. *Proc Natl Acad Sci USA* **107**, 3776-3781.

Cai, H., RodrÃguez, B. T., Zhang, W., Broadbent, J. R. & Steele, J. L. (2007). Genotypic and phenotypic characterization of *Lactobacillus casei* strains isolated from different ecological niches suggests frequent recombination and niche specificity. *Microbiology* **153**, 2655-2665.

Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M.-A., Barrell, B. G. & Parkhill, J. (2005). ACT: the Artemis comparison tool. *Bioinformatics* **21**, 3422-3423.

Champagne, C. P., Gardner, N. J. & Roy, D. (2005). Challenges in the addition of probiotic cultures to foods. *Crit Rev Food Sci Nutr* **45**, 61-84.

Chao, S.-H., Tomii, Y., Sasamoto, M., Fujimoto, J., Tsai, Y.-C. & Watanabe, K. (2008). *Lactobacillus capillatus* sp. nov., a motile bacterium isolated from stinky tofu brine. *Int J Sys Evol Microbiol* **58**, 2555-2559.

Daniels, R., Vanderleyden, J. & Michiels, J. (2004). Quorum sensing and swarming migration in bacteria. *FEMS Microbiol Rev* **28**, 261-289.

Danielsen, M. & Wind, A. (2003). Susceptibility of *Lactobacillus* spp. to antimicrobial agents. *Int J Food Microbiol* **82**, 1-11.

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.

De Man, J. C., M. Rogosa & Sharpe, M. E. (1960). A medium for the cultivation of lactobacilli. *Journal of Applied Microbiology* **23**, 130-135.

de Palencia, P. F., López, P., Corbí, A. L., Peláez, C. & Requena, T. (2008). Probiotic strains: survival under simulated gastrointestinal conditions, *in vitro* adhesion to Caco-2 cells and effect on cytokine secretion. *Eur Food Res Technol* **227**, 1475-1484.

de Vrese, M., Stegelmann, A., Richter, B., Fenselau, S., Laue, C. & Schrezenmeir, J. (2001). Probiotics - compensation for lactase insufficiency. *Am J Clin Nutr* **73**, 421s-429s.

Deibel, R. H. & Niven Jr, C. F. (1958). Microbiology of meat curing: I. The occurrence and significance of a motile microorganism of the genus *Lactobacillus* in ham curing brines. *Appl Microbiol* **6**, 323.

Delgado, S., O'Sullivan, E., Fitzgerald, G. & Mayo, B. (2007a). Subtractive screening for probiotic properties of *Lactobacillus* species from the human gastrointestinal tract in the search for new probiotics. *J Food Sci* **72**, M310-M315.

Delgado, S., Suarez, A. & Mayo, B. (2007b). Dominant cultivable *Lactobacillus* species from the feces of healthy adults in northern Spain. *Int Microbiol* **10**, 141-145.

Desai, A. R., Musil, K. M., Carr, A. P. & Hill, J. E. (2009). Characterization and quantification of feline fecal microbiota using cpn60 sequence-based methods and investigation of animal-to-animal variation in microbial population structure. *Veterinary Microbiology* **137**, 120-128.

Diancourt, L., Passet, V., Chervaux, C., Garault, P., Smokvina, T. & Brisse, S. (2007). Multilocus sequence typing of *Lactobacillus casei* reveals a clonal population structure with low levels of homologous recombination. *Appl Environ Microbiol* **73**, 6601-6611.

Dunne, C., Murphy, L., Flynn, S. & other authors (1999). Probiotics: from myth to reality. Demonstration of functionality in animal models of disease and in human clinical trials. In *Lactic Acid Bacteria: Genetics, Metabolism and Applications*, pp.

279-292. Edited by Wil Konings, Otto Kuipers, O.P. Kuipers & J. H. J. H. i. t. Veld: Springer.

EFSA (2012). Guidance on the assessment of bacterial susceptibility to antimicrobials of human and veterinary importance. In *EFSA Panel on Additives and Products or Substances used in Animal Feed (FEEDAP)* pp. 2740. Edited by EFSA: EFSA J.

Endo, A., Futagawa-Endo, Y. & Dicks, L. M. T. (2010). Diversity of *Lactobacillus* and *Bifidobacterium* in feces of herbivores, omnivores and carnivores. *Anaerobe* **16**, 590-596.

Forde, B. M., Neville, B. A., O' Donnell, M. M., Riboulet-Bisson, E., Claesson, M. J., Coghlan, A., Ross, R. P. & O' Toole, P. W. (2011). Genome sequences and comparative genomics of two *Lactobacillus ruminis* strains from the bovine and human intestinal tracts. *Microb Cell Fact* **10**, S13.

Greetham, H. L., Giffard, C., Hutson, R. A., Collins, M. D. & Gibson, G. R. (2002). Bacteriology of the Labrador dog gut: a cultural and genotypic approach. *Journal of Applied Microbiology* **93**, 640-646.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In *Nucleic acids symposium series*, pp. 95-98.

Hänninen, M. L. (1991). Sensitivity of *Helicobacter pylori* to different bile salts. *Euro J Clin Microbiol* **10**, 515-518.

Harrison Jr, A. P. & Hansen, P. A. (1950). A motile *Lactobacillus* from the cecal feces of turkeys. *J Bacteriol* **59**, 444.

Harshey, R. M. (2003). BACTERIAL MOTILITY ON A SURFACE: Many Ways to a Common Goal. *Annual Review of Microbiology* **57**, 249-273.

Hayashi, F., Smith, K. D., Ozinsky, A. & other authors (2001). The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. *Nature* **410**, 1099-1103.

Heilig, H. G., Zoetendal, E. G., Vaughan, E. E., Marteau, P., Akkermans, A. D. & de Vos, W. M. (2002). Molecular diversity of *Lactobacillus* spp. and other lactic acid bacteria in the human intestine as determined by specific amplification of 16S ribosomal DNA. *Appl Environ Microbiol* **68**, 114-123.

Huson, D. H. & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**, 254-267.

Jacobsen, L., Wilcks, A., Hammer, K., Huys, G., Gevers, D. & Andersen, S. R. (2007). Horizontal transfer of tet(M) and erm(B) resistance plasmids from food strains of *Lactobacillus plantarum* to *Enterococcus faecalis* JH2-2 in the gastrointestinal tract of gnotobiotic rats. *FEMS Microbiol Ecol* **59**, 158-166.

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

Karasu, N., Şimşek, Ö. & Çon, A. H. (2010). Technological and probiotic characteristics of *Lactobacillus plantarum* strains isolated from traditionally produced fermented vegetables. *Ann Microbiol* **60**, 227-234.

Kim, W. & Surette, M. G. (2004). Metabolic differentiation in actively swarming *Salmonella*. *Mol Microbiol* **54**, 702-714.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* **16**, 111-120.

Kovalenko, N. K., Golovach, T. N. & Kvasnikov, E. I. (1989). Lactic bacteria in the digestive tract of poultry. *Mikrobiologiia* **58**, 137-143.

Langmead, B. & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Met* **9**, 357-359.

Lawley, B., Sims, I. M. & Tannock, G. W. (2013). Whole-Transcriptome Shotgun Sequencing (RNA-seq) Screen Reveals Upregulation of Cellobiose and Motility Operons of *Lactobacillus ruminis* L5 during Growth on Tetrasaccharides Derived from Barley β-Glucan. *Appl Environ Microbiol* **79**, 5661-5669.

Lerche, M. & Reuter, G. (1961). Isolierung und Differenzierung anaerober *Lactobacilleae* aus Darm erwachsener Menschen (Beitrag zum *Lactobacillus* bifidus Problem). *Zentralbl Bakteriol* **182**, 324.

Lohse, M., Bolger, A. M., Nagel, A., Fernie, A. R., Lunn, J. E., Stitt, M. & Usadel, B. (2012). RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Res* **40**, W622-W627.

Lux, R., Munasinghe, V. R. N., Castellano, F., Lengeler, J. W., Corrie, J. E. T. & Khan, S. (1999). Elucidation of a PTS-carbohydrate chemotactic signal pathway in *Escherichia coli* using a time-resolved behavioral assay. *Mol Biol Cell* **10**, 1133-1146.

Makivuokko, H., Tiihonen, K., Tynkkynen, S., Paulin, L. & Rautonen, N. (2010). The effect of age and non-steroidal anti-inflammatory drugs on human intestinal microbiota composition. *Br J Nutr* **103**, 227-234.

Mathiesen, S. D., Orpin, C. G., Greenwood, Y. & Blix, A. S. (1987). Seasonal changes in the cecal microflora of the high-arctic Svalbard reindeer (*Rangifer tarandus platyrhynchus*). *Appl Environ Microbiol* **53**, 114-118.

Neville, B. A., Forde, B. M., Claesson, M. J., Darby, T., Coghlan, A., Nally, K., Ross, R. P. & O' Toole, P. W. (2012). Characterization of pro-inflammatory flagellin proteins produced by *Lactobacillus ruminis* and related motile lactobacilli. *PLOS One* **7**, e40592.

Nielsen, D. S., Schillinger, U., Franz, C. M. A. P., Bresciani, J., Amoa-Awua, W., Holzapfel, W. H. & Jakobsen, M. (2007). *Lactobacillus ghanensis* sp. nov., a motile lactic acid bacterium isolated from Ghanaian cocoa fermentations. *Int J Sys Evol Microbiol* **57**, 1468-1472.

Niu, C., Graves, J. D., Mokuolu, F. O., Gilbert, S. E. & Gilbert, E. S. (2005). Enhanced swarming of bacteria on agar plates containing the surfactant Tween 80. *Journal of microbiological methods* **62**, 129-132.

Noguchi, H., Park, J. & Takagi, T. (2006). MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Res* **34**, 5623-5630.

O' Donnell, M. M., O' Toole, P. W. & Ross, R. P. (2013). Catabolic flexibility of mammalian-associated lactobacilli. *Microb Cell Fact* **12**, 48.

O' Donnell, M. M., Forde, B. M., Neville, B. A., Ross, R. P. & O' Toole, P. W. (2011). Carbohydrate catabolic flexibility in the mammalian intestinal commensal *Lactobacillus ruminis* revealed by fermentation studies aligned to genome annotations. *Microb Cell Fact* **10**, S12.

Partridge, J. D. & Harshey, R. M. (2013). Swarming: flexible roaming plans. *J Bacteriol* **195**, 909-918.

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.

Raftis, E. J., Salvetti, E., Torriani, S., Felis, G. E. & O'Toole, P. W. (2011). Genomic Diversity of *Lactobacillus salivarius*. *Appl Environ Microbiol* **77**, 954-965.

Rather, P. N. (2005). Swarmer cell differentiation in Proteus mirabilis. *Environmental Microbiology* **7**, 1065-1073.

Reuter, G. (2001). The *Lactobacillus* and *Bifidobacterium* microflora of the human intestine: composition and succession. *Curr Issues Intest Microbiol* **2**, 43-53.

Ritchie, L. E., Burke, K. F., Garcia-Mazcorro, J. F., Steiner, J. M. & Suchodolski, J. S. (2009). Characterization of fecal microbiota in cats using universal 16S rRNA gene and group-specific primers for *Lactobacillus* and *Bifidobacterium* spp. *Vet Microbiol* **In Press, Corrected Proof**.

Roumagnac, P., Weill, F. o.-X., Dolecek, C. & other authors (2006). Evolutionary history of *Salmonella typhi*. *Science* **314**, 1301-1304.

Salyers, A. A., Gupta, A. & Wang, Y. (2004). Human intestinal bacteria as reservoirs for antibiotic resistance genes. *Trends Microbiol* **12**, 412-416.

Sharma, M. & Anand, S. K. (2002). Swarming: a coordinated bacterial activity. *Curr Sci* **83**, 707-715.

Sharpe, M. E., Latham, M. J., Garvie, E. I., Zirngibl, J. & Kandler, O. (1973). Two new species of *Lactobacillus* isolated from the bovine rumen, *Lactobacillus ruminis* sp.nov. and *Lactobacillus vitulinus* sp.nov. *J Gen Microbiol* **77**, 37-49.

Stewart, C. S., Fonty, G. & Gouet, P. (1988). The establishment of rumen microbial communities. *Animal Feed Science and Technology* **21**, 69-97.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**, 2731-2739.

Tanigawa, K. & Watanabe, K. (2011). Multilocus sequence typing reveals a novel subspeciation of *Lactobacillus delbrueckii*. *Microbiology* **157**, 727-738.

Tannock, G. W., Munro, K., Harmsen, H. J., Welling, G. W., Smart, J. & Gopal, P. K. (2000). Analysis of the fecal microflora of human subjects consuming a

probiotic product containing *Lactobacillus rhamnosus* DR20. *Appl Environ Microbiol* **66**, 2578-2588.

Taweechotipatr, M., Iyer, C., Spinler, J. K., Versalovic, J. & Tumwasorn, S. (2009). *Lactobacillus saerimneri* and *Lactobacillus ruminis*: novel human-derived probiotic strains with immunomodulatory activities. *FEMS Microbiology Letters* **293**, 65-72.

Verstraeten, N., Braeken, K., Debkumari, B., Fauvart, M., Fransaer, J., Vermant, J. & Michiels, J. (2008). Living on a surface: swarming and biofilm formation. *Trends in Microbiology* **16**, 496-506.

Vörös, A. (2008). Diet related changes in the gastrointestinal microbiota of horses. In *Dept of Animal Nutrition and Management*, pp. 36. Uppsala: Swedish University of Agricultural Sciences.

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 Microbiol Ecol* **59**, 127-137.

Wang, Y., Ahmed, Z., Feng, W., Li, C. & Song, S. (2008). Physicochemical properties of exopolysaccharide produced by *Lactobacillus kefiranofaciens* ZW3 isolated from Tibet kefir. *Int J Biol Macromol* **43**, 283-288.

Willing, B., Vörös, A., Roos, S., Jones, C., Jansson, A. & Lindberg, J. E. (2009). Changes in faecal bacteria associated with concentrate and forage-only diets fed to horses in training. *Equine Vet J* **41**, 908-914.

Xenoulis, P. G., Gray, P. L., Brightsmith, D., Palculict, B., Hoppes, S., Steiner, J. M., Tizard, I. & Suchodolski, J. S. Molecular characterization of the cloacal microbiota of wild and captive parrots. *Veterinary Microbiology* **In Press, Accepted Manuscript**.

Yu, C., Zavaljevski, N., Desai, V. & Reifman, J. (2011). QuartetS: a fast and accurate algorithm for large-scale orthology detection. *Nucleic Acids Res* **39**, e88.

Yun, J.-H., Yim, D.-s., Kang, J.-Y. & other authors (2005). Identification of *Lactobacillus ruminus* SPM0211 isolated from healthy Koreans and its antimicrobial activity against some pathogens. *Archives of pharmacal research* **28**, 660-666.

Yunis, A. A. (1989). Chloramphenicol toxicity: 25 years of research. *Am J Med* **87**, 44N-48N.

6.6 Supplementary Information

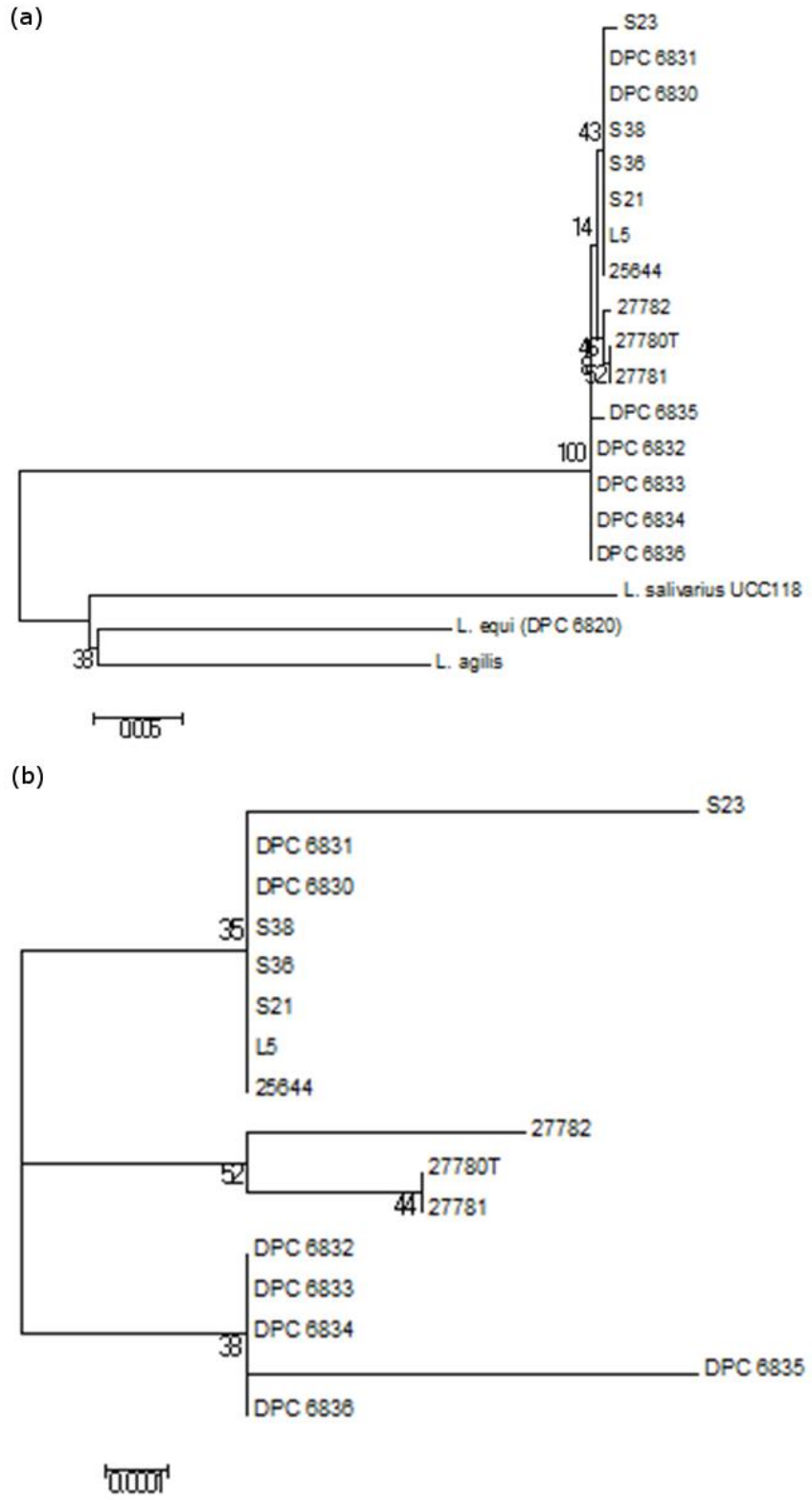


Figure S6.1. Neighbour-joining phylogenetic trees (a) Neighbour joining Phylogenetic tree for the *L. ruminis* cluster and other *Lactobacillus salivarius* clade species. (b) Neighbour joining phylogenetic tree for the *L. ruminis* isolates.

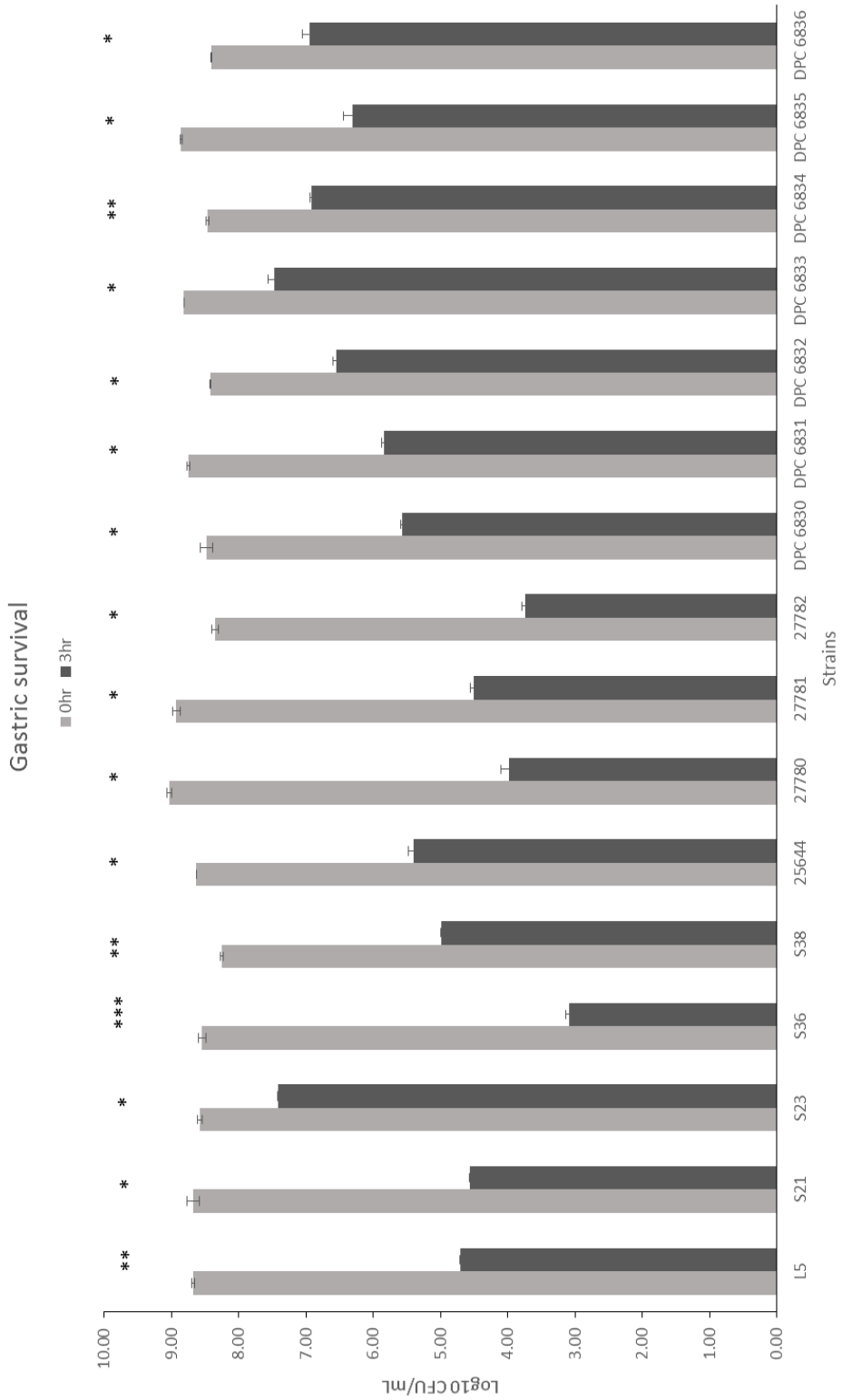


Figure S6.2. Gastric survival chart for all the sixteen *L. ruminis* isolates over a 3hr time period. Data plotted are plate counts after each incubation on MRS-glucose * P>0.05; ** P>0.01; *** P>0.001

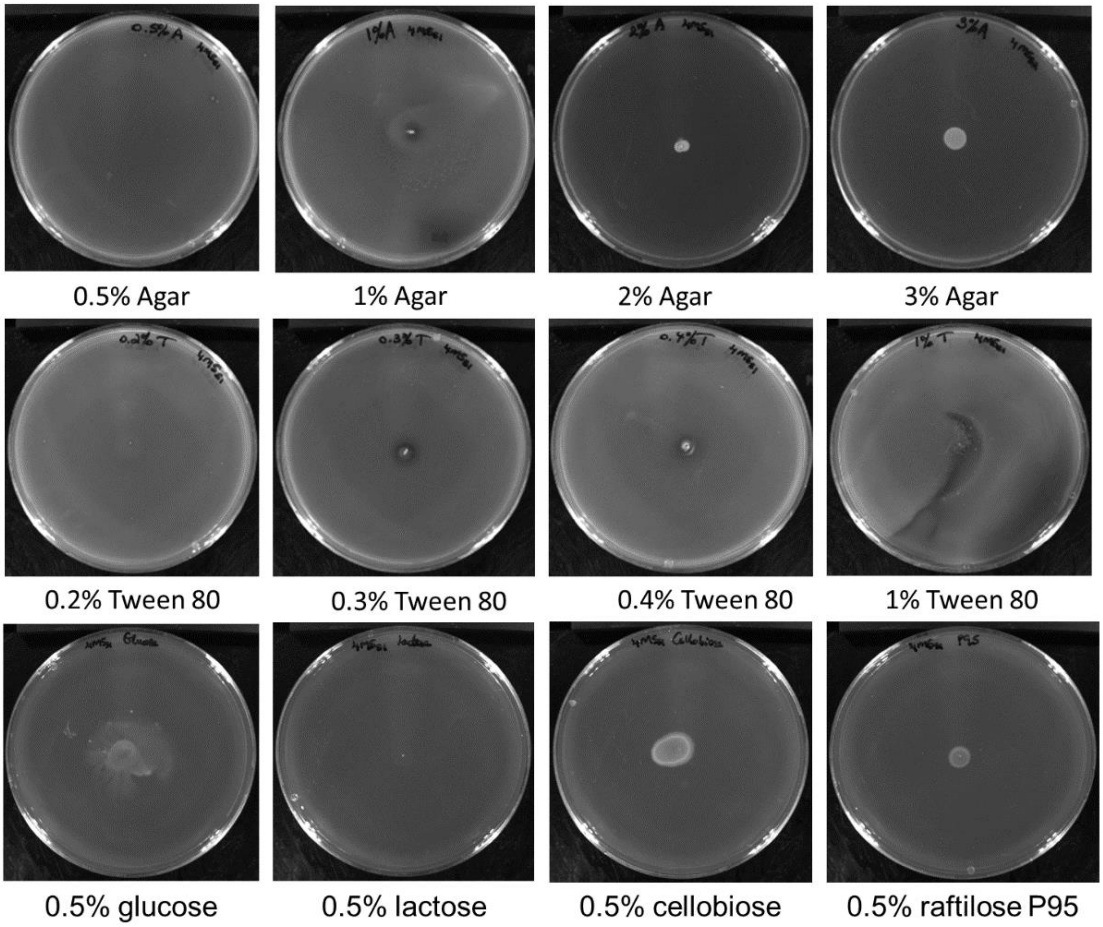


Figure S6.3. Screening of equine *L. ruminis* isolates for their swarming phenotype in the presence of varying percentages of agar, the bio-surfactant Tween 80 and minimal carbohydrates. Note - All of the images above are from the equine strain DPC 6833 but are typical of the results seen with the other equine and porcine strains tested.

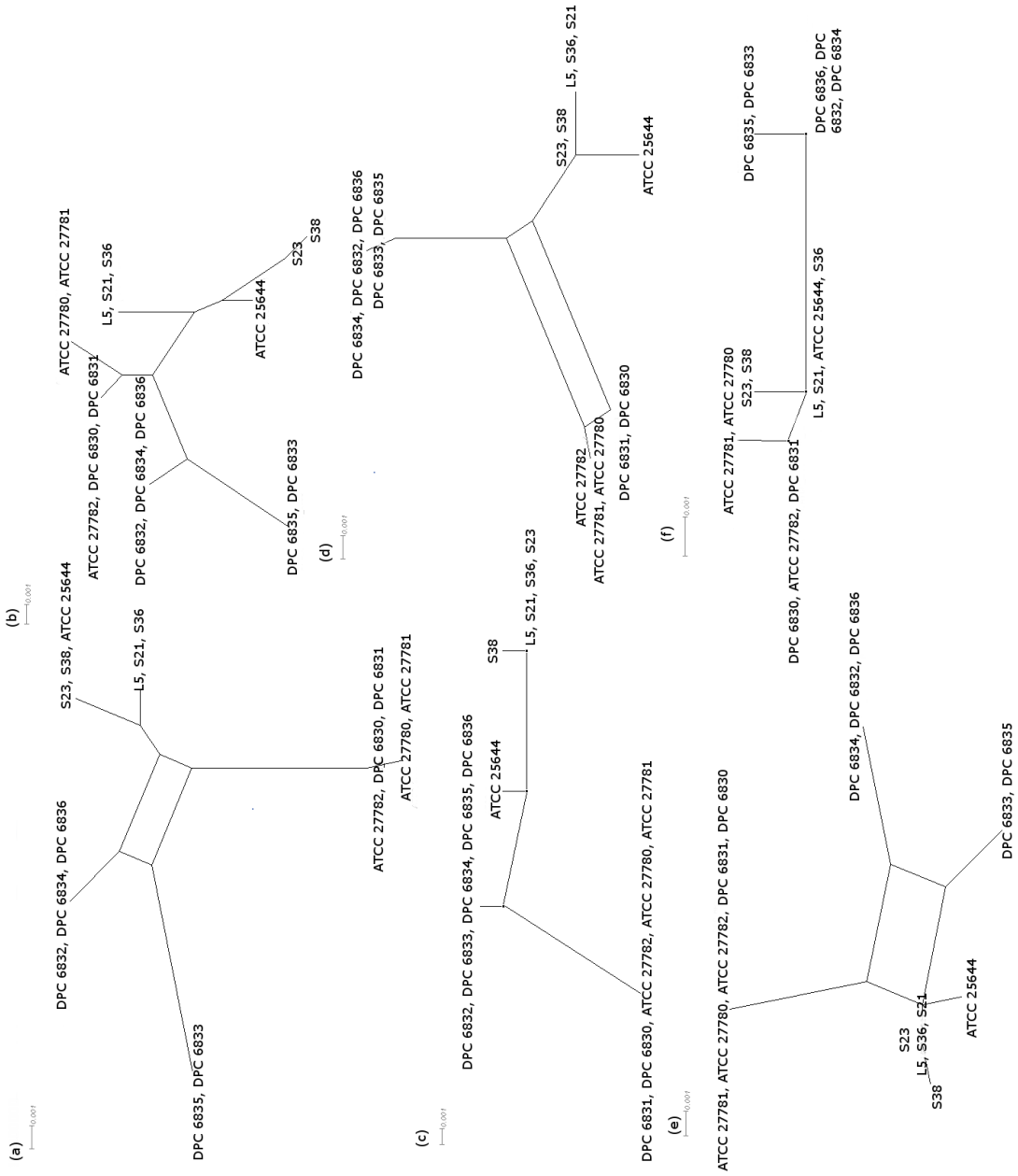


Figure S6.4. Splits decomposition trees generated from the housekeeping genes used in the MLST (a) *ftsQ*, (b) *nrdB*, (c) *parB*, (d) *pheS*, (e) *pstB*, (f) *rpoA*.

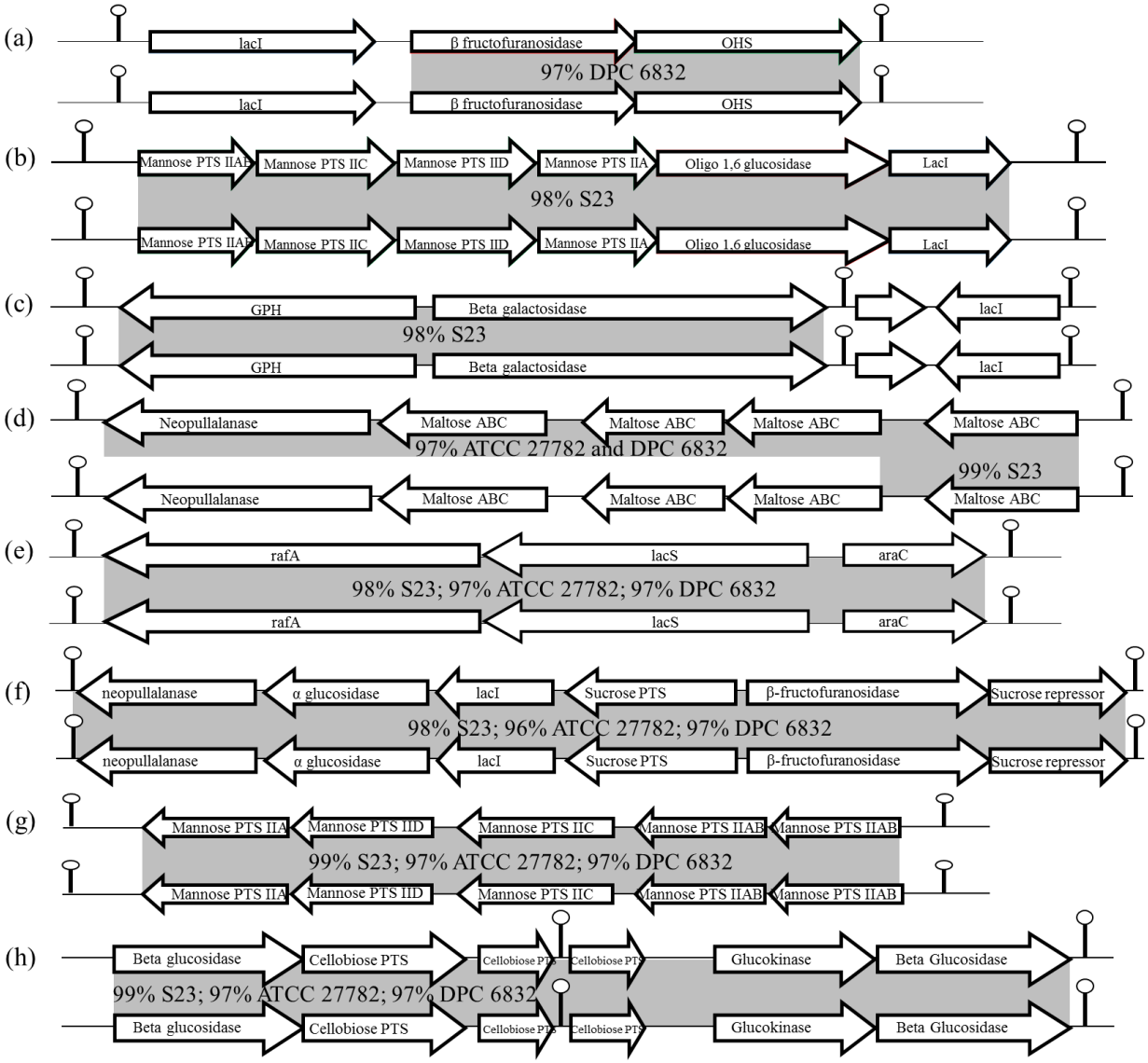


Figure S6.5. Prebiotic utlisation operon comparisons between *L. ruminis* ATCC 25644, ATCC 27782, S23 and DPC 6832. (a) FOS operon (b) Mannose PTS operon 1 (c) Lactose operon 1 (*lacZ*1) (d) Maltose ABC operon (e) Raffinose operon (f) Sucrose PTS operon (g) Mannose PTS operon 3 (h) Cellobiose PTS operon

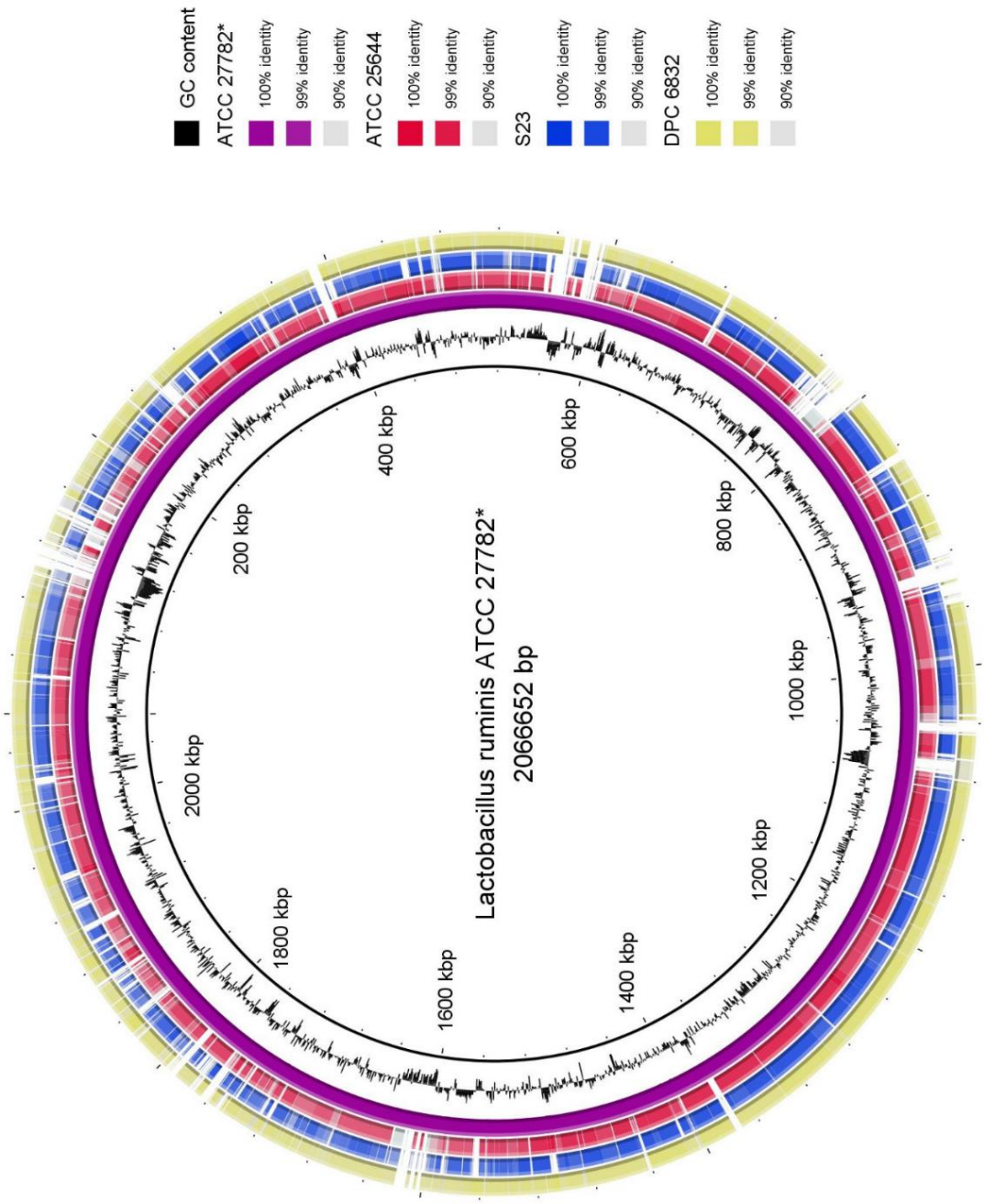


Figure S6.6. BRIG comparison between ATCC 27782, ATCC 25644, S23 and DPC 6832. The rings from the central nucleotide scale marker outwards, are GC% content, GC skew, ATCC 27882 (reference genome), ATCC 25644 genome, S23 genome and DPC 6832 genome with % identity to the reference genome colour coded as per the legend to the right.

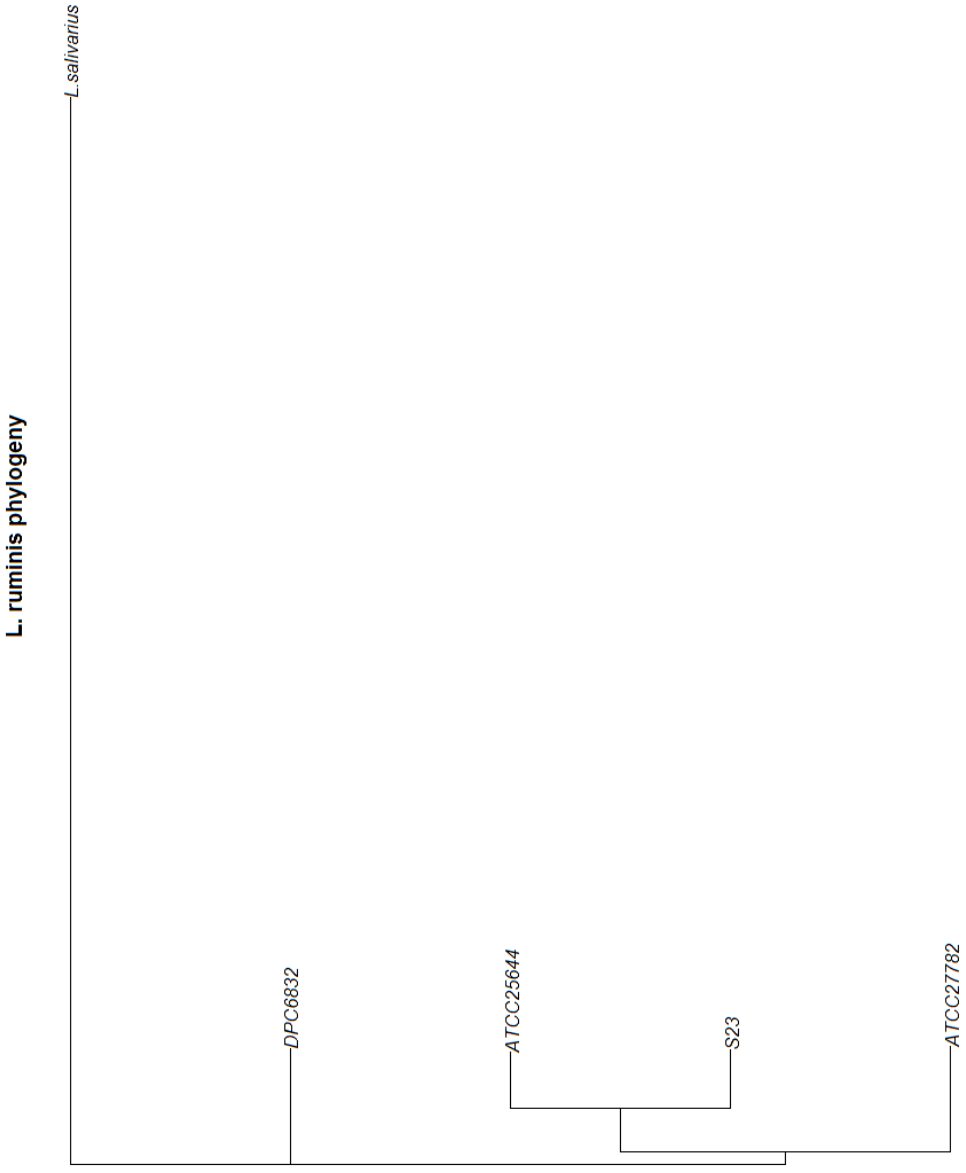


Figure S6.7. Whole genome phylogenetic tree for the four sequenced *L. ruminis* genomes and 33 *L. salivarius* species forming an out-group.

Table S6.1. Primers used in this study

Primers	Sequence 5'-3'	Region amplified	Size of amplicon (bp)	Source
16S-Lru_F	ACCATGAACACCGCATGATGTTC			
16S-Lru_R	TTCCATCTCTGGAATTGTCAGAAG	16S rRNA	849	This study
27F	AGAGTTTGATCMTGGCTCAG			
1492R	TACGGCACCTTGTTACGACTT	16S rRNA	1500	Peace *et al*., 19
ftsQ-F	GTGCAGCACGTTGGACGATATCATC			
ftsQ-R	TTTTAGGATATGCGTAAGCTCCGACT	ftsQ	745	This study
nrdB-F	AAGTTTTCGGAGGGCTGAC			
nrdB-R	CCGTTTCCGACCTGAGAGAA	nrdB	733	This study
parB-F	CGGACTTGACGCATTATTCACTGAA			
parB-R	GCTCTTGATTGAAACCTTCGTACTGA	parB	754	This study
pheS-F	GGACCTATTACTGAAGTGCTCCG			
pheS-R	TCCGGTCCAAGACCAAATGC	pheS	839	This study
pstB-F	GACGTTCATCTGTACTATGGCAAA			
pstB-R	TTTGTTGTCCGGCGTCACAA	pstB	696	This study
rpoA-F	CGCTTGAACGTGGCTATGGT			
rpoA-R	CCAAGATCTGCCAACTTAGCC	rpoA	846	This study
rpsB-F	TCGTCGTTGGAACCCAAAGA			
rpsB-R	AGTCTTCTTTACCTTCAACG	rpsB	728	This study
RT-PCR_1-F	AAGATCGGGAGTTTGTTGC	LRN_87/LRC_0064		
RT-PCR_1-R	CCGAAAAGCTCATCTGAATC	0	82	This study
RT-PCR_2-F	TCAAGCTTCAGGAAATCTGC	LRN_108/LRC_007		
RT-PCR_2-R	CCTGCTGAATATGTTTTGCC	80	219	This study
RT-PCR_3-F	GGCGAAAGTTTGATGAAGAC			
RT-PCR_3-R	GCGCATATGAACGATAGACC	LRN_109/pfkB	220	This study
RT-PCR_4-F	AGCCTGCACATCTCTTCTTC	LRN_110/LRC_008		
RT-PCR_4-R	GTTTTCAGCTTCCTTCCTTG	00	188	This study
RT-PCR_5-F	GTCATGTCAAGGTTTTGCG	LRN_324/LRC_032		
RT-PCR_5-R	TGCTCCGAGAATAAGATTGC	50	218	This study
RT-PCR_6-F	AGGGGAACGTACCGAAAAG	LRN_409/LRC_043		
RT-PCR_6-R	GCATGGTCCAAATCAATGTC	70	113	This study
RT-PCR_7-F	TTATCGTCTCGGCTACCATC			
RT-PCR_7-R	AATCATGTCCCTGCTTCTTG	LRN_520	163	This study
RT-PCR_8-F	GACGCTTGCCTATCTTTCC			
RT-PCR_8-R	CAGATCCGATCCAGAACAG	LRN_521	182	This study
RT-PCR_9-F	GATGACCTCAGCCAAAAGC	LRN_561/LRC_057		
RT-PCR_9-R	CGTACGTGTCCAAGAAAACC	80	130	This study
RT-PCR_10-F	CAGCAGCCAATTCAATACG	LRN_598/LRC_061		
RT-PCR_10-R	GCTGAGTTCGACATCCATC	70	103	This study
RT-PCR_11-F	TGATGACGAACGCTTGAAC			
RT-PCR_11-R	CTCTTCCCAATGCTGACTTG	LRN_933	110	This study
RT-PCR_12-F	ACGTCGCAGCTATGAACAC	LRN_1405&1777/fl		
RT-PCR_12-R	AACCACCGATTTGTGACTTC	iC	159	This study
RT-PCR_13-F	CAGGTTTGCGTATCAACAAG	LRN_1410/LRC_15		
RT-PCR_13-R	GAATGCTGTGAGTTTCGTTC	700	165	This study
RT-PCR_14-F	CGAACGGTCAATACCAAATC	LRN_1655/LRC_18		
RT-PCR_14-R	GATCGGAACGAAAACATCAG	780	185	This study
RT-PCR_15-F	GTGGCTTGTAATGCTATTCC			
RT-PCR_15-R	CTAACTGATTGTTTCGGCC	LRC_16260	96	This study
RecA-F	TTGGGAATCGTGTTCGTATC			
RecA-R	TTCACCGGTCTTGGAAATC	RecA	156	This study

Table S6.2. 16S rRNA sequencing results

Origin	Name	Species identification using blastn	% similarity to *L. ruminis* ATCC 27780^T	Fragment size (bp)	Reference
Human	L5	*L. ruminis*	99	1397	G.W. Tannock
	S21	*L. ruminis*	99	1355	G.W. Tannock
	S23	*L. ruminis*	99	1426	G.W. Tannock
	S36	*L. ruminis*	99	1408	G.W. Tannock
	S38	*L. ruminis*	99	1438	G.W. Tannock
	ATCC 25644	*L. ruminis*	99	1447	Lerche and Reuter, 1961
Bovine	ATCC 27780^T	*L. ruminis*	100	1444	Sharpe *et al*., 1973
	ATCC 27781	*L. ruminis*	100	1447	Sharpe *et al*., 1973
	ATCC 27782	*L. ruminis*	99	1394	Sharpe *et al*., 1973
Porcine	DPC 6830	*L. ruminis*	99	1385	This study
	DPC 6831	*L. ruminis*	99	1392	This study
	AR110	*Streptococcus alactolyticus*	87	1469	This study
	AR114	*Lactobacillus acidophilus*	88	1450	This study
	WR215	*Lactobacillus johnsonii*	89	1490	This study
	W308	*Lactobacillus amylovorus*	87	1469	This study
	W312	*Lactobacillus amylovorus*	88	1474	This study
Equine	DPC 6832	*L. ruminis*	99	1440	This study
	DPC 6836	*L. ruminis*	99	1427	This study
	DPC 6833	*L. ruminis*	99	1412	This study
	DPC 6834	*L. ruminis*	99	1398	This study
	DPC 6835	*L. ruminis*	99	1416	This study
	4R51	*Streptococcus equinus*	90	490	This study
	5R4S1	*Streptococcus equinus*	87	1198	This study
	5R6S1	*Streptococcus equinus*	88	1177	This study

^{T} – type strain

Table S6.3. Statistically significantly differentially expressed genes in swimming and swarming *Lactobacillus ruminis* ATCC 27782 cells

id	ATCC 27782 swimming vs stationary	pval	ATCC 27782 swarming vs stationary	pval	Fold change	GENBANK FUNCTION
LRC_18740	-5.45	*	0.15	>0.05	48.51	maltose/maltodextrin ABC transporter binding protein
LRC_18780	5.11	***	-0.48	>0.05	48.06	PTS system sucrose-specific transporter subunit IIABC
LRC_0703l	-1.54	>0.05	3.18	*	26.40	tRNA-Tyr
LRC_02110	3.41	*	-1.31	>0.05	26.29	DegV family protein
LRC_0638a	5.59	*	0.94	>0.05	25.19	tRNA-Arg
LRC_16560	4.62	**	0.30	>0.05	19.94	ribose-5-phosphate isomerase
LRC_07250	-3.74	*	0.33	>0.05	16.84	hypothetical protein
serS	-3.25	*	0.79	>0.05	16.48	seryl-tRNA synthetase
LRC_1388a	4.50	**	0.47	>0.05	16.36	tRNA-Glu
LRC_15450	3.78	*	-0.15	>0.05	15.25	glutaredoxin
LRC_05410	3.70	*	-0.18	>0.05	14.68	toxin/antitoxin system, Toxin component
LRC_19580	3.35	*	-0.49	>0.05	14.36	transcriptional regulator
LRC_18160	4.47	**	0.71	>0.05	13.53	isochorismatase family protein
lytR	4.73	**	1.00	>0.05	13.26	LytR family transcriptional regulator
LRC_03890	4.11	**	0.44	>0.05	12.71	NlpC/P60
LRC_0373f	4.73	**	1.07	>0.05	12.62	tRNA-Thr
LRC_05420	2.82	*	-0.83	>0.05	12.59	toxin/antitoxin system, Antitoxin component
LRC_255m	4.55	*	0.94	>0.05	12.22	tRNA-Pro
LRC_08030	-3.37	*	0.23	>0.05	12.19	hypothetical protein
LRC_0425a	4.01	**	0.46	>0.05	11.73	tRNA-Thr
LRC_04960	3.27	*	-0.06	>0.05	10.08	hypothetical protein
LRC_18260	3.24	*	-0.08	>0.05	10.00	AraC family transcriptional regulator
LRC_07100	-3.11	*	0.21	>0.05	9.98	hypothetical protein
LRC_18150	4.07	*	0.83	>0.05	9.47	hypothetical protein
LRC_07510	-2.88	*	0.28	>0.05	8.96	hypothetical protein
rpsL	3.94	**	0.79	>0.05	8.92	30S ribosomal protein S12
rplC	3.13	*	-0.02	>0.05	8.89	50S ribosomal protein L3
LRC_05780	4.04	**	0.89	>0.05	8.85	hypothetical protein
LRC_255n	4.39	**	1.28	>0.05	8.65	tRNA-Pro
LRC_11220	3.51	*	0.41	>0.05	8.57	hypothetical protein
LRC_17850	3.17	*	0.09	>0.05	8.46	glycosyltransferase
LRC_18790	3.56	*	0.53	>0.05	8.16	sucrose-6-phosphate hydrolase
rpsJ	3.22	*	0.29	>0.05	7.65	30S ribosomal protein S10
LRC_12520	3.56	*	0.63	>0.05	7.62	30S ribosomal protein S15
LRC_17160	3.25	*	0.37	>0.05	7.36	D-Ala-teichoic acid biosynthesis protein
LRC_0373e	5.27	*	2.52	>0.05	6.72	tRNA-Glu
LRC_17840	3.53	*	0.78	>0.05	6.71	hypothetical protein
LRC_02780	3.14	*	0.40	>0.05	6.65	50S ribosomal protein L4
LRC_00560	4.04	**	1.31	>0.05	6.64	Deoxyguanosine kinase
LRC_06960	3.54	*	0.82	>0.05	6.59	glycosyltransferase
rplK	3.52	*	0.81	>0.05	6.56	50S ribosomal protein L11
s6	3.38	*	0.68	>0.05	6.50	30S ribosomal protein S6
LRC_00520	3.56	*	0.86	>0.05	6.49	cytidine deaminase
rplV	2.86	*	0.16	>0.05	6.47	50S ribosomal protein L22
rpsG	3.30	*	0.67	>0.05	6.22	30S ribosomal protein S7
rpsS	2.81	*	0.19	>0.05	6.12	30S ribosomal protein S19
rplB	2.99	*	0.39	>0.05	6.08	50S ribosomal protein L2
rpl23p	3.19	*	0.59	>0.05	6.05	50S ribosomal protein L23
LRC_0419b	5.35	***	2.76	>0.05	6.05	tRNA-Ala
ccpA	2.84	*	0.29	>0.05	5.88	Catabolite control protein A
LRC_18800	3.05	*	0.50	>0.05	5.87	Sucrose operon repressor
LRC_16260	3.70	*	1.16	>0.05	5.84	hypothetical protein
LRC_06030	3.67	*	1.22	>0.05	5.49	CAAX family protease
infA	3.12	*	0.67	>0.05	5.45	translation initiation factor IF-1
LRC_14830	3.90	*	1.47	>0.05	5.39	hypothetical protein
rpoA	2.84	*	0.42	>0.05	5.36	DNA-directed RNA polymerase subunit alpha
ezrA	2.98	*	0.60	>0.05	5.20	septation ring formation regulator
LRC_19820	2.99	*	0.73	>0.05	4.78	ribonuclease P
LRC_18350	3.24	*	1.03	>0.05	4.62	multidrug/hemolysin transport system ATP-binding protein
LRC_01390	2.88	*	0.74	>0.05	4.40	UDP-galactopyranose mutase
rplA	3.14	*	1.05	>0.05	4.26	50S ribosomal protein L1
LRC_1447a	0.85	*	2.91	*	4.19	5S ribosomal RNA
LRC_17530	2.98	*	0.97	>0.05	4.03	transposase
LRC_16270	3.37	*	1.37	>0.05	3.99	Secreted LysM-domain containing protein
LRC_0703p	2.86	*	0.94	>0.05	3.77	tRNA-Cys

id	ATCC 27782 swimming vs stationary	pval	ATCC 27782 swarming vs stationary	pval	Fold change	GENBANK FUNCTION
LRC_12720	3.15	*	1.27	>0.05	3.68	hypothetical protein
LRC_0703k	3.91	*	2.05	>0.05	3.62	tRNA-Phe
LRC_06490	2.84	*	1.00	>0.05	3.58	hypothetical protein
trnA	3.67	*	1.86	>0.05	3.51	tRNA-Ala
LRC_01380	2.83	*	1.14	>0.05	3.22	oligosaccharide translocase
LRC_12530	3.53	*	1.93	>0.05	3.03	30S ribosomal protein S20
LRC_0703q	3.08	*	1.59	>0.05	2.80	tRNA-Leu
LRC_10980	2.82	*	1.37	>0.05	2.75	hypothetical protein
LRC_255k	3.63	*	2.79	*	1.79	tRNA-Leu

Table S6.4. Statistically significantly differentially expressed genes in swimming and swarming *Lactobacillus ruminis* DPC6832 cells

id	DPC 6832 Swimming vs Stationary[a]	pval	DPC 6832 Swarming vs Stationary[b]	pval	Fold change	GENBANK FUNCTION
LRN_108	-5.11	***	2.66	*	218	DeoR fructose transcriptional regulator
LRN_109	-3.94	**	3.62	**	189	1-phosphofructokinase
LRN_110	-2.93	*	4.29	**	149	PTS_system,_fructose_specific_IIABC_component
LRN_0409	-4.62	***	2.28	>0.05	120	hypothetical_protein_LRU_02075
LRN_0721	-3.50	**	1.71	>0.05	37	phosphoenolpyruvate_carboxykinase_(ATP)
LRN_0087	-3.36	**	1.81	>0.05	36	hypothetical_protein_HMPREF0542_11617
LRN_1651	-2.66	*	2.31	>0.05	31	maltose|maltodextrin_ABC_superfamily
LRN_1460	2.48	*	-2.47	*	31	LysM_domain_protein
LRN_1355	-2.06	>0.05	2.64	*	26	ArsR_family_transcriptional_regulator
LRN_329	-1.97	>0.05	2.68	*	25	myosin-cross-reactive_antigen
LRN_1539	-2.81	*	1.79	>0.05	24	aldose_1-epimerase
LRN_376	-3.54	**	0.99	>0.05	23	transposase_ISSoc7
LRN_521	0.06	>0.05	4.49	***	22	MFS Transporter Beta fructofuranosidase
LRN_1540	-3.00	*	1.35	>0.05	20	aldose_1-epimerase
LRN_1620	-2.86	*	1.38	>0.05	19	sugar_ABC_superfamily_ATP_binding_cassette
LRN_520	-0.74	>0.05	3.39	*	18	beta-fructofuranosidase
LRN_1692	-3.19	*	0.74	>0.05	15	endonuclease|exonuclease|phosphatase_family
LRN_1708	2.74	*	-1.15	>0.05	15	membrane_protein
LRN_0692	-2.58	*	1.25	>0.05	14	hypothetical_function_DUF299
LRN_1751	3.07	*	-0.56	>0.05	12	hypothetical_protein_LRC_02660
LRN_1784	2.81	*	-0.56	>0.05	10	transposase
LRN_0324	2.50	*	-0.86	>0.05	10	hypothetical_protein_HMPREF0542_10780
LRN_1236	2.71	*	-0.57	>0.05	10	inositol-phosphate phosphatase
LRN_1587	-2.43	*	0.69	>0.05	9	Peroxiredoxin (PRX) family
LRN_1452	2.60	*	-0.49	>0.05	9	flagellar_basal_body_rod_protein
LRN_1764	2.46	*	-0.62	>0.05	8	transposase
LRN_1338	-2.41	*	0.63	>0.05	8	ferritin,_Dps_family_protein
LRN_1527	3.39	*	0.35	>0.05	8	putative_peptide|deacylase
LRN_1800	3.05	*	0.19	>0.05	7	transposase
LRN_1337	-2.45	*	0.36	>0.05	7	cytochrome_b5
LRN_1756	-0.10	>0.05	2.70	*	7	hypothetical_protein_LGG_01889
LRN_1574	2.88	*	0.10	>0.05	7	hypothetical_protein_LRC_17840
LRN_1451	2.41	*	-0.33	>0.05	7	flagellar_basal-body_rod_protein_FlgC
LRN_0058	0.30	>0.05	2.83	*	6	pyruvate_formate-lyase_activating_enzyme
LRN_526	2.77	*	0.27	>0.05	6	MFS_transporter DBSA oxidoreductase
LRN_1265	2.34	*	-0.12	>0.05	6	(3R)-hydroxyacyl-[acyl_carrier_proteindehydratase
LRN_1746	3.13	*	0.71	>0.05	5	ferulic_acid_esterase
LRN_1788	3.98	**	1.59	>0.05	5	transposase
LRN_1218	2.67	*	0.35	>0.05	5	F0F1_ATP_synthase_subunit_A
LRN_1217	2.53	*	0.22	>0.05	5	ATP_synthase_F0_sector_subunit_C
LRN_0056	0.32	>0.05	2.64	*	5	pyruvate formate-lyase activating enzyme
LRN_0756	2.57	>0.05	0.31	>0.05	5	phosphatidylserine_decarboxylase_proenzyme_2
LRN_1424	2.72	*	0.49	>0.05	5	chemotaxis_protein_methyltransferase
LRN_1425	2.50	*	0.27	>0.05	5	chemotaxis_response_regulator_protein-glutamate_methylesterase
LRN_1700	3.40	**	1.17	>0.05	5	inosine guanoisine nucleoside hydrolase
LRN_1426	2.47	*	0.29	>0.05	5	CheW chemotaxis protein
LRN_0783	3.11	*	1.01	>0.05	4	hypothetical_protein_HMPREF0542_10419
LRN_0784	2.52	*	0.47	>0.05	4	arginyl-tRNA_synthetase
LRN_1331	2.37	*	0.33	>0.05	4	VanZ_family_protein
LRN_1590	2.80	*	0.77	>0.05	4	HIT_family_protein
LRN_1523	2.66	*	0.65	>0.05	4	D-alanine--poly(phosphoribitol)_ligase_subunit_1
LRN_1422	2.40	*	0.41	>0.05	4	chemotaxis_protein_CheC
LRN_1438	2.53	*	0.54	>0.05	4	flagellar_biosynthesis_protein_FliO
LRN_1658	2.52	*	0.53	>0.05	4	endonuclease|exonuclease|phosphatase_family
LRN_1576	2.77	*	0.82	>0.05	4	LysE_family_L-lysine_permease
LRN_1420	2.77	*	0.84	>0.05	4	chemotaxis_signal_transduction_protein_CheW
LRN_0741	2.56	*	0.64	>0.05	4	GMP_reductase
LRN_0933	5.15	***	3.24	*	4	hypothetical_protein_HMPREF0542_11529
LRN_1421	2.47	*	0.56	>0.05	4	chemotaxis_protein_CheY
LRN_1423	2.38	*	0.48	>0.05	4	histidine_kinase
LRN_0932	5.17	***	3.32	*	4	transposase,_ISSmi4
LRN_1215	2.40	*	0.55	>0.05	4	ATP_synthase_F1_sector_delta_subunit
LRN_1437	2.47	*	0.71	>0.05	3	flagellar_biosynthetic_protein_FliP

id	DPC 6832 Swimming vs Stationary[a]	pval	DPC 6832 Swarming vs Stationary[b]	pval	Fold change	GENBANK FUNCTION
LRN_1789	2.47	*	0.75	>0.05	3	MutR family transcriptional regulator
LRN_1020	2.82	*	1.11	>0.05	3	transposase
LRN_1522	2.79	*	1.09	>0.05	3	D-alanine-poly(phosphoribitol)_ligase_subunit_2
LRN_1659	3.28	**	1.68	>0.05	3	PTS_family_glucose_porter,_IICBA_
LRN_1416	2.39	*	0.83	>0.05	3	flagellar_motor_switch_protein
LRN_0277	2.32	*	0.81	>0.05	3	50S_ribosomal_protein_L23
LRN_1455	3.02	*	1.72	>0.05	2	methyl-accepting_chemotaxis_protein
LRN_0070	1.41	>0.05	2.70	*	2	membrane protein
LRN_1454	2.70	*	1.42	>0.05	2	flagellar_motor_protein_A
LRN_1410	1.81	>0.05	3.04	*	2	flagellin
LRN_0032	2.47	*	1.24	>0.05	2	30S_ribosomal_protein_S6
LRN_1777	1.89	>0.05	3.09	*	2	flagellin
LRN_1405	1.70	>0.05	2.89	*	2	flagellin
LRN_0466	3.22	*	2.10	>0.05	2	xanthine_phosphoribosyltransferase
LRN_1521	2.38	*	1.30	>0.05	2	D-alanine_transfer_protein_DltD
LRN_1762	3.28	*	2.23	>0.05	2	N-acetyltransferase
LRN_1401	1.47	>0.05	2.50	*	2	hypothetical_protein_HMPREF0542_12012
LRN_0467	1.96	>0.05	2.60	*	2	xanthine_permease
LRN_1384	2.45	*	2.04	>0.05	1	methyl_accepting_chemotaxis_protein
LRN_0904	3.22	*	3.13	*	1	hypothetical_protein_ANHS_1530

Table S6.5. Allele frequencies for all of the sixteen *L. ruminis* isolates

Allele	***ftsQ***	***nrdB***	***parB***	***pheS***	***pstB***	***rpoA***
1	3	2	3	1	5	3
2	3	3	4	3	3	4
3	3	1	1	2	1	2
4	2	1	1	1	1	2
5	3	1	2	2	1	3
6	2	2	5	2	3	2
7	-	3	-	3	2	-
8	-	2	-	2	-	-
9	-	-	-	-	-	-
Unique	6	8	6	8	7	6

Table S6.6. Sequence characteristics of the internal gene fragments used for multilocus sequence typing analysis

Gene	Fragment analysed (nt)	Mean % GC of fragment	% GC of complete gene	Number of		Nucleotide diversity per site	SSCF (p value)	MCF (p value)
				Alleles	Polymorphic sites			
ftsQ	658	41.99	40.71	6	22	0.01008	263 (0.343)	9 (1.000)
nrdB	660	44.38	44.65	8	23	0.01096	1425 (0.106)	16 (0.119)
parB	673	46.23	45.27	6	22	0.00942	462 (0.658)	13 (1.000)
pheS	748	46.73	45.18	8	19	0.00806	1615 (0.160)	16 (1.000)
pstB	616	48.86	47.09	7	18	0.00930	622 (0.595)	12 (1.000)
rpoA	765	42.03	41.69	6	9	0.00356	241 (0.126)	7 (1.000)

Chapter VII

The genome of the predominant equine lactobacillus species, *Lactobacillus equi* is reflective of its lifestyle adaptations to a herbivorous host

Michelle M. O' Donnell[1,2], Hugh M.B. Harris[2], Paul W. O' Toole[2] and R. Paul Ross[1*]
Teagasc Food Research Centre, Moorepark, Fermoy, Co. Cork, Ireland[1]
Department of Microbiology, University College Cork, Co. Cork, Ireland[2]

A shortened version of this chapter was published in *Genome announcements*:

O'Donnell M. M, Harris H. M. B, O'Toole P. W., Ross R. P. Genome of the predominant equine *Lactobacillus* species, *Lactobacillus equi* is reflective of its lifestyle adaptations to a herbivorous host. Genome Announc. 2014 2(1):e01155-13. doi:10.1128/genomeA.01155-13.

Note:
Sample collection, DNA isolation and purification, manual curation of the draft assembled genome, bioinformatic analysis including TMHMM 2.0, Arnold, CW-PRED and KAAS, identification of enzymes and genes of interest was carried out by M.M O' Donnell (author of this thesis)

Contig assembly, ORF idenfication using Metagene, genome synteny plots and calculations and generation of ACT comparison files was carried out by H. M. B Harris

Chapter VII

Table of contents

Abstract

We report here the draft genome sequence of *Lactobacillus equi* strain DPC 6820 isolated from equine faeces. We determined the 2.19 Mb genome sequence of this racehorse-derived strain of *Lactobacillus equi*, and identified 2263 potential coding sequences in a 39% G+C chromosome. We identified a relatively large repertoire of proteins associated with carbon catabolism including those involved in fructan degradation. The predicted ability to transport and metabolize nutrients from the GIT likely contributes to the competitiveness and colonization capability of *L. equi*. The *L. equi* genome sequence will improve understanding of the microbial ecology of the equine hindgut and the influence lactobacilli have therein. This identification of host interaction characteristics and carbohydrates utilised by *L. equi* may inform on rational nutritional and/or probiotic approaches to promote this species and enhance its performance *in vivo* in the future.

7.1 Introduction

Lactobacilli are Gram-positive bacteria that are naturally present in the gastrointestinal tract (GIT) of humans, other mammals and fowl. *Lactobacillus equi* is a lactic acid bacterium found particularly in the gastrointestinal tracts of horses (Morotomi *et al.*, 2002) and which, along with *Lactobacillus hayakitensis* and *Lactobacillus equigenerosi*, has been identified as the predominant *Lactobacillus* of the equine hindgut (Morita *et al.*, 2009a).

7.2 Materials and methods

7.2.1 Strain isolation

Lactobacillus equi DPC 6820 was isolated from a faecal sample from a healthy Irish Thoroughbred racehorse fed a diet of haylage supplemented with starch concentrate. The faecal sample was diluted in maximum recovery diluent and a dilution series was plated on cf-MRS supplemented with raffinose as previously described (O' Donnell *et al.*, 2011). The isolates were then identified to species level by sequencing 16S rRNA gene amplicons generated by the 27F and 1492R bacterial primers (Lane, 1991).

7.2.2 Genome sequencing

The sequence data was obtained using the Illumina HiSeq 2000 reversible dye terminator system (Macrogen, Seoul, Korea) with average read lengths of 101 bp. The functional assignment of predicted genes was performed using Metagene (Noguchi *et al.*, 2006) to predict open reading frames (ORFs) and BLASTP to annotate them using the NCBI database (Altschul *et al.*, 1990). The tRNA genes in the *L. equi* genome were predicted using tRNA scan (Schattner *et al.*, 2005).

7.2.3 Genome synteny plots and orthologue comparison

Using the PROmer (Delcher *et al.*, 2002) feature of MUMmer (Kurtz *et al.*, 2004) the draft genome of *L. equi* was compared to the published genomes of *Lactobacillus ruminis* ATCC 25644, *Lactobacillus ruminis* ATCC 27782, *Lactobacillus salivarius* UCC118, *Lactobacillus mali* KCTC 3596 = DSM 20444, *Lactobacillus plantatrum*

WCFS1, *Lactobacillus acidophilus* NCFM, *Lactobacillus rhamnosus* GG, *Streptococcus mutans* UA159 and *Streptococcus equi* subsp. *equi* 4047. *L. equi* along with *L. ruminis*, *L. salivarius* and *L. mali* are all members of the *L. salivarius* clade (Neville & O'Toole, 2010). For the comparison plots, the scaffolds from the genome assembly of *L. equi* (y-axis) were ordered and orientated relative to the comparative genome (x-axis) using Mauve contig mover (Rissman *et al.*, 2009). For easier visual comparison, isolated contigs were joined together, and for complete genomes containing plasmids, the plasmids were placed to the end of the chromosome. We also sought to computationally/mathematically assess the proportion of genes shared (orthologues) by *L. equi* in a pairwise manner with the other genomes mentioned earlier. However, due to variation in the number of predicted genes within the 9 genomes the data required normalization. This procedure is summarized by the following equation: $P = (\bar{g}/g_i).(O_i/g_{L.equi})$

where P is the normalized proportion; $\bar{g}$ is the median gene count of the nine strains (excluding *L. equi*); g_i is the individual gene count for genome i; O_i is the number of shared orthologs between genome i and *L. equi* and $g_{L.equi}$ is the gene count for *L. equi*.

Un-weighted Pair Group Method with Arithmetic Mean (UPGMA) phylogenetic trees were also generated for the 9 species outlined above and *L. equi* using the 16S ribosomal RNA nucleotide sequences for each species. Sequences were aligned using the ClustalW algorithm of Bioedit (Hall, 1999) and sequences were trimmed to remove any discrepancies caused by additional sequence in one species versus another. SplitsTree4 was then used to generate the UPGMA 16S rRNA phylogenetic trees. A Principle Coordinate Analysis plot was also generated using the 16S rRNA aligned genes in R (Team, 2008). Further whole genome comparisons were made between the *L. salivarius* clade species using the Artemis Comparison Tool (ACT) (Carver *et al.*, 2005).

7.2.4 Genome sequence analysis

Structural information was elucidated using various prediction programs including tRNAscan-SE to identify tRNAs (Schattner *et al.*, 2005). Signal peptide cleavage sites were predicted using SignalP 4.1 (Petersen *et al.*, 2011), transmembrane domains using TMHMM 2.0 (Krogh *et al.*, 2001), and terminator-like structures

using Arnold (Naville *et al.*, 2011). Potential LPXTG-like motifs (targets for cleavage and covalent coupling to peptidoglycan by sortase enzymes) were predicted using CW-PRED (Fimereli *et al.*, 2012). The Transmembrane classification database was used to categorize potential transport proteins (Saier *et al.*, 2006). Carbohydrate utilisation pathways were also predicted using the KEGG Automatic annotation server (KAAS) (Moriya *et al.*, 2007) and the assignments manually curated into the genome following BLASTP comparisons.

7.2.5 Nucleotide sequence accession numbers.

This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession AWWH00000000. The version described in this paper is version AWWH01000000.

7.3 Results

7.3.1 General genome features

The HiSeq system paired-end sequencing strategy generated 36,133,338 reads (3,649,467,138 bp). 254 scaffolds containing 273 contigs were assembled, corresponding to 34,664,201 reads from the HiSeq system (3,501,084,301bp), which represents 1,608-fold genome coverage based on an estimated genome size of 2.19 Mb. The N50 score for the assembly estimating contig length was 39,802 bp. The draft *L. equi* genome includes 2,187,681 bases (G+C content of 39.16%). It comprises 2,263 predicted genes or coding sequences (CDS). Eight rRNA operons and 68 predicted tRNAs, representing all 20 amino acids were identified in the genome. Functions could be predicted for 76% of the *L. equi* chromosomal genes. The remaining genes were either homologous to conserved hypothetical proteins in other species or had no match to any known protein. No plasmids were detected in the draft assembly. Three hundred and sixty-nine predicted CDS were annotated as hypothetical proteins. The genome harboured 87 predicted transposable elements. Also identified within the genome were 55 proteins predicted as being bacteriophage or prophage-related proteins and these were identified in 8 clusters. A potential origin of replication (*ori*C) was identified with the gene *DnaA* (LEQ0793) flanked by "perfect" DnaA boxes (Mackiewicz *et al.*, 2004). Genes encoding CRISPR-related

proteins were also predicted in the *L. equi* genome (LEQ0046-0047, LEQ1986). To date no bacteriocin-related genes have been identified in the draft genome of *L. equi*. The genome also encodes a predicted subtilase (LEQ1490). This enzyme contains the peptidase S8 family domain and can be classed as a Streptococcal C5a peptidase (SCP). The SCP have been shown to be important Streptococcal virulence factors (Chen & Cleary, 1990; Cleary *et al.*, 1992). This protein also contains a sortase LPxTG anchor indicating its function as an extracellular peptidase. However, the function of this peptidase in *L. equi* is currently unknown.

7.3.2 Genome synteny

Comparison of synteny using MUMmerplots revealed that the *L. equi* genome is closely related to that of other *L. salivarius* clade members [Figure 7.1 (a-d)]. It is most closely related to *L. ruminis* ATCC 27782 [Figure 7.1 (b)]; very few translocation or insertion events are present. The genomes of *L. plantarum* WCFS1, *L. acidophilus* NCFM and *L. rhamnosus* GG [Figure 7.1 (e-g) are less similar to that of *L. equi* and show more translocations, insertions and some minor inversions. The genomes of both streptococcal species analysed are even less like that of *L. equi* than the other *Lactobacillus* species. This expected phylogenetic pattern (Canchaya *et al.*, 2006) is also obvious in the 16S rRNA phylogenetic tree (Figure 7.2) in which there is a clustering of the *L. salivarius* clade species, a clustering of the Streptococci, and loose grouping of the remaining lactobacilli. Examination of the orthologous genes present in the nine genomes corroborated this similarity profile in which *L. ruminis* ATCC 27782 shares a higher proportion of genes with *L. equi* than any other bacteria analysed (Table 7.1), perhaps reflective of adaptation of the herbivore gut. Furthermore, of particular note was the proportion of genes shared between *S. mutans* UA159 and *L. equi* which is even higher than the proportion of genes shared between *L. equi* and other non-*L. salivarius* clade lactobacillus species like *L. plantarum* and *L. rhamnosus*. The ACT comparison of the *L. salivarius* clade members (Fig 7.3 [a-d]) further supports the trend of synteny with *L. ruminis* ATCC 27782 (Fig 7.3 [a]) being the most similar and *L. mali* (Fig 7.3 [d]) least.

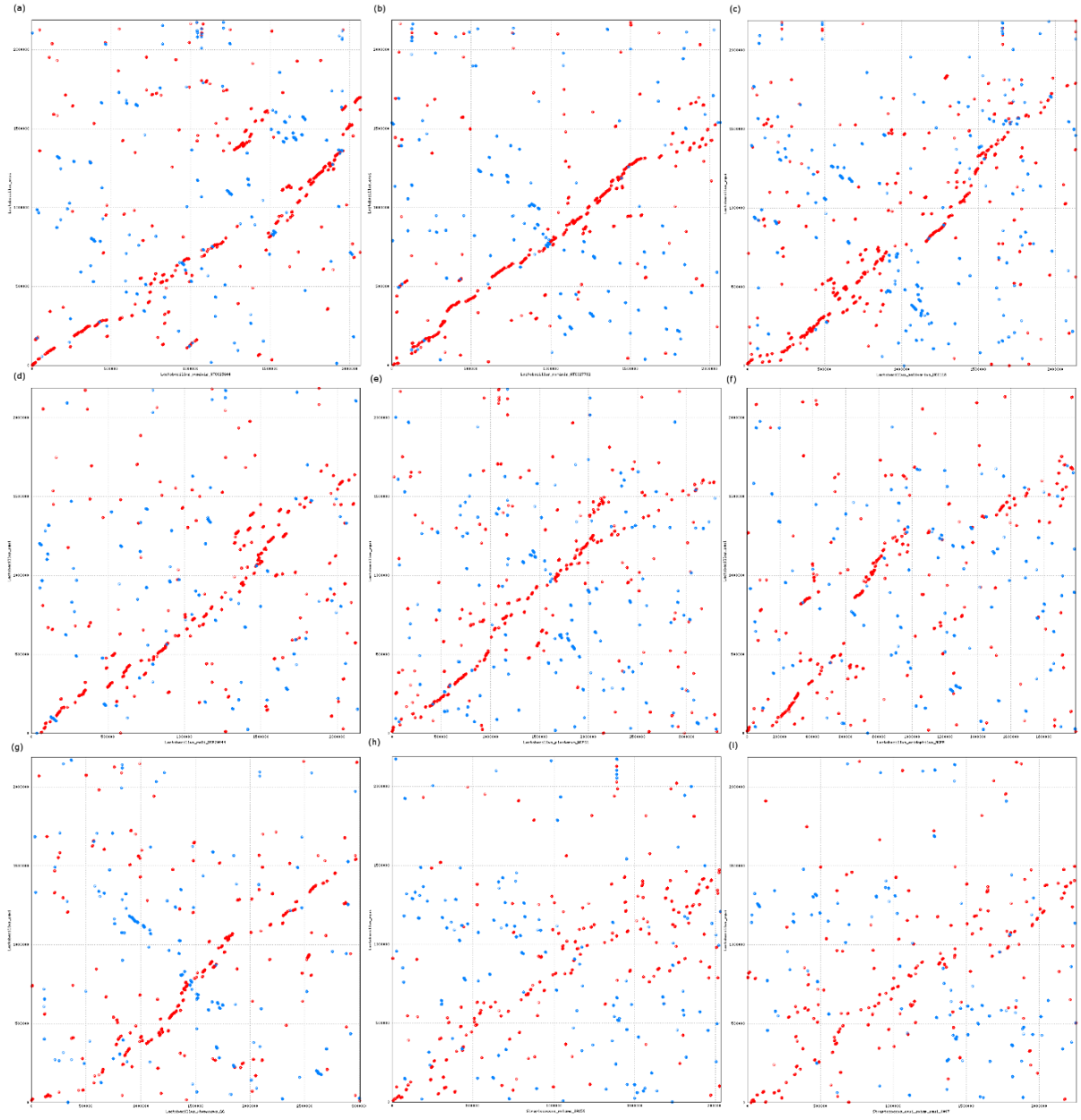


Fig. 7.1 Genome synteny of lactobacillus genomes analyzed. (a) *L. ruminis* ATCC 25644; (b) *L. ruminis* ATCC 27782; (c) *L. salivarius* UCC118; (d) *L. mali* KCTC 3596 (DSM 20444); (e) *L. plantatrum* WCFS1; (f) *L. acidophilus* NCFM; (g) *L. rhamnosus* GG; (h) *S. mutans* UA159 and (i) *S. equi* subsp. *equi* 4047

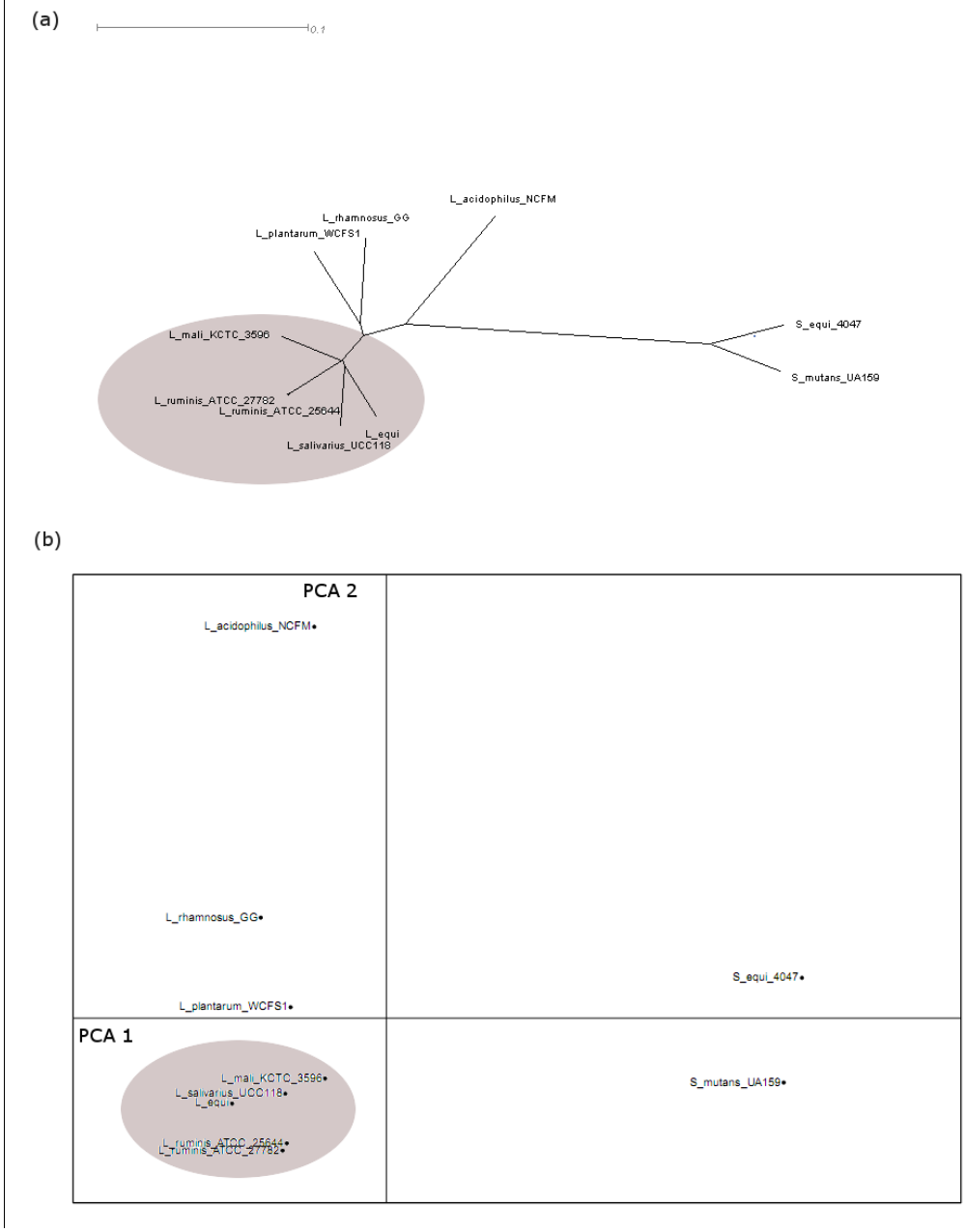


Fig. 7.2 Phylogenetic comparison of *L. equi* with nine other LAB bacteria. (a) UPGMA 16S rRNA phylogenetic trees. (b) PCoA plots. Grey oval; *L. salivarius* clade members.

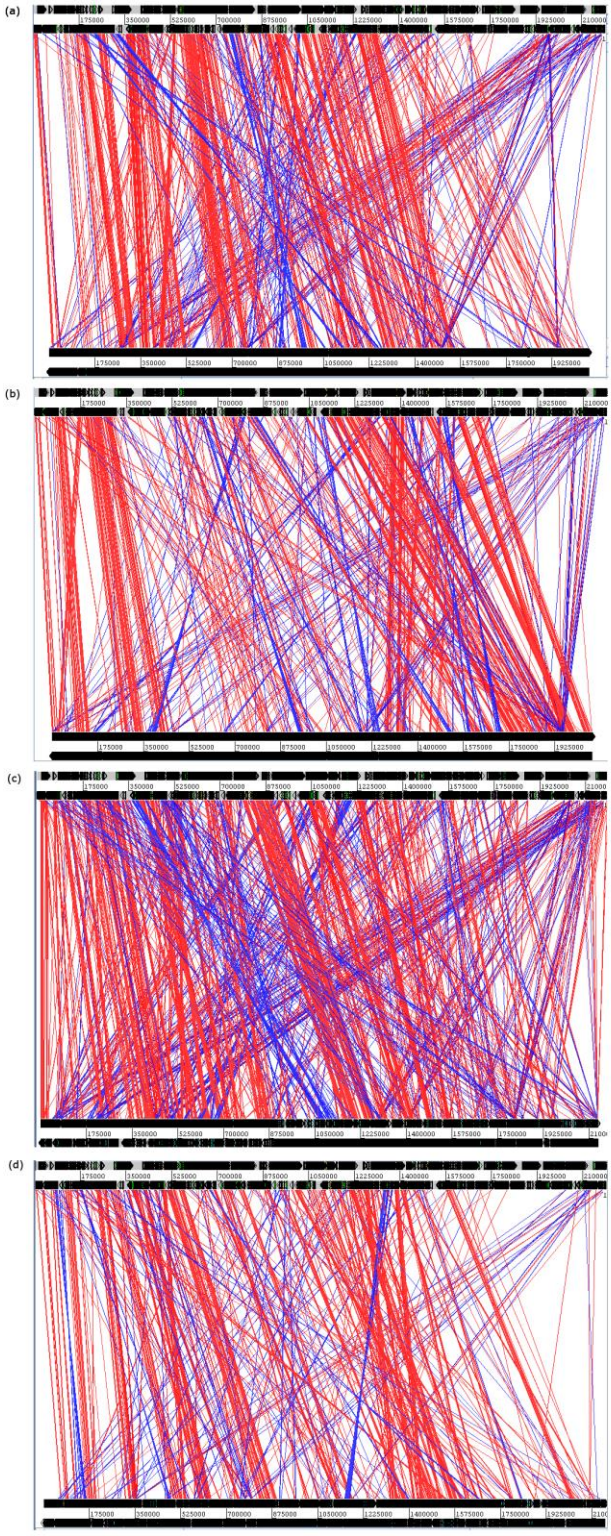


Figure 7.3 *Lactobacillus salivarius* clade species whole genome ACT comparisons. (a) *L. equi* DPC 6820 vs. *L. ruminis* ATCC 27782 (b) *L. equi* DPC 6820 vs. *L. ruminis* ATCC 25644 (c) *L. equi* DPC 6820 vs. *L. salivarius* UCC 118 (d) *L. equi* DPC 6820 vs. *L. mali* KCTC 3596 (DSM 20444).

Table 7.1 Orthologue comparison of *Lactobacillus equi* and nine other LAB bacteria

Genome	Proportion[a]	Gene No.[b]	*L. equi* gene No.[c]	Reference
L. ruminis ATCC 27782	0.5083	1835	2263	(Forde *et al.*, 2011)
L. ruminis ATCC 25644	0.505	1890	2263	(Forde *et al.*, 2011)
L. salivarius UCC118	0.505	2092	2263	(Claesson *et al.*, 2006)
L. acidophilus NCFM	0.4263	1882	2263	(Altermann *et al.*, 2005)
L. mali DSM 20444	0.3974	2177	2263	(Neville *et al.*, 2012)
S. mutans UA159	0.3792	1889	2263	(Ajdić *et al.*, 2002)
L. rhamnosus GG	0.3504	2680	2263	(Kankainen *et al.*, 2009)
L. plantarum WCFS1	0.3483	3015	2263	(Kleerebezem *et al.*, 2003)
S. equi 4047	0.3164	2121	2263	(Holden *et al.*, 2009)

[a] Normalized proportions of shared genes between each pair of genomes.

[b] Total number of predicted genes present in the genomes of each LAB bacteria listed

[c] Total number of predicted genes present in the genome of *Lactobacillus equi*.

7.3.3 Carbohydrate metabolism and transport for a herbivore-associated niche.

The genome of *L. equi* contains genes consistent with those normally seen in the commensal microbiota of herbivores. The carbohydrate utilisation pathways are presented in Supplemental Figures 7.1-7.13. We identified 12 operons likely involved in the utilisation of carbohydrates, five of which have the potential ability to degrade complex carbohydrates (also referred to as prebiotics in human diet). Genes for a number of glycosidases (EC 3.2.1) are present in the *L. equi* genome including those for α-glucosidase, β-glucosidase, α-galactosidase, β-galactosidase, fructan hydrolases and β-fructofuranosidase (Henrissat, 1991; Henrissat & Bairoch, 1996). The *L. equi* genome encodes the enzyme tagatose 1,6 diphosphate aldolase (EC 4.1.2.40) (LEQ1436). This enzyme is involved in the reversible conversion of D-tagatose 1,6-bisphosphate to glycerone phosphate and D-Glyceraldehyde 3-phosphate. Both of these products are important in multiple intracellular pathways including glycolysis, fructose, galactose and pentose sugar metabolism. Tagatose is a hexose monosaccharide found in dairy products and fruits and this enzyme is present in other lactobacilli mainly those involved in dairy fermentations (*L. casei*, *L. buchneri*, *L.*

bulgaricus). However, the genomes of some mammalian-associated lactobacilli like *L. gasseri*, *L. johnsonii* and *L. rhamnosus* also encode this enzyme. However, it must be noted that no member of the *L. salivarius* clade encodes this enzyme and that this activity may confer a competitive advantage over other closely related GIT lactobacilli. The *L. equi* genome encodes two predicted fructan hydrolases/levanases (LEQ1367 and LEQ1643) (EC 3.2.1.65). Both of these predicted enzymes contain single transmembrane helices and an LPXTG motif (sortase associated Gram positive cell wall anchor) which suggests that these enzymes are anchored in the cell wall of *L. equi*. This extracellular enzyme, unlike other fructan hydrolases/beta-fructofuranosidases, does not require an additional transporter or transport cascade to facilitate fructan transport/utilisation. The levanases are responsible for the hydrolysis of 2,6-β-D-fructofuranosides like levan. Both enzymes are classified as glycosyl hydrolase family 32 (Henrissat, 1991). The most similar database sequence matches correspond to members of the *Streptococcus* genus which suggests that this gene may have been acquired via horizontal transfer from a *Streptococcus in vivo*. The presence of these fructan hydrolases as well as a predicted fructan hydrolase pseudogene (LEQ830) and two beta-fructofuranosidases (LEQ0945, 1428) suggests that fructans, inulins and fructooligosaccharides form part of the primary carbon source of *L. equi* DPC 6820. Fructans are important storage carbohydrates found in plants and grasses. Horses eat a high polysaccharide-based diet and *L. equi* may have acquired fructan utilisation genes via horizontal transfer as a method of competing for nutrients in the equine gut.

The *L. equi* genome encodes an incomplete citrate acid cycle. Like *L. ruminis* and *Lactobacillus animalis* (other members of the *L. salivarius* clade (Felis & Dellaglio, 2007; O' Donnell *et al.*, 2011), *L. equi* lacks a transaldolase and transketolase, key enzymes of the pentose phosphate pathway. This is in agreement with data from Morita and colleagues who identified the species as homofermentative (Morita *et al.*, 2009a). *L. equi* also encodes a complete pyruvate metabolism pathway which allows for the recycling of pyruvate. The genome also encodes genes for the predicted conversion of pyruvate metabolic products into (R)-2-Acetoin and (R,R)-Butane-2,3-diol. A gene was annotated that encodes the enzyme diacetyl reductase (EC 1.1.1.303) which converts diacetyl into (R)-2-Acetoin. This particular conversion is mainly found in food-borne lactobacilli (Chaillou *et al.*, 2005; Makarova *et al.*, 2006;

Tompkins *et al.*, 2012; Vogel *et al.*, 2011) and the only other mammalian-commensal species which contains this enzyme is *Lactobacillus reuteri* (Morita *et al.*, 2008). Unlike other sequenced mammalian lactobacilli there are no predicted carbohydrate-associated ATP-binding cassette (ABC) transporters encoded in the draft genome of *L. equi*. This suggests that *L. equi* like its *L. salivarius* clade counterpart *L. ruminis* (O' Donnell *et al.*, 2011) is primarily dependent on symporters and to a lesser extent phosphoenolpyruvate phosphotransferase system (PEP-PTS) to facilitate carbohydrate transport. The genome encodes PEP-PTS transporters to facilitate the transport of mannitol, mannose, sorbitol and galactitol, as well those for β-glucoside transport (Saier *et al.*, 2006; Saier *et al.*, 2009; Saier, 2000).

7.3.4 Amino acid metabolism of *L. equi*.

The genome of *L. equi* contains enzymes for the *de novo* biosynthesis or inter-conversions from intermediates of 15 amino acids and is auxotrophic for a further six amino acids. The level of auxotrophy is less than that of its *L. salivarius* clade co-members (Claesson *et al.*, 2006; Forde *et al.*, 2011) and more distantly related lactobacilli for example *L. acidophilus* (Altermann *et al.*, 2005) but not as low as *L. plantarum* (Kleerebezem *et al.*, 2003). This suggests that *L. equi* is not as dependent on extracellular amino acids for its growth as other mammalian-associated lactobacilli. The *L. equi* genome lacks the enzyme L-serine dehydratase (EC 4.3.1.17) present in both of the completed genomes of other *L. salivarius* clade members (Claesson *et al.*, 2006; Forde *et al.*, 2011). This enzyme is responsible for the conversion of pyruvate to serine and therefore, *L. equi* is predicted to be unable to synthesise the amino acid serine. Similarly, the genomes of *L. equi*, *L. ruminis* (Forde *et al.*, 2011) and *L. salivarius* (Claesson *et al.*, 2006) lack the enzyme threonine aldolase (EC 4.1.2.5) which means it likely that *L. equi* cannot synthesise glycine from threonine. However, *L. salivarius* (Claesson *et al.*, 2006) contains the enzyme glycine hydroxymethyltransferase which facilitates the conversion of serine to glycine but this enzyme is absent from *L. equi* and *L. ruminis* (Forde *et al.*, 2011).

7.3.5 Cell surface structures.

The surfaces of Gram Positive bacterial cells are often decorated with structures which influence their interaction with the environment and with other bacteria. The

genome of *L. equi* encodes 67 CDS with predicted signal protein sequences. The genome also encodes a potential exopolysaccharide (EPS) cluster. A 9-gene cluster (LEQ1594-1602) predicted to be involved in exopolysaccharide (EPS) production and transport was identified and this correlated with a deviation in the local GC content of the genome, typical for EPS operons in lactobacilli (Altermann *et al.*, 2005; Azcarate-Peril *et al.*, 2008). Genes for a further 13 glycosyltransferases proteins were identified in the genome as being potentially involved in EPS production, indicating the likely importance for this surface macromolecule in *L. equi*. Interestingly, a single predicted sortase enzyme was identified in the *L. equi* genome (LEQ0479). Sortase enzymes function to covalently anchor surface proteins to peptidoglycan and are found in all Gram-positive bacteria. The Sortase type A enzymes (SrtA) anchor proteins containing the characteristic substrate. From analysis of the genome, 4 proteins were identified as containing an LPxTG sortase anchor motif (LEQ1367, 1446, 1490, 1643). Two of these are fructan hydrolases and a PrtP-like subtilase discussed above, while LEQ1446 is a hypothetical protein which contains a surface exclusion domain and a structural maintenance of chromosomes (SMC) domain. The overall sortase-anchored protein repertoire of *L. equi* is significantly smaller than that of human-associated lactobacilli (Boekhorst *et al.*, 2005), suggesting that interaction with mucins, dendritic cells and mammalian receptors is less important in this species.

7.3.6 Stress proteins

An intestinal bacterium must be able to survive and protect itself from a variety of often harsh conditions such as pH, salinity and oxygen. The *L. equi* genome encodes a comprehensive array of stress survival proteins. We identified a putative cold shock protein (LEQ0317), heat shock protein (LEQ1726) and an alkaline shock protein (LEQ0226). An abortive phage resistance protein was also predicted in the *L. equi* genome. The genome of *L. equi* also harbours genes for a number of Clp proteases, (clpX and clpP), which previous studies have shown to be involved in the degradation of mis-folded proteins (Krüger *et al.*, 2000). The ability of an organism to survive, respond and eliminate reactive oxygen species is important (Fridovich, 1998). The *L. equi* genome encodes a number of thioredoxins (LEQ0856, 1096, 1693), a class of protein which act as antioxidants through the reduction of other proteins by cysteine thiol-disulfide exchange (Carmel-Harel & Storz, 2000). Encoded also within the *L.*

equi genome is a predicted thioredoxin reductase (LEQ0579) which has been shown to be a key enzyme in the oxidative stress response of *L. plantarum* WCFS1 catalyzing the regeneration of oxidised thioredoxin (Serrano *et al.*, 2007). The genome also encodes two glutaredoxins (NrdH) (LEQ0374, 0716) which function with thioredoxin reductase to catalyze the reduction of ribonucleotide reductase as part of cell redox homeostasis. A glutathione peroxidase (LEQ2366) was also identified from the genome and catalyzes the reduction of various hydroperoxides. A gene for a single peroxiredoxin (Prxs) (LEQ2501) was identified in the genome. Peroxiredoxins like thioredoxins are a ubiquitous family of antioxidant proteins which use thioredoxin (Trx) to aid its recycling following the detoxification of hydrogen peroxide (H_2O_2). This array of redox enzymes identified from an *in silico* examination of the genome suggest that *L. equi* has the ability to combat various oxidative stress exposures as part of a lifestyle involving intestinal and presumably also extra-intestinal phases.

7.3.7 Motility

Unlike some members of the *L. salivarius* clade for example, *L. ruminis*, *L. mali* and *Lactobacillus ghanensis* (Neville *et al.*, 2012). *L. equi* is non-motile and lacks all genes necessary for chemotaxis and flagellar formation.

7.4 Conclusion

The genome sequence of *L. equi* DPC 6820 provides a foundation for understanding the lactobacilli in an important but poorly characterised habitat, the equine hindgut. From the sequence analysis of the genome we noted a large repertoire of carbohydrate metabolism proteins (~5% of total gene content) which suggest that *L. equi* is adapted to a polysaccharide-rich environment. A large collection of encoded stress proteins and a relatively high amino acid autotrophy would also suggest that *L. equi* is able to readily adapt to changes in environmental conditions. With these factors in mind, *Lactobacillus equi* DPC 6820 has the potential to be harnessed as a probiotic modulator of both human and animal gut health.

7.5 References

Ajdić, D., McShan, W. M., McLaughlin, R. E. & other authors (2002). Genome sequence of *Streptococcus mutans* UA159, a cariogenic dental pathogen. *Proceedings of the National Academy of Sciences* **99**, 14434-14439.

Altermann, E., Russell, W. M., Azcarate-Peril, M. A. & other authors (2005). Complete genome sequence of the probiotic lactic acid bacterium *Lactobacillus acidophilus* NCFM. *Proc Natl Acad Sci USA* **102**, 3906-3912.

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

Canchaya, C., Claesson, M. J., Fitzgerald, G. F., van Sinderen, D. & O'Toole, P. W. (2006). Diversity of the genus *Lactobacillus* revealed by comparative genomics of five species. *Microbiology* **152**, 3185-3196.

Carmel-Harel, O. & Storz, G. (2000). Roles of the glutathione-and thioredoxin-dependent reduction systems in the *Escherichia coli* and *Saccharomyces cerevisiae* responses to oxidative stress. *Annual Reviews in Microbiology* **54**, 439-461.

Chaillou, S., Champomier-Verges, M.-C., Cornet, M. & other authors (2005). The complete genome sequence of the meat-borne lactic acid bacterium *Lactobacillus sakei* 23K. *Nat Biotech* **23**, 1527-1533.

Chen, C. C. & Cleary, P. P. (1990). Complete nucleotide sequence of the streptococcal C5a peptidase gene of *Streptococcus pyogenes*. *Journal of Biological Chemistry* **265**, 3161-3167.

Claesson, M. J., Li, Y., Leahy, S. & other authors (2006). Multireplicon genome architecture of *Lactobacillus salivarius*. *Proc Natl Acad Sci USA* **103**, 6718-6723.

Cleary, P. P., Prahbu, U., Dale, J. B., Wexler, D. E. & Handley, J. (1992). Streptococcal C5a peptidase is a highly specific endopeptidase. *Infection and Immunity* **60**, 5219-5223.

Delcher, A. L., Phillippy, A., Carlton, J. & Salzberg, S. L. (2002). Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Research* **30**, 2478-2483.

Felis, G. E. & Dellaglio, F. (2007). Taxonomy of Lactobacilli and Bifidobacteria. *Curr Issues Intest Microbiol* **8**, 44-61.

Fimereli, D. K., Tsirigos, K. D., Litou, Z. I., Liakopoulos, T. D., Bagos, P. G., & & Hamodrakas, S. J. (2012). CW-PRED: a HMM-Based method for the classification of cell wall-anchored proteins of gram-positive bacteria. *In Artificial Intelligence: Theories and Applications*, 285-290.

Forde, B. M., Neville, B. A., O' Donnell, M. M., Riboulet-Bisson, E., Claesson, M. J., Coghlan, A., Ross, R. P. & O' Toole, P. W. (2011). Genome sequences and

comparative genomics of two *Lactobacillus ruminis* strains from the bovine and human intestinal tracts. *Microb Cell Fact* **10**, S13.

Fridovich, I. (1998). Oxygen toxicity: a radical explanation. *Journal of Experimental Biology* **201**, 1203-1209.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In *Nucleic acids symposium series*, pp. 95-98.

Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J* **280**, 309-316.

Henrissat, B. & Bairoch, A. (1996). Updating the sequence-based classification of glycosyl hydrolases. *Biochem J* **316 (Pt 2)**, 695-696.

Holden, M. T. G., Heather, Z., Paillot, R. & other authors (2009). Genomic Evidence for the Evolution of *Streptococcus equi*: Host Restriction, Increased Virulence, and Genetic Exchange with Human Pathogens. *PLoS Pathog* **5**, e1000346.

Kankainen, M., Paulin, L., Tynkkynen, S. & other authors (2009). Comparative genomic analysis of *Lactobacillus rhamnosus* GG reveals pili containing a human-mucus binding protein. *Proc Natl Acad Sci USA* **106**, 17193-17198.

Kleerebezem, M., Boekhorst, J., van Kranenburg, R. & other authors (2003). Complete genome sequence of *Lactobacillus plantarum* WCFS1. *Proc Natl Acad Sci USA* **100**, 1990-1995.

Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *Journal of Molecular Biology* **305**, 567-580.

Krüger, E., Witt, E., Ohlmeier, S., Hanschke, R. & Hecker, M. (2000). The Clp Proteases of *Bacillus subtilis* Are Directly Involved in Degradation of Misfolded Proteins. *Journal of Bacteriology* **182**, 3259-3265.

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C. & Salzberg, S. L. (2004). Versatile and open software for comparing large genomes. *Genome Biol* **5**, R12.

Lane, D. (1991). *16s/23s rRNA sequencing*: John Wiley & Sons, West Sussex, United Kingdom.

Mackiewicz, P., Zakrzewska-Czerwińska, J., Zawilak, A., Dudek, M. R. & Cebrat, S. (2004). Where does bacterial replication start? Rules for predicting the oriC region. *Nucleic Acids Research* **32**, 3781-3791.

Makarova, K., Slesarev, A., Wolf, Y. & other authors (2006). Comparative genomics of the lactic acid bacteria. *Proc Natl Acad Sci USA* **103**, 15611-15616.

Morita, H., Toh, H., Fukuda, S. & other authors (2008). Comparative Genome Analysis of *Lactobacillus reuteri* and *Lactobacillus fermentum* Reveal a Genomic Island for Reuterin and Cobalamin Production. *DNA Res* **15**, 151-161.

Morita, H., Nakano, A., Shimazu, M. & other authors (2009). *Lactobacillus hayakitensis*, *L. equigenerosi* and *L. equi*, predominant lactobacilli in the intestinal flora of healthy thoroughbreds. *Anim Sci J* **80**, 339-346.

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C. & Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* **35**, W182-185.

Morotomi, M., Yuki, N., Kado, Y., Kushiro, A., Shimazaki, T., Watanabe, K. & Yuyama, T. (2002). Lactobacillus equi sp. nov., a predominant intestinal Lactobacillus species of the horse isolated from faeces of healthy horses. *Int J Syst Evol Micr* **52**, 211.

Naville, M., Ghuillot-Gaudeffroy, A., Marchais, A. & Gautheret, D. (2011). ARNold: a web tool for the prediction of Rho-independent transcription terminators. *RNA biology* **8**, 11-13.

Neville, B. A. & O'Toole, P. W. (2010). Probiotic properties of *Lactobacillus salivarius* and closely related *Lactobacillus* species. *Future Microbiology* **5**, 759-774.

Neville, B. A., Forde, B. M., Claesson, M. J., Darby, T., Coghlan, A., Nally, K., Ross, R. P. & O' Toole, P. W. (2012). Characterization of Pro-Inflammatory Flagellin Proteins Produced by *Lactobacillus ruminis* and Related Motile Lactobacilli. *PloS one* **7**, e40592.

Noguchi, H., Park, J. & Takagi, T. (2006). MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Res* **34**, 5623-5630.

O'Donnell, M. M., Forde, B. M., Neville, B. A., Ross, R. P. & O'Toole, P. W. (2011). Carbohydrate catabolic flexibility in the mammalian intestinal commensal *Lactobacillus ruminis* revealed by fermentation studies aligned to genome annotations. *Microb Cell Fact* **10**, S12.

Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods* **8**, 785-786.

Rissman, A. I., Mau, B., Biehl, B. S., Darling, A. E., Glasner, J. D. & Perna, N. T. (2009). Reordering contigs of draft genomes using the Mauve Aligner. *Bioinformatics* **25**, 2071-2073.

Saier, M. H., Tran, C. V. & Barabote, R. D. (2006). TCDB: the Transporter Classification Database for membrane transport protein analyses and information. *Nucleic Acids Res* **34**, D181-D186.

Saier, M. H., Yen, M. R., Noto, K., Tamang, D. G. & Elkan, C. (2009). The Transporter Classification Database: recent advances. *Nucleic Acids Res* **37**, D274-D278.

Saier, M. H., Jr. (2000). A functional-phylogenetic classification system for transmembrane solute transporters. *Microbiol Mol Biol Rev* **64**, 354-411.

Schattner, P., Brooks, A. N. & Lowe, T. M. (2005). The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Research* **33**, W686-W689.

Serrano, L. M., Molenaar, D., Wels, M., Teusink, B., Bron, P., de Vos, W. & Smid, E. (2007). Thioredoxin reductase is a key factor in the oxidative stress response of *Lactobacillus plantarum* WCFS1. *Microbial Cell Factories* **6**, 29.

Tompkins, T. A., Barreau, G. & Broadbent, J. R. (2012). Complete Genome Sequence of *Lactobacillus helveticus* R0052, a Commercial Probiotic Strain. *Journal of Bacteriology* **194**, 6349.

Vogel, R. F., Pavlovic, M., Ehrmann, M. A. & other authors (2011). Genomic analysis reveals *Lactobacillus sanfranciscensis* as stable element in traditional sourdoughs. *Microb Cell Fact* **10**, S6.

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191.

7.6 Supplementary information

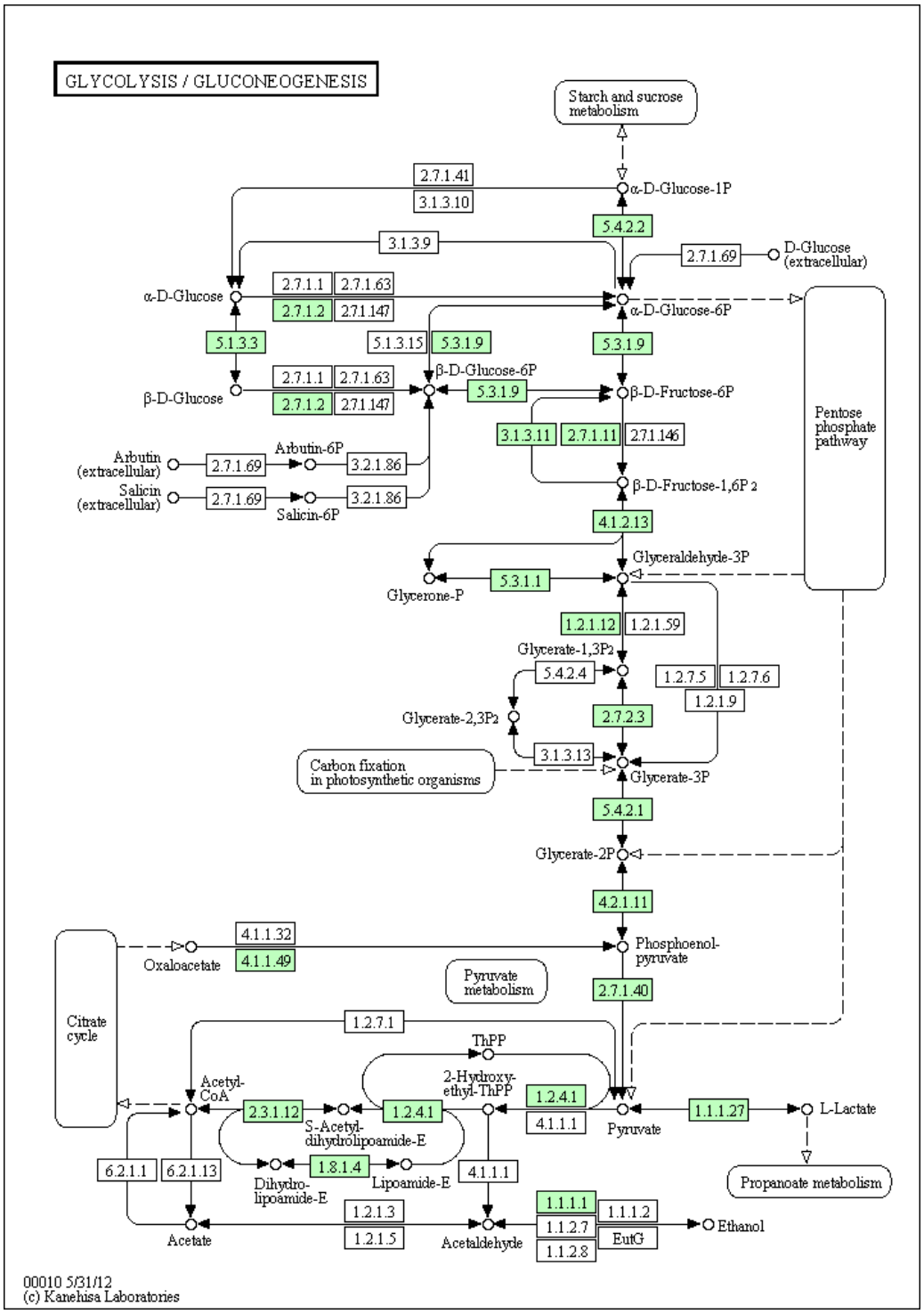


Fig S7.1 Glycolysis pathway in *L. equi* DPC 6820

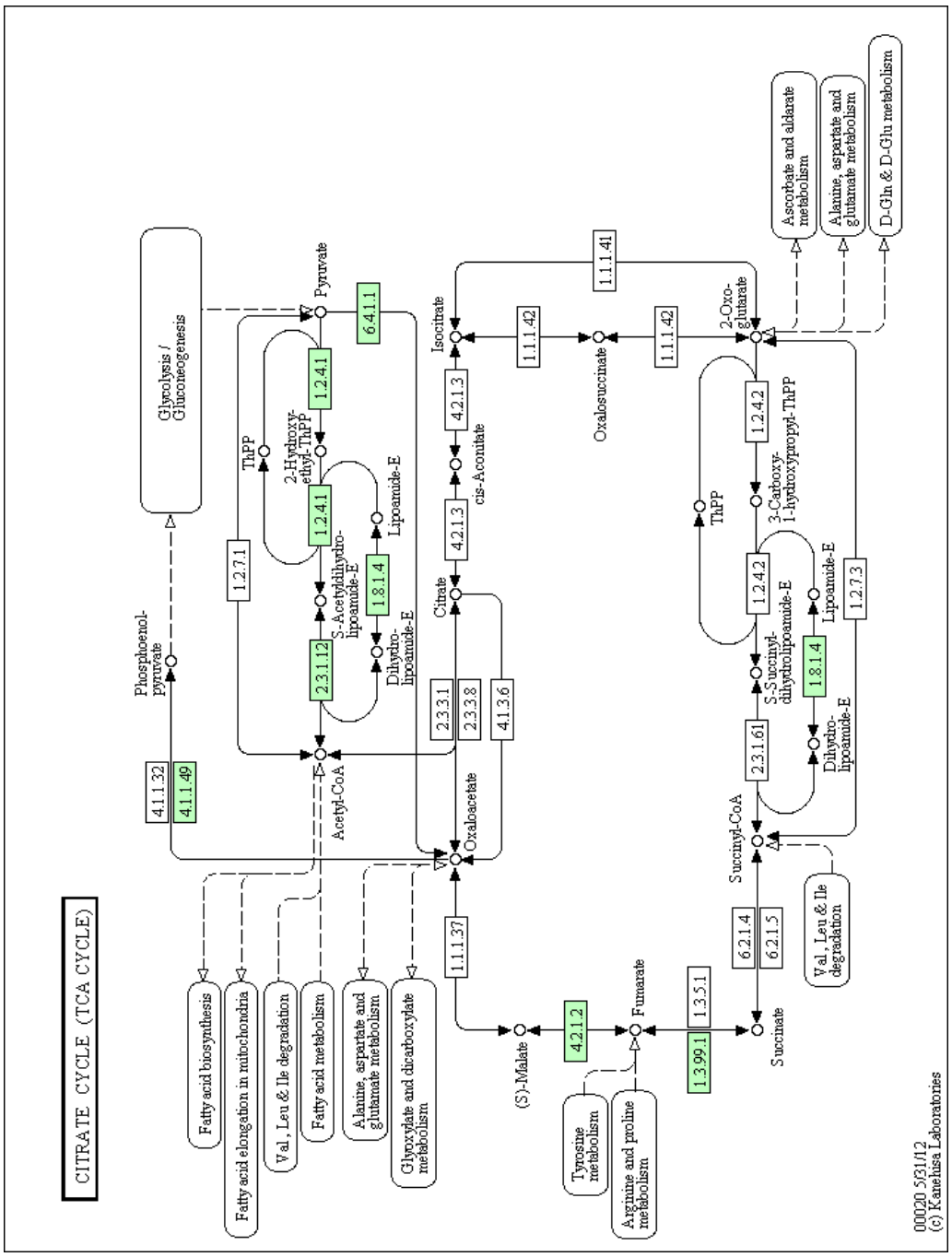


Fig S7.2 Citrate (TCA) cycle in *L. equi* DPC 6820

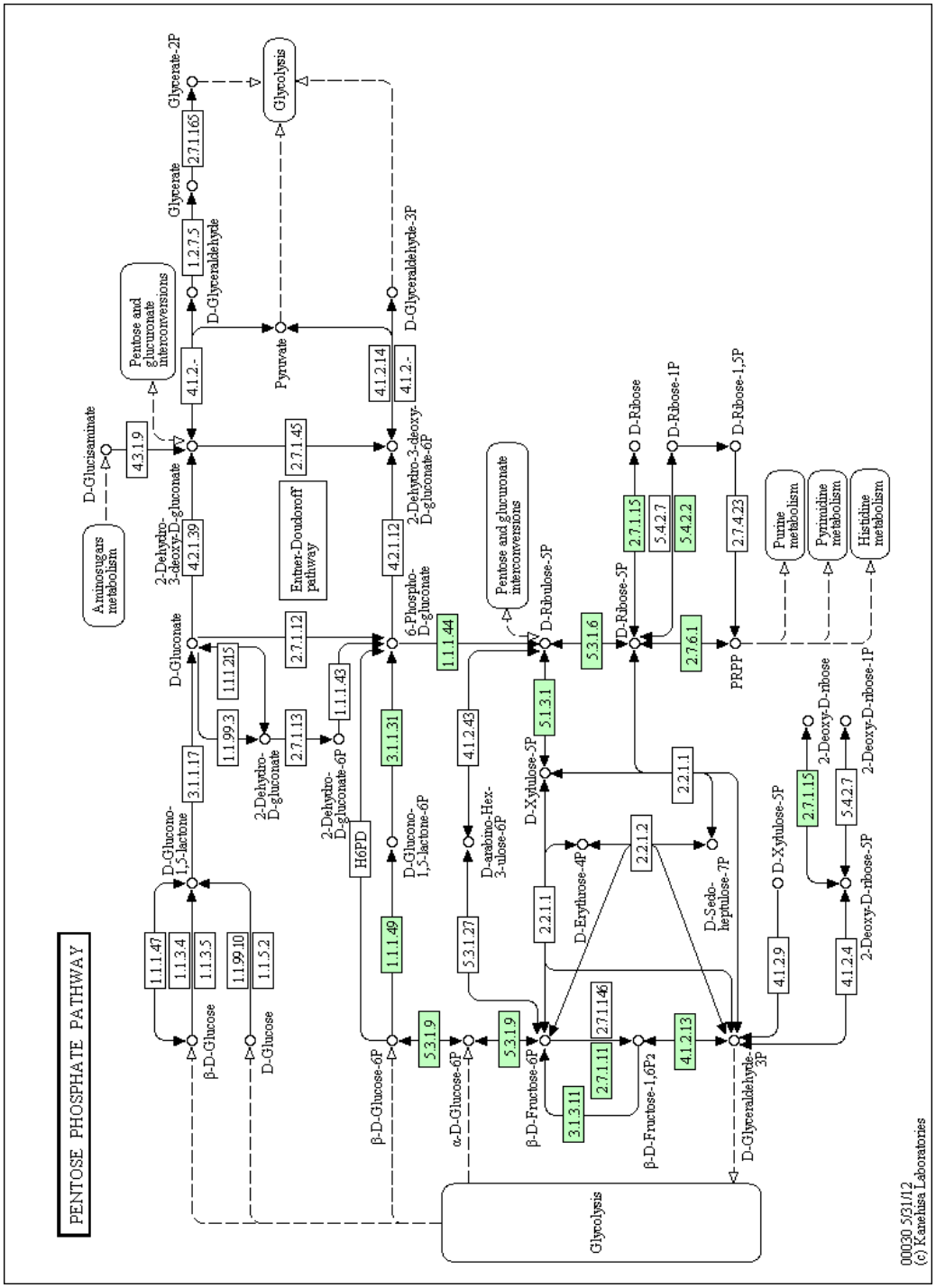


Fig S7.3 Pentose phosphate utilisation pathway in *L. equi* DPC 6820

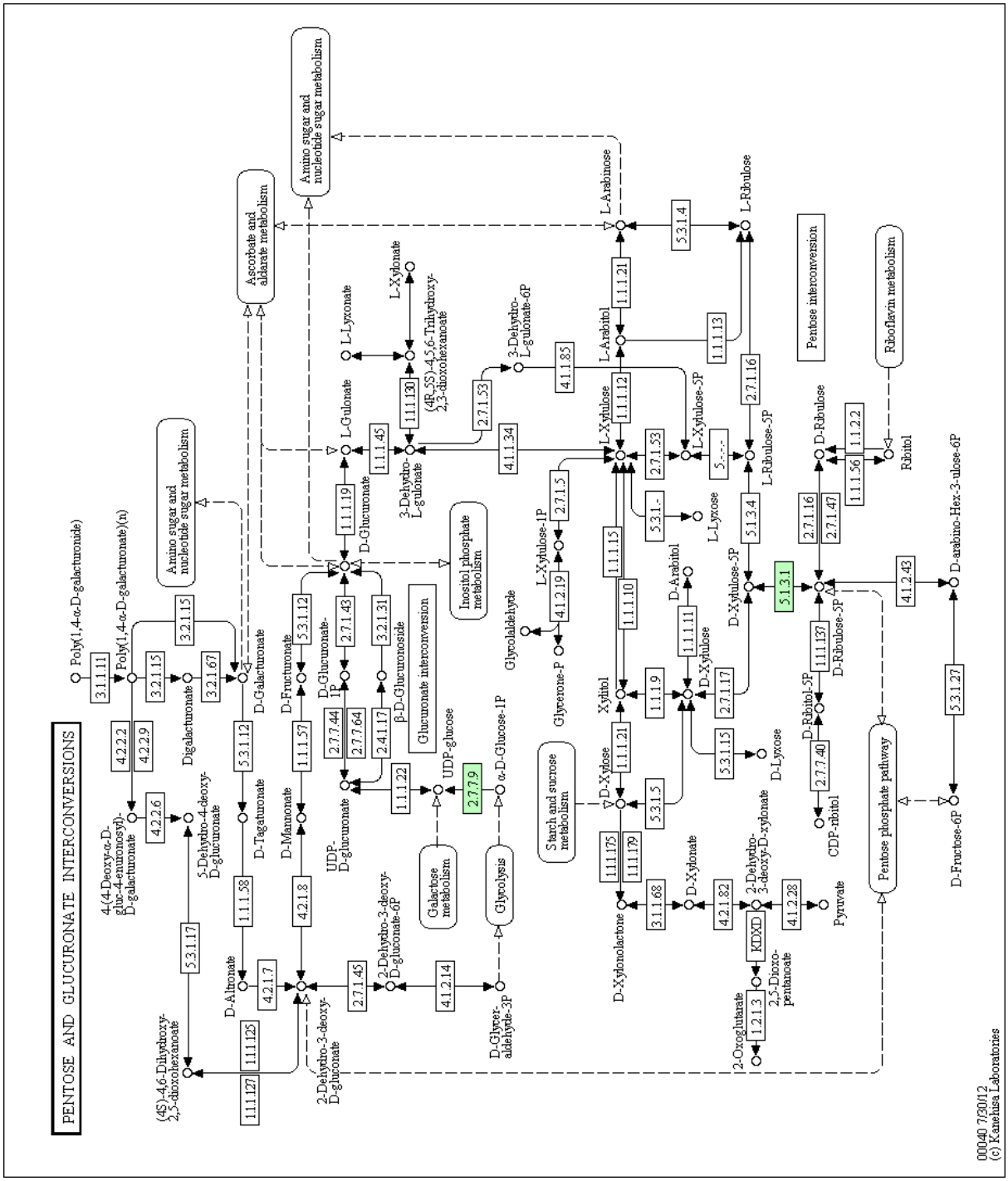


Fig S7.4 Pentose and glucuronate interconversions pathway in *L. equi* DPC 6820

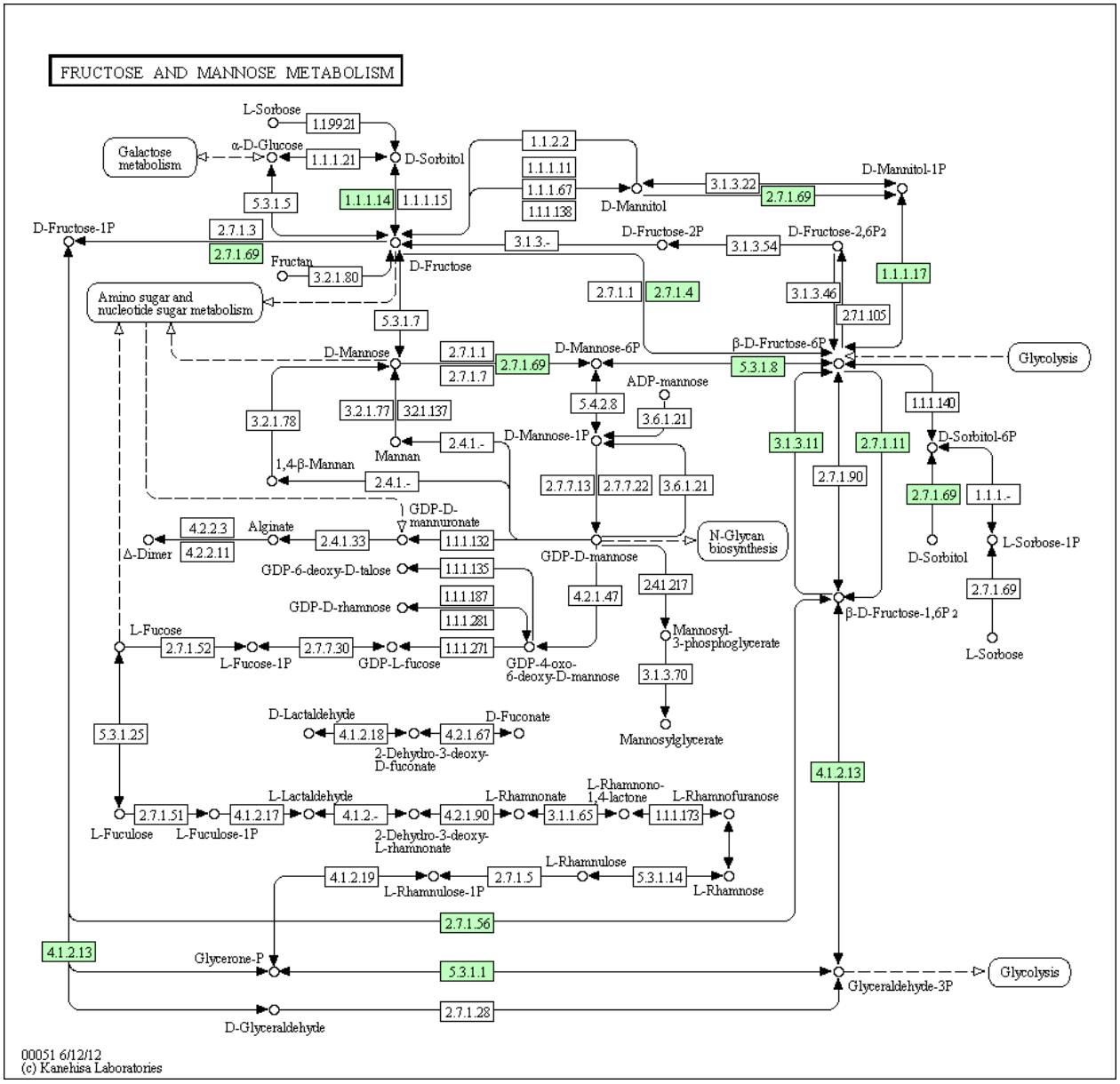


Fig S7.5 Fructose and mannose utilisation pathway in *L. equi* DPC 6820

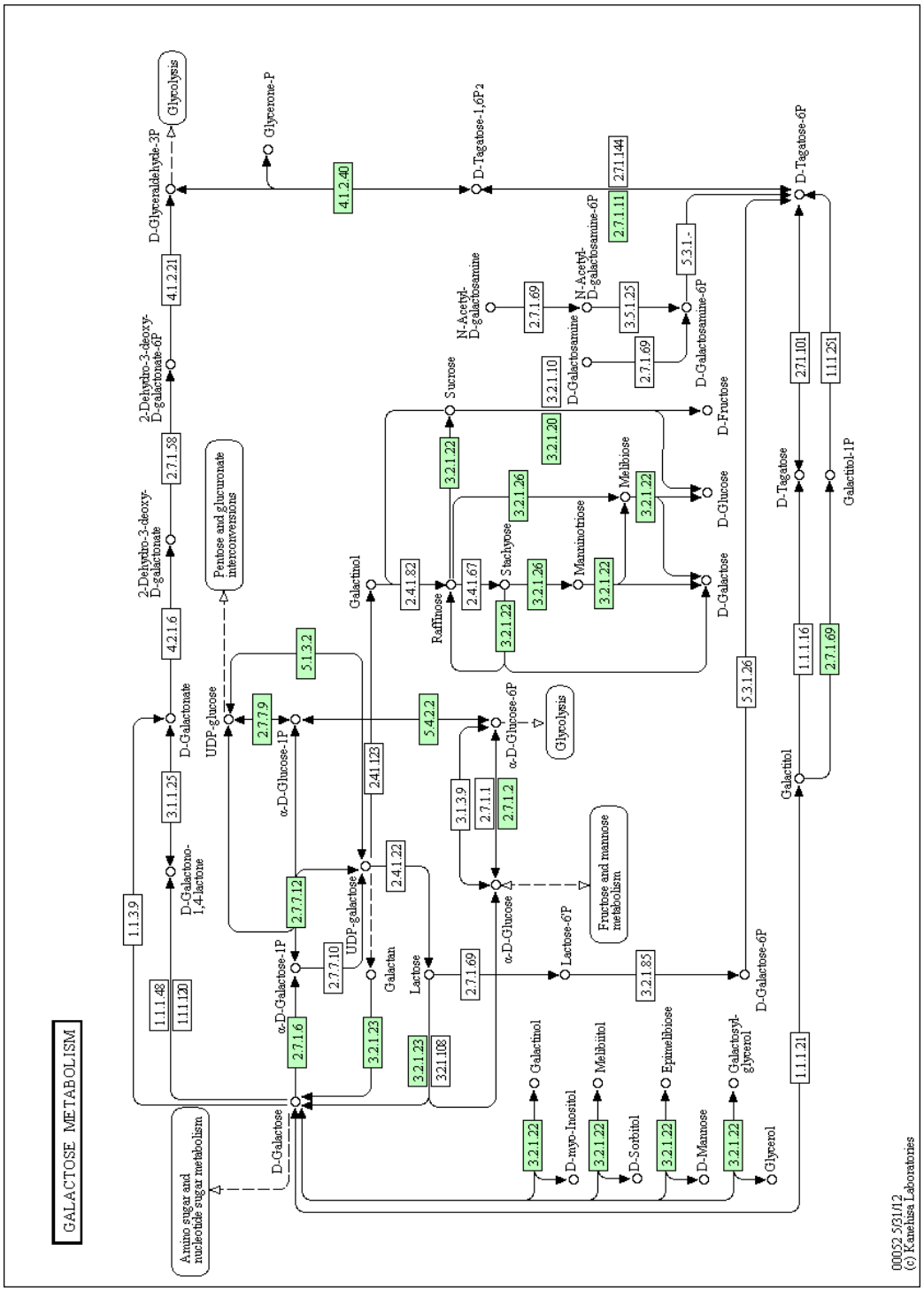


Fig S7.6 Galactose utilisation pathway in *L. equi* DPC 6820

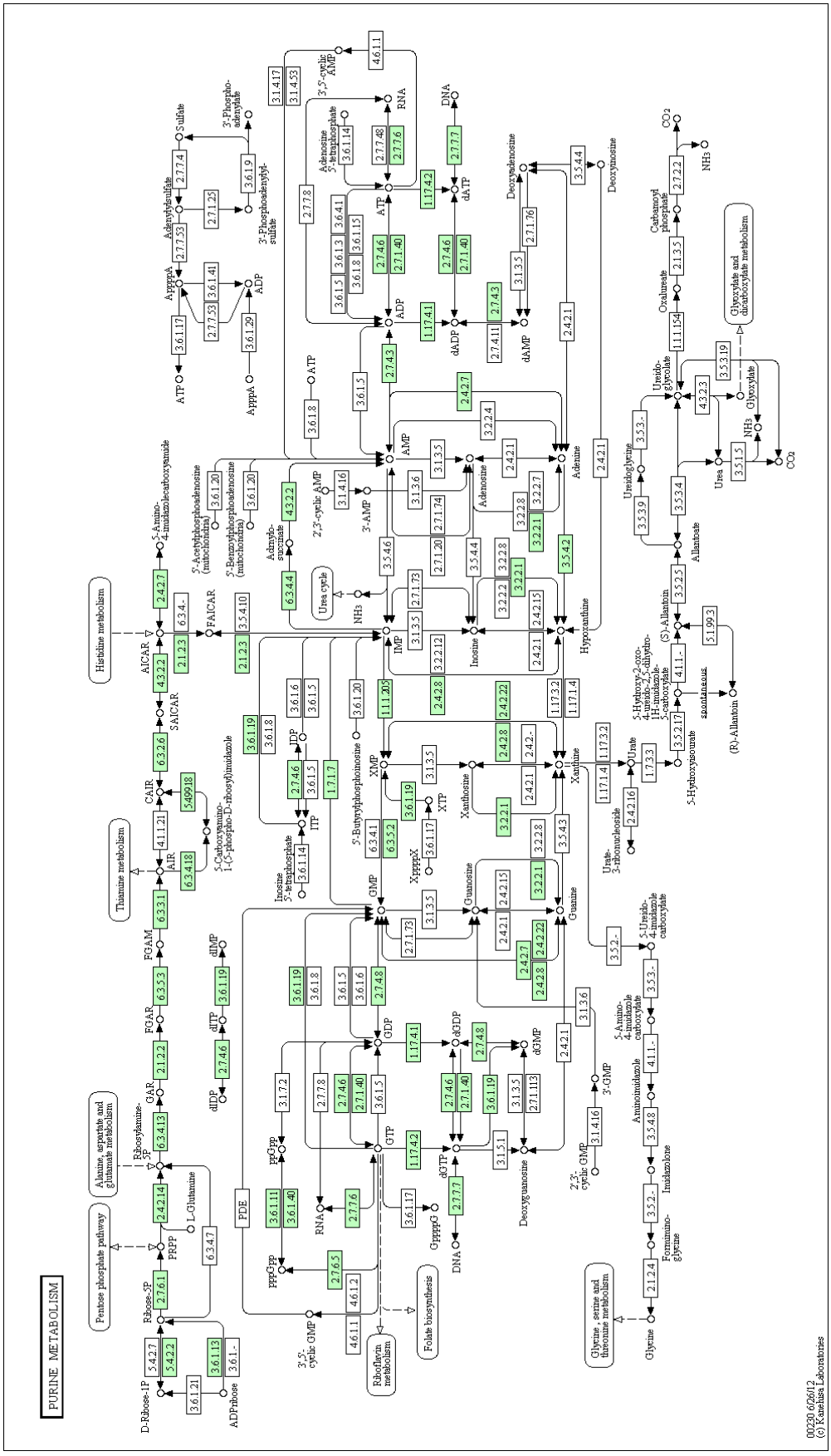


Fig S7.7 Purine metabolism pathways in *L. equi* DPC 6820

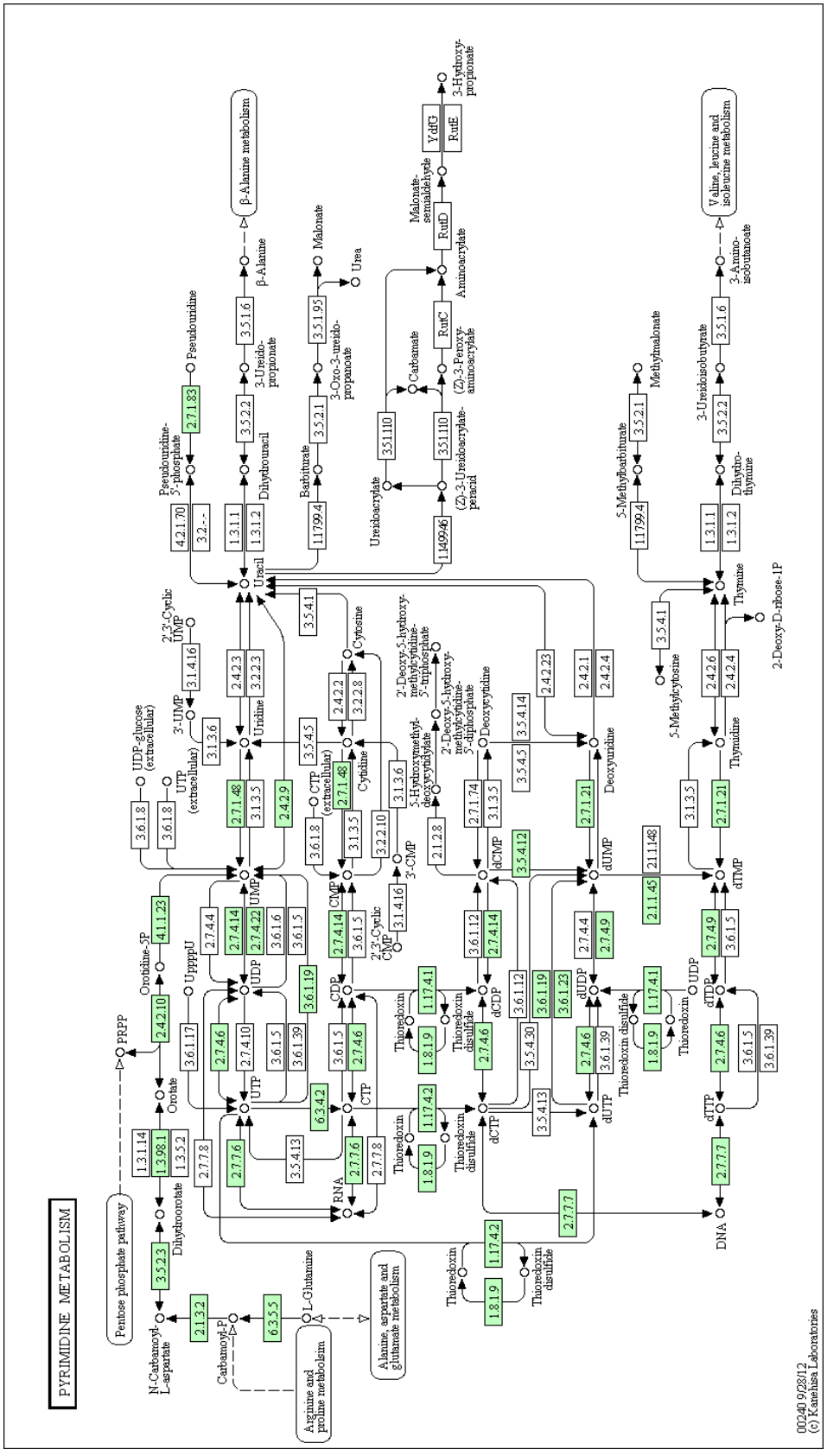


Fig S7.8 Pyrimidine metabolim pathways in *L .equi* DPC 6820

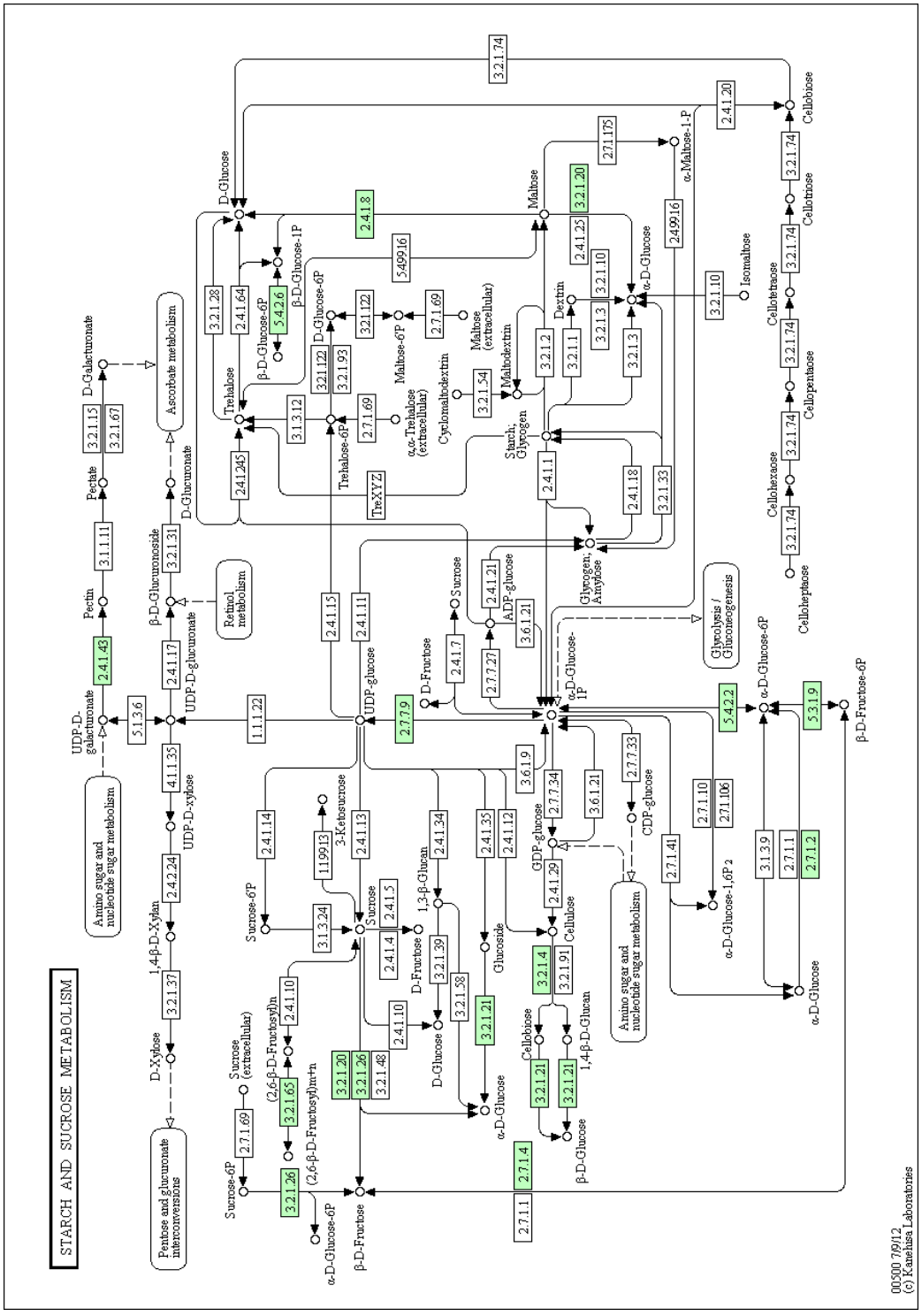


Fig S7.9 Starch and sucrose metabolic pathways in *L. equi* DPC 6820

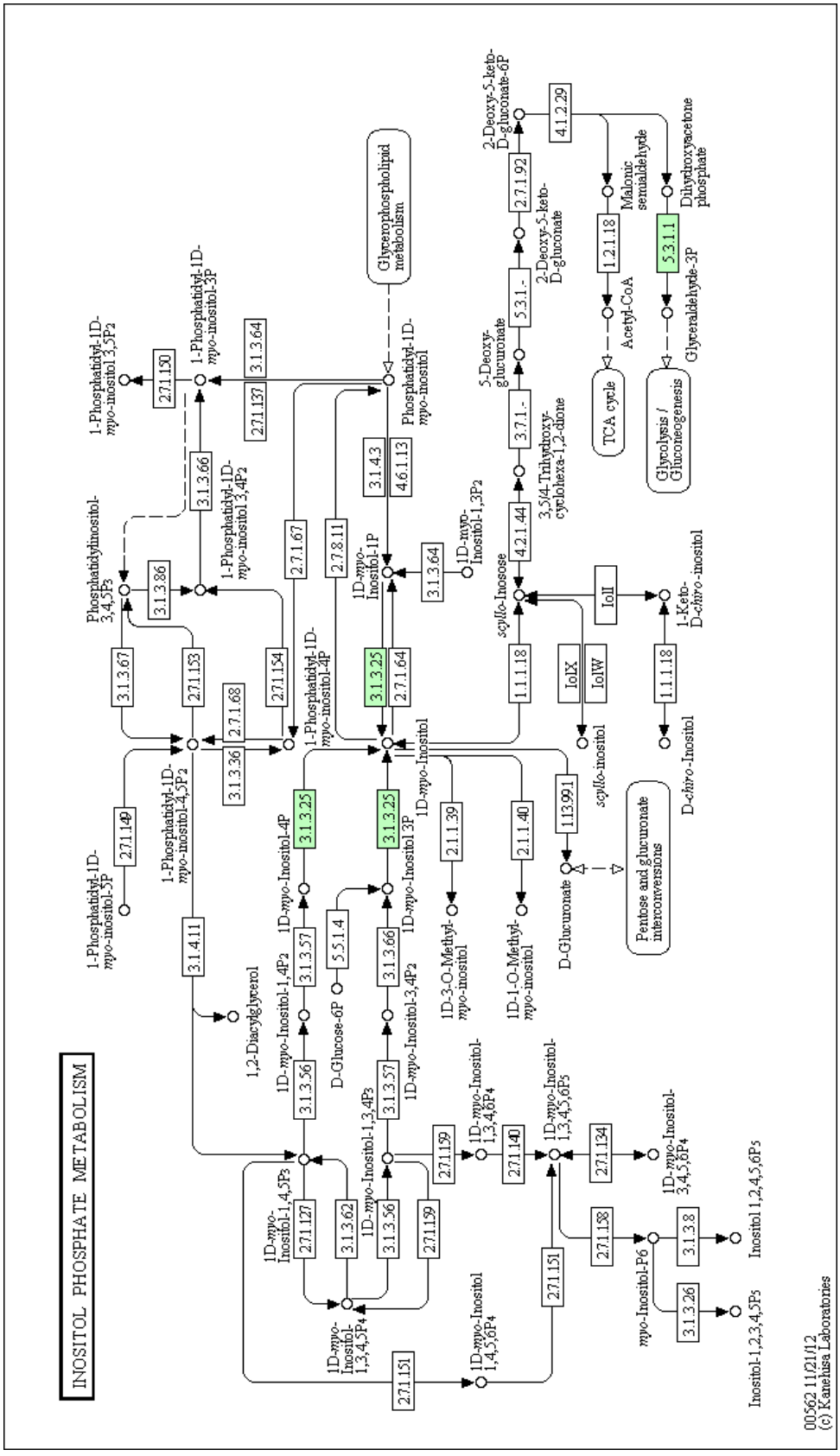


Fig S7.10 Inositol phosphate metabolic pathways in *L. equi* DPC 6820

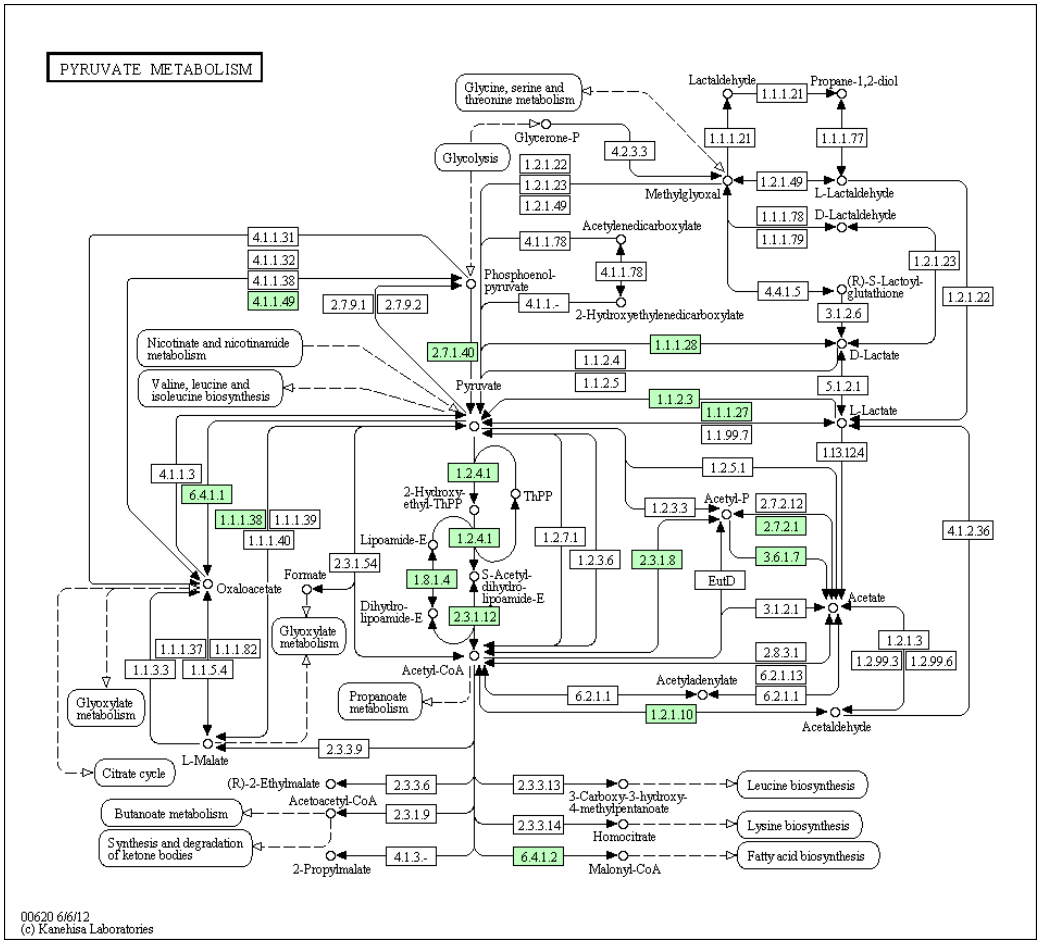


Fig S7.11 Pyruvate metabolism pathway in *L. equi* DPC 6820

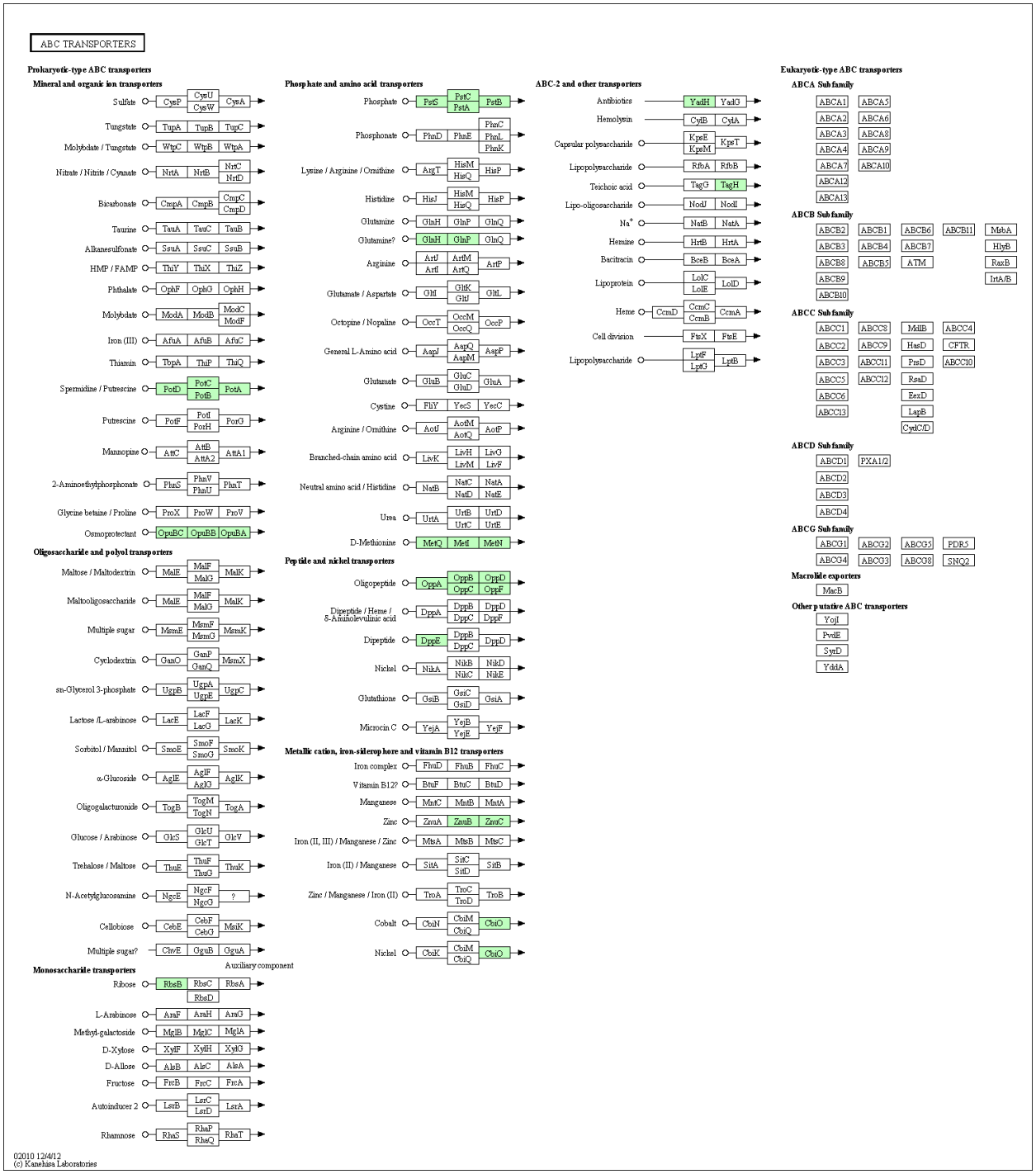


Fig S7.12 ABC transporters in *L. equi* DPC 6820

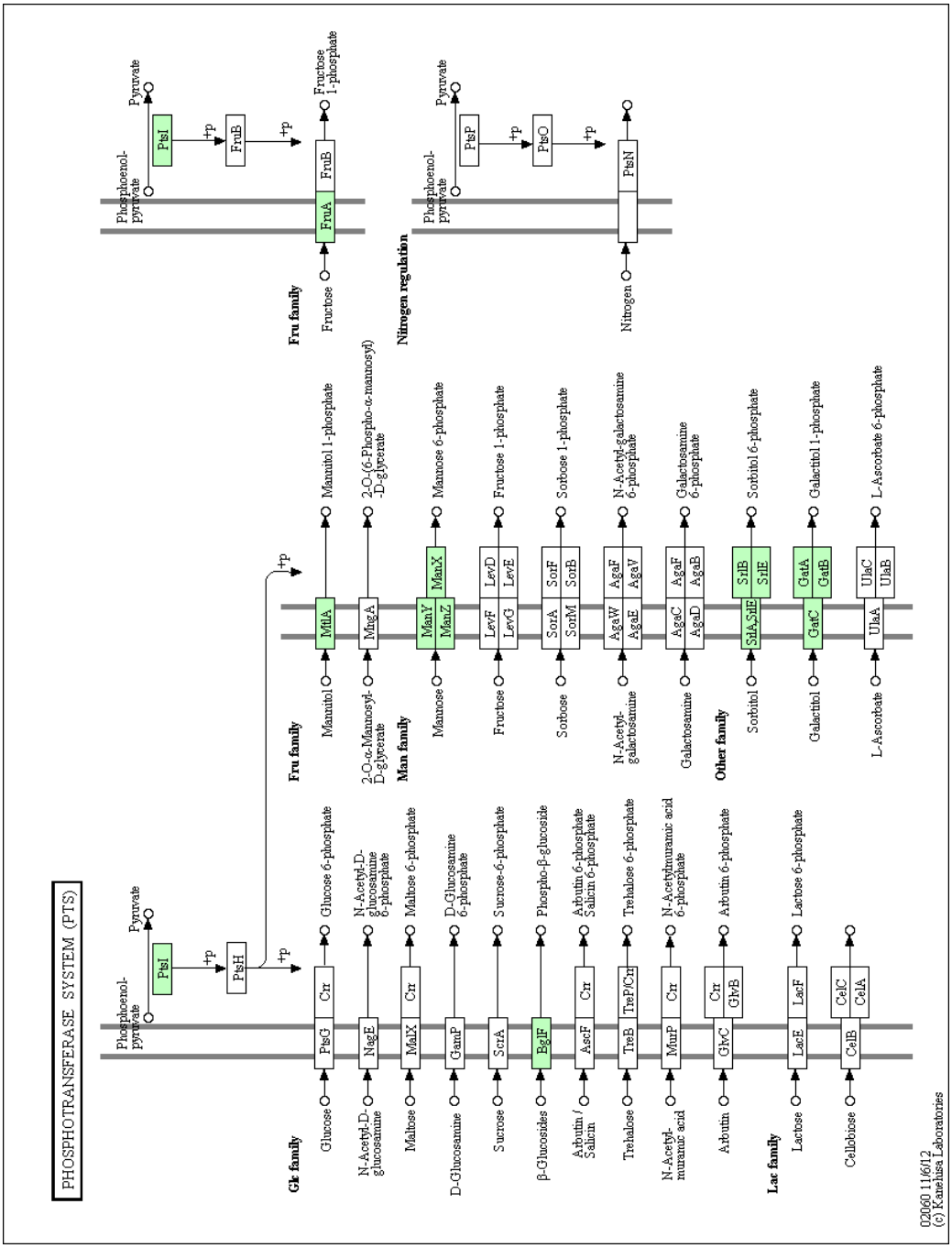


Fig S7.13 Phosphotransferase system (PTS) tranporters present in *L. equi* DPC 6820

Chapter VIII

Discussion/Conclusion

8.1 General discussion

This thesis aimed to characterise the catabolic flexibility of commensal lactobacilli, primarily *Lactobacillus ruminis*. It also aimed to gain a better understanding of the effects that dietary carbohydrates and carbohydrate supplementation can have on the microbiota of humans and animals. Catabolic flexibility is vital for the *Lactobacillus* genus especially for those species exposed to a variety of carbohydrates on a regular basis. The ability to degrade and transport complex carbohydrates is one of the mechanisms that allows *Lactobacillus* spp. to survive and flourish in niche environments like the gastrointestinal tract. Diets can have a profound effect on the diversity in the intestinal microbiota of both humans and animals (Claesson *et al.*, 2012; De Filippo *et al.*, 2010; Muegge *et al.*, 2011).

Chapter II provided the first in-depth characterisation of the fermentation capabilities of *L. ruminis* from two mammalian hosts. The research described in Chapter III, created a platform for future dietary interventions in the microbiota of economically important, hindgut fermenting equids, by identifying a core microbiota in the equine gut. Chapter IV marked the first comparative study into the bacterial diversity and the core microbiota of domesticated herbivores. The research described in Chapter V elucidated the genetic diversity and survival characteristics of *L. ruminis* strains from different mammalian hosts. In Chapter VI, the effect of diets supplemented with a prebiotic or synbiotic on the porcine gut microbiota was assessed. However, the characterisation of the effect of the dietary supplements was hindered by the lack of a control group. In Chapter VII, the genome of the equine associated *L. salivarius* clade species, *Lactobacillus equi* was annotated and sequenced for the first time, providing an insight into the ecology of an equine-associated lactobacillus.

Chapter II, Chapter VI and Chapter VII of this thesis succeeded in expanding the knowledge base of carbohydrate degradation potential in the mammalian-associated commensal lactobacilli of the *L. salivarius* clade. The catabolic flexibility of non-*L. salivarius* species within the clade was not well characterised until now. Genome sequencing and analysis of commensal lactobacilli have revealed that glycosyl hydrolases (EC 3.2.1) are vital for catabolic flexibility (Barrangou *et al.*, 2006). The carbohydrates Degree of Polymerisation (DP) is a limiting factor for carbohydrate fermentation by many species of commensal bacteria (Crittenden *et al.*, 2002; Gopal *et al.*, 2001; Hopkins *et al.*, 1998). Many of the carbohydrates classed as

prebiotics (fructooligosaccharides, galactooligosaccharides and lactulose) have a DP < 10 and are readily fermented by lactobacilli and bifidobacteria (Cummings *et al.*, 2001). Longer carbohydrate polymers (β glucans, starches and inulins) are unlikely to be degraded by commensal lactobacilli, but they have the potential to utilise the hydrolysates of these longer polymers as carbon sources (Jaskari *et al.*, 1998; Snart *et al.*, 2006). *L. ruminis* is a resident species in the lower intestinal tract in humans and animals (Reuter, 2001), while *L. equi* is present in the equine hindgut microbiota (Hidetoshi *et al.*, 2009; Morotomi *et al.*, 2002). The localisation of *L. ruminis* and *L. equi* to the colon and caecum means that there is limited availability of simpler carbohydrates (mono and di-saccharides) to ferment (Wong & Jenkins, 2007) but a readily available supply of non-digestible oligosaccharides. Chapter II confirmed that catabolic flexibility in *L. ruminis* was dependent on a cadre of glycosyl hydrolases. The carbohydrate fermentation capacity of *L. ruminis* was also dictated by the DP of the carbohydrates analysed. In Chapter VII, predicted glycosyl hydrolases associated with degradation of plant polysaccharides (fructans) were predicted in the genome of *L. equi*. The presence of such glycosyl hydrolases indicated that *L. equi* has adapted to its niche within the hindgut of horses where large volumes of plant matter transit daily.

The carbohydrate flexibility studied both *in vitro* and *in silico* in membrane of the *L. salivarius* clade is strongly indicative of niche adaptation to life within a polysaccharide consuming host. It suggests a degree of mutualism between the host and the *L. salivarius* clade species, where the host benefits from the energy generated from the carbohydrate degradation and the bacterium benefits from the access to nutrients within the gastrointestinal niche environment.

Since the domestication of animals thousands of years ago animal health and welfare have long been a priority for researchers and animal husbandry practitioners alike. Horses are domesticated herbivores with uses in farm labour, sport and recreation. In order to expand upon the knowledge of the equine microbiota researchers have recently examined the effects of diet, health status and breed on the microbiota of horses (Costa *et al.*, 2012; Milinovich *et al.*, 2006; Shepherd *et al.*, 2012; Steelman *et al.*, 2012; Willing *et al.*, 2009b; Yamano *et al.*, 2008). Complementing this literature Chapter III identified the core faecal microbiota of healthy Thoroughbred racehorses irrespective of internal (diet) and environmental (geographic location or activity) factors. The taxa identified within the faecal

microbiota are a valuable addition to equine gut health research. Although slightly limited in our conclusions by the sample size, it is the opinion of the author of this thesis that dietary supplementation of racehorses has potential negative health implications. This data may form the basis for future studies into equine dietary supplements and allow researchers to combat common dysbiosis-related diseases of the equine microbiota (Milinovich *et al.*, 2010; Pollitt, 2004; Shirazi-Beechey, 2008).

The intestinal microbiota of mammals is a complex environment populated by millions of bacteria. Many factors can influence the diversity of the mammalian microbiota including age, gender, health status and antibiotics (Keegan *et al.*, 2005; Ley *et al.*, 2005; Mueller *et al.*, 2006; O'Toole & Claesson, 2010). Recent evidence suggests however, that diet is a major factor affecting the mammalian microbiota (Claesson *et al.*, 2012; Ley *et al.*, 2008). The effect of diet and by extension digestion type on the microbiota was confirmed in Chapter IV. The effect of dietary modulation using prebiotics was examined in Chapter V; however, no effect on the diversity of the microbiota was measured. But the synbiotic treatment used in Chapter V reduced the microbiota diversity following the 14 day study. The mechanism behind this reduction is unknown at this time but a possible factor was the age of the pigs. A newly weaned pigs intestinal microbiota is in a state of instability due to the dietary transition from digesting milk to solid feed (Richards *et al.*, 2005). The addition of a synbiotic containing readily fermentable, low DP carbohydrates and live *L. ruminis* cells may have a large, albeit temporary effect on the porcine microbiota during this time. A temporal study monitoring the effect of the synbiotic over a number of weeks post weaning would be needed to assess the effect of synbiotic supplementation on the porcine microbiota. Chapters III, IV and V also revealed that a large proportion of the mammalian microbiota remains uncharacterised and poorly understood. However, understandably the progress in characterising the bacteria of the intestinal microbiota is slow due to the difficulty in isolating, culturing and characterising new isolates.

Genome sequencing and comparative studies are extremely important for identifying the phylogeny of a species and can be used to infer and inform future studies (Felis & Dellaglio, 2007). Genome sequence analysis and annotation is vital to understand the regulation and function pathways within the cell. In Chapter II, *in silico* metabolic pathway mapping was used to infer and predict the mechanisms used by *L. ruminis* to degrade carbohydrates. The *in silico* predictions were combined with *in vitro* growth assays to facilitate an in-depth characterisation of the catabolic

flexibility of *L. ruminis*. In Chapter VI, the genome sequences of four *L. ruminis* strains (S23, ATCC 25644, ATCC 27782 and DPC 6832) were compared to identify regions of synteny and homology between each and also to identify the whole genome phylogenetic relationship of the four strains. In Chapter VII whole genome sequencing and annotation was used to identify the similarity of *L. equi* to other *L. salivarius* clade species and other commensal bacteria.

The field of high throughput metagenomics has allowed researchers to identify and characterise the microbiota of diverse habitats, for example the human gastrointestinal tract (Claesson *et al.*, 2012; De Filippo *et al.*, 2010; Nam *et al.*, 2011; Qin *et al.*, 2010). Knowledge of the microbial consortium present in the gastrointestinal tracts of human and animals is of potentially great importance for maintaining the health and welfare of both. Many gastrointestinal diseases are thought to be caused or exacerbated by bacteria in the microbiota (Codling *et al.*, 2010; Joossens *et al.*, 2011; Pollitt, 2004). Identification of disease causing bacteria in the microbiota is essential for the development of methods to modify the microbiota and return the microbiota to a homeostatic state. Meta-transcriptomics may be used to identify particular biological functions encoded by the microbiota which could be exploited by the food and beverage industries to modify the microbiota. Targeted prebiotics could be used to ameliorate microbial dysbiosis which current prebiotics, to date, have been unable to do (Whelan, 2013).

In the future, carbohydrate degrading enzymes identified in commensal lactobacilli could be used in the biotechnology industry to generate novel prebiotics and prebiotic hydrolysates (Díez-Municio *et al.*, 2013; Iqbal *et al.*, 2010; Maischberger *et al.*, 2010). The development of prebiotics from commensal carbohydratases may allow researchers to expand upon the definition of a prebiotic and target specific beneficial bacteria within the microbiota. Many carbohydrates currently classed as prebiotics can be fermented by bacteria outside the generally defined parameters (*Lactobacillus* and *Bifidobacterium*) for a prebiotic target organism (Rada *et al.*, 2008).

Dietary supplementation of performance horses in the future should take the microbiota influence of readily fermentable carbohydrates in the diets into account. While supplementation of the equine diet may be beneficial for endurance and training (Lawrence, 1990), it may also have negative consequences for the composition and function of the equine microbiota (Shirazi-Beechey, 2008). This

information is most relevant to large industrial mills producing horse feed and supplements. The effect of dietary supplementation in conjunction with other factors such as breed, gender and age should be taken into account before feeding an animal vast quantities of fermentable carbohydrates. Furthermore, a return to a more natural grazing style of feeding may be of greater long term benefit both to the individual animals but also to the bloodstock industry.

In conclusion, this thesis has used a combination of classical microbiology, molecular microbiology, genomics and culture independent microbiota analysis to study *L. ruminis*, another member of the *L. salivarius* clade lactobacilli, *L. equi*. The same methodologies were used to elucidate the effect of diet on the mammalian intestinal microbiota. The findings have contributed to our understanding of the gut ecology of intestinal lactobacilli and helped elucidate the genetic basis for interaction of intestinal lactobacilli with diet, microbiota and host.

8.2 References

Barrangou, R., Azcarate-Peril, M. A., Duong, T., Conners, S. B., Kelly, R. M. & Klaenhammer, T. R. (2006). Global analysis of carbohydrate utilization by *Lactobacillus acidophilus* using cDNA microarrays. *P Natl Acad Sci USA* **103**, 3816-3821.

Claesson, M. J., Jeffery, I. B., Conde, S. & other authors (2012). Gut microbiota composition correlates with diet and health in the elderly. *Nature* **488**, 178-184.

Codling, C., O'Mahony, L., Shanahan, F., Quigley, E. M. & Marchesi, J. (2010). A Molecular Analysis of Fecal and Mucosal Bacterial Communities in Irritable Bowel Syndrome. *Digestive Diseases and Sciences* **55**, 392-397.

Costa, M. C., Arroyo, L. G., Allen-Vercoe, E., StÃ¤mpfli, H. R., Kim, P. T., Sturgeon, A. & Weese, J. S. (2012). Comparison of the fecal microbiota of healthy horses and horses with colitis by high throughput sequencing of the V3-V5 region of the 16S rRNA gene. *PLOS One* **7**, e41484.

Crittenden, R., Karppinen, S., Ojanen, S., Tenkanen, M., Fagerström, R., Mättö, J., Saarela, M., Mattila-Sandholm, T. & Poutanen, K. (2002). *In vitro* fermentation of cereal dietary fibre carbohydrates by probiotic and intestinal bacteria. *Journal of the Science of Food and Agriculture* **82**, 781-789.

Cummings, J. H., Macfarlane, G. T. & Englyst, H. N. (2001). Prebiotic digestion and fermentation. *The American Journal of Clinical Nutrition* **73**, 415s-420s.

De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J. B., Massart, S., Collini, S., Pieraccini, G. & Lionetti, P. (2010). Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *P Natl Acad Sci USA* **107**, 14691-14696.

Díez-Municio, M., de las Rivas, B., Jimeno, M. L., MuÃ±oz, R., Moreno, F. J. & Herrero, M. (2013). Enzymatic Synthesis and Characterization of Fructooligosaccharides and Novel Maltosylfructosides by Inulosucrase from *Lactobacillus gasseri* DSM 20604. *Applied and environmental microbiology* **79**, 4129-4140.

Felis, G. E. & Dellaglio, F. (2007). Taxonomy of Lactobacilli and Bifidobacteria. *Curr Issues Intest Microbiol* **8**, 44-61.

Gopal, P. K., Sullivan, P. A. & Smart, J. B. (2001). Utilisation of galacto-oligosaccharides as selective substrates for growth by lactic acid bacteria including *Bifidobacterium lactis* DR10 and *Lactobacillus rhamnosus* DR20. *International Dairy Journal* **11**, 19-25.

Hidetoshi, M., Akiyo, N., Mitsuharu, S. & other authors (2009). *Lactobacillus hayakitensis, L. equigenerosi* and *L. equi*, predominant lactobacilli in the intestinal flora of healthy thoroughbreds. *Anim Sci J* **80**, 339-346.

Hopkins, Cummings & Macfarlane (1998). Inter-species differences in maximum specific growth rates and cell yields of bifidobacteria cultured on oligosaccharides and other simple carbohydrate sources. *Journal of Applied Microbiology* **85**, 381-386.

Iqbal, S., Nguyen, T.-H., Nguyen, T. T., Maischberger, T. & Haltrich, D. (2010). β-Galactosidase from *Lactobacillus plantarum* WCFS1: biochemical characterization and formation of prebiotic galacto-oligosaccharides. *Carbohydrate Research* **345**, 1408-1416.

Jaskari, J., Kontula, P., Siitonen, A., Jousimies-Somer, H., Mattila-Sandholm, T. & Poutanen, K. (1998). Oat beta-glucan and xylan hydrolysates as selective substrates for *Bifidobacterium* and *Lactobacillus* strains. *Appl Microbiol Biotechnol* **49**, 175-181.

Joossens, M., Huys, G., Cnockaert, M., De Preter, V., Verbeke, K., Rutgeerts, P., Vandamme, P. & Vermeire, S. (2011). Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. *Gut* **60**, 631-637.

Keegan, T. P., Dritz, S. S. & Nelssen, J. L. (2005). Effects of in-feed antimicrobial alternatives and antimicrobials on nursery pig performance and weight variation. *J Swine Health Prod* **13**, 12-18.

Lawrence, L. M. (1990). Nutrition and fuel utilization in the athletic horse. *Veterinary Clinics of North America, Equine Practice* **6**, 393-418.

Ley, R. E., Bäckhed, F., Turnbaugh, P., Lozupone, C. A., Knight, R. D. & Gordon, J. I. (2005). Obesity alters gut microbial ecology. *PNAS* **102**, 11070-11075.

Ley, R. E., Hamady, M., Lozupone, C. & other authors (2008). Evolution of mammals and their gut microbes. *Science* **320**, 1647.

Maischberger, T., Leitner, E., Nitisinprasert, S., Juajun, O., Yamabhai, M., Nguyen, T.-H. & Haltrich, D. (2010). β-Galactosidase from *Lactobacillus pentosus*: Purification, characterization and formation of galacto-oligosaccharides. *Biotechnology Journal* **5**, 838-847.

Milinovich, G., Trott, D., Burrell, P., Van Eps, A., Thoefner, M., Blackall, L., Al Jassim, R., Morton, J. & Pollitt, C. (2006). Changes in equine hindgut bacterial populations during oligofructose induced laminitis. *Environ Microbiol* **8**, 885-898.

Milinovich, G. J., Klieve, A. V., Pollitt, C. C. & Trott, D. J. (2010). Microbial events in the hindgut during carbohydrate-induced equine laminitis. *Vet Clin North Am-Equine* **26**, 79-94.

Morotomi, M., Yuki, N., Kado, Y., Kushiro, A., Shimazaki, T., Watanabe, K. & Yuyama, T. (2002). Lactobacillus equi sp. nov., a predominant intestinal Lactobacillus species of the horse isolated from faeces of healthy horses. *Int J Syst Evol Micr* **52**, 211.

Muegge, B. D., Kuczynski, J., Knights, D., Clemente, J. C., GonzÃ¡lez, A., Fontana, L., Henrissat, B., Knight, R. & Gordon, J. I. (2011). Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science* **332**, 970-974.

Mueller, S., Saunier, K., Hanisch, C. & other authors (2006). Differences in fecal microbiota in different European study populations in relation to age, gender, and country: a cross-sectional study. *Appl Environ Microbiol* **72**, 1027-1033.

Nam, Y.-D., Jung, M.-J., Roh, S. W., Kim, M.-S. & Bae, J.-W. (2011). Comparative analysis of Korean human gut microbiota by barcoded pyrosequencing. *PLOS One* **6**, e22109.

O'Toole, P. W. & Claesson, M. J. (2010). Gut microbiota: Changes throughout the lifespan from infancy to elderly. *Int Dairy J* **20**, 281-291.

Pollitt, C. C. (2004). Equine laminitis. *Clin Tech Equine Pract* **3**, 34-44.

Qin, J., Li, R., Raes, J. & other authors (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59-65.

Rada, V., Nevoral, J., Trojanová, I., Tománková, E., Šmehilová, M. & Killer, J. (2008). Growth of infant faecal bifidobacteria and clostridia on prebiotic oligosaccharides in *in vitro* conditions. *Anaerobe* **14**, 205-208.

Reuter, G. (2001). The *Lactobacillus* and *Bifidobacterium* microflora of the human intestine: composition and succession. *Curr Issues Intest Microbiol* **2**, 43-53.

Richards, J. D., Gong, J. & de Lange, C. F. M. (2005). The gastrointestinal microbiota and its role in monogastric nutrition and health with an emphasis on pigs: Current understanding, possible modulations, and new technologies for ecological studies. *Can J Anim Sci* **85**, 421-435.

Shepherd, M. L., Swecker Jr, W. S., Jensen, R. V. & Ponder, M. A. (2012). Characterization of the fecal bacteria communities of forage fed horses by pyrosequencing of 16S rRNA V4 gene amplicons. *FEMS Microbiol Lett* **326**, 62-68.

Shirazi-Beechey, S. P. (2008). Molecular insights into dietary induced colic in the horse. *Equine Vet J* **40**, 414-421.

Snart, J., Bibiloni, R., Grayson, T. & other authors (2006). Supplementation of the diet with high-viscosity beta-glucan results in enrichment for lactobacilli in the rat cecum. *Appl Environ Microbiol* **72**, 1925-1931.

Steelman, S. M., Chowdhary, B. P., Dowd, S., Suchodolski, J. & JaneÄ • ka, J. E. (2012). Pyrosequencing of 16S rRNA genes in fecal samples reveals high diversity of hindgut microflora in horses and potential links to chronic laminitis. *BMC Vet Res* **8**, 231.

Whelan, K. (2013). Mechanisms and effectiveness of prebiotics in modifying the gastrointestinal microbiota for the management of digestive disorders. *Proceedings of the Nutrition Society* **72**, 288-298.

Willing, B., Vörös, A., Roos, S., Jones, C., Jansson, A. & Lindberg, J. (2009). Changes in faecal bacteria associated with concentrate and forage only diets fed to horses in training. *Equine Vet J* **41**, 908-914.

Wong, J. M. W. & Jenkins, D. J. A. (2007). Carbohydrate Digestibility and Metabolic Effects. *The Journal of Nutrition* **137**, 2539S-2546S.

Yamano, H., Koike, S., Kobayashi, Y. & Hata, H. (2008). Phylogenetic analysis of hindgut microbiota in Hokkaido native horses compared to light horses. *Anim Sci J* **79**, 234-242.

Acknowledgements

First and foremost I would like to thank Prof. Paul Ross and Prof. Paul O'Toole for their guidance and support throughout my entire Ph.D. I appreciated their input, encouragement and their expertise over the years. Thank you both for being there day or night to answer all of my queries.

I would like to also extend my thanks to the people who gave me the benefit of their advice and help over the years, especially Dr. Anne Neville, Dr. Brian Forde, Hugh Harris, Dr. Marcus Claesson, Dr. Ian Jeffrey, Dr. Mary Rea, Dr. Alleson Dobson, Dr. Felicia Ciocia, Dr. Sheila Morgan, Bhuvaneswari Lakshminarayanan, Robert McCarthy, Sinead Mackey, Fiona Fuohy, Dr. Gillian Gardiner, Dr. Peadar Lawlor, Dr. Bridget Younge and Dr. Tomás Ryan.

A special thanks to Hugh Harris, without his help I am not sure I would have ever finished this thesis.

To my friends: Felicia Ciocia, Alleson Dobson, Buna Laks, Paul Joyce, Michelle Smith, Elaine Langley and Amy Watkins, thank you for always being there for me. Many thanks also to all my friends and labmates in APC One, especially Robert Kent and Daniel Burke for our many interesting and varied chats over the years.

To all the visiting students, past and present that I have had the pleasure of meeting over the years. Thank you all for the many interesting conversations we had together.

I would like to express my deepest gratitude and thanks my family especially my mother Rita for her support during my postgraduate studies. I am not sure what I would have done without her. My heartfelt thanks to both of my sisters Julie and Sarah for being there for me and always making me laugh. Many thanks also to my grandmother, aunts and uncles who helped me along the way. Also to my boyfriend Stephen for always patiently listening to me talk about my various trials and experiments and giving me his support.

Finally, I would also like to express my deepest sorrow that my late father Michael is not here to celebrate this achievement with me and the rest of my family. But I know that he was and always will be proud of me and my sisters.

Appendices – Published articles

O'Donnell *et al. Microbial Cell Factories* 2013, **12**:48
http://www.microbialcellfactories.com/content/12/1/48

MICROBIAL CELL FACTORIES

REVIEW Open Access

Catabolic flexibility of mammalian-associated lactobacilli

Michelle M O'Donnell[1,2], Paul W O'Toole[2] and Reynolds Paul Ross[1*]

Abstract

Metabolic flexibility may be generally defined as "the capacity for the organism to adapt fuel oxidation to fuel availability". The metabolic diversification strategies used by individual bacteria vary greatly from the use of novel or acquired enzymes to the use of plasmid-localised genes and transporters. In this review, we describe the ability of lactobacilli to utilise a variety of carbon sources from their current or new environments in order to grow and survive. The genus *Lactobacillus* now includes more than 150 species, many with adaptive capabilities, broad metabolic capacity and species/strain variance. They are therefore, an informative example of a cell factory capable of adapting to new niches with differing nutritional landscapes. Indeed, lactobacilli naturally colonise and grow in a wide variety of environmental niches which include the roots and foliage of plants, silage, various fermented foods and beverages, the human vagina and the mammalian gastrointestinal tract (GIT; including the mouth, stomach, small intestine and large intestine). Here we primarily describe the metabolic flexibility of some lactobacilli isolated from the mammalian gastrointestinal tract, and we also describe some of the food-associated species with a proven ability to adapt to the GIT. As examples this review concentrates on the following species - *Lb. plantarum*, *Lb. acidophilus*, *Lb. ruminis*, *Lb. salivarius*, *Lb. reuteri* and *Lb. sakei*, to highlight the diversity and inter-relationships between the catabolic nature of species within the genus.

Introduction

The human gut is an ecological niche where biotransformations of dietary ingredients occur, catalysed by gut bacteria including lactobacilli. With that in mind, this review describes, compares and summarises the catabolic machinery present in the mammalian-associated lactobacilli. Lactobacilli are well-characterised members of the Lactic Acid Bacteria (LAB) that are found throughout the gastrointestinal tract of humans and other mammals, and although generally sub dominant in the colon, can be present at proportionately high levels in the upper GIT [1].

The LAB are low G+C Gram positive bacteria and have multiple uses in the food industry. Those associated with foods include the *Lactobacillus* and *Bifidobacterium* genera [2]. Bifidobacteria are phylogenetically distant from all of the other low [G+C%]–genome LAB, but are pragmatically included in the LAB group based on their functionality and habitat [3]. In this respect, LAB are integral inhabitants of the microbiota of the gastrointestinal tract where they contribute to intestinal barrier integrity and have roles in immunomodulation and pathogen resistance [4]. This adds impetus to their inclusion in functional food products.

The growth of all living organisms is dependent on efficient cycling and recovery of energy from the environment. Carbohydrates are the primary source of carbon and energy for the growth of microorganisms [5]. Glycolysis is the most important carbohydrate metabolic cycle in the majority of bacteria and constitutes the main energy generating mechanism. In many of the commensal *Lactobacillus* species, four of the main glycolytic genes along with a regulator are encoded by the *gap* operon. Such *gap* operons have previously been reported for other Gram positive bacteria including bacilli and clostridia [6,7]. The *gap* operon in mammalian lactobacilli generally encodes the central glycolytic gene regulator (*cggR*), glyceraldehydes-3-phosphate dehydrogenase (*gap*), phosphoglycerate kinase (*pgk*), triosephosphate isomerase (*tpi*) and an enolase (*eno*). This operon arrangement was first noted in the genomes of *Lactobacillus plantarum* NC8 and *Lactobacillus sakei* Lb790 [8]. However, this particular arrangement of the *gap* operon has also since been

* Correspondence: paul.ross@teagasc.ie
[1]Teagasc Food Research Centre, Moorepark, Fermoy, Co. Cork, Ireland
Full list of author information is available at the end of the article

identified in a variety of other *Lactobacillus* species genomes [9-13], while some other genomes contain only partial operons [14-18]. The conservation of this operon arrangement (and fragments thereof) in the genomes of a number of mammalian-associated lactobacilli has a number of implications. It suggests that, through evolution and adaptation, this glycolytic operon gene arrangement has been optimised for functionality and that there is a strong selective pressure against nucleotide, gene and operon change.

The ability of lactobacilli to efficiently utilise both of the glycolytic pathways facilitates the degradation of a wider range of carbohydrates present in a given niche, but is also information relevant for their industrial exploitation. For example, *Lactobacillus reuteri* is a commensal, facultatively hetero-fermentative species able to use both the Embden-Meyerhof pathway (EMP) and the phosphoketolase pathway (PKP) to ferment carbohydrates, exemplified by *Lb. reuteri* ATCC 55730 [19]. However, examination of the genome sequences of other heterofermentative lactobacilli has also revealed genes corresponding to both glycolytic pathways [10,14]. A number of genes for enzymes involved in both glycolytic cycles were identified in the genome of *Lb. reuteri* ATCC 55730; however, no recognisable *Lactobacillus*-like *pfkA* gene could be annotated. Metabolic flux analysis identified PKP as the main glycolytic pathway with EMP acting as a shunt [19]. Of the two glycolytic pathways, PKP yields less energy production overall. However, it seems that the EMP functions to provide a net gain in ATP in conjunction with the main energy production by the PKP. It is believed that the use of PKP as the main glycolytic pathway is an adaptation of *Lb. reuteri* and other heterofermentative lactobacilli to an environment rich in carbohydrates [19]. Since *Lb. reuteri* can be used as a cell factory to produce industrially exploitable metabolic intermediates or end products such as 3-hydroxypropionaldehyde for nylons and plastics, the ability to culture lactobacilli such as *Lb. reuteri* efficiently and cost-effectively will undoubtedly be informed by knowledge of its metabolism [20].

The structure of carbohydrates and their degrees of polymerisation determine the complexity of the sugar as well as the enzymes capable of degrading them. The building blocks of the majority of complex carbohydrates metabolised by LAB are glucose, fructose, xylose and galactose, while the linkages between monosaccharide residues are what determine carbohydrate digestibility in the small intestine [21]. Related to these parameters, prebiotics are defined as "selectively fermented ingredients that allow specific changes both in the composition and/or activity in the gastrointestinal microbiota that confers benefits upon host well-being and health" [22]. The lactobacilli of the mammalian microbiota are capable of fermenting a range of carbohydrates including oligosaccharides, starch, non-starch polysaccharides and many more carbohydrates [23-26]. Many different bacterial enzymes are used in the degradation of simple and complex carbohydrates; prominent among them are the glycosyl hydrolase (EC 3.2.1) family of enzymes [27,28]. Table 1 shows a list of glycosyl hydrolases commonly identified in and utilised by lactobacilli.

In a more health conscious society, there has been a growing interest in recent years in the use of prebiotics as modulators of intestinal health [22], and prebiotics have become economically and industrially important as nutritional supplements for adults and as components in the burgeoning infant milk formula market. Lactose, soy oligosaccharides (stachyose and raffinose), lactulose and fructooligosaccharides are some of the carbohydrates that can be classed as prebiotics and that are commonly consumed as dairy, fruits and vegetables [29]. The microbiota is under constant pressure to adapt to the variety of foods consumed on a daily basis, especially in omnivores like humans. Lactobacilli present in the mammalian GIT have developed an array of adaptations to facilitate their continued presence in the human intestinal microbiota, examples of which will now be discussed. These case studies illustrate how knowledge of *Lactobacillus* metabolism is useful for optimizing their growth in the laboratory or factory, or promoting their retention in the intestinal tract by functional foods.

Carbon metabolic machinery encoded by *lactobacillus* genomes and COG assignments

In the last decade, there has been a dramatic expansion in the number of available *Lactobacillus* genome sequences from organisms isolated from a variety of environments including the mammalian GIT, dairy products and fermented foods. Based on the Integrated Microbial Genomes (IMG) website (http://img.jgi.doe.gov/cgi-bin/w/main.cgi), as of April 2013 there are 46 completed *Lactobacillus* genome sequences, comprising 18 unique species. This expansion in the number of genome sequences available has facilitated the use of comparative genomic approaches to examine the machinery involved in growth and survival of lactobacilli with unprecedented rigour.

The genome size of a *Lactobacillus* is often a determinant of the organism's capacity to metabolise a wide range of carbohydrates. Bacterial species with larger genomes are often capable of utilising a wider range of complex carbohydrates like prebiotics while those with smaller genomes are often associated with more restricted niche habitats, for example milk, and are only capable of utilising simple sugars like lactose and galactose. A comparison of the genome size and gene content for the majority of mammalian lactobacilli is shown in Table 2. *Lb. plantarum* WCFS1 has the largest genome of any

Table 1 Common glycosyl hydrolases present in mammalian lactobacilli

Enzyme	EC number	Gene	Reaction	Associated pathways	References
Alpha-amylase	3.2.1.1	amyA	Endo-hydrolysis of (1->4)-alpha-D-glucosidic linkages in polysaccharides containing three or more (1->4)-alpha-linked D-glucose units	Starch and sucrose metabolism	[9,10,14,26]
Oligo-1,6-glucosidase	3.2.1.10	malL	Hydrolysis of (1->6)-alpha-D-glucosidic linkages in some oligosaccharides produced from starch and glycogen by EC 3.2.1.1 (alpha-amylase), and in isomaltose	Starch and sucrose metabolism	[9-11,13,14,26]
Maltose-6'-phosphate glucosidase	3.2.1.122	glvA	Hydrolysis of maltose 6'-phosphate	Starch and sucrose metabolism	[18]
Alpha-glucosidase	3.2.1.20	malZ	Hydrolysis of terminal, non-reducing (1->4)-linked alpha-D -glucose residues with release of D-glucose	Galactose, starch and sucrose metabolism	[9-11,13,18,26,46]
Beta-glucosidase	3.2.1.21	bglX	Hydrolysis of terminal, non-reducing beta-D-glucosyl residues with release of beta-D-glucose	Starch and sucrose metabolism	[9-11,13,18,26]
Alpha-galactosidase	3.2.1.22	rafA	Hydrolysis of terminal, non-reducing alpha-D-galactose residues in alpha-D-galactosides, including galactose oligosaccharides, galactomannans and galactolipids	Galactose metabolism	[9-11,13,14,18,26,46]
Beta-galactosidase	3.2.1.23	lacZ	Hydrolysis of terminal non-reducing beta-D-galactose residues in beta-D-galactosides	Galactose metabolism	[9-11,14,18,26,46]
Beta-fructofuranosidase	3.2.1.26	sacA	Hydrolysis of terminal non-reducing beta-D-fructofuranoside residues in beta-D-fructofuranosides	Galactose, starch and sucrose metabolism	[9-11,13,14,18,26]
Beta-N-acetylhexosaminidase	3.2.1.52	nagZ	Hydrolysis of terminal non-reducing N-acetyl-D-hexosamine residues in N-acetyl-beta-D-hexosaminides	Amino sugar and nucleotide sugar metabolism	[9,10,14,26]
6-phospho-beta-galactosidase	3.2.1.85	lacG	Hydrolysis of 6-phospho-beta-D-galactosides	Galactose metabolism	[10,13,18]
6-phospho-beta-glucosidase	3.2.1.86	bglA	Hydrolysis of 6-phospho-beta-D-glucosyl-(1->4)-D-glucose	Glycolysis	[9-11,13,18,26]
Trehalose-6-phosphate hydrolase	3.2.1.93	treC	Hydrolysis of alpha,alpha-trehalose 6-phosphate	Starch and sucrose metabolism	[10,11,13,14,18]

Table 2 Genome statistics of various mammalian *Lactobacillus* species

Genome name	Reference	Genome size (Mb)	Gene count	GC (%)
Lb. acidophilus NCFM	[18]	1.99	1970	35
Lb. amylovorus GRL 1118	[16]	2.07	2126	38
Lb. fermentum CECT 5716	[35]	2.1	1125	51
Lb. gasseri ATCC 33323	[13]	1.9	1874	35
Lb. johnsonii FI9785	[36]	1.8	1804	34
Lb. johnsonii NCC 533	[11]	1.99	1941	35
Lb. plantarum JDM1	[37]	3.2	3026	45
Lb. plantarum WCFS1	[10]	3.35	3230	44
Lb. reuteri F275, JCM 1112	[17]	2.04	1901	39
Lb. rhamnosus GG	[12]	3.01	3016	47
Lb. rhamnosus GG, ATCC 53103	[38]	3.00	2905	47
Lb. rhamnosus Lc 705	[12]	3.03	3068	47
Lb. ruminis ATCC 25644	[9]	2.14	1901	44
Lb. ruminis ATCC 27782	[9]	2.01	2251	44
Lb. salivarius CECT 5713	[15]	2.13	1672	33
Lb. salivarius UCC118	[14]	2.13	2196	33

Lactobacillus genome sequenced to date. This organism uses the phosphoketolase pathway as a central metabolic pathway. *Lb. plantarum* has been isolated from a variety of environments including soil, vegetables, meat, dairy and from the gastrointestinal tract of humans and animals and has been used as a model *Lactobacillus* for metabolic studies [30,31]. Indeed, the genome of *Lb. plantarum* encodes a large contingent of PTS transporters, ABC transporters and glycosyl hydrolases associated with carbohydrate metabolic flexibility [10]. In contrast, *Lactobacillus gasseri* has a much smaller genome and is considered to be part of the autochthonous species present in the human gastrointestinal tract, frequently isolated from the mouth, intestines, faeces and vagina of juveniles and adults [13,32]. This homofermentative organism is unable to ferment polyols (sugar alcohols), pentoses or deoxysugars, and in this respect resembles other obligate homofermenters [33,34]. Its inability to ferment pentoses is because of the absence of two key enzymes of the pentose phosphate pathway namely transketolase and transaldolase. Absence of either or both of these enzymes results in the inability to utilise pentose sugars. This limitation is also clearly illustrated by two members of the *Lb. salivarius* clade; *Lb. salivarius* itself (heterofermentative) produces both enzymes and is capable of utilising pentoses while *Lactobacillus ruminis* (homofermentative) lacks a transaldolase gene in its genome and as a result is unable to utilise pentose sugars [14,26].

It should be emphasized, however, that examination of *Lactobacillus* genomes alone provides a limited quality of information. Functional genomics studies provide empirical experimental evidence for the functionality, mechanisms and pathways involved in carbohydrate metabolism. The fields of proteomics and transcriptomics in combination with genomics have been exploited to elucidate the mechanisms involved in carbohydrate metabolism in the host and this will be discussed in the next section.

Metabolic potential of lactobacilli – adaptation to the environment

A wide range of adaptations can potentially develop within a genus or species based on the availability of nutrients and the complexity and competition within their current environment. Adaptation to a particular environment is of great importance for survival especially in a diverse and complex milieu like the mammalian gastrointestinal tract where a wide variety of carbon sources are often present.

Lb. reuteri has previously been used as a model organism for developing and testing microbe/host symbiosis theories [39]. Along with other mammalian associated lactobacilli, *Lb. reuteri* is reliant on the fermentable carbohydrates and amino acids present in the mammalian gut digesta. However, some strains of *Lb. reuteri* also have the ability to degrade 1,2-propanediol using the cobamide-enzyme-requiring propanediol dehydratase (EC 4.2.1.28), which may constitute a primary human colonisation parameter for the species. Propanediol dehydratase is a multifunctional enzyme with roles in glycerol utilisation, glycerolipid metabolism, vitamin B_{12} biosynthesis and reuterin formation [39]. Interestingly, an enzyme with a potentially similar function has been previously identified in *Lactobacillus brevis* ATCC 367 [40]. Glycerol is used in food and beverage manufacture as a sweetener, humectant, preservative, filler, thickening agent and solvent. It has also applications in the manufacture of mono/di-glycerides and poly-glycerol for margarine production. Therefore, glycerol can form a significant part of the foods consumed daily, particularly in the western world. The capability to hydrolyse glycerol may provide lactobacilli a competitive advantage in the gastrointestinal tract.

Some *Lactobacillus* species utilise differentially present or differentially expressed features of their carbohydrate metabolic machinery in order to facilitate their colonisation and persistence in the mammalian gut. For example, *Lactobacillus johnsonii* and *Lb. reuteri* do not compete in the mouse fore-stomach because the former utilizes glucose and the latter maltose, even though both species have the genes for metabolizing both substrates [41]. This is an example of niche sharing by way of resource partitioning. Using a mouse model Denou *et al*., 2008 showed that *Lb. johnsonii* strains use a number of genes (carbohydrate utilisation genes included) for long-term gut persistence. Correlating the datasets from the genomic hybridisation of two strains (ATCC 33200 and NCC533) and the *in vivo* microarray transcription data from strain NCC533 identified six genes, forming three loci that are *Lb. johnsonii* NCC533 strain specific. Two of the loci are involved in carbohydrate metabolism namely exopolysaccharide biosynthesis (glycosyltransferases) and a mannose phosphoenolpyruvate phosphotransferase system PTS (transporter) [42].

A similar transcriptomic study, focusing on the adaptations of *Lb. plantarum,* demonstrated the capacity of a *Lactobacillus* to alter its metabolism in response to the human or murine intestine [43,44]. In those studies, a number of genes required for carbohydrate metabolism were identified as differentially transcribed in the human and mouse gastrointestinal tract under different dietary conditions. The genes up-regulated included those encoding glycosyl hydrolases, glycolytic enzymes and various carbohydrate transporter classes [43,44]. An overlap in the enzymes induced in the mammalian GIT included those involved in the degradation and transport of lactose and the plant derived-disaccharides melibiose, cellobiose and maltose. In animals fed a Western diet there was also a noteworthy up-regulation of glycerol metabolism-related enzymes, which relates to the presence of glycerol in many

foods discussed above. The induction of carbohydrate metabolism genes highlights the importance of metabolic flexibility in the adaptation of *Lactobacillus* and other bacteria to the human and mammalian intestine [43,44].

Metabonomic studies using Nuclear Magnetic Resonance (NMR) spectroscopy have identified the metabolites most affected by supplementation of the human diet with fructooligosaccharides (FOS) and *Lactobacillus acidophilus* and *Bifidobacterium longum* based synbiotics [45]. Beneficial short chain fatty acids (SCFA) namely propionate and butyrate were identified in faeces of individuals receiving the synbiotic treatments. There was also a marked decrease in the recoverable amino acids in the samples. The increase in *lactobacillus* numbers over the month-long period as well as the increase in SCFA levels and decrease in amino acid concentrations indicate that the feeding of a synbiotic resulted in a shift of the intestinal metabolome from an overall proteolytic pattern to a saccharolytic one. The presence of FOS in the diet, which is indigestible in the upper GIT, had the ability to affect the SCFA profile of the lower GIT when fermented by bacterial species like lactobacilli and bifidobacteria [45].

Another recent study focussed on the adaptation by *Lb. reuteri* to the GIT of mice [46]. *In vivo* studies using *Lactobacillus*-free (LF) mice and different vertebrate-derived *Lb. reuteri* isolates established that only the rodent isolates were capable of reaching colonising numbers in the LF mice, supporting the theory of host specialisation. Using comparative genome hybridisation, the genome of an *Lb. reuteri* mouse isolate was compared to that of 24 other *Lb. reuteri* strains from various sources. A xylose utilisation operon was conserved in the strains of rodent and porcine origin [46] but absent in the others. Xylose forms a large percentage of the hemi-cellulose found in some plants and so is consumed as part of animal diet.

Other examples of niche-specific genes or host specialisation genes between dairy and gastrointestinal lactobacilli have also been revealed using comparative genomic approaches. For example, mannose-6-phosphate glucosidase (EC 3.2.1.122), a mannose catabolic enzyme, was identified as a solely gut-specific gene in the genome sequences of a number of frequently present mammalian lactobacilli [11,12,16,18,37,47]. This enzyme works in conjunction with a maltose phosphotransferase system to import phosphorylated maltose into the cell. Once internalised the enzyme converts maltose-6-phosphate into glucose and glucose-6-phosphate, and it is this method of transport and degradation that is thought to be specific to strains of gut origin. However, this mechanism of maltose utilisation is not ubiquitous among the gut lactobacilli [9,10,13,14,47]. Genome decay, due to gene loss, seems to operate in the dairy lactobacilli that have higher numbers of pseudogenes in their genomes than other lactobacilli. The majority of the pseudo-genes present are related to carbon catabolism, amino acid metabolism and transport, reflecting the fact that these organisms (for example *Lactobacillus helveticus* [48]) have less need for these processes in a milk environment. However, it must be noted that even for an organism like *Lb. plantarum* with a diverse range of habitats, continual passage in a nutrient rich medium can lead to genome contraction and loss of certain types of carbohydrate transporters and enzymes [37]. A genome level comparison of *Lb. plantarum* JDM1 with *Lb. plantarum* WCFS1 revealed that certain saccharolytic genes and transporters present in strain WCFS1 were absent in the closely related strain JDM1 [10,37]. Examples of the absent enzymes include alpha-amylase, alpha-L-rhamnosidase, beta-N-acetylhexosaminidase, mannosyl-glycoprotein, endo-beta-N-acetylglucosaminidase and glucan 1,4-alpha-maltohydrolase [37]. This variability of saccharolytic capability within a species is also clearly illustrated by the work of Molenaar *et al.*, 2005 who compared over 20 *Lb. plantarum* species using microarray genotyping technology [49]. These were clear examples of a species adapting to their environment and altering their metabolic profile to suit the new environment either by gene acquisition or in this case gene loss.

Recent studies have also focussed on the cellular response of certain lactobacilli to complex carbohydrates. For example, Majumder and colleagues identified a number of proteins involved in the adaptation of *Lactobacillus acidophilus* NCFM to growth in the presence of the prebiotic lactitol (a synthetic sugar alcohol derived from lactose, used in the food industry and in some medications) [50]. Examination of the late exponential phase whole-cell extract proteome revealed a number of proteins present which may be involved in utilization of lactitol including a β-galactosidase subunit, galactokinase and other galactose utilisation proteins. The majority of enzymes identified in lactitol utilisation were the same enzymes involved in the Leloir pathway (the lactose utilisation pathway) and transportation of lactitol into the cell was facilitated by LacS (a glycoside-pentoside-hexuronide cation symporter). While transport of lactitol is facilitated by a permease, it is the phosphotransferase system that transports and metabolises sorbitol [50]. *Lb. reuteri* (as well as the other mammalian lactobacilli) also possess the genetic determinants for enzymes associated with the utilisation of raffinose family oligosaccharides (RFO). RFOs are present in many vegetables namely legumes and are associated with flatulence and gastrointestinal upset [51]. Alpha galactosidase (EC 3.2.1.20) and to a lesser extent levansucrase (EC 2.4.1.10) are the main enzymes commonly encoded in the genome sequences of mammalian derived lactobacilli, which are responsible for the hydrolysis and partial hydrolysis of RFO, respectively [10-14,16,18,26,35,52]. Interestingly, the genome sequences of dairy lactobacilli such

as *Lactobacillus bulgaricus* and *Lb. helveticus* [48,53] are devoid of RFO degradation associated enzymes, consistent with the fact that milk generally contains negligible amounts of RFO.

Dairy derived lactobacilli, however, can possess considerable and demonstrable metabolic flexibility. Burns *et al.*, 2010 investigated the "progressive adaptation" of dairy *Lactobacillus delbrueckii* strains to bile (a bio-surfactant produced in the liver for emulsifying fats in the diet). The proteomes of *Lb. delbrueckii* and an enhanced bile resistant derivative were examined using cells grown in the presence and absence of bile. A total of 35 proteins were affected by the inclusion of bile. Three of the proteins were found to be part of the glycolytic cycle with phosphoglycerate mutase (*pgm*) and glyceraldehyde-3P-dehydrogenase genes up-regulated, while fructose-bisphosphate aldolase was down-regulated at the protein level [54]. *Lactobacillus casei*, a predominantly dairy associated isolate, is frequently isolated from a range of other niches, including plants, and the human GIT [55,56]. Examination of the *Lb. casei* strain fermentation profiles from these various niches identified several trends, for example the increased utilisation of polyols by strains of plant and human origin. Not surprisingly, strains of cheese origin also were found to have an increased capacity for lactose utilisation when compared to non-dairy isolates. The data suggest that *Lb. casei* can adjust its metabolic capabilities in order to adapt to the carbon sources available in a particular niche.

Lactobacilli also have the capacity to alter their metabolism to adapt to a new environment. This is clearly exemplified by a study of *Lb. sakei* where Chiaramonte and colleagues (2010) showed that the meat-borne *Lactobacillus sakei* is capable of colonizing the GIT of mice [57]. Analysis of *Lb. sakei* wild-type and morphological mutants revealed an increased capacity for the utilisation of some carbon sources (fructose, ribose and galactose) when compared to the original meat-borne parent strain. Up-regulation of the genes encoding 6-phosphofructokinase, L-lactate dehydrogenase and fructose-bisphosphate aldolase was considered to be the likely cause of this capacity to colonize the mouse GIT. Two genes involved in nucleotide metabolism, CTP synthase and xanthine phosphoribosyltransferase were also up-regulated in the mutants derived from the passage of meat-borne *Lb. sakei* strain through the GIT of axenic mice [57].

Transporters and their importance in metabolic flexibility and regulation of metabolism

Carbohydrate transporters or permeases are an essential component in carbohydrate metabolism to facilitate permeability of the cell to carbon metabolites, and may be the rate limiting step in their utilization [58]. Transporters involved in carbohydrate metabolism include proton coupled active transport and group translocators [59]. A summary of those systems most commonly found in lactobacilli is presented in Table 3.

Within the LAB, the ATP binding cassette (ABC) transporters form the largest group [60], whereby a metabolite or macromolecule is transported using energy derived

Table 3 Common carbohydrate transporters utilised by mammalian lactobacilli

Superfamily	Transport family	Transporter class	Transporter subclass	Transport classification system	Transmembrane domain range
MFS	Major Facilitator Superfamily (MFS)	Electrochemical Potential-driven Transporters	Porters (uniporters, symporters, antiporters)	TC 2.A.1	12-24
GPH	Glycoside-Pentoside-Hexuronide (GPH):Cation Symporter Family	Electrochemical Potential-driven Transporters	Porters (uniporters, symporters, antiporters)	TC 2.A.2	12
ATP Binding Cassette	ATP-binding Cassette (ABC)	Primary Active Transporters	P-P-bond-hydrolysis-driven transporters	TC 3.A.1	5-6
PTS-GFL	PTS Glucose-Glucoside (Glc) Family	Group Translocators	Phosphotransfer-driven Group Translocators	TC 4.A.1	8
PTS-GFL	PTS Fructose-Mannitol (Fru) Family	Group Translocators	Phosphotransfer-driven Group Translocators	TC 4.A.2	8
PTS-GFL	PTS Lactose-N,N'-Diacetylchitobiose-β-glucoside (Lac) Family	Group Translocators	Phosphotransfer-driven Group Translocators	TC 4.A.3	8
PTS-GFL	PTS Glucitol (Gut) Family	Group Translocators	Phosphotransfer-driven Group Translocators	TC 4.A.4	8
PTS-GFL	PTS Galactitol (Gat) Family	Group Translocators	Phosphotransfer-driven Group Translocators	TC 4.A.5	8
PTS-GFL	PTS Mannose-Fructose-Sorbose (Man) Family	Group Translocators	Phosphotransfer-driven Group Translocators	TC 4.A.6	8
PTS-GFL	PTS L-Ascorbate (L-Asc) Family	Group Translocators	Phosphotransfer-driven Group Translocators	TC 4.A.7	8

from ATP hydrolysis [61]. ABC transporters are capable of transporting mono, di, tri, poly and oligosaccharide as well as polyols [62]. ABC transporters encoded by the genome sequences of mammalian lactobacilli include those for maltose, lactose, arabinose, sorbitol, mannitol, glucose, N-acetylglucosamine and cellobiose transport together with ribose, xylose, fructose and rhamnose, all of which are commonly found in the mammalian digesta, especially of omnivores [9,10,13,14,18]. However, genomes from strains of dairy and meat origin so far examined harbour only gene fragments of carbon-transport-related ABC transporters and do not therefore encode a complete transporter protein [53,63].

Transporters that use chemo-osmosis in order to import carbohydrates are called secondary active transporters and are categorised as either uni-porters, symporters or anti-porters [64]. The majority of uni/sym/anti-porters are part of a large group called the Major Facilitator Superfamily (MFS) with over 40 recognised MFS families [65]. MFS transporters are capable of transporting the majority of micro-molecules (like low DP carbohydrates) but are unable to transport macromolecules. Glycoside-pentoside -hexuronide (GPH) transporters are a class of sodium ion symporters that are used by both homo and heterofermentative lactobacilli to transport carbohydrates [10,23,26,43,66]. Lactobacilli found exclusively in the gastrointestinal tract, for instance *Lb. ruminis*, have been found to harbour a lower number of complete PTS transporters but a higher number of symporters otherwise known as secondary active transporters [26]. In contrast, *Lb. gasseri*, another autochthonous species in the human gut, encodes two glucose permeases but does not encode a lactose/galactose permease [13]. The reliance of some lactobacilli on symporters may be due in part to the fact that the gastrointestinal tract is a nutrient-rich, complex environment. Thus the cells do not have to expend as much energy in order to internalize carbohydrates; instead a carbohydrate is transported into the cell using simultaneous sodium ion exchange. Often the sugars found in the GIT are of a high degree of polymerisation like inulin and starches which require alternate transportation methods to the PTS system.

The majority of carbohydrate transport in lactobacilli isolated from a variety of environments, for example *Lb. plantarum* and *Lb. acidophilus*, is done using PTS systems [10,18]. This method of transport involves the coupling of energy molecules with phosphorylation, to bring the phosphorylated carbohydrates into the cell, and is of particular importance in the transport of low complexity hexose sugars [67]. PTS transporters are characterised by a phosphate transfer cascade involving phosphoenolpyruvate (PEP), enzyme I (EI), histidine protein (HPr) and various EIIABC's. HPr is phosphorylated at site serine 46 by HPrK/P which is only present in the low [G+C%] Gram positives [68]. PEP-dependent phosphorylation of HPr by EI yields HPr-His-P, which is required for PTS-mediated transport of carbon sources [69].

Many mammalian lactobacilli rely on the PEP-PTS to facilitate nutrient uptake in the gastrointestinal tract and contain a number of PTS classes. This is best exemplified by *Lb. plantarum* and members of the acidophilus complex [11,13,18]. The *Lb. plantarum* WCFS1 genome encodes 25 predicted complete PTS EII complexes; it also encodes some incomplete complexes [10]. This high number of PTS genes is one of the largest counts in a sequenced microbial genome and currently comes second only to *Listeria monocytogenes* [70]. The genome of *Lb. acidophilus* NCFM encodes 20 PEP-PTS; the transporters have predicted specificity for trehalose, fructose, sucrose, glucose, mannose, melibiose, gentiobiose, cellobiose, salicin, arbutin and N-acetylglucosamine PTS [18]. The genome of *Lb. gasseri* ATCC 33323, another acidophilus complex bacterium, encodes 21 PEP-PTS transporters including those for predicted transport of fructose, mannose, glucose, cellobiose, lactose, sucrose, trehalose, β-glucosides and N-acetylglucosamine [13]. The genome of *Lb. johnsonii* NCC 533 encodes 16 PEP-PTS which is a large number for a genome of its size; allowing the predicted transport of sugars such as mannose, melibiose, cellobiose, raffinose, N-acetylglucosamine, trehalose and sucrose ,which is supported experimentally by physiological (API CH50, Biomerieux, France) data [11].

As mentioned above, bacterial species will often preferentially utilise one carbohydrate prior to utilising another by means of the phosphotransferase system. This system requires strict regulation to ensure the ability to preferentially utilise the particular carbohydrate, for example glucose, before any other carbon source. This type of control is called carbon catabolite repression (CCR). CCR is defined as "a regulatory phenomenon by which the expression of functions for the use of secondary carbon sources and the activities of the corresponding enzymes are reduced in the presence of a preferred carbon source" [71]. Various methods of CCR are present in nearly all free living microorganisms. In phylum Firmicutes, the main components are catabolite control protein A (CcpA), HPr, HPr kinase/phosphorylase (HPrK) and the glycolytic enzymes fructose 1,6-bisphosphate and glucose-6-phosphate. In *Enterobacteriaceae* the phosphorylation state of EIIA is crucial for CCR, whereas in Firmicutes the phosphorylation state of HPr is essential [72]. HPr phosphorylation can occur at two sites, at Histidine-15 by EI and at Serine-46 by HPrK. In the presence of glucose, there is an increase in the level of fructose 1,6-bisphosphate which indicates a high level of glycolytic activity. HPrK kinase activity is triggered by this increase causing phosphorylated HPr to bind to CcpA, which then binds to the cre site on the DNA thereby repressing transcription of the catabolic

genes. When glucose levels are low there is a decreased level of Fructose 1,6-bisphosphate, which dephosphorylates HPrK/P at Ser-46 [73,74]. The outcome from CCR is the same with the preferential use of a carbon source.

Regulation of carbohydrate metabolism (especially lactose) has also been identified in *Lb. acidophilus* NCFM [50]. In the presence of lactose there was an increase in the abundance of pyruvate kinase, a noted indicator of regulation via carbon catabolite repression, and the down regulation of genes for nucleotide metabolism proteins [50]. A similar phenomenon was noted in the proteome of *Lactococcus lactis* when grown in the presence of lactose as a carbon source [75]. Similarly, in *Lb. plantarum* CCR has been shown to control the expression of phospho-β-glucosidase [76]. Lactobacilli like *Lb. brevis* and *Lb. pentosus* which have relaxed control of their carbon catabolite machinery are being investigated for their carbon degradation potential for industry [77,78]. This alternative or relaxed mechanism of carbon catabolite control is being used in industrial fermentations of cellulolytic and ligno-celluloytic materials to form lactic acid and ethanol, respectively [77,78]. The use of lactobacilli that are capable of using mixed carbohydrate sources for growth is of great importance for industries utilising lignocellulose hydrolysate-like biomass containing hexose and pentose sugars like glucose, arabinose and xylose.

Horizontal gene transfer and plasmid-encoded carbon metabolism genes

Horizontal gene transfer (HGT) has long been recognised as a method by which bacteria receive genes and other genetic elements conferring new abilities from another species, for example *Escherichia coli* transferring ampicillin resistance to *Shigella flexneri* [79]. Mobile genetic elements include transposons, bacteriophages and plasmids [80]. While examining the genomes of two species of GIT-associated lactobacilli and a dairy isolate in particular (*Lb. delbrueckii* ssp. *bulgaricus*), it was noted that extensive horizontal gene transfer (HGT) had occurred between the three species [81]. Comparison of phylogenetic trees for over four hundred proteins highlighted the variance between the members of the acidophilus complex. In many cases, the acquisition of new genetic capabilities can include a new method of solute transportation. Mannose PTS transporters are a class of PTS transporters (TC 4.A.6) associated with mammalian associated *Lactobacillus* species with the exception of *Lb. reuteri* [17]. Comparison of phylogenetic trees created from the ClustalW alignment of mannose PTS transporters from twenty five bacteria including *Lb. plantarum*, highlighted the likelihood of HGT having occurred [82]. The study identified the lack of concordance between evolutionary data from 16S ribosomal RNA gene sequences and the evolutionary data generated from the mannose PTS sequences. The analysis also noted that within the mannose transporters in particular, there was a high level of sequence variation among the bacteria studied. Sequence analysis and comparison of the 58 mannose PTS proteins identified the varying patterns caused by HGT and allowed organising the species into six groups [82].

A plasmid is defined as "a linear or circular double-stranded DNA that is capable of replicating independently of the chromosomal DNA". Plasmids are very common within the *Lactobacillus* genus with approximately 38% of all species containing one or more plasmids of varying sizes [83], including most of the species routinely used for industrial applications. Regions of homology have been identified in plasmids from the same species, genus and from other genera [84]. Plasmids contribute to horizontal gene transfer, with plasmids often containing genes for carbohydrate, citrate and amino acid utilisation, production of bacteriocins or other biosynthetic genes [83]. This is best exemplified by *Lb. salivarius* UCC118 which contains 2 cryptic plasmids and one megaplasmid [85]. The megaplasmid (pMP118) harbours genes for the utilisation of pentoses and polyols. It also carries genes involved in glycolysis (FBP) and genes for two pentose pathway essential enzymes, transketolase and transaldolase. The plasmid pMP118 encodes an additional copy of the enzyme ribose-5-phosphate isomerase which may contribute to its metabolic flexibility and adaptive capabilities. Thus, for *Lb. salivarius* to survive in an environment dominated by pentose sugars these plasmid acquired genes would be essential [14,85]. However, the most striking example in the mammalian derived lactobacilli of the importance of plasmids in carbohydrate metabolism is the case of the *Lactobacillus rhamnosus* Lc705 plasmid pLC1 [12]. This 64 Kbp plasmid sequence encodes proteins predicted for the fructose PTS, glucose uptake proteins, a glycosyl hydrolase and a number of genes involved in alpha and beta-galactoside utilisation and transport [12]. It is obvious that without the presence of these plasmid-borne genes, *Lb. rhamnosus* Lc705 would be at a severe competitive disadvantage in the mammalian GIT compared to other *Lactobacillus* species that have these genes integrated in the chromosome. The presence of these genes in the plasmid presumably allows *Lb. rhamnosus* to compete for the alpha galactosides and fructose from plant sources and also for the beta-galactosides from dairy products. It is clear from the available plasmid sequences that, while not always present, carbohydrate genes carried by plasmids are important mobile genetic elements for lactobacilli.

The presence of carbohydrate metabolic genes located on plasmids is also common in food, plant and dairy lactobacilli. Another example of plasmid encoded pentose sugar utilisation genes is the xylose utilisation cluster present in plasmids isolated from *Lactobacillus pentosus*

[86], a plant derived *Lactobacillus*. A study comparing 34 sequenced *Lactobacillus* plasmids revealed that the carbohydrate and amino acid transport category was that most frequently encoded among the plasmids analysed [87]. The presence of a larger cohort of carbohydrate and amino acid transporters is possibly a niche adaptation. *Lb. casei* 64H lacking the plasmid pLZ64, which contains a lactose PEP-PTS and phospho-β-galactosidase, is unable to utilize lactose. There is limited knowledge on the true extent of plasmids from mammalian derived lactobacilli and their impact on gut health. However, there is detailed knowledge on the presence and function of plasmids in dairy-derived lactobacilli for example *Lb. casei* [88].

Conclusions

Carbon metabolism is essential for life and the survival of many bacterial species depends on their ability to exert some degree of metabolic flexibility. *Lactobacillus* as a genus, has a broad range of environmental niches and is equipped with an intricate array of enzymatic systems and adaptive responses to cope with differing carbohydrate sources. This poses challenges for examining the effect of lactobacilli on the gut microbiota but also opportunities for their efficient industrial exploitation. Although there is an extensive amount of information on the *in vitro* and *in silico* catabolic flexibility of mammalian lactobacilli, additional studies and investigations are required to elucidate all the factors and systems that are involved in carbohydrate degradation mechanisms *in vivo* in the mammalian GIT. Further metabolomic, metabonomic and metatranscriptomic studies along with concerted effort are needed to fully elucidate all of the effects that carbohydrate metabolism has on strain phenotypes. With advances in sequencing technologies it is now possible and "affordable" to use RNA-seq (whole transcriptome shotgun sequencing) rather than using microarrays. Microarrays have shortcomings that including for example requiring prior sequence information of a strain, and the need to use of pure cultures which makes it difficult to assess the effect of species or carbohydrate on the microbiome as a system of interconnected genera and species. Metatranscriptomics can identify the gene expression of mixed communities of organisms *in vivo* under a wide range of parameters including diet, stresses, disease state and other environmental and health factors. The use of metatranscriptomics in conjunction with animal model feeding studies would allow a more accurate measurement of the effect diet has on the *Lactobacillus* component of the microbiota. For *in vivo* studies the use of a "standard" mammalian GIT model, for example the pig, whose physiology is similar to that of humans would be advantageous in allowing more rigorous comparisons of *in vivo* feeding studies. The use of mouse models, while convenient and relatively inexpensive, should be viewed as a "small-scale" step before transitioning the research into a larger human GIT analogue model like the pig. Further investigations using some of the techniques outline above on a wider number of mammalian derived lactobacilli will provide information that will lead to a greater understanding of *in vivo* carbohydrate metabolism of mammalian derived lactobacilli and the implications for human and animal health. The industrial usage of lactobacilli for production of metabolites and process ingredients will benefit from progress in metabolic modelling, exemplified to date by *Lb. plantarum* WCFS1 [89], but not yet applied to many relevant lactobacillus species. Success of these modelling experiments will be aided by empirical data provided by complementary "omics" analyses, generating greater precision in establishing and fine-tuning models for lactobacillus growth in the laboratory and in the factory.

Competing interest
The authors declare that they have no competing interests.

Authors' contributions
MMOD drafted the manuscript and participated in the title conception, PWOT and RPR conceived of the title, and helped to draft the manuscript. All authors read and approved the final manuscript.

Author details
[1]Teagasc Food Research Centre, Moorepark, Fermoy, Co. Cork, Ireland. [2]Microbiology Department, University College Cork, College Road, Co. Cork, Ireland.

Received: 14 February 2013 Accepted: 8 May 2013
Published: 16 May 2013

References

1. Holzapfel WHN, Wood BJ: **The Lactic Acid Bacteria: the genera of lactic acid bacteria**. In *The Lactic Acid Bacteria: The genera of lactic acid bacteria*, Volume 2. Springer; 1995:420.
2. Stiles ME, Holzapfel WH: **Lactic acid bacteria of foods and their current taxonomy.** *Int J Food Microbiol* 1997, **36**(1):1–29.
3. Vaughan EE, de Vries MC, Zoetendal EG, Ben-Amor K, Akkermans AD, de Vos WM: **The intestinal LABs.** *A Van Leeuw* 2002, **82**(1–4):341–352.
4. Stiles ME: **Biopreservation by lactic acid bacteria.** *A Van Leeuw* 1996, **70**(2):331–345.
5. Gunsalus IC, Horecker BL, Wood WA: **Pathways of carbohydrate metabolism in microorganisms.** *Microbiol Mol Biol Rev* 1955, **19**(2):79–128.
6. Ludwig H, Homuth G, Schmalisch M, Dyka FM, Hecker M, Stülke J: **Transcription of glycolytic genes and operons in *Bacillus subtilis*: evidence for the presence of multiple levels of control of the gapA operon.** *Mol Microbiol* 2001, **41**(2):409–422.
7. Schreiber W, Dürre P: **Differential Expression of Genes Within the gap Operon of *Clostridium acetobutylicum*.** *Anaerobe* 2000, **6**(5):291–297.
8. Naterstad K, Rud I, Kvam I, Axelsson L: **Characterisation of the gap operon from *Lactobacillus plantarum* and *Lactobacillus sakei*.** *Curr Microbiol* 2007, **54**(3):180–185.
9. Forde BM, Neville BA, O' Donnell MM, Riboulet-Bisson E, Claesson MJ, Coghlan A, Ross RP, O' Toole PW: **Genome sequences and comparative genomics of two *Lactobacillus ruminis* strains from the bovine and human intestinal tracts.** *Microb Cell Fact* 2011, **10**(1):S13.
10. Kleerebezem M, Boekhorst J, van Kranenburg R, Molenaar D, Kuipers OP, Leer R, Tarchini R, Peters SA, Sandbrink HM, Fiers MW, *et al*: **Complete genome sequence of *Lactobacillus plantarum* WCFS1.** *Proc Natl Acad Sci USA* 2003, **100**(4):1990–1995.
11. Pridmore RD, Berger B, Desiere F, Vilanova D, Barretto C, Pittet AC, Zwahlen MC, Rouvet M, Altermann E, Barrangou R, *et al*: **The genome sequence of the probiotic intestinal bacterium *Lactobacillus johnsonii* NCC 533.** *Proc Natl Acad Sci USA* 2004, **101**(8):2512–2517.

12. Kankainen M, Paulin L, Tynkkynen S, von Ossowski I, Reunanen J, Partanen P, Satokari R, Vesterlund S, Hendrickx APA, Lebeer S, *et al*: **Comparative genomic analysis of *Lactobacillus rhamnosus* GG reveals pili containing a human- mucus binding protein.** *Proc Natl Acad Sci USA* 2009, **106**(40): 17193–17198.
13. Azcarate-Peril MA, Altermann E, Goh YJ, Tallon R, Sanozky-Dawes RB, Pfeiler EA, O'Flaherty S, Buck BL, Dobson A, Duong T, *et al*: **Analysis of the genome sequence of *Lactobacillus gasseri* ATCC 33323 reveals the molecular basis of an autochthonous intestinal organism.** *Appl Environ Microbiol* 2008, **74**(15):4610–4625.
14. Claesson MJ, Li Y, Leahy S, Canchaya C, van Pijkeren JP, Cerdeno-Tarraga AM, Parkhill J, Flynn S, O'Sullivan GC, Collins JK, *et al*: **Multireplicon genome architecture of *Lactobacillus salivarius*.** *Proc Natl Acad Sci USA* 2006, **103**(17):6718–6723.
15. Jiménez E, Martín R, Maldonado A, Martín V, Fernández L, Rodríguez JM, Gómez de Segura A: **Complete Genome Sequence of *Lactobacillus salivarius* CECT 5713, a Probiotic Strain Isolated from Human Milk and Infant Feces.** *J Bacteriol* 2010, **192**(19):5266–5267.
16. Kant R, Paulin L, Alatalo E, de Vos WM, Palva A: **Genome Sequence of *Lactobacillus amylovorus* GRL1112.** *J Bacteriol* 2011, **193**(3):789–790.
17. Morita H, Toh H, Fukuda S, Horikawa H, Oshima K, Suzuki T, Murakami M, Hisamatsu S, Kato Y, Takizawa T, *et al*: **Comparative Genome Analysis of *Lactobacillus reuteri* and *Lactobacillus fermentum* Reveal a Genomic Island for Reuterin and Cobalamin Production.** *DNA Res* 2008, **15**(3):151–161.
18. Altermann E, Russell WM, Azcarate-Peril MA, Barrangou R, Buck BL, McAuliffe O, Souther N, Dobson A, Duong T, Callanan M, *et al*: **Complete genome sequence of the probiotic lactic acid bacterium *Lactobacillus acidophilus* NCFM.** *Proc Natl Acad Sci USA* 2005, **102**(11):3906–3912.
19. Årsköld E, Lohmeier-Vogel E, Cao R, Roos S, Rådström P, van Niel EWJ: **Phosphoketolase Pathway Dominates in *Lactobacillus reuteri* ATCC 55730 Containing Dual Pathways for Glycolysis.** *J Bacteriol* 2008, **190**(1):206–212.
20. Vollenweider S, Grassi G, König I, Puhan Z: **Purification and Structural Characterization of 3-Hydroxypropionaldehyde and Its Derivatives.** *J Agr Food Chem* 2003, **51**(11):3287–3293.
21. Manning TS, Gibson GR: **Microbial-gut interactions in health and disease. Prebiotics.** *Best Pract Res Clin Gastroenterol* 2004, **18**(2):287–298.
22. Gibson GR, Probert HM, Loo JV, Rastall RA, Roberfroid MB: **Dietary modulation of the human colonic microbiota: updating the concept of prebiotics.** *Nutr Res Rev* 2004, **17**(2):259–275.
23. Barrangou R, Azcarate-Peril MA, Duong T, Conners SB, Kelly RM, Klaenhammer TR: **Global analysis of carbohydrate utilization by *Lactobacillus acidophilus* using cDNA microarrays.** *Proc Natl Acad Sci USA* 2006, **103**(10):3816–3821.
24. Saulnier DM, Molenaar D, de Vos WM, Gibson GR, Kolida S: **Identification of prebiotic fructooligosaccharide metabolism in *Lactobacillus plantarum* WCFS1 through microarrays.** *Appl Environ Microbiol* 2007, **73**(6):1753–1765.
25. Barrangou R, Altermann E, Hutkins R, Cano R, Klaenhammer TR: **Functional and comparative genomic analyses of an operon involved in fructooligosaccharide utilization by *Lactobacillus acidophilus*.** *Proc Natl Acad Sci USA* 2003, **100**(15):8957–8962.
26. O'Donnell MM, Forde BM, Neville BA, Ross RP, O'Toole PW: **Carbohydrate catabolic flexibility in the mammalian intestinal commensal *Lactobacillus ruminis* revealed by fermentation studies aligned to genome annotations.** *Microb Cell Fact* 2011, **10**(Suppl 1):S12.
27. Henrissat B: **A classification of glycosyl hydrolases based on amino acid sequence similarities.** *Biochem J* 1991, **280:**309–316. 0264–6021 (Print)).
28. Henrissat B, Bairoch A: **Updating the sequence-based classification of glycosyl hydrolases.** *Biochem J* 1996, **316**(Pt 2):695–696.
29. Gibson GR, Roberfroid MB: **Dietary modulation of the human colonic microbiota: Introducing the concept of prebiotics.** *J Nutr* 1995, **125**(6): 1401–1412.
30. Siezen RJ, Tzeneva VA, Castioni A, Wels M, Phan HTK, Rademaker JLW, Starrenburg MJC, Kleerebezem M, Molenaar D, Van Hylckama Vlieg JET: **Phenotypic and genomic diversity of *Lactobacillus plantarum* strains isolated from various environmental niches.** *Environ Microbiol* 2010, **12**(3):758–773.
31. Siezen R, van Hylckama VJ: **Genomic diversity and versatility of *Lactobacillus plantarum*, a natural metabolic engineer.** *Microb Cell Fact* 2011, **10**(Suppl 1):S3.
32. Reuter G: **The *Lactobacillus* and *Bifidobacterium* microflora of the human intestine: composition and succession.** *Curr Issues Intest Microbiol* 2001, **2**(2):43–53.
33. Felis GE, Dellaglio F: **Taxonomy of Lactobacilli and Bifidobacteria.** *Curr Issues Intest Microbiol* 2007, **8**(2):44–61.
34. Kandler O: **Carbohydrate metabolism in lactic acid bacteria.** *A Van Leeuw* 1983, **49**(3):209–224.
35. Jiménez E, Langa S, Martín V, Arroyo R, Martín R, Fernández L, Rodríguez JM: **Complete Genome Sequence of *Lactobacillus fermentum* CECT 5716, a Probiotic Strain Isolated from Human Milk.** *J Bacteriol* 2010, **192**(18):4800.
36. Wegmann U, Overweg K, Horn N, Goesmann A, Narbad A, Gasson MJ, Shearman C: **Complete Genome Sequence of *Lactobacillus johnsonii* FI9785, a Competitive Exclusion Agent against Pathogens in Poultry.** *J Bacteriol* 2009, **191**(22):7142–7143.
37. Zhang Z, Liu C, Zhu Y, Zhong Y, Zhu Y, Zheng H, Zhao G-P, Wang S-Y, Guo X: **Complete genome sequence of *Lactobacillus plantarum* JDM1.** *J Bacteriol* 2009, **191**(15):5020–5021.
38. Morita H, Toh H, Oshima K, Murakami M, Taylor TD, Igimi S, Hattori M: **Complete Genome Sequence of the Probiotic *Lactobacillus rhamnosus* ATCC 53103.** *J Bacteriol* 2009, **191**(24):7630–7631.
39. Walter J, Britton RA, Roos S: **Host-microbial symbiosis in the vertebrate gastrointestinal tract and the *Lactobacillus reuteri* paradigm.** *Proc Natl Acad Sci USA* 2010, **108**(suppl. 1):4645–4652.
40. Makarova K, Slesarev A, Wolf Y, Sorokin A, Mirkin B, Koonin E, Pavlov A, Pavlova N, Karamychev V, Polouchine N, *et al*: **Comparative genomics of the lactic acid bacteria.** *Proc Natl Acad Sci USA* 2006, **103**(42):15611–15616.
41. Tannock GW, Wilson CM, Loach D, Cook GM, Eason J, O'Toole PW, Holtrop G, Lawley B: **Resource partitioning in relation to cohabitation of *Lactobacillus* species in the mouse forestomach.** *ISME J* 2012, **6**(5):927–938.
42. Denou E, Pridmore RD, Berger B, Panoff J-M, Arigoni F, Brüssow H: **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* 2008, **190**(9):3161–3168.
43. Marco ML, de Vries MC, Wels M, Molenaar D, Mangell P, Ahrne S, de Vos WM, Vaughan EE, Kleerebezem M: **Convergence in probiotic *Lactobacillus* gut-adaptive responses in humans and mice.** *ISME J* 2010, **4**(11):1481–1484.
44. Marco ML, Peters THF, Bongers RS, Molenaar D, Van Hemert S, Sonnenburg JL, Gordon JI, Kleerebezem M: **Lifestyle of Lactobacillus plantarum in the mouse caecum.** *Environ Microbiol* 2009, **11**(10):2747–2757.
45. Ndagijimana M, Laghi L, Vitali B, Placucci G, Brigidi P, Guerzoni ME: **Effect of a synbiotic food consumption on human gut metabolic profiles evaluated by 1H Nuclear Magnetic Resonance spectroscopy.** *Int J Food Microbiol* 2009, **134**(1–2):147–153.
46. Frese SA, Benson AK, Tannock GW, Loach DM, Kim J, Zhang M, Oh PL, Heng NCK, Patil PB, Juge N, MacKenzie DA, Pearson BM, Lapidus A, Dalin E, Tice H, Goltsman E, Land M, Hauser L, Ivanova N, Kyrpides NC, Walter J: **The Evolution of Host Specialization in the Vertebrate Gut Symbiont *Lactobacillus reuteri*.** *PLOS Genet* 2011, **7**(2):e1001314.
47. O' Sullivan O, O' Callaghan J, Sangrador-Vegas A, McAuliffe O, Slattery L, Kaleta P, Callanan M, Fitzgerald GF, Ross RP, Beresford T: **Comparative genomics of lactic acid bacteria reveals a niche-specific gene set.** *BMC Microbiol* 2009, **9**(1):50.
48. Callanan M, Kaleta P, O'Callaghan J, O'Sullivan O, Jordan K, McAuliffe O, Sangrador-Vegas A, Slattery L, Fitzgerald GF, Beresford T, Ross RP: **Genome sequence of *Lactobacillus helveticus*, an organism distinguished by selective gene loss and insertion sequence element expansion.** *J Bacteriol* 2008, **190**(2):727–735.
49. Molenaar D, Bringel F, Schuren FH, de Vos WM, Siezen RJ, Kleerebezem M: **Exploring *Lactobacillus plantarum* genome diversity by using microarrays.** *J Bacteriol* 2005, **187**(17):6119–6127.
50. Majumder A, Sultan A, Jersie-Christensen RR, Ejby M, Schmidt BG, Lahtinen SJ, Jacobsen S, Svensson B: **Proteome reference map of Lactobacillus acidophilus NCFM and quantitative proteomics towards understanding the prebiotic action of lactitol.** *Proteomics* 2011, **11**(17):3470–3481.
51. Rackis Joseph J: **Oligosaccharides of Food Legumes: Alpha-Galactosidase Activity and the Flatus Problem.** In *Physiological Effects of Food Carbohydrates*, Volume 15. AMERICAN CHEMICAL SOCIETY; 1975:207–222.
52. Teixeira JS, McNeill V, Gänzle MG: **Levansucrase and sucrose phoshorylase contribute to raffinose, stachyose, and verbascose metabolism by lactobacilli.** *Food Microbiol* 2012, **31**(2):278–284.
53. Hao P, Zheng H, Yu Y, Ding G, Gu W, Chen S, Yu Z, Ren S, Oda M, Konno T, *et al*: **Complete Sequencing and Pan-Genomic Analysis of *Lactobacillus***

delbrueckii **subsp. *bulgaricus* Reveal Its Genetic Basis for Industrial Yogurt Production.** *PLoS One* 2011, **6**(1):e15964.
54. Burns P, Sánchez B, Vinderola G, Ruas-Madiedo P, Ruiz L, Margolles A, Reinheimer J, de los Reyes-Gavilán CG: **Inside the adaptation process of *Lactobacillus delbrueckii* subsp. *lactis* to bile.** *Int J Food Microbiol* 2010, **142**(1–2):132–141.
55. Cai H, Rodríguez BT, Zhang W, Broadbent JR, Steele JL: **Genotypic and phenotypic characterization of *Lactobacillus casei* strains isolated from different ecological niches suggests frequent recombination and niche specificity.** *Microbiology* 2007, **153**(8):2655–2665.
56. Kandler O, Weiss N: **Genus Lactobacillus.** In *Bergey's manual of systematic bacteriology*, Volume 2; 1986:1209–1234.
57. Chiaramonte F, Anglade P, Baraige F, Gratadoux J-J, Langella P, Champomier-Vergès M-C, Zagorec M: **Analysis of *Lactobacillus sakei* Mutants Selected after Adaptation to the Gastrointestinal Tracts of Axenic Mice.** *Appl Environ Microbiol* 2010, **76**(9):2932–2939.
58. Lengeler JW: **Carbohydrate transport in bacteria under environmental conditions, a black box?** *Antonie Van Leeuwenhoek* 1993, **63**(3–4):275–288.
59. Dills SS, Apperson A, Schmidt MR, Saier MH Jr: **Carbohydrate transport in bacteria.** *Microbiol Mol Biol Rev* 1980, **44**(3):385–418.
60. Poolman B: **Transporters and their roles in LAB cell physiology.** *A Van Leeuw* 2002, **82**(1):147–164.
61. Jojima T, Omumasaba C, Inui M, Yukawa H: **Sugar transporters in efficient utilization of mixed sugar substrates: current knowledge and outlook.** *Appl Microbiol Biot* 2010, **85**(3):471–480.
62. Saier MH, Tran CV, Barabote RD: **TCDB: the Transporter Classification Database for membrane transport protein analyses and information.** *Nucleic Acids Res* 2006, **34**(suppl 1):D181–D186.
63. Chaillou S, Champomier-Verges M-C, Cornet M, Crutz-Le Coq A-M, Dudez A-M, Martin V, Beaufils S, Darbon-Rongere E, Bossy R, Loux V, *et al*: **The complete genome sequence of the meat-borne lactic acid bacterium *Lactobacillus sakei* 23K.** *Nat Biotech* 2005, **23**(12):1527–1533.
64. Konings WN: **Microbial transport: adaptations to natural environments.** *A Van Leeuw* 2006, **90**(4):325–342.
65. Chang AB, Lin R, Studley WK, Tran CV, Saier JMH: **Phylogeny as a guide to structure and function of membrane transport proteins (Review).** *Mol Membr Biol* 2004, **21**(3):171–181.
66. Andersen JM, Barrangou R, Abou Hachem M, Lahtinen S, Goh YJ, Svensson B, Klaenhammer TR: **Transcriptional and functional analysis of galactooligosaccharide uptake by lacS in *Lactobacillus acidophilus*.** *Proc Natl Acad Sci USA* 2011, **108**(43):17785–17790.
67. Postma PW, Lengeler JW, Jacobson GR: **Phosphoenolpyruvate: carbohydrate phosphotransferase systems of bacteria.** *Microbiol Mol Biol Rev* 1993, **57**(3):543–594.
68. Saier MH Jr, Chauvaux S, Cook GM, Deutscher J, Paulsen IT, Reizer J, Ye JJ: **Catabolite repression and inducer control in Gram-positive bacteria.** *Microbiology* 1996, **142**:217.
69. Titgemeyer F, Hillen W: **Global control of sugar metabolism: a gram-positive solution.** *A Van Leeuw* 2002, **82**(1–4):59–71.
70. Glaser P, Frangeul L, Buchrieser C, Rusniok C, Amend A, Baquero F, Berche P, Bloecker H, Brandt P, Chakraborty T, *et al*: **Comparative Genomics of Listeria Species.** *Science* 2001, **294**(5543):849–852.
71. Deutscher J: **The mechanisms of carbon catabolite repression in bacteria.** *Curr Opin Microbiol* 2008, **11**(2):87–93.
72. Brückner R, Titgemeyer F: **Carbon catabolite repression in bacteria: choice of the carbon source and autoregulatory limitation of sugar utilization.** *FEMS Microbiol Lett* 2002, **209**(2):141–148.
73. Gorke B, Stulke J: **Carbon catabolite repression in bacteria: many ways to make the most out of nutrients.** *Nat Rev Microbiol* 2008, **6**(8):613–624.
74. Stulke J, Hillen W: **Carbon catabolite repression in bacteria.** *Curr Opin Microbiol* 1999, **2**(2):195–201.
75. Guillot A, Gitton C, Anglade P, Mistou M-Y: **Proteomic analysis of *Lactococcus lactis*, a lactic acid bacterium.** *Proteomics* 2003, **3**(3):337–354.
76. Marasco R, Muscariello L, Varcamonti M, De Felice M, Sacco M: **Expression of the bglH Gene of *Lactobacillus plantarum* Is Controlled by Carbon Catabolite Repression.** *J Bacteriol* 1998, **180**(13):3400–3404.
77. Kim J-H, Block DE, Shoemaker SP, Mills DA: **Atypical ethanol production by carbon catabolite derepressed lactobacilli.** *Bioresource Technol* 2010, **101**(22):8790–8797.
78. Kim J-H, Shoemaker SP, Mills DA: **Relaxed control of sugar utilization in *Lactobacillus brevis*.** *Microbiology* 2009, **155**(4):1351–1359.
79. Tauxe RV, Cavanagh TR, Cohen ML: **Interspecies Gene Transfer In Vivo Producing an Outbreak of Multiply Resistant Shigellosis.** *J Infect Diseases* 1989, **160**(6):1067–1070.
80. Rankin DJ, Rocha EPC, Brown SP: **What traits are carried on mobile genetic elements, and why[quest].** *Heredity* 2011, **106**(1):1–10.
81. Nicolas P, Bessières P, Ehrlich SD, Maguin E, Van De Guchte M: **Extensive horizontal transfer of core genome genes between two Lactobacillus species found in the gastrointestinal tract.** *BMC Evol Biol* 2007, **7**(1):141.
82. Zúñiga M, Comas I, Linaje R, Monedero V, Yebra MJ, Esteban CD, Deutscher J, Pérez-Martínez G, González-Candelas F: **Horizontal gene transfer in the molecular evolution of mannose PTS transporters.** *Mol Biol Evol* 2005, **22**(8):1673–1685.
83. Wang TT, Lee BH: **Plasmids in *Lactobacillus*.** *Crit Rev Biotechnol* 1997, **17**(3):227–272.
84. Vogel RF, Lohmann M, Weller AN, Hugas M, Hammes WP: **Structural similarity and distribution of small cryptic plasmids of *Lactobacillus curvatus* and *L. sake*.** *FEMS Microbiol Lett* 1991, **84**(2):183–190.
85. Li Y, Canchaya C, Fang F, Raftis E, Ryan KA, van Pijkeren JP, van Sinderen D, O'Toole PW: **Distribution of megaplasmids in *Lactobacillus salivarius* and other lactobacilli.** *J Bacteriol* 2007, **189**(17):6128–6139.
86. Posno M, Heuvelmans PT, van Giezen MJ, Lokman BC, Leer RJ, Pouwels PH: **Complementation of the inability of Lactobacillus strains to utilize D-xylose with D-xylose catabolism-encoding genes of *Lactobacillus pentosus*.** *Appl Environ Microbiol* 1991, **57**(9):2764–2766.
87. Zhang W, Yu D, Sun Z, Chen X, Bao Q, Meng H, Hu S, Zhang H: **Complete nucleotide sequence of plasmid plca36 isolated from Lactobacillus casei Zhang.** *Plasmid* 2008, **60**(2):131–135.
88. Lee LJ, Hansen JB, Jagusztyn-Krynicka EK, Chassy BM: **Cloning and expression of the beta-D-phosphogalactoside galactohydrolase gene of Lactobacillus casei in Escherichia coli K-12.** *J Bacteriol* 1982, **152**(3):1138–1146.
89. Teusink B, Wiersma A, Jacobs L, Notebaart RA, Smid EJ: **Understanding the Adaptive Growth Strategy of *Lactobacillus plantarum* by *In Silico* Optimisation.** *PLoS Comput Biol* 2009, **5**(6):e1000410.

doi:10.1186/1475-2859-12-48
Cite this article as: O'Donnell *et al.*: Catabolic flexibility of mammalian-associated lactobacilli. *Microbial Cell Factories* 2013 **12**:48.

O' Donnell et al. Microbial Cell Factories 2011, 10(Suppl 1):S12
http://www.microbialcellfactories.com/content/10/S1/S12

MICROBIAL CELL FACTORIES

PROCEEDINGS Open Access

Carbohydrate catabolic flexibility in the mammalian intestinal commensal *Lactobacillus ruminis* revealed by fermentation studies aligned to genome annotations

Michelle M O' Donnell[1,2], Brian M Forde[2], B Neville[2], Paul R Ross[1], Paul W O' Toole[2*]

From 10th Symposium on Lactic Acid Bacterium
Egmond aan Zee, the Netherlands. 28 August - 1 September 2011

Abstract

Background: *Lactobacillus ruminis* is a poorly characterized member of the *Lactobacillus salivarius* clade that is part of the intestinal microbiota of pigs, humans and other mammals. Its variable abundance in human and animals may be linked to historical changes over time and geographical differences in dietary intake of complex carbohydrates.

Results: In this study, we investigated the ability of nine *L. ruminis* strains of human and bovine origin to utilize fifty carbohydrates including simple sugars, oligosaccharides, and prebiotic polysaccharides. The growth patterns were compared with metabolic pathways predicted by annotation of a high quality draft genome sequence of ATCC 25644 (human isolate) and the complete genome of ATCC 27782 (bovine isolate). All of the strains tested utilized prebiotics including fructooligosaccharides (FOS), soybean-oligosaccharides (SOS) and 1,3:1,4-β-D-gluco-oligosaccharides to varying degrees. Six strains isolated from humans utilized FOS-enriched inulin, as well as FOS. In contrast, three strains isolated from cows grew poorly in FOS-supplemented medium. In general, carbohydrate utilisation patterns were strain-dependent and also varied depending on the degree of polymerisation or complexity of structure. Six putative operons were identified in the genome of the human isolate ATCC 25644 for the transport and utilisation of the prebiotics FOS, galacto-oligosaccharides (GOS), SOS, and 1,3:1,4-β-D-Gluco-oligosaccharides. One of these comprised a novel FOS utilisation operon with predicted capacity to degrade chicory-derived FOS. However, only three of these operons were identified in the ATCC 27782 genome that might account for the utilisation of only SOS and 1,3:1,4-β-D-Gluco-oligosaccharides.

Conclusions: This study has provided definitive genome-based evidence to support the fermentation patterns of nine strains of *Lactobacillus ruminis*, and has linked it to gene distribution patterns in strains from different sources. Furthermore, the study has identified prebiotic carbohydrates with the potential to promote *L. ruminis* growth *in vivo*.

Background

Immediately following birth, humans are colonised by a variety of bacteria which form the gastrointestinal tract microbiota [1]. Lactic Acid bacteria (LAB), which include *Lactobacillus* spp., are a subdominant element of the microbiota of humans and animals [2].

Lactobacillus ruminis is a LAB which is part of the autochthonous microbiota in the intestines of both humans [3], and pigs [4] and it has also been isolated from the bovine rumen [5]. *L. ruminis* is a low G+C Gram positive bacillus [6]. It is a candidate probiotic organism (see below), since it has been reported to have immunomodulatory characteristics [7], specifically the ability to induce Nuclear Factor Kappa B (NF-κB) in the absence of lipopolysaccharide production and to activate Tumour Necrosis Factor alpha (TNFα) production in

* Correspondence: pwotoole@ucc.ie
[2]Department Microbiology, University College Cork, Ireland
Full list of author information is available at the end of the article

THP-1 monocytes [7]. Unusually, some strains of *L. ruminis* are motile [5]. Limited studies have identified some of the carbohydrates utilised by *L. ruminis* which include cellobiose and raffinose [5,6,8]. However, little information is available about the fermentation of oligosaccharides/prebiotics by *Lactobacillus ruminis*.

There is growing interest in modulating the human microbiota using dietary supplements including probiotics and prebiotics. Probiotics are defined as "live microorganisms which when administered in adequate amounts confer a health benefit on the host" [9]. However, maintained ingestion of probiotic cultures is generally required to sustain the probiotic effect, with only some of the inoculum surviving gastrointestinal transit, and the vast majority of surviving bacteria shed days after ingestion [10]. For this reason there has been an increasing research effort expended in the area of prebiotics in order to extend the persistence of particular bacteria (mainly bifidobacteria) in the intestine. Prebiotics are "selectively fermented ingredients that result in specific changes in the composition and/or activity of the gastrointestinal microbiota, thus conferring benefit (s) upon host health" [11]. To be considered a prebiotic, the compound has to resist hydrolysis by gastrointestinal tract enzymes and pass into the large intestine, where ideally it promotes the growth of commensal bacteria [12]. The fermentation of prebiotics in the colon is largely influenced by the type of sugar monomer, the degree of polymerisation and the nature of the glycosidic bonds between the sugar moieties [13]. The constituent sugars of the majority of prebiotics are monosaccharides such as glucose, fructose, galactose and xylose [14]. The degree of polymerisation (DP) of prebiotics can vary from as low as two for lactulose and in excess of 23 for chicory-derived inulin [15]. Humans lack the gastrointestinal enzymes necessary to degrade many of the glycosidic bonds between the sugar units of compounds that are prebiotics, which accounts for their resistance to hydrolysis [14]. A number of enzymes produced by colonic commensal bacteria may hydrolyse these bonds. These glycosyl hydrolase (GH) enzymes include β-Glucosidases, α-Glucosidases, β-Fructofuranosidases, β-Galactosidases and α-Galactosidases [16-18].

Studies of other *Lactobacillus* species have identified a variety of genetic systems that encode the ability to utilize carbohydrates of varying complexity. β-fructofuranosidase is responsible for the hydrolysis of FOS, and this activity was identified in *L. plantarum* WCFS1 [19], *L. acidophilus* NCFM [20], and *L. paracasei* 1195 [21]. β-galactosidases involved in lactose degradation were characterised in *L. sakei*[22], *L. bulgaricus*[23], *L. coryniformis*[24] and *L. reuteri*[25]. β-glucosidase activity (which is responsible for the hydrolysis of 1,4-β-D-Glucans like cellobiose) has been identified in *L. plantarum*[26]. α-galactosidases, which hydrolyse α-galactosides like raffinose, stachyose and melibiose, were identified in *L. plantarum* ATCC 8014 [27] and *L. reuteri*[28]. Moreover, several α-glucosidases have been characterised in *L. brevis*[29], *L. acidophilus*[30] and *L. pentosus*[31].

In this study, we describe the fermentation profiles of nine strains of *Lactobacillus ruminis*. The interpretation of the carbohydrate utilisation profiles generated was complemented by the annotation of carbohydrate utilisation genes in the genomes of *L. ruminis* ATCC 25644 and ATCC 27782.

Results

Growth of *L. ruminis* in media containing diverse carbon sources

A carbohydrate utilisation profile for each of nine strains of *L. ruminis* on fifty carbohydrates was established as described in Methods. Additional file 1 summarizes the data, with individual strain data in Additional Files 2, 3, 4, 5, 6, 7, 8, 9, 10. In summary, there was significant variation with respect to carbohydrate fermentation profiles at the strain level. Moderate growth was observed for strains L5 and S21 when grown on α-galactosides (melibiose, raffinose, stachyose) and β-glucosides (β-glucotriose B, cellobiose) (Additional file 1). The majority of bovine isolates could poorly utilize fructooligosaccharides, except for ATCC 27781 with Beneo P95 and Raftilose P95. Moderate growth was observed for the majority of isolates with galactooligosaccharides (GOS, GOS-inulin, lactose, lactulose). All strains were able to ferment β-Glucotriose B, cellobiose, galactose, glucose, maltose, mannose, melibiose, raffinose, stachyose and sucrose (Additional file 1). Some strains showed a distinctly higher ability to utilize specific carbohydrates e.g. fructose by strains L5 and S21, (Additional Files 2 and 3); lactose by strains S23, ATCC 25644 and ATCC 27780T (Additional Files 4, 7 and 8); raffinose by ATCC 27781 (Additional File 8); and Raftilose P95 by strain S36 (Additional File 5).

Growth and fermentation analysis of human and bovine-derived *L. ruminis* type strains

Table 1 shows the final cell numbers and culture-medium pH values reached for the two strains ATCC 25644 (human isolate) and ATCC 27782 (bovine isolate), in the presence of various carbohydrates and prebiotics for 24 h. *L. ruminis* ATCC 25644 reached the highest cell density (8.9 x 10^8 cfu/ml) when grown on Raftilose Synergy 1 which coincided with the lowest culture medium pH value of 4.86. ATCC 27782 reached the highest cell density values (2.7x10^8 cfu/ml) when grown on Beta Glucotriose B, and fermentation resulted in a culture medium pH value of 5.19 following 24 hours incubation. This was far higher than cellobiose, the other beta-

Table 1 Growth and fermentation analysis of *L. ruminis* strains ATCC 25644 (human isolate) and ATCC 27782 (bovine isolate).

Carbohydrate type	Carbohydrate	ATCC 25644		ATCC 27782	
		Cfu/ml	pH*	Cfu/ml	pH*
	Cellobiose	2.40 x 10^8	5.21	7.00 x 10^6	5.13
Disaccharide	Lactulose	3.20 x 10^8	4.99	0	6.53
	Lactose	2.76 x 10^8	4.76	0	6.57
Monosaccharide	Glucose	4.39 x 10^8	4.86	1.53 x 10^8	4.85
	Beta Glucotriose B	4.05 x 10^8	5.17	2.66 x 10^8	5.19
Oligosaccharide	Raftilose Synergy 1	8.90 x 10^8	5.01	1.35 x 10^7	6.04
	Raftilose P95	2.91 x 10^8	5.28	2.51 x 10^6	5.42
Tetrasaccharide	Stachyose	3.94 x 10^8	5.13	2.37 x 10^8	5.11
Trisaccharide	Raffinose	3.24 x 10^8	5.2	1.40 x 10^8	5.2

*. pH value of culture medium after 24 h growth in indicated carbon source. Values tabulated are the average of two replicates carried out on separate days.

glucoside tested, although the final pH of both cultures was very similar, and the medium was buffered in the same way as MRS.

Annotation of carbohydrate pathways in the *L. ruminis* genome

A high-quality draft genome sequence was generated for *L. ruminis* ATCC 25644 and a finished genome sequence was generated for ATCC 27782, as described in Methods. The complete functional and comparative analysis of these genomes will be described elsewhere (Forde *et al.*, in preparation; Neville *et al.*, in preparation). A draft sequence of ATCC 25644 has also been generated by the Human Microbiome Project [32]; however it has a different scaffold structure and assembly statistics to that which we generated for ATCC 25644, and for that reason was not used in the current study. The carbohydrate utilisation genes of ATCC 25644 and ATCC 27782 were annotated by manual curation in conjunction with KEGG Automatic Annotation Server (KAAS). *L. ruminis*-specific Kyoto Encyclopaedia of Genes and Genomes (KEGG) maps were generated based upon our annotated genome sequences that we analyzed with KAAS. As a representative example, the galactose metabolic pathway (for both sequenced *L. ruminis* genomes) is presented in Figure 1. It demonstrates the predicted reliance on glycosyl hydrolases to ferment carbohydrates in *L. ruminis* as well as highlighting the fermentable α and β-galactosides.

Sixteen major pathways or systems involved in carbohydrate utilization were annotated in both genomes, and are shown in Additional Files 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26. These include those for glycolysis, pentose and glucuronate interconversions, fructose and mannose utilization, starch and sucrose. Of the sixteen pathways identified, eight are considered partial pathways (Additional Files 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26).

Identification of Glycosyl Hydrolases

Glycosyl hydrolases are key to prebiotic utilization, and can also be manipulated to synthesize prebiotics. Twenty glycosyl hydrolases were annotated in the genome of ATCC 25644, and fourteen were annotated in the genome of ATCC 27782. The glycosyl hydrolases include α-amylase (EC 3.2.1.1), endo-1,4-β-xylanase (EC 3.2.1.8), oligo-1,6-glucosidase (EC 3.2.1.10), lysozyme (EC 3.2.1.17), α-glucosidase (EC 3.2.1.20), β-glucosidase (EC 3.2.1.21), α-galactosidase (EC 3.2.1.22), β-galactosidase (EC 3.2.1.23), β-fructofuranosidase (EC 3.2.1.26), β-N-acetylhexosaminidase (EC 3.2.1.52), glucan 1,6-α-glucosidase (EC 3.2.1.70), 6-phospho-β−glucosidase (EC 3.2.1.86) and neopullalanase (EC 3.2.1.135). The majority of these enzymes are present in ATCC 27782 with the exceptions of α-amylase, oligo-1,6-glucosidase and β-galactosidase.

Identification of putative genes and operons involved in prebiotic utilisation

The sequenced *L. ruminis* genomes were extensively scrutinized to identify putative operons involved in carbohydrate transport and utilisation. Specificity of substrate was based upon manual curation of the annotated region, including reference to BLAST identity to functionally characterized homologues, genetic neighbourhood analysis, and protein motif matching. Six putative prebiotic utilisation operons were annotated in the *L. ruminis* ATCC 25644 genome (human isolate; Figure 2),

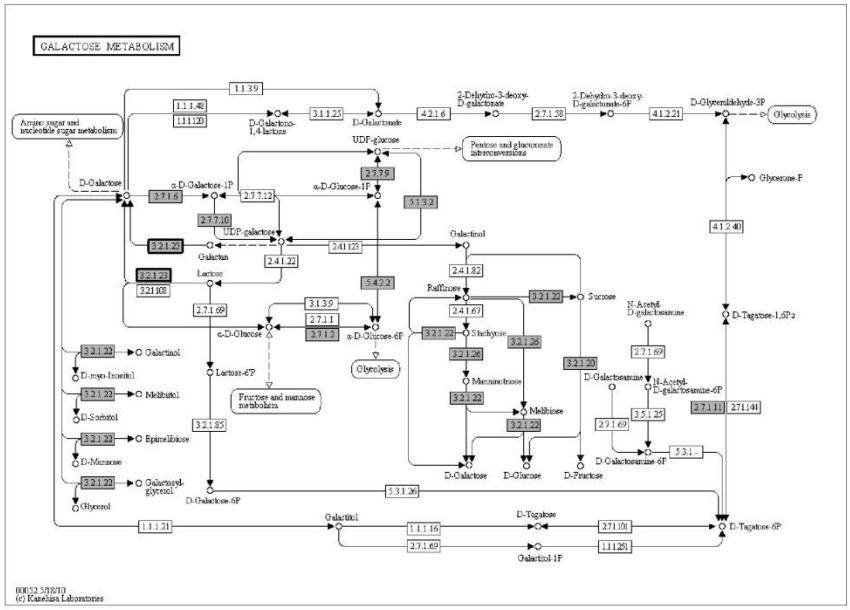


Figure 1 Galactoside utilisation metabolic map for *L. ruminis* ATCC 25644 and ATCC 27782. Grey boxes, enzymes present in both ATCC 25644 and ATCC 27782; Grey boxes with emphasised black border, enzymes present in ATCC 25644 and absent from ATCC 27782.

only three of which were identified in the bovine isolate ATCC 27782 (Additional File 27). Most of the operons are flanked by predicted rho-independent transcriptional terminators (Figure 2), and these operons constitute one to two transcriptional units, with a gene for a LacI-type transcriptional regulator in four of six cases.

We annotated a predicted FOS utilization operon only in the human isolate *L. ruminis* ATCC 25644. β-fructofuranosidase, a Glycosyl hydrolase (GH) family 32 enzyme [16], has been identified as the key enzyme in operons involved in FOS utilisation in other *Lactobacillus* species [19-21]. This activity is predicted to be encoded by the *L. ruminisbfrA* gene, which is linked to a presumptive oligosaccharide symporter gene. The ATCC 25644 genome was also distinguished by having two additional operons for lactose/galactose utilization (Figure 2). The genomes of both strains harboured operons predicted to confer utilization of sucrose, cellobiose and raffinose. As well as the β-fructofuranosidase (*sacA*) in the sucrose operon, genes for an amylopullalanase (*amyB*) and an α-glucosidase (*malZ*) are also contiguous and are potentially co-transcribed with the sucrose operon, but do not have a predicted function in the hydrolysis of sucrose or FOS (Figure 2B).

The cellobiose operon is predicted to be responsible for the transport and hydrolysis of both cellobiose and 1,3:1,4-β-D-Glucan hydrolysates, and in *L. ruminis* it appears to involve two β glucosidases (Figure 2) that belong to the GH1 family of glycosyl hydrolases [16]. The amino acid sequence of BglB and $BglB_2$ showed 70% and 77% identity to the β-glucosidases identified in the genomes of *L. helveticus* DPC 4571 and *L. ultunensis* DSM 16047, respectively. The products of the raffinose operon (Fig. 2D; also present in ATCC 27782) are predicted to have the additional ability to breakdown melibiose and stachyose. All of the glycolytic enzymes discussed above lack predicted transmembrane domains (TMD) and therefore most likely require import of their respective substrates.

Predicted carbohydrate transporters

A relationship exists between the genomic association of genes and the functional interaction of the proteins they

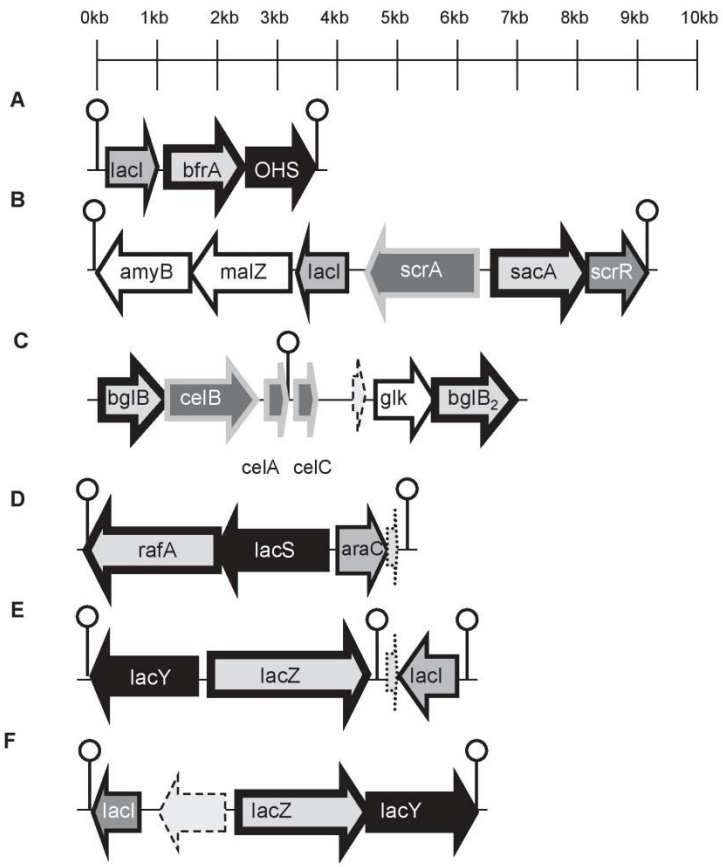


Figure 2 Putative operons for the predicted utilisation of carbohydrates in *L. ruminis* ATCC 25644. Predicted substrates are A, FOS; B, Sucrose; C, Cellobiose; D, raffinose; E, lactose/galactose; F, lactose/galactose operon. Light grey arrows with thick black border, glycosyl hydrolase family enzyme; Black arrows, major facilitator superfamily transporters; Medium grey arrows, transcriptional regulators; Dark grey arrows with thick grey border, phosphotransferase system transporters; Lollipops, rho-dependent transcriptional terminators; White arrows with dashed surround, transposases; white arrows with dotted surround, hypothetical proteins; White arrows with black continuous surround, potentially co-transcribed enzymes. Operons B, C and D were also annotated in the ATCC 27782 genome (see Additional File 27: Figure 1).

encode [33]. To refine our annotation of the carbohydrate utilisation operons, we therefore performed a detailed analysis of the predicted transporter proteins encoded by the contiguous genes. As for hydrolases, specificity of substrate was predicted based upon an integrated analysis of the annotated region, including reference to BLAST identity to functionally characterized homologues, linked genes, and protein motif matching. Putative carbohydrate transporters were analysed with transmembrane prediction software, with 14 and 10 transporters identified in the genome sequences of *L. ruminis* ATCC 25644 and 27782, respectively (Table 2). The predicted carbohydrate

Table 2 Transmembrane domains (TMD) of the predicted carbohydrate transport proteins in *Lactobacillus ruminis*

Family	Gene	Locus number[a]		Predicted substrate	TMD[b]	
		ATCC 25644	ATCC 27782		ATCC 25644	ATCC 27782
OHS	*lacY*	ANHS_218	-	FOS	12	-
GPH	*lacY*	ANHS_744c ANHS_924	-	Lactose, galactose, galactan	12	-
	lacS	ANHS_783	LRC_18250	Raffinose, stachyose, melibiose	12	12
	ugpE	ANHS_648	LRC_16940	Glycerol	6	6
ABC	*ugpA*	ANHS_649c	LRC_16950	Glycerol	6	6
	malG	ANHS_839c	LRC_18720	Maltose	6	6
	malF	ANHS_840c	LRC_18730	Maltose	8	8
	manY	ANHS_242	LRC_18860	Mannose	7	7
	manZ	ANHS_243	LRC_18850	Mannose	5	4
PTS	*scrA*	ANHS_846c	LRC_18780	Sucrose, FOS	8	8
	fruA	ANHS_1075	LRC_00800	Fructose	9	9
	celB	ANHS_1218	LRC_02240	Cellobiose	10	10
	gluA	ANHS_851c	LRC_18820	Glucose	9	9

a. Locus number in draft genome sequences

b. TMD: predicted trans-membrane domains, as described in Materials & Methods

transporters belong to the ATP-binding Cassette family (ABC), the Glycoside-Pentoside-Hexuronide cation symporter family (GPH), the Oligosaccharide H^+ Symporter (OHS) and the Phosphotransferase System (PTS). Transmembrane domain (TMD) numbers are generally indicative of the type of carbohydrate transporter, with some exceptions [34]. ABC transporters have on average 10-12 TMD but this can be highly variable. PTS transporters have been identified with up to 10 TMD (this study). GPH and OHS transporters (both being Major Facilitator Superfamily transporters) generally have 12 TMD [34]. In ATCC 25644, three GPH transporters were identified (Table 2) and these are predicted to transport the β-galactosides (lactose, galactose, lactulose and GOS) and the α-galactosides (raffinose, melibiose and stachyose). However, in ATCC 27782 only one GPH transporter was identified, which was predicted to transport α-galactosides. The OHS identified in the genome of ATCC 25644 is adjacent to a β-fructofuranosidase and may be involved in transporting FOS. Both genomes encode six predicted PTS transporters, which potentially transport mannose, sucrose, fructose, cellobiose and glucose. In both *L. ruminis* genome sequences, four ABC transporters were identified, with the putative substrates identified as mannose and glycerol-3-phosphate. All of the transporters identified in each genome had associated metabolic genes located either upstream or downstream in the genome, and the majority were arranged in operons. Both genomes also encoded proteins for glucose uptake (with TMD counts of 5 and 9 in ATCC 25644 and ATCC 27782, respectively), and a simple sugar transport system permease protein which was predicted to transport monosaccharides like galactose.

Discussion

We consider *L. ruminis* as a candidate probiotic, which we are also investigating as a potential responder for prebiotic/symbiotic supplementation in humans and animals. Several studies have identified *L. ruminis* in the gastrointestinal tract of humans [35-37]. *L. ruminis* was isolated from the bovine rumen [5], from the pig [4,8]], chickens [38], sheep [39], Svalbard reindeer [40], horses [41-43], cats [44,45], dogs [46] and parrots [47]. *L. ruminis* thus appears to be variably present in the microbiota of humans and many domesticated animals.

L. ruminis was previously described as a homofermentative bacterium, with the ability to ferment amygdalin, cellobiose, galactose, maltose, mannose, melibiose, raffinose, salicin, sorbitol and sucrose [48]. In the current study, the nine strains of *L. ruminis* were unable to utilise sorbitol as a carbon source. *L. ruminis* has also been reported to have the ability to ferment D-ribose [49]. However, we observed no growth for any of the nine *L. ruminis* strains when cultured in cfMRS supplemented with ribose. ATCC 27782 lacks a transaldolase gene (and the draft genome sequence suggests ATCC 25644 also lacks this gene), which would account for inability to utilise any of the pentose sugars tested. All of the *L. ruminis* strains tested (with the exception of ATCC 27782 which lacks a *lacZ* gene) had strong growth in lactose. This contrasts with a previous study, where

moderate growth was recorded on lactose [48]. It has also been reported that *L. ruminis* showed a strain dependent fermentation of starch [50], and very little growth was recorded for any of the strains tested here.

As a species, *L. ruminis* is generally able to ferment prebiotic compounds including FOS, GOS, lactulose, 1,3:1,4 β-D-Glucooligosaccharides, raffinose and stachyose. Only one strain, S36 was capable of (weakly) fermenting the prebiotic disaccharide palatinose. Palatinose is made by enzymatic rearrangement of the glycosidic linkages present in sucrose from an α-1,2-fructoside to an α-1,6-fructoside [51]. This suggests that the catalytic enzymes involved in sucrose utilisation may no longer be able to degrade the α-1,6-fructoside linkage in this disaccharide. The majority of *L. ruminis* strains achieved higher cell densities when grown on the prebiotic carbohydrates raffinose, lactulose, FOS, GOS and stachyose than when grown in other mono- and disaccharide carbohydrates tested. This growth pattern may be attributed to a niche for *L. ruminis* in the lower gastrointestinal tract (GIT). Mono and disaccharides are often unable to resist the hydrolytic action of the upper GIT, unlike prebiotics, and would not therefore be as freely available as carbon sources for *L. ruminis* in the large intestine. Lactulose, a disaccharide derivative of lactose, has previously been shown to support high level growth of other lactobacilli namely *L. rhamnosus*, *L. paracasei* and *L. salivarius*[52]. Lactulose also supported a high level of growth for the majority of *L. ruminis* strains. The β-galactosides lactulose and GOS are predicted to be transported and hydrolysed in ATCC 25644 by LacY and LacZ as part of the lactose operon. Two operons for β-galactoside utilisation were identified in the genome of ATCC 25644; however neither of these operons or any potential genetic determinants could be identified for lactose utilisation in ATCC 27782. The absence of a lactose operon in the genome may suggest an ecological niche adaptation by ATCC 27782 to an environment devoid of milk sugars.

β-glucooligosaccharides such as cellobiose are generally transported and hydrolysed using the cellobiose PTS and β-glucosidase enzymes. Both cellobiose and β-glucotriose B are 1,4-β-D-glucooligosaccharides with a similar structure which allows the transport and utilisation of these carbohydrates by the products of the cellobiose operon. The bovine *L. ruminis* isolates, ATCC 27780T, 27781 and 27782 were previously reported to utilise β-glucan hydrolysates as a carbohydrate source [53], and in that study, all bovine isolates utilised β-glucan hydrolysates of DP3, and only ATCC 27780T was unable to utilise DP4 oligosaccharide. ATCC 27781 was distinguished by being able to utilise the highest percentage of both DP3 and DP4 β glucan. We have shown that all the strains tested in this study were able to utilise the DP3 β-glucan hydrolysates to a moderate degree. The bovine isolate ATCC 27780T achieved the highest growth (data not shown) when utilizing β glucan hydrolysate, in contrast to a previous study which identified ATCC 27781 as having the highest percentage utilisation of β-glucan oligosaccharide [53].

In previous analysis of sixteen *Lactobacillus* species, only *L. acidophilus* L3, *L. acidophilus* 74-2 and *L. casei* CRL431 were able to utilise Raftilose P95, an oligofructose [54]. In the current study, eight strains of *L. ruminis* were capable of utilizing Raftilose P95. In addition, *L. ruminis* was capable of moderate to strong fermentation of Raftilose Synergy 1, an oligofructose-enriched inulin. *L. paracasei* subsp. *paracasei* 8700:2 was previously shown to be the only strain, out of ten strains tested, that was capable of strong growth on Raftilose Synergy 1, while three other species were capable of moderate growth [55]. Based on these comparisons, *L. ruminis* may have a growth advantage over other lactobacilli in the presence of fructooligosaccharides.

A novel β-fructofuranosidase was identified in the genome of *L. ruminis* ATCC 25644 that potentially hydrolyses the linkages present in chicory derived fructooligosaccharides. The cognate transporter OHS was identified only in the strains isolated from humans. Transport of FOS may be transported using the sucrose PTS transporter in the bovine strains ATCC 27780 and 27781. The human isolates of *L. ruminis* apparently use an OHS to transport FOS into the cell. Both sequenced strains likely use the ABC transport system to transport simple carbohydrates like maltose and glycerol. The most populated class of transporter identified was the phosphotransferase system transporter, with six such systems present. However, in *L. ruminis* many of the fermentable carbohydrates including α-galactosides and β-galactosides are predicted to be transported by GPH symporters. GPH transporters contain a C-terminal hydrophilic domain which interacts with the PTS system [34], which may thus be an important regulatory mechanism in *L. ruminis*.

Conclusions

Lactobacillus ruminis is a saccharolytic member of the intestinal microbiota capable of degrading a variety of prebiotics. Genes and operons were identified in the genomes of two sequenced strains for the hydrolysis and transport of the utilisable prebiotics. This work is the first step in the characterisation of carbohydrate metabolism, transportation and regulation in *L. ruminis*. Further studies will focus on the functional characterisation of the putative operons identified in this study and also *in vivo* studies with dietary supplementation by selected carbohydrates. Characterisation of the novel FOS degrading enzyme BfrA may facilitate applications

including reverse engineering of the FOS degradation pathway to allow the biosynthesis of a potentially novel fructooligosaccharide.

Methods

Bacterial strains and culture conditions

Nine *Lactobacillus ruminis* strains were used in this study, and were obtained courtesy of Prof. Gerald Tannock, University of Otago, New Zealand. Four of these are American Type Culture Collection strains: ATCC 25644 (human isolate), ATCC 27780T, ATCC 27781 and ATCC 27782 (bovine isolates). Five human-derived *L. ruminis* strains, L5, S21, S23, S36 and S38 were also studied. All strains were stored at -80°C in de Man-Rogosa-Sharpe (MRS) broth (Difco, BD, Ireland), supplemented with 25% (vol/vol) glycerol as a cryoprotectant. *Lactobacillus* strains were grown anaerobically on MRS agar plates at 37°C for two days. Growth tests were initiated by growing *Lactobacillus* strains anaerobically in MRS-glucose broth at 37°C overnight and unless otherwise stated, all further incubations were also performed under anaerobic conditions at 37°C.

Growth medium

Modifications were made to the de Man-Rogosa-Sharpe (MRS) [56] medium by omitting the carbohydrate source (glucose) and meat extract. Carbohydrate-free MRS (cfMRS) was used as a basal growth medium to study the ability of *Lactobacillus ruminis* strains to utilise various carbohydrates, because it contains no additional carbohydrates and lacks Lab Lembco as a source of carbohydrates. The cfMRS medium contained the following components (gL^{-1}): bacteriological peptone (Oxoid) 10.0, yeast extract (Fluka) 5.0, sodium acetate (Sigma) 5.0, ammonium citrate (Sigma) 2.0, potassium phosphate (Sigma) 2.0, magnesium sulphate (BDH Chemical) 0.2, Manganese sulphate (BDH Chemical) 0.05. The medium also includes Tween 80 (Sigma) 1 ml $litre^{-1}$. The pH was adjusted to between 6.2 and 6.5 and the medium was sterilised at 121°C for 15 minutes. Carbohydrate-free MRS was unable to support bacterial growth above an OD_{600nm} of 0.1 for any of the strains tested.

Carbohydrates and prebiotics

Fifty- two carbohydrates were used in this study (Additional file 28). Stock solutions of the 50 carbohydrates were filter-sterilized (0.45μm) (Sarstedt) into the cfMRS basal medium to yield a concentration of 0.5% (v/v) for use in the fermentation tests.

Growth measurements

The fermentation profiles of the various strains were determined using optical density (OD) measurements. The sterile carbohydrate supplemented MRS media was added to the wells of 96 well microtiter plates. The medium in the wells was inoculated with 1% (v/v) of the overnight bacterial culture in MRS-glucose. The OD values of the 96 well microtiter plate wells were read using a Synergy 2 plate reader (BioTek Instruments, Inc., Vermont, US). The inoculated microtiter plates were incubated anaerobically at 37°C and OD readings were taken before and after a 48 hour period [57]. The mean OD readings, standard deviations and standard errors were calculated using technical triplicate data from biological duplicate experiments.

Lactobacillus ruminis genome sequencing and assembly

The genome sequencing, assembly and detailed annotation of the *L. ruminis* ATCC 27782 and 25644 genomes will be described elsewhere in this volume (Forde *et al*, manuscript in preparation). In brief, a hybrid next-generation strategy generated 28-fold coverage of the ATCC 27782 genome by 454 pyrosequencing, complimented by 217-fold coverage with Illumina paired-end sequences. The assembly of *L. ruminis* ATCC 27782 is a finished genome; the genome assembly of *L. ruminis* ATCC 25644 a high-quality draft [58].

Bioinformatic analysis and gene annotation

The Artemis program [59] was used to visualise and identify carbohydrate metabolism genes in the genome of *Lactobacillus ruminis* ATCC 25644 and ATCC 27782 [60]. Open reading frames were predicted using Glimmer 3 [61]. Each carbohydrate utilisation enzyme, predicted from opening reading frames (ORF), was assigned a KEGG orthology (KO) identifier by KAAS and graphical representations for each metabolic pathway were generated [62]. The TMHMM 2.0 server was used to predict the transmembrane helices of proteins, which were identified from annotation as putative carbohydrate transporters. THHMM 2.0 uses Hidden Markov models to predict the proteins topology with a high degree of accuracy [63]. TransTermHP [64] was used to predict rho-independent transcriptional terminators. Comparisons to other Lactobacillus genomes were made using data available from both NCBI [65] and KEGG Organisms [66].

Sequence data availability and accession numbers

The finished genome of ATCC 27782 is available under accession number XXYYZZ123. The draft genome of ATCC 25644 is available under accession number CCGGHHIIUU.

Additional material

Additional file 1: Fermentation profiles for nine *Lactobacillus ruminis* strains

Additional file 2: Growth profile for *L. ruminis* strain L5

Additional file 3: Growth profile for *L. ruminis* strain S21

Additional file 4: Growth profile for *L. ruminis* strain S23

Additional file 5: Growth profile for *L. ruminis* strain S36

Additional file 6: Growth profile for *L. ruminis* strain S38

Additional file 7: Growth profile for *L. ruminis* strain ATCC 25644

Additional file 8: Growth profile for *L. ruminis* strain ATCC 27780T

Additional file 9: Growth profile for *L. ruminis* strain ATCC 27781

Additional file 10: Growth profile for *L. ruminis* strain ATCC 27782

Additional file 11: Glycolysis map representing enzymes present in *L. ruminis* ATCC 25644 or ATCC 27782.

Additional file 12: Citrate cycle map representing enzymes present in *L. ruminis* ATCC 25644 or ATCC 27782.

Additional file 13: Pentose phosphate pathway map representing enzymes present in *L. ruminis* ATCC 25644 or ATCC 27782.

Additional file 14: Pentose and glucuronate interconversions map representing enzymes present in *L. ruminis* ATCC 25644 or ATCC 27782.

Additional file 15: Fructose and Mannose metabolism map representing enzymes present in *L. ruminis* ATCC 25644 or ATCC 27782.

Additional file 16: Galactose metabolism map representing enzymes present in *L. ruminis* ATCC 25644 or ATCC 27782.

Additional file 17: Ascorbate and aldarate metabolism map representing enzymes present in *L. ruminis* ATCC 25644 or ATCC 27782.

Additional file 18: Starch and sucrose metabolism map representing enzymes present in *L. ruminis* ATCC 25644 or ATCC 27782.

Additional file 19: Amino and nucleotide sugar metabolism map representing enzymes present in *L. ruminis* ATCC 25644 or ATCC 27782.

Additional file 20: Inositol Phosphate metabolism map representing enzymes present in *L. ruminis* ATCC 25644 or ATCC 27782.

Additional file 21: Pyruvate metabolism map representing enzymes present in *L. ruminis* ATCC 25644 or ATCC 27782.

Additional file 22: Glyoxylate and Dicarboxylate metabolism map representing enzymes present in *L. ruminis* ATCC 25644 or ATCC 27782.

Additional file 23: Propanoate metabolic map representing enzymes present in *L. ruminis* ATCC 25644 or ATCC 27782.

Additional file 24: Butanoate metabolic map representing enzymes present in *L. ruminis* ATCC 25644 or ATCC 27782.

Additional file 25: ABC transporters map representing enzymes present in *L. ruminis* ATCC 25644 or ATCC 27782.

Additional file 26: Phosphotransferase system map representing enzymes present in *L. ruminis* ATCC 25644 or ATCC 27782.

Additional file 27: Operons in the genome of *L. ruminis* ATCC 27782 associated with prebiotic utilisation

Additional file 28: Carbohydrates used in this study

Acknowledgements
This work was supported by a Principal Investigator Award (07/IN.1/B1780) from Science Foundation Ireland to P.W. O'Toole.
This article has been published as part of *Microbial Cell Factories* Volume 10 Supplement 1, 2011: Proceedings of the 10th Symposium on Lactic Acid Bacterium. The full contents of the supplement are available online at http://www.microbialcellfactories.com/supplements/10/S1.

Author details
[1]Teagasc Food Research Centre, Moorepark, Fermoy, Co. Cork, Ireland.
[2]Department Microbiology, University College Cork, Ireland.

Authors' contributions
MMOD designed the experiments and drafted the manuscript. BMF carried out the genome sequencing and assembly of both *L. ruminis* genomes and provided the output of both the TMHMM 2.0 server and TransTerm HP. BAN initiated the genome sequencing and participated in assembly of the genomes. PWOT and RPR conceived the study, designed the research, and contributed to writing the manuscript.

Competing interests
The authors declare they have no competing interest.

Published: 30 August 2011

References

1. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, *et al*: **A human gut microbial gene catalogue established by metagenomic sequencing.** *Nature* 2010, **464**:59-65.
2. O'Toole PW, Claesson MJ: **Gut microbiota: Changes throughout the lifespan from infancy to elderly.** *International Dairy Journal* **20**:281-291.
3. Reuter G: **The *Lactobacillus* and *Bifidobacterium* microflora of the human intestine: composition and succession.** *Curr Issues Intest Microbiol* 2001, **2**:43-53.
4. Al Jassim RA: ***Lactobacillus ruminis* is a predominant lactic acid producing bacterium in the caecum and rectum of the pig.** *Lett Appl Microbiol* 2003, **37**:213-217.
5. Sharpe ME, Latham MJ, Garvie EI, Zirngibl J, Kandler O: **Two new species of *Lactobacillus* isolated from the bovine rumen, *Lactobacillus ruminis* sp. nov. and *Lactobacillus vitulinus* sp.nov.** *J Gen Microbiol* 1973, **77**:37-49.
6. Krieg NR, Holt JG: **Bergey's manual of systematic bacteriology.** 1984.
7. Taweechotipatr M, Iyer C, Spinler JK, Versalovic J, Tumwasorn S: ***Lactobacillus saerimneri* and *Lactobacillus ruminis*: novel human-derived probiotic strains with immunomodulatory activities.** *FEMS Microbiology Letters* 2009, **293**:65-72.
8. Yin Q, Zheng Q: **Isolation and identification of the dominant *Lactobacillus* in gut and faeces of pigs using carbohydrate fermentation and 16S rDNA analysis.** *J Biosci Bioeng* 2005, **99**:68-71.
9. **Report on Joint FAO/WHO expert consultation on evaluation of health and nutritional properties of probiotics in food Including powder milk with live lactic acid bacteria.** [ftp://ftp.fao.org/es/esn/food/probio_report_en.pdf].
10. Bezkorovainy A: **Probiotics: determinants of survival and growth in the gut.** *Am J Clin Nutr* 2001, **73**:399S-405.
11. Gibson GR, Scott KP, Rastall RA, Tuohy KM, Hotchkiss A, Dubert-Ferrandon A, Gareau M, Murphy EF, Saulnier D, Loh G: **Dietary prebiotics: current status and new definition.** *Food Science & Technology Bulletin: Functional Foods* 2010, **7**:1-19.
12. Gibson GR, Roberfroid MB: **Dietary modulation of the human colonic microbiota: Introducing the concept of prebiotics.** *J Nutr* 1995, **125**:1401-1412.
13. Swennen K, Courtin C, Delcour J: **Non-digestible oligosaccharides with prebiotic properties.** *Critical Reviews in Food Science & Nutrition* 2006, **46**:459-471.
14. Manning TS, Gibson GR: **Microbial-gut interactions in health and disease. Prebiotics.** *Best Pract Res Clin Gastroenterol* 2004, **18**:287-298.
15. Gibson GR, Probert HM, Loo JV, Rastall RA, Roberfroid MB: **Dietary modulation of the human colonic microbiota: updating the concept of prebiotics.** *Nutr Res Rev* 2004, **17**:259-275.
16. Henrissat B: **A classification of glycosyl hydrolases based on amino acid sequence similarities.** *Biochem J* 1991, **280**:309-316.
17. Henrissat B, Bairoch A: **New families in the classification of glycosyl hydrolases based on amino acid sequence similarities.** *Biochem J* 1993, **293(Pt 3)**:781-788.
18. Henrissat B, Bairoch A: **Updating the sequence-based classification of glycosyl hydrolases.** *Biochem J* 1996, **316(Pt 2)**:695-696.

19. Saulnier DM, Molenaar D, de Vos WM, Gibson GR, Kolida S: **Identification of prebiotic fructooligosaccharide metabolism in *Lactobacillus plantarum* WCFS1 through microarrays.** *Appl Environ Microbiol* 2007, **73**:1753-1765.
20. Barrangou R, Altermann E, Hutkins R, Cano R, Klaenhammer TR: **Functional and comparative genomic analyses of an operon involved in fructooligosaccharide utilization by *Lactobacillus acidophilus*.** *Proc Natl Acad Sci U S A* 2003, **100**:8957-8962.
21. Goh YJ, Zhang C, Benson AK, Schlegel V, Lee J-H, Hutkins RW: **Identification of a putative operon involved in fructooligosaccharide utilization by *Lactobacillus paracasei*.** *Appl Environ Microbiol* 2006, **72**:7518-7530.
22. Stentz R, Loizel C, Malleret C, Zagorec M: **Development of genetic tools for *Lactobacillus sakei*: Disruption of the beta-galactosidase gene and use of lacZ as a reporter gene to study regulation of the putative copper ATPase, AtkB.** *Appl Environ Microbiol* 2000, **66**:4272-4278.
23. Schmidt BF, Adams RM, Requadt C, Power S, Mainzer SE: **Expression and nucleotide sequence of the *Lactobacillus bulgaricus* beta-galactosidase gene cloned in Escherichia coli.** *J Bacteriol* 1989, **171**:625-635.
24. Corral J, Bañuelos O, Adrio J, Velasco J: **Cloning and characterization of a β-galactosidase encoding region in *Lactobacillus coryniformis* CECT 5711.** *Applied microbiology and biotechnology* 2006, **73**:640-646.
25. Nguyen T-h, Splechtna B, Steinböck M, Kneifel W, Lettner HP, Kulbe KD, Haltrich D: **Purification and characterization of two novel beta-galactosidases from *Lactobacillus reuteri*.** *Journal of Agricultural and Food Chemistry* 2006, **54**:4989-4998.
26. Spano G, Rinaldi A, Ugliano M, Moio L, Beneduce L, Massa S: **A beta-glucosidase gene isolated from wine *Lactobacillus plantarum* is regulated by abiotic stresses.** *Journal of Applied Microbiology* 2005, **98**:855-861.
27. Silvestroni A, Connes C, Sesma F, de Giori GS, Piard J-C: **Characterization of the melA locus for α-galactosidase in *Lactobacillus plantarum*.** *Appl Environ Microbiol* 2002, **68**:5464-5471.
28. Tzortzis G, Jay AJ, Baillon MLA, Gibson GR, Rastall RA: **Synthesis of α-galactooligosaccharides with α-galactosidase from *Lactobacillus reuteri* of canine origin.** *Applied microbiology and biotechnology* 2003, **63**:286-292.
29. De Cort S, Kumara HMCS, Verachtert H: **Localization and characterization of alpha-glucosidase activity in *Lactobacillus brevis*.** *Appl Environ Microbiol* 1994, **60**:3074-3078.
30. Li KB, Chan KY: **Production and properties of alpha-glucosidase from *Lactobacillus acidophilus*.** *Appl Environ Microbiol* 1983, **46**:1380-1387.
31. Chaillou S, Lokman BC, Leer RJ, Posthuma C, Postma PW, Pouwels PH: **Cloning, sequence analysis, and characterization of the genes involved in isoprimeverose metabolism in *Lactobacillus pentosus*.** *J Bacteriol* 1998, **180**:2312-2320.
32. **Human Microbiome Project.** [http://nihroadmap.nih.gov/hmp/].
33. Snel B, Bork P, Huynen MA: **The identification of functional modules from the genomic association of genes.** *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99**:5890-5895.
34. Saier MH Jr.: **A functional-phylogenetic classification system for transmembrane solute transporters.** *Microbiol Mol Biol Rev* 2000, **64**:354-411.
35. Antonio MA, Hawes SE, Hillier SL: **The identification of vaginal *Lactobacillus* species and the demographic and microbiologic characteristics of women colonized by these species.** *J Infect Dis* 1999, **180**:1950-1956.
36. Delgado S, Suarez A, Otero L, Mayo B: **Variation of microbiological and biochemical parameters in the faeces of two healthy people over a 15 day period.** *Eur J Nutr* 2004, **43**:375-380.
37. Kimura K, Nishio T, Mizoguchi C, Koizumi A: **Analysis of the composition of lactobacilli in humans.** *Bioscience and Microflora* 2010, **29**:47-50.
38. Kovalenko NK, Golovach TN, Kvasnikov EI: **Lactic bacteria in the digestive tract of poultry.** *Mikrobiologiia* 1989, **58**:137-143.
39. Mueller RE, Iannotti EL, Asplund JM: **Isolation and identification of adherent epimural bacteria during succession in young lambs.** *Appl Environ Microbiol* 1984, **47**:724-730.
40. Mathiesen SD, Orpin CG, Greenwood Y, Blix AS: **Seasonal changes in the cecal microflora of the high-arctic Svalbard reindeer (*Rangifer tarandus platyrhynchus*).** *Appl Environ Microbiol* 1987, **53**:114-118.
41. Hidetoshi M, Akiyo N, Mitsuharu S, Hidehiro TOH, Fumihiko N, Masahiro N, Shin H, Yukio K, Misako T, Hideto T, *et al*: ***Lactobacillus hayakitensis*, *L. equigenerosi* and *L. equi*, predominant lactobacilli in the intestinal flora of healthy thoroughbreds.** *Animal Science Journal* 2009, **80**:339-346.
42. Vörös A: **Diet related changes in the gastrointestinal microbiota of horses.** *Masters* Swedish University of Agricultural Sciences, Dept of Animal Nutrition and Management; 2008.
43. Willing B, Voros A, Roos S, Jones C, Jansson A, Lindberg JE: **Changes in faecal bacteria associated with concentrate and forage-only diets fed to horses in training.** *Equine Vet J* 2009, **41**:908-914.
44. Desai AR, Musil KM, Carr AP, Hill JE: **Characterization and quantification of feline fecal microbiota using cpn60 sequence-based methods and investigation of animal-to-animal variation in microbial population structure.** *Veterinary Microbiology* 2009, **137**:120-128.
45. Ritchie LE, Burke KF, Garcia-Mazcorro JF, Steiner JM, Suchodolski JS: **Characterization of fecal microbiota in cats using universal 16S rRNA gene and group-specific primers for *Lactobacillus* and *Bifidobacterium* spp.** *Veterinary Microbiology* 2009, Corrected Proof.
46. Greetham HL, Giffard C, Hutson RA, Collins MD, Gibson GR: **Bacteriology of the Labrador dog gut: a cultural and genotypic approach.** *Journal of Applied Microbiology* 2002, **93**:640-646.
47. Xenoulis PG, Gray PL, Brightsmith D, Palculict B, Hoppes S, Steiner JM, Tizard I, Suchodolski JS: **Molecular characterization of the cloacal microbiota of wild and captive parrots.** *Veterinary Microbiology* 2010, **146**:320-325.
48. Kandler O, Weiss N: **Genus Lactobacillus.** *Bergey's manual of systematic bacteriology* 1986, **2**:1209-1234.
49. Tanasupawat S, Shida O, Okada S, Komagata K: ***Lactobacillus acidipiscis* sp. nov. and *Weissella thailandensis* sp. nov., isolated from fermented fish in Thailand.** *Int J Syst Evol Microbiol* 2000, **50**:1479-1485.
50. Kato Y, Sakala RM, Hayashidani H, Kiuchi A, Kaneuchi C, Ogawa M: ***Lactobacillus algidus* sp. nov., a psychrophilic lactic acid bacterium isolated from vacuum-packaged refrigerated beef.** *Int J Syst Evol Microbiol* 2000, **50**:1143-1149.
51. Lina BAR, Jonker D, Kozianowski G: **Isomaltulose (Palatinose®): a review of biological and toxicological studies.** *Food and Chemical Toxicology* 2002, **40**:1375-1381.
52. Saarela M, Hallamaa K, Mattila-Sandholm T, Mättö J: **The effect of lactose derivatives lactulose, lactitol and lactobionic acid on the functional and technological properties of potentially probiotic *Lactobacillus* strains.** *International Dairy Journal* 2003, **13**:291-302.
53. Snart J, Bibiloni R, Grayson T, Lay C, Zhang H, Allison GE, Laverdiere JK, Temelli F, Vasanthan T, Bell R, Tannock GW: **Supplementation of the diet with high-viscosity beta-glucan results in enrichment for lactobacilli in the rat cecum.** *Appl Environ Microbiol* 2006, **72**:1925-1931.
54. Kneifel W, Rajal A, Kulbe KD: ***In vitro* growth behaviour of probiotic bacteria in culture media with carbohydrates of prebiotic importance.** *Microbial Ecology in Health & Disease* 2000, **12**:27-34.
55. Makras L, Van Acker G, De Vuyst L: ***Lactobacillus paracasei* subsp. *paracasei* 8700:2 degrades inulin-type fructans exhibiting different degrees of polymerization.** *Appl Environ Microbiol* 2005, **71**:6531-6537.
56. De Man JC, Rogosa M, Sharpe ME: **A medium for the cultivation of lactobacilli.** *Journal of Applied Microbiology* 1960, **23**:130-135.
57. Brewster JD: **A simple micro-growth assay for enumerating bacteria.** *Journal of Microbiological Methods* 2003, **53**:77-86.
58. Chain PSG, Grafham DV, Fulton RS, FitzGerald MG, Hostetler J, Muzny D, Ali J, Birren B, Bruce DC, Buhay C, *et al*: **Genome Project Standards in a New Era of Sequencing.** *Science* 2009, **326**:236-237.
59. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream M-Al, Barrell B: **Artemis: sequence visualization and annotation.** *Bioinformatics* 2000, **16**:944-945.
60. Mural RJ: **ARTEMIS: a tool for displaying and annotating DNA sequence.** *Brief Bioinform* 2000, **1**:199-200.
61. Delcher AL, Bratke KA, Powers EC, Salzberg SL: **Identifying bacterial genes and endosymbiont DNA with Glimmer.** *Bioinformatics* 2007, **23**:673-679.
62. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M: **KAAS: an automatic genome annotation and pathway reconstruction server.** *Nucleic Acids Res* 2007, **35**:W182-185.
63. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL: **Predicting transmembrane protein topology with a hidden markov model: application to complete genomes.** *Journal of Molecular Biology* 2001, **305**:567-580.
64. Kingsford C, Ayanbule K, Salzberg S: **Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake.** *Genome Biology* 2007, **8**:R22.

65. NCBI Genome Homepage. [http://www.ncbi.nlm.nih.gov/sites/genome].
66. KEGG Organisms: Complete Genomes. [http://kegg.jp/kegg/catalog/org_list.html].

doi:10.1186/1475-2859-10-S1-S12
Cite this article as: O' Donnell *et al.*: Carbohydrate catabolic flexibility in the mammalian intestinal commensal *Lactobacillus ruminis* revealed by fermentation studies aligned to genome annotations. *Microbial Cell Factories* 2011 **10**(Suppl 1):S12.

Letters in Applied Microbiology

Letters in Applied Microbiology ISSN 0266-8254

ORIGINAL ARTICLE

The core faecal bacterial microbiome of Irish Thoroughbred racehorses

M.M. O' Donnell[1,2], H.M.B. Harris[2], I.B. Jeffery[2], M.J. Claesson[2], B. Younge[3], P.W. O' Toole[2] and R.P. Ross[1]

1 Teagasc Food Research Centre, Moorepark, Fermoy, Ireland
2 Department of Microbiology, University College Cork, Cork, Ireland
3 Department of Life Sciences, Faculty of Science and Engineering, University of Limerick, Limerick, Ireland

Significance and Impact of the Study: Although Irish thoroughbreds are used nationally and internationally as performance animals, very little is known about the core faecal microbiota of these animals. This is the first study to characterize the bacterial microbiota present in the Irish thoroughbred racehorse faeces and elucidate a core microbiome irrespective of diet, animal management and geographical location.

Keywords
Intestinal microbiology, *Lactobacillus*, metabolism, metagenomics, *Streptococci*.

Correspondence
Reynolds P. Ross, Teagasc Food Research Centre, Moorepark, Fermoy, Co. Cork, Ireland.
E-mail: paul.ross@teagasc.ie

2013/0736: received 18 April 2013, revised 14 July 2013 and accepted 22 July 2013

doi:10.1111/lam.12137

Abstract

In this study, we characterized the gut microbiota in six healthy Irish thoroughbred racehorses and showed it to be dominated by the phyla *Firmicutes*, *Bacteroidetes*, *Proteobacteria*, *Verrucomicrobia*, *Actinobacteria*, *Euryarchaeota*, *Fibrobacteres* and *Spirochaetes*. Moreover, all the horses harboured *Clostridium*, *Fibrobacter*, *Faecalibacterium*, *Ruminococcus*, *Eubacterium*, *Oscillospira*, *Blautia Anaerotruncus*, *Coprococcus*, *Treponema* and *Lactobacillus* spp. Notwithstanding the sample size, it was noteworthy that the core microbiota species assignments identified *Fibrobacter succinogenes*, *Eubacterium coprostanoligenes*, *Eubacterium hallii*, *Eubacterium ruminantium*, *Oscillospira guillermondii*, *Sporobacter termiditis*, *Lactobacillus equicursoris*, *Treponema parvum* and *Treponema porcinum* in all the horses. This is the first study of the faecal microbiota in the Irish thoroughbred racehorse, a significant competitor in the global bloodstock industry. The information gathered in this pilot study provides a foundation for veterinarians and other equine health-associated professionals to begin to analyse the microbiome of performance of racehorses. This study and subsequent work may lead to alternate dietary approaches aimed at minimizing the risk of microbiota-related dysbiosis in these performance animals.

Introduction

The horse is a member of the family *Equidae* and is 'a monogastric' or nonruminant herbivore whose physiology is suited to digesting and utilizing high-fibre diets as a result of continual microbial fermentation within the hindgut. Ireland is now the third largest producer of thoroughbreds in the world after the USA and Australia, with approximately 40% of European thoroughbreds originating from Ireland (Leadon and Herholz 2009), and the equine sector is worth an estimated €100 billion a year to the European economy.

Until recently, the equine hindgut microbiota had remained relatively poorly characterized. Previous studies have used culture and molecular methods to identify the bacterial genera present in the equine gastrointestinal microbiota affected by laminitis and colic (Pollitt 2004; Milinovich *et al.* 2006; Respondek *et al.* 2008; Shirazi-Beechey 2008). Recent studies have used next-generation sequencing to investigate the faecal microbiota of two Arabian geldings (Shepherd *et al.* 2012). Comparison of the microbiota of healthy and unhealthy horses suffering from colitis (Costa *et al.* 2012) revealed a shift in the predominant phyla. The *Firmicutes* phylum predominated in

healthy horses, while in colitis-affected horses, *Bacteroidetes* predominated. A similar investigation comparing the microbiota of healthy horses and those with laminitis revealed an increase in the *Verrucomicrobia* phylum for those horses with the disease (Steelman *et al.* 2012).

The link between altered gastrointestinal microbiota and disease risk is becoming a well-established concept in both humans and animals (Yatsunenko *et al.* 2012). Identification of the core microbiota present in the faeces of horses would allow for a better understanding of the dietary requirements needed to prevent or to inhibit microbiota-related diseases and to promote gut health. Fructans and starches are present at varying levels in grasses depending on the growing season and the cultivar (Hoffman *et al.* 2001; Superchi *et al.* 2010) and thus can have seasonal effects on the composition of the microbiota of the grazing horse. Knowledge of the effect that different grasses, types of forage, concentrates and supplements can have on the horse microbiota is therefore very valuable especially to the bloodstock industry.

The objective of this study was to characterize the microbiota of Irish thoroughbred horses fed various commonly consumed diets to elucidate the core microbiome of the Irish thoroughbred racehorse independent of diet, management regime, geographical location or age.

Results and discussion

Horses feed naturally by what is termed 'trickle feeding' (Hill 2002); however, modern practices have necessarily altered this feeding pattern. Thoroughbreds and other performance racehorses are often fed a high-energy, carbohydrate-enriched feed twice a day. The microbiological impact of this alteration to the natural grazing-based feeding pattern of the horse has yet to be fully elucidated. Starch concentrate was chosen as a representative of 'high-sugar' feeds that are often detrimental to equine health. Feeding excessive carbohydrates to horses in the form of either starch or fructooligosaccharides may result in laminitis (Milinovich *et al.* 2010).

We applied 16S rRNA gene (V4 region) amplicon pyrosequencing to determine the faecal microbiota composition in six thoroughbred racehorses. Following the removal of low-quality reads, a total of 178 975 sequences were obtained from the six samples. Read numbers ranged from 17 757 to 38 378 (SD = 8002; Table 1). The average read length following quality trimming was 224·8 bp (SD = 3·23). A total of 19 phyla, 229 genera and 143 bacterial species were identified across the six horses. At the phylum level, an average of 93% of the reads from the trial animals were classified as bacterial phyla with 6% of the reads remaining unclassified and <0·3% Archaea. An average 43% of the reads identified were assigned to bacteria at the genus level, while 57% (average) remain unclassified, and a small proportion was assigned as Archaea (<0·3% on average). The high level of unclassified read assignments may suggest that the equine faecal samples contain many genera that are distinct and novel from those isolated from other mammals and the wider environment. The Archaea present in all horses were identified as *Methanobrevibacter* and *Methanocorpusculum*. *Methanobrevibacter woesei* was identified at the species level in all of the horses. However, it must be noted that a previous study has shown that the V4 region of the 16S rRNA gene may underestimate the true population levels of Archaea present in a faecal sample (Yu *et al.* 2008).

Four measures of alpha diversity (microbiota diversity within a subject) were calculated to assess the diversity of faecal microbiota in the 6 racehorses (Table S3). In each metric, the average diversity of samples TCM 1-2 was the lowest, followed by samples TCM 3-4, with samples TCM 5-6 having the highest alpha diversity in all four metrics. This may be due to the different feeding regimes and diets; however, a larger sample size would be needed for a conclusive analysis. Rarefaction curves were also generated for three of the alpha diversity metrics: observed species, phylogenetic diversity and the Shannon index (Fig. 1). The observed OTUs and the phylogenetic diversity metric curves have not reached a plateau at 17 000 reads, which suggests that the equine faecal microbiota is more diverse than that measured in this pilot study. Curves for the Shannon index plateau at relatively low read numbers. However, the saturation of microbial diversity at these read numbers is unlikely as the addition of low-abundance OTUs has a minor effect on the value of the Shannon index.

The beta diversity (i.e. between animals) of the six faecal microbiota samples was measured by generating PCoA plots based on the rarefied OTU table (Fig. S1). Figure S1 (a) shows an unweighted PCoA plot of the six samples, coloured by feed received. The first two principle axes, which explain 50% of the variation in the samples, show a grouping of the samples according to diet group. This suggests that the two samples from each diet group are more similar to each other in terms of the presence/absence of microbial taxa than they are to the samples from the other diet groups. Unweighted PCoA plots therefore may be affected by the inclusion of low-abundance reads; however, it is difficult to quantify the severity of this effect. While the weighted PCoA plot includes proportional data, the inclusion of low-abundance reads will probably have a negligible effect. Figure S1 (b) shows a weighted PCoA plot of the six samples, coloured by feed received. The first two principle axes, which explain 68% of the variation in the samples, do not group the samples according to diet.

Table 1 The core genera[a] and relative abundance identified in the hindgut microbiota of Irish thoroughbred racehorses

	% of total reads per animal						
Genus	TCM 1 (%)	TCM 2 (%)	TCM 3 (%)	TCM 4 (%)	TCM 5 (%)	TCM 6 (%)	Order › Family
Methanocorpusculum	0·0	0·6	0·1	0·1	0·1	0·5	*Methanomicrobiales › Methanocorpusculaceae*
Anaerophaga	0·4	0·3	1·2	0·8	0·2	1·3	*Bacteroidales › Marinilabiaceae*
Paludibacter	0·3	0·4	0·2	0·5	1·5	1·5	*Bacteroidales › Porphyromonadaceae*
Paraprevotella	0·1	0·1	0·9	0·3	0·2	0·3	*Bacteroidales › Prevotellaceae*
Prevotella	0·3	1·1	1·4	0·5	0·3	1·2	*Bacteroidales › Prevotellaceae*
Galbibacter	0·3	0·2	0·3	0·6	0·1	0·3	*Flavobacteriales › Flavobacteriaceae*
Fibrobacter	0·3	0·3	0·5	5·2	3·7	2·3	*Fibrobacterales › Fibrobacteraceae*
Anaerosporobacter	0·1	0·1	0·2	0·3	0·3	0·2	*Clostridiales › Clostridiaceae*
Clostridium	0·6	0·7	0·4	0·1	1·3	1·0	*Clostridiales › Clostridiaceae*
Lactonifactor	0·2	0·4	0·1	0·9	0·4	0·2	*Clostridiales › Clostridiaceae*
Eubacterium	0·0	0·5	0·0	0·1	0·5	0·1	*Clostridiales › Eubacteriaceae*
Acetitomaculum	0·1	0·5	0·0	0·5	1·3	0·4	*Clostridiales › Lachnospiraceae*
Blautia	0·3	0·8	0·1	1·9	1·8	1·1	*Clostridiales › Lachnospiraceae*
Coprococcus	0·5	0·5	0·6	1·3	1·4	0·7	*Clostridiales › Lachnospiraceae*
Dorea	0·1	0·3	0·1	0·3	0·4	0·4	*Clostridiales › Lachnospiraceae*
Pseudobutyrivibrio	0·1	0·3	0·2	1·0	0·5	0·5	*Clostridiales › Lachnospiraceae*
Robinsoniella	0·1	0·0	0·1	0·5	0·2	0·5	*Clostridiales › Lachnospiraceae*
Roseburia	0·1	0·2	0·1	0·5	0·8	0·5	*Clostridiales › Lachnospiraceae*
Oscillibacter	1·7	5·3	2·4	0·6	1·0	2·0	*Clostridiales › Oscillospiracea*
Acetivibrio	0·4	1·1	0·7	1·2	1·0	1·0	*Clostridiales › Ruminococcaceae*
Anaerotruncus	0·0	2·2	1·2	0·4	0·2	0·5	*Clostridiales › Ruminococcaceae*
Faecalibacterium	0·2	0·1	0·2	0·5	0·4	1·1	*Clostridiales › Ruminococcaceae*
Papillibacter	0·3	0·4	0·9	0·3	0·3	1·0	*Clostridiales › Ruminococcaceae*
Ruminococcus	0·7	0·9	1·3	3·2	2·3	1·2	*Clostridiales › Ruminococcaceae*
Sporobacter	1·3	6·3	3·5	3·6	3·1	3·3	*Clostridiales › Ruminococcaceae*
Holdemania	0·1	0·1	0·3	0·2	0·2	0·3	*Erysipelotrichales › Erysipelotrichaceae*
Lactobacillus	2·6	0·8	0·0	0·4	0·4	0·0	*Lactobacillales › Lactobacillaceae*
Acidaminococcus	0·2	0·3	0·7	0·2	0·2	0·4	*Selenomonadales › Acidaminococcaceae*
Phascolarctobacterium	0·1	0·1	0·1	0·2	0·1	0·0	*Selenomonadales › Acidaminococcaceae*
Acinetobacter	0·2	0·0	9·1	0·4	1·0	0·0	*Pseudomonadales › Moraxellaceae*
Treponema	1·8	2·3	4·7	5·7	5·8	2·9	*Spirochaetales › Spirochaetaceae*
Anaeroplasma	0·0	0·1	0·0	0·2	0·2	0·5	*Anaeroplasmatales › Anaeroplasmataceae*
Akkermansia	0·2	0·2	0·0	0·1	0·1	0·0	*Verrucomicrobiales › Verrucomicrobiaceae*
Subdivision5_*incertae_sedis*	2·1	2·4	3·0	3·9	6·4	10·8	–

[a]Core as defined by presence in the microbiota of 4 or more of the 6 racehorses at ≥0·1% of the total reads.

When relative abundance of taxa is taken into account, TCM 4 is grouped closer to the grass-fed samples and TCM 3 and TCM 2 are grouped together, while TCM 1 lies a considerable distance away from both groups. A possible reason for this 'outlier' status may be due to the significantly younger age of the animal TCM 1, but it is more likely due to this horse only being housed at the sample collection stable for approximately a month. The previous feeding regime and management style experienced by this horse may have greatly influenced its microbiota and thus our beta diversity indices. This outlier status can also be seen in the large number of reads assigned to the *Streptococcus* genus in this horse compared with the other animals in the study.

We measured greater phylotype diversity in the equine faecal microbiota compared with data from the distal bowel microbiota of other animals (Pitta *et al.* 2010) (Lamendella *et al.* 2011). Our phylotype number estimations for the equine faecal microbiota (1755–2736) are higher than those estimated for the human microbiota (Claesson *et al.* 2009; Nam *et al.* 2011) and other horses (Shepherd *et al.* 2012; Steelman *et al.* 2012). However, this difference in observed phylotypes might be influenced by the metric used to generate the phylotype numbers (Kemp and Aller 2004). Additionally, we opted to not to use a de-noise step in the QIIME pipeline, and this too can have an influence on the alpha diversity metrics (Reeder and Knight 2010). The diversity indices indicated

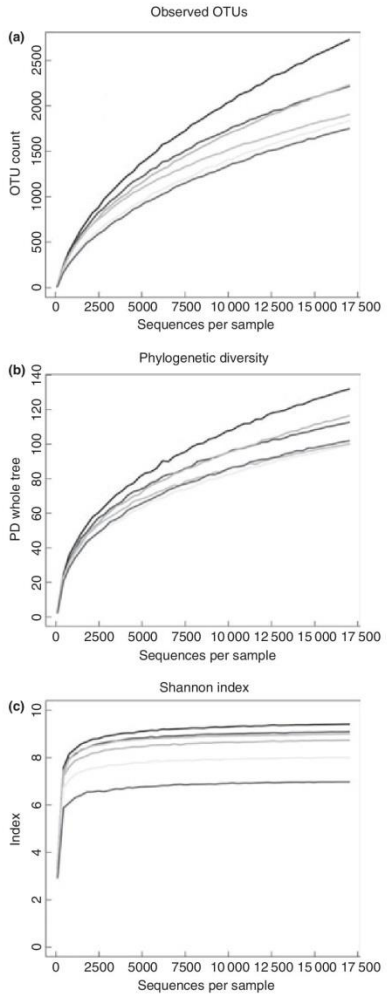


Figure 1 (a) Observed species (OTUs); (b) phylogenetic diversity; (c) Shannon index identified from the faecal samples of each horse used in this study. Each of the plots was generated by multiple rarefaction where subsamples of different depths (read number) were taken from each sample in increments of 100 reads (*x*-axis) and the appropriate diversity metric at each subsampling was calculated (*y*-axis). The colour scheme is the same for all three plots. (■) TCM 1, (■) TCM 2, (■) TCM 3, (■) TCM 4, (■), TCM 5 and (■) TCM 6.

that consumption of the starch concentrate in conjunction with haylage reduced the faecal microbiota diversity, where the forage-fed horses harboured the most diverse microbiota. However, further study is needed to confirm this trend.

The relative phylum abundance in the faecal microbiota of the six racehorses is shown Fig. 2. A total of 19 phyla were identified, twelve of which were present in all horses. In addition to the phyla shown in Fig. 2, these 12 phyla include (*Chlamydiae*, *Chloroflexi*, *Deferribacteres*, *Cyanobacteria* and *Synergistetes* present at low-abundance levels). Phyla *Firmicutes* and *Bacteroidetes* were dominant in all the horses, with all microbiota displaying a *Firmicutes* to *Bacteroidetes* ratio of >2 : 1. Collectively, these two phyla accounted for 73–85% (SD = 5·7%) of the sequences. Although the current analysis of the thoroughbred microbiota identified 19 phyla, only five were present in all horses above a 0·5% cut-off. The dominance of *Firmicutes* and *Bacterioidetes* phyla in the faecal microbiome is similar to that measured in humans and cows (van den Bogert *et al.* 2011; Jami and Mizrahi 2012). The *Firmicutes* range (47–74%) is consistent with other equine studies, which attributed 15–83% (Daly *et al.* 2001; Willing *et al.* 2009; Costa *et al.* 2012; Perkins *et al.* 2012; Shepherd *et al.* 2012; Steelman *et al.* 2012) of the total reads to the *Firmicutes* phylum. The relative abundance of *Bacteroidetes* (3·65–9·94%) identified by other equine microbiota studies (Shepherd *et al.* 2012; Steelman *et al.* 2012) is far lower than the average relative abundance identified by this study, but is similar to the levels identified by Willing *et al.* 2009 (Willing *et al.* 2009). The relative abundances of the *Proteobacteria*, *Verrucomicrobia*, *Spirochaetes*, *Tenericutes* and *Fibrobacteres* phyla identified in this study were statistically significantly higher ($P < 0{\cdot}001$) in the horses fed the forage-based diets. To our knowledge, this is also the first time the *Euryarchaeota* phylum has been identified in horses, albeit at low levels; however, it is as of yet unknown what function members of this phylum have in the microbiome of horses. However, as stated earlier, this may be an underestimation of the true extent of the presence of the Archaea in the horse faecal samples due to the use of the V4 region primer pair.

Consideration of the assignment of sequences to phylogenetic orders within the *Firmicutes* phylum (Fig. S2) revealed that the increased *Firmicutes* abundance in the faeces of samples TCM 1-2 was due to increased abundance of the *Lactobacillales*, especially in sample TCM 1. At the order level, the microbiota of the horses was dominated by *Lactobacillales*, *Clostridia* and *Erysipelotrichi*. In our study, *Streptococcus* and *Lactobacillus* were identified at significantly elevated levels ($P < 0{\cdot}001$) in the horses from the Limerick stable, samples TCM 1-2. However, at the species level, only *Streptococcus caballi* (Milinovich

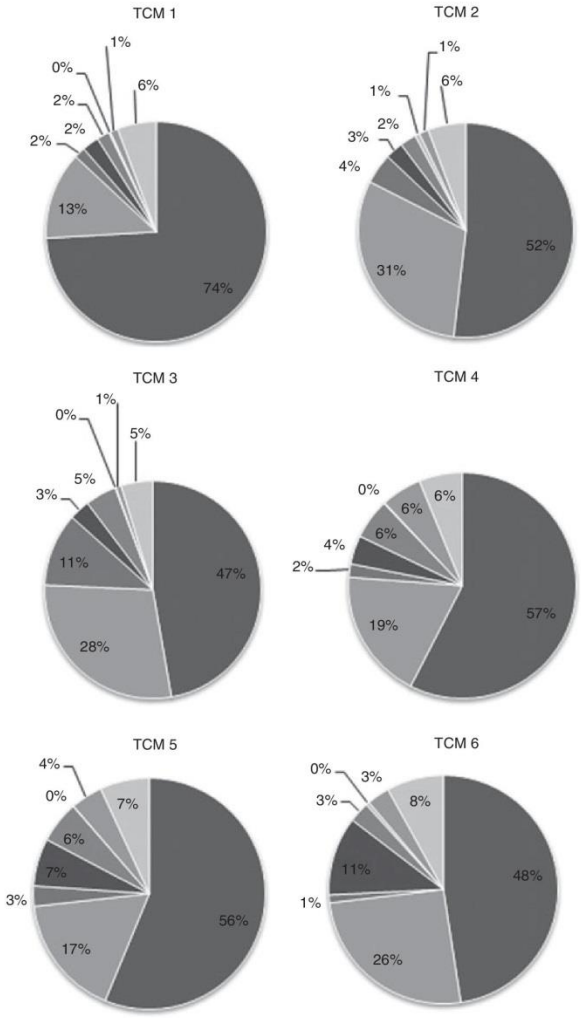


Figure 2 Phylum-level assignment of V4 16S rRNA sequences from individual horses, according to the RDP classifier (CI ≥ 97%). Reading clockwise: (■) *Firmicutes*, (■) *Bacteroidetes*, (■) *Proteobacteria*, (■) *Verrucomicrobia*, (■) *Spirochaetes*, (■) *Euryarchaeota*, (■) Other and (■) unclassified.

et al. 2008) was identified at <0·1% of the total reads from samples TCM 1-2 only. This loss of resolution at the species level is probably firstly due to the fact that some sequences are too short to accurately identify to species level. Secondly, the limited size of the RDP database (i.e. more sequences would lead to a greater representation of bacterial diversity, and more sequences would be assigned to species level).

Two hundred and twenty-nine genera were identified across the six horse samples, and Table 2 lists the

Table 2 Genus-level diversity of the faecal microbiota between the horses used in this study

	% of total reads per animal					
Genus	TCM 1 (%)	TCM 2 (%)	TCM 3 (%)	TCM 4 (%)	TCM 5 (%)	TCM 6 (%)
Streptococcus	26·85	6·46	0·02	0·15	0·00	0·00
Subdivision5_incertae_sedis	2·13	2·40	2·99	3·91	6·39	10·83
Treponema	1·82	2·28	4·65	5·72	5·84	2·86
Sporobacter	1·31	6·31	3·52	3·64	3·12	3·28
Oscillibacter	1·69	5·26	2·43	0·57	1·02	1·99
Acinetobacter	0·19	0·01	9·13	0·35	0·99	0·00
Fibrobacter	0·30	0·28	0·45	5·21	3·69	2·31
Ruminococcus	0·70	0·91	1·34	3·20	2·30	1·22
Allobaculum	7·93	0·02	0·01	0·08	0·01	0·00
Lysinibacillus	0·04	0·00	4·62	0·09	2·05	0·00
Blautia	0·28	0·80	0·09	1·85	1·77	1·10
Acetivibrio	0·40	1·09	0·68	1·17	1·03	1·03
Prevotella	0·34	1·11	1·36	0·52	0·33	1·21
Coprococcus	0·46	0·46	0·62	1·26	1·39	0·66
Clostridium	0·60	0·66	0·42	0·10	1·32	0·99
Anaerophaga	0·44	0·26	1·19	0·78	0·18	1·28
Anaerotruncus	0·02	2·17	1·21	0·39	0·17	0·51
Lactobacillus	2·58	0·76	0·03	0·40	0·38	0·02
Other	6·08	7·41	8·12	10·02	11·31	10·83
Unclassified	45·86	61·37	57·12	60·60	56·71	59·88

genus-level diversity of the faecal microbiota between the horses used in this study; 93 were found in ≥4 of the data sets and 64 of those were present in all the horses. We can thus consider these genera as being part of the core faecal microbiota of thoroughbreds. This means that approximately 41% of the genera identified were consistently found in the majority of the horses sampled and 28% of the genera are present at varying levels in all the horses. The genera most commonly found at relatively high levels (≥0·2% of reads) in the majority of the samples include *Prevotella, Fibrobacter, Clostridium, Ruminococcus, Sporobacter, Acinetobacter* and *Trepomena*. Further scrutiny of the genus data revealed that 34 genera were present in the faecal microbiota of ≥4 racehorses, the identities of which are listed in Table 1. Therefore, in this study, 15% of the faecal microbiota of the majority of horses tested was consistently found irrespective of feed or geographical locale.

Thirty-five species were present as core microbiota in four or more racehorses present at ≥0·1%, which are listed in Table S5. Of these 35 species, 19 belonged to the *Clostridiales* order. The majority of reads were assigned to the *Eubacteriaceae* and *Ruminococcaceae* families. However, this core species is calculated from the reads assigned to the species level and not the total reads for each horse. There are 13 species of bacteria present as a core microbiome when we calculated the species of bacteria present at ≥0·1% of the total read assignments for each horse. The majority of reads are assigned to the *Clostridiales* and *Spirochaetales* order. The read assignments for the total reads assigned can be seen in Table 3. *Sporobacter termiditis* was the most abundant species identified (2·9% average aggregate proportion across the six samples). When the species-level assignments generated from the total reads are compared with the genera-level assignments, *S. termiditis* accounts for between 52 and 100% of those reads assigned to the *Sporobacter* genus. The cellulolytic species *Fibrobacter succinogenes* was the second most abundant species identified in this study and accounted for 88–100% of the reads assigned to the Fibrobacter genus. *Lactobacillus equicursoris,* a predominant equine lactobacillus, accounts for between 13 and 94% of the reads assigned to the *Lactobacillus* genus. *Lb. equicursoris* was present in all samples, but statistically higher levels were present in samples TCM 1-2. On average, the percentage of the total reads in the study we could identify to the species level in each horse was *c.* 13%, which is in line with previous studies from our laboratory (Claesson *et al.* 2009).

Major *Clostridium* clusters in humans have been linked to changes in diet, short-chain fatty acid production and anti-inflammatory effects (O'Toole and Claesson 2010). *Clostridium* Clusters IV and XIVa dominate in all horses in this study (Fig S3), similar to studies on the human faecal microbiota (Claesson *et al.* 2011). The microbiota of samples TCM 5-6 had a higher proportion of Cluster I (1·76% average) clostridia. All racehorses examined had a similar proportion of *Clostridium* Cluster III. Sixteen *Clostridium* species were identified across the six racehorse microbiota data sets, although the efficiency of

Table 3 The 13 species that form the core microbiome accounting for ≥0·1% of the total reads for four or more animals used in this study

	% of the total reads per animal						
Species	TCM 1 (%)	TCM 2 (%)	TCM 3 (%)	TCM 4 (%)	TCM 5 (%)	TCM 6 (%)	Order › Family › Genus
Paludibacter propionicigenes	0·2	0·1	0·0	0·1	0·9	0·4	*Bacteroidales › Porphyromonadaceae › Paludibacter*
Fibrobacter succinogenes	0·3	0·3	0·4	5·2	3·7	2·3	*Fibrobacterales › Fibrobacteraceae › Fibrobacter*
Eubacterium coprostanoligenes	0·2	0·1	0·3	0·3	0·3	0·1	*Clostridiales › Eubacteriaceae › Eubacterium*
Eubacterium hallii	0·2	0·6	0·0	0·9	1·2	0·3	*Clostridiales › Eubacteriaceae › Eubacterium*
Eubacterium ruminantium	0·2	0·2	0·2	0·8	0·3	0·1	*Clostridiales › Eubacteriaceae › Eubacterium*
Oscillospira guilliermondii	0·8	3·9	1·4	0·3	1·0	1·1	*Clostridiales › Ruminococcaceae › Oscillospira*
Sporobacter termitidis	1·3	5·6	3·2	2·2	1·6	3·4	*Clostridiales › Ruminococcaceae › Sporobacter*
Lactobacillus equicursoris	1·8	0·5	0·0	0·4	0·1	0·0	*Lactobacillales › Lactobacillaceae › Lactobacillus*
Phascolarctobacterium faecium	0·1	0·0	0·1	0·1	0·1	0·0	*Selenomonadales › Acidaminococcaceae › Phascolarctobacterium*
Treponema brennaborense	0·1	0·0	0·0	0·2	0·1	0·1	*Spirochaetales › Spirochaetaceae › Treponema*
Treponema parvum	0·1	0·1	0·2	0·4	0·6	0·3	*Spirochaetales › Spirochaetaceae › Treponema*
Treponema porcinum	0·5	0·6	0·5	1·8	0·6	0·7	*Spirochaetales › Spirochaetaceae › Treponema*
Treponema saccharophilum	0·1	0·1	0·0	0·0	0·3	0·3	*Spirochaetales › Spirochaetaceae › Treponema*

assignment using this approach is not high (Claesson *et al.* 2009, 2010). *Clostridium butyricum*, *Clostridium caenicola*, *Clostridium hathewayi*, *Clostridium hylemonae*, *Clostridium lactatifermentans*, *Clostridium leptum* and *Clostridium methylpentosum* were present in the microbiota independent of the diet.

We investigated the microbiota at a single time point, from a single breed of horse housed in two stables close to Limerick City, Ireland. Seasonal and geographical influences on the forages consumed may also prevail; for example, Yamano *et al.* (2008) monitored the faecal bacteria from two horse breeds, the Hokkaido native horse dominated by cellulolytic species and a light horse breed dominated by soluble sugar utilizers grazing on hilly winter woodland pasture (Yamano *et al.* 2008). Costa *et al.* (2012) also noted that the two thoroughbreds that were housed similarly on the same farm with the same feeding regime had a similar microbiota and that feeding regimes, location and other management factors may influence the microbiota (Costa *et al.* 2012).

This study clearly outlines that the horse faecal microbiome is a diverse and practically unknown habitat, and as such, further large-scale studies are required to identify 'unclassified' genera and species present in the core microbiome. Although not the primary focus of this work, we noted that the feed consumed by the horses did have an effect on the levels of certain genera in the faecal microbiome. As a multimillion Euro industry and given the high monetary value of performance horses, future work should also concentrate on identifying the effect that diet has on the microbiome. A practical future application of this study and corroborated by future work might be, for example, that horses transferred to starch concentrate feed for performance enhancement might be supplemented with a microbiota cocktail corresponding to that typical for forage animals or with forage extracts to maintain levels of associated genera. This might offset or preclude the observed increases in *Streptococcus* or *Lactobacillus* abundance.

Materials and methods

Animals and diets

Faecal samples were collected from six mature Irish thoroughbred horses that were housed in two stables; horse weights, ages and genders are shown in Table S1. All faecal sample collection and analysis were consistent with the current animal welfare legislation in Ireland. The horses were each assigned the abbreviation TCM (thoroughbred core microbiome) and numbered from one to six. Faecal samples were collected, and all the faecal samples were held anaerobically at 4°C prior to DNA extraction within 24 h. Grass and haylage were chosen as diets to represent racehorses at rest, while haylage supplemented with starch concentrate represents diets of those performance horses in active training. Each horse had been receiving their respective feed for a month.

DNA extraction, PCR amplification and 454 pyrosequencing

Total genomic faecal DNA was isolated from the six faecal samples using the Isolate faecal DNA kit (myBio, Kilkenny, Ireland). The V4 region PCR conditions were outlined previously by Claesson *et al.* 2009 (Claesson *et al.* 2009). Table S2 contains a full list of the primers used in the study. PCR products were purified and quantified using the Agencourt

AMPure XP PCR (Beckman Coulter, High Wycombe, UK) purification beads and the Quant-It Picogreen dsDNA kit (Invitrogen, Biosciences, Dun Laoghaire, Dublin, Ireland), respectively. The 16S rRNA V4 amplicons were sequenced on a 454 Genome Sequencer FLX platform (Teagasc Food Research Centre, Moorepark, Ireland).

DNA sequence processing and statistical analysis

Raw sequencing reads were quality-trimmed using a locally installed version of the RDP Pyrosequencing Pipeline (Claesson *et al.* 2009). The following analysis of the pyrosequencing data was performed in QIIME (Caporaso *et al.* 2010b). All of the sequences from the six samples were clustered into OTUs (operational taxonomic units) of 97% sequence identity using UCLUST (Edgar 2010). The representative sequences for each OTU were aligned using PyNAST (Caporaso *et al.* 2010a), using the best match from the Greengenes (DeSantis *et al.* 2006) core set (http://greengenes.lbl.gov/). Taxonomy was assigned to the unaligned representative set using the RDP classifier (Cole *et al.* 2005) with a minimum confidence value of 0·8. Chimeras were identified in the aligned representative set using ChimeraSlayer (Haas *et al.* 2011), and the same core set of Greengenes aligned sequences used to align the representative set. A phylogenetic tree was constructed from the aligned, filtered representative set using FastTree (Price *et al.* 2009). Before rarefaction, the OTU table was filtered for OTUs represented by a single read in a single sample. If an OTU represented by a single read was identified in more than 1 sample, it was included in the study. The OTU table was rarefied to account for variations in sequencing depth among the samples, and a subsample of 17 000 sequences was taken from each sample. Weighted and unweighted UniFrac (Lozupone and Knight 2005) distance metrics were constructed from the rarefied OTU table. Single rarefaction was carried out on the OTU table, and the rarefied samples from this table were subjected to Uni Frac, principle coordinates analysis and statistics. Multiple rarefaction was used on the OTU table to generate the rarefaction plots. 2D and 3D PCoA plots were constructed from the weighted and unweighted distance metrics. The 2D plots were generated in R (version 2.13.1) from collated alpha diversity values imported from QIIME (Caporaso *et al.* 2010b). Due to the small sample sizes, statistical analysis of the data was carried out using Fisher's exact test (Hynes *et al.* 2002; Ruijter *et al.* 2002; Clayton *et al.* 2012). The method used to assign reads to *Clostridium* clusters is outlined in Claesson *et al.* (2011).

For species-level assignments, all the sequences in the RDP database were blasted against themselves in an all-against-all BLAST (Altschul *et al.* 1990). As multiple strains of the same species are present in the database, the BLAST score varied slightly for the multiple within-species BLAST alignments. Any sequence from our analysis blasted against the RDP database that had a score $\geq$ the lowest within-species BLAST score was assigned to that species as a 'strict' species assignment (Jeffery *et al.* 2012). If the BLAST score was lower than the lowest within-species BLAST score, but higher than the next highest BLAST score to another species, it was assigned to the species as a 'relaxed' species assignment. Otherwise, no species was assigned to the read. A core genus or species was assigned if it was present in the microbiota of $\geq$4 horses at 0·1% of the total read assignments. Relaxed species assignments were primarily used in this study.

Alpha and beta diversity metrics

Four alpha diversity metrics were calculated to measure the microbial diversity in each of the six horses. These metrics are observed OTUs, phylogenetic diversity, the Shannon index ($H' = -\sum p_i \log p_i$) and the species or Pielou's evenness ($E = H'/H'_{max}$). Each metric was calculated from a rarefied OTU table consisting of subsamples of 17 000 reads per sample. The last index used was phylogenetic diversity using a phylogenetic tree created from all the reads in the six samples. The phylogenetic diversity for any one sample was then the sum of the branch lengths that lead to every read in the tree that belongs to that sample. Rarefaction curves for each sample were based upon the calculated alpha diversity metric for subsamples ranging from 100 to 17 000 reads at increments of 100 reads.

Beta diversity was calculated using weighted and unweighted Uni Frac distance in QIIME and displayed graphically using principle coordinates analysis in R. Uni Frac distance is calculated by constructing a phylogenetic tree from all the OTUs and, for each pair of samples, calculating a distance measure using the equation (sum of unshared branch lengths)/(sum of total branch lengths).

Acknowledgements

This work was supported by a Principal Investigator Award (07/IN.1/B1780) from Science Foundation Ireland to P.W. O'Toole.

Conflict of Interest

No conflict of interest declared.

References

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

van den Bogert, B., de Vos, W.M., Zoetendal, E.G. and Kleerebezem, M. (2011) Microarray analysis and Barcoded pyrosequencing provide consistent microbial profiles depending on the source of human intestinal samples. *Appl Environ Microbiol* **77**, 2071–2080.

Caporaso, J.G., Bittinger, K., Bushman, F.D., DeSantis, T.Z., Andersen, G.L. and Knight, R. (2010a) PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26**, 266–267.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G. *et al.* (2010b) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**, 335–336.

Claesson, M.J., O'Sullivan, O., Wang, Q., Nikkilä, J., Marchesi, J.R., Smidt, H., De Vos, W.M., Ross, R.P. *et al.* (2009) Comparative analysis of pyrosequencing and a phylogenetic microarray for exploring microbial community structures in the human distal intestine. *PLoS ONE* **4**, e6669.

Claesson, M.J., Wang, Q., O'Sullivan, O., Greene-Diniz, R., Cole, J.R., Ross, R.P. and O'Toole, P.W. (2010) Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Res* **38**, e200.

Claesson, M.J., Cusack, S., O'Sullivan, O., Greene-Diniz, R., De Weerd, H., Flannery, E., Marchesi, J.R., Falush, D. *et al.* (2011) Composition, variability, and temporal stability of the intestinal microbiota of the elderly. *Proc Natl Acad Sci USA* **108**, 4586.

Clayton, E.M., Rea, M.C., Shanahan, F., Quigley, E.M.M., Kiely, B., Ross, R.P. and Hill, C. (2012) Carriage of *Clostridium difficile* in outpatients with irritable bowel syndrome. *J Med Microbiol* **61**, 1290–1294.

Cole, J., Chai, B., Farris, R., Wang, Q., Kulam, S., McGarrell, D., Garrity, G. and Tiedje, J. (2005) The ribosomal database project (RDP-II): sequences and tools for high-throughput rRNA analysis. *Nucleic Acids Res* **33**, D294–D296.

Costa, M.C., Arroyo, L.G., Allen-Vercoe, E., StÃ¤mpfli, H.R., Kim, P.T, Sturgeon, A. and Weese, J.S. (2012) Comparison of the fecal microbiota of healthy horses and horses with colitis by high throughput sequencing of the V3-V5 region of the 16S rRNA gene. *PLoS ONE* **7**, e41484.

Daly, K., Stewart, C.S., Flint, H.J. and Shirazi Beechey, S.P. (2001) Bacterial diversity within the equine large intestine as revealed by molecular analysis of cloned 16S rRNA genes. *FEMS Microbiol Ecol* **38**, 141–151.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D. *et al.* (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72**, 5069–5072.

Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461.

Haas, B.J., Gevers, D., Earl, A.M., Feldgarden, M., Ward, D.V., Giannoukos, G., Ciulla, D., Tabbaa, D. *et al.* (2011) Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res* **21**, 494–504.

Hill, J. (2002) Effect of level of inclusion and method of presentation of a single distillery by-product on the processes of ingestion of concentrate feeds by horses. *Livest Prod Sci* **75**, 209–218.

Hoffman, R., Wilson, J., Kronfeld, D., Cooper, W., Lawrence, L., Sklan, D. and Harris, P. (2001) Hydrolyzable carbohydrates in pasture, hay, and horse feeds: direct assay and seasonal variation. *J Anim Sci* **79**, 500.

Hynes, S.O., Broutet, N., Eurohepygast Study Group, Wadström, T., Mikelsaar, M., O'Toole, P.W., Telford, J., Engstrand, L., Kamiya, S., Mentis, A.F. and Moran, A.P. (2002) Phenotypic variation of *Helicobacter pylori* isolates from geographically distinct regions detected by lectin typing. *J Clin Microbiol* **40**, 227–232.

Jami, E. and Mizrahi, I. (2012) Composition and similarity of bovine rumen microbiota across individual animals. *PLoS ONE* **7**, e33306.

Jeffery, I.B., O'Toole, P.W., Öhman, L., Claesson, M.J., Deane, J., Quigley, E.M.M. and Simrén, M. (2012) An irritable bowel syndrome subtype defined by species-specific alterations in faecal microbiota. *Gut* **61**, 997–1006.

Kemp, P.F. and Aller, J.Y. (2004) Bacterial diversity in aquatic and other environments: what 16S rDNA libraries can tell us. *FEMS Microbiol Ecol* **47**, 161–177.

Lamendella, R., Santo Domingo, J., Ghosh, S., Martinson, J. and Oerther, D. (2011) Comparative fecal metagenomics unveils unique functional capacity of the swine gut. *BMC Microbiol* **11**, 103.

Leadon, D. and Herholz, C. (2009) Globalisation of trade and the spread of infectious disease. *Equine Vet Educ Man* **8**.

Lozupone, C. and Knight, R. (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* **71**, 8228–8235.

Milinovich, G., Trott, D., Burrell, P., Van Eps, A., Thoefner, M., Blackall, L., Al Jassim, R., Morton, J. *et al.* (2006) Changes in equine hindgut bacterial populations during oligofructose induced laminitis. *Environ Microbiol* **8**, 885–898.

Milinovich, G.J., Burrell, P.C., Pollitt, C.C., Bouvet, A. and Trott, D.J. (2008) *Streptococcus henryi* sp. nov. and *Streptococcus caballi* sp. nov., isolated from the hindgut of horses with oligofructose-induced laminitis. *Int J Syst Evol Microbiol* **58**, 262–266.

Milinovich, G.J., Klieve, A.V., Pollitt, C.C. and Trott, D.J. (2010) Microbial events in the hindgut during carbohydrate-induced equine laminitis. *Vet Clin North Am-Equine* **26**, 79–94.

Nam, Y.-D., Jung, M.-J., Roh, S.W., Kim, M.-S. and Bae, J.-W. (2011) Comparative analysis of Korean human gut microbiota by barcoded pyrosequencing. *PLoS ONE* **6**, e22109.

O'Toole, P.W. and Claesson, M.J. (2010) Gut microbiota: changes throughout the lifespan from infancy to elderly. *Int Dairy J* **20**, 281–291.

Perkins, G.A., den Bakker, H.C., Burton, A.J., Erb, H.N., McDonough, S.P., McDonough, P.L., Parker, J., Rosenthal, R.L. *et al.* (2012) The equine stomach harbors an abundant and diverse mucosal microbiota. *Appl Environ Microbiol* **78**, 2522–2532.

Pitta, D., Pinchak, W., Dowd, S., Osterstock, J., Gontcharova, V., Youn, E., Dorton, K., Yoon, I. *et al.* (2010) Rumen bacterial diversity dynamics associated with changing from Bermudagrass Hay to Grazed winter wheat diets. *Microbiol Ecol* **59**, 511–522.

Pollitt, C.C. (2004) Equine laminitis. *Clin Tech Equine Pract* **3**, 34–44.

Price, M.N., Dehal, P.S. and Arkin, A.P. (2009) FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* **26**, 1641–1650.

Reeder, J. and Knight, R. (2010) Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. *Nat Meth* **7**, 668–669.

Respondek, F., Goachet, A.G. and Julliand, V. (2008) Effects of dietary short-chain fructooligosaccharides on the intestinal microflora of horses subjected to a sudden change in diet. *J Anim Sci* **86**, 316–323.

Ruijter, J.M., Van Kampen, A.H.C. and Baas, F. (2002) Statistical evaluation of SAGE libraries: consequences for experimental design. *Physiol Genomics* **11**, 37–44.

Shepherd, M.L., Swecker, W.S. Jr, Jensen, R.V. and Ponder, M.A. (2012) Characterization of the fecal bacteria communities of forage fed horses by pyrosequencing of 16S rRNA V4 gene amplicons. *FEMS Microbiol Lett* **326**, 62–68.

Shirazi-Beechey, S.P. (2008) Molecular insights into dietary induced colic in the horse. *Equine Vet J* **40**, 414–421.

Steelman, S.M., Chowdhary, B.P., Dowd, S., Suchodolski, J. and Janečka, J.E. (2012) Pyrosequencing of 16S rRNA genes in fecal samples reveals high diversity of hindgut microflora in horses and potential links to chronic laminitis. *BMC Vet Res* **8**, 231.

Superchi, P., Sabbioni, A., Beretti, V., Vecchi, I. and Vaccari Simonini, F. (2010) Hydrolyzable and fermentable carbohydrates in North Italian pastures for horses. *Ital J Anim Sci* **9**, e66.

Willing, B., Vörös, A., Roos, S., Jones, C., Jansson, A. and Lindberg, J. (2009) Changes in faecal bacteria associated with concentrate and forage only diets fed to horses in training. *Equine Vet J* **41**, 908–914.

Yamano, H., Koike, S., Kobayashi, Y. and Hata, H. (2008) Phylogenetic analysis of hindgut microbiota in Hokkaido native horses compared to light horses. *Anim Sci J* **79**, 234–242.

Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., Magris, M., Hidalgo, G. *et al.* (2012) Human gut microbiome viewed across age and geography. *Nature* **486**, 222–227.

Yu, Z., García-González, R., Schanbacher, F.L. and Morrison, M. (2008) Evaluations of different hypervariable regions of archaeal 16S rRNA genes in profiling of methanogens by archaea-specific PCR and denaturing gradient gel electrophoresis. *Appl Environ Microbiol* **74**, 889–893.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1 Panel (a) unweighted, and panel (b) weighted, PCoA plots, of the six microbiota samples.

Figure S2 *Firmicutes* order-level read distribution in the hindgut microbiota between the six horses (TCM 1-6) used in this study.

Figure S3 *Clostridium* cluster assignment in the hindgut microbiota between the thoroughbred racehorses used in this study.

Table S1 Animals and diets used in this study.

Table S2 Barcode primers used in this study.

Table S3 The number of sequences obtained from faecal samples from the six racehorses and species-richness estimates (using 97% CI).

Table S4 The 229 genera identified and total read assignments for each horse used in the study.

Table S5 The core species and relative abundance identified in the hindgut microbiota of Irish thoroughbred racehorses.

genomeA
Journals.ASM.org

The Genome of the Predominant Equine *Lactobacillus* Species, *Lactobacillus equi*, Is Reflective of Its Lifestyle Adaptations to an Herbivorous Host

Michelle M. O'Donnell,[a,b] Hugh M. B. Harris,[b] Paul W. O'Toole,[b] R. Paul Ross[a]

Teagasc Food Research Centre, Moorepark, Fermoy, County Cork, Ireland[a]; Department of Microbiology, University College Cork, Cork, County Cork, Ireland[b]

We report the draft genome sequence of *Lactobacillus equi* strain DPC6820, isolated from equine feces. *L. equi* is a predominant *Lactobacillus* species in the horse hindgut microbiota. An examination of the genome identified genes and enzymes highlighting *L. equi* adaptations to the herbivorous gastrointestinal tract of the horse, including fructan hydrolases. This genome sequence may help us further understand the microbial ecology of the equine hindgut and the influence lactobacilli have on it.

Received 4 December 2013 **Accepted** 11 December 2013 **Published** 16 January 2014

Citation O'Donnell MM, Harris HMB, O'Toole PW, Ross RP. 2014. The genome of the predominant equine *Lactobacillus* species, *Lactobacillus equi*, is reflective of its lifestyle adaptations to an herbivorous host. Genome Announc. 2(1):e01155-13. doi:10.1128/genomeA.01155-13.

Address correspondence to R. Paul Ross, paul.ross@teagasc.ie.

Lactobacillus equi is a lactic acid bacterium found in the gastrointestinal tracts of horses (1) and which, along with *Lactobacillus hayakitensis* and *Lactobacillus equigenerosi*, has been identified as the predominant lactobacillus of the equine hindgut (2). *L. equi* was isolated from a fecal sample of a healthy Irish Thoroughbred racehorse. The sequence data were obtained using the Illumina HiSeq 2000 reversible dye terminator system (Macrogen, Seoul, South Korea), with read lengths of 101 bp. The HiSeq system paired-end sequencing strategy generated 36,133,338 reads (3,649,467,138 bp). Two hundred fifty-four scaffolds containing 273 contigs were assembled, corresponding to 34,664,201 reads from the HiSeq system (3,501,084,301 bp), which represents 1,608-fold genome coverage based on an estimated genome size of 2.19 Mb. The N_{50} score for the assembly estimating contig length was 39,802 bp. The draft *L. equi* genome includes 2,187,681 bases (G+C content of 39.16%). It comprises 2,263 predicted genes or coding sequences (CDS). Eight rRNA operons and 68 predicted tRNAs, representing all 20 amino acids, were identified in the genome. Functions were predicted for 76% of the *L. equi* chromosomal genes. The functional assignment of the predicted genes was completed using Metagene (3) to predict open reading frames (ORFs) and BLASTp to annotate them using the NCBI database (4).

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. AWWH00000000. The version described in this paper is version AWWH01000000.

ACKNOWLEDGMENT

This work was supported by a Principal Investigator Award (07/IN.1/B1780) from Science Foundation Ireland to P. W. O'Toole.

REFERENCES

1. **Morotomi M, Yuki N, Kado Y, Kushiro A, Shimazaki T, Watanabe K, Yuyama T.** 2002. *Lactobacillus equi* sp. nov., a predominant intestinal *Lactobacillus* species of the horse isolated from faeces of healthy horses. Int. J. Syst. Evol. Microbiol. **52:**211–214.
2. **Morita H, Nakano A, Shimazu M, Toh H, Nakajima F, Nagayama M, Hisamatsu S, Kato Y, Takagi M, Takami H, Akita H, Matsumoto M, Masaoka T, Murakami M.** 2009. *Lactobacillus hayakitensis*, L. *equigenerosi* and *L. equi*, predominant lactobacilli in the intestinal flora of healthy thoroughbreds. Anim. Sci. J. **80:**339–346. http://dx.doi.org/10.1111/j.1740-0929.2009.00633.x.
3. **Noguchi H, Park J, Takagi T.** 2006. Metagene: prokaryotic gene finding from environmental genome shotgun sequences. Nucleic Acids Res. **34:** 5623–5630. http://dx.doi.org/10.1093/nar/gkl723.
4. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410. http://dx.doi.org/10.1016/S0022-2836(05)80360-2.