CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Genomic diversity of Lactobacillus salivarius
Authors	Raftis, Emma J.
Publication date	2015
Original Citation	Raftis, E. J. 2015. Genomic diversity of Lactobacillus salivarius. PhD Thesis, University College Cork.
Type of publication	Doctoral thesis
Rights	© 2015, Emma J. Raftis. - http://creativecommons.org/licenses/by-nc-nd/3.0/
Download date	2026-11-15 02:12:52
Item downloaded from	https://hdl.handle.net/10468/3133

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh



GENOMIC DIVERSITY OF *Lactobacillus salivarius*

Candidate: Emma J. Raftis (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.

May 2015

Supervisors: Prof. Paul W. O'Toole & Prof. Douwe van Sinderen

Examiners: Prof. Peter Mullany (External) & Prof. Gerald Fitzgerald (Internal)

Head of Dept.: Prof. Gerald Fitzgerald

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.



Emma J. Raftis

Dedication

This thesis is dedicated to Seán Brown.

General Table of Contents

Abstract..... **5**
Chapter I: General Introduction..... **7**
Chapter II: Genomic Diversity of *Lactobacillus salivarius*..... **129**
Chapter III: Unusual Genome Complexity in *Lactobacillus salivarius* JCM1046..... **179**
Chapter IV: General Discussion and Future Perspectives..... **239**
Acknowledgements..... **258**

ABSTRACT

Lactobacillus salivarius is unusual among the lactobacilli due to its multireplicon genome architecture. The circular megaplasmids harboured by *L. salivarius* strains encode strain-specific traits for intestinal survival and probiotic activity. *L. salivarius* strains are increasingly being exploited for their probiotic properties in humans and animals. In terms of probiotic strain selection, it is important to have an understanding of the level of genomic diversity present in this species.

Comparative genomic hybridization (CGH) and multilocus sequence typing (MLST) were employed to assess the level of genomic diversity in *L. salivarius*. The well-characterised probiotic strains *L. salivarius* UCC118 was employed as a genetic reference strain. The group of test strains were chosen to reflect the range of habitats from which *L. salivarius* strains are frequently recovered, including human, animal, and environmental sources.

Strains of *L. salivarius* were found to be genetically diverse when compared to the UCC118 genome. The most conserved strains were human GIT isolates, while the greatest level of divergence were identified in animal associated isolates. MLST produced a better separation of the test strains according to their isolation origins, than that produced by CGH-based strain clustering. The exopolysaccharide (EPS) associated genes of *L. salivarius* strains were found to be highly divergent. The EPS-producing phenotype was found to be carbon-source dependent and inversely related to a strain's ability to produce a biofilm.

The genome of the porcine isolate *L. salivarius* JCM1046 was shown by sequencing to harbour four extrachromosomal replicons, a circular megaplasmid (pMP1046A), a putative chromid (pMP1046B), a linear megaplasmid (pLMP1046) and a smaller circular plasmid (pCTN1046) which contains an integrated Tn*916*-like element (Tn*6224*), which carries the tetracycline resistance gene *tetM*. pLMP1046 represents the first sequence of a linear plasmid in a *Lactobacillus* species. Dissemination of antibiotic resistance genes among species with food or probiotic-association is undesirable, and the identification of Tn*6224*-like elements in this species has implications for strain selection for probiotic applications.

In summary, this thesis used a comparative genomics approach to examine the level of genotypic diversity in *L. salivarius,* a species which contains probiotic strains. The genome sequence of strain JCM1046 provides additional insight into the spectrum of extrachromosomal replicons present in this species.

CHAPTER I

General Introduction

Chapter I Table of Contents

1.1 The biology and ecology of lactobacilli. 10
1.1.1 General introduction to lactobacilli. 10
1.1.2 Phylogeny and taxonomy 10
1.1.3 Natural habitats of lactobacilli 11
1.1.4 Safety assessment of lactobacilli 13
1.1.5 Industrial uses of lactobacilli. 14
1.1.6 Probiotic lactobacilli 15
1.1.7 Functions of exopolysaccharides produced by lactobacilli. 15
1.2 *Lactobacillus salivarius* 18
1.2.1 The *Lactobacillus salivarius* clade 18
1.2.2 Probiotic and strain specific properties of *L. salivarius*. 21
1.3 The importance of studying bacterial strain diversity 24
1.3.1 The importance of studying strain diversity. 24
1.3.2 Intraspecific genotyping of the *Lactobacillus* species 25
1.3.2.1 Population reconstruction using Multilocus Sequence Typing. 26
1.3.2.2 Whole genome comparisons utilizing comparative genome hybridization. 29
1.3.2.3 Next generation sequencing technologies: improvements and applications 31
1.4 Genome architecture of *Lactobacillus salivarius* and other bacteria. 39
1.4.1 Unusual genomic architectures in bacterial species. 39
1.4.2 Circular megaplasmids, defining characteristics. 42
1.4.2.1 Distribution and encoded properties of circular megaplasmids. 42
1.4.2.2 Evolutionary origin. 49
1.4.3 Megaplasmids in *Lactobacillus salivarius* 52
1.4.3.1 Contribution of pMP118 to the lifestyle of *L. salivarius* UCC118. 52
1.4.3.2 Discovery of pMP118. 52
1.4.3.3 Replication region of pMP118. 53
1.4.3.4 GC content and coding capacity. 54
1.4.3.6 Pseudogenes. 55
1.4.3.7 Contribution of pMP118 to genotype and phenotype. 55
1.4.3.8 Amino acid metabolism. 56
1.4.3.9 Purine and pyrimidine anabolism. 57
1.4.3.10 Acquisition of core chromosomal genes by pMP118. 57
1.4.3.11 Role of pMP118 in carbohydrate metabolism. 58
1.4.3.12 Sensing and regulation. 59
1.4.3.13 Host interaction. 61
1.4.3.14 Bile salt hydrolase (BSH) activity. 61
1.4.3.15 Surface proteins. 61
1.4.3.16 Microbe-Microbe interactions 62
1.4.3.17 Megaplasmid distribution in L. salivarius and other species. 63
1.4.3.18 Dissemination mechanism. 63
1.4.3.19 Concluding remarks. 64
1.4.4 Linear plasmids 65
1.4.4.1 Distribution of linear plasmids in prokaryotes. 65
1.4.4.2 Structural organization of linear plasmids. 66

1.4.4.3 Linear plasmid replication mechanisms. 68
1.4.4.4 Effects of linear plasmids on host adaptation. 69
1.4.4.5 Biotechnological applications. 70
1.5 Conjugative transposons: Tn*916* as a model 75
1.5.1 Introduction to conjugative transposons 75
1.5.2 Core genome. 77
1.5.3 Tn*916*-like elements 78
1.5.3.1 Distribution of Tn*916*-like elements. 78
1.5.3.2 Transposition mechanism of Tn*916*-like elements. 78
1.5.3.3 Antibiotic resistance and accessory genes. 80
1.6 Summary of Thesis contents 82
1.6.1 Aims and Objectives 82
1.7 References 83

1.1 The biology and ecology of lactobacilli.

1.1.1 General introduction to lactobacilli.

The genus *Lactobacillus* encompasses a large heterogeneous group of Gram-positive, low G+C, nonsporulating bacteria which form the most numerous group in the *Lactobacteriaceae* (Felis & Dellaglio, 2007). Lactobacilli are sub-dominant members of the functionally defined group of bacteria known as the lactic acid bacteria (LAB) which form lactic acid as the main end product of carbohydrate metabolism. Currently members of the genus *Lactobacillus* are usually considered to have a fermentative metabolism (Hammes & Vogel, 1995) and are subdivided into three groups based on their fermentation characteristics: homofermentative, facultative heterofermentative, and obligate heterofermentative, as outlined by (Hammes & Vogel, 1995; Pot *et al.*, 1994).

1.1.2 Phylogeny and taxonomy

The genus *Lactobacillus* currently contains 152 recognised species (Salvetti *et al.*, 2012) and belongs to the phylum *Firmicutes*, class *Bacilli*, order *Lactobacillales*, family *Lactobacillaceae*. The closest phylogenetic relatives of the lactobacilli are the genera *Paralactobacillus* and *Pediococcus.* Species of *Lactobacillus* are characterized by an extraordinary level of genotypic, phenotypic and ecological diversity (Berger *et al.*, 2007; Felis & Dellaglio, 2007; Makarova & Koonin, 2007) which is mirrored in the taxonomical diversity of the genus. A lack of correlation between the phylogenetic structure and metabolic properties of *Lactobacillus* species remains a major discrepancy in the taxonomy of the genus (Salvetti *et al.*, 2012), which is generally considered to be unsatisfactory (Canchaya *et al.*, 2006; Claesson *et al.*, 2008; Felis & Dellaglio, 2007; Makarova *et al.*, 2006; Salvetti *et al.*, 2012). This is highlighted by the fact that most lactobacilli are in breach of the traditional definition of a genus, when sequence similarities are based on the 16S rRNA gene (Claesson *et al.*, 2008). Additionally, despite being described as low G+C bacteria, the observed range is from 33 mol % to an upper limit of 59.2 mol % GC (Cai *et al.*, 2012); twice the span which is generally accepted for a well defined genus (Schleifer & Ludwig, 1995).

Revision of the phylogenetic structure of the Family *Lactobacillaceae* is ongoing (Claesson *et al.*, 2008; Felis & Dellaglio, 2007; Hammes & Hertel, 2006;

Pot *et al.*, 1994) as novel species are identified. Most recently, Salvetti *et al.* (2012) employed the 16s RNA gene, the most commonly used bacterial phylogenetic marker, to determine the phylogenetic structure of the Family *Lactobacillaceae* (Salvetti *et al.*, 2012). The Family is divided into 15 groups of 3 or more species in addition to 4 couples and 10 single lines of descent (Salvetti *et al.*, 2012). The phylogenetic structure inferred by Felis and Dellaglio (2007) (Felis & Dellaglio, 2007) was confirmed and it was found that rather than forming distinct clades (Salvetti *et al.*, 2012), the *Lactobacillus* and *Pediococcus* genera are paraphyletic and intermixed within a 16S rRNA gene-based tree (Felis & Dellaglio, 2007). Split decomposition analysis of the 16S genes indicates that the Family *Lactobacillaceae* are likely to have a complex evolutionary history, having undergone recombination and or horizontal gene transfer rather than speciation events alone (Salvetti *et al.*, 2012).

Despite the 16S rRNA gene being by far the most common single-gene phylogenetic marker employed for prokaryotes; the *groEL* gene has previously been shown to be a more robust and suitable phylogenetic marker for unsequenced members of the genus *Lactobacillus* and should be considered as an alternative to the 16S rRNA gene in large-scale phylogenetic studies (Salvetti *et al.*, 2012). Phylogenetic analyses and reconstruction of ancestral gene sets have revealed several trends in the evolution of *Lactobacillales* (Makarova *et al.*, 2006; Makarova & Koonin, 2007). For example, members of this order have undergone extensive gene loss together with key gene acquisitions *via* horizontal gene transfer during their evolution, which is thought to have been heavily influenced by their environmental niches (Azcarate-Peril *et al.*, 2008; Boekhorst *et al.*, 2004; Callanan *et al.*, 2008; Makarova *et al.*, 2006).

1.1.3 Natural habitats of lactobacilli

Lactobacilli are nutritionally fastidious (Kandler & Weiss, 1986), which is reflected in the diverse, carbon-rich habitats in which they are found (Hammes & Vogel, 1995; Marteau *et al.*, 2001; Tannock, 2004). Different species have adapted to grow under a wide range of environmental conditions and a detailed overview of the various species associated with individual environmental, food, human, and animal habitats is described in Hammes & Hertel (2006) (Hammes & Hertel, 2006). Lactobacilli are often used as starter bacteria and are frequently recovered from dairy

products, fermented drinks, and foodstuffs including silage, olives, pickles, and preserved meats etc. (Hammes & Hertel, 2006). In addition, they have been isolated from a range of environmental sources including grains, soil, sewage, and plant material (Hammes & Hertel, 2006). Although food-associated lactobacilli can often survive gut-passage, but they do not persist unless they are consumed regularly (Tannock *et al.*, 2000).

Lactobacilli are often associated with the mucosal surfaces of humans and animals (Tannock, 2004) and form part of the indigenous microflora of the human oral cavity (Dal Bello & Hertel, 2006), gastrointestinal tract (GIT) (Tannock, 1990) and the female urogenital tract (Redondo-Lopez *et al.*, 1990). The number and species of lactobacilli that inhabit the GIT of humans and animals is dependent on a number of factors, including the species of the host, the health and age of the host, and the location of the lactobacilli within the GIT (Tannock, 1999; Tannock, 2004; Tannock, 2005).

Lactobacilli have been shown to be present in human saliva in numbers exceeding 10^5 CFU per ml, thus the oral cavity can act as a natural reservoir for gastrointestinal lactobacilli (Dal Bello & Hertel, 2006). When examined using cultivation techniques, lactobacilli are among the most common bacteria found in the stomach, duodenum, and jejunum of humans (Reuter, 2001), however molecular analyses indicate that lactobacilli form a minor proportion of the human GIT microbiota (Walter, 2008). Lactobacilli are found in the gastro-intestinal tracts soon after birth and are estimated to be present in the oral cavity (10^3-10^4 cfu/g), ileum (10^3-10^7 cfu/g), and the colon (10^4-10^8 cfu/g) in a healthy adult human. However, the composition and number of lactobacilli that form part of the human fecal and oral microbiota has been shown to fluctuate over time and from person to person (Codling *et al.*, 2010; Dal Bello & Hertel, 2006; Tannock *et al.*, 2000). Lactobacilli are commonly identified in biopsy samples from the stomach, small intestine, and colon but in variable and usually at low numbers (Tannock, 1990; Walter, 2005). *L. gasseri*, *L. crispatus*, *L. plantarum*, *L. salivarius*, *L. brevis*, *L. rhamnosus*, *L. paracasei*, and *L. vaginalis*, predominate in the oral cavity of humans and are also frequently isolated from human feces (Walter, 2008). The following species have also been shown to be associated with the human GIT, *L. fermentum*, *L. casei*, *L. rhamnosus*, *L. johnsonii*, *L. plantarum*, *L. delbrueckii*, *L. curvatus*, and *L. sakei* (Heilig *et al.*, 2002; Walter, 2008).

It has been proposed that only a small number of *Lactobacillus* species are truly autochthonous to the mammalian intestinal tract, but are in fact allochthonous and are derived from the oral cavity, consumed foodstuffs and other proximal parts of the GIT (Walter, 2008). For example, several *Lactobacillus* species, including *L. salivarius* and *L. gasseri*, may be allochthonous to the human intestinal tract but autochthonous to the oral cavity (Reuter, 2001).

The infant micrbiota is considered highly unstable when compared to the adult microbiota and contain lactobacilli in variable amounts, 10^5 CFU/g of feces in neonates and between 10^6 to 10^8 CFU/g of feces in inflants older than 1 month (Mackie *et al.*, 1999). Sequences retrieved from infant feces group with the species *L. salivarius, L. rhamnosus,* and *L. paracasei* (Heilig *et al.*, 2002).

In contrast to the GIT, lactobacilli are the dominant members of the healthy microbiota of the female urogenital tract (Redondo-Lopez *et al.*, 1990). The female urogenital tract is a dynamic ecosystem with the prevalence of each bacterial species varying with repetitive temporal sampling, pregnancy, and with stage in the menstrual cycle (Redondo-Lopez *et al.*, 1990). However, lactobacilli have been shown to dominate the microbiota of a female urogenital tract (Redondo-Lopez *et al.*, 1990), with the most common species found being *L. crispatus*, *L. gasseri*, *L. iners and L. jensenii* (Burton *et al.*, 2003; Pavlova *et al.*, 2002). A healthy and stable lactobacilli population plays an important role in the control of the vaginal microflora, often through antagonistic properties and production of metabolites which help control opportunistic pathogens (Pavlova *et al.*, 2002; Redondo-Lopez *et al.*, 1990) and prevent bacterial vaginosis (Brook & Frazier, 1993; Homayouni *et al.*, 2014).

Many of the *Lactobacillus* species present in the GIT of mice, rats, pigs, and chickens are considered to be autochthonous, as can they can be cultured in large numbers, are present in almost all animals and form stable populations throughout the life of the host (Tannock, 1992).

1.1.4 Safety assessment of lactobacilli

Lactobacilli are frequently described as having a "generally regarded as safe (GRAS)" status. However, they are occasionally found in clinical samples (Brook & Frazier, 1993) and have been associated with cases of endocarditis, bacteremia and peritonitis (Cannon *et al.*, 2005; Salvana & Frank, 2006), although these infections

are often associated with immunocompromised patients. Considering the huge numbers of lactobacilli that are ingested daily, isolated cases of lactobacillemia remained relatively rare (Bernardeau *et al.*, 2008). A study by Bernardeau *et al.* recommends that "the safety requirements for *Lactobacillus* strains of technological interest should be limited to an antibiotic profile and a determination of whether any antibiotic resistance(s) of medical interest is detected (or are) transferable" (Bernardeau *et al.*, 2008). Safety assessments of individual probiotic strains are often carried out in detail (Abe *et al.*, 2010; Lara-Villoslada *et al.*, 2007; Zhang *et al.*, 2013) and an overview of the safety profile of probiotic bacteria is outlined in Lara-Villoslada *et al.* (2010) and Sanders *et al.* (2010) (Lara-Villoslada *et al.*, 2010; Sanders *et al.*, 2010).

1.1.5 Industrial uses of lactobacilli.

Lactobacilli have been used historically as natural acidifiers of fermented products. They are employed extensively in food preservation, as starters or complementary cultures for dairy products (cheese, yoghurt and fermented milk), and play a essential role in the production of fermented vegetables, fish, meats, sourdough bread, and as silage inoculants (Giraffa *et al.*, 2010; Hammes & Hertel, 2006; Leroy & De Vuyst, 2004; Leroy *et al.*, 2006; Seegers, 2002; Stiles, 1996; Tannock, 2004).

In more recent times lactobacilli have attracted the interest due to their expanding biotechnological applications and have been developed as probiotics, functional foods or as vaccine carriers because of their therapeutic and prophylactic attributes (Forssten *et al.*, 2011; Gaspar *et al.*, 2013). Lactobacilli have been employed as cell factories for the production of metabolites including lactic acid, β-galactosidase and lantibiotics through the usage of modern fermentation technology (Gaspar *et al.*, 2013).

Given their industrial importance and economic value the metabolism and genetic of lactobacilli have been studied intensively. This data has been crucial in the selection and generation of robust starter cultures with desirable phenotypic traits as well as providing insight into the molecular basis of industrially important traits including those that govern sugar metabolism, flavor formation, and stress responses, adaptation and interactions (Chamba & Irlinger., 2004). Starter cultures can also contribute to the inherent nutrition of these foods as well as contributing to flavour

enhancement and texture. Production of aromatic compounds and exopolysaccharides are relevant for the quality and nutritional value of the end product (Leroy & De Vuyst, 2004) Additionally, lactobacilli have been exploited for their ability to produce antimicrobial substances (Balciunas *et al.*, 2013), acids and enzymes (Gaspar *et al.*, 2013). Bacteriocin producing lactobacilli are thought to have a protective effect on fermented meats, fermented olives, and dairy products (Leroy & De Vuyst, 2004).

1.1.6 Probiotic lactobacilli

Lactobacilli are regularly consumed as probiotics, which are defined as "Living micro-organisms which upon ingestion in certain numbers exert health benefits beyond inherent general nutrition" (reviewed in (Guarner & Schaafsma, 1998; Klaenhammer, 2000). Numerous probiotic studies with different strains of *Lactobacillus* have been performed both *in vitro* and *in vivo* to investigate their probiotic potential. Studies have shown that probiotics have the ability to benefit the host (Dunne *et al.*, 1999) through a wide variety of mechanisms (Lebeer *et al.*, 2008), which are often described as being species- and strain-specific (Lebeer *et al.*, 2008; Maassen *et al.*, 2000; Remus *et al.*, 2011; Servin, 2004).

Desirable characteristics and proposed mechanisms of action of probiotics have been thoroughly described elsewhere (Lebeer *et al.*, 2008; Ljungh & Wadstrom, 2006; Wells, 2011) as have their effect on the host (Servin & Coconnier, 2003; Servin, 2004). Clinical applications of probiotics are often most effective when treating diseases associated with dysbiosis of the microbiota (Dotterud *et al.*, 2010; McLoughlin & Mills, 2011). Lactoacilli have demonstrated significant and promising results in the treatment of acute infectious diarrhoea and in the prevention of antibiotic-associated diarrhoea in human clinical trials (Sazawal *et al.*, 2006).

1.1.7 Functions of exopolysaccharides produced by lactobacilli.

Like many other bacteria, lactobacilli produce several forms of polysaccharides which are classified based on their placement within or proximity to the cell (see review (Chapot-Chartier & Kulakauskas, 2014). Exopolysaccharides (EPSs) are long-chain polysaccharides which are excreted outside the cell wall of a variety of microorganisms including many gram positive and gram negative bacteria during growth. EPS can form an adherent cohesive polysaccharide layer which encapsulates the bacteria and is often covalently bound to cell surface, alternatively

it can be released into the surrounding environment while remaining loosely attached or unbound to the bacterial cell (Ruas-Madiedo & de los Reyes-Gavilan, 2005). There is some debate with regards to the definition of exopolysaccharides as the various types of polysaccharides are often difficult to distinguish experimentally, and the nomenclature of these biopolymers is not uniform within the literature. For example, EPS can also refer to extracellular polysaccharide while bound EPS is sometimes referred to as capsular ploysaccharides (CPS). For the purposes of this thesis, capsular EPS will be referred to as bound EPS (EPS-b) and EPS that is excreted into the environment will be referred to as released EPS (EPS-r).

Research into microbial EPS is driven by the potential applications of EPS in a range of sectors including food and biotechnology industries, cosmetics, or medicine and human health (Freitas *et al.*, 2011). Exopolysaccharides are increasingly being utilised as components of functional foods as they interact with the immune system of consumers (Vinderola *et al.*, 2006) in a strain-dependant manner and have been shown to act as prebiotics in the GIT (Bello *et al.*, 2001). The composition and amount of EPS produced by a bacterium are genetically determined traits, and the quantity of EPS produce is influced by media components and cultivation conditions (Freitas *et al.*, 2011). Approximately 30 species of lactobacilli are described as EPS producers. Generally the yield of production is under 1 g/L for homopolysaccharides (HoPSs) when culture conditions are not optimized and even less for the majority of heteropolysaccharides (HePSs) (Badel *et al.*, 2011).

EPS of some bacterial species have been reported to act as virulence factors in several infectious diseases (Koo *et al.*, 2010). However, EPS produced by probiotic strains of bacteria are increasingly being examined for their health promoting properties including their capacity to interact with and modulate the host immune response (Hidalgo-Cantabrana *et al.*, 2012; Lebeer *et al.*, 2011). The exopolysaccharides produced by *Lactobacillus* and other genera of lactic acid bacteria (LAB) have been extensively reviewed in terms of their structure (De Vuyst & Degeest, 1999; Laws *et al.*, 2001), biosynthesis (Boels *et al.*, 2001; Jolly & Stingele, 2001; Ruas-Madiedo, 2014), and their industrial uses (Freitas *et al.*, 2011; Jolly *et al.*, 2002; Ruas-Madiedo *et al.*, 2002) and their health promoting effects (Patel & Prajapati, 2012; Ruas-Madiedo *et al.*, 2010; Ruas-Madiedo, 2014). There have been an increasing number of reports claiming health benefits associated with the consumption of EPS derived from LAB (Patten *et al.*, 2014). More recently

Badel *et al*. (2011) reviewed the properties associated with the *Lactobacillus*-derived EPS (Badel *et al.*, 2011). The potential role that *Lactobacillus* EPS production plays in host-interaction, biofilm-formation and colonisation has been described (Lebeer *et al.*, 2008; Lebeer *et al.*, 2010; Lebeer *et al.*, 2011), and demonstrated (Tannock *et al.*, 2005; Vélez *et al.*, 2010; Walter *et al.*, 2008) in some detail, but has yet to be fully explained.

1.2 *Lactobacillus salivarius*

1.2.1 The *Lactobacillus salivarius* clade

The *Lactobacillus salivarius* clade is a cohesive, peripheral clade within the 16S rRNA gene phylogeny of the *Lactobacillus* genus. *L. salivarius* is the most thoroughly characterised species of this clade, and thus gives the clade its name (Felis & Dellaglio, 2007; Hammes & Hertel, 2006; Salvetti *et al.*, 2012). The GC content of the *L. salivarius* clade ranges from 32 to 44 mol% and currently comprises 25 species, 11 of which were added to this clade in recent years (Salvetti *et al.*, 2012) (Table 1.1). There is a strong correlation between phylogenetic groupings in the *L. salivarius* clade and the source from which the species was isolated (Salvetti *et al.*, 2012). 16 of the 25 species in this clade are homofermenters, while 9 of the species are facultative heterofermenters. Each member produces L(+)-lactic acid or both L(+)- and D(-)-lactic acids, but no one species produces the D(-)-isomer exclusively. The peptidoglycan types of these species are Lys-D-Asp and meso-Dpm-direct. The majority of the motile species (*L. agilis, L. aquaticus, L. capillatus, L. ghanensis, L. mali, L. nagelii, L. oeni, L. ruminis, L. sucicola, L. satsumensis, L. uvarum and L. vini*) of the *Lactobacillus* genus are clustered within this clade. Comparative genomic analysis of *L. ruminis* (Forde *et al.*, 2011), has facilitated the determination of the genetic basis of motility in the *L. salivarius* clade (Neville & O'Toole, 2010).

Table 1.1 Species of the *Lactobacillus salivarius* phylogenetic clade, modified from (Neville *et al.*, 2012)

Species	Source	Reference
Lactobacillus apodemi	Feces of Japanese large wood mouse	(Osawa *et al.*, 2006)
Lactobacillus animalis	Dental plaque of primates	(Dent & Williams, 1982)
Lactobacillus aviarius	Chicken intestines	(Fujisawa *et al.*, 1984)
Lactobacillus ceti	Lungs of beaked whale	(Vela *et al.*, 2008)
Lactobacillus equi	Horse feces	(Morotomi *et al.*, 2002)
Lactobacillus hayakitensis	Horse intestines	(Morita *et al.*, 2007)
Lactobacillus murinus	GI tract of rat and mouse	(Hemme D *et al.*, 1980)
Lactobacillus saerimneri	Pig feces	(Pedersen & Roos, 2004)
Lactobacillus ruminis	Bovine rumen	(Reuter, 2001; Sharpe *et al.*, 1973; Tannock *et al.*, 2000)
Lactobacillus salivarius	Oral cavity of human	(Rogosa *et al.*, 1953)
Lactobacillus agilis	Piglets, humans, pigeons, sewage	(Baele *et al.*, 2001; Kandler & Weiss, 1986)
Lactobacillus aquaticus	Freshwater pond	(Mañes-Lázaro *et al.*, 2009b)
Lactobacillus caconum	Cocoa fermentations	(De Bruyne *et al.*, 2009)
Lactobacillus ghanensis	Cocoa fermentations	(Nielsen *et al.*, 2007)
Lactobacillus hordei	Malted barley	(Rouse *et al.*, 2008)
Lactobacillus sucicola	Sap of oak trees	(Irisawa & Okada, 2009)
Lactobacillus mali	Pressed apple juice	(Carr & Davies, 1970)
Lactobacillus nagelii	Fermenting grape wine	(Edwards *et al.*, 2000)
Lactobacillus oeni	Bobal grape wines	(Mañes-Lázaro *et al.*, 2009a)
Lactobacillus satsumensis	Shochu mashes (by-product of distilled spirit production)	(Endo & Okada, 2005)
Lactobacillus uvarum	Fermenting grape musts	(Mañes-Lázaro *et al.*, 2008)
Lactobacillus vini	Fermenting grape musts	(Rodas *et al.*, 2006)
Lactobacillus acidipiscis	Fermented fish	(Tanasupawat *et al.*, 2000)
Lactobacillus capillatus	Fermented stinky tofu brine	(Chao *et al.*, 2008)
Lactobacillus pobuzihi	Fermented cummingcordia	(Chen *et al.*, 2009)

L. salivarius forms part of the indigenous microbiota of the gastrointestinal tract (GIT) and the oral cavity of humans and hamsters (Molin *et al.*, 1993a; Rogosa *et al.*, 1953; Wall *et al.*, 2007) and also from human breast milk (Martín *et al.*, 2006). This species has also been isolated from the intestinal tract of swine and chickens (Abbas *et al.*, 2007; Casey *et al.*, 2004; Mitsuoka, 1969). *Lactobacillus salivarius* was first isolated by Rogosa *et al.* in 1953 as part of a survey of human oral lactobacilli (Rogosa *et al.*, 1953). In the original characterization of *L. salivarius*, two varieties of were described, '*Lactobacillus salivarius* var. *salivarius*' and *Lactobacillus salivarius* var. *salicinius*' which were distinguished by their respective fermentation profiles. Both subspecies were described as homofermentative, producing lactic acid as the primary product of hexose carbohydrate fermentation. *L. salivarius* var. *salivarius*, characterized by the ability to ferment *inter alia*, rhamnose, but not salicin and *L. salivarius* var. *salicinius*, characterized by ability to ferment *inter alia* salicin, but not rhamnose (Rogosa *et al.*, 1953). In order to re-examine the infraspecific classification of *L. salivarius*, Li *et al* studied the relatedness of strains from both subspecies of *L. salivarius* (Li *et al.*, 2006). A diverse panel of *L. salivarius* strains, isolated from multiple human, animal and environmental sources gathered from various geographical regions were included in the study. An examination of the carbohydrate fermentation profiles together with a comparison of the rRNA gene sequences determined that a subdivision of the species was unwarranted. Sequencing of the 16S-23S rRNA intergenic spacer region (ISR) of the type strains of the two subspecies revealed 100% identity between them and even the sequences of ISR flanking the 23S rRNA genes were found to be identical (Song *et al.*, 2000). However, the *groEL* gene phylogeny of *L. salivarius* was found to be sufficiently discriminatory to resolve the *L. salivarius* strains into distinct clades and clusters but no concordance was found between the *groEL*-based phylogeny with phenotypic traits. Polyphasic analysis of this species by Li and colleagues indicated that intraspecies differentiation could not be justified on the basis of either phenotypic or molecular typing (Li *et al.*, 2006) and it was suggested that its use be discontinued in the literature. A revised version (Li *et al.*, 2006) of the original *L. salivarius* species description (Rogosa *et al.*, 1953) was produced, an abridged form of this is outlined below.

Cells of *L. salivarius* owe their name to the salivary properties of the oral cavity from which it was originally isolated (Rogosa *et al.*, 1953) and are

predominately isolated from the mouth and intestinal tract of humans and hamsters and the intestinal tract of chicken and swine. They are Gram-positive, non-motile, non-spore forming cells with a DNA G+C content is 33–36 mol%. They are catalase-negative, nitrate-reduction-negative rods with rounded ends of 0.6-0.9 x 1.5-5 μm, occuring both singly and in chains of varying length, slightly bent and swollen on one end. All strains grow at 45 ^{o}C, but not at 15 ^{o}C, with an optimal growth temperature of 37 ^{o}C. and growth is stimulated in the presence of 5 % CO_2. Colonies are convex milky-white or light-brown colour on MRS agar, with smooth edges and smooth surfaces. The species contains both homofermentative and facultatively heterofermentative strains. All strains ferment glucose, fructose, mannose, mannitol, N-acetylglucosamine and sucrose. Most strains produce L- lactic acid from available hexoses by homofermentation. Some strains can also ferment ribose and produce heterofermentative end products (lactic acid, acetic acid and ethanol) from hexoses. The type strains of the two subspecies do not ferment pentoses or pentitols.

Megaplasmids are a general feature of species *L. salivarius,* and to date have been found to range in size from 100 kb to 380 kb (discussed in detail below). Additionally, some strains also harbour smaller plasmids (Fang *et al.*, 2008). 16S rRNA gene sequences or 16S–23S rRNA intergenic spacer region sequences do not provide reliable discrimination between strains. The species can be separated into two major clades, based on *groEL* phylogeny. Each clade contains one of the type strains of the two previously recognized subspecies, DSM 20554T (=ATCC 11742T=JCM 1150T) (the type strain of *L. salivarius* subsp. *salicinius*) and DSM 20555T (=ATCC 11741T=JCM 1231T) (the type strain of *L. salivarius* and of *L. salivarius* subsp. *salivarius*).

1.2.2 Probiotic and strain specific properties of *L. salivarius*.

The first genome of this species to be published (Claesson *et al.*, 2006; Flynn *et al.*, 2002) was that of the well-characterised strain *L. salivarius* UCC118 (Claesson *et al.*, 2006; Corr *et al.*, 2007; Fang *et al.*, 2009; Li *et al.*, 2006; Li *et al.*, 2007; van Pijkeren *et al.*, 2006) which was originally isolated from the terminal ileum of a healthy human patient (Dunne *et al.*, 1999). Genome sequencing of the strain revealed a 1.83 Mb chromosome, a 242 kb megaplasmid (pMP118) and two smaller plasmids (Claesson *et al.*, 2006; Flynn *et al.*, 2002). Prior to their discovery in strain UCC118, megaplasmids had not previously been verified in LAB or

intestinal lactobacilli. A detailed description of pMP118 and its contribution to the biology of *L. salivarius* are outlined in sections 1.4.3.

At least 10 additional *L. salivarius* genomes have been sequenced since that of strain UCC118. Additionally, of the megaplasmids of *L. salivarius* discovered subsequently (Li *et al.*, 2007), three of these have been sequenced (strains CECT 5713 (Jimenez *et al.*, 2010) NIAS840 (Ham *et al.*, 2011) and SMXD51 (Kergourlay *et al.*, 2012) with two being finished to a draft quality status (Cho *et al.*, 2011). The strain-specific properties of *L. salivarius* are have been previously studied in relation to their immunomodulatory activity and probiotic traits (see review (MacKenzie *et al.*, 2014; Neville & O'Toole, 2010), plasmid profile (Li *et al.*, 2007), bile resistance (Fang *et al.*, 2008; Fang *et al.*, 2009) and bacteriocin production (Li *et al.*, 2007), capacity for intestinal barrier protection (O'Callaghan *et al.*, 2012) and their antimicrobial activity (Ryan *et al.*, 2009). This introduction does not aim to cover all aspects of the strain-specific characteristics of *L. salivarius* but will briefly outline some current areas of interest with regard to their probiotic features and safety.

The probiotic potential of strain UCC118 has been extensively studied and there have been a number of beneficial effects reported, including anti-inflammatory (Sibartie *et al.*, 2009), anti-infectivity (Corr *et al.*, 2007), amelioration of the effects of induced rheumatoid arthritis (RA) in mice (Sheil *et al.*, 2004), and reduction of *Helicobacter pylori*-induced IL-8 secretion by AGS gastric cells (Ryan *et al.*, 2009). Strains of *L. salivarius*, including the recently sequenced strain SMXD51 (Kergourlay *et al.*, 2012), have demonstrated antimicrobial activity against pathogens including *Campylobacter* (Messaoudi, 2011), *Helicobacter pylori* (Ryan *et al.*, 2008) and *Listeria monocytogenes* (Corr *et al.*, 2007). Strains of *L. salivarius* are well characterized bacteriocin producers (Dobson *et al.*, 2012) and can produce unmodified bacteriocins of the sub-classes IIa, IIb and IId (O'Shea *et al.*, 2011). It has recently been demonstrated that there is a high level of sequence diversity between the bacteriocin-producing loci of human and porcine intestinal orgins and that production of multiple bacteriocins from a single locus is possible in these strains (O'Shea *et al.*, 2011). A recently updated review of the probiotic action of *L. salivarius* strains, with particular attention paid to their bacteriocin activity, is outlined in detail in the following review by Messaoudi *et al.* (2013) (Messaoudi *et al.*, 2013).

The human isolate, *Lactobacillus salivarius* CECT5713 has been demonstrated to be safe in animal models and is considered to be safe for human consumption (Lara-Villoslada *et al.*, 2007; Olivares *et al.*, 2006) and fulfils the safety criteria defined for probiotic bacteria (Martín *et al.*, 2006) and it has been demonstrated to be safe in animal models (Lara-Villoslada *et al.*, 2007; Olivares *et al.*, 2006). Indeed, it has been demonstrated that, a dose 500–10, 000 times higher (per gram of body weight) than what is normally consumed by humans, *L. salivarius* CECT5713 was not able to translocate from the gut to other tissues (Lara-Villoslada *et al.*, 2007). The probiotic isolate *L. salivarius* SMXD51 was applied to Caco-2/TC7 monolayers for 24 h and the cytotoxicity was evaluated by levels of lactate dehydrogenase (LDH) production and microscopic observation. The results showed that the integrity of the monolayer was fully preserved (Messaoudi *et al.*, 2012). Oral administration of the probiotic REN strain to BALB/c mice had no adverse effects on body weight, food intake, hematological and serum biochemical parameters, and showed no treatment-associated bacterial translocation and was therefore deemed to be safe for human consumption (Zhang *et al.*, 2013). Antibiotic susceptibility profiles showed that the strains of *L. salivarius* tested were sensitive to the majority of the antibiotics used in the study (Kirtzalidou *et al.*, 2011).

1.3 The importance of studying bacterial strain diversity

1.3.1 The importance of studying strain diversity.

The concept of bacterial species, as a form of taxonomic unit remains a subject of debate (Stackebrandt E *et al.*, 2002). However, the ability to definitively identify and describe bacteria to the strain level is of critical importance as it facilitates the tracking of strains with industrially relevant properties such as probiotic, sensorial or antimicrobial attributes. The dynamic nature of bacterial genomes is well recognised and generates inter- and intra-specific genotypic and phenotypic variation (Taboada *et al.*, 2008). Gene loss and metabolic simplification is often the cause of sequence diversity (Makarova & Koonin, 2007), which also arises from horizontal gene transfer (HGT) of genes from distantly and closely related organisms (Nicolas *et al.*, 2007; Ochman *et al.*, 2000). The mobilome of bacteria includes genes and regions which are frequently transferred laterally between bacterial species and strains, and includes mobile genetic elements such as transposons, phage-related genes, plasmids, and other "jumping genes" and has been described as “the driver” of prokaryotic evolution (Siefert, 2009).

Genetic variation gives rise to industrially (Berger *et al.*, 2007) and medically relevant (Fukiya *et al.*, 2004) phenotypic differences in lactobacilli and many industrial processes and medical treatment strategies are based on these phenotypic differences. The need to characterize strain-specific differences has become increasingly important when producing and guarding intellectual property. Biotechnology industries have a clear need to monitor the use of patented strains in addition to tracking genomic changes within the isolates. From a safety aspect, it is crucial to be able to monitor the genetic stability and evolution of clinical isolates and industrial strains and to ensure, as far as it is possible, that candidate probiotics are free from potentially pathogenic traits (Donohue & Salminen., 1996; Klein *et al.*, 1998).

Species and strain-specific properties contribute to the economical and medical importance of lactobacilli species; therefore the discrimination of lactobacilli to the strain level is of critical importance. The mechanisms through which lactobacilli exert probiotic effects are known to be species and strain specific characteristics (Lebeer *et al.*, 2008; Lee *et al.*, 2013; Maassen *et al.*, 2000; Ryan *et*

al., 2009; Servin, 2004). For example, the specific anti-infective activities of *Lactobacillus* strains in the human intestinal microbiota has recently been reviewed (Liévin-Le Moal & Servin, 2014). Additionally, Fernandez *et al.* (2011) determined that the protective effect of *Lactobacillus* peptidoglycan (PG) is strain-specific and is linked to the induction of diverse immune regulatory pathways (Fernandez *et al.*, 2011). *L. salivarius* Ls33 was found to have an IL-10 mediated protective effect in a chemically induced colitis mouse model, whereas *Lactobacillus acidophilus* NCFM-isolated PG lacked this capacity (Macho Fernandez *et al.*, 2011). A detailed analysis of their peptidoglycan layers revealed that subtle differences in their structures proved to be pivotal for the strain-specific immunomodulatory properties observed (Fernandez *et al.*, 2011; Macho Fernandez *et al.*, 2011).

1.3.2 Intraspecific genotyping of the *Lactobacillus* species

Identification of lactobacilli has been traditionally based primarily on their morphology, selective culturing techniques, and their biochemical and metabolic profiles (Kandler & Weiss, 1986), however, the identification of several *Lactobacillus* species by biochemical methods alone is not reliable (Charteris *et al.*, 1997; Klein *et al.*, 1998; Schleifer *et al.*, 1995). Although these methods are still used, particularly for the routine screening of samples and specimens for the presence of *Lactobacillus* species, they have now been superseded in both accuracy and convenience by molecular techniques. In more recent times, a polyphasic approach has been employed to explore the genetic and metabolic characteristics of the *Lactobacillus* genus (Cai *et al.*, 2007; Li *et al.*, 2006) with comparative genomics leading the charge in this era of high-throughput sequencing technologies (Canchaya *et al.*, 2006; Kok *et al.*, 2005; Makarova *et al.*, 2006).

A wide array of molecular methods are available to investigate the epidemiology of bacteria (reviewed by (Li *et al.*, 2009)). Generally, typing methods can be classified into three main categories: DNA banding pattern, DNA hybridization-based methods, and DNA sequencing. DNA banding pattern-based genotyping methods discriminate the studied strains based on differences in the size of the DNA bands (fragments) generated by amplification of genomic DNA or by cleavage of DNA using restriction enzymes (REs). Genotypic typing methods are typically PCR-based techniques including RAPD (randomly amplified polymorphic) analysis (Welsh & McClelland, 1990), whereby short arbitrary sequences are used as

primers in PCR, which yields strain-specific amplification product patterns, which can be used to distinguish between strains. Variations in the restriction profiles of strains has been used as the "golden standard" of restriction enzyme analysis (e.g., pulsed-field gel electrophoresis [PFGE] and ribotyping) (Vandamme *et al.*, 1996). In PFGE rare-cutting enzymes are used to digest genomic DNA and the resulting large genomic fragments are separated using gel electrophoresis (Schwartz & Cantor, 1984), while ribotyping, utilises rRNA genes and/or their spacer regions as probes to hybridize to restricted fragments of genomic DNA (Stull *et al.*, 1988). DNA hybridization-based methods are mainly referred to as DNA microarray studies, whereby strains are discriminated by analyzing the hybridization of their DNA to probes of known sequences. MALDI-TOF MS is a novel high-throughput identification method relying on the analysis of whole cell proteins. MALDI-TOF MS has rapid turnaround times and low sample volume requirements and modest reagent costs when compared with conventional phenotypic or PCR-based identification, and has been used to identify *Lactobacillus* strains with some success (Dec *et al.*, 2014). Classification relies on mass spectral patterns, mostly composed of highly abundant proteins, including many ribosomal proteins, which are assumed to be characteristic for each bacterial species. DNA sequencing-based genotyping discriminate among bacterial strains directly from polymorphisms in their DNA. The discriminatory capability of each genotyping method is species dependent, with the exception of genome sequencing (Li *et al.*, 2009). Thus, whole-genome sequencing, followed by comparative genomics, is an ideal way to elucidate the genetic variability within a bacterial species. A recent review has assessed the time, cost and discriminatory power of several pheno-, physico-, and genotyping methods for their ability to "type" probiotic *Lactobacillus* strains to the genus, species, and subspecies level (Herbel *et al.*, 2013).

1.3.2.1 Population reconstruction using Multilocus Sequence Typing.

With the exception of some monomorphic (displaying low levels of DNA sequence diversity) pathogens, levels of sequence diversity are sufficiently high in most microbial taxa so that the sequences of several housekeeping gene fragments can provide a medium-resolution overview of their population phylogenetic structure (Maiden, 1998; Maiden, 2006). Multilocus sequence typing (MLST) was devised in 1998 as a method for discriminating and characterising bacterial strains (Maiden,

2006). In comparison to the typing methods mentioned above, multilocus sequence typing has more utility in determining underlying phylogenetic relationships of bacterial species (Enright & Spratt, 1999).

MLST utilises the internal nucleotide sequences of multiple housekeeping genes, as they are generally conserved within a species and mutations within these vital genes are thought to accumulate slowly and be selectively neutral, thus reflecting the true phylogeny of the strains. MLST characterizes the alleles present at several relatively conserved genomic loci and involves the partial sequencing (usually 450 bp to 700 bp) of an internal portion of a small number (between 5 and 7) of housekeeping genes, which do not show evidence of recombination. The general criteria for gene selection include their presence in all strains as a single copy (and >1 kb), their chromosomal locations, which should be evenly dispersed across the entire genome; and additionally, the functions of the encoded proteins should be conserved and well characterized. Several housekeeping genes have been used to study intraspecies relationships in LAB species (*fusA, gdh, gyrB, ddl, mutS, purK1, pgm, hsp60, ileS, pyrG, recA, recG*) (Cremonesi *et al.*, 2011; Strus *et al.*, 2012; Tanigawa & Watanabe, 2011). MLST is characterised by a universal nomenclature scheme, for each house-keeping gene, the different sequences present within a bacterial species are assigned as distinct alleles and are assigned an integer. The number of nucleotide differences between alleles is ignored and sequences are given different allele numbers whether they differ at a single nucleotide site or at many sites. The alleles at each of the seven loci define the allelic profile or sequence type (ST) of each isolate. Each isolate of a species is therefore unambiguously characterised by a series of seven integers which correspond to the alleles at the seven house-keeping loci.

There are two basic approaches that can be employed in the analysis of MLST data, those that determine the relationships of strains on the basis of their alleles' designations and STs, and those that analyze nucleotide sequences directly. Each analysis method has both strengths and weaknesses and the most appropriate means of analysis is determined by the structure of the particular population under investigation, for example the level of clonality and recombination within a species (Maiden, 2006). It is important to determine the degree of clonality of a given bacterial population as this directs the type of analysis that is most appropriate to carry out for the population in question. A test of congruence of phylogenetic signal at each of the loci is carried out, which tests the data against the null model of

clonality (Dykhuizen *et al.*, 1993; Holmes *et al.*, 1999). If the gene trees are found to be congruent, or nearly so, then a direct phylogenetic analysis of the nucleotide sequence approach is preferable, whereas in the case of incongruence an allele-based clustering technique is more appropriate (Maiden, 2006). The method of split decomposition can be used to assess the degree of tree-like structure present in the alleles found for each locus (Huson & Bryant, 2006). A tree-like structure is created when the descent is clonal, but an interconnecting network or a parallelogram appears when recombination has been involved in the evolution of the analysed gene (Huson & Bryant, 2006). Analysis which is based only on the allele assignment data, can be subjected to a number of clustering algorithms including including UPGMA (unweighted pair group method with arithmetic mean) (Prager & Wilson, 1978), split decomposition (Huson & Bryant, 2006), and BURST (based upon related sequence types) and eBURST analyses (Feil *et al.*, 2004), that were developed specifically with ST data in mind. This approach works well if the predominant mechanism for intra-strain variation is recombination, as each allele change is accorded equal weight, regardless of the number of nucleotides involved in each allelic change. Conversely, for clonal isolates, organisms for which recombination is rare or absent, it is preferable to analyze nucleotide sequences either as individual loci or as concatenated sequences containing all loci (Priest *et al.*, 2004). A test of congruence of the phylogenetic signal at each of the loci, which are displayed in the form of phylogenetic trees. If gene trees for each of the loci are congruent or almost so, the gene sequences are concatenated head-to-tail and a super-gene alignment is carried out prior to phylogenetic reconstruction to generate the so-called "supertree" (Gadagkar *et al.*, 2005). In practice many bacteria have some clonal element in their populations and this is accommodated in MLST analyses (Maiden, 2006) A range of sequence analysis software is available on the MLST website (http:// pubmlst.org/software/analysis/) and within the START2 package (Jolley *et al.*, 2001).

MLST is highly discriminatory and provides unambiguous results that are directly comparable between laboratories. MLST is particularly useful as it provides a standardised and portable method of characterizing bacteria from multiple labs and multiple sources of isolation. MLST also overcomes many of the problems associated with inter-lab variation associated with subjective techniques (e.g PFGE) (Li *et al.*, 2009). MLST generates unambiguous results that can be universally

accessed through curated online databases of MLST sequence data (http://pubmlst.org and http://mlst.net). MLST has been applied to more than 48 taxa (Achtman, 2008) and has been extensively used to study the genetic evolution, epidemiology and population structure and of a wide range of plant (Yan *et al.*, 2008) and human pathogens (Achtman, 2008; Dingle *et al.*, 2001; Enright *et al.*, 2001; Ruiz-Garbajosa *et al.*, 2006) and has more recently been applied to industrially relevant LAB strains (Calmin *et al.*, 2008; de las Rivas *et al.*, 2004) and species of lactobacilli including *L. plantarum* (de las Rivas *et al.*, 2006), *L. casei* (Cai *et al.*, 2007; Diancourt *et al.*, 2007) and *L. sanfranciscensis* (Picozzi *et al.*, 2010). As mentioned, a previous study (Li *et al.*, 2006) used a polyphasic approach to re-examine the relatedness of *L. salivarius* strains, which resulted in the unification of the original subspecies and demonstrated the suitability of the housekeeping gene *groES* as a means of resolving intra-specific differences within the *L. salivarius* species.

Reconstructing the evolutionary relationships of species and strains of bacteria through phylogenetic analysis is often central to many areas of microbiological research, such as comparative genomics, functional prediction, detection of lateral gene transfer, or the identification of new micro-organisms. Phylogenetic analysis generally follows a number of sequential successive steps which include the identification of homologous sequences, multiple alignment, phylogenetic reconstruction, and graphical representation of the inferred tree. A huge variety of models, approaches and computer programs are generally employed in the application of phylogenomics and many of these are included in Joe Felsenstein's phylogeny software inventory
(http://evolution.genetics.washington.edu/phylip/software.html).

1.3.2.2 Whole genome comparisons utilizing comparative genome hybridization.

The rapid acquisition of genetic data was one of the original motivations for the development of the orginal Affymetrix microarray technology (Fodor *et al.*, 1991). Modern DNA microarray technology has been used in a range of biological applications, which have been comprehensibly reviewed elsewhere (see (Stoughton, 2005)). From the perspective of studying bacterial genomic diversity, comparative genomic hybridisation (CGH) using DNA microarrays (Kallioniemi *et al.*, 1992) has proven invaluable in the field of bacterial comparative genomics. CGH enables the

comparison of unsequenced test strains to a sequenced reference strain on a genome-wide level, in addition to allowing correlation of phenotypic variance to genomic content (Siezen *et al.*, 2011). CGH arrays can also be used to provide insight into evolutionary processes by analyzing the diversity among strains of the same species (Rasmussen *et al.*, 2008) or different species (Fukiya *et al.*, 2004). Each experiment involves two genomes, a reference genome (often a sequenced and annotated model organism) and a test genome (often a close relative of the reference), and provides intensity ratios for each gene present in the reference genome. CGH provides an estimate of the relative abundance of genomic DNA (gDNA) in a both a target and reference strain or sample. This is determined by labelling genomic DNA from a test and reference sample with different fluorescent dyes followed by competitive co-hybridisation of the labelled samples to a microarray containing gene probes that represent the genome of the reference organism. The inclusion of reference gDNA in co-hybridisation experiments acts as a control measure for potential defects in slide printing and hybridisation conditions (Schoolnik, 2002). Bacterial plant pathogens (Aittamaa *et al.*, 2008; Sarkar *et al.*, 2006), bacterial human pathogens (Guidot *et al.*, 2009; Lepage *et al.*, 2006; Witney *et al.*, 2005), environmental isolates (Ihssen *et al.*, 2007), and industrially-relevant strains(van Hijum *et al.*, 2008) have been studied using CGH to identify intraspecific variation and to infer strain-specific virulence factors as well as potential host-interaction and metabolic capabilities. HGT is often associated with niche adaption and is evident in many CGH studies of lactobacilli species that have revealed strain-specific traits associated with mobile elements; and industrially-relavant strains, carbohydrate utilization, bacteriocin and, exopolysaccharide (EPS) production (Berger *et al.*, 2007; Cai *et al.*, 2009; Denou *et al.*, 2008; Molenaar *et al.*, 2005).

There are numerous methods (Carter *et al.*, 2008; Carter, 2005; Snipen *et al.*, 2006; Snipen *et al.*, 2009), algorithms and software programs (Chu *et al.*, 2001) which have been developed in order to normalise CGH data in prokaryotes (van Hijum *et al.*, 2008). Software programs are available which can dynamically choose hybridsation value cut-off points, which are then used to determine which genes are 'present', 'absent' or divergent' in the test strain (Davey *et al.*, 2007; Kim *et al.*, 2002).

There is a significant limitation to using a microarray design based on a single-reference genome sequence. Genes that are present in the test strains but

absent from the reference strain will not be incorporated in the CGH dataset. This limitation is mitigated when using multi-strain arrays (Siezen *et al.*, 2011) or a single reference genome sequence with the inclusion of additional niche-specific genes. Pangenome arrays harbour DNA oligomers probes which target several sequenced reference genomes from the same species. These are particularly useful when investigating species with high levels of genotypic variability. Biological interpretation of pangenome CGH data is often done on a gene-function level and often examines the presence or absence of orthologous genes (Siezen *et al.*, 2011).

Next-generation sequencing technology is revolutionizing the way in which we examine microbial diversity. As the cost of whole genome sequencing continues to fall, the sequencing of multiple strains of individual species has become a more realistic prospect, even for smaller laboratory groups. As a result, genomic diversity is increasingly being examined using whole genome sequencing, as recently demonstrated by the comparative genomic analysis of 100 *L. rhamnosus* strains (Douillard *et al.*, 2013) and 34 *L. paracasei* strains (Smokvina *et al.*, 2013). Whole genome sequencing provides information regarding genome synteny and content that are not produced by traditional microarray-based CGH. Undoubtedly, whole genome sequencing will play an increasingly important role in the study of genome diversity and the field of comparative bacterial genomics.

1.3.2.3 Next generation sequencing technologies: improvements and applications

The past 40 years has seen DNA sequencing technologies and applications undergoing significant advancements. There has been a dramatic shift away from the use of automated Sanger sequencing, which had dominated the industry since the mid 1970's. Novel sequencing methodologies are now reducing the cost per base of raw sequence as well as dramatically increasing sample throughput. Areas of research that were until recently cost prohibitive, particularly for individual laboratories, are now being carried out routinely. Notably, research into systematics, taxonomy, biotechnology, microbiology, medical diagnostics, forensic biology, and personalised medicine have benefited greatly from the vast amount of genome data that is now publically available.

Sequencing technologies and their respective chemistries can be divided into distinct "generations". The chain termination method developed by Sanger and Coulson in 1975 (Sanger *et al.*, 1977) and the chemical method of DNA sequencing

developed by Maxam and Gilbert in 1976-1977 (Maxam & Gilbert, 1977). While automated Sanger capillary-based sequencing is suitable for the sequencing of small genomic regions are both considered "first-generation" sequencing (FGS) technologies and were directly responsible for the 30 plus years of sequence-derived research that followed their invention. These techniques and their various applications have been reviewed in detail in previous articles (Ansorge, 2009; França *et al.*, 2002) it is far too labour intensive, expensive and impractical for whole genome sequencing projects (Metzker, 2010). Despite its limitations, Sanger sequencing is likely to be remain an essential tool for the majority of labs involved in molecular biology due to its accuracy and scalability.

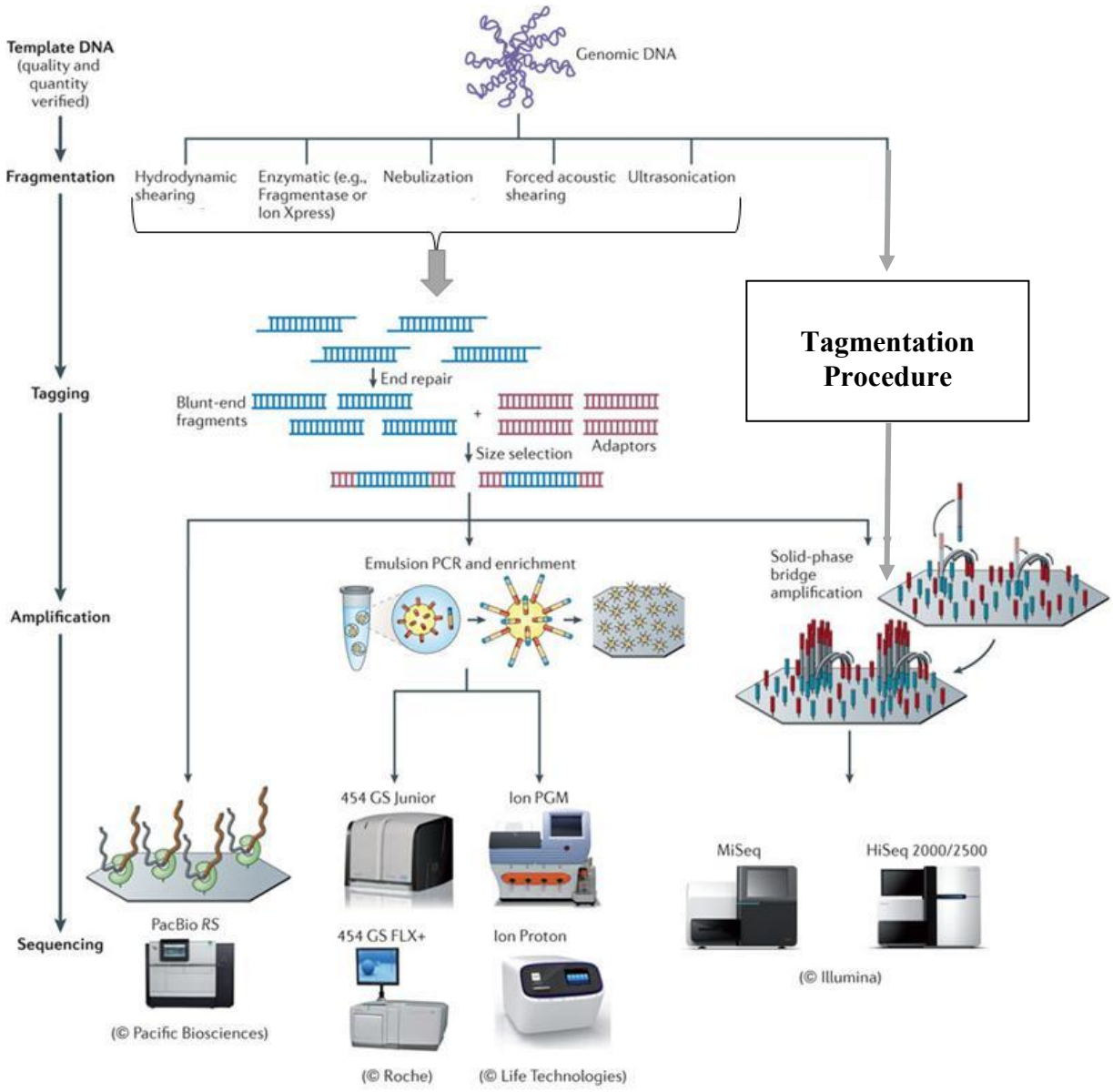


Figure 1.1. Schematic overview of the currently available, high-throughput sequencing platforms, with their associated sample preparation and template amplification procedures (modified from (Loman *et al.*, 2012)).

The high cost per base and low throughput of the first generation sequencing platforms prompted the development of so-called next-generation sequencing (NGS)/second generation sequencing (SGS) technologies during the 1990s with the first NGS equipment becoming available in 2004. Second generation sequencing/NGS technologies have overcome the low-throughput issues associated with FGS and have also significantly lowered the cost per sequenced base generated. The costs of NGS continues to fall as the capacity to multiplex sequencing reactions continues to rise (Mardis, 2011). Unlike Sanger sequencing, NGS platforms do not require a cloning step, instead using synthetic DNA fragments (adapters) which are platform-specific, to amplify the DNA library on a solid support matrix followed by cyclic sequencing. Furthermore, the steps involved in Sanger sequencing are sequential, whereas commercial NGS instruments are capable of performing parallel sequencing reactions. Although the read length in Sanger sequencing is limited by gel electrophoresis-related issues, it is determined by the signal-to-noise ratio when NGS is employed (Mardis, 2013). All NGS chemistries have in common the cyclic parallel reading of clonally amplified and spatially separated amplicons (Mardis, 2013). The technical details of NGS technologies (Loman *et al.*, 2012; Mardis, 2013; Morey *et al.*, 2013; Shendure & Ji, 2008), the associated bioinformatic and assembly programs (Horner *et al.*, 2010; Lee *et al.*, 2012) and the applications to which NGS have been applied (Didelot *et al.*, 2012; Jessri & Farah, 2014; McCormack *et al.*, 2013) have been extensively reviewed elsewhere and are beyond the scope of this introduction. An outline of the main high-throughput sequencing platforms that are currently available and the associated sample and template preparation steps are illustrated in Figure 1.1. Additionally, Table 1.2 presents a summary of the current commercially available NGS instruments available and includes the key factors to be considered in the definition and evaluation of NGS technology platforms, namely: read length, and throughput.

Table 1.2 Current commercially available NGS instruments, table adapted from (Didelot *et al.*, 2012; McCormack *et al.*, 2013).

Platforms	Sequencing principle	Read Length (bp)	Run Time	Max output	Pros	Cons	Application
Roche 454 FLX+	Pyrosequencing	700	23 h	0.7 GB	High throughput, long read, short run time	High error rate in homopolymers regions, long hand-on time	*De novo* whole genome sequencing of microbes, exome sequencing
Roche 454 GS Junior (BT)	Pyrosequencing	400	10 h	0.035 GB	Long reads, short run time	High error rate in homopolymers regions, long hand-on time	*De novo* whole genome sequencing of microbes, exome sequencing
Illumina® HiSeq™ 2500/1500	Sequencing by synthesis	36/50/100	11 days 27 h (rapid run)	600 GB	High throughput, cost-effectiveness	Short read length(Jessri & Farah, 2014) high error rate	Human whole genome sequencing, exome sequencing, RNA-seq methylation
Illumina® MiSeq™ (BT)	Sequencing by synthesis	25/36/100	4-27 h	8.5 GB	High throughput, cost-effectiveness, short run time, high coverage	Short read length	Microbial discovery, exome sequencing, targeted capture
Life Technologies™ SOLiD™ 5500XL	Sequencing by ligation	75+35	7 days to 4 weeks	180 GB	Very high throughput, low error rate	Long run times, short reads, complexity of library preparation	Human whole genome sequencing, whole exome sequencing, RNA-seq, methylation
Life Technologies™ Ion PGM™ (BT)	Hydrogen ion sensitive transistor	35/200/400	2 h	2 GB	Short run time, low cost per sample	High reagent cost, high error rate in homopolymers region	Exome sequencing, targeted gene sequencing, microbial discovery
Life Technologies™ Ion Proton™ ChipI/II (BT)	Hydrogen ion sensitive transistor	Up to 200	2 h	10 GB/ 100 GB	Short run times, flexible chip reagents	High reagent cost, high error rate in homopolymers region, limited body of literature to support	Exome sequencing, targeted gene sequencing, microbial discovery
Pacific Biosciences Inc. PacBio RS I1 (TGS)	Single molecule real-time sequencing	250-10,000	30-120 min	217 Mb	Short run times, no PCR amplification, long reads, low cost	Not optimized for human genome	Virus, bacteria, lower eukaryote genome sequencing, target sequencing

BT: Bench-top instrument.

TGS: Third generation sequencing.

Important applications of NGS include: whole genome re-sequencing (Suzuki *et al.*, 2011), de novo sequencing (Ghosh *et al.*, 2011), complete RNA transcriptomics (Gibbons *et al.*, 2009) and bacterial epidemiology (Parkhill & Wren, 2011). Of particular interest to the study of microbial diversity is the study of those bacteria which are generally considered to be unculturable (Pace, 2009). The study of metagenomics involves the study of genetic diversity, population structures and interactions of microbial communities in their ecosystems, using culture-independent techniques. The study of metagenomics has come to the forefront of microbiological research in recent years with the maturation of NGS technologies, due to the depth of coverage and low sequencing costs associated with these techniques. Compositional 16S rRNA gene sequencing using NSG has enabled comprehensive quantitative and qualitative analysis of microbial diversity in a variety of ecosystems (Claesson *et al.*, 2012; Crielaard *et al.*, 2011; Kong, 2011). The advent of second-generation sequencing technologies has advanced our ability to identify and characterize microbes in a more timely and accurate manner. Instead of typing strains based on relatively few selected loci or phenotypes, we can now rapidly sequence bacterial genomes (Chin *et al.*, 2011), thereby facilitating the detection of subtle inter-strain differences that might otherwise not be recognized (Mutreja *et al.*, 2011). The speed of NGS techniques also enables the investigation of infectious outbreaks in real time (Mellmann *et al.*, 2011). While the low cost and high-throughput nature of NGS platforms is suited to the re-sequencing of whole genomes. *De novo* genome assembly of microbial genomes generated with short read technologies often results in highly fragmented assemblies, because of the reduction in assembly quality with decreasing read lengths (Whiteford *et al.*, 2005). Gaps in assemblies derived from NGS reads are generally a direct result of unresolved repeats (Cahill *et al.*, 2010). Despite many computational advances in genome assembly, complete and accurate assembly from second-generation, short-read data is a major challenge (Alkan *et al.*, 2011). In spite of this, NGS technologies have been applied to *de novo* bacterial genome sequencing projects (Margulies, 2005). The read lengths and throughput of NGS technologies (Table 1.2) have steadily increased in the last 10 years which has resulted in increased contig sizes and a reduced number of sequencing gaps (Cahill *et al.*, 2010). However, the most significant advance in NGS technologies was the

introduction of paired reads, which is critical for resolving repeats within sequencing assemblies (Chaisson *et al.*, 2009). Draft genomes based on NGS data alone contain gaps which are often caused by extreme GC- or AT-rich regions or palindromic sequences, which cause problems for NGS methods. Sanger sequencing is often used to 'finish' microbial genomes, but this process is laborious, slow, and expensive, and is the reason why many genome sequences are not being assembled to completion. Large scale structural rearrangements, gene duplications or sequence inversions, in addition to mobile elements such plasmids and phages can often be overlooked by re-sequencing approaches to genome assemblies. These sequence elements often represent the only genetic differences between closely related strains and provide valuable information when studying microbial diversity (Ricker *et al.*, 2012).

Third Generation Sequencing technologies (TGS) are still in their infancy in term of their routine use in the laboratory. However, one of the main advantages of TGS techniques over NGS are their ability to sequence single molecules of DNA with no need for clonal amplification prior to sequencing (see review by (Morey *et al.*, 2013). This avoids the introduction of artifacts from PCR and requires less manipulation of the sample in comparison with NGS. TGS also usually involves sequencing-by-synthesis chemistries, but detection techniques are based mainly on the physical recognition of DNA bases in an unmodified DNA strand, rather than on detection of chemical incorporation (Pettersson *et al.*, 2009).

Interestingly, Single Molecule, Real-Time (SMRT) DNA sequencing (Pacific Biosciences) has been shown to generate sequencing reads that are significantly longer than either second-generation or Sanger sequencing reads, thus facilitating high quality *de novo* genome assembly and genome finishing (English *et al.*, 2012). For a detailed description of TGS techniques which are currently under development and use, see reviews (Mardis, 2013; Morey *et al.*, 2013).

Hybrid assembly strategies, combining sequencing data from both FGS and 454 platforms have been used successfully to sequence complete microbial genomes (Goldberg *et al.*, 2006). This hybrid approach improves assemblies by increasing coverage and reducing gaps resulting from cloning bias (Schatz *et al.*, 2010). Hybrid sequencing strategies have been developed to use only NGS such as Illumina and 454

(Aury *et al.*, 2008), or a combinations of NGS and TGS platforms (Bashir *et al.*, 2012). TGS methods can generate reads that span many thousands of bases, however, the raw accuracy per read generated from third-generation technologies is currently markedly lower than that from second-generation methods (Schadt *et al.*, 2011). Therefore, strategies which incorporate both second- and third-generation data are proposed to be capable of generating high-quality, near-complete genomes (Bashir *et al.*, 2012). It is thought that hybrid sequencing and assembly strategies are most effective when complementary sequencing technologies are combined, for example homopolymer errors which are inherent in 454 generated reads can be detected and corrected using the higher coverage Illumina platform (Aury *et al.*, 2008; Forde *et al.*, 2011). Furthermore, it has been shown that hybrid assemblies using both 454 and Illumina platforms produce *de novo* assemblies whose quality is at least on a par with those produced using only Sanger sequencing (Aury *et al.*, 2008).

There is an ongoing need to develop statistical methods and bioinformatics tools which are suitable for the analysis and management of the NGS and TGS data. It is also necessary to develop programs capable of translating the results of DNA sequencing into useable information which can be readily interpreted by researchers, physicians, patients, and others (Schadt *et al.*, 2010). There are numerous pipelines for the automated annotation of *de novo* assemblies, including RAST (Aziz *et al.*, 2008), IMG/ER (Markowitz *et al.*, 2009) and the IGS Annotation Engine (developed by the Institute of Genome Sciences, University of Maryland School of Medicine, USA). An overview of the programs and algorithms which are commonly used to analyse and process NGS and TGS data has been outlined in detail in the following reviews by Loman *et al.* (2012) and Morey *et al.* (2013) (Loman *et al.*, 2012; Morey *et al.*, 2013). Edwards & Holt (2013) have also provided a comprehensive beginner's guide to comparative bacterial genome analysis using next-generation sequence data (Edwards & Holt, 2013).

1.4 Genome architecture of *Lactobacillus salivarius* and other bacteria.

1.4.1 Unusual genomic architectures in bacterial species.

Classically, bacterial genomes are thought to be comprised of a single indispensable (Mb-sized) circular chromosome and zero to many dispensable plasmids (Casjens, 1998). Advances in Pulsed Field Gel electrophoresis (PFGE) (Schwartz & Cantor, 1984) methodologies and the recent boom in available genome sequences from related species has begun to challenge this paradigm, by identifying bacterial genome sizes ranging from approximately 139 kb (McCutcheon & von Dohlen, 2011) to 13.7 Mb (Chang *et al.*, 2011), as well as a growing number of bacterial species which have a multipartite genomic organization (Bentley & Parkhill, 2004; Casjens, 1998). Technological advances have provided information about the genome architecture and evolution of an increasing number of multireplicon genomes and revealed a wide distributions of both circular and linear plasmids in eukaryotic and prokaryotic species (see review by (Bentley & Parkhill, 2004)).

The term "plasmid" was introduced in 1952 by J. Lederberg and is used to describe all extra-chromosomal genetic elements (Lederberg, 1952) which are widespread among prokaryotes and which often confer selective advantages on the host cell (Bentley & Parkhill, 2004). While plasmids were originally considered to be exclusively circular in nature, the first report of a linear extra-chromosomal element (discussed below) was described in 1967 (Ravin & Golub, 1967). Although a linear topology has more frequently been described for extra-chromosomal elements, linear chromosomes have also been frequently identified in *Borrelia burgdorferi* (Ferdows *et al.*, 1996), *Agrobacterium tumefaciens* (Allardet-Servent *et al.*, 1993), *Streptomyces coelicolor* (Kieser *et al.*, 1992). While *Streptomycetes* generally harbours exclusively linear replicons, *Agrobacterium* and *Borrelia* species have a combination of linear and circular replicons (Bentley & Parkhill, 2004). In addition, the *Rhizobiaceae* family of the *Alphaproteobacteria* harbours both singular and multiple chromosome

arrangements, in addition to linear replicons and plasmids of various sizes (Slater *et al.*, 2009).

Primary chromosomes of bacteria are generally the largest replicon resident in the cell, they employ a *dnaA*-based replication process and encode the vast majority of the core genes of the species, including those that govern the essential cellular processes (MacLellan *et al.*, 2004). "Second chromosomes" were originally identified by Suwanto (Suwanto & Kaplan, 1989) and have since been identified in a wide range of bacterial genera (Landeta *et al.*, 2011). Secondary chromosomes are generally smaller than the primary chromosome, harbor *repABC*-type plasmid-associated replication systems (Downie & Young, 2001; Harrison *et al.*, 2010) and contain few essential genes which generally code for niche-specific attributes (Egan *et al.*, 2005; Harrison *et al.*, 2010; Slater *et al.*, 2009) or are duplicates of essential genes, such as rRNAs or tRNAs (Landeta *et al.*, 2011). Indeed, empirical determination of their necessity for cell viability is also used to identify replicons as secondary chromosomes (Landeta *et al.*, 2011). A recent study by Harrison *et al* proposed a novel naming scheme for these "secondary chromosomes", suggesting that they be called "chromids" (Harrison *et al.*, 2010). These elements are distinguished from both primary chromosomes and plasmids by a combination of specific features: (I) "chromids have plasmid-type maintenance and replication systems". Using a defined set of criteria, this study identified 82 chromids, varying in both size and composition, from a test set of 897 replicons (replicating independently of the chromosome) (Harrison *et al.*, 2010) (II) "chromids have a nucleotide composition close to that of the chromosome" (generally within 1%), and (III) "chromids carry core genes that are found on the chromosome in other species", which make the chromid indispensible to the cell (Harrison *et al.*, 2010). In addition to these primary criteria, chromids are described as being larger than other co-resident plasmids and also encode a greater proportion of accessory genes when compared to the chromosome (Harrison *et al.*, 2010). These accessory genes are often encoded by other species within the same genus or may be shared by other chromids but otherwise have a limited phylogenetic distribution (Harrison *et al.*, 2010). Harrison *et al* (2010) describes the possession of chromids as a genus-specific characteristic and their distribution appears to be a reflection of the hosts' phylogeny rather than their ecological setting

(Harrison *et al.*, 2010). There is still much confusion surrounding the designation of many extra-chromosomal replicons as either megaplasmids (see below) or chromids. Although the new nomenclature suggested by Harrison *et al*. (2010) has not yet been universally adopted, it may over time be used to replace the term secondary chromosome (Harrison *et al.*, 2010).

1.4.2 Circular megaplasmids, defining characteristics.

The term megaplasmid was first coined by Rosenberg *et al* in 1981 when studying large plasmids in *Sinorhizobium meliloti* and was used to describe plasmids with a molecular weight greater than 300 x 10^6 (~450kb) (Rosenberg *et al.*, 1981). This size threshold was arbitrary and had no particular biological significance for distinguishing large plasmids from their smaller coresident counterparts. However, perhaps the most important distinction between the newly described chromids and *bona fide* megaplasmids is the carriage of essential genes, which is a factor that is often used to distinguish megaplasmids from secondary chromosomes (Ng *et al.*, 1998). This is emphasized by Landeta *et al.* who refers to chromids as megaplasmids that have acquired sets of essential genes, thus becoming secondary chromosomes (Landeta *et al.*, 2011). Indeed it is often historical precedence within a species that determines whether a large secondary element is designated as a megaplasmid or mini/secondary chromosome (Bentley & Parkhill, 2004). In many cases the essential nature of a replicon is difficult to prove because of technical difficulties encountered when trying to cure these elements (Smillie *et al.*, 2010). Claesson *et al.* (2006) defined the megaplasmids of *L. salivarius* as such because they: (I) are larger than 100 kb in size. (II) utilise plasmid associated replication and maintenance systems and (III) do not encode any single-copy essential housekeeping genes (Claesson *et al.*, 2006).

1.4.2.1 Distribution and encoded properties of circular megaplasmids.

Megaplasmids harboured by plant associated bacteria are some of the most thoroughly studied and were first described over 30 years ago (Rosenberg *et al.*, 1981). The presence of large (~2.1 Mb) megaplasmids is characteristic of strains of *Ralstonia solanacearum* (Rosenberg *et al.*, 1982) a major soil-borne plant pathogen that causes bacterial wilt disease in over 200 plant species (Genin & Boucher, 2004). Their megaplasmids are thought to have coevolved with the chromosome and have gradually evolved to become an indispensable component of the genome (Coenye & Vandamme, 2003). The megaplasmids of this species are thought to be still evolving through translocation and duplication of essential genes normally resident on the chromosome (Genin & Boucher, 2002). The megaplasmids also encode genes associated with increased fitness and survival as they encode several duplications of essential

housekeeping genes as well as genes that encode enzymes controlling the biosynthesis of amino acids, nucleotides, and cofactors (Salanoubat *et al.*, 2002) including those involved in flagellum biosynthesis, pathogenicity, catabolism of aromatic compounds, and putative metal resistance (Genin & Boucher, 2002).

Many of the largest megaplasmids identified to date have been isolated from rhizobia and often carry genes responsible for their hosts' symbiotic properties, including nodulation and the ability to fix atmospheric nitrogen (Bartosik *et al.*, 2002; Pueppke & Broughton, 1999). The genus *Rhizobium* (previously members of the genus *Agrobacterium* (Young *et al.*, 2001; Young *et al.*, 2003)) contains plant-pathogenic species *R. tumefaciens*, *R. rhizogenes* and *R. vitis*, strains of which invade the crown (*via* wounds), roots and stems of dicotyledonous and gymnospermous plants (Young *et al.*, 2001). Most of the genes essential for the pathogenicity (the vir and T-DNA gene regions) of these species are encoded by their tumour-inducing (Ti) or hairy-root-inducing (Ri) megaplasmids, which cause crown gall and hairy root diseases, respectively (Stanton, 1990). These megaplasmids generally have a broad host range, are approximately 200 kpb in size and are commonly used as vectors in the construction of transgenic plants (Gelvin, 2003) pSymA (1.35Mb-1.63Mb) and pSymB (1.68 Mb-1.82Mb) (Barloy-Hubler & Jebbar, 2009; Suzuki *et al.*, 2009).

All characterized strains of *Sinorhizobium meliloti*, a gram negative soil dwelling bacteria, harbor two stable extra-chromosomal replicons with different evolutionary histories, pSymA (1.35 Mb-1.63 Mb) and pSymB (1.68 Mb-1.82 Mb) (Barloy-Hubler & Jebbar, 2009). pSymB is considered to be a secondary chromosome while pSymA is described as an accessory megaplasmid (Barloy-Hubler & Jebbar, 2009) and has been successfully cured without affecting the host viability (Oresnik & Long, 2000). The nature of the genome of *S. meliloti* is highly dynamic with pSymA, pSymB and the host chromosome undergoing frequent co-integration and excision. pSymA encodes genes responsible for the formation of nitrogen-fixing root nodules and their associated nitrogen fixation capabilities (Banfalvi *et al.*, 1981; Rosenberg *et al.*, 1981). In addition to a complete de-nitrification pathway, pSymA also encodes stress-response genes and a number of redundant housekeeping genes and those involved in the uptake and assimilation of organic and inorganic nutrients. Thus, pSymA enhances the metabolic

versatility and symbiotic nature of *S. meliloti* and enables this species to colonize plants and thrive in the rhizosphere.

A wide variety of bacteria have been found to harbour megaplasmids which are solely or partially responsible for the industrial and ecological attributes in their hosts (Schwartz, 2009) (Table 1.3) by encoding genes or complex metabolic pathways that, though nonessential, enable the host to adapt or flourish in specialized niches. In complex environments, where dynamic metabolic capabilities are of benefit, it is thought that there are selective pressures towards large or multireplicon genomes, where large extra-chromosomal elements often encode paralogous metabolic contingency genes and pathways. Catabolic and anabolic pathways are frequently, either fully or partially, encoded on megaplasmids in a range of soil and water associated bacteria (Vedler, 2009), predominately *Pseudomonads* (Nojiri *et al.*, 2009) and are generally large, low copy number replicons (up to and over 1Mb) (Table 1.3). In addition, these megaplasmids often encode a full complement of the genes necessary for conjugal transfer of these elements, in addition to numerous transposons and ARE elements (Vedler, 2009). The continuous evolution of these pathways enables soil and water associated bacteria to utilize rarely encountered and novel xenobiotics, transient nutrient resources and environmental pollutants including polycyclic aromatic hydrocarbons and heteroaromatic compounds which are known mutagens and carcinogens (Marston *et al.*, 2001; Mastrangelo *et al.*, 1996; Xue & Warshawsky, 2005). Sequence comparisons of catabolic megaplasmids reveal a complex evolutionary history, which results from transposition and recombination events (Dennis, 2005). For a detailed overview of catabolic megaplasmids, see (Nojiri *et al.*, 2009; Vedler, 2009). Haloarchaea are widely distributed geographically in hypersaline environments and typically harbour one or more megaplasmids which encode genes which provide alternate or expanded metabolic pathways within the host organism (DasSarma *et al.*, 2009). However megaplasmids associated with archaeal species will not be discussed further in this introduction.

Table 1.3 Selection of circular megaplasmids associated with niche adaptation.

Host strain	Plasmid	size	Associated niche/Isolation source	Associated phenotype	Reference
Enterococcus faecium E980	Unnamed	200 kbp	Mammalian gut	Raffinose metabolism	(Zhang *et al.*, 2011)
Streptococcus salivarius K12	pSsal-K12	190 kbp	Oral Cavity	Lantibiotic bacteriocin (salivaricin A2 and salivaricin B) production	(Hyink *et al.*, 2007)
Ralstonia eutropha H16	pHG1	452 kbp	Soil and water	Hydrogenotrophy, chemolithotrophy, CO2 fixation and denitrification	(Pohlmann *et al.*, 2006)
Oligotropha carboxidovorans OM5	pHCG3	133 kbp	Soil of waste water sewage treatment ponds	H_2- and CO-based lithoautotrophy CO2 fixation	Gerstenberg 1982, Kraut and Meyer 1988, Fuhrmann 2003
Arthrobacter nicotinovorans	pAO1	165 kbp	Soil around tobacco plants	Nicotine catabolism	Igloi and Brandsch 2003
Gordonia westfalica Kb1	pKB1	101 kbp	Foul water inside a degrading tyre	Rubber biodegradation Cadmium resistance	(Broker *et al.*, 2004)
Cupriavidus metallidurans CH34	pMOL30	234 kbp	Sludge of a zinc decantation tank	Resistance to zinc, cadmium, cobalt, copper silver, lead and mercury	(Janssen *et al.*, 2010; Monchy *et al.*, 2007; von Rozycki & Nies, 2009) and references therein
Cupriavidus metalliduransi CH34	pMOL28	171 kbp	Sludge of a zinc decantation tank	Resistance to nickel, cobalt chromate mercury and thallium	(Janssen *et al.*, 2010; Monchy *et al.*, 2007; von Rozycki & Nies, 2009) and references therein
Cupriavidus. taiwanensis LMG19424	pRALTA	557 kbp	Root nodule of Mimosa pudica	Nitrogen fixation and legume symbiosis	(Amadou *et al.*, 2008)
Burkholderia xenovorans LB400	box_M	1.47 Mbp	Polychlorinated biphenyls-contaminated soil	Benzoate catabolism	(Denef *et al.*, 2006)
Ralstonia solanacearum	PSI07	5.6 Mbp	Soil	Oxidize arsenite, antimitotic toxin rhizoxin	(Remenant *et al.*, 2010)
Pantoea vagans C9-1	pPag3	530 kbp	An apple	Carotenoid pigmentation siderophore biosynthesis, ampicillin resistance	(Smits *et al.*, 2010)

In addition to plant-pathogens, megaplasmids are associated with a range of human-pathogen bacterial species. *Bacillus anthracis* is the etiological agent of anthrax (Mock & Fouet, 2001) which leads to toxemia and septicemia in infected mammals. Virulent strains of this species harbor two megaplasmids, pXO1 (approx 182 kb) (Okinaka *et al.*, 1999) and pXO2 (95 kb) which encode major virulence factors (Candela & Fouet, 2005; Makino *et al.*, 1989; Mikesell *et al.*, 1983), whereas strains lacking either of these megaplasmids show attenuated virulence in animal models (see (Fouet & Moya, 2009) and references therein). Both linear and circular extra-chromosomal elements have been described in the genus *Mycobacterium* and generally range in sizes between 13 kb and 330 kb (Pidot *et al.*, 2009). *M. ulcerans* is the causative agent of Buruli ulcer and virulence in *M. ulcerans* has been associated with the presence of a 174 kb megaplasmid, pMUM001 (Stinear *et al.*, 2004). The acquisition of pMUM001 has been proposed to be the result of horizontal transfer (Stinear *et al.*, 2004) and is directly responsible for the emergence of *M. ulcerans* as a significant human pathogen (Pidot *et al.*, 2009; Stinear *et al.*, 2004). Even more significant from a medical point of view are the species of the genus *Yersinia* and their associated plasmids. Of the 11 species belonging to this genus, three are pathogenic to humans including the causative agent of plague *Yersinia pestis* and the enteropathogenic Yersiniae, *Y. pseudotuberculosis* and *Y. enterocolitica* which cause gastroenteritis (see (Wren, 2003) (Table 1.3). *Y. pestis* strains harbour two further plasmids and references therein). All three species carry a ~70-kb virulence plasmid (pYV), which is essential for infection of lymphatic tissues and for overcoming host defense mechanisms (Howard *et al.*, 2009). pMT1 (100 kb) and a pPCP1 (9.6 kb) which are absent in *Y. pseudotuberculosis* and *Y. enterocolitica*. Plasmids have played a role in the development and evolution of virulence in *Yersiniae* species and their transition from benign soil-dwelling bacteria to human pathogens (Howard *et al.*, 2009; Wren, 2003).

Shigellosis is a disease of humans and is characterized by the bacterial invasion of the colonic epithelium which causes inflammation and destruction of the colonic mucosa. It is caused by the following species of *Shigella*: *S. boydii*, *S. dysenteriae*, *S. flexneri* and *S. sonnei*, and the enteroinvasive pathogen *Escherichia coli* (EIEC). All *Shigella* and EIEC strains carry a virulence plasmid (VP) of approximately 200 kb that

encodes essential virulence determinants (see (Dorman, 2009; Parsot, 2005) and references therein). Although chromosomal genes are necessary for full expression of *Shigella* virulence determinants (Hale, 1991; Parsot, 2005) it is widely accepted that the acquisition of a precursor of the present day virulence plasmid was the critical step in the creation of the pathogenic *Shigella* species. (Jin *et al.*, 2002; Peng *et al.*, 2009). Shiga toxigenic *Escherichia coli* (STEC) (also referred to as enterohemorrhagic *E. coli* (EHEC)) are food-borne pathogens (Feng *et al.*, 2011) which are associated with the development of hemolytic uremic syndrome (HUS) in humans and animals (Nataro & Kaper, 1998). In addition to well documented chromosomally-encoded virulence genes (Sperandio *et al.*, 1998) the majority of STEC strains of human origin harbor large, genetically diverse, virulent plasmids (>90kb) (dos Santos *et al.*, 2010; Fratamico *et al.*, 2011; Srimanote *et al.*, 2002), some of which are conjugative (Srimanote *et al.*, 2002). *Enterococcus faecium* is a commensal species of both animal and human intestines, but may act as a facultative pathogen under specific conditions and in patients with severe underlying diseases. Conjugative megaplasmids (~150 kb) are commonly identified in *E. faecium* strains and are known to carry various antibiotic resistance genes and GI tract colonization factors (Freitas *et al.*, 2010; Garcia-Migura *et al.*, 2007; Rice *et al.*, 2009; Zhu *et al.*, 2009). The putative virulence associated gene hyl(Efm), encodes a hyaluronidase (Arias *et al.*, 2009; Freitas *et al.*, 2010; Rice *et al.*, 2003) and is located on *E. faecium* megaplasmids (150–350 kb) and forms part of a highly conserved genetic structure on these plasmids (Laverde Gomez *et al.*, 2011). This gene is found almost exclusively in *E. faecium* clinical isolates and is involved in peritoneal invasion and intestinal colonization (Arias *et al.*, 2009; Freitas *et al.*, 2010). Pilus production in *E. faecium* is known to play a role in biofilm formation and the ability to adhere to mammalian cells (Jouko Sillanpää *et al.*, 2010). PilA-type pili have been identified on the surfaces of both bacteremia and fecal isolates of *E. faecium* (Hendrickx *et al.*, 2010) and are encoded by a pilin gene which is colocalised with hyl Efm on the megaplasmid pLG1 (Kim *et al.*, 2010) In addition to the megaplasmids of *E. faecium* enhancing their pathogenic capacity, carbohydrate transport and metabolism genes, as well as heavy metal resistance are also encoded by many *E. faecium* megaplasmids (Laverde Gomez *et al.*, 2011). The genes associated with raffinose utilisation were also recently localized to

megaplasmids of environmental and commensal *E. faecium* which range in size from 150 and 300 kb (Zhang *et al.*, 2011). It is thought that the conjugative megaplasmids of *E. faecium* facilitate horizontal gene transfer and are therefore responsible for the spread of pathogenicity factors and antibiotic resistance genes in clinical isolates of *E. faecium*.

The megaplasmid pMP118 of *L. salivarius* UCC118 was originally described and sequenced by Claesson *et al* (2006) and at the time was the largest plasmid of LAB to be sequenced (Claesson *et al.*, 2006), but has since been surpassed in size by the enterococcal plasmids described above. A detailed description of the genes encoded pMP118 and the role that it plays in the lifestyle of *L. salivarius* is described in detail below. Prior to the discovery of pMP118 (Claesson *et al.*, 2006), large plasmids (>100 kb) had been identified in the *Lactobacillus* species, *L. acidophilus* (Roussel *et al.*, 1993) and *L. gasseri* (Muriana & Klaenhammer, 1987) but had not been studied in any detail. In addition to pMP118 (242 kb), Claesson *et al* (2006) identified megaplasmids in a further nine strains of *L. salivarius*, including the species type-strains DSM20554 and DSM20555. The megaplasmids varied in size (100kb-390kb) and were identified in strains that were isolated from a variety of ecological niches, including gastrointestinal tract isolates, clinical strains, animal and food sources (Claesson *et al.*, 2006). Subsequently, Li *et al* (2007) expanded the study of *L. salivarius* megaplasmids, identifying "*repA*-type" megaplasmids in 22 additional *L. salivarius* isolates (Li *et al.*, 2007) (see section *1.4.3* below). The *L. salivarius* circular megaplasmids were shown to have highly conserved replication backbones consisting of the replication associated genes *repA* and *repE* and the chromosome partitioning ATPase, *parA* (Li *et al.*, 2007). Thus, the circular megaplasmids were designated *repA*-type megaplasmids (Li *et al.*, 2007) and were asserted to be "a general feature of *L. salivarius*" (Li *et al.*, 2007). The *L. salivarius* megaplasmids were also surveyed for the conservation of phenotypic features associated with pMP118, including bacteriocin production and carbohydrate metabolism associated genes, which are thought to offer selective advantages to the host strain. Results indicated that the *repA*-type megaplasmids of *L. salivarius* were genetically diverse (Li *et al.*, 2007) and no correlation between gene presence/absence could be attributed to specific niche associations (Li *et al.*, 2007). Li *et al* (2007) went on to examine the distribution of megaplasmids throughout the genus *Lactobacillus*. 91

strains belonging to 47 species were examined and megaplasmids were identified in seven strains belonging to six of the species examined (Li *et al.*, 2007). All strains that harboured megaplasmids were gastrointestinal isolates of either human or animal origin and belonged to the following species: *Lactobacillus hamsteri* DSM5661 (pMP5661, 390 kb), *L. intestinalis* DSM6629 (pMP6629, 320 kb), *L. kalixensis* DSM16043 (pMP16043, 490 kb), *L. ingluviei* DSM14792 (pMP14792, 190kb), *L. ingluviei* DSM15946 (pMP15946, 120 kb), and *L. acidophilus* ATCC 4356 (pMP4356, 150 kb) (Li *et al.*, 2007)

The replication machinery employed by *L. salivarius* megaplasmids is genetically distinct from that of the other known *Lactobacillus*' megaplasmids, all of which lack a homolog of the pMP118-encoded *repA* gene (Li *et al.*, 2007). The distribution pattern of the circular megaplasmids of *Lactobacillus* does not concord with the phylogenetic structure of the genus. In contrast, both *L. salivarius* and *L. equi* belong to group E of the 16S rRNA phylogeny (Canchaya *et al.*, 2006) and were the only two species found to harbour linear megaplasmids (Li *et al.*, 2007). The correlation between these phylogenetically related species and the presence of linear megaplasmids may however be purely coincidental and warrants further investigation. It is somewhat clearer though that all megaplasmid-containing species identified in the study by Li *et al* (2006) are of intestinal origin and it was therefore proposed that "megaplasmids are uncommon in extraintestinal lactobacilli, food-associated lactobacilli, and free-living species" (Li *et al.*, 2007).

1.4.2.2 Evolutionary origin.

Understanding the evolution of multipartite genomes has been greatly facilitated in recent years by the sequencing of multiple strains of multiple species with multireplicon architectures. However, it is not yet known what, if any, evolutionary advantage multipartite genomes provide to their hosts. It has been speculated that maintenance of multiple large replicons may allow for replicon-specific gene regulation (Heidelberg *et al.*, 2000) which may be advantageous under particular environmental conditions. In addition secondary replicons may offer a repository for niche-specific genes which can remain independent of the chromosome replication cycle (Gonzalez *et al.*, 2010). Slater *et al* (2009) proposed a generalized model for the formation of

secondary chromosomes among bacteria (Slater *et al.*, 2009). The authors suggest that ancestral plasmids (of the *repABC* type) were recipients of intragenomic gene transfer of essential genes from progenitor primary chromosomes, thus enlarging the secondary replicons and making them essential to cell viability (Slater *et al.*, 2009). The secondary chromosomes and large plasmids of many species display conservation and synteny of gene clusters with their primary chromosomes, supporting the aforementioned hypothesis (Slater *et al.*, 2009). Unsurprisingly, a similar proposal has been suggested for the origins of the newly described (but long observed) chromids. Harrison *et al* (2010) suggests that the formation of chromids happens as a result of resident plasmids acquiring core genes, through intragenic recombination with the chromosome followed by the loss of the original copies from the hosts' chromosomes (Harrison *et al.*, 2010). Alternatively, the authors also propose that the formation of a chromid could arise from a "rebirth" event whereby an existing, essential, chromid-like replicon adopts the replication system of a newly acquired plasmid (Harrison *et al.*, 2010). There is ongoing confusion surrounding the naming schemes and criteria used to distinguish large secondary replicons from one another. The megaplasmid of *Halobacterium* sp NRC-1 is thought to have evolved from one of its smaller plasmids, carrying Rep-like proteins which was subsequently followed by the acquisition and duplication of chromosomal genes, ensuring the maintenance of the newly formed replicon (DasSarma *et al.*, 2009) that "secondary chromosomes", "chromids" "megaplasmids" and "large plasmids" represent a continuous spectrum of evolution. Megaplasmids are most likely formed through homologous or site-specific recombination events involving multiple smaller plasmids or by the integration of chromosomal (essential and nonessential) genes or regions into resident plasmids or a combination of both. The acquisition of genes *via* horizontal gene transfer, gene loss and gene degradation is also likely to have influenced the subsequent evolution of these replicons considerably. A study of the molecular phylogeny of the chromosomes and megaplasmids of *L. salivarius* indicated that they have a shared evolutionary history (Li *et al.*, 2007). Maximum-likelihood trees were constructed, employing the *repE* genes of the *repA*-type megaplasmids and the *groEL* genes of the *L. salivarius* panel of strains described above (Li *et al.*, 2007). Interestingly, human-associated strains clustered more closely with each other than with those of

animal origin. The concordance between the *repE*-derived and *groEL* trees was incomplete each group forming subclusters within the two major branches of the *repE* and *groEL* trees (Li *et al.*, 2007) however, there was a strong agreement between both trees and thus the authors proposed that the acquisition of megaplasmids was a "relatively early event in the evolution of *L. salivarius*" (Li *et al.*, 2007).

Section 1.4.3 is published as a book chapter (Raftis & O'Toole, 2009)

1.4.3 Megaplasmids in *Lactobacillus salivarius*

1.4.3.1 Contribution of pMP118 to the lifestyle of L. salivarius UCC118.

Lactobacillus salivarius UCC118 habours a multiple replicon genome, including a 242 kb megaplasmid designated pMP118. pMP118 carries a number of contingency genes which work in conjunction with chromosomally encoded genes and pathways to broaden the metabolic flexibility of this strain. This increases the potential viability of UCC118 in the competitive environment of the gastrointestinal tract. Annotation and functional studies have indicated that pMP118 contributes significantly to the probiotic properties of UCC118, encoding a bile salt hydrolase gene and a potent broad-spectrum bacteriocin. pMP118-related megaplasmids have been established as a general feature of the species *Lactobacillus salivarius.* Megaplasmids have also been identified in six other *Lactobacillus* species of intestinal origin. Dissemination of the pMP118-related megaplasmids may have occurred by a conjugation apparatus which is now non-functional in pMP118. The analysis of pMP118 highlights the contribution of this replicon to the biology and ecology of this commensal species.

1.4.3.2 Discovery of pMP118.

Smaller plasmids are a common feature of LAB and specifically the genus *Lactobacillus* (Wang & Lee, 1997), with many strains harbouring multiple replicons. The LAB plasmids have been the focus of much study as plasmid encoded traits have been shown to contribute significantly to the phenotype of industrially important strains of species belonging to the LAB group (Mills *et al.*, 2006; Siezen *et al.*, 2005). Despite this, relatively few large plasmids (>100kb) had ever been described for the genus *Lactobacillus* (Roussel *et al.*, 1993) and beyond these initial reports, there has been no further characterization. Megaplasmid pMP118 was the first to be definitively identified and characterized in a probiotic species, and is, to date, the largest plasmid to be sequenced in a LAB (Mayo *et al.*, 1989; Muriana & Klaenhammer, 1987; Roussel *et al.*, 1993). The original study by Claesson *et al* designated pMP118 as a megaplasmid because: (A) it does not encode the unique copy of any gene considered essential for viability of the cell (Claesson *et al.*, 2006) (B) it encodes neither rRNA nor tRNA genes

and (C) it harbours plasmid-related replication genes (Claesson *et al.*, 2006). The genome of *L. salivarius* comprises four replicons, two of which, plasmids pSF118-20 (20.4 kb) and pSF118-44 (44 kb), had previously been sequenced (Flynn, 2001). To further assess the plasmid complement of this strain, genomic DNA was treated with S1 nuclease, which preferentially linearises megaplasmid DNA and converts it into its equivalent unit length (Barton *et al.*, 1995). Pulsed-field gel electrophoresis (PFGE) of the treated DNA resolved a discrete band, below the chromosomal band, which showed consistent migration under various PFGE conditions, relative to a linear DNA marker (Claesson *et al.*, 2006). This electrophoretic mobility was consistent with that of a linearised plasmid. The newly found 242 kb plasmid was designated pMP118 (Claesson *et al.*, 2006). Annotation of the pMP118 sequence identified the presence of a plasmid-associated replication gene, *repA*. This gene was experimentally localized to pMP118 using *repA* as a probe for Southern hybridisation analysis of the resolved plasmid profile of *L. salivarius* UCC 118 (Claesson *et al.*, 2006).

1.4.3.3 Replication region of pMP118.

Replication of pMP118 is thought to proceed by theta replication, consistent with expectations for a plasmid of this size. Plasmid encoded Rep proteins are involved in initiating replication and facilitating binding of host proteins (Claesson *et al.*, 2006). The pMP118 *ori* region includes two Rep genes, *repA* and *repE* (del Solar *et al.*, 1998). The replication region also includes the presumptive chromosome partitioning ATPase, ParA, the hypothetical gene *parB* and three pseudogenes. The *repA* gene product displays 33% identity to the RepA/RepE protein of the *Enterococcus faecalis* theta-replicating plasmid pS86 (Claesson *et al.*, 2006; Martínez-Bueno, 2000) also the downstream *repE* gene is weakly related to staphylococcal plasmid replication proteins (Claesson *et al.*, 2006). However, neither RepA boxes nor interons were identified in the replication region of pMP118 and it has not yet been assigned to a plasmid replication family (Claesson *et al.*, 2006). The product of *parA* is 41% identical to a copy number control protein of the *Listeria innocua* plasmid pLI100 (Claesson *et al.*, 2006). The copy number of pMP118 was originally determined relative to the chromosome and was estimated by quantitative PCR to be 4.7+/- 0.6 copies (Claesson *et al.*, 2006). However more recent comparative genomic hybridisation data (Raftis and O' Toole, unpublished)

suggests that pMP118, is in fact present in only one copy per cell. In contrast both the chromosome and pMP118 have very similar GC contents (32.9% and 32.09%, respectively), suggesting that these two replicons are likely to have a coordinated evolutionary history whereas the two smaller plasmids may have been a more recent acquisition.

1.4.3.5 Insertion sequence elements and transposons.

pMP118 harbours twenty-five percent of the IS content of the genome, representing seven different IS families. We identified two pairs of tandem IS elements present on both the chromosome and pMP118, in near identical arrangements, including flanking DNA sequence but this remains to be verified by another method.

Table 1.4 Physical and genetic features of the *L. salivarius* UCC118 genome

Replicon	*L.salivarius* UCC118 chromosome	pMP118	pSF118-44	pSF118-20
Replicon size (bp)	1,827,111	242,436	44,013	20,417
Topology	Circular	Circular	Circular	Circular
% of total genome size	85.62	11.36	0.02	0.01
G+C content (%)	32.94	32.09	39.58	39.11
Gene density (%)	84.14	75.62	66.89	75.13
Predicted no. functional genes	1,117	222	48	27
Pseudogenes	49	20	4	0
tRNAs	78	0	0	0
rRNA operons	7	0	0	0
IS elements	32	11	1	2

1.4.3.4 GC content and coding capacity.

The major features of the *L. salivarius* UCC118 genome are outlined in Table 1.3. The 1.83 Mb chromosome is the smallest of the *Lactobacillus* chromosomes published to date. However, including the plasmid content of this strain, the 2.13 Mb genome is of a comparable size to the other sequenced lactobacilli, with the exceptions of *L. plantarum* (3.3 Mb) and *L. casei* (2.9 Mb). In comparison to the overall GC content of the genome (33.04%), both pSF118-20 and pSF118-44 have elevated GC contents (39.58 % and 39.11 % respectively) (Flynn, 2001) on this basis we proposed

that IS elements may have contributed to the formation or expansion of pMP118 by an integration of chromosomal genes into an earlier state of the pMP118 replicon (Claesson *et al.*, 2006), as previously suggested for the 191 kb archaeal megaplasmid pNRC100 (Ng *et al.*, 1998).

1.4.3.6 Pseudogenes.

A recent comparative genomics study of the LAB has shown that gene loss and metabolic simplification are major trends within the evolution of the *Lactobacillales* (Makarova *et al.*, 2006). At 49, the chromosome of strain UCC118 has one of the highest number (equal to *L. brevis*) of pseudogenes of the commensal lactobacilli of a similar genomic size. On a genome wide scale strain UCC118 includes 73 pseudogenes, with pMP118 making a substantial contribution (27%) to this number (Claesson *et al.*, 2006). A high number of pseudogenes may be indicative of a replicon in an active state of gene inactivation and elimination which is often associated with niche adaptation (Makarova *et al.*, 2006), as noted recently for the sequenced strain *L. delbrueckii* ssp. *bulgaricus* (van de Guchte, 2006). Of the 20 pseudogenes identified on pMP118, genes encoding proteins of unknown function are the most prevalent, followed by genes predicted to encode restriction modification systems, transposases, surface proteins and ABC transporters. Frame shifting was found to be the most common form of inactivation of genes associated with both pMP118 and the chromosome (Claesson *et al.*, 2006).

1.4.3.7 Contribution of pMP118 to genotype and phenotype.

Figure 1.2 illustrates the COG (clusters of orthologous groups) assignments of the genes harboured by pMP118. Over half of these genes encode hypothetical proteins, and so are not assigned to a COG category. A further thirteen are determined to have poor or no COG assignment predictions. Further functional analysis may help to characterize these proteins and determine their functional roles and by extension their role in the biology of UCC18. Of the genes that were assigned to COG categories, those associated with carbohydrate and amino acid transport and metabolism are highly represented in the genes carried by pMP118, as are genes associated with signal transduction and transcription (Fig. 1.2).

1.4.3.8 Amino acid metabolism.

pMP118 contributes neither rRNA nor tRNA genes to the strain, and so does not impart the advantages (e.g. rapid response to environmental change) that are associated with these genes (Klappenbach *et al.*, 2000). This enzyme catalyses the reversible reductive amination of pyruvate into alanine in the presence of NAD+. Adjacent to LSL_ 1768 is a gene encoding a putative alanine permease (LSL_1767). The chromosome of UCC118 encodes the enzymes *ldhL* and *ldhD*, which are involved in the generation of D-lactate. pMP118 harbors an additional copy of the *ldhD* gene (LSL_1887). However, pMP118 encodes a number of genes associated with increased biosynthetic capabilities. Among these (Fig. 1.2) is that for alanine dehydrogenase (LSL_1768, EC1.4.1.1) which is commonly found in the genus *Bacillus*, but is more rare among the sequenced lactobacilli having only been identified in *L. casei, L. brevis* (http://img.jgi.doe.gov/cgi-bin/pub/main.cgi. Cited 14 July 2008), and *L. salivarius* (Claesson *et al.*, 2006) Fig. 1.2, which may increase the efficiency of D-lactate production. This molecule is an important component of cell wall precursors in *L. plantarum* (Goffin *et al.*, 2005). pMP118 encodes both the alpha and beta subunits of L-serine dehydratase (EC4.3.1.17). This enzyme catalyses the interconversion of pyruvate to serine. pMP118 also harbors a paralog (LSL_1927) for one of two enzymes, encoded by the chromosome, that are required for the conversion of pyruvate to L-aspartate (Claesson *et al.*, 2006) serine can subsequently be converted to glycine by a hydroxymethyltransferase, which is chromosomally encoded. Serine may also be thiolated to cysteine by the chromosomally encoded enzyme CysK. Cysteine can then be further converted into methionine, again using chromosomally encoded genes. pMP118 encodes five genes involved in the biosynthesis of the aromatic amino acids, phenylalanine, tyrosine, and tryptophan. Three of these are paralogs of genes encoded by the chromosome and two are uniquely coded for by pMP118, but none of these complete the pathway. *In silico* analysis of the complete genome predicts that *L. salivarius* UCC118 can synthesize *de novo* or by interconversion nine amino acids, is capable of converting glutamine to three more and is auxotrophic for the remaining eight essential amino acids. By virtue of the unique and contingency genes that pMP118 harbours, *L. salivarius* UCC118 has an elevated level of prototrophy in comparison to

other enteric lactobacilli of a similar chromosomal size. This may confer a competitive advantage to strain UCC118 over other lactobacilli in the GIT and broaden the range of intestinal ecological niches in which it can survive, in contrast to other small genome lactobacilli such as *L. johnsonii* (Pridmore *et al.*, 2004).

1.4.3.9 Purine and pyrimidine anabolism.

The annotated genome suggests that *L. salivarius* UCC118 can synthesize purines and pyrimidines *de novo* from 5-phosphoribosyl-1-pyrophosphate (PRPP). Of note in this context, pMP118 harbours the genes that complete the pentose phosphate pathway (PPP) in strain UCC118 (see below for details), which may be used for the biosynthesis of PRPP. *L. salivarius* is predicted to generate UMP *via* a chromosomally encoded pathway, which can then be further converted to UTP and CTP by utilizing the megaplasmid encoded gene *dnk* (LSL_1936 diphosphate kinase) (Claesson *et al.*, 2006). It is also possible for these conversions to be catalyzed by the chromosomally encoded gene product pyruvate kinase (LSL_0867) (Kilstrup *et al.*, 2005).The ability to survive independently of host provisions of purines and pyrimidines may allow *L. salivarius* UCC118 to survive in a wider range of environments, than those lactobacilli with a more auxotrophic nature.

1.4.3.10 Acquisition of core chromosomal genes by pMP118.

By comparing the predicted proteomes of five *Lactobacillus* species, Canchaya *et al.* (2006) identified 75 *Lactobacillus* core proteins (Canchaya *et al.*, 2006). Interestingly, there were four genes present in the core that are carried by pMP118. LSL_1901 (Fig. 1.2) encodes a bifunctional acetaldehyde/ alcohol dehydrogenase which is the only enzyme present in UCC118 that catalyses the formation of ethanol from acetyl-CoA *via* acetaldehyde, thus providing an additional reductive pathway in UCC118 (Claesson *et al.*, 2006). The three remaining core genes harbored by pMP118 include those for L-serine dehydratase α subunit, ribulose-phosphate 3-epimerase (discussed below) and a paralog of a chromosomally encoded ribose-5-phosphate isomerase gene. Thus the presence of these genes on the megaplasmid may be indicative of mobilization of genes between the chromosome and pMP118 (Canchaya *et al.*, 2006).

1.4.3.11 Role of pMP118 in carbohydrate metabolism.

L. salivarius was originally described as homofermentative (Rogosa *et al.*, 1953). This would infer that like *L. gasseri* (Azcarate-Peril *et al.*, 2008)*, L. bulgaricus* (van de Guchte, 2006), *L. acidophilus* (Altermann *et al.*, 2005), sugars would only be fermented *via* the Embden-Meyerhof-Parnas pathway to ultimately produce lactate as a sole product (Kandler, 1983). The genes encoding the complete gylcolysis pathway were found to be encoded by the chromosome of strain UCC118 (Claesson *et al.*, 2006). However, we also annotated genes in the genome involved in the pentose phosphate pathway. This suggested the capability of fermenting pentose sugars as a sole carbon source, producing lactate, acetate and ethanol and CO_2 (Kandler, 1983), which was subsequently proven experimentally for strain UCC118 (Claesson *et al.*, 2006). pMP118 was also discovered to harbour a fructose-1,6- bisphosphate gene (LSL_1903). pMP118 uniquely encodes two enzymes (Fig. 1.2), transketolase (LSL_1946) and transaldolase (LSL_1888, LSL1947) which complete the pentose phosphate pathway that is partially encoded by the chromosome of *L. salivarius* UCC118. The contingency genes encoded by pMP118 are not essential for the biosynthesis of nucleotides when strain UCC118 is grown on glucose. However, the additional copy of a gene encoding ribose-5-phosphate isomerase (EC5.3.1.6) may increase the flexibility and flux of the pentose phosphate pathway (Claesson *et al.*, 2006) (Fig. 1. 2) which catalyses the formation of fructose-6-phosphate and which completes a gluconeogenesis pathway partly encoded by the chromosome. The presence of a complete gluconeogenesis pathway is unusual, with *L. case*i being the only other sequenced *Lactobacillus* genome to harbor both of these enzymes. The presence of a functioning pentose phosphate pathway, together with a complete gluconeogenesis pathway, may be an adaptation to pentose-based growth (Claesson *et al.*, 2006). A functioning pentose phosphate pathway would allow *L. salivarius* to utilize exogenous deoxyriboses, which are likely abundant in the GIT. The gluconeogenesis and the pentose phosphate pathways are prime examples of the synergistic functioning of megaplasmid and chromosomally encoded genes in the central metabolism of *L. salivarius* UCC118. *In silico* analysis also predicted that genes encoded by pMP118 would enable strain UCC118 to assimilate sorbitol and rhamnose. The rhamnose fermentation pathway utilises rhamnulokinase, L-rhamnose isomerase

and rhamnulose-1-phosphate aldolase, all of which are megaplasmid encoded genes (Claesson *et al.*, 2006). Sorbitol-6-phosphate 2-dehydrogenase is also a megaplasmid encoded gene, necessary for the catabolism of sorbitol in this strain, which is a common component of the human diet. Putative phosphotransferase transporters of sorbitol were also identified upstream of this enzyme.

1.4.3.12 Sensing and regulation.

UCC118 harbors nine two-component regulatory systems, two of which are encoded by pMP118, and one of which is the system involved in control of expression of the bacteriocin Abp118 (discussed below), AbpK-AbpR (Fig. 1.2)

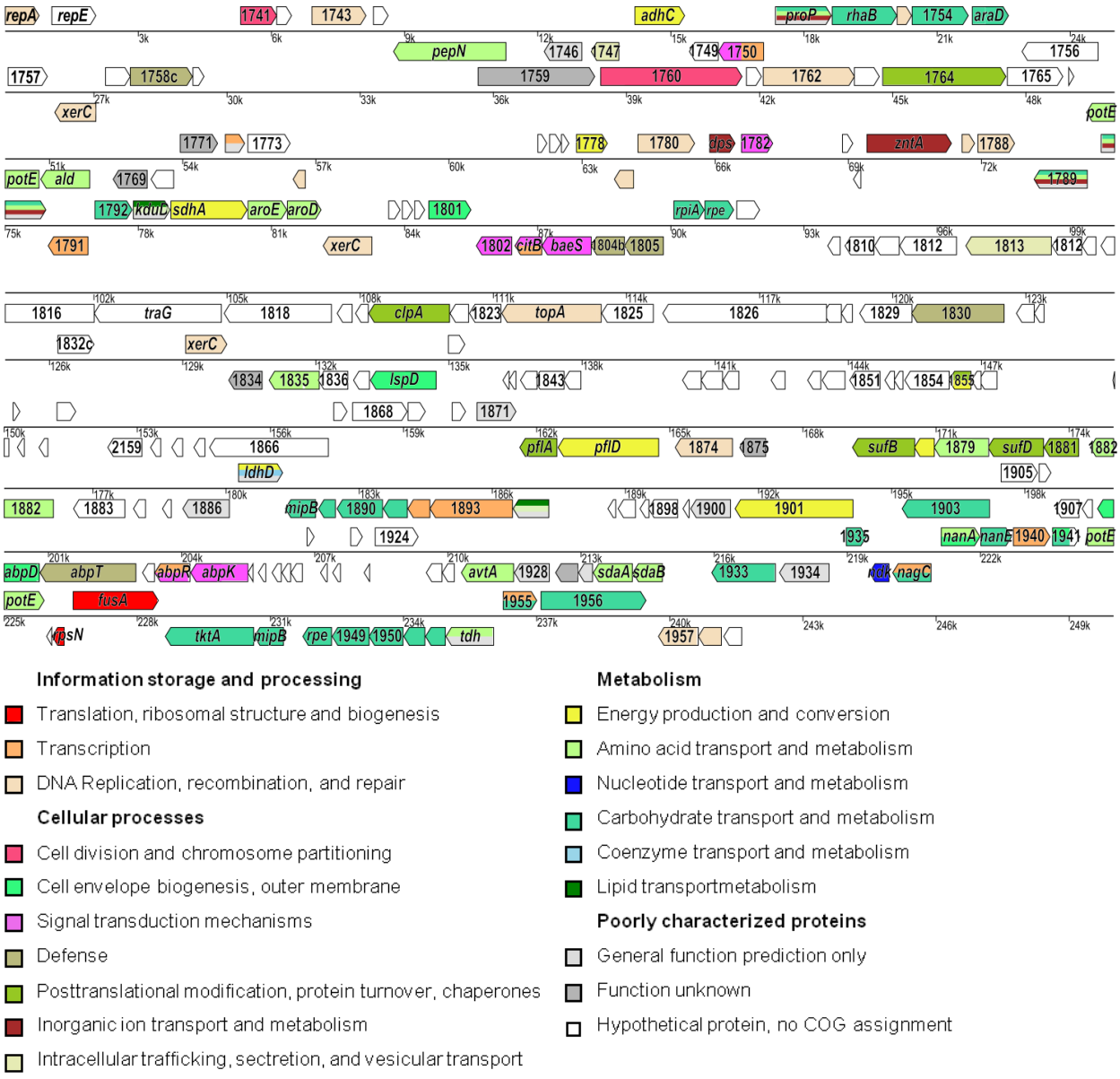


Figure 1.2: Linear genome map of pMP118, modified from an image generated by Microbial Genome Viewer v1.0 (Kerkhoven *et al.*, 2004), representing the COG assignments of genes harboured by pMP118. Each four digit number corresponds to the locus tag of the represented gene, which includes the prefix LSL_. For clarity of presentation, a number of genes remain unlabeled. Pseudogenes are excluded from COG assignment. The accompanying colour key indicates the COG assignments of each gene.

Two chromosomally encoded orphan sensors were also identified, one of which (LSL_1454), is a paralog of the megaplasmid located gene LSL_1802 (Claesson *et al.*, 2006). Seven of the 62 transcriptional regulator genes in the genome are present on pMP118 (Claesson *et al.*, 2006), thus emphasizing the potential effect that pMP118 may have on the regulated genome expression and phenotype of this strain.

1.4.3.13 Host interaction.

One of the most interesting properties of some *L. salivarius* strains including UCC118 is their probiotic nature. The contribution of extra-chromosomal genes to this phenotype is therefore interesting from biological and evolutionary perspectives.

1.4.3.14 Bile salt hydrolase (BSH) activity.

Enteric species encounter many host defense mechanisms in the GIT. Microbial exclusion *via* host production of bile is one such mechanism. The ability to tolerate bile imparts a selective advantage in the GIT environment (Tanaka *et al.*, 1999) and has proven essential for GIT persistence (Dussurget *et al.*, 2002). Many intestinal lactobacilli counteract the damaging effects of bile by encoding bile salt hydrolase enzymes (BSH) (Tanaka *et al.*, 1999). BSH catalyses the breakdown of bile salts, causing the release of taurine or glycine (Tanaka *et al.*, 1999). UCC118 is resistant to acid and bile (Dunne *et al.*, 1999). UCC118 was originally described as having both chromosomally encoded and megaplasmid encoded bile-inactivating enzymes (Claesson *et al.*, 2006). However, recent analysis suggests that pMP118 harbours the sole gene primarily responsible for bile-salt hydrolase activity in this strain (LSL_1801, choloylglycine hydrolase). LSL_1801 shares 53% sequence identity with a BSH gene in *L. monocytogenes*. The survival rate of a *L. salivarius* UCC118 LSL_1801-knock out (KO) strain, grown in porcine bile (0.02%), was shown to be reduced by greater than 2 logs in comparison to the wild type strain (Fang and O' Toole, unpublished). This illustrates the potential contribution that the megaplasmid encoded BSH gene plays in the survival and adaptation of UCC118 to the GIT environment.

1.4.3.15 Surface proteins.

The interaction of bacteria with the intestinal epithelium is thought to modulate immune responses and improve mucosal integrity (Tannock, 1999). Strain UCC118

encodes 108 secreted proteins by chromosomal genes. Ten of the surface proteins encoded by UCC118 are sortase dependent. A number of candidate host interaction proteins have been functionally analyzed in lactobacilli (Buck *et al.*, 2005) meaning they are covalently linked to peptidoglycan by the sortase protein (Pretzer *et al.*, 2005). Four of these proteins are encoded by pMP118. Only one of these four genes was annotated as functional (LSL_1838, *lspD*) (van Pijkeren *et al.*, 2006) (Fig. 1.2). However the LspD gene-product was not found to be a significant component in adherence of UCC118 to Caco C2 cells, whereas the disruption of the chromosomally encoded gene *lspA*, resulted in a significant adhesion reduction to this cell line (van Pijkeren *et al.*, 2006). Interestingly, one of the sortase-dependent pseudogenes (LSL_1774b) present on pMP118 was annotated as a caseinolytic cell-surface protease, PrtP (Claesson *et al.*, 2006) and shows homology to a *prtR* gene in *L. rhamnosus* (van Pijkeren *et al.*, 2006). The *prtP* gene is non-functional, and may in part explain the inability of strain UCC118 to grow in milk (Raftis and O'Toole, unpublished). Niche adaptation and selective pressures may have induced the loss of *prtP* function in this GIT associated organism (Makarova *et al.*, 2006; van Pijkeren *et al.*, 2006).

1.4.3.16 Microbe-Microbe interactions

A number of plasmids in *Lactobacillus* species are associated with bacteriocin production (Wang & Lee, 1997), which is a desirable trait in probiotic strains. *L. salivarius* UCC118 produces Abp118, a broad spectrum, heat labile, class II bacteriocin (Flynn, 2001). The genes responsible for the production of Abp118 were originally described as chromosomally encoded (Flynn *et al.*, 2002), but were later localized to pMP118 (Fig. 1. 2) when the genome was sequenced (Claesson *et al.*, 2006). Abp118 is regulated by a quorum-sensing mechanism, governed by the induction of peptide AbpIP (Flynn, 2001). The gene conferring immunity to the action of Abp118 is encoded downstream of the genes encoding the Abp118 α and Abp118 β components. Abp118 has been established as the primary mechanism by which protection against *L. monocytogenes* infection is mediated in the mouse model (Corr *et al.*, 2007). Furthermore, Abp118 has broad-spectrum activity (Flynn *et al.*, 2002) and is likely to enhance the competitiveness of UCC118 in the GIT and to potentially modulate the microbiota.

1.4.3.17 Megaplasmid distribution in L. salivarius and other species.

Prior to the identification of pMP118, many plasmids of the lactobacilli were determined by methods that pre-dated the application of PFGE for separation of large DNA molecules. The procedure outlined above, that resulted in the discovery of pMP118 was first developed by Barton *et al* and was later tailored for examining the complete plasmid profiles of *L. salivarius* and other species of the genus *Lactobacillus* (Barton *et al.*, 1995; Claesson *et al.*, 2006; Li *et al.*, 2007). A recent study by Li *et al* (2007) has revealed that the presence of megaplasmids is a general feature of *L. salivarius* (Li *et al.*, 2007). The presence of pMP118-related megaplasmids (100kb-380kb) was confirmed in 33 strains from diverse sources (Li *et al.*, 2007). These megaplasmids all hybridised to the *repA* gene probe derived from pMP118. Preliminary characterization of the strains indicated a diversity of genomic content, phenotypic characteristics, and megaplasmid size (100 kb -380 kb). Subsequent comparative genomic hybridisation analysis of this panel of *L. salivarius* has elucidated a high level of genome plasticity within this species (Raftis and O' Toole, unpublished data). Li *et al* (2007) also confirmed the presence of non-*repA*-type megaplasmids in an additional six *Lactobacillus* species, all of which can be found in the GIT (Li *et al.*, 2007).

1.4.3.18 Dissemination mechanism.

The mode of dissemination of megaplasmids in *L. salivarius* is as yet unclear. Interestingly, pMP118 was found to harbor a tract of genes (*tra* locus) that show relatedness to known or suspected conjugation genes in other species, including *E. faecalis* (Claesson *et al.*, 2006) and *L. lactis* (Fang *et al.*, 2008). The *tra* locus of pMP118 spans a 29.5 kb region and appears to represent a remnant plasmid transfer locus (Claesson *et al.*, 2006; Fang *et al.*, 2008). A recent study from this laboratory showed pMP118 is incapable of mobilizing smaller plasmids and that successful conjugation cannot be achieved with the complement of genes that are present on pMP118 (Fang *et al.*, 2008). Conjugation is a plausible mechanism through which pMP118-related plasmids have disseminated in *L. salivarius* (Fang *et al.*, 2008) and the tra region encoded by pMP118 may represent a remnant of a previously functioning conjugation pathway.

1.4.3.19 Concluding remarks.

It is evident from the *L. salivarius* genome annotation and functional studies that pMP118 contributes significantly to the biology of its bacterial host cell. pMP118 encodes a variety of proteins that work in conjunction with pathways encoded by the chromosome, to increase the metabolic flexibility of UCC118, while also contributing significantly to its probiotic properties. Megaplasmids of varying size are a general feature of the species *L. salivarius*. As a genetically well characterized strain-plasmid system, *L. salivarius* UCC118 provides a unique platform for investigating the potential development of megaplasmid-based replicons as cloning vectors for this species, or for cloning large operons in LAB. Further comparative genomic and functional analyses of environmentally diverse strains will help define the impact that size variation and genetic variation have on the genomic content and biological properties of *L. salivarius*.

1.4.4 Linear plasmids

*1.4.4.1 Distribution of linear plasmids in prokaryotes***.**

The discovery of a linear extra-chromosomal element was first reported in 1967 (Ravin & Golub, 1967). However, it is only over the past 15 years that linear plasmids have become better characterized in prokaryotes. A growing number of studies have examined the structure and functionality of linear plasmids in a range of gram positive and gram negative species (see review (Meinhardt *et al.*, 1997)). Perhaps the most extensively studied linear plasmids belong to the genus *Streptomyces*, members of which garnered particular interest due to their ability to produce antibiotics. The first description of a linear plasmid in a gram-positive species was reported in *Streptomyces rochei* (Hayakawa *et al.*, 1979) and they have subsequently been identified in numerous *Streptomyces* sp. (Table 1.3) (Chater & Kinashi, 2007; Kinashi *et al.*, 1987; Medema *et al.*, 2010; Netolitzky *et al.*, 1995). Noteworthy is strain *Streptomyces clavuligerus* ATCC 27064, which harbours one of the largest linear plasmids ever identified and sequenced, pSCL4 (1.8 Mb) (Medema *et al.*, 2010).

Linear plasmids have also been documented in the following genera: *Agrobacterium* (Goodner *et al.*, 2001), *Borrelia* (Fraser *et al.*, 1997)*, Bacillus* (Carlson *et al.*, 1992; Stromsten *et al.*, 2003; Verheust *et al.*, 2003), *Rhodococcus* (Crespi *et al.*, 1992; Kalkus *et al.*, 1990; Kalkus *et al.*, 1993; Konig *et al.*, 2004; Kosono *et al.*, 1997; Shimizu *et al.*, 2001; Stecker *et al.*, 2003)*, Mycobacterium* (Coleman & Spain, 2003; Le Dantec *et al.*, 2001; Picardeau & Vincent, 1997; Picardeau & Vincent, 1998) *Pseudomonas* (Danko *et al.*, 2004)*, Xanthobacter* (Krum & Ensign, 2001)*, Ochrobactrum* (Danko *et al.*, 2004)*, Clavibacter* (Brown *et al.*, 2002)*, Planobispora* (Polo *et al.*, 1998)*, Thiobacillus* (Wlodarczyk & Nowicka, 1988)*, Klebsiella* (Stoppel *et al.*, 1995) and more recently in *Micrococcus* (Dib *et al.*, 2010a) and *Brevibactierium* (Dib *et al.*, 2010b). Linear plasmids are also commonly found as extra-chromosomal elements in both the planta and fungi kingdoms of Eukaryota (Klassen & Meinhardt, 2007). A detailed compilation of microbial linear plasmids is presented in Meinhardt and Klassen (2007) (Meinhardt & Klassen, 2007) but will not be discussed further in this introduction.

Despite the widespread presence of linear replicons among Gram positive bacteria, they are not generally associated with LAB species. Prior to the discovery of linear megaplasmids in *L. salivarius* (Li *et al.*, 2007) there had been a single report of a 150 kb linear plasmid present in *Lactobacillus gasseri* CNRZ222 (Roussel *et al.*, 1993). Further characterisation of this element was not carried out and the genetic compliment and replication mechanisms utilized by linear elements in *Lactobacillus* remain unknown. The linear plasmid identified in *L. gasseri* is in the same size range as the three linear plasmids of *L. salivarius* described by Li *et al* in 2007 (JCM1046 (140 kb), JCM1047 (140 kb) and AH43348 (175 kb).

1.4.4.2 Structural organization of linear plasmids.

The structural organization of linear plasmids together with their modes of replications are used to functionally define the two major groups under which these elements are most often described (Meinhardt *et al.*, 1997) the hairpin and the invertron-type linear plasmids. Hairpin plasmids have short covalently closed terminal inverted repeats (TIRs) at their telomeres (Barbour & Garon, 1987; Casjens *et al.*, 1997; Hinnebusch & Barbour, 1991; Kobryn, 2007), in contrast to the invertron-type linear plasmids which are open ended with 44 bp (Chen *et al.*, 1993) to 95 kbp (Gravius *et al.*, 1994) TIRs at their telomeres and have covalently attached proteins linked to their 5' termini (Fig. 1.3). Hairpin plasmids are best represented by the linear plasmids of *Borrelia* species (sp.) (Chaconas & Kobryn, 2010; Kobryn, 2007; Schutzer *et al.*, 2011; Stewart *et al.*, 2005; Takahashi *et al.*, 2000) but are also commonly identified in *Agrobacterium tumefaciens* sp (Goodner *et al.*, 2001), and in a small number of bacteriophages (Chaconas, 2005). Invertron-like plasmids are structurally similar to the linear genomes of viruses and transposons (Sakaguchi, 1990) and range in size from less than 10kb (Chater & Kinashi, 2007) to 1.8 Mb (Medema *et al.*, 2010)**.** The invertron-type plasmids are the largest group of extra-chromosomal linear replicons and are commonly found in the phylum Actinobacteria (Chater & Kinashi, 2007; Chen, 2007; Meinhardt *et al.*, 1997) and are best studied in the genus *Streptomyces* (Chater & Kinashi, 2007; Chen, 2007) (Table 1.3) due in part to their value as prolific antibiotic producers (Chater, 2006; Kinashi, 2011). Several linear phage genomes have also been well-studied, including N15 of *E. coli* (Ravin *et al.*, 2000), PY54 of *Yersinia*

enterocolitica (Hertwig *et al.*, 2003; Popp *et al.*, 2000), and KO2 of *Klebsiella oxytoca* (Casjens *et al.*, 2004). N15 is largely homologous to the genomes of other lamdoid phage (Ravin *et al.*, 2000) but is unusual among the bacteriophages of *E. coli* as it forms a linear double-stranded plasmid, with single stranded cohesive (12-bp) overhangs during lysogeny (Ravin, 2011). ΦKO2 and PY54 share homology with genes that are associated with the linear plasmid lifestyle of N15, including the protelomerase, plasmid-partitioning and packaging genes and as well as a *repA*-like replicase (Hertwig, 2007). Given their similar genome organizations, it has been proposed that this group of N15-like phage have diverged from a common ancestor (Ravin, 2011). The progenitor is thought to be a lambdoid phage that acquired additional genetic modules or a plasmid that acquired a lambdoid set of ‘‘virion’’ genes (Ravin, 2011). More recently, a 142 kb linear bacteriophage, φLb338,was identified in the probiotic cheese strain *Lactobacillus paracasei* NFBC 338 (Alemayehu *et al.*, 2009).

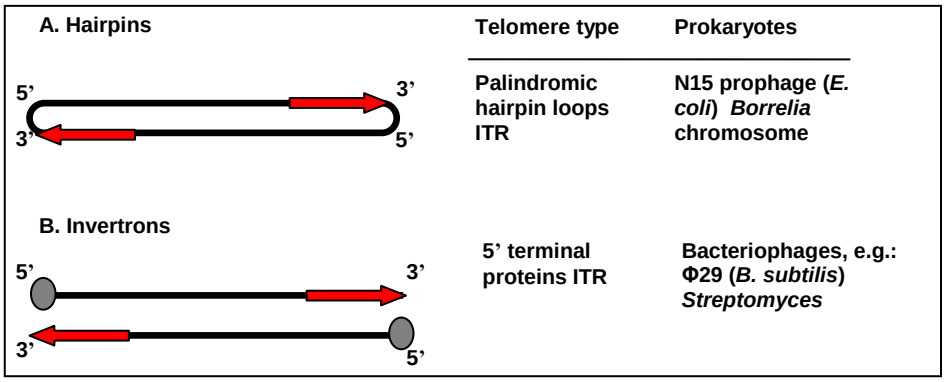


Figure 1.3. Structure of bacterial linear plasmids. Inverted terminal repeat are indicated by red arrows. Filled circles represent covalently attached proteins. Adapted from (Hinnebusch & Tilly, 1993).

1.4.4.3 Linear plasmid replication mechanisms.

During replication, linear replicons encounter several problems which their circular counterparts do not (Chen, 1996; Meinhardt *et al.*, 1997) and the modes by which linear chromosomes and plasmids replicate has been the subject of increased study over the last 20 years (Chang & Cohen, 1994; Chang & Cohen, 1996; Chen, 1996; Hiratsu *et al.*, 2000; Huang *et al.*, 1998; Molin *et al.*, 1993b; Qin & Cohen, 1998; Qin *et al.*, 2003; Redenbach *et al.*, 1999; Shiffman & Cohen, 1992)}. However, discussion of the modes of replication employed by the two major structural types of linear replicons is beyond the scope of this introduction and is discussed in detail in the following references (Chen, 2007; Hertwig, 2007; Kobryn, 2007).

In order to identify the genes present in the replication region of the linear plasmids of *L. salivarius* (Chapter III), Southern hybridisation experiments were carried out using probes associated with the replication regions of the *rep*A-type circular megaplasmids of strain UCC118 (Li *et al.*, 2007). The results indicated that the partitioning of plasmid pLMP43348 is likely to be dependent upon a *parA*-homolog which is harboured by the circular megaplasmids of *L. salivarius* UCC118 (Claesson *et al.*, 2006). Replication-related probes (*repA* and *repE*), associated with the circular megaplasmids of *L. salivarius* failed to hybridize to the DNA of any of the linear megaplasmids of *L. salivarius*. Thus it was considered likely that a novel mechanism of replication was employed by these linear elements (Li *et al.*, 2007), or at the very least, that replication-associated genes that are related to the circular megaplasmids of *L. salivarius* are not encoded by their coresident linear plasmids (Li *et al.*, 2007).

1.4.4.4 Effects of linear plasmids on host adaptation.

A number of linear plasmids have been shown to carry genes that encode important metabolic functions which have the capacity to effect the fitness of their hosts (Norris *et al.*, 2010). Within the genus *Borrelia* are the human pathogens *B. burgdorferi*, which causes Lyme borreliosis (Burgdorfer *et al.*, 1982) and *B. hermsii* which causes relapsing fever maladies (Dworkin *et al.*, 2002). Both of these species harbor numerous circular and linear plasmids in addition to linear chromosomes (~ 1000 kbp) (Fraser *et al.*, 1997). The importance of plasmids for the lifestyle of *Borrelia* species is underscored by the fact that all isolates examined have been found to be rich in extra-chromosomal elements, some of which have been described as essential genetic elements (Stewart *et al.*, 2005), or mini-chromosomes (Barbour & Zöckert, 1997). The genome of *B. burgdorferi* type-strain B31 (Casjens *et al.*, 2000; Fraser *et al.*, 1997) was found to be comprised of a small linear chromosome (911kb), twelve linear and nine circular extra-chromosomal elements, contributing 40 percent of the total coding capacity of the genome (Bentley & Parkhill, 2004; Casjens, 1998). In addition, the recombination potential of the telomeric regions of the linear plasmids is known to facilitate the antigenic variation systems found in both Lyme disease (Norris, 2006; Zhang *et al.*, 1997) and relapsing fever spirochetes (Barbour & Guo, 2010; Dai *et al.*, 2006; Kitten & Barbour, 1990; Tabuchi *et al.*, 2002) and contribute to the dramatic genetic variability often observed in the linear plasmids of this and related species which harbor analogous linear plasmids (Chaconas & Kobryn, 2010; Iyer *et al.*, 2003). The presence and topology of the linear plasmids of pathogenic *Borrelia* species are known to contribute to the pathogenic lifestyle of the host organism (Casjens *et al.*, 2010). *Borrelia* species have been shown to evade the immune recognition and clearance mechanisms of their mammalians hosts by continually varying cell surface proteins through elaborate gene conversions (Bankhead & Chaconas, 2007; Zhang *et al.*, 1997; Zhang & Norris, 1998b). VlsE is a surface-localized lipoprotein that is located in the subtelomeric region of plasmid lp28-1 and is thought to be involved in chronic infection and immune evasion (Zhang & Norris, 1998a) by facilitating antigenic variation (Zhang *et al.*, 1997; Zhang & Norris, 1998b). Many of the pathogenicity determinants associated with dissemination, infection and persistence are plasmid localized in *B.*

burgdorferi which have been found in sizes which range between 5 kb and 220 kb (Chaconas & Kobryn, 2010; Purser & Norris, 2000; Stewart *et al.*, 2005).

Bacteria of the phylum *Actinobacteria* harbor conjugative linear plasmids (Meinhardt *et al.*, 1997; Picardeau & Vincent, 1998; Ravel *et al.*, 1998; Ravel *et al.*, 2000) which often confer advantageous phenotypes to their hosts, including enhanced catabolic capabilities (Coleman & Spain, 2003; Dabrock *et al.*, 1994; Danko *et al.*, 2004; Fetzner *et al.*, 2007; Konig *et al.*, 2004; Kosono *et al.*, 1997; Overhage *et al.*, 2005; Shimizu *et al.*, 2001), hydrogen autotrophy (Kalkus *et al.*, 1990; Kalkus *et al.*, 1993), virulence determinants (Crespi *et al.*, 1992; Goethals *et al.*, 2001), heavy metal resistance genes (Dabrock *et al.*, 1994; Ravel *et al.*, 1998; Stecker *et al.*, 2003), and secondary metabolite production (Suwa *et al.*, 2000). In some cases linear replicons have also been associated with the production of antibiotics and the transfer of antibiotic associate genes (Kinashi *et al.*, 1987; Kinashi, 2011). Similar to their circular counterparts, linear plasmids are not generally essential for the viability of their hosts but often extend their metabolic versatility, thus conferring a selective advantage.

1.4.4.5 Biotechnological applications.

N15 and N15-like linear replicons have many biotechnological applications, which are partly responsible for the extensive characterisation that has been carried out for these replicons since their discovery. The linear nature of the N15 prophage has led to its exploitation as a cloning vector (Mardanov & Ravin, 2007). The topology of this phage together with the associated lack of supercoiling reduces the possibility of secondary structure and cruciform formation which favors sequence deletion and recombination (Ravin, 2011). Many N15-based linear plasmids are proving particularly useful as cloning vectors for sequence containing inverted repeats (Godiska *et al.*, 2009). The application of new sequencing technologies offers new opportunities to study the genome structure and functions of linear replicons in detail. However, as emphasized in a recent paper by Wagenknecht *et al* (2010), these elements pose specific challenges to next generation sequencing technologies (Wagenknecht *et al.*, 2010). Structural peculiarities of the linear megaplasmid (pLMA1) in *Micrococcus luteus* were shown to interfere with the assembly of pyrosequencing reads (Genome Sequencer FLX system (Roche Applied Science), making assembly of the replicon impossible.

Alignment of 454 sequences to the Sanger-sequenced clones of the pLMA1 fragments, spanning repeat regions and a large number of putative transposase genes was necessary for completion of sequencing project (Wagenknecht *et al.*, 2010).

Table 1. 4 Selected examples of species that frequently harbour linear plasmids

Organism	Plasmid	Size (kbp)	Phenotype	Reference
Bacteria				
Rhodococcus sp (formerly *Nocardia opaca*)	pHG201	270	Autotrophy	(Kalkus *et al.*, 1990; Kalkus *et al.*, 1993)
	pHG204	180	Thallium resistance	(Kalkus *et al.*, 1990; Kalkus *et al.*, 1993)
	pHG205	280	Autotrophy	(Kalkus *et al.*, 1990; Kalkus *et al.*, 1993)
Rhodococcus erythropolis	pBD2	210	Isopropylbenzene and trichlorethene catabolism, arsenite and mercury resistance	(Dabrock *et al.*, 1994; Kesseler *et al.*, 1996)
Rhodococcus fascians	pFiD188	200	Induction of fasciation	(Crespi *et al.*, 1992; Crespi *et al.*, 1994)
Rhodococcus RHA1	pRHL1	1,123	Catabolism of aromatic compounds	(McLeod *et al.*, 2006)
	pRHL2	442.5	Catabolism of aromatic compounds	(McLeod *et al.*, 2006)
	pRHL3	332.4	Catabolism of aromatic compounds	(McLeod *et al.*, 2006)
Streptomyces clavuligerus	pSCL1	11.7	Cryptic	(Wu & Roy, 1993)

Organism	Plasmid	Size (kbp)	Phenotype	Reference
Bacteria				
Streptomyces clavuligerus	pSCL4	1,796	Biosynthesis of secondary metabolites	(Medema *et al.*, 2010)
Streptomyces coelicolor	SCP1	350	Methylenomycin synthesis	(Kinashi & Shimaji, 1987; Kinashi *et al.*, 1993)
Streptomyces fradiae		420	Tylosin synthesis	(Kinashi & Shimaji, 1987)
Streptomyces lasaliensis	pKSL	520	Lasalocid A synthesis	(Kinashi *et al.*, 1987)
Streptomyces parvulus		520	Actinomycin D synthesis	(Kinashi & Shimaji, 1987)
Streptomyces rimosus	pPZG101	387	Cryptic	(Gravius *et al.*, 1994)
Streptomyces venezuelae	Unnamed	130	Chloramphenicol synthesis	(Kinashi & Shimaji, 1987)
Lactobacillus gasseri		150	Cryptic	(Roussel *et al.*, 1993)
Lactobacillus salivarius	pLMP43348		Cryptic	(Li *et al.*, 2007)
	pLMP1046		Cryptic	(Li *et al.*, 2007)
	pLMP1047		Cryptic	(Li *et al.*, 2007)
Lactobacillus equi	pLMP15833A	130	Cryptic	(Li *et al.*, 2007)
	pLMP15833B	230	Cryptic	(Li *et al.*, 2007)
	pLMP15833C	440	Cryptic	(Li *et al.*, 2007)
Micrococcus luteus	pLMA1	110	Putative erythromycin resistance	(Dib *et al.*, 2010a)
Micrococcus	pLMH5	110	Cryptic	(Dib *et al.*, 2010a)

Organism	Plasmid	Size (kbp)	Phenotype	Reference
Bacteria				
Micrococcus	pLMV7	90	Cryptic	(Dib *et al.*, 2010a)
Micrococcus luteus	pLMA1	110	Cryptic	(Wagenknecht *et al.*, 2010)
*Brevibacterium*Ap13	pAP13	87	Cryptic	(Dib *et al.*, 2010b)
Borrelia burgdorferi B31	13 linear plasmids	5.2-22.6	Virulence determinants, de novo pyrimidine synthesis	(Casjens *et al.*, 2000; Fraser *et al.*, 1997)
Borrelia hermsii	9 linear plasmids	2.6-22	Virulence determinants	(Dai *et al.*, 2006)

1.5 Conjugative transposons: Tn*916* as a model

1.5.1 Introduction to conjugative transposons

Horizontal gene transfer contributes greatly to the evolution of bacterial populations and is often mediated by the transfer of mobile genetic elements, including plasmids, bacteriophage, and transposons (Frost *et al.*, 2005; Juhas *et al.*, 2009; Ochman *et al.*, 2000). Conjugative transposons (CTs) were first discovered in the late 1970s (Franke & Clewell, 1981) and form a subgroup within a larger class of mobile elements known as Integrative Conjugative Elements (ICEs) (Burrus *et al.*, 2002). CTs vary widely in their genetic organization and in the numbers and types of functional and accessory genes that they encode (Roberts *et al.*, 2008). CTs are often difficult to classify as they combine many features of transposons, plasmids, and phage (Osborn & Böltner, 2002). A large number of genetic elements are covered by the term CT and ICE, most of which are beyond the scope of this review (see reviews (Burrus *et al.*, 2002; Salyers *et al.*, 1995) and references therein) The distinction between these elements is made ambiguous because of their structural similarities. Despite the introduction of a robust nomenclature scheme for transposons by Roberts *et al* (Roberts *et al.*, 2008), the terms ICEs and conjugative transposons are often used interchangeably. Roberts *et al* describes groups of mobile genetic elements as a "genetic continuum" in which elements interact and undergo frequent reshuffling and recombination which then generates novel chimeric elements (Roberts & Mullany, 2009).

The transfer intermediates of conjugative transposons resemble conjugative plasmids as they are both covalently closed circular elements that are transferred by conjugation (Salyers *et al.*, 1995). In addition, many ICEs and conjugative plasmids utilise similar mechanisms and machinery for DNA transfer between cells (Toussaint & Merlin, 2002). Conjugative plasmids, however, are known to replicate autonomously within a host cell whereas until recently it was accepted that CTs were incapable of autonomous replication (Burrus & Waldor, 2004). However, this is not the case for Actinomycete integrative and conjugative elements which are capable of autonomous replication (te Poele *et al.*, 2008) and in a recent study by Lee *et al* (2010), autonomous plasmid-like replication of a CT (ICEBs1) in Bacillus subtilis (Lee *et al.*, 2010) was demonstrated. This study also proposes that autonomous

plasmid-like replication is a common property of ICEs and that it contributes to the maintenance of mobile genetic elements in bacterial species (Lee *et al.*, 2010). Also considered a discriminating factor is the routine integration of CTs into either the chromosome or plasmid of a host genome (Salyers *et al.*, 1995). There are however also many examples of plasmids that reside in Streptomyces sp. that are known to integrate into the chromosome of some strains while replicating as plasmids in others (Hopwood, 2006). Thus the distinction between CTs and microbial plasmids is still somewhat difficult to apply consistently.

Transposable elements are defined by Roberts *et al* (2008) as being "specific DNA segments that can repeatedly insert into one or more sites in one or more genomes" (Roberts *et al.*, 2008) and references therein). Movement and integration of conjugative transposons either within or between cells involves their transposition from one site to another, as illustrated in Fig 1.4. CTs are somewhat similar to transposons as they are excised from and integrate into DNA. However, they differ from the well studied transposons (Tn5 and Tn10) in a number of ways: (I) the mechanism that they use to achieve transposition (Salyers *et al.*, 1995) (II) CTs do not induce duplication of the target DNA into which they insert (Salyers *et al.*, 1995) and (III) CTs form covalently closed circular transposition intermediates following excision. CTs employ a similar excision and integration system as temperate bacteriophage and many CTs encode integrases that share limited sequence homology with the lambda integrase family (Poyart-Salmeron *et al.*, 1990).

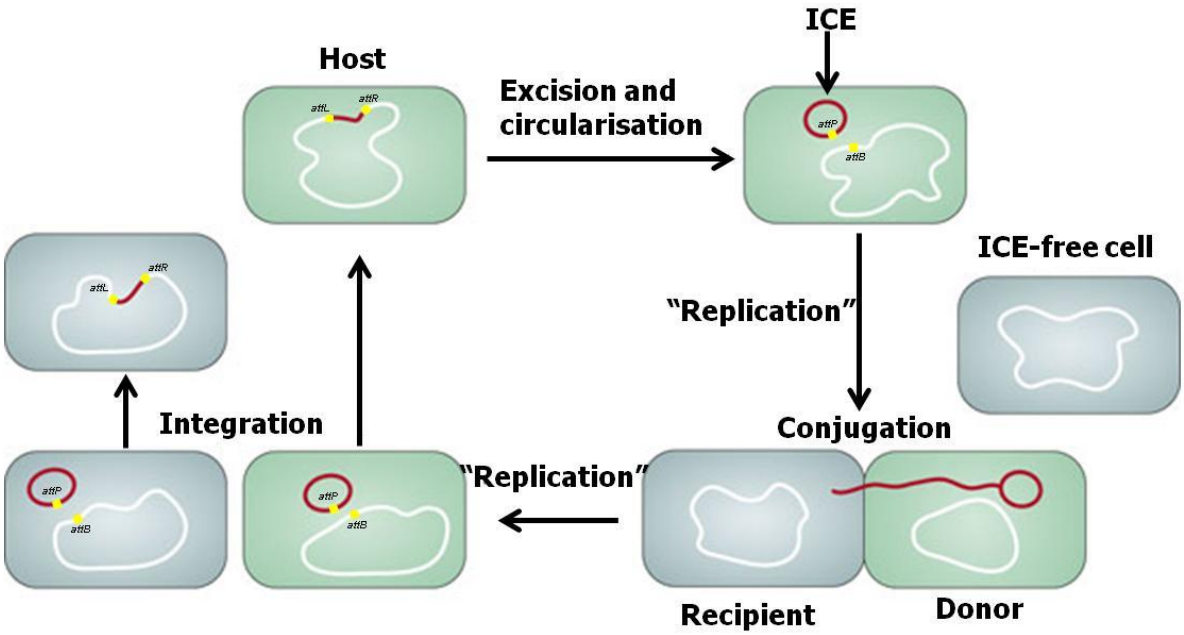


Figure 1.4. **Schematic of a typical integrative and conjugative element life cycle. Figure modified from (Wozniak & Waldor, 2010). An integrative and conjugative element (ICE) is integrated into the host replicon and is bound by specific sequences on the right (*attR*) and left (*attL*) of the element. Excision of the ICE occurs as a result of recombination between *attL* and *attR* to yield *attP* (in the ICE) and *attB* (in the host) and produces a covalently closed circular molecule. During conjugation, a single DNA strand is transferred from a donor cell to the new ICE-free "recipient" by the action of rolling circle replication. Following transfer, the double-stranded circular form of the ICE is regenerated using DNA polymerase in the recipient cell. Integration into the host replicon occurs following a recombination event between *attP* and *attB* (Wozniak & Waldor, 2010).**

1.5.2 Core genome.

The global structure of all self-transmissible ICEs follows a similar pattern, which is composed of a minimum of three functional modules, which can be exchanged with other mobile elements or with their host genome (Burrus & Waldor, 2004) (Fig 1.5). These include regions of the genome that govern specific functions including maintenance, dissemination, and regulation; for further details see Burrus *et al* (2004) (Burrus & Waldor, 2004).

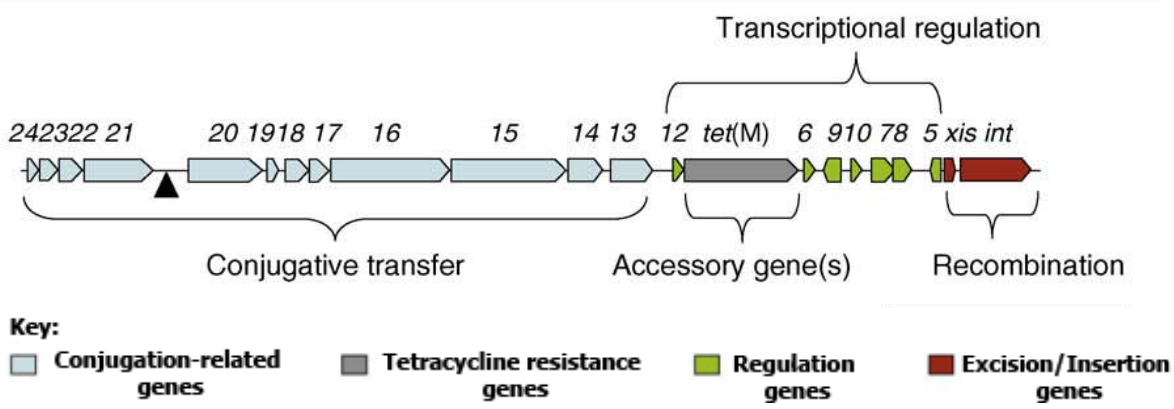


Figure 1.5. **A schematic representation of the four functional modules of Tn*916*. The *oriT* is represented by a black triangle, the accessory module is flanked by genes involved in the transcriptional regulation of the element, though this varies among other Tn*916*-like elements. The *ORFs* have been named as previously described (Roberts *et al.*, 2001b). Modified from (Roberts & Mullany, 2009)**

1.5.3 Tn*916*-like elements

1.5.3.1 Distribution of Tn916-like elements.

One of the best characterized CT is Tn*916* (18.5 kb) which was first described in *Enterococcus faecalis* DS16 (Franke & Clewell, 1981) and later sequenced (Flannagan *et al.*, 1994) and transfer between a wide variety of bacterial phyla and genera including Firmicutes, Proteobacteria and Actinobacteria. Tn*916* is the prototype of a family of conjugative transposons, which are highly promiscuous (Clewell *et al.*, 1995) both in the lab and in natural environments (Bertram *et al.*, 1991; Courvalin, 1994). Transfer frequencies are generally lower between gram-positive and negative bacteria (Scott, 2002; Trieu-Cuot *et al.*, 1988). Tn*916*-like elements are defined by a commonality of certain characteristics. The DNA sequence and functional modules of their core region, which govern conjugation, regulation, recombination and accessory genes are similar to that of the original Tn*916* element and generally resemble the organization seen in Fig 1.5 (Roberts & Mullany, 2009).

1.5.3.2 Transposition mechanism of Tn916-like elements.

Conjugation has allowed Tn*916*-like elements to disseminate into a wide range of host cells (Clewell *et al.*, 1995; Franke & Clewell, 1981). Transfer of Tn*916* is most often accompanied by integration into the recipient chromosomes but integration into the plasmids of recipients has also been reported (Scott *et al.*, 1994) (Fig 1.4). Transposition of Tn*916* is thought to occur in a manner similar to that of

conjugative plasmids. Excision and integration of Tn*916* (Fig. 1.4) requires the integrase protein Int and the excisionase Xis. The *int* gene encodes a tyrosine recombinase which is necessary for the integration and excision of the element. *xis* encodes the accessory protein Xis, which regulates excision of the CT (Hinerfeld & Churchward, 2001) and is responsible for of the directionality of the transposition reaction (Roberts & Mullany, 2009), but is not essential under all conditions (Rudy *et al.*, 1997). The integrase (*int*) and excision (*xis*) genes are located at one end of the CT in all Tn*916*-related elements and are downstream of *tetM* (Roberts & Mullany, 2009). During excision staggered cuts are made at each termini of the integrated transposons producing coupling sequences at the 5'hydroxyl ends of the element which are generally 6-nucleotide single-stranded stretches of nonhomologous DNA (Manganelli *et al.*, 1996; Rudy & Scott, 1994) but the lengths of the coupling sequences have been shown to vary among Tn*916*-like elements (Rice & Carrias, 1994). The single stranded sequences are then ligated to form a nonreplicating, double stranded circular intermediate (Caparon & Scott, 1989). The joint between the ends of the transposon in the circular intermediate is often described as forming a heteroduplex (Caparon & Scott, 1989; Manganelli *et al.*, 1996) although homoduplex joints have also been identified in a limited number of species (Manganelli *et al.*, 1997). A DNA relaxase then nicks the conjugative element (Rocco & Churchward, 2006) at the origin of transfer (*oriT*) (Jaworski & Clewell, 1995), which is followed by the transfer of a single strand of the intermediate into a recipient cell (Scott *et al.*, 1994). A complimentary strand is then synthesized in both the donor and recipient cells, thus regenerating the double-stranded circular element, which must be completed before the transposons can integrate into a target site (Fig. 1.4).

Semi-conservative replication or mismatch repair then resolves the heteroduplexes present at the sites of integration and excision (Rice & Carrias, 1994). Integration is thought to be a reversal of the excision process (Caparon & Scott, 1989). Tn*916* demonstrates semi-random integration, however, distinct preferences for AT-rich regions have been demonstrated, with a generic consensus target sequence of 5'-TT/ ATTTT(N6)AAAAAA/TA-3 (Wilson, 2006). Tn*916*-like elements that encode serine instead of tyrosine recombinases show different target site specificities (Wang *et al.*, 2000).

Tn*916*-like CTs are described as "sociable" mobile elements, which do not exclude each other (Norgren & Scott, 1991) as is the case for many incompatible plasmids and a number of standard transposons (Clewell *et al.*, 1995). Although Tn*916* is generally incapable of mobilising plasmids in *cis* it can mobilize co-resident plasmids in *trans* (see review (Salyers *et al.*, 1995). The regulatory mechanisms of Tn*916*-like elements are not well understood and what is known is mostly limited to Tn*916* (Celli & Trieu-Cuot, 1998; Roberts & Mullany, 2009; Su *et al.*, 1992). Transmission of Tn*916* is governed by regulatory networks, which responds to environmental cues. In conjunction with host factors, Tn*916* responds to these signals by altering gene expression and transfer levels (Mullany *et al.*, 2002). There are reports that tetracycline, in subinhibitory concentrations, stimulates the transfer Tn*916*-related conjugative transposons (Doucet-Populaire *et al.*, 1991; Showsh & Andrews, 1992; Torres *et al.*, 1991), in one case up to a 100-fold increase (Showsh & Andrews, 1992). Transcriptional analyses (Celli & Trieu-Cuot, 1998) support the idea that this region is governed by a transcriptional attenuation mechanism (Su *et al.*, 1992). The Tn*916* CT is able to respond to the presence or absence of tetracycline rapidly and alter transcription and translation events accordingly (Roberts & Mullany, 2009).

1.5.3.3 Antibiotic resistance and accessory genes.

The introduction of a novel CT can have both beneficial and negative effects on the recipient cell. In addition to this, CTs can further effect the evolution of a host cell through genomic rearrangements, deletion of nucleotides upon excision (Manganelli *et al.*, 1996), disruption of the integration site (Smidt *et al.*, 1999), can cause transcriptional disruption of upstream and downstream genes through polar effects (Ike *et al.*, 1992) and can form gene fusions upon insertion (Sebaihia *et al.*, 2006). Thus, in addition to the introduction of functional modules to the host cell, CTs have further potential influence on natural selection within population of bacteria through their physical integration with the host genome.

The overuse of antibiotics is regarded as one of the major causes of the accumulation and spread of antibiotic resistance genes within the environment (Levy & Marshall, 2004) and it has been suggested that commensal bacteria may act as natural reservoirs for antibiotic resistance determinants (Perreten *et al.*, 1997; Schjørring & Krogfelt, 2011) which are often disseminated by mobile elements

including plasmids and conjugative transposons (Ammor *et al.*, 2007). Most Tn*916*-like elements encode the antibiotic resistance gene *tetM* (Flannagan *et al.*, 1994), which mediates resistance through a ribosomal protection mechanism (Taylor & Chau, 1996), with others conferring resistance to other drugs such as kanamycin and macrolides (Courvalin & Carlier, 1986). The acquisition of antibiotic resistance in bacteria is generally associated with a decrease in biological fitness which generally results in a reduced bacterial growth rate (Andersson & Levin, 1999). However, studies have shown that "compensatory evolution can stabilize resistant bacterial populations in the absence of antibiotics by making them as fit as susceptible clones" (Andersson & Hughes, 2010). It is thought that the cost of resistance may also be alleviated by the acquisition of additional fitness-compensatory mutations (Andersson & Hughes, 2010). In Tn*916*-like elements, the accessory gene *tetM* is replaced in some cases (not always at the same position) by a range of other accessory genes in many Tn*916*-like elements, including: alternative antibiotic resistance genes (Lancaster *et al.*, 2004), putative subtilisin immunity genes (Rice *et al.*, 2007), ABC transporters (Sebaihia *et al.*, 2006), mercury resistance genes (Soge *et al.*, 2008), collagen-binding genes (Rice *et al.*, 2007), and putative cell-surface proteins (Sebaihia *et al.*, 2006).

Genes which confer resistance to commonly used antibiotics including vancomycin, erythromycin and tetracycline have all be identified in species that are commonly used as starter cultures and in the food manufacturing industry including *Enterococcus*, *Lactobacillus* and *Lactococcus* (Ammor *et al.*, 2007; Gevers *et al.*, 2003; Mathur & Singh, 2005; Rizzotti *et al.*, 2009). In addition, Tn*916*-like elements have demonstrated intra- and inter species transfer from *L. lactis* food isolates (Boguslawska *et al.*, 2009), and between streptococcal species within dental biofilms (Roberts *et al.*, 2001a). There is a growing concern that beneficial and commensal bacterial populations may contribute to the transfer of antibiotic resistance to pathogenic and opportunistic bacteria (Wozniak *et al.*, 2007). Dissemination of antibiotic resistance genes *via* the food chain to either the resident microbiota of the human gut or indeed pathogenic bacteria would have far reaching effects on both human and animal health and present a major financial cost (Panel, 2005). The need to characterize the vehicles of dissemination is therefore of great importance.

1.6 Summary of Thesis contents

Lactobacillus salivarius is a bacterial species which has potential probiotic applications in humans and animals. As described herein, genomic diversity is a significant feature of *L. salivarius* and contributes to the varying phenotypic characteristics observed in this species. In Chapter II an in-depth genotypic and phenotypic analysis of *L. salivarius* was carried out and the phylogenetic structure of this species was studied in detail. The ability of a strain to form biofilms and produce EPS has potential implications for microbe-host interactions. EPS production levels were determined for a group of *L. salivarius* strains and the relationship between an EPS-producing phenotype and the capacity of *L. salivarius* strains to form a biofilm was assessed.

Chapter III presents the genome sequence of the porcine *L. salivarius* isolate JCM1046, which includes the first sequence of a linear plasmid in a *Lactobacillus* species and the first description of a Tn*916*-like element (Tn6224) in a *L. salivarius* genome. The potential implications for probiotic strain selection are also considered in light of the discovery of Tn6224, which confers tetracycline resistance on its host.

1.6.1 Aims and Objectives

-To characterize the level of genomic diversity in *Lactobacillus salivarius* using comparative genomics.

-To determine the potential effects of genotypic variation on the phenotypic traits of strains from human, animal and environmental habitats

-To assess the potential role of niche adaptation in *L. salivarius* and to determine if strains from different habitats harbour niche-specific gene sets.

-Devise a suitable Multilocus Sequence Typing scheme using appropriate housekeeping genes in order to determine the phylogenetic structure of *L. salivarius*.

-To determine the genome sequence and gene content of the porcine *L. salivarius* isolate JCM1046

-To establish the number and topology of the extra-chromosomal replicons harboured by strain JCM1046.

1.7 References

Abbas, H., Hanan, T., Surakka, A., Apajalahti, J. & Saris, P. E. J. (2007). Identification of the most abundant *Lactobacillus* species in the crop of 1- and 5-week-old broiler chickens. *Appl Environ Microbiol* 73, 7867-7873.

Abe, F., Muto, M., Yaeshima, T., Iwatsuki, K., Aihara, H., Ohashi, Y. & Fujisawa, T. (2010). Safety evaluation of probiotic bifidobacteria by analysis of mucin degradation activity and translocation ability. *Anaerobe* 16, 131-136.

Achtman, M. (2008). Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annu Rev Microbiol* 62, 53-70.

Aittamaa, M., Somervuo, P., Pirhonen, M., Mattinen, L., Nissinen, R., Auvinen, P. & Valkonen, J. P. T. (2008). Distinguishing bacterial pathogens of potato using a genome-wide microarray approach. *Mol Plant Pathol* 9, 705-717.

Alemayehu, D., Ross, R. P., O'Sullivan, O., Coffey, A., Stanton, C., Fitzgerald, G. F. & McAuliffe, O. (2009). Genome of a virulent bacteriophage Lb338-1 that lyses the probiotic *Lactobacillus paracasei* cheese strain. *Gene* 448, 29-39.

Alkan, C., Sajjadian, S. & Eichler, E. E. (2011). Limitations of next-generation genome sequence assembly. *Nat Meth* 8, 61-65.

Allardet-Servent, A., Michaux-Charachon, S., Jumas-Bilak, E., Karayan, L. & Ramuz, M. (1993). Presence of one linear and one circular chromosome in the Agrobacterium tumefaciens C58 genome. *J Bacteriol* 175, 7869-7874.

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.

Amadou, C., Pascal, G., Mangenot, S. & other authors (2008). Genome sequence of the β-rhizobium *Cupriavidus taiwanensis* and comparative genomics of rhizobia. *Genome Res* 18, 1472-1483.

Ammor, M. S., Belén Flórez, A. & Mayo, B. (2007). Antibiotic resistance in non-enterococcal lactic acid bacteria and bifidobacteria. *Food Microbiol* 24, 559-570.

Andersson, D. I. & Levin, B. R. (1999). The biological cost of antibiotic resistance. *Curr Opin Microbiol* 2, 489-493.

Andersson, D. I. & Hughes, D. (2010). Antibiotic resistance and its cost: is it possible to reverse resistance? *Nat Rev Micro* 8, 260-271.

Ansorge, W. (2009). Next-generation DNA sequencing techniques. *New Biotechnology* 25, 195-203.

Arias, C. A., Panesso, D., Singh, K. V., Rice, L. B. & Murray, B. E. (2009). Co-Transfer of Antibiotic Resistance Genes and a hylEfm-Containing Virulence Plasmid in Enterococcus faecium. *Antimicrob Agents Chemother*, AAC.00242-00209.

Aury, J.-M., Cruaud, C., Barbe, V. & other authors (2008). High quality draft sequences for prokaryotic genomes using a mix of new sequencing technologies. *BMC Genomics* 9, 603.

Azcarate-Peril, M., Altermann, E., Goh, Y. & 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.

Aziz, R., Bartels, D., Best, A. & other authors (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9, 75.

Badel, S., Bernardi, T. & Michaud, P. (2011). New perspectives for lactobacilli exopolysaccharides. *Biotech Adv* 29, 54-66.

Baele, M., Devriese, L. A. & Haesebrouck, F. (2001). *Lactobacillus agilis* is an important component of the pigeon crop flora. *J Appl Microbiol* 91, 488-491.

Balciunas, E. M., Castillo Martinez, F. A., Todorov, S. D., Franco, B. D. G. d. M., Converti, A. & Oliveira, R. P. d. S. (2013). Novel biotechnological applications of bacteriocins: A review. *Food Control* 32, 134-142.

Banfalvi, Z., Sakanyan, V., Koncz, C., Kiss, A., Dusha, I. & Kondorosi, A. (1981). Location of nodulation and nitrogen fixation genes on a high molecular weight plasmid of *R meliloti Mol Gen Genet*, 318-325.

Bankhead, T. & Chaconas, G. (2007). The role of VlsE antigenic variation in the Lyme disease spirochete: persistence through a mechanism that differs from other pathogens. *Mol Microbiol*, 1547–1558.

Barbour, A. & Garon, C. (1987). Linear plasmids of the bacterium *Borrelia burgdorferi* have covalently closed ends. *Science* 237, 409-411.

Barbour, A. G. & Zöckert, W. R. (1997). Genome sequencing. New tricks of a tick-borne pathogen. *Nature*, 580–586.

Barbour, A. G. & Guo, B. P. (2010). Pathogenesis of relapsing fever. In *Borrelia: Molecular Biology, Host Interaction and Pathogenesis* pp. 333–357. Edited by D. S. Samuels & J. D. Radolf. Norwich, UK: Horizon Scientific Press.

Barloy-Hubler, F. & Jebbar, M. (2009). *Sinorhizobium meliloti*; megaplasmids and symbiosis in;*S. meliloti*. In *Microbial Megaplasmids*, pp. 91-118. Edited by E. Schwartz: Springer Berlin / Heidelberg.

Barton, B. M., Harding, G. P. & Zuccarelli, A. J. (1995). A general method for detecting and sizing large plasmids. *Analyt Biochem* 226, 235-240.

Bartosik, D., Baj, J., Bartosik, A. A. & Wlodarczyk, M. (2002). Characterization of the replicator region of megaplasmid pTAV3 of *Paracoccus versutus* and search for plasmid-encoded traits. *Microbiology* 148, 871-881.

Bashir, A., Klammer, A. A., Robins, W. P. & other authors (2012). A hybrid approach for the automated finishing of bacterial genomes. *Nat Biotech* 30, 701-707.

Bello, F. D., Walter, J., Hertel, C. & Hammes, W. P. (2001). In vitro study of prebiotic properties of levan-type exopolysaccharides from lactobacilli and non-digestible carbohydrates using denaturing gradient gel electrophoresis. *Syst App Micro* 24, 232-237.

Bentley, S. D. & Parkhill, J. (2004). Comparative genomic structure of prokaryotes. *Ann Rev Genet* 38, 771-791.

Berger, B., Pridmore, R., Barretto, C., Delmas-Julien, F., Schreiber, K., Arigoni, F. & Brüssow, H. (2007). Similarity and differences in the *Lactobacillus acidophilus* group identified by polyphasic analysis and comparative genomics. *J Bacteriol* 189, 1311-1321.

Bernardeau, M., Vernoux, J. P., Henri-Dubernet, S. & Guéguen, M. (2008). Safety assessment of dairy microorganisms: The *Lactobacillus* genus. *Int J Food Microbiol* 126, 278-285.

Bertram, J., Stratz, M. & Durre, P. (1991). Natural transfer of conjugative transposon Tn916 between gram-positive and gram-negative bacteria. *J Bacteriol* 173, 443-448.

Boekhorst, J., Siezen, R., Zwahlen, M. & other authors (2004). The complete genomes of *Lactobacillus plantarum* and *Lactobacillus johnsonii* reveal extensive differences in chromosome organization and gene content. *Microbiology* 150, 3601 - 3611.

Boels, I. C., Kranenburg, R. v., Hugenholtz, J., Kleerebezem, M. & de Vos, W. M. (2001). Sugar catabolism and its impact on the biosynthesis and engineering of exopolysaccharide production in lactic acid bacteria. *Int Dairy J* 11, 723-732.

Boguslawska, J., Zycka-Krzesinska, J., Wilcks, A. & Bardowski, J. (2009). Intra- and interspecies conjugal transfer of Tn916-like elements from *Lactococcus lactis in vitro* and *in vivo*. *Appl Environ Microbiol* 75, 6352-6360.

Broker, D., Arenskotter, M., Legatzki, A., Nies, D. H. & Steinbuchel, A. (2004). Characterization of the 101-kilobase-pair megaplasmid pKB1, isolated from the rubber-degrading bacterium *Gordonia westfalica* Kb1. *J Bacteriol* 186, 212-225.

Brook, I. & Frazier, E. H. (1993). Significant recovery of nonsporulating anaerobic rods from clinical specimens. *Clin Infect Dis* 16, 476-480.

Brown, S. E., Knudson, D. L. & Ishimaru, C. A. (2002). Linear Plasmid in the Genome of Clavibacter michiganensis subsp. sepedonicus. *J Bacteriol* 184, 2841-2844.

Buck, B. L., Altermann, E., Svingerud, T. & Klaenhammer, T. R. (2005). Functional analysis of putative adhesion factors in *Lactobacillus acidophilus* NCFM. *Appl Environ Microbiol* 71, 8344–8351.

Burgdorfer, W., Barbour, A., Hayes, S., Benach, J., Grunwaldt, E. & Davis, J. (1982). Lyme disease-a tick-borne spirochetosis? *Science* 216, 1317-1319.

Burrus, V., Pavlovic, G., Decaris, B. & Guédon, G. (2002). Conjugative transposons: the tip of the iceberg. *Molecular Microbiology* 46, 601-610.

Burrus, V. & Waldor, M. K. (2004). Shaping bacterial genomes with integrative and conjugative elements. *Res Microbiol* 155, 376-386.

Burton, J. P., Cadieux, P. A. & Reid, G. (2003). Improved understanding of the bacterial vaginal microbiota of women before and after probiotic instillation. *Appl Environ Microbiol* 69, 97-101.

Cahill, M. J., Köser, C. U., Ross, N. E. & Archer, J. A. C. (2010). Read length and repeat resolution: exploring prokaryote genomes using next-generation sequencing technologies. *PLoS ONE* 5, e11518.

Cai, H., Rodriguez, 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.

Cai, H., Thompson, R., Budinich, M. F., Broadbent, J. R. & Steele, J. L. (2009). Genome sequence and comparative genome analysis of *Lactobacillus casei*: insights into their niche-associated evolution. *Genome Biol Evol* 2009, 239-257.

Cai, Y., Pang, H., Kitahara, M. & Ohkuma, M. (2012). *Lactobacillus nasuensis* sp. nov., a lactic acid bacterium isolated from silage, and emended

description of the genus *Lactobacillus*. *Int J Syst Evol Microbiol* 62, 1140-1144.

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.

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

Canchaya, C., Claesson, M., Fitzgerald, G., van Sinderen, D. & O'Toole, P. (2006). Diversity of the genus *Lactobacillus* revealed by comparative genomics of five species. *Microbiology (Reading, England)* 152, 3185-3196.

Candela, T. & Fouet, A. (2005). *Bacillus anthracis CapD*, belonging to the gamma-glutamyltraspeptidase family, is required for the covalent anchoring of capsule to peptidoglycan. *Mol Microbiol*, 717-726.

Cannon, J. P., Lee, T. A., Bolanos, J. T. & Danziger, L. H. (2005). Pathogenic relevance of *Lactobacillus*: a retrospective review of over 200 cases. *Eur J Clin Microbiol* 24, 31-40.

Caparon, M. G. & Scott, J. R. (1989). Excision and insertion of the conjugative transposon Tn916 involves a novel recombination mechanism. *Cell* 59, 1027-1034.

Carlson, C. R., Grønstad, A. & Kolstø, A.-B. (1992). Physical maps of the genomes of three *Bacillus cereus* strains. *J Bacteriol*, 3750–3756.

Carr, J. G. & Davies, P. A. (1970). Homofermentative lactobacilli of ciders including *Lactobacillus mali* nov. spec. *J Appl Bacteriol* 33, 768-774.

Carter, B., Wu, G., Woodward, M. & Anjum, M. (2008). A process for analysis of microarray comparative genomics hybridisation studies for bacterial genomes. *BMC Genomics* 9, 53.

Carter, B. R. (2005). Statistical methodology for the analysis of microarray data. *PhD Thesis*.

Casey, P., Casey, G., Gardiner, G., Tangney, M., Stanton, C., Ross, R., Hill, C. & Fitzgerald, G. (2004). Isolation and characterization of anti-*Salmonella* lactic acid bacteria from the porcine gastrointestinal tract. *Lett Appl Microbiol*, 431-438.

Casjens, S., Murphy, M., DeLange, M., Sampson, L., Vugt, R. v. & Huang, W. M. (1997). Telomeres of the linear chromosomes of Lyme disease spirochaetes: nucleotide sequence and possible exchange with linear plasmid telomeres.
. *Mol Microbiol* 26.

Casjens, S. (1998). The diverse and dynamic structure of bacterial genomes. *Ann Rev Genet* 32, 339-377.

Casjens, S., Palmer, N., Van Vugt, R. & other authors (2000). A bacterial genome in flux: the twelve linear and nine circular extrachromosomal DNAs in an infectious isolate of the Lyme disease spirochete *Borrelia burgdorferi*. *Mol Microbiol* 35, 490-516.

Casjens, S. R., Gilcrease, E. B., Huang, W. M., Bunny, K. L., Pedulla, M. L., Ford, M. E., Houtz, J. M., Hatfull, G. F. & Hendrix, R. W. (2004). The pKO2 linear plasmid prophage of *Klebsiella oxytoca*. *J Bacteriol* 186, 1818-1832.

Casjens, S. R., Eggers, C. H. & Schwartz, I. (2010). *Borrelia* genomics: chromosome, plasmids, bacteriophages and genetic variation. In *Borrelia: Molecular Biology, Host Interaction and Pathogenesis*, pp. 27–53. Edited by D. S. Samuels & J. D. Radolf. Norwich, UK: Horizon Scientific Press.

Celli, J. & Trieu-Cuot, P. (1998). Circularization of Tn916 is required for expression of the transposon-encoded transfer functions: characterization of long tetracycline-inducible transcripts reading through the attachment site. *Mol Microbiol* 28, 103-117.

Chaconas, G. (2005). Hairpin telomeres and genome plasticity in *Borrelia*: all mixed up in the end. *Mol Microbiol* 58, 625-635.

Chaconas, G. & Kobryn, K. (2010). Structure, function, and evolution of linear replicons in *Borrelia*. *Ann Rev Microbiol* 64, 185-202.

Chaisson, M., Brinza, D. & Pevzner, P. (2009). *De novo* fragment assembly with short mate-paired reads: Does the read length matter? *Genome Res* 19, 336-346.

Chamba & Irlinger. (2004).Secondary and adjunct cultures. In *Cheese, Chemistry, Physics and Microbiology, General Aspects*. Edited by P. F. Fox, P. L. H. McSweeney, T. M. Cogan & T. P. Guinee. London: Elsevier Academic.

Chang, P.-C. & Cohen, S. N. (1994). Birdirectional replication from an internal origin in a linear *Streptomyces* plasmid. *Science*, 952-954.

Chang, P.-C. & Cohen, S. N. (1996). *Streptomyces* linear plasmids that contain a phage-like, centrally located replication origin *Mol Microbiol*, 789-800.

Chang, Y.-j., Land, M., Hauser, L. & other authors (2011). Non-contiguous finished genome sequence and contextual data of the filamentous soil bacterium *Ktedonobacter racemifer* type strain (SOSP1-21(T)). *SIGS* 5, 97-111.

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 Syst Evol Microbiol* 58, 2555-2559.

Chapot-Chartier, M.-P. & Kulakauskas, S. (2014). Cell wall structure and function in lactic acid bacteria. *Microb Cell Fact* 13, S9.

Charteris, W. P., Kelly, P. M., Morelli, L. & Collins, J. K. (1997). Selective detection, enumeration and identification of potentially probiotic *Lactobacillus* and *Bifidobacterium* species in mixed bacterial populations *Int J Food Microbiol*, 1-27.

Chater, K. & Kinashi, H. (2007). *Streptomyces*; linear plasmids: their discovery, functions, interactions with other replicons, and evolutionary significance. In *Microbial Linear Plasmids*, pp. 1-31. Edited by F. Meinhardt & R. Klassen: Springer Berlin / Heidelberg.

Chater, K. F. (2006). *Streptomyces* inside-out: a new perspective on the bacteria that provide us with antibiotics. *Philos Trans R Soc Lond B Biol Sci* 361, 761-768.

Chen, C. (2007). *Streptomyces*; linear plasmids: replication and telomeres. In *Microbial Linear Plasmids*, pp. 33-61. Edited by F. Meinhardt & R. Klassen: Springer Berlin / Heidelberg.

Chen, C. W., Yu, T.-W., Lin, Y.-S., Kieser, H. M. & Hopwood, D. A. (1993). The conjugative plasmid SLP2 of *Streptomyces lividans* is a 50-kb linear molecule. *Mol Microbiol*, 925-932.

Chen, C. W. (1996). Complications and implications of linear bacterial chromosomes. *Trends Genet* 12, 192-196.

Chen, Y.-s., Miyashita, M., Suzuki, K.-i., Sato, H., Hsu, J.-s. & Yanagida, F. (2009). *Lactobacillus pobuzihi* sp. nov., isolated from pobuzih (fermented cummingcordia) in Taiwan. *Int J Syst Evol Microbiol*.

Chin, C.-S., Sorenson, J., Harris, J. B. & other authors (2011). The origin of the Haitian cholera outbreak strain. *N Engl J Med* 364, 33-42.

Cho, Y.-J., Choi, J. K., Kim, J.-H., Lim, Y.-S., Ham, J.-S., Kang, D.-K., Chun, J., Paik, H.-D. & Kim, G.-B. (2011). Genome sequence of *Lactobacillus salivarius* GJ-24, a probiotic strain isolated from healthy adult intestine. *J Bacteriol* 193, 5021-5022.

Chu, L., Scharf, E. & Kondo, T. (2001). GeneSpring™: Tools for Analyzing Microarray Expression Data. *Genome Inform* 12.

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

Claesson, M., van Sinderen, D. & O'Toole, P. (2008). *Lactobacillus* phylogenomics-towards a reclassification of the genus. *Int J Syst Evol Microbiol* 58, 2945-2954.

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.

Clewell, D. B., Flannagan, S. E. & Jaworski, D. D. (1995). Unconstrained bacterial promiscuity: the Tn916-Tn1545 family of conjugative transposons. *Trends Microbiol* 3, 229-236.

Codling, C., O'Mahony, L., Shanahan, F., Quigley, E. M. M. & Marchesi, J. R. (2010). A molecular analysis of fecal and mucosal bacterial communities in irritable bowel syndrome. *Dig Dis Sci* 55, 392-397.

Coenye, T. & Vandamme, P. (2003). Simple sequence repeats and compositional bias in the bipartite *Ralstonia solanacearum* GMI1000 genome. *BMC Genomics* 4, 10.

Coleman, N. V. & Spain, J. C. (2003). Distribution of the coenzyme M pathway of epoxide metabolism among ethene- and vinyl chloride-degrading mycobacterium strains. *Appl Environ Microbiol* 69, 6041-6046.

Corr, S. C., Li, Y., Riedel, C. U., O'Toole, P. W., Hill, C. & Gahan, C. G. M. (2007). From the Cover: Bacteriocin production as a mechanism for the antiinfective activity of *Lactobacillus salivarius* UCC118. *Proc Natl Acad Sci U S A* 104, 7617-7621.

Courvalin, P. & Carlier, C. (1986). Transposable multiple antibiotic resistance in *Streptococcus pneumoniae*. *Mol Gen Genet MGG* 205, 291-297.

Courvalin, P. (1994). Transfer of antibiotic resistance genes between gram-positive and gram-negative bacteria. *Antimicrob Agents Chemother* 38, 1447-1451.

Cremonesi, P., Vanoni, L., Morandi, S., Silvetti, T., Castiglioni, B. & Brasca, M. (2011). Development of a pentaplex PCR assay for the simultaneous detection of *Streptococcus thermophilus*, *Lactobacillus delbrueckii* subsp. *bulgaricus*, *L. delbrueckii* subsp. *lactis*, *L. helveticus*, *L. fermentum* in whey starter for Grana Padano cheese. *Int J Food Microbiol* 146, 207-211.

Crespi, M., Messens, E., Caplan, A. B., Montagu, M. v. & Desomer, J. (1992). Fasciation induction by the phytopathogen *Rhodococcus fascians* depends upon a linear plasmid encoding a cytokinin synthase gene. *EMBO J*, 795-804.

Crespi, M., Vereecke, D., Temmerman, W., Van Montagu, M. & Desomer, J. (1994). The fas operon of *Rhodococcus fascians* encodes new genes required for efficient fasciation of host plants. *J Bacteriol* 176, 2492-2501.

Crielaard, W., Zaura, E., Schuller, A., Huse, S., Montijn, R. & Keijser, B. (2011). Exploring the oral microbiota of children at various developmental stages of their dentition in the relation to their oral health. *BMC Medical Genomics* 4, 22.

Dabrock, B., Kesseler, M., Averhoff, B. & Gottschalk, G. (1994). Identification and characterization of a transmissible linear plasmid from *Rhodococcus erythropolis* BD2 that encodes isopropylbenzene and trichloroethene catabolism. *Appl Environ Microbiol* 60, 853-860.

Dai, Q., Restrepo, B. I., Porcella, S. F., Raffel, S. J., Schwan, T. G. & Barbour, A. G. (2006). Antigenic variation by *Borrelia hermsii* occurs through recombination between extragenic repetitive elements on linear plasmids. *Mol Microbiol*, 1329–1343.

Dal Bello, F. & Hertel, C. (2006). Oral cavity as natural reservoir for intestinal lactobacilli. *Systematic and Applied Microbiology* 29, 69-76.

Danko, A. S., Luo, M., Bagwell, C. E., Brigmon, R. L. & Freedman, D. L. (2004). Involvement of linear plasmids in aerobic biodegradation of vinyl chloride. *Appl Environ Microbiol* 70, 6092-6097.

DasSarma, S., Capes, M. & DasSarma, P. (2009). Haloarchaeal Megaplasmids. In *Microbial Megaplasmids*, pp. 3-30. Edited by E. Schwartz: Springer Berlin / Heidelberg.

Davey, R., Savva, G., Dicks, J. & Roberts, I. N. (2007). MPP: a microarray-to-phylogeny pipeline for analysis of gene and marker content datasets. *Bioinformatics* 23, 1023-1025.

De Bruyne, K., Camu, N., De Vuyst, L. & Vandamme, P. (2009). *Lactobacillus fabifermentans* sp. nov. and *Lactobacillus cacaonum* sp. nov., isolated from Ghanaian cocoa fermentations. *Int J Syst Evol Microbiol* 59, 7-12.

de las Rivas, B., Marcobal, A. & Munoz, R. (2004). Allelic diversity and population structure in *Oenococcus oeni* as determined from sequence analysis of housekeeping genes. *Appl Environ Microbiol* 70, 7210-7219.

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 Vuyst, L. & Degeest, B. (1999). Heteropolysaccharides from lactic acid bacteria. *FEMS Microbiology Reviews* 23, 153-177.

Dec, M., Urban-Chmiel, R., Gnat, S., Puchalski, A. & Wernicki, A. (2014). Identification of *Lactobacillus* strains of goose origin using MALDI-TOF mass spectrometry and 16S–23S rDNA intergenic spacer PCR analysis. *Res Microbiol* 165, 190-201.

del Solar, G., Giraldo, R., Ruiz-Echevarría, M. J., Espinosa, M. & Dí az-Orejas, R. (1998). Replication and control of circular bacterial plasmids. *Microbiol Mol Biol Rev* 62, 434-464.

Denef, V. J., Klappenbach, J. A., Patrauchan, M. A., Florizone, C., Rodrigues, J. L. M., Tsoi, T. V., Verstraete, W., Eltis, L. D. & Tiedje, J. M. (2006). Genetic and genomic insights into the role of benzoate-catabolic pathway redundancy in *Burkholderia xenovorans* LB400. *Appl Environ Microbiol* 72, 585-595.

Dennis, J. (2005). The evolution of IncP catabolic plasmids. *Curr Opin Biotechnol*, 291-298.

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

Dent, V. E. & Williams, R. A. D. (1982). *Lactobacillus animalis* sp. nov., a new species of *Lactobacillus* from the alimentary canal of animals. *Zentralblatt für Bakteriologie Mikrobiologie und Hygiene: I Abt Originale C: Allgemeine, angewandte und ökologische Mikrobiologie* 3, 377-386.

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.

Dib, J. R., Wagenknecht, M., Hill, R. T., Farías, M. E. & Meinhardt, F. (2010a). First report of linear megaplasmids in the genus *Micrococcus*. *Plasmid* 63, 40-45.

Dib, J. R., Wagenknecht, M., Hill, R. T., Farías, M. E. & Meinhardt, F. (2010b). Novel linear megaplasmid from *Brevibacterium* sp. isolated from extreme environment. *J Basic Microbiol* 50, 280-284.

Didelot, X., Bowden, R., Wilson, D. J., Peto, T. E. A. & Crook, D. W. (2012). Transforming clinical microbiology with bacterial genome sequencing. *Nat Rev Genet* 13, 601-612.

Dingle, K. E., Colles, F. M., Wareing, D. R. A. & other authors (2001). Multilocus sequence typing system for *Campylobacter jejuni*. *J Clin Microbiol* 39, 14-23.

Dobson, A., Cotter, P. D., Ross, R. P. & Hill, C. (2012). Bacteriocin production: a probiotic trait? *Appl Environ Microbiol* 78, 1-6.

Donohue, D. C. & Salminen., S. (1996). Safety of probiotic bacteria. *Asia Pac J Clin Nutr*, 25–28.

Dorman, C. (2009). The virulence plasmids of ;*Shigella flexneri*. In *Microbial Megaplasmids*, pp. 151-170. Edited by E. Schwartz: Springer Berlin / Heidelberg.

dos Santos, L. F., Irino, K., Vaz, T. M. I. & Guth, B. E. C. (2010). Set of virulence genes and genetic relatedness of O113 : H21 Escherichia coli strains isolated from the animal reservoir and human infections in Brazil. *J Med Microbiol* 59, 634-640.

Dotterud, C. K., Storrø, O., Johnsen, R. & Øien, T. (2010). Probiotics in pregnant women to prevent allergic disease: a randomized, double-blind trial. *Brit J Dermatol* 163, 616-623.

Doucet-Populaire, F., Trieu-Cuot, P., Dosbaa, I., Andremont, A. & Courvalin, P. (1991). Inducible transfer of conjugative transposon Tn1545 from *Enterococcus faecalis* to *Listeria monocytogenes* in the digestive tracts of gnotobiotic mice. *Antimicrob Agents Chemother* 35, 185-187.

Douillard, F. P., Ribbera, A., Kant, R. & other authors (2013). Comparative genomic and functional analysis of 100 *Lactobacillus rhamnosus* strains and their comparison with strain GG. *PLoS Genet* 9, e1003683.

Downie, J. A. & Young, J. P. W. (2001). Genome sequencing: The ABC of symbiosis. *Nature* 412, 597-598.

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. *Antonie van Leeuwenhoek* 76, 279-292.

Dussurget, O., Cabanes, D., Dehoux, P., Lecuit, M., Buchrieser, C., Glaser, P. & Cossart, P. (2002). *Listeria monocytogenes* bile salt hydrolase is a PrfA regulated virulence factor involved in the intestinal and hepatic phases of listeriosis. *Mol Microbiol* 45, 1095–1106.

Dworkin, M. S., Schwan, T. G. & Anderson, D. E. (2002). Tick-borne relapsing fever in North America. *Med Clin North Am* 86, 417-433.

Dykhuizen, D. E., Polin, D. S., Dunn, J. J., Wilske, B., Preac-Mursic, V., Dattwyler, R. J. & Luft, B. J. (1993). *Borrelia burgdorferi* is clonal: implications for taxonomy and vaccine development. *Proc Natl Acad Sci* 90, 10163-10167.

Edwards, C. G., Collins, M. D., Lawson, P. A. & Rodriguez, A. V. (2000). *Lactobacillus nageli* sp. nov., an organism isolated from a partially fermented wine. *Int J Syst Evol Microbiol* 50, 699-702.

Edwards, D. & Holt, K. (2013). Beginner's guide to comparative bacterial genome analysis using next-generation sequence data. *Microb Inform Exp* 3, 2.

Egan, E. S., Fogel, M. A. & Waldor, M. K. (2005). MicroReview: Divided genomes: negotiating the cell cycle in prokaryotes with multiple chromosomes. *Molecular Microbiology* 56, 1129-1138.

Endo, A. & Okada, S. (2005). *Lactobacillus satsumensis* sp. nov., isolated from mashes of shochu, a traditional Japanese distilled spirit made from fermented rice and other starchy materials. *Int J Syst Evol Microbiol* 55, 83-85.

English, A. C., Richards, S., Han, Y. & other authors (2012). Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing yechnology. *PLoS ONE* 7, e47768.

Enright, M. C. & Spratt, B. G. (1999). Multilocus sequence typing. *Trends Microbiol* 7, 482–487.

Enright, M. C., Spratt, B. G., Kalia, A., Cross, J. H. & Bessen, D. E. (2001). Multilocus sequence typing of *Streptococcus pyogenes* and the relationships between emm type and clone. *Infect Immun* 69, 2416-2427.

Fang, F., Flynn, S., Li, Y., Claesson, M. J., van Pijkeren, J.-P., Collins, J. K., van Sinderen, D. & O'Toole, P. W. (2008). Characterization of endogenous plasmids from *Lactobacillus salivarius* UCC118. *Appl Environ Microbiol* 74, 3216-3228.

Fang, F., Li, Y., Bumann, M., Raftis, E. J., Casey, P. G., Cooney, J. C., Walsh, M. A. & O'Toole, P. W. (2009). Allelic variation of bile salt hydrolase genes in *Lactobacillus salivarius* does not determine bile resistance levels. *J Bact* 191, 5743-5757.

Feil, E. J., Li, B. C., Aanensen, D. M., Hanage, W. P. & Spratt, B. G. (2004). eBURST: Inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *J Bacteriol* 186, 1518-1530.

Felis, G. & Dellaglio, F. (2007). Taxonomy of lactobacilli and bifidobacteria. *Curr Issues Mol Biol* 8, 44-61.

Feng, P. C. H., Councell, T., Keys, C. & Monday, S. R. (2011). Virulence characterization of shiga-toxigenic *Escherichia coli* isolates from wholesale produce. *Appl Environ Microbiol* 77, 343-345.

Ferdows, M. S., Serwer, P., Griess, G. A., Norris, S. J. & Barbour, A. G. (1996). Conversion of a linear to a circular plasmid in the relapsing fever agent *Borrelia hermsii*. *J Bacteriol*, 793–800.

Fernandez, E. M., Valenti, V., Rockel, C., Hermann, C., Pot, B., Boneca, I. G. & Grangette, C. (2011). Anti-inflammatory capacity of selected lactobacilli in experimental colitis is driven by NOD2-mediated recognition of a specific peptidoglycan-derived muropeptide. *Gut* 60, 1050-1059.

Fetzner, S., Kolkenbrock, S. & Parschat, K. (2007). Catabolic linear plasmids. In *Microbial Linear Plasmids*, pp. 63-98. Edited by F. Meinhardt & R. Klassen: Springer Berlin / Heidelberg.

Flannagan, S. E., Zitzow, L. A., Su, Y. A. & Clewell, D. B. (1994). Nucleotide sequence of the 18-kb conjugative transposon Tn916 from *Enterococcus faecalis*. *Plasmid* 32, 350-354.

Flynn, S. (2001). Molecular characterisation of bacteriocin producing genes and plasmid encoded functions of the probiotic strain *Lactobacillus salivarius* subsp. *salivarius* UCC118. Ph.D Thesis. *University College Cork, Cork, Ireland*.

Flynn, S., van Sinderen, D., Thornton, G. M., Holo, H., Nes, I. F. & Collins, J. K. (2002). Characterization of the genetic locus responsible for the production of ABP-118, a novel bacteriocin produced by the probiotic bacterium *Lactobacillus salivarius* subsp. *salivarius* UCC118. *Microbiology* 148, 973-984.

Fodor, S., Read, J., Pirrung, M., Stryer, L., Lu, A. & Solas, D. (1991). Light-directed, spatially addressable parallel chemical synthesis. *Science* 251, 767-773.

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

Forssten, S. D., Sindelar, C. W. & Ouwehand, A. C. (2011). Probiotics from an industrial perspective. *Anaerobe* 17, 410-413.

Fouet, A. & Moya, M. (2009). Virulence megaplasmids in *Bacillus anthracis* and their relatives in the *Bacillus cereus* group. In *Microbial Megaplasmids*, pp. 187-206. Edited by E. Schwartz: Springer Berlin / Heidelberg.

França, L., Carrilho, E. & Kist, T. (2002). A review of DNA sequencing techniques. *Q Rev Biophys* 35, 169-200.

Franke, A. E. & Clewell, D. B. (1981). Evidence for a chromosome-borne resistance transposon (Tn916) in *Streptococcus faecalis* that is capable of

"conjugal" transfer in the absence of a conjugative plasmid. *J Bacteriol* 145, 494-502.

Fraser, C., Casjens, S., Huang, W. & other authors (1997). Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature* 390, 580 - 586.

Fratamico, P. M., Yan, X., Caprioli, A., Esposito, G., Needleman, D. S., Pepe, T., Tozzoli, R., Cortesi, M. L. & Morabito, S. (2011). The complete DNA sequence and analysis of the virulence plasmid and of five additional plasmids carried by Shiga toxin-producing *Escherichia coli* O26:H11 strain H30. *Int J Med Microbiol* 301, 192-203.

Freitas, A. R., Tedim, A. P., Novais, C. & other authors (2010). Global spread of the hylEfm colonization-virulence gene in megaplasmids of the *Enterococcus faecium* CC17 polyclonal subcluster. *Antimicrob Agents Chemother* 54, 2660-2665.

Freitas, F., Alves, V. D. & Reis, M. A. M. (2011). Advances in bacterial exopolysaccharides: from production to biotechnological applications. *Trends Biotechnol* 29, 388-398.

Frost, L. S., Leplae, R., Summers, A. O. & Toussaint, A. (2005). Mobile genetic elements: the agents of open source evolution. *Nat Rev Micro* 3, 722-732.

Fujisawa, T., Shirasaka, S., Watabe, J. & Mitsuoka, T. (1984). *Lactobacillus aviarius* sp. nov.: A new species isolated from the intestine of chickens. *Syst Appl Microbiol* 5, 414-420.

Fukiya, S., Mizoguchi, H., Tobe, T. & Mori, H. (2004). Extensive genomic diversity in pathogenic *Escherichia coli* and *Shigella* strains revealed by comparative genomic hybridization microarray. *J Bacteriol* 186, 3911 - 3921.

Gadagkar, S. R., Rosenberg, M. S. & Kumar, S. (2005). Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. *J Exp Zool B Mol Dev Evol* 304B, 64-74.

Garcia-Migura, L., Liebana, E. & Jensen, L. B. (2007). Transposon characterization of vancomycin-resistant *Enterococcus faecium* (VREF) and dissemination of resistance associated with transferable plasmids. *J Antimicrob Chemother* 60, 263-268.

Gaspar, P., Carvalho, A. L., Vinga, S., Santos, H. & Neves, A. R. (2013). From physiology to systems metabolic engineering for the production of biochemicals by lactic acid bacteria. *Biotechnol Adv* 31, 764-788.

Gelvin, S. B. (2003). Agrobacterium-mediated plant transformation: the biology behind the "gene-jockeying" tool. *Microbiol Mol Biol Rev* 67, 16-37.

Genin, S. & Boucher, C. (2002). *Ralstonia solanacearum*: secrets of a major pathogen unveiled by analysis of its genome. *Mol Plant Pathology* 3, 111-118.

Genin, S. & Boucher, C. (2004). Lessons learned from the genome analysis of Ralstonia solanacearum. *Annu Rev Phytopathol* 42, 107 - 134.

Gevers, D., Danielsen, M., Huys, G. & Swings, J. (2003). Molecular characterization of *tet*(M) genes in *Lactobacillus* isolates from different types of fermented dry sausage. *Appl Environ Microbiol* 69, 1270-1275.

Ghosh, W., George, A., Agarwal, A., Raj, P., Alam, M., Pyne, P. & Das Gupta, S. K. (2011). Whole-genome shotgun sequencing of the sulfur-oxidizing chemoautotroph *Tetrathiobacter kashmirensis*. *J Bacteriol* 193, 5553-5554.

Gibbons, J. G., Janson, E. M., Hittinger, C. T., Johnston, M., Abbot, P. & Rokas, A. (2009). Benchmarking next-generation transcriptome sequencing for functional and evolutionary genomics. *Mol Biol Evol* 26, 2731-2744.

Giraffa, G., Chanishvili, N. & Widyastuti, Y. (2010). Importance of lactobacilli in food and feed biotechnology. *Res Microbiol* 161, 480-487.

Godiska, R., Mead, D., Dhodda, V., Wu, C., Hochstein, R., Karsi, A., Usdin, K., Entezam, A. & Ravin, N. (2009). Linear plasmid vector for cloning of repetitive or unstable sequences in *Escherichia coli*. *Nucleic Acids Res.*

Goethals, K., Vereecke, D., Jaziri, M., Van Montagu, M. & Holsters, M. (2001). Leafy gall formation by *Rhodococcus Fascians Annu Rev Plant Physiol* 39, 27-52.

Goffin, P., M. Deghorain, J. L. Mainardi, I. Tytgat, M. C. Champomier-Verges, Kleerebezem, M. & Hols, P. (2005). Lactate racemization as a rescue pathway for supplying D-lactate to the cell wall biosynthesis machinery in *Lactobacillus plantarum*. *J Bacteriol* 187, 6750-6761.

Goldberg, S., Johnson, J., Busam, D. & other authors (2006). A Sanger/pyrosequencing hybrid approach for the generation of high-quality draft assemblies of marine microbial genomes. *Proc Natl Acad Sci USA* 103, 11240-11245.

Gonzalez, V., Acosta, J. L., Santamaria, R. I. & other authors (2010). Conserved symbiotic plasmid DNA sequences in the multireplicon pangenomic structure of *Rhizobium etli*. *Appl Environ Microbiol* 76, 1604-1614.

Goodner, B., Hinkle, G., Gattung, S. & other authors (2001). Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens C58. *Science* 294, 2323-2328.

Gravius, B., Glocker, D., Pigac, J., Pandza, K., Hranueli, D. & Cullum, J. (1994). The 387-kb linear plasmid pPZ101 of *Streptomyces rimosus* and its interactions with the chromosome. *Microbiology*, 2271-2277.

Guarner, F. & Schaafsma, G. J. (1998). Probiotics. *Int J Food Microbiol* 39, 237-238.

Guidot, A., Coupat, B., Fall, S., Prior, P. & Bertolla, F. (2009). Horizontal gene transfer between *Ralstonia solanacearum* strains detected by comparative genomic hybridization on microarrays. *Isme J* 3, 549-562.

Hale, T. L. (1991). Genetic basis of virulence in *Shigella* species. *Microbiol Mol Biol Rev* 55, 206-224.

Ham, J.-S., Kim, H.-W., Seol, K.-H. & other authors (2011). Genome sequence of *Lactobacillus salivarius* NIAS840, isolated from chicken intestine. *J Bacteriol* 193, 5551-5552.

Hammes, W. & Hertel, C. (2006). The Genera *Lactobacillus* and *Carnobacterium* In *The Prokaryotes*, pp. 320-403. Edited by M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer & E. Stackebrandt: Springer New York.

Hammes, W. P. & Vogel, R. F. (1995). The genus *Lactobacillus* In *The lactic acid bacteria, vol 2* pp. 19-54. Edited by B. J. B. Wood & W. H. Holzapfel. London: Blackie Academic and Professional.

Harrison, P. W., Lower, R. P. J., Kim, N. K. D. & Young, J. P. W. (2010). Introducing the bacterial `chromid': not a chromosome, not a plasmid. *Trends Microbiol* 18, 141-148.

Hayakawa, T., Tanaka, T., Sakaguchi, K., Otake, N. & Yonehara, H. (1979). A linear plasmid-like DNA in *Streptomyces* species producing lankacidin group antibiotics *J Gen Appl Microbiol* 255-260.

Heidelberg, J., Eisen, J., Nelson, W. & other authors (2000). DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* 406, 477 - 483.

Heilig, H. G. H. J., Zoetendal, E. G., Vaughan, E. E., Marteau, P., Akkermans, A. D. L. & 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.

Hemme D, Raibaud P, Ducluzeau R, Galpin JV, Sicard P & J, v. H. (1980). *Lactobacillus murinus* n. sp., une nouvelle espece de la flore dominante autochtone du tube digestif du rat et de la souris. *Ann Microbiol* 131, 297–308.

Hendrickx, A. P. A., Schapendonk, C. M. E., van Luit-Asbroek, M., Bonten, M. J. M., van Schaik, W. & Willems, R. J. L. (2010). Differential PilA pilus assembly by a hospital-acquired and a community-derived *Enterococcus faecium* isolate. *Microbiology* 156, 2649-2659.

Herbel, S., Vahjen, W., Wieler, L. & Guenther, S. (2013). Timely approaches to identify probiotic species of the genus *Lactobacillus*. *Gut Pathog* 5, 27.

Hertwig, S., Klein, I., Lurz, R., Lanka, E. & Appel, B. (2003). PY54, a linear plasmid prophage of *Yersinia enterocolitica* with covalently closed ends. *Mol Microbiol* 48, 989-1003.

Hertwig, S. (2007). Linear plasmids and prophages in gram-negative bacteria. In *Microbial Linear Plasmids*, pp. 141-162. Edited by F. Meinhardt & R. Klassen: Springer Berlin / Heidelberg.

Hidalgo-Cantabrana, C., López, P., Gueimonde, M., de los Reyes-Gavilán, C., Suárez, A., Margolles, A. & Ruas-Madiedo, P. (2012). Immune modulation capability of exopolysaccharides synthesised by lactic acid bacteria and bifidobacteria. *Probiotics Antimicrob Proteins* 4, 227-237.

Hinerfeld, D. & Churchward, G. (2001). Xis protein of the conjugative transposon Tn916 plays dual opposing roles in transposon excision. *Mol Microbiol* 41, 1459-1467.

Hinnebusch, J. & Barbour, A. G. (1991). Linear plasmids of *Borrelia burgdorferi* have a telomeric structure and sequence similar to those of a eukaryotic virus. *J Bacteriol* 173, 7233-7239.

Hinnebusch, J. & Tilly, K. (1993). Linear plasmids and chromosomes in bacteria. *Mol Microbiol* 10, 917-922.

Hiratsu, K., Mochizuki, S. & Kinashi, H. (2000). Cloning and analysis of the replication origin and the telomeres of the large linear plasmid pSLA2-L in *Streptomyces rochei Mol Gen Genet*, 1015-1021.

Holmes, E. C., Urwin, R. & Maiden, M. C. (1999). The influence of recombination on the population structure and evolution of the human pathogen *Neisseria meningitidis*. *Mol Biol and Evol* 16, 741-749.

Homayouni, A., Bastani, P., Ziyadi, S., Mohammad-Alizadeh-Charandabi, S., Ghalibaf, M., Mortazavian, A. M. & Mehrabany, E. V. (2014). Effects of probiotics on the recurrence of bacterial vaginosis: a review. *J Low Genit Tract Dis* 18, 79-86 10.1097/LGT.1090b1013e31829156ec.

Hopwood, D. A. (2006). Soil to genomics: the *Streptomyces* chromosome. *Annu Rev Genet* 40, 1-23.

Horner, D., Pavesi, G., Castrignanò, T., De Meo, P., Liuni, S., Sammeth, M., Picardi, E. & Pesole, G. (2010). Bioinformatics approaches for genomics and

post genomics applications of next-generation sequencing. *Brief Bioinform* 11, 181-197.

Howard, S., Strong, P. & Wren, B. (2009). The three bears and virulence-associated plasmids in the genus *Yersinia*. In *Microbial Megaplasmids*, pp. 171-186. Edited by E. Schwartz: Springer Berlin / Heidelberg.

Huang, C.-H., Lin, Y.-S., Yang, Y.-L., Huang, S.-W. & Chen, C. W. (1998). The telomeres of *Streptomyces* chromosomes contain conserved palindromic sequences with potential to form complex secondary structures. *Mol Microbiol* 28.

Huson, D. & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 23, 254-267.

Hyink, O., Wescombe, P. A., Upton, M., Ragland, N., Burton, J. P. & Tagg, J. R. (2007). Salivaricin A2 and the novel lantibiotic salivaricin B are encoded at adjacent loci on a 190-Kilobase transmissible megaplasmid in the oral probiotic strain *Streptococcus salivarius* K12. *Appl Environ Microbiol* 73, 1107-1113.

Ihssen, J., Grasselli, E., Bassin, C., François, P., Piffaretti, J.-C., Köster, W., Schrenzel, J. & Egli, T. (2007). Comparative genomic hybridization and physiological characterization of environmental isolates indicate that significant (eco-)physiological properties are highly conserved in the species *Escherichia coli*. *Microbiology* 153, 2052-2066.

Ike, Y., Flannagan, S. E. & Clewell, D. B. (1992). Hyperhemolytic phenomena associated with insertions of Tn916 into the hemolysin determinant of *Enterococcus faecalis* plasmid pAD1. *J Bacteriol* 174, 1801-1809.

Irisawa, T. & Okada, S. (2009). *Lactobacillus sucicola* sp. nov., a motile lactic acid bacterium isolated from oak tree (Quercus sp.) sap. *Int J Syst Evol Microbiol* 59, 2662-2665.

Iyer, R., Kalu, O., Purser, J., Norris, S., Stevenson, B. & Schwartz, I. (2003). Linear and circular plasmid content in *Borrelia burgdorferi* clinical isolates. *Infect Immun* 71, 3699-3706.

Janssen, P. J., Van Houdt, R., Moors, H. & other authors (2010). The complete genome sequence of *Cupriavidus metallidurans* Strain CH34, a master survivalist in harsh and anthropogenic environments. *PLoS ONE* 5, e10433.

Jaworski, D. & Clewell, D. (1995). A functional origin of transfer (*oriT*) on the conjugative transposon Tn916. *J Bacteriol* 177, 6644-6651.

Jessri, M. & Farah, C. S. (2014). Next generation sequencing and its application in deciphering head and neck cancer. *Oral Oncology* 50, 247-253.

Jimenez, E., Martin, R., Maldonado, A., Martin, V., Gomez de Segura, A., Fernandez, L. & Rodriguez, J. M. (2010). Complete genome sequence of *Lactobacillus salivarius* CECT 5713, a probiotic strain isolated from human milk and infant feces. *J Bacteriol* 192, 5266-5267.

Jin, Q., Yuan, Z., Xu, J. & other authors (2002). Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity through comparison with genomes of *Escherichia coli* K12 and O157. *Nucleic Acids Res* 30, 4432-4441.

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.

Jolly, L. & Stingele, F. (2001). Molecular organization and functionality of exopolysaccharide gene clusters in lactic acid bacteria. *Int Dairy J* 11, 733-745.

Jolly, L., Vincent, S., Duboc, P. & Neeser, J.-R. (2002). Exploiting expolysaccharides from lactic acid bacteria. *Antonie van Leeuwenhoek* 82, 367-374.

Jouko Sillanpää, Sreedhar R. Nallapareddy, Kavindra V. Singh, Vittal P. Prakash, Timothy Fothergill, Hung Ton-That & Murray, B. E. (2010). Characterization of the ebpfm pilus-encoding operon of Enterococcus faecium and its role in biofilm formation and virulence in a murine model of urinary tract infection. 1, 236 - 246.

Juhas, M., Van Der Meer, J. R., Gaillard, M., Harding, R. M., Hood, D. W. & Crook, D. W. (2009). Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol Rev* 33, 376-393.

Kalkus, J., Reh, M. & Schlegel, H. G. (1990). Hydrogen autotrophy of *Nocardia opaca* strains is encoded by linear megaplasmids. *J Gen Microbiol* 136, 1145-1151.

Kalkus, J., Dorrie, C., Fischer, D., Reh, M. & Schlegel, H. G. (1993). The giant linear plasmid pHG207 from *Rhodococcus* sp. encoding hydrogen autotrophy: characterization of the plasmid and its termini. *J Gen Microbiol* 139, 2055-2065.

Kallioniemi, A., Kallioniemi, O., Sudar, D., Rutovitz, D., Gray, J., Waldman, F. & Pinkel, D. (1992). Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* 258, 818-821.

Kandler, O. (1983). Carbohydrate metabolism in lactic acid bacteria. *Antonie van Leeuwenhoek*, 209–224.

Kandler, O. & Weiss, N. (1986). Genus *Lactobacillus*. In *Bergey's manual of systematic bacteriology*, pp. 1209-1234. Edited by P. H. A. Sneath, N. S. Mair, M. E. Sharpe & J. G. Holt. Baltimore: Williams Wilkins.

Kergourlay, G., Messaoudi, S., Dousset, X. & Prévost, H. (2012). Genome sequence of *Lactobacillus salivarius* SMXD51, a potential probiotic strain isolated from chicken cecum, showing anti-*Campylobacter* activity. *J Bacteriol* 194, 3008-3009.

Kerkhoven, R., van Enckevort, F. H. J., Boekhorst, J., Molenaar, D. & Siezen, R. J. (2004). Visualization for genomics: the microbial genome viewer. *Bioinformatics* 20, 1812-1814.

Kesseler, M., Dabbs, E. R., Averhoff, B. & Gottschalk, G. (1996). Studies on the isopropylbenzene 2,3-dioxygenase and the 3-isopropylcatechol 2,3-dioxygenase genes encoded by the linear plasmid of *Rhodococcus erythropolis* BD2. *Microbiology* 142, 3241-3251.

Kieser, H. M., Kieser, T. & Hopwood, D. A. (1992). A combined genetic and physical map of the Streptomyces coelicolor A3(2) chromosome. *J Bacteriol* 174, 5496-5507.

Kilstrup, M., Hammer, K., Ruhdal Jensen, P. & Martinussen, J. (2005). Nucleotide metabolism and its control in lactic acid bacteria. *FEMS Microbiol Rev* 29, 555-590.

Kim, C. C., Joyce, E. A., Chan, K. & Falkow, S. (2002). Improved analytical methods for microarray-based genome-composition analysis. *Genome Biol* 3, research0065.0061-0065.0017

Kim, D. S., Singh, K. V., Nallapareddy, S. R., Qin, X., Panesso, D., Arias, C. A. & Murray, B. E. (2010). The fms21 (*pilA*)-fms20 locus encoding one of four distinct pili of *Enterococcus faecium* is harboured on a large transferable plasmid associated with gut colonization and virulence. *J Med Microbiol* 59, 505-507.

Kinashi, H. & Shimaji, M. (1987). Detection of giant linear plasmids in antibiotic producing strains of *Streptomyces* by the OFAGE technique. *J Antibiot* 40, 913-916.

Kinashi, H., Shimaji, M. & Sakai, A. (1987). Giant linear plasmids in *Streptomyces* which code for antibiotic biosynthesis genes. *Nature* 328, 454-456.

Kinashi, H., Murayama, M., Matsushita, H. & Nimi, O. (1993). Structural analysis of the giant linear plasmid SCP1 in various *Streptomyces coelicolor* strains. *J Gen Microbiol* 139, 1261-1269.

Kinashi, H. (2011). Giant linear plasmids in *Streptomyces*: a treasure trove of antibiotic biosynthetic clusters. *J Antibiot* 64, 19-25.

Kirtzalidou, E., Pramateftaki, P., Kotsou, M. & Kyriacou, A. (2011). Screening for lactobacilli with probiotic properties in the infant gut microbiota. *Anaerobe* 17.

Kitten, T. & Barbour, A. G. (1990). Juxtaposition of expressed variable antigen genes with a conserved telomere in the bacterium *Borrelia hermsii*. *Proc Natl Acad Sci USA*, 6077–6081.

Klaenhammer, T. R. (2000). Probiotic bacteria: today and tomorrow. *J Nutr* 130, 415-416.

Klappenbach, J. A., Dunbar, J. M. & Schmidt, T. M. (2000). rRNA operon copy number reflects ecological strategies of bacteria. *Appl Environ Microbiol* 66, 1328-1333.

Klassen, R. & Meinhardt, F. (2007). Linear protein-primed replicating plasmids in eukaryotic microbes. In *Microbial Linear Plasmids*, pp. 187-226. Edited by F. Meinhardt & R. Klassen: Springer Berlin / Heidelberg.

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

Kobryn, K. (2007). The linear hairpin replicons of *Borrelia burgdorferi*. In *Microbial Linear Plasmids*, pp. 117-140. Edited by F. Meinhardt & R. Klassen: Springer Berlin / Heidelberg.

Kok, J., Buist, G., Zomer, A. L., Hijum, S. A. F. T. & Kuipers, O. P. (2005). Comparative and functional genomics of lactococci. *FEMS Microbiol Rev* 29, 411-433.

Kong, H. H. (2011). Skin microbiome: genomics-based insights into the diversity and role of skin microbes. *Trends Mol Med* 17, 320-328.

Konig, C., Eulberg, D., Groning, J., Lakner, S., Seibert, V., Kaschabek, S. R. & Schlomann, M. (2004). A linear megaplasmid, p1CP, carrying the genes for chlorocatechol catabolism of *Rhodococcus opacus* 1CP. *Microbiology* 150, 3075-3087.

Koo, H., Xiao, J., Klein, M. I. & Jeon, J. G. (2010). Exopolysaccharides produced by *Streptococcus mutans* glucosyltransferases modulate the establishment of microcolonies within multispecies biofilms. *Journal of Bacteriology* 192, 3024-3032.

Kosono, S., Maeda, M., Fuji, F., Arai, H. & Kudo, T. (1997). Three of the seven *bphC* genes of *Rhodococcus erythropolis* TA421, isolated from a termite ecosystem, are located on an indigenous plasmid associated with biphenyl degradation. *Appl Environ Microbiol* 63, 3282-3285.

Krum, J. G. & Ensign, S. A. (2001). Evidence that a linear megaplasmid encodes enzymes of aliphatic alkene and epoxide metabolism and coenzyme M (2-Mercaptoethanesulfonate) biosynthesis in *Xanthobacter* strain Py2. *J Bacteriol* 183, 2172-2177.

Lancaster, H., Roberts, A. P., Bedi, R., Wilson, M. & Mullany, P. (2004). Characterization of Tn916S, a Tn916-like element containing the tetracycline resistance determinant *tet*(*S*). *J Bacteriol* 186, 4395-4398.

Landeta, C., Davalos, A., Cevallos, M. A., Geiger, O., Brom, S. & Romero, D. (2011). Plasmids with a chromosome-like role in rhizobia. *J Bacteriol* 193, 1317-1326.

Lara-Villoslada, F., Sierra, S., Díaz-Ropero, M. P., Olivares, M. & Xaus, J. (2007). Safety assessment of the human milk-isolated probiotic *Lactobacillus salivarius* CECT5713. *J Dairy Sci* 90, 3583-3589.

Lara-Villoslada, F., Olivares, M. & Xaus, J. (2010). Safety of probiotic bacteria. pp. 47-58: Elsevier Inc.

Laverde Gomez, J. A., van Schaik, W., Freitas, A. R., Coque, T. M., Weaver, K. E., Francia, M. V., Witte, W. & Werner, G. (2011). A multiresistance megaplasmid pLG1 bearing a hylEfm genomic island in hospital *Enterococcus faecium* isolates. *Int J Med Microbiol* 301, 165-175.

Laws, A., Gu, Y. & Marshall, V. (2001). Biosynthesis, characterisation, and design of bacterial exopolysaccharides from lactic acid bacteria. *Biotechnol Adv* 19, 597-625.

Le Dantec, C., Winter, N., Gicquel, B., Vincent, V. & Picardeau, M. (2001). Genomic sequence and transcriptional analysis of a 23-kilobase mycobacterial linear plasmid: evidence for horizontal transfer and identification of plasmid maintenance systems. *J Bacteriol* 183, 2157-2164.

Lebeer, S., Vanderleyden, J. & De Keersmaecker, S. C. J. (2008). Genes and molecules of lactobacilli supporting probiotic action. *Microbiol Mol Biol Rev* 72, 728-764.

Lebeer, S., Vanderleyden, J. & De Keersmaecker, S. (2010). Host interactions of probiotic bacterial surface molecules: comparison with commensals and pathogens. *Nature Rev Microbiol* 8, 171 - 184.

Lebeer, S., Claes, I. J. J., Verhoeven, T. L. A., Vanderleyden, J. & De Keersmaecker, S. C. J. (2011). Exopolysaccharides of *Lactobacillus rhamnosus* GG form a protective shield against innate immune factors in the intestine. *Microbial Biotech* 4, 368-374.

Lederberg, J. (1952). Cell genetics and hereditary symbiosis. *Physiol Rev*, 403–430.

Lee, C. A., Babic, A. & Grossman, A. D. (2010). Autonomous plasmid-like replication of a conjugative transposon. *Mol Microbiol* 75, 268-279.

Lee, H., Lai, K., Lorenc, M., Imelfort, M., Duran, C. & Edwards, D. (2012). Bioinformatics tools and databases for analysis of next-generation sequence data. *Brief Funct Genomics* 11, 12-24.

Lee, J.-K., Chang, S. W., Perinpanayagam, H. & other authors (2013). Antibacterial efficacy of a human β-defensin-3 peptide on multispecies biofilms. *J Endodont* 39, 1625-1629.

Lepage, E., Brinster, S., Caron, C., Ducroix-Crepy, C., Rigottier-Gois, L., Dunny, G., Hennequet-Antier, C. & Serror, P. (2006). Comparative genomic hybridization analysis of *Enterococcus faecalis*: identification of genes absent from food strains. *J Bacteriol* 188, 6858-6868.

Leroy, F. & De Vuyst, L. (2004). Lactic acid bacteria as functional starter cultures for the food fermentation industry. *Trends Food Sci Tech* 15, 67-78.

Leroy, F., Verluyten, J. & De Vuyst, L. (2006). Functional meat starter cultures for improved sausage fermentation. *Int J Foof Microbiol* 106, 270-285.

Levy, S. B. & Marshall, B. (2004). Antibacterial resistance worldwide: causes, challenges and responses. *Nat Med* 10.

Li, W., Raoult, D. & Fournier, P.-E. (2009). Bacterial strain typing in the genomic era. *FEMS Microbiol Rev* 33, 892-916.

Li, Y., Raftis, E., Canchaya, C., Fitzgerald, G. F., van Sinderen, D. & O'Toole, P. W. (2006). Polyphasic analysis indicates that *Lactobacillus salivarius* subsp. *salivarius* and *Lactobacillus salivarius* subsp. *salicinius* do not merit separate subspecies status. *Int J Syst Evol Microbiol*, 2397-2403.

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.

Liévin-Le Moal, V. & Servin, A. L. (2014). Anti-Infective activities of *Lactobacillus* strains in the human intestinal microbiota: from probiotics to gastrointestinal anti-infectious biotherapeutic agents. *Clin Microbiol Rev* 27, 167-199.

Ljungh, A. & Wadstrom, T. (2006). Lactic acid bacteria as probiotics. *Curr Issues Intest Microbiol* 7, 73 - 90.

Loman, N., Constantinidou, C., Chan, J., Halachev, M., Sergeant, M., Penn, C., Robinson, E. & Pallen, M. (2012). High-throughput bacterial genome

sequencing: an embarrassment of choice, a world of opportunity. *Nat Rev Microbiol* 10, 599 - 606.

Maassen, C. B. M., van Holten-Neelen, C., Balk, F., Heijne den Bak-Glashouwer, M.-J., Leer, R. J., Laman, J. D., Boersma, W. J. A. & Claassen, E. (2000). Strain-dependent induction of cytokine profiles in the gut by orally administered *Lactobacillus* strains. *Vaccine* 18, 2613-2623.

Macho Fernandez, E., Pot, B. & Grangette, C. (2011). Beneficial effect of probiotics in IBD: Are peptidogycan and NOD2 the molecular key effectors? *Gut Microbes* 2, 280-286.

MacKenzie, D. A., McLay, K., Roos, S., Walter, J., Swarbreck, D., Drou, N., Crossman, L. C. & Juge, N. (2014). Draft genome sequence of a novel *Lactobacillus salivarius* strain isolated from piglet. *Genome Announc* 2.

Mackie, R. I., Sghir, A. & Gaskins, H. R. (1999). Developmental microbial ecology of the neonatal gastrointestinal tract. *Am J Clin Nutr* 69, 1035s-1045s.

MacLellan, S. R., Sibley, C. D. & Finan, T. M. (2004). Second chromosomes and megaplasmids in bacteria. In *Plasmid biology*, pp. 529–542. Edited by B. E. Funnell & G. J. Phillips. Washington, DC: ASM Press.

Maiden, M. (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A* 95, 3140 - 3145.

Maiden, M. C. J. (2006). Multilocus sequence typing of bacteria. *Annu Rev Microbiol* 60, 561-588.

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

Makarova, K. S. & Koonin, E. V. (2007). Evolutionary genomics of lactic acid bacteria. *J Bacteriol* 189, 1199-1208.

Makino, S., Uchida, I., Terakado, N., Sasakawa, C. & Yoshikawa, M. (1989). Molecular characterisation and protein analysis of the cap region, which is essential for encapsulation in *Bacillus anthracis J Bacteriol*, 722-730.

Mañes-Lázaro, R., Ferrer, S., Rosselló-Mora, R. & Pardo, I. (2008). *Lactobacillus uvarum* sp. nov. – A new lactic acid bacterium isolated from Spanish bobal grape must. *Syst Appl Microbiol* 31, 425-433.

Mañes-Lázaro, R., Ferrer, S., Rosselló-Mora, R. & Pardo, I. (2009a). *Lactobacillus oeni* sp. nov., from wine. *Int J Syst Evol Microbiol* 59, 2010-2014.

Mañes-Lázaro, R., Song, J., Pardo, I., Cho, J.-C. & Ferrer, S. (2009b). *Lactobacillus aquaticus* sp. nov., isolated from a Korean freshwater pond. *Int J Syst Evol Microbiol* 59, 2215-2218.

Manganelli, R., Ricci, S. & Pozzi, G. (1996). Conjugative transposon Tn916: evidence for excision with formation of 5'-protruding termini. *J Bacteriol* 178, 5813-5816.

Manganelli, R., Ricci, S. & Pozzi, G. (1997). The joint of Tn916 circular intermediates Is a homoduplex in *Enterococcus faecalis*. *Plasmid* 38, 71-78.

Mardanov, A. V. & Ravin, N. V. (2007). The Antirepressor Needed for Induction of Linear Plasmid-Prophage N15 Belongs to the SOS Regulon. *J Bacteriol* 189, 6333-6338.

Mardis, E. (2011). A decade's perspective on DNA sequencing technology. *Nature* 470, 198-203.

Mardis, E. R. (2013). Next-Generation Sequencing Platforms. *Annu Rev Anal Chem* 6, 287-303.

Margulies, M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376-380.

Markowitz, V. M., Mavromatis, K., Ivanova, N. N., Chen, I.-M. A., Chu, K. & Kyrpides, N. C. (2009). IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* 25, 2271-2278.

Marston, C., Pereira, C., Ferguson, J., Fischer, K., Hedstrom, O., Dashwood, W. & Baird, W. (2001). Effect of a complex environmental mixture from coal tar containing polycyclic aromatic hydrocarbons (PAH) on the tuor initiation, PAH-DNA binding and metabolic activation of carcinogenic PAH in mouse epidermis. . *Carcinogenesis*, 1077-1086.

Marteau, P., Pochart, P., Dore, J., Bera-Maillet, C., Bernalier, A. & Corthier, G. (2001). Comparative study of bacterial groups within the human cecal and fecal microbiota. *Appl Environ Microbiol* 67, 4939-4942.

Martín, R., Jiménez, E., Olivares, M., Marín, M. L., Fernández, L., Xaus, J. & Rodríguez, J. M. (2006). *Lactobacillus salivarius* CECT 5713, a potential probiotic strain isolated from infant feces and breast milk of a mother-child pair. *Int J Food Microbiol* 112, 35-43.

Martínez-Bueno, M., Valdivia, E., Gálvez, A., Mmqueda, M. (2000). pS86, a new theta-replicating plasmid from *Enterococcus faecalis*. *Current Microbiology* 41, 257-261.

Mastrangelo, G., Fadda, E. & Marzia, V. (1996). Polycyclic aromatic hydrocarbons and cancer in man *Environ Health Perspect*, 1166-1170.

Mathur, S. & Singh, R. (2005). Antibiotic resistance in food lactic acid bacteria--a review. *Int J Food Microbiol* 105, 281-295.

Maxam, A. & Gilbert, W. (1977). A new method for sequencing DNA. *Proc Natl Acad Sci USA* 74, 560-564.

Mayo, B., Hardisson, C. & Braña, A. (1989). Selected characteristics of several strains of *Lactobacillus plantarum. Microbiologia* 5, 105-112.

McCormack, J. E., Hird, S. M., Zellmer, A. J., Carstens, B. C. & Brumfield, R. T. (2013). Applications of next-generation sequencing to phylogeography and phylogenetics. *Mol Phylogenet Evol* 66, 526-538.

McCutcheon, J. P. & von Dohlen, C. D. (2011). An Interdependent metabolic patchwork in the nested symbiosis of mealybugs. *Curr Biol* 21, 1366-1372.

McLeod, M. P., Warren, R. L., Hsiao, W. W. L. & other authors (2006). The complete genome of *Rhodococcus* sp. RHA1 provides insights into a catabolic powerhouse. *Proc Nat Acad Sci* 103, 15582-15587.

McLoughlin, R. M. & Mills, K. H. G. (2011). Influence of gastrointestinal commensal bacteria on the immune responses that mediate allergy and asthma. *J Allergy Clin Immunol* 127, 1097-1107.

Medema, M. H., Trefzer, A., Kovalchuk, A. & other authors (2010). The sequence of a 1.8-Mb bacterial linear plasmid reveals a rich evolutionary reservoir of secondary metabolic pathways. *Genome Biol Evol* 2, 212-224.

Meinhardt, F., Schaffrath, R. & Larsen, M. (1997). Microbial linear plasmids. *Appl Microbiol Biotechnol* 47, 329-336.

Meinhardt, F. & Klassen, R. (2007).Microbial Linear Plamids. In *Microbial Monographs* pp. 252. Edited by A. Steinbuchel. Berlin-Heidelberg-New York. : Springer.

Mellmann, A., Harmsen, D., Cummings, C. A. & other authors (2011). Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. *PLoS ONE* 6, e22751.

Messaoudi, S. (2011). Identification of lactobacilli residing in chicken ceca with antagonism against *Campylobacter*. *Int Microbiol*, 103–110.

Messaoudi, S., Madi, A., Prévost, H., Feuilloley, M., Manai, M., Dousset, X. & Connil, N. (2012). *In vitro* evaluation of the probiotic potential of *Lactobacillus salivarius* SMXD51. *Anaerobe* 18, 584-589.

Messaoudi, S., Manai, M., Kergourlay, G., Prévost, H., Connil, N., Chobert, J. M. & Dousset, X. (2013). *Lactobacillus salivarius*: bacteriocin and probiotic activity. *Food Microbiology* 36, 296-304.

Metzker, M. L. (2010). Sequencing technologies-the next generation. *Nat Rev Genet* 11, 31-46.

Mikesell, P., Ivins, B., Ristroph, J. & Dreier, T. (1983). Evidence for plasmid-mediated toxin production in *Bacillus anthracis Infect Immun*, 371-376.

Mills, S., McAuliffe, O. E., Coffey, A., Fitzgerald, G. F. & Ross, R. P. (2006). Plasmids of lactococci - genetic accessories or genetic necessities? *FEMS Microbiology Reviews* 30, 243-273.

Mitsuoka, T. (1969). Vergleichende untersuchungen über die *Laktobazillen* aus den faeces von menschen, schweinen und hühnern. *Bakteriol* 210, 32-51.

Mock, M. & Fouet, A. (2001). Anthrax. *Annu Rev Microbiol*, 647-671.

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.

Molin, G., Jeppsson, B., Johansson, M. L., Ahrne, S., Nobaek, S., Stahl, M., Bengmark, S. & (1993a). Numerical taxonomy of *Lactobacillus* spp. associated with healthy and diseased mucosa of the human intestines. *J Appl Bacteriol*, 314-323.

Molin, G., Jeppsson, B., Johansson, M. L., Ahrne, S., Nobaek, S., Stahl, M., S Bengmark . 74 & (1993b). Numerical taxonomy of *Lactobacillus* spp. associated with healthy and diseased mucosa of the human intestines. *J Appl Bacteriol*.

Monchy, S., Benotmane, M. A., Janssen, P., Vallaeys, T., Taghavi, S., van der Lelie, D. & Mergeay, M. (2007). Plasmids pMOL28 and pMOL30 of *Cupriavidus metallidurans* are specialized in the maximal viable response to heavy metals. *J Bacteriol* 189, 7417-7425.

Morey, M., Fernández-Marmiesse, A., Castiñeiras, D., Fraga, J., Couce, M. & Cocho, J. (2013). A glimpse into past, present, and future DNA sequencing. *Mol Genet Metab* 110, 3-24.

Morita, H., Shiratori, C., Murakami, M. & other authors (2007). *Lactobacillus hayakitensis* sp. nov., isolated from intestines of healthy thoroughbreds. *Int J Syst Evol Microbiol* 57, 2836-2839.

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.

Mullany, P., Roberts, A. P. & Wang, H. (2002). Mechanism of integration and excision in conjugative transposons. *Cell Mol Life Sci* 59, 2017-2022.

Muriana, P. M. & Klaenhammer, T. R. (1987). Conjugal transfer of plasmid-encoded determinants for bacteriocin production and immunity in *Lactobacillus acidophilus* 88. *Appl Environ Microbiol* 53, 553-560.

Mutreja, A., Kim, D. W., Thomson, N. R. & other authors (2011). Evidence for several waves of global transmission in the seventh cholera pandemic. *Nature* 477, 462-465.

Nataro, J. P. & Kaper, J. B. (1998). Diarrheagenic *Escherichia coli*. *Clin Microbiol Rev* 11, 403-.

Netolitzky, D. J., Wu, X., Jensen, S. E. & Roy, K. L. (1995). Giant linear plasmids of β-lactam antibiotic producing *Streptomyces*. *FEMS Microbiol Lett* 131, 27-34.

Neville, B. A. & O'Toole, P. W. (2010). Probiotic properties of *Lactobacillus salivarius* and closely related *Lactobacillus* species. *Future Microbiol* 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.

Ng, W. V., Ciufo, S. A., Smith, T. M. & other authors (1998). Snapshot of a large dynamic replicon in a Halophilic archaeon: megaplasmid or minichromosome? *Genome Res* 8, 1131-1141.

Nicolas, P., Bessieres, 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.

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 Syst Evol Microbiol* 57, 1468-1472.

Nojiri, H., Sota, M. & Shintani, M. (2009). Catabolic plasmids involved in the degradation of polycyclic aromatic hydrocarbons and heteroaromatic compounds. In *Microbial Megaplasmids*, pp. 55-87. Edited by E. Schwartz: Springer Berlin / Heidelberg.

Norgren, M. & Scott, J. R. (1991). The presence of conjugative transposon Tn916 in the recipient strain does not impede transfer of a second copy of the element. *J Bacteriol* 173, 319-324.

Norris, S. J. (2006). Norris Antigenic variation with a twist—the *Borrelia* story. *Mol Microbiol*, 1319–1322.

Norris, S. J., Coburn, J., Leong, J. M., Hu, L. T. & Hook, M. (2010). Pathobiology of Lyme disease *Borrelia*. In *Borrelia: Molecular Biology, Host Interaction and Pathogenesis* pp. 300–331. Edited by D. S. Samuels & J. D. Radolf. Norwich UK: Horizon Scientific Press.

O'Callaghan, J., Buttó, L. F., MacSharry, J., Nally, K. & O'Toole, P. W. (2012). Influence of adhesion and bacteriocin production by *Lactobacillus salivarius* on the intestinal epithelial cell transcriptional response. *Applied and Environmental Microbiology* 78, 5196-5203.

O'Shea, E. F., O'Connor, P. M., Raftis, E. J., O'Toole, P. W., Stanton, C., Cotter, P. D., Ross, R. P. & Hill, C. (2011). Production of multiple bacteriocins from a single locus by gastrointestinal strains of *Lactobacillus salivarius*. *Journal of Bacteriology*.

Ochman, H., Lawrence, J. & Groisman, E. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* 405, 299 - 304.

Okinaka, R., Cloud, K., Hampton, O. & other authors (1999). Sequence and organisation of pXO1, the large *Bacillus anthracis* plasmid harboring the anthrax toxin genes *J Bacteriol*, 6509-6515.

Olivares, M., Diaz-Ropero, M. P., Martin, R., Rodriguez, J. M. & Xaus, J. (2006). Antimicrobial potential of four *Lactobacillus* strains isolated from breast milk. *J Appl Microbiol* 101, 72-79.

Oresnik, I. & Long, S. (2000). Megaplasmid pRme2011a of *Simorhizobium meliloti* is not required for viability *J Bacteriol*, 3582-3586.

Osawa, R., Fujisawa, T. & Pukall, R. (2006). *Lactobacillus apodemi* sp. nov., a tannase-producing species isolated from wild mouse faeces. *Int J Syst Evol Microbiol* 56, 1693-1696.

Osborn, M. A. & Böltner, D. (2002). When phage, plasmids, and transposons collide: genomic islands, and conjugative- and mobilizable-transposons as a mosaic continuum. *Plasmid* 48, 202-212.

Overhage, J., Sielker, S., Homburg, S., Parschat, K. & Fetzner, S. (2005). Identification of large linear plasmids in *Arthrobacter* spp. encoding the degradation of quinaldine to anthranilate. *Microbiology* 151, 491-500.

Pace, N. (2009). Mappng the tree of life: progress and prospects *Microbiol Mol Biol Rev* 73, 565-576.

Panel, F. (2005). Opinion of the Scientific Panel on additives and products or substances used in animal feed (FEEDAP) on the updating of the criteria

used in the assessment of bacteria for resistance to antibiotics of human or veterinary importance *EFSA Journal*.

Parkhill, J. & Wren, B. (2011). Bacterial epidemiology and biology - lessons from genome sequencing. *Genome Biol* 12, 230.

Parsot, C. (2005). *Shigella* spp. and enteroinvasive *Escherichia coli* pathogenicity factors. *FEMS Microbiol Lett* 252, 11-18.

Patel, A. & Prajapati, J. (2012). Food and health applications of exopolysaccharides produceed by lactic acid bacteria. *J Adv Dairy Res* 1.

Patten, D. A., Leivers, S., Chadha, M. J., Maqsood, M., Humphreys, P. N., Laws, A. P. & Collett, A. (2014). The structure and immunomodulatory activity on intestinal epithelial cells of the EPSs isolated from *Lactobacillus helveticus* sp. Rosyjski and *Lactobacillus acidophilus* sp. 5e2. *Carbohydr Res* 384, 119-127.

Pavlova, S. I., Kilic, A. O., Kilic, S. S., So, J.-S., Nader-Macias, M. E., Simoes, J. A. & Tao, L. (2002). Genetic diversity of vaginal lactobacilli from women in different countries based on 16S rRNA gene sequences. *J Appl Microbiol*, 451-459.

Pedersen, C. & Roos, S. (2004). *Lactobacillus saerimneri* sp. nov., isolated from pig faeces. *Int J Syst Evol Microbiol* 54, 1365-1368.

Peng, J., Yang, J. & Jin, Q. (2009). The molecular evolutionary history of *Shigella* spp. and enteroinvasive *Escherichia coli*. *Infect Genet Evol* 9, 147-152.

Perreten, V., Schwarz, F., Cresta, L., Boeglin, M., Dasen, G. & Teuber, M. (1997). Antibiotic resistance spread in food. *Nature* 389, 801-802.

Pettersson, E., Lundeberg, J. & Ahmadian, A. (2009). Generations of sequencing technologies. *Genomics* 93, 105-111.

Picardeau, M. & Vincent, V. (1997). Characterization of large linear plasmids in mycobacteria. *J Bacteriol* 179, 2753-2756.

Picardeau, M. & Vincent, V. (1998). Mycobacterial linear plasmids have an invertron-like structure related to other linear replicons in actinomycetes. *Microbiology*, 1981–1988.

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.

Pidot, S., Tobias, N. & Stinear, T. (2009). The pMUM megaplasmid of; *Mycobacterium ulcerans*; and closely related mycobacteria: a blueprint for

the synthesis of mycolactones. In *Microbial Megaplasmids*, pp. 283-296. Edited by E. Schwartz: Springer Berlin / Heidelberg.

Pohlmann, A., Fricke, W. F., Reinecke, F. & other authors (2006). Genome sequence of the bioplastic-producing Knallgas bacterium *Ralstonia eutropha* H16. *Nat Biotech* 24, 1257-1262.

Polo, S., Guerini, O., Sosio, M. & Dehb, G. (1998). Identification of two linear plasmids in the actinomycete *Planobispora rosea*. *Microbiology* 144, 2819-2825.

Popp, A., Hertwig, S., Lurz, R. & Appel, B. (2000). Comparative study of temperate bacteriophages isolated from *Yersinia Syst Appl Microbiol* 23, 469-478.

Pot, B., Ludwig, W., Kersters, K. & Schleifer, K.-H. (1994). Taxonomy of lactic acid bacteria. In *Bacteriocins of lactic acid bacteria: microbiology, genetics and applications*, pp. 13-90. Edited by L. d. Vuyst & E. J. Vandamme. UK: Blackie Academic & Professional.

Poyart-Salmeron, C., Trieu-Cuot, P., Carlier, C. & Courvalin, P. (1990). The integration-excision system of the conjugative transposon Tn 1545 is structurally and functionally related to those of lambdoid phages. *Mol Microbiol* 4, 1513-1521.

Prager, E. & Wilson, A. (1978). Construction of phylogenetic trees for proteins and nucleic acids: empirical evaluation of alternative matrix methods. *J Mol Evol* 11, 129-142.

Pretzer, G., Snel, J., Molenaar, D. & other authors (2005). Biodiversity-based identification and functional characterization of the mannose-specific adhesin of *Lactobacillus plantarum*. *J Bacteriol* 187, 6128–6136.

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

Priest, F. G., Barker, M., Baillie, L. W. J., Holmes, E. C. & Maiden, M. C. J. (2004). Population structure and evolution of the *Bacillus cereus* group. *J Bacteriol* 186, 7959-7970.

Pueppke, S. G. & Broughton, W. J. (1999). Rhizobium sp. strain NGR234 and *R. fredii* USDA257 share exceptionally broad, nested host ranges. *Mol Plant Microbe Interact* 12, 293-318.

Purser, J. E. & Norris, S. J. (2000). Correlation between plasmid content and infectivity in *Borrelia burgdorferi*. *Proc Nat Acad Sci USA* 97, 13865-13870.

Qin, Z. & Cohen, S. N. (1998). Replication at the telomeres of the *Streptomyces* linear plasmid pSLA2. *Mol Microbiol*, 893-904.

Qin, Z., Shen, M. & Cohen, S. N. (2003). Identification and characterization of a pSLA2 plasmid locus required for linear DNA replication and circular plasmid stable inheritance in *Streptomyces lividans*. *J Bacteriol* 185, 6575-6582.

Raftis, E. J. & O'Toole, P. W. (2009). Megaplasmids in *Lactobacillus salivarius*. In *Microbial Megaplasmids*, pp. 311-325. Edited by E. Schwartz. Berlin: Springer Science and Business Media.

Rasmussen, T., Danielsen, M., Valina, O., Garrigues, C., Johansen, E. & Pedersen, M. (2008). *Streptococcus thermophilus* core genome: comparative genome hybridization study of 47 strains. *Appl Environ Microbiol* 74, 4703-4710.

Ravel, J., Schrempf, H. & Hill, R. T. (1998). Mercury resistance Is encoded by transferable giant linear plasmids in two chesapeake bay *Streptomyces* strains. *Appl Environ Microbiol* 64, 3383-3388.

Ravel, J., Wellington, E. M. H. & Hill, R. T. (2000). Interspecific transfer of *Streptomyces* giant linear plasmids in sterile amended soil microcosms. *Appl Environ Microbiol* 66, 529-534.

Ravin, N. V. (2011). N15: The linear phage-plasmid. *Plasmid* 65, 102-109.

Ravin, V. & Golub, E. (1967). A study of phage conversion in *Escherichia coli*. I.The aquisition of resistance to bacteriophage T1 as a result of lysogenization. *Genetika*, 113–121.

Ravin, V., Ravin, N., Casjens, S., Ford, M. E., Hatfull, G. F. & Hendrix, R. W. (2000). Genomic sequence and analysis of the atypical temperate bacteriophage N15. *Journal of Molecular Biology* 299, 53-73.

Redenbach, M., Bibb, M., Gust, B., Seitz, B. & Spychaj, A. (1999). The linear plasmid SCP1 of *Streptomyces coelicolor* A3(2) possesses a centrally located replication origin and shows significant homology to the transposon Tn4811. *Plasmid*, 174-185.

Redondo-Lopez, V., Cook, R. L. & Sobel, J. D. (1990). Emerging role of lactobacilli in the control and maintenance of the vaginal bacterial microflora. *Rev Infect Dis* 12.

Remenant, B., Coupat-Goutaland, B., Guidot, A. & other authors (2010). Genomes of three tomato pathogens within the Ralstonia solanacearum species complex reveal significant evolutionary divergence. *BMC Genomics* 11, 379.

Remus, D. M., Kleerebezem, M. & Bron, P. A. (2011). An intimate tête-à-tête — How probiotic lactobacilli communicate with the host. *Eur J Pharmacol* 668, Supplement 1, S33-S42.

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

Rice, L. B. & Carrias, L. L. (1994). Studies on excision of conjugative transposons in enterococci: evidence for joint sequences composed of strands with unequal numbers of nucleotides. *Plasmid*, 312–316.

Rice, L. B., Carias, L., Rudin, S. & other authors (2003). A potential virulence gene, *hylEfm*, predominates in *Enterococcus faecium* of clinical origin. *J Infect Dis* 187, 508-512.

Rice, L. B., Carias, L. L., Marshall, S. H., Hutton-Thomas, R. & Rudin, S. (2007). Characterization of Tn5386, a Tn916-related mobile element. *Plasmid* 58, 61-67.

Rice, L. B., Laktičova, V., Carias, L. L., Rudin, S., Hutton, R. & Marshall, S. H. (2009). Transferable capacity for gastrointestinal colonization in *Enterococcus faecium* in a mouse model. *J Infect Dis* 199, 342-349.

Ricker, N., Qian, H. & Fulthorpe, R. R. (2012). The limitations of draft assemblies for understanding prokaryotic adaptation and evolution. *Genomics* 100, 167-175.

Rizzotti, L., La Gioia, F., Dellaglio, F. & Torriani, S. (2009). Molecular diversity and transferability of the tetracycline resistance gene *tet*(*M*), carried on Tn916-1545; family transposons, in enterococci from a total food chain. *Antonie van Leeuwenhoek* 96, 43-52.

Roberts, A. P., Cheah, G., Ready, D., Pratten, J., Wilson, M. & Mullany, P. (2001a). Transfer of Tn916-like elements in microcosm dental plaques. *Antimicrob Agents Chemother* 45, 2943-2946.

Roberts, A. P., Johanesen, P. A., Lyras, D., Mullany, P. & Rood, J. I. (2001b). Comparison of Tn5397 from *Clostridium difficile*, Tn916 from *Enterococcus faecalis* and the CW459*tet*(*M*) element from *Clostridium perfringens* shows that they have similar conjugation regions but different insertion and excision modules. *Microbiology* 147, 1243-1251.

Roberts, A. P., Chandler, M., Courvalin, P. & other authors (2008). Revised nomenclature for transposable genetic elements. *Plasmid* 60, 167-173.

Roberts, A. P. & Mullany, P. (2009). A modular master on the move: the Tn916 family of mobile genetic elements. *Trends Microbiol* 17, 251-258.

Rocco, J. M. & Churchward, G. (2006). The integrase of the conjugative transposon Tn916 directs strand- and sequence-specific cleavage of the origin of conjugal transfer, *oriT*, by the endonuclease Orf20. *J Bacteriol* 188, 2207-2213.

Rodas, A. M., Chenoll, E., Macián, M. C., Ferrer, S., Pardo, I. & Aznar, R. (2006). *Lactobacillus vini* sp. nov., a wine lactic acid bacterium homofermentative for pentoses. *Int J Syst Evol Microbiol* 56, 513-517.

Rogosa, M., Wiseman, R. F., Mitchell, J. A., Disraely, M. N. & Beaman, A. J. (1953). Species differentiation of oral lactobacilli from man including descriptions of *Lactobacillus salivarius* nov spec and *Lactobacillus cellobiosus* nov spec. *J Bacteriol* 65, 681-699.

Rosenberg, C., Boistard, P., Denarie, J. & Casse-Delbart, F. (1981). Genes controlling early and late functions in symbiosis are located on a megaplasmid in *Rhizobium meliloti*. *Mol Gen Genet* 184, 326 - 333.

Rosenberg, C., Casse-Delbart, F., Dusha, I., David, M. & Boucher, C. (1982). Megaplasmids in the plant-associated bacteria *Rhizobium meliloti* and *Pseudomonas solanacearum*. *J Bacteriol* 150, 402-406.

Rouse, S., Canchaya, C. & van Sinderen, D. (2008). *Lactobacillus hordei* sp. nov., a bacteriocinogenic strain isolated from malted barley. *Int J Syst Evol Microbiol* 58, 2013-2017.

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

Ruas-Madiedo, P., Hugenholtz, J. & Zoon, P. (2002). An overview of the functionality of exopolysaccharides produced by lactic acid bacteria. *Int Dairy J* 12, 163-171.

Ruas-Madiedo, P. & de los Reyes-Gavilan, C. G. (2005). Invited review: methods for the screening, isolation, and characterization of exopolysaccharides produced by lactic acid bacteria. *J Dairy Sci* 88, 843-856.

Ruas-Madiedo, P., Medrano, M., Salazar, N., De Los Reyes-Gavilán, C. G., Pérez, P. F. & Abraham, A. G. (2010). Exopolysaccharides produced by *Lactobacillus* and *Bifidobacterium* strains abrogate in vitro the cytotoxic effect of bacterial toxins on eukaryotic cells. *J Appl Microbiol* 109, 2079-2086.

Ruas-Madiedo, P. (2014). Biosynthesis and bioactivity of exopolysaccharides produced by probiotic bacteria. In *Food Oligosaccharides*, pp. 118-133: John Wiley & Sons, Ltd.

Rudy, C., Taylor, K. L., Hinerfeld, D., Scott, J. R. & Churchward, G. (1997). Excision of a conjugative transposon in vitro by the Int and Xis proteins of Tn916. *Nucleic Acids Res* 25, 4061-4066.

Rudy, C. K. & Scott, J. R. (1994). Length of the coupling sequence of Tn916. *J Bacteriol* 176, 3386-3388.

Ruiz-Garbajosa, P., Bonten, M. J. M., Robinson, D. A. & other authors (2006). Multilocus sequence typing scheme for *Enterococcus faecalis* reveals hospital-adapted genetic complexes in a background of high rates of recombination. *J Clin Microbiol* 44, 2220-2228.

Ryan, K., O'Hara, A., van Pijkeren, J.-P., Douillard, F. & O'Toole, P. (2009). *Lactobacillus salivarius* modulates cytokine induction and virulence factor gene expression in *Helicobacter pylori*. *J Med Microbiol* 58, 996-1005.

Ryan, K. A., Daly, P., Li, Y., Hooton, C. & O'Toole, P. W. (2008). Strain-specific inhibition of *Helicobacter pylori* by *Lactobacillus salivarius* and other lactobacilli. *J Antimicrob Chemoth* 61, 831-834.

Sakaguchi, K. (1990). Invertrons, a class of structurally and functionally related genetic elements that includes linear DNA plasmids, transposable elements, and genomes of adeno-type viruses. *Microbiol Mol Biol Rev* 54, 66-74.

Salanoubat, M., Genin, S., Artiguenave, F. & other authors (2002). Genome sequence of the plant pathogen *Ralstonia solanacearum*. *Nature* 415, 497 - 502.

Salvana, E. M. T. & Frank, M. (2006). *Lactobacillus* endocarditis: Case report and review of cases reported since 1992. *J Infect* 53, e5-e10.

Salvetti, E., Torriani, S. & Felis, G. (2012). The genus *Lactobacillus*: A taxonomic update. *Probiotics & Antimicro Prot*, 1-10.

Salyers, A., Shoemaker, N., Stevens, A. & Li, L. (1995). Conjugative transposons: an unusual and diverse set of integrated gene transfer elements. *Microbiol Rev* 59, 579-590.

Sanders, M. E., Akkermans, L. M. A., Haller, D. & other authors (2010). Safety assessment of probiotics for human use. *Gut Microbes* 1, 1-22.

Sanger, F., Nicklen, S. & Coulson, A. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Nat Acad Sci USA* 74, 5463-5467.

Sarkar, S. F., Gordon, J. S., Martin, G. B. & Guttman, D. S. (2006). Comparative genomics of host-specific virulence in *Pseudomonas syringae*. *Genetics* 174, 1041-1056.

Sazawal, S., Hiremath, G., Dhingra, U., Malik, P., Deb, S. & Black, R. (2006). Efficacy of probiotics in prevention of acute diarrhoea: a meta-analysis of masked, randomised, placebo-controlled trials. *Lancet Infect Dis* 6, 374-382.

Schadt, E. E., Linderman, M. D., Sorenson, J., Lee, L. & Nolan, G. P. (2010). Computational solutions to large-scale data management and analysis. *Nat Rev Genet* 11, 647-657.

Schadt, E. E., Turner, S. & Kasarskis, A. (2011). A window into third generation sequencing. *Hum Mol Genet* 20, 853.

Schatz, M. C., Delcher, A. L. & Salzberg, S. L. (2010). Assembly of large genomes using second-generation sequencing. *Genome Res* 20, 1165-1173.

Schjørring, S. & Krogfelt, K. A. (2011). Assessment of bacterial antibiotic resistance transfer in the gut. *Int J Microbiol* 2011, 10.

Schleifer, K.-H., Ehrmann, M., Beimfohr, C., Brockmann, E., Ludwig, W. & Amann., R. (1995). Application of molecular methods for the classification and identification of lactic acid bacteria. *Int Dairy J*, 1081–1094.

Schleifer, K. H. & Ludwig, W. (1995). Phylogeny of the genus *Lactobacillus* and related genera. *System Appl Microbiol* 18, 461 - 467.

Schoolnik, G. K. (2002). Microarray analysis of bacterial pathogenicity. *Adv Microb Physiol* 46, 1 - 45.

Schutzer, S. E., Fraser-Liggett, C. M., Casjens, S. R., Qiu, W.-G., Dunn, J. J., Mongodin, E. F. & Luft, B. J. (2011). Whole-genome sequences of thirteen isolates of *Borrelia burgdorferi*. *J Bacteriol* 193, 1018-1020.

Schwartz, D. C. & Cantor, C. R. (1984). Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell* 37, 67-75.

Schwartz, E. (2009).Microbial Megaplasmids. In *Microbiology Monographs*. Edited by A. Steinbüchel. Berlin and Münster: Springer.

Scott, J. R., Bringel, F., Marra, D., Van Alstine, G. & Rudy, C. K. (1994). Conjugative transposition of Tn916: preferred targets and evidence for conjugative transfer of a single strand and for a double-stranded circular intermediate. *Mol Microbiol* 11, 1099-1108.

Scott, K. P. (2002). The role of conjugative transposons in spreading antibiotic resistance between bacteria that inhabit the gastrointestinal tract. *Cellular and Molecular Life Sciences* 59, 2071-2082.

Sebaihia, M., Wren, B. W., Mullany, P. & other authors (2006). The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nat Genet* 38, 779-786.

Seegers, J. F. (2002). Lactobacilli as live vaccine delivery vectors: progress and prospects. *Trends Biotechnol* 20, 508-515.

Servin, A. L. & Coconnier, M.-H. (2003). Adhesion of probiotic strains to the intestinal mucosa and interaction with pathogens. *Best Pract Res Clin Gastroenterol* 17, 741-754.

Servin, A. L. (2004). Antagonistic activities of lactobacilli and bifidobacteria against microbial pathogens. *FEMS Microbiol Rev* 28, 405-440.

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. *Journal of General Microbiology* 77, 37-49.

Sheil, B., McCarthy, J., O'Mahony, L., Bennett, M. W., Ryan, P., Fitzgibbon, J. J., Kiely, B., Collins, J. K. & Shanahan, F. (2004). Is the mucosal route of administration essential for probiotic function? Subcutaneous administration is associated with attenuation of murine colitis and arthritis. *Gut* 53, 694-700.

Shendure, J. & Ji, H. (2008). Next-generation DNA sequencing. *Nat Biotech* 26, 1135-1145.

Shiffman, D. & Cohen, S. N. (1992). Reconstruction of *Streptomyces* linear plasmid replication from separately cloned DNA fragment: existence pf a cryptic origin of circular replication within the linear plasmid *Proc Natl Acac Sci USA*, 6129-6133.

Shimizu, S., Kobayashi, H., Masai, E. & Fukuda, M. (2001). Characterization of the 450-kb linear plasmid in a polychlorinated biphenyl degrader, *Rhodococcus* sp. strain RHA1. *Appl Environ Microbiol* 67, 2021-2028.

Showsh, S. A. & Andrews, R. E. (1992). Tetracycline enhances Tn916-mediated conjugal transfer. *Plasmid* 28, 213-224.

Sibartie, S., O'Hara, A., Ryan, J., Fanning, A., O'Mahony, J., O'Neill, S., Sheil, B., O'Mahony, L. & Shanahan, F. (2009). Modulation of pathogen-induced CCL20 secretion from HT-29 human intestinal epithelial cells by commensal bacteria. *BMC Immunol* 10.

Siefert, J. L. (2009). Defining the mobilome. In *Horizontal gene transfer*, pp. 13-27.

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

Siezen, R. J., Bayjanov, J. R., Felis, G. E., van der Sijde, M. R., Starrenburg, M., Molenaar, D., Wels, M., van Hijum, S. A. F. T. & van Hylckama Vlieg, J. E. T. (2011). Genome-scale diversity and niche adaptation analysis of *Lactococcus lactis* by comparative genome hybridization using multi-strain arrays. *Microbial Biotech* 4, 383-402.

Slater, S. C., Goldman, B. S., Goodner, B. & other authors (2009). Genome Sequences of Three Agrobacterium Biovars Help Elucidate the Evolution of Multichromosome Genomes in Bacteria. *J Bacteriol* 191, 2501-2511.

Smidt, H., Song, D., van der Oost, J. & de Vos, W. M. (1999). Random transposition by Tn916 in desulfitobacterium dehalogenans allows for isolation and characterization of halorespiration-deficient mutants. *J Bacteriol* 181, 6882-6888.

Smillie, C., Garcillan-Barcia, M. P., Francia, M. V., Rocha, E. P. C. & de la Cruz, F. (2010). Mobility of Plasmids. *Microbiol Mol Biol Rev* 74, 434-452.

Smits, T. H. M., Rezzonico, F., Pelludat, C., Goesmann, A., Frey, J. E. & Duffy, B. (2010). Genomic and phenotypic characterization of a nonpigmented variant of *Pantoea vagans* biocontrol strain C9-1 lacking the 530-kb megaplasmid pPag3. *FEMS Microbiol Lett* 308, 48-54.

Smokvina, T., Wels, M., Polka, J., Chervaux, C., Brisse, S., Boekhorst, J., Vlieg, J. E. T. v. H. & Siezen, R. J. (2013). *Lactobacillus paracasei* comparative genomics: towards species pan-genome definition and exploitation of diversity. *PLoS ONE* 8, e68731.

Snipen, L., Repsilber, D., Nyquist, L., Ziegler, A., Aakra, A. & Aastveit, A. (2006). Detection of divergent genes in microbial aCGH experiments. *BMC Bioinformatics* 7, 181.

Snipen, L., Nyquist, O., Solheim, M., Aakra, A. & Nes, I. (2009). Improved analysis of bacterial CGH data beyond the log-ratio paradigm. *BMC Bioinformatics* 10, 91.

Soge, O. O., Beck, N. K., White, T. M., No, D. B. & Roberts, M. C. (2008). A novel transposon, Tn6009, composed of a Tn916 element linked with a *Staphylococcus aureus* mer operon. *J Antimicrob Chemother* 62, 674-680.

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

Sperandio, V., Kaper, J. B., Bortolini, M. R., Cruz Neves, B., Keller, R. & Trabulsi, L. R. (1998). Characterization of the locus of enterocyte effacement (LEE) in different enteropathogenic *Escherichia coli* (EPEC) and

Shiga-toxin producing *Escherichia coli* (STEC) serotypes. *FEMS Microbiol Lett* 164, 133-139.

Srimanote, P., Paton, A. W. & Paton, J. C. (2002). Characterization of a novel type IV pilus locus encoded on the large plasmid of locus of enterocyte effacement-negative shiga-toxigenic *Escherichia coli* strains that Are virulent for humans. *Infect Immun* 70, 3094-3100.

Stackebrandt E, Frederiksen W, Garrity G M, Grimont P A & Kampfer P (2002). Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int J Syst Bacteriol*, 1043–1047.

Stanton, B. G. (1990). Crown gall disease and hairy root disease: a sledgehammer and a tackhammer. *Plant Physiol* 92, 281-285.

Stecker, C., Johann, A., Herzberg, C., Averhoff, B. & Gottschalk, G. (2003). Complete nucleotide sequence and genetic organization of the 210-kilobase linear plasmid of *Rhodococcus erythropolis* BD2. *J Bacteriol* 185, 5269-5274.

Stewart, P. E., Byram, R., Grimm, D., Tilly, K. & Rosa, P. A. (2005). The plasmids of *Borrelia burgdorferi*: essential genetic elements of a pathogen. *Plasmid* 53, 1-13.

Stiles, M. (1996). Biopreservation by lactic acid bacteria. *Antonie Leeuwenhoek*, 331–345.

Stinear, T. P., Mve-Obiang, A., Small, P. L. C. & other authors (2004). Giant plasmid-encoded polyketide synthases produce the macrolide toxin of *Mycobacterium ulcerans*. *Proc Nat Acad Sci USA* 101, 1345-1349.

Stoppel, R. D., Meyer, M. & Schlegel, H. G. (1995). The nickel resistance determinant cloned from the enterobacterium *Klebsiella oxytoca*: conjugational transfer, expression, regulation and DNA homologies to various nickel-resistant bacteria. *Biometals*
70–79.

Stoughton, R. B. (2005). Applications of DNA microarrays in biology. *Annu Rev Biochem* 74, 53-82.

Stromsten, N. J., Benson, S. D., Burnett, R. M., Bamford, D. H. & Bamford, J. K. H. (2003). The *Bacillus thuringiensis* linear double-stranded DNA phage Bam35, which is highly similar to the *Bacillus cereus* linear plasmid pBClin15, has a prophage state. *J Bacteriol* 185, 6985-6989.

Strus, M., Chmielarczyk, A., Kochan, P., Adamski, P., Chełmicki, Z., Chełmicki, A., Pałucha, A. & Heczko, P. B. (2012). Studies on the effects of probiotic *Lactobacillus* mixture given orally on vaginal and rectal colonization and on parameters of vaginal health in women with intermediate vaginal flora. *Eur J Obstet Gynecol Reprod Biol* 163, 210-215.

Stull, T. L., LiPuma, J. J. & Edlind, T. D. (1988). A Broad-Spectrum Probe for Molecular Epidemiology of Bacteria: Ribosomal RNA. *Journal of Infectious Diseases* 157, 280-286.

Su, Y. A., He, P. & Clewell, D. B. (1992). Characterization of the tet(M) determinant of Tn916: evidence for regulation by transcription attenuation. *Antimicrob Agents Chemother* 36, 769-778.

Suwa, M., Sugino, H., Sasaoka, A., Mori, E., Fujii, S., Shinkawa, H., Nimi, O. & Kinashi, H. (2000). Identification of two polyketide synthase gene clusters on the linear plasmid pSLA2-L in *Streptomyces rochei*. *Gene* 246, 123-131.

Suwanto, A. & Kaplan, S. (1989). Physical and genetic mapping of the *Rhodobacter sphaeroides* 2.4.1 genome: presence of two unique circular chromosomes. *J Bacteriol* 171, 5850-5859.

Suzuki, K., Tanaka, K., Yamamoto, S., Kiyokawa, K., Moriguchi, K. & Yoshida, K. (2009). Ti and Ri Plasmids. In *Microbial Megaplasmids*, pp. 133-147. Edited by E. Schwartz: Springer Berlin / Heidelberg.

Suzuki, S., Ono, N., Furusawa, C., Ying, B.-W. & Yomo, T. (2011). Comparison of sequence reads obtained from three next-generation sequencing platforms. *PLoS ONE* 6, e19534.

Taboada, E. N., Luebbert, C. C. & Nash, J. H. E. (2008). Studying bacterial genome dynamics using microarray-based comparative genomic hybridization. In *Comparative Genomics*, pp. 223-253.

Tabuchi, N., Mitani, H., Seino, S. & Fukunaga, M. (2002). The 44-kb linear plasmid molecule in the relapsing fever agent *Borrelia duttonii* strain Ly serve as a preservation of *vmp* genes. . *Microbiol Immunol*, 159–165.

Takahashi, Y., Cutler, S. J. & Fukunaga, M. (2000). Size conversion of a linear plasmid in the relapsing fever agent *Borrelia duttonii*. *Microbiol Immunol*, 1071–1074.

Tanaka, H., Doesburg, K., Iwasaki, T. & Mierau, I. (1999). Screening of lactic acid bacteria for bile salt hydrolase activity. *J Dairy Sci* 82, 2530-2535.

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.

Tanigawa, K. & Watanabe, K. (2011). Multilocus sequence typing reveals a novel subspeciation of *Lactobacillus delbrueckii*. *Microbiology* 157, 727-738.

Tannock, G. (1990). The microecology of lactobacilli inhabiting the gastrointestinal tract. In *Advances in Microbial Ecology*, pp. 147-171. Edited by K. C. Marshall: Springer US.

Tannock, G. (1992). The lactic microflora of pigs, mice and rats. In *The Lactic Acid Bacteria Volume 1*, pp. 21-48. Edited by B. B. Wood: Springer US.

Tannock, G., Ghazally, S., Walter, J. & other authors (2005). Ecological behavior of *Lactobacillus reuteri* 100-23 is affected by mutation of the *luxS* gene. *Appl Environ Microbiol* 71, 8419-8425.

Tannock, G. W. (1999). Analysis of the intestinal microflora: a renaissance. *Antonie van Leeuwenhoek* 76, 265-278.

Tannock, G. W., Munro, K., Harmsen, H. J. M., Welling, G. W. & Smart, P. K. J. (2000). Gobal analyses of the fecal microflora of human subjects consuming a probiotic product containing *Lactobacillus rhamnosus* DR20 *Appl Environ Microbiol*, 2578–2588.

Tannock, G. W. (2004). A special fondness for lactobacilli. *Appl Environ Microbiol* 70, 3189-3194.

Tannock, G. W. (2005). New perceptions of the gut microbiota: implications for future research. *Gastroenterol Clin North Am* 34, 361-382.

Taylor, D. & Chau, A. (1996). Tetracycline resistance mediated by ribosomal protection. *Antimicrob Agents Chemother* 40, 1-5.

te Poele, E., Bolhuis, H. & Dijkhuizen, L. (2008). Actinomycete integrative and conjugative elements. *Antonie van Leeuwenhoek* 94, 127-143.

Torres, O. R., Korman, R. Z., Zahler, S. A. & Dunny, G. M. (1991). The conjugative transposon *Tn925* enhancement of conjugal transfer by tetracycline in *Enterococcus faecalis*; and mobilization of chromosomal genes in *Bacillus subtilis* and *E. faecalis*. *Mol Gen Genet MGG* 225, 395-400.

Toussaint, A. & Merlin, C. (2002). Mobile elements as a combination of functional modules. *Plasmid* 47, 26-35.

Trieu-Cuot, P., Carlier, C. & Courvalin, P. (1988). Conjugative plasmid transfer from *Enterococcus faecalis* to *Escherichia coli*. *J Bacteriol* 170, 4388-4391.

van de Guchte, M. (2006). The complete genome sequence of *Lactobacillus bulgaricus* reveals extensive and ongoing reductive evolution. *Proc Natl Acad Sci USA* 103, 9274-9279.

van Hijum, S., Baerends, R., Zomer, A., Karsens, H., Martin-Requena, V., Trelles, O., Kok, J. & Kuipers, O. (2008). Supervised Lowess normalization of comparative genome hybridization data - application to lactococcal strain comparisons. *BMC Bioinformatics* 9, 93.

van Pijkeren, J., Canchaya, C., Ryan, K. & other authors (2006). Comparative and functional analysis of sortase-dependent proteins in the predicted secretome of *Lactobacillus salivarius* UCC118. *Appl Environ Microbiol* 72, 4143 - 4153.

Vandamme, P., Pot, B., Gillis, M., de Vos, P., Kersters, K. & Swings, J. (1996). Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol Rev* 60, 407-438.

Vedler, E. (2009). Megaplasmids and the Degradation of Aromatic Compounds by Soil Bacteria. In *Microbial Megaplasmids*, pp. 33-53. Edited by E. Schwartz: Springer Berlin Heidelberg.

Vela, A. I., Fernandez, A., Espinosade los Monteros, A., Goyache, J., Herraez, P., Tames, B., Cruz, F., Domínguez, L. & Fernández-Garayzábal, J. F. (2008). *Lactobacillus ceti* sp. nov., isolated from beaked whales (Ziphius cavirostris). *Int J Syst Evol Microbiol* 58, 891-894.

Vélez, M. P., Petrova, M. I., Lebeer, S., Verhoeven, T. L. A., Claes, I., Lambrichts, I., Tynkkynen, S., Vanderleyden, J. & De Keersmaecker, S. C. J. (2010). Characterization of MabA, a modulator of *Lactobacillus rhamnosus* GG adhesion and biofilm formation. *FEMS Immunol Med Microbiol* 59, 386-398.

Verheust, C., Jensen, G. & Mahillon, J. (2003). pGIL01, a linear tectiviral plasmid prophage originating from *Bacillus thuringiensis* serovar *israelensis*. *Microbiology* 149, 2083-2092.

Vinderola, G., Perdigón, G., Duarte, J., Farnworth, E. & Matar, C. (2006). Effects of the oral administration of the exopolysaccharide produced by *Lactobacillus kefiranofaciens* on the gut mucosal immunity. *Cytokine* 36, 254-260.

von Rozycki, T. & Nies, D. (2009). *Cupriavidus metallidurans* evolution of a metal-resistant bacterium. *Antonie van Leeuwenhoek* 96, 115-139.

Wagenknecht, M., Dib, J., Thürmer, A., Daniel, R., Farías, M. & Meinhardt, F. (2010). Structural peculiarities of linear megaplasmid, pLMA1, from *Micrococcus luteus*; interfere with pyrosequencing reads assembly. *Biotechnology Letters* 32, 1853-1862.

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.

Walter, J. (2005). The microecology of lactobacilli in the gastrointestinal tract. *Probiotics and Prebiotics: Scientific Aspects*, 51-82.

Walter, J. (2008). Ecological role of *lactobacilli* in the gastrointestinal tract: implications for fundamental and biomedical research. *Applied and Environmental Microbiology* 74, 4985-4996.

Walter, J., Schwab, C., Loach, D. M., Ganzle, M. G. & Tannock, G. W. (2008). Glucosyltransferase A (*GtfA*) and inulosucrase (*Inu*) of *Lactobacillus reuteri* TMW1.106 contribute to cell aggregation, in vitro biofilm formation, and colonization of the mouse gastrointestinal tract. *Microbiology* 154, 72-80.

Wang, H., Roberts, A. P., Lyras, D., Rood, J. I., Wilks, M. & Mullany, P. (2000). Characterization of the ends and target sites of the novel conjugative transposon Tn5397 from clostridium difficile: excision and circularization Is mediated by the large resolvase, TndX. *J Bacteriol* 182, 3775-3783.

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

Wells, J. (2011). Immunomodulatory mechanisms of lactobacilli. *Microb Cell Fact* 10, 1-15.

Welsh, J. & McClelland, M. (1990). Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res*, 7213-7218.

Whiteford, N., Haslam, N., Weber, G., Prügel-Bennett, A., Essex, J. W., Roach, P. L., Bradley, M. & Neylon, C. (2005). An analysis of the feasibility of short read sequencing. *Nucleic Acids Research* 33, e171.

Wilson, J. (2006). Genetic exchange in bacteria and the modular structure of mobile DNA elements. In *Molecular Paradigms of Infectious Disease*, pp. 34-77. Edited by C. Nickerson & M. Schurr: Springer US.

Witney, A. A., Marsden, G. L., Holden, M. T., Stabler, R. A., Husain, S. E., Vass, J. K., Butcher, P. D., Hinds, J. & Lindsay, J. A. (2005). Design, validation, and application of a seven-strain *Staphylococcus aureus* PCR product microarray for comparative genomics. *Appl Environ Microbiol* 71, 7504 - 7514.

Wlodarczyk, M. & Nowicka, B. (1988). Preliminary evidence for the linear nature of *Thiobacillus versutus* pTAV2 plasmid. *FEMS Microbiol Lett* 55.

Wozniak, A., Sciot, R., Guillou, L. & other authors (2007). Array CGH analysis in primary gastrointestinal stromal tumors: cytogenetic profile correlates with anatomic site and tumor aggressiveness, irrespective of mutational status. *Genes, Chromosomes & Cancer* 46, 261 - 276.

Wozniak, R. A. F. & Waldor, M. K. (2010). Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat Rev Micro* 8, 552-563.

Wren, B. W. (2003). The *Yersiniae*- a model genus to study the rapid evolution of bacterial pathogens. *Nat Rev Micro* 1, 55-64.

Wu, X. & Roy, K. L. (1993). Complete nucleotide sequence of a linear plasmid from *Streptomyces clavuligerus* and characterization of its RNA transcripts. *J Bacteriol* 175, 37-52.

Xue, W. & Warshawsky, D. (2005). Metabolic activation of polycyclic and heterocyclic aromatic hydrocarbons and DNA damage: a review. *Toxicol Appl Pharmacol*, 73-93.

Yan, S., Liu, H., Mohr, T. J., Jenrette, J., Chiodini, R., Zaccardelli, M., Setubal, J. C. & Vinatzer, B. A. (2008). Role of recombination in the evolution of the model plant pathogen *Pseudomonas syringae* pv. tomato DC3000, a very atypical tomato strain. *Appl Environ Microbiol* 74, 3171-3181.

Young, J., Kuykendall, L., Martinez-Romero, E., Kerr, A. & Sawada, H. (2001). A revision of *Rhizobium* Frank 1889, with an emended description of the genus, and the inclusion of all species of *Agrobacterium* Conn 1942 and *Allorhizobium undicola* de Lajudie *et al.* 1998 as new combinations: *Rhizobium radiobacter*, *R. rhizogenes*, *R. rubi*, *R. undicola* and *R. vitis*. *Int J Syst Evol Microbiol* 51, 89-103.

Young, J. M., Kuykendall, L. D., Martinez-Romero, E., Kerr, A. & Sawada, H. (2003). Classification and nomenclature of *Agrobacterium* and *Rhizobium* - a reply to Farrand *et al.* (2003). *Int J Syst Evol Microbiol* 53, 1689-1695.

Zhang, H., Sun, J., Wang, Q. Y., He, Y. T., Gu, H. Y., Guo, H. Y., Ding, Q. B., Yang, J. S. & Ren, F. Z. (2013). Safety assessment of *Lactobacillus salivarius* ren, a probiotic strain isolated from centenarian feces. *Food Sci Technol Res* 19, 1037-1043.

Zhang, J.-R. & Norris, S. J. (1998a). Kinetics and *in vivo* induction of genetic variation of *vlsE* in *Borrelia burgdorferi*. *Infect Immun* 66, 3689-3697.

Zhang, J. R., Hardham, J. M., Barbour, A. G. & Norris, S. J. (1997). Antigenic variation in Lyme disease *Borreliae* by promiscuous recombination of VMP-like sequence cassettes. *Cell*, 275–285.

Zhang, J. R. & Norris, S. J. (1998b). Genetic variation of the *Borrelia burgdorferi* gene *vlsE* involves cassette-specific, segmental gene conversion. *Infect Immun*, 3698–3704.

Zhang, X., Vrijenhoek, J. E. P., Bonten, M. J. M., Willems, R. J. L. & van Schaik, W. (2011). A genetic element present on megaplasmids allows *Enterococcus faecium* to use raffinose as carbon source. *Environ Microbiol* 13, 518-528.

Zhu, X., Zheng, B., Wang, S., Willems, R. J. L., Xue, F., Cao, X., Li, Y., Bo, S. & Liu, J. (2009). Molecular characterisation of outbreak-related strains of vancomycin-resistant *Enterococcus faecium* from an intensive care unit in Beijing, China. *J Hosp Infect* 72, 147-154.

CHAPTER II

Genomic Diversity of *Lactobacillus salivarius*

This chapter has been published as a research article:

Genomic Diversity of *Lactobacillus salivarius*

Emma J. Raftis,[1,2] Elisa Salvetti,[1,2,3] Sandra Torriani,[3] Giovanna E. Felis,[3] and Paul W. O'Toole [1,2*]

Department of Microbiology[1] and Alimentary Pharmabiotic Centre,[2] University College Cork, Cork, Ireland, and Dipartimento di Biotecnologie, Universita` degli Studi di Verona, Verona, Italy[3]

***Appl. Environ. Microbiol.* 2011, 77(3):954. DOI: 10.1128/AEM.01687-10.**

*Corresponding author.

Mailing address: School of Microbiology, University College Cork, Ireland.

Phone: +353 21 490 3997.

E-mail:pwotoole@ucc.ie

NOTE:

Some of the following sections contain work that has been carried out, mainly or partly, by people other than me. To maintain clarity and flow of the overall project, these sections or sentences are left in, but marked with number sign [#.]

Chapter II Table of Contents

ABSTRACT **132**
2.1 INTRODUCTION **133**
2.2 MATERIALS AND METHODS **135**
2.2.1 Bacterial strains and culture conditions. 135
2.2.2 Microarray description, labeling, and hybridization. 137
2.2.3 Microarray data acquisition and analysis. 138
2.2.4 Multilocus sequence typing (MLST). 139
2.2.5 Exopolysaccharide (EPS) isolation and quantification. 140
2.2.6 Confocal imaging of biofilms. 141
2.2.7 Statistical analysis. 141
2.2.8 Microarray data accession number. 141
2.3 RESULTS **142**
2.3.1 Comparative genome hybridization reveals unusually high-level diversity in *L. salivarius*. 142
2.3.2 Concordant CGH clustering and MLST phylogeny of *L. salivarius*. 147
2.3.3 Separation of *L. salivarius* strains by EPS gene content. 149
2.3.4 EPS production varies independently of CGH-based groups. 151
2.3.5 Effect of EPS production on biofilm formation. 153
2.4 DISCUSSION **156**
2.5 ACKNOWLEDGEMENTS **160**
2.6 REFERENCES **161**
2.7 SUPPLEMENTARY MATERIAL **166**

ABSTRACT

Strains of *Lactobacillus salivarius* are increasingly employed as probiotic agents for humans or animals. Despite the diversity of environmental sources from which they have been isolated, the genomic diversity of *L. salivarius* has been poorly characterized, and the implications of this diversity for strain selection have not been examined. To tackle this, we applied comparative genomic hybridization (CGH) and multilocus sequence typing (MLST) to 33 strains derived from humans, animals, or food. The CGH, based on total genome content, including small plasmids, identified 18 major regions of genomic variation, or hot spots for variation. Three major divisions were thus identified, with only a subset of the human isolates constituting an ecologically discernible group. Omission of the small plasmids from the CGH or analysis by MLST provided broadly concordant fine divisions and separated human-derived and animal-derived strains more clearly. The two gene clusters for exopolysaccharide (EPS) biosynthesis corresponded to regions of significant genomic diversity. The CGH-based groupings of these regions did not correlate with levels of production of bound or released EPS. Furthermore, EPS production was significantly modulated by available carbohydrate. In addition to proving difficult to predict from the gene content, EPS production levels correlated inversely with production of biofilms, a trait considered desirable in probiotic commensals. *L. salivarius* displays a high level of genomic diversity, and while selection of *L. salivarius* strains for probiotic use can be informed by CGH or MLST, it also requires pragmatic experimental validation of desired phenotypic traits.

2.1 INTRODUCTION

Lactobacillus spp. are lactic acid bacteria (LAB) that display phylogenetic, phenotypic and ecological heterogeneity which is reflected in their taxonomic diversity (13). Lactobacilli have complex nutritional requirements which are reflected in the diverse, carbon-rich habitats in which they are found (59). Lactobacilli have been studied extensively because of their importance for the production of fermented foods and beverages (59). Some well characterized lactobacilli are generally regarded as safe (GRAS) and in more recent times they have been used as probiotics and vaccine carriers (30). Administration of probiotic cultures benefits the host through a wide variety of mechanisms which are increasingly recognized as being species- and strain-specific (34). Knowledge of the genetic basis for strain diversity in potentially-probiotic species is thus called for. Comparative genomics has emerged as a powerful approach in this era of high-throughput sequencing technologies and it provides a technological platform to identify strain-specific traits (13).

Lactobacillus salivarius (38) is part of the indigenous microbiota of the gastrointestinal tract (GIT) and oral cavity of humans and hamsters (49). This species has also been isolated from human breast milk (41) and from the intestinal tract of swine (10) and chickens (1). There has been a recent increase in the number of studies in which the probiotic utility of diverse *L. salivarius* strains has been explored (44). However, there is no detailed information on the genomic variability of this species, to serve as a reference for identifying strain-specific properties. In this study, we examined the diversity of *L. salivarius* by applying MLST and CGH to a collection of strains. These were derived from a range of ecological niches, and were diverse in plasmid content and phenotypic traits (21, 37-38). The panel included the strain UCC118, whose genome has been sequenced (12) and which has been extensively studied for its probiotic properties (44). MLST is a powerful sequence-based typing method, that has been applied to more than 48 bacterial taxa (2). It utilizes the internal nucleotide sequences of multiple housekeeping genes to infer genetic relatedness of bacterial strains and species. MLST has been applied to industrially relevant LAB strains (15) and to *Lactobacillus* species including *L.*

plantarum (16), *L. casei* (8, 18) and *L. sanfranciscensis* (46). CGH facilitates comparison of unsequenced strains on a genome-wide level, and can enable correlation of phenotypic patterns within a species to genomic content. Horizontal gene transfer (HGT) is often associated with niche adaptation and was detected by CGH studies of other *Lactobacillus* species, which have revealed strain-specific traits including carbohydrate utilization, and bacteriocin and exopolysaccharide (EPS) production (5, 43). One of the aims of the current study was therefore to investigate whether niche adaptation or probiotic potential was evident in the general *L. salivarius* population.

EPS produced by LAB has been used in the dairy industry to improve the texture, viscosity and rheological properties of fermented products (26). EPS is also attributed with health-promoting properties such as cholesterol-lowering, immunomodulation, anti-tumorogenic effects and prebiotic effects (26). Thus *L. salivarius* strain clusters, defined by EPS gene content, were further analyzed for EPS production and surface properties.

2.2 MATERIALS AND METHODS

2.2.1 Bacterial strains and culture conditions.

L. salivarius strains used in this study are listed in Table 2.1. *Lactobacillus rhamnosus* GG (54) was also employed as a positive control for biofilm formation in this study. Strains were routinely cultured at 37^{o}C under micro-aerophilic conditions (5% CO_2) in de Man-Rogosa-Sharpe (MRS) medium (Oxoid Ltd, Basingstoke, Hampshire, UK). Bacterial strains were maintained as frozen stocks at -80^{o}C in 25% glycerol.

Table 2.1. *L. salivarius* strains used in this study.

Strain	Origin	Source or Reference
UCC118	Human ileal-cecal region	(12, 61)
UCC119	Chicken caecum	(51)
AH4231	Human ileal-cecal region	(61)
AH4331	Human ileal-cecal region	(61)
AH43310	Human ileal-cecal region	(61)
AH43324	Human ileal-cecal region	(61)
AH43348	Human ileal-cecal region	(61)
JCM1040	Human intestine	(42)
JCM1042	Human intestine	(42)
JCM1044	Human intestine	(42)
JCM1045	Human intestine	(42)
JCM1046	Swine intestine	(42)
JCM1047	Swine intestine	(42)
JCM1230	Chicken intestine	(42)
CCUG47825	Human blood, 55-year-old female	CCUG
CCUG47826	Human blood, 55-year-old female	CCUG
CCUG45735	Human blood	CCUG
CCUG43299	Human blood	CCUG
CCUG38008	Human gall, 73-year-old male	CCUG
CCUG47171	Human tooth plaque	CCUG
CCUG44481	Bird	CCUG
CCUG27530B	Human abdomen, abscess	CCUG
DSM20554	Human saliva	(49)
DSM20555^{T}	Human saliva	(49)
DSM20492	Human saliva	DSM
NCIMB8816	Italian human saliva	NCIMB
NCIMB8817	Turkey feces	NCIMB
NCIMB8818	St. Ivel cheese	NCIMB
NCIMB702343	Unknown	NCIMB
LMG14476	Cat with myocarditis	LMG
LMG14477	Parakeet with sepsis	LMG
01M14315	Human gall bladder pus	(64)
L21	Human feces	*

CCUG, Culture Collection University Göteborg; DSM, Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH; JCM, Japan Collection of microorganisms; LMG, Laboratorium voor Microbiologie, Universiteit Gent; NCIMB, National Collections of Industrial Food and Marine Bacteria. * Provided by Professor Gerald Tannock Univ. Otago, New Zealand. The type strain is indicated by a superscript T.

2.2.2 Microarray description, labeling, and hybridization.

Genomic DNA (gDNA) isolation was performed as described previously (38). The array-CGH platform was a customized, high definition microarray manufactured by Agilent Technologies, as described by Fang *et al*. 2009 (21). Briefly, the *L. salivarius* array contained 60-mer nucleotide oligonucleotides corresponding to 2,184 genes (including annotated pseudogenes) in the genome of *L. salivarius* UCC118. A maximum of four probes (21 replicates) for each gene were designed eArray (Agilent Technologies) from each open reading frame (smaller genes had fewer probes) and were spaced throughout the coding region. 1,500 Agilent quality control spots were also included on the array. gDNA to be analyzed was fragmented to an average size of 100-600 base pairs (bp) by sonication at 10 Amps for a total of 20 cycles (one cycle equals 30 seconds on, 30 seconds off) in iced water. Fragmentation was confirmed by agarose gel electrophoresis. Test and reference gDNA was labeled using randomly primed fluorescent polymerization reactions. Briefly, 2 μg of template DNA and 3 μg of random nonamers (MWG), were combined in a reaction volume of 41.5 μL, and were denatured at 95°C for 5 min and snap cooled on ice. 1 μL of dNTP solution (5 mM each dATP, dGTP, and dTTP, plus 2 mM dCTP), 5 μL of 10X exo-Klenow buffer, 1 μl of exo-Klenow (3-9 U/μl) and 1.5 μL of the fluorescent nucleotide analogs, Cyanine 3-dCTP or Cyanine 5-dCTP (Amersham Pharmacia Biotech), were added to the labeling mixture containing either test or reference template DNA. Labeling reactions were incubated at 37°C for 120 min, followed by incubation at 75°C for 15 min to stop the reaction. Labeled test and reference target DNA was purified by Qiagen MiniElute Purification kit (QIAGEN, Valencia, CA) according to the manufacturer's instructions, with the following modifications: labeled test, and reference DNA samples were individually purified, and quantified using the NanoDrop ND-1000 UV-Vis spectrophotometer (NanoDrop Technologies. Rockland, DE). The labeled samples were then mixed with the components of an Agilent Oligo CGH/Chip-on chip hybridization kit, according to the manufacturer's instructions with the addition of 0.54 μg of salmon sperm DNA prior to denaturation. Hybridizations were performed in an Agilent hybridization oven (G2545A) at 65°C for 24 h. Each array was washed in a 50 ml conical polypropylene tube, using: 50 ml of wash buffer 1 (Agilent Technologies, CA, USA) for 5 mins at room temperature; 50 mls of wash

Buffer 2 (Agilent); prewarmed to 37°C for 1 min, 50 mls Acetonitrile (Sigma, >99.5% Analytical grade) at room temperature, and then stabilization and drying solution (Agilent) for 30 sec at room temperature. Two dye-swap replicate hybridizations were performed for each strain tested.

2.2.3 Microarray data acquisition and analysis.

Slides were scanned using the Agilent Microarray Scanner System (G2505B) with Agilent scan control software version 7.0 for the 44k microarray at a resolution of 5 μm and Red and Green PMT at 10. Agilent Feature Extraction software version 9.1 was used for feature extraction. Default settings were employed, except that Linear and Lowess based normalization was performed on the data set and background-subtracted signals were extracted for both the red (Cy5) and green (Cy3) channels. Low signal spots were removed, using the software-defined flag IsWellAboveBG and control spots were omitted from further analysis. Microarray data outliers were removed with the Grubbs test (22), and the mean of replicate probe values was calculated. P-values were calculated according to the Cyber-t test (4). Normalized signal ratios were transformed to their base 2 logarithm value, $\log_2$ (T/R). The distribution of the $\log_2$-transformed signal ratios for each hybridization reaction was analyzed separately. The mode of the normal distribution of the ratio values fitting the main peak was calculated. The $\log_2$ ratio value of each averaged replicate probe was then modified in order to shift the main peak of each hybridization reaction to center around zero. The method modified from Chen *et al*. (11), was performed to further reduce the effects of hybridisation-efficiency variation on the dataset.

The genes used for MLST analysis in this study were employed to validate the normalization method used for this dataset. Cutoff values for genes presumed to be "present", "divergent" and "absent" genes in our analysis were chosen based on the BLASTN (3) alignments resulting from a comparison of the probe sequences for each open reading frame on the *L. salivarius* array to the draft genome sequence of the type strain DSM20555^T (http://www.ncbi.nlm.nih.gov/sites/entrez?db=genome&cmd=Retrieve&dopt=Overview&list_uids=6258). The resulting BLASTN Bit scores and E-values were compared to the empirically determined signal ratios, and the following $\log_2$ cutoff values were chosen: highly conserved ≥ -1.5 ≥ conserved ≥-2.4 ≥ divergent ≥ -4.5 ≥

highly divergent ≥ -5.8 ≥ absent. Genes that gave low signal intensity values in comparison to the reference strain UCC118 were considered to contain sufficient sequence divergence so that strong hybridization did not occur under the conditions of stringency used. Genes that gave an amplified signal in the test strain in comparison to the reference strain were considered to be present in additional copies in the test strain. CGH-based results were also validated against a subset of genes, whose presence or absence had previously been determined by Southern blot analysis of all strains tested (data not shown).

Clustering of the CGH data was performed using consensus present and divergent (coupled), and highly divergent and absent (coupled) calls for all CGH experiments were converted to the integers 1 (present) and absent (0). CGHdist (14) software was then used to estimate distance matrices based on the gene content derived from the results of the gene calls. A neighbour-joining tree was then generated using SplitsTree 4.8 software (24). Clustering of the strains based on CGH results was performed by Hierarchical clustering using the complete linkage clustering method, implemented by Genesis software (55).

2.2.4 Multilocus sequence typing (MLST).

The nucleotide sequences of an intragenic region of each of the following genes were used for MLST analysis: *parB, pstB, rpsB, pheS, ftsQ, nrdB* and *rpoA* (Table S2.1). These genes were chosen on the basis of the essential nature of their gene products, their size, chromosomal locations and in some cases, their effective use, or genes of a similar function, in previous *Lactobacillus* MLST studies (5, 8, 19, 50). Primers were designed using Primer 3 software (32). An approximately 900 bp internal fragment of each gene was amplified[#] to allow accurate sequencing of a 700–800 bp fragment within each amplicon using the primers specified in Table S2.2, on both the forward and reverse strands. DNA was sequenced by MWG Biotech (Ebersberg, Germany). 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). Strains with the same ST are considered to be members of a single clone or lineage. A similarity matrix was generated from the ST data using a web version of SplitsTree 4 (http://pubmlst.org/perl/mlstanalyse/mlstanalyse.pl?site=pubmlst). Split

decomposition analysis of the allelic profile data and individual alleles was performed using SplitsTree 4.8 (24). Network-like structures revealed incompatible phylogenetic signals at the *pheS* and *ftsQ* loci. Subsequent phylogenetic analysis was carried out using the in-frame concatenation of the following genes: *parB, rspB, rpoA, pstB,* and *nrdB* for each strain. A 3,862 bp-long concatenated artificial sequence (CAS) was created for each strain by the in-frame concatenation of the sequences of the five gene fragments. MEGA software version 4.8 (57) was used to perform multiple sequence alignments, and phylogenetic tree generation of both the single gene, and the CAS of each strain. Neighbor-joining trees were generated using the Kimura two parameter method (29), for each of the five genes and a supertree of the CAS was generated using the same parameters. The reliability of the groups was evaluated by bootstrap testing with 1000 resamplings. All of the MLST data from this study have been deposited in the *L. salivarius* MLST database (http://pubmlst.org/lsalivarius/).

2.2.5 Exopolysaccharide (EPS) isolation and quantification.

Semi defined media (SDM) was used to assess the capacity of *L. salivarius* to produce EPS when grown on different carbon sources. The SDM was modified from that of Kimmel and Roberts *et al.* (1998), (Table S2.3), and contained either glucose (glu-SDM), galactose (gal-SDM), or sucrose (suc-SDM), as the carbon source (28). 10ml of each of the media were inoculated with each of the test strains which had been pre-cultured in the relevant SDM. Cultures were harvested for EPS extraction following static incubation at 30^{o}C untill they reached early stationary phase. Differentiation of bound EPS (EPS-b) from released EPS (EPS-r) was by the method of Tallon *et al* (56) with minor modifications. EPS-r was precipitated from the culture supernatant of pelleted cells (15,000 *g*, 15min, 4^{o}C) with two volumes of cold (-20^{o}C) ethanol, while EPS-b was first extracted from PBS-washed, pelleted-cells by overnight incubation of the PBS-washed cells at 4^{o}C in 0.05 M EDTA, prior to ethanol precipitation of the resulting culture supernatant. EPS precipitates were centrifuged (6,000 *g*, 30 mins, 4^{o}C), dried briefly at 50^{o}C and re-suspended in double distilled sterile water. EPS samples of *L. salivarius* strains UCC118 and CCUG44481 were subsequently dialyzed (molecular weight cut-off: 6000–8000 Da) against 1L of distilled water for 24hr with three water changes per day. The total amount of carbohydrate in the EPS was determined using the phenol/sulfuric acid

method (20) with glucose as a standard. Results were expressed in µg equivalent of glucose per milliliter of growth medium. The concentration of EPS was determined in triplicate for each strain.

2.2.6 Confocal imaging of biofilms.

Glass coverslips (22 by 22 mm) were surface-sterilized and placed horizontally into a six-well tissue culture plates (Nunc, Roskilde, Denmark). Overnight cultures of *L. salivarius* strains and *L. rhamnosus* GG were harvested and PBS-washed cells were adjusted to an OD_{600nm} of 1.0. 100 µl of each O.D.-adjusted culture was inoculated into 4ml of medium (glu-SDM, suc-SDM or AOAC). Inoculated medium was dispensed (4 ml) into each well so that the coverslip was fully submerged, with independent biological duplicates of each test strain and medium carried out. Uninoculated media were used as a negative control in the described experimental setup. Following 72 hr static incubation at 30 ^{o}C, coverslips were washed three times by immersion and agitation in PBS solution and stained with SYTO 9 (7.5 mM/ml in Ringers solution) for 30 min under light limiting conditions at 4^{o}C. Coverslips were then washed a further three times in PBS solution. Confocal imaging was performed with a multiphoton confocal scanning laser microscope (Zeiss LSM 510, inverted microscope). The objective was an Plan-Apochromat (63X/1.4 oil) (Carl Zeiss Microimaging , Inc, Thornwood NJ). An argon laser with a maximum-emission line at 488 nm was used as the excitation source and a long-pass filter was applied at 505 nm. Z-stack sections were collected at 0.5 µm intervals. 3-D reconstitutions of the biofilms were generated by Zen MicroImaging software (2009). Background subtraction of images was performed using IMARIS software (Bitplane, Zürich, Switzerland) and biofilm thickness was quantified using Comstat2 (23) under the ImageJ shell. Thicknesses of biofilms were expressed as the mean of replicate measurements +/- the standard deviation.

2.2.7 Statistical analysis.

A one-way analysis of variance was performed in order to test the significance of differences in the thickness of biofilms when grown in different media. Differences in measurements were considered significant when $P < 0.05$.

2.2.8 Microarray data accession number.

The microarray data can be found at EMBL-EBI Array Express under accession number E-MEXP-3036.

2.3 RESULTS

2.3.1 Comparative genome hybridization reveals unusually high-level diversity in *L. salivarius*.

Using an array based on the genes annotated in the *L. salivarius* UCC118 genome, we performed CGH on 32 additional strains of diverse origin. For the initial data analysis, the smaller plasmids pSF118-20 and pSF118-44, and the megaplasmid pMP118, were not excluded. A heat map constructed from hybridization signals (Fig. 2.1) clearly illustrates the presence of 18 regions at which genomic diversity is concentrated. These functions involve transposases, bacteriophage genes, CRISPR loci, EPS biosynthesis, and carbohydrate metabolism, which have been recognized as being encoded by hypervariable regions in other lactobacilli (5).These hypervariable regions did not always align with regions of anomalous [G+C] mol% content (Fig. 2.1) suggesting that these regions were not acquired by horizontal gene transfer but may have been inherited from the ancestral *L. salivarius* genome and subsequently lost over time. The putative conjugation region of the megaplasmid pMP118, previously noted as being non-functional, was also highly divergent. Based upon hierarchical clustering, three major divisions (Fig. 2.1. A-C) were distinguished and strains were designated to clusters which occurred at the first major branching point of the dendrogram. One of these contained seven out of eight animal isolates, but no other discrete strain clusters were identified, apart from the grouping of recent local intestinal isolates (Fig. 2.1). Among these, strains AH43310 and AH43324 showed complete conservation of all loci tested except those on the 20 kb plasmid.

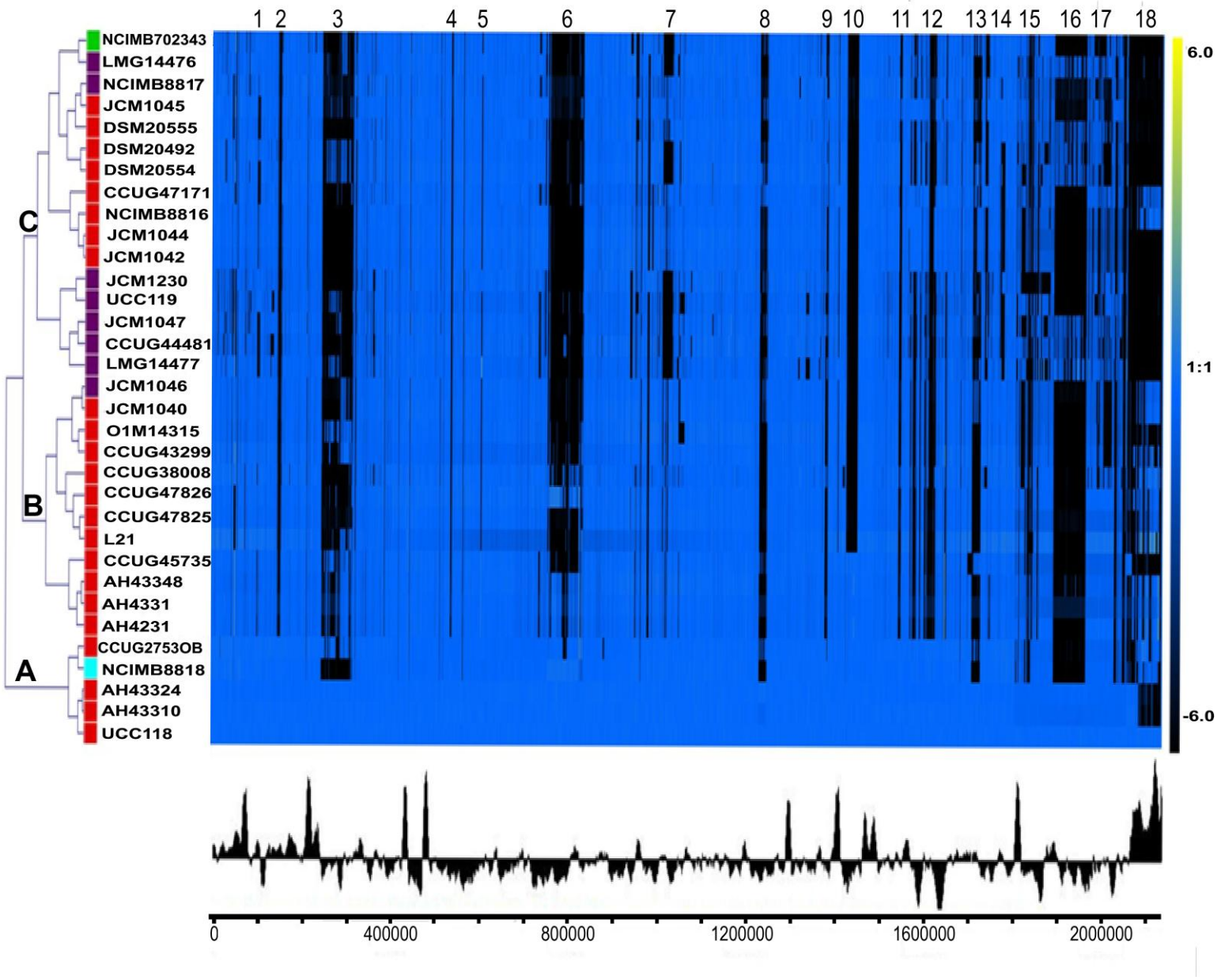

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
NCIMB702343
LMG14476
NCIMB8817
JCM1045
DSM20555
DSM20492
DSM20554
CCUG47171
NCIMB8816
JCM1044
JCM1042
JCM1230
UCC119
JCM1047
CCUG44481
LMG14477
JCM1046
JCM1040
O1M14315
CCUG43299
CCUG38008
CCUG47826
CCUG47825
L21
CCUG45735
AH43348
AH4331
AH4231
CCUG2753OB
NCIMB8818
AH43324
AH43310
UCC118
C
B
A
6.0
1:1
-6.0
0
400000
800000
1200000
1600000
2000000

FIG. 2.1 CGH analysis of 33 *L. salivarius* strains. CGH data is ordered according to the organisation of the UCC118 genome, with replicons ordered left to right: chromosome, pMP118, pSF118-20 and pSF118-44. The colour legend corresponds to the log_2 values of normalized hybridisation signal ratios (test strain/reference strain) to the right of the figure. The gradient goes from black to blue to yellow to depict the absence, conservation, or overrepresentation of a gene in the test strain. The dendrogram shows the relationship of the test strains when compared to the UCC118 genome, using hierarchical clustering of CGH data using a Euclidean distance. GC% of the UCC118 sequence was mapped onto the concatenated replicons under the genomic diversity map. The sources from which the strains originated from are as follows: ■ Unknown, ■ Animal, ■ Human, ■ Food. Numbers 1-18 above the figure indicate hypervariable genomic regions in *L. salivarius*: 1,CRISPR genes; 2, carbohydrate metabolism; 3, Sal2; 4, hypothetical proteins; 5, transposases; 6, Sal1; 7, EPS-cluster 1; 8, Sal4; 9, mucus-binding protein; 10, hypothetical proteins; 11, hypothetical proteins; 12, EPS-cluster 2; 13, Sal3; 14, mannose PTS system; 15, ABC transporter; 16, conjugation region; 17, bacteriocin locus; 18, small plasmids.

The genomic diversity revealed by CGH is summarized in Table 2.2, which shows the conservation level of genes or regions of interest in the test strains relative to UCC118. Conservation in pseudogene numbers varied among the tested isolates, with some strains lacking more than 20 of the pseudogenes identified in UCC118. This may be indicative of genome decay and may be an indication of ongoing adaptation within the *L. salivarius* species. Only one strain, other than the two near-identical AH isolates, harbored a complete bacteriophage identical to that in UCC118. Strain NCIMB8818 harbored Sal1; this strain is a cheese isolate from the UK, and is unlikely to be clonally related to the AH or UCC strains. All other strains substantially lacked Sal1 and Sal2 prophages. Restriction-modification (RM) systems act as a barrier to bacteriophage infection (25); the UCC118 genome includes an unusual shufflon that provides potential for encoding multiple type I systems. Where CGH indicated divergence of any gene within the RM locus, as occurred in 20 out of 32 strains, this divergence was invariably in the gene encoding the specificity-determining substrate. A further 4 strains completely lacked this RM shufflon, including DSM20555; this strain has recently been sequenced by the Human Microbiome Consortium, and may represent a useful transformation recipient.

Table 2.2. Strain-specific characteristics of 33 *L. salivarius* strains as determined by CGH analysis.

		Characteristics[a]														
Origin	**Strains**	**Pseudogenes**		**Prophage[b]**		**RMS[c]**	***bsh1*[d]**	**Abp[e]118**	**2-C-R-S[f]**		**Regulators**		**Mannose PTS**		**MBP [g]**
		HD*	**Abs***	**Sal1**	**Sal2**				**HD***	**Abs***	**CS***	**MP***	**A**	**B**	***lspA***
Human	UCC118	0	0	+	+	+	+	+	0	0	53	7	+	+	+
	AH43310	0	0	+	+	+	+	+	0	0	53	7	+	+	+
	AH43324	0	0	+	+	+	+	+	0	0	53	7	+	+	+
	AH4231	3	12	~	~	~	+	+	0	0	49	4	+	-	+
	AH4331	15	0	~	~	~	+	+	0	0	50	4	+	~	+
	AH43348	8	6	~	~	~	+	+	0	0	50	4	+	~	+
	L21	3	19	~	~	~	+	+	0	0	50	4	+	~	-
	DSM20554	4	18	~	~	~	+	-	0	1	48	4	-	-	-
	DSM20555	6	12	~	-	-	+	~	0	1	49	5	+	~	+
	DSM20492	1	18	~	~	~	+	-	0	1	47	4	-	-	-
	NCIMB8816	2	19	~	~	~	+	~	0	0	48	3	-	-	-
	CCUG47171	5	16	~	~	~	+	~	1	0	47	4	-	-	~
	JCM1040	2	20	~	~	+	+	-	0	1	51	3	+	-	-
	JCM1042	3	18	~	-	~	+	~	1	0	48	3	-	-	-
	JCM1044	3	14	~	-	~	+	~	1	0	48	3	-	-	-
	JCM1045	6	13	~	~	~	+	~	1	0	51	4	+	~	+
	CCUG47825	11	14	~	~	~	+	+	0	0	51	4	+	~	-
	CCUG47826	4	19	~	~	~	+	+	0	0	50	4	+	-	-
	CCUG45735	4	12	~	~	~	+	+	0	0	49	4	+	-	+
	CCUG43299	4	17	~	~	+	+	-	0	1	51	3	+	-	-
	CCUG38008	3	20	~	~	~	+	~	0	0	50	3	+	-	-
	CCUG2753OB	2	9	~	~	+	+	+	0	0	52	4	+	~	+
	O1M14315	3	19	~	~	+	+	-	2	2	51	3	+	-	-
Animal	NCIMB8817	16	8	~	~	~	+	~	1	0	49	5	+	~	+
	JCM1046	5	14	~	~	+	+	~	1	0	51	3	+	~	-
	JCM1047	6	21	~	-	~	+	~	0	0	47	3	-	-	+
	LMG14477	4	15	~	~	~	~	~	0	0	46	3	-	-	+
	LMG14476	4	16	~	~	-	+	~	0	0	47	5	+	-	+
	CCUG44481	7	18	~	~	~	+	~	0	0	47	5	-	-	~
	UCC119	2	27	~	~	-	+	+	0	1	45	4	-	-	~
	JCM1230	5	26	~	-	~	-	~	0	0	45	1	+	-	+
Food	NCIMB8818	2	12	+	~	+	+	+	0	0	52	4	+	-	+
Unknown	NCIMB 702343	4	22	~	~	-	+	~	1	0	49	4	+	-	~

[a]∼, highly divergent and absent locus/region; −, absent locus/region; +, conserved locus/region. Pseudogenes were compared to 64/72 pseudogenes of UCC118 that are represented on the microarray. HD, highly divergent; Abs, absent; CS, genes/loci associated with the chromosome of UCC118; MP, genes/loci associated with pMP118. All calculations excluded genes associated with pSF118-20 and pSF118-44.
[b]Conservation of intact prophage SalI (40,288 bp) and Sal2 (12,529 bp).
[c]Conservation of type 1 restriction modification system (LSL_0915-LSL_0920) encoded by UCC118.
[d]Bile salt hydrolase gene (*bsh1*; LSL_1801).
[e]Region associated with bacteriocin production (Abp 118, LSL_1909-LSL_1920).
[f]Two-component response regulatory systems encoded by UCC118 (encodes 9).
[g]Mucus-binding protein, *lspA* (LSL_0311).

Many of the variable traits (Table 2.2) relate to niche adaptation or survival. Consistent with PCR based screening (21), all *L. salivarius* strains harboured a gene for Bile Salt Hydrolase except for JCM1230 (a chicken gut isolate) and LMG14477 (parakeet isolate). The megaplasmid-encoded structural genes for bacteriocin Abp118 production were highly conserved in 20 of the strains tested. However, of these 20 strains, 8 lacked one or other of the two genes associated with bacteriocin export (LSL_1909 and LSL_1910), and 1 strain showed divergence of the regulator gene of the 2 component regulatory system which governs transcriptional regulation of Abp118. This diversity corroborates the observed lack of bacteriocin production in many strains, despite harbouring many of the associated genes (37).

The ability to sense and respond to environmental cues is an important survival trait for many bacteria. Although the repertoire of two-component systems is relatively conserved across the strain panel, genes for individual transcriptional regulators are very divergent or absent. The presence of genes associated with mannose uptake has been associated with intestinal persistence in *L. johnsonii* (17). Only a minority of strains lacked either of the two mannose utilization systems and the candidate probiotic strains had at least one, or sometimes two mannose utilization loci. However, some strains described as being intestinal in origin lacked genes for mannose utilization. Both *in vitro* (7, 62) and bioinformatic analyses (6) have indicated that residence in the human gut may be promoted by expression of mucin-binding proteins. Mutation of the *lspA* gene of UCC118 significantly reduced adhesion to HT29 cells (62). The *lspA* gene was conserved in all of the AH strains,

but also in almost all of the animal isolates, so its role in intestinal persistence warrants further investigation.

When the CGH data was analyzed by COG assignment (Table S2.4), the widest variation was seen in COG category G for Carbohydrate Transport and Metabolism. Genes resident on the megaplasmid of UCC118 are required to complete the Pentose Phosphate Pathway (PPP) for heterofermentation. CGH data indicated that all strains harboured the chromosomally-encoded PPP-related genes, but the pMP118-located PPP-related genes were variably present, emphasizing the importance of the megaplasmid as a reservoir for contingency metabolism genes.

2.3.2 Concordant CGH clustering and MLST phylogeny of *L. salivarius*.

To further investigate *L. salivarius* strain relatedness, we performed MLST. The resulting phylogeny was compared to CGH-based clustering from which the smaller plasmids were excluded (Fig. 2.2). The MLST-based tree (Fig. 2.2a) was robust, supported by high bootstrap values. Three major clades were evident, one of which (clade C) included five of the animal isolates. Although the topology and primary nodes of the CGH tree were not identical, the fine grouping of strains was broadly concordant, with 15 strains sharing the same grouping pattern in both the MLST and CGH trees, as shown by the numbering (1-6) in Fig. 2.2 and in addition, a further 7 strains show similar grouping patterns in both trees. In the MLST tree, all but one blood isolate (CCUG43299) clusters with AH candidate probiotic isolates of intestinal origin, and this blood isolate also clusters with JCM1040, which is of human intestinal origin. Thus *L. salivarius* strains that are isolated from blood or tissue (e.g. gall bladder, pus) are not genetically distinct based upon the tested comparisons.

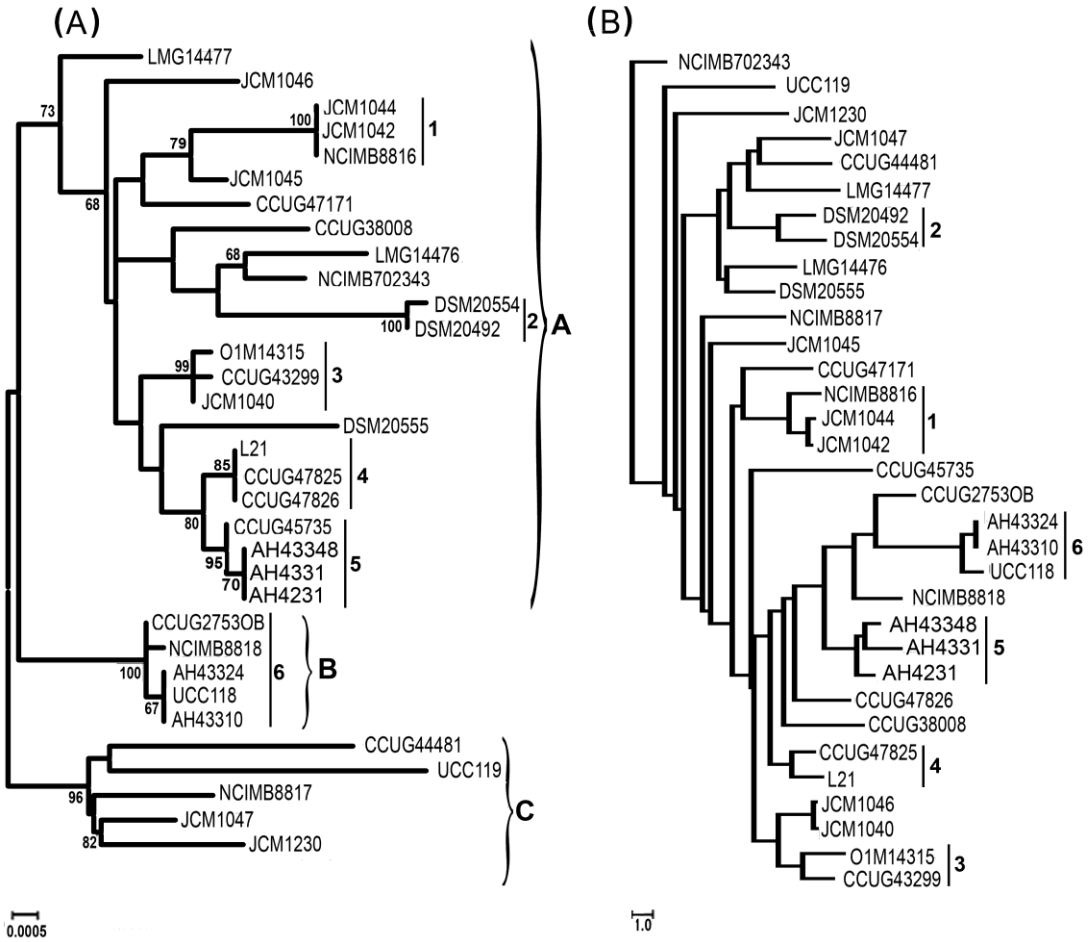


Fig. 2.2 (A) SplitsTree (v.4.8) (24) was used to generate a neighbour-joining tree of maximum likelihood-based distances generated by CGHdist (v.1) (http://cbr.jic.ac.uk/dicks/software/cghdist/index.html, based on CGH). Scale bar represents number of gene differences (present or divergent/absent) per gene site. (B) Supertree generated from the concatenation of 5 MLST gene fragments. The tree was generated using the Kimura two parameter method and neigbour-joining algorithm. Bootstrap values (1000 replicates) over 60% are shown at the nodes. The scale bar represents the number of substitutions per site. See text for explanation of major clades, indicated alpha-numerically.

2.3.3 Separation of *L. salivarius* strains by EPS gene content.

The overall CGH analysis (Fig. 2.1) indicated that the two gene clusters for EPS production in the UCC118 genome were highly divergent among the tested isolates. EPS has a number of biologically significant roles in commensal lactobacilli, including stress resistance, adhesion, and interaction with the immune system (34). The distribution of genes in EPS clusters 1 and 2 was therefore examined in greater detail (Fig. 2.3). Based upon Cluster 1, four groups of strains were distinguished. Group D contained 6 of the 8 animal isolates, and essentially lacked the entire EPS Cluster 1. Groups A and B contained most of the human isolates. The genetic diversity in groups B and C was concentrated in two regions encoding functionally related sets of genes, both involving genes that govern EPS sugar content. Two of the AH candidate probiotic strains in Group A displayed identical gene profiles to UCC118, as did the food isolate NCIMB8818 and the clinical isolate CCUG2753OB. These five strains also showed complete conservation of the EPS gene cluster 2 (Fig. 2.3, panel B). Even though the remaining strains groupings were characterized by divergence or total absence of the central seventeen genes in EPS cluster 2, strain group F corresponded closely to strain group B for EPS cluster 1. This strain group is most proximal to that (group A/E) includes UCC118 and two of the AH strains. The strain group including the animal isolates was most distant. Thus the distribution and relatedness of genes for EPS biosynthesis is rationally related to probiotic potential and/or human intestinal origin.

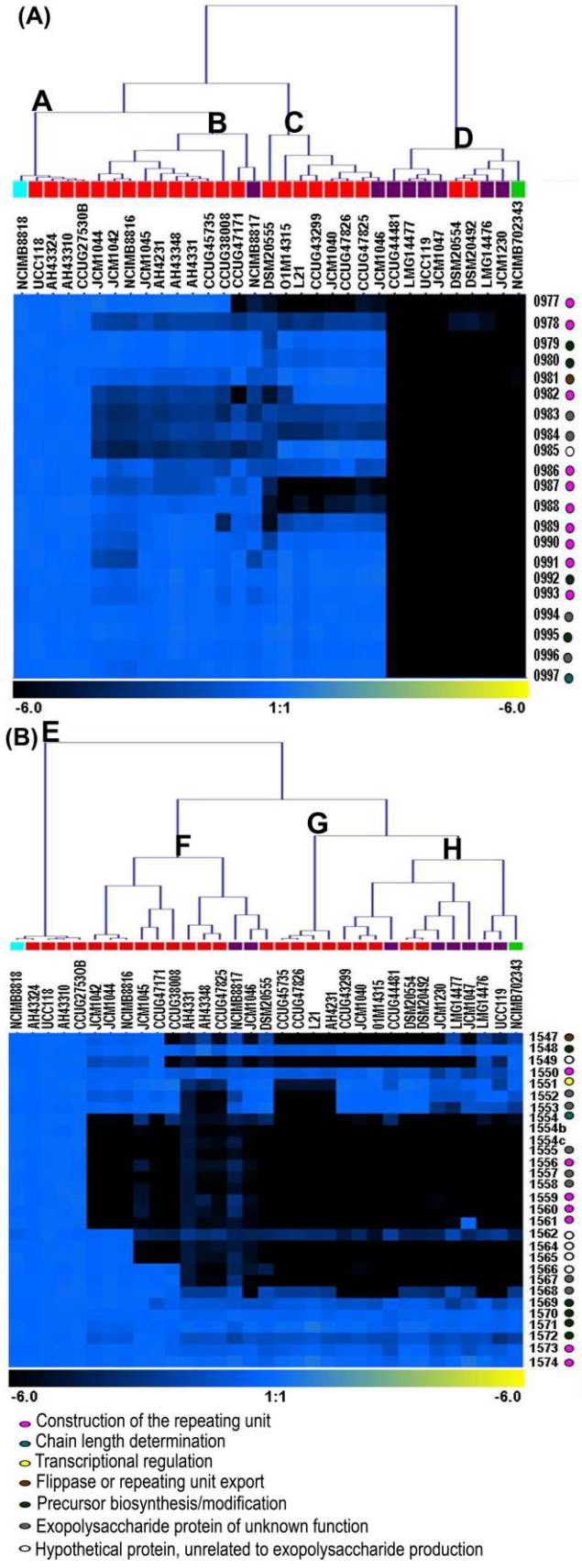


FIG. 2.3 (A) CGH data for the EPS cluster 1 and (B) cluster 2 regions of UCC118 is represented in Fig. 2.3.a and Fig. 2.3.b respectively. A dendrogram

was generated from the EPS cluster 1 and EPS cluster 2 CGH data using hierarchical clustering with an euclidean distance metric. The color legend corresponds to the log_2 values of normalized hybridisation signal ratios (test strain/reference strain) to the right of the figure. The gradient goes from black to blue to yellow to depict the absence, conservation, or overrepresentation of a gene sequence within the test strain. Each four digit number corresponds to the locus identity of the represented gene, which includes the prefix LSL_.

2.3.4 EPS production varies independently of CGH-based groups.

Production of EPS has been well characterized for commensal lactobacilli but not for *L. salivarius*. To search for correlations with CGH-based strain groups, we screened cell-bound and released EPS production levels in the panel of strains. This was performed in the presence of three different sugars, since available carbon source can limit EPS production (26). The EPS levels for all 33 strains are shown in Fig. S2.1; data for eight strains are shown in Fig. 2.4. These strains were selected to cover the major CGH strain-groupings plus additional isolation sources. For all strains, the EPS production level varied depending on the available carbon source, and no single carbon source supported high-level EPS production in all strains. Levels of cell-bound and released EPS varied independently by strain. Perhaps surprisingly since EPS production might be considered a probiotic-related trait, UCC118 and a closely related group A/E strain of food origin, NCIMB8818, both produced relatively low levels of EPS on the sugars tested (Fig. 2.4). Highest levels of bound EPS were produced by a blood isolate (CCUG47826) and a saliva isolate (DSM20492). The bird isolate CCUG44481 produced dramatically higher levels of released EPS than any other strain, at 28 times higher levels than the next highest producing strain, CCUG47826. The released EPS levels for the 33 strains had been determined by the described screening method. Subsequently the CCUG44481 EPS-R was purified by dialysis to remove residual monosaccharide carryover (resulting form the described screening method) from the fermentation media, which may artificially inflate EPS values. Dialysis reduced the measured EPS level by 23% in CCUG44481.

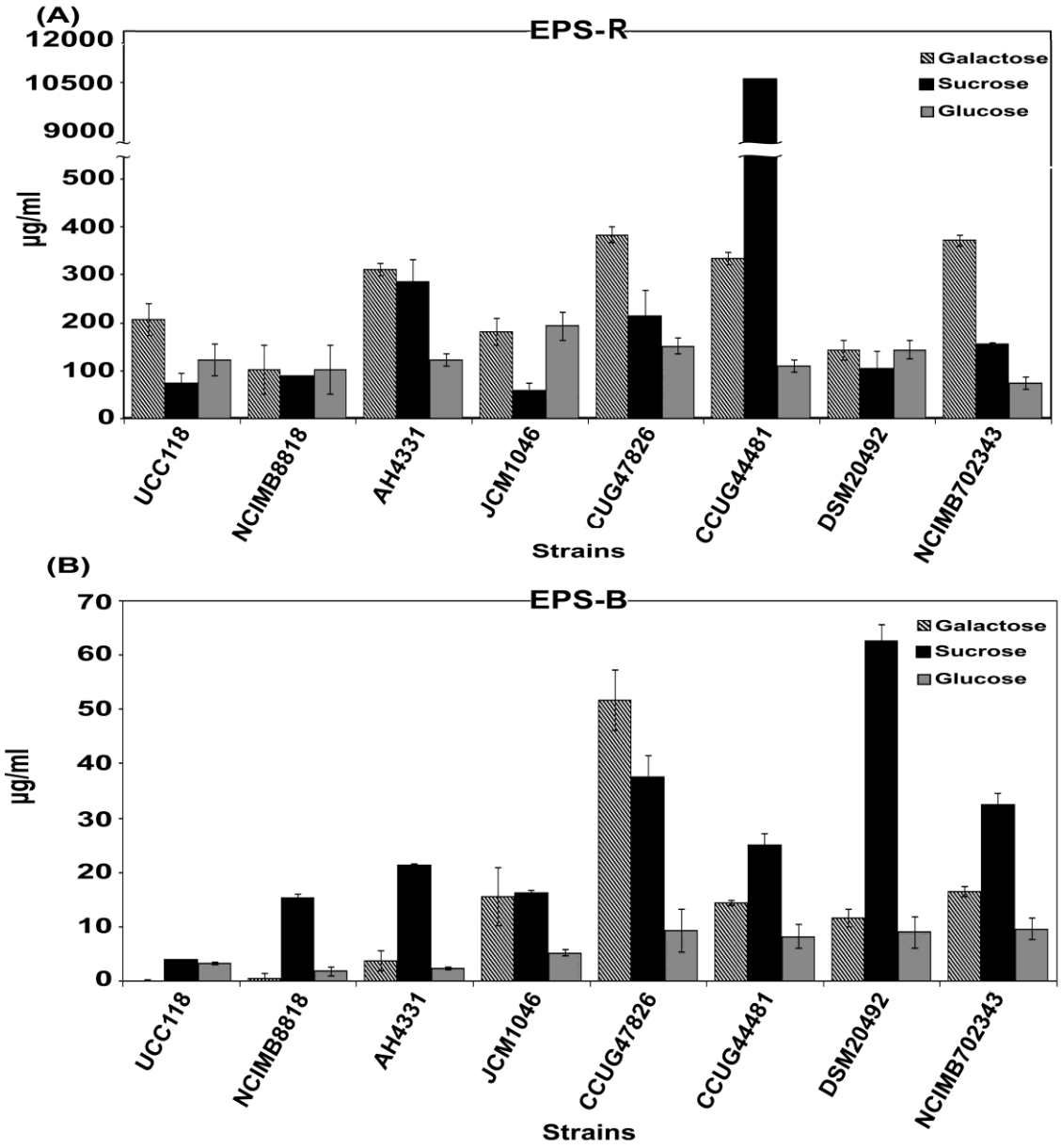


Fig. 2.4. (A) Released exopolysaccharide (EPS-r) and (B) cell-bound exopolysaccharide (EPS-b) production levels of 8 *L. salivarius* strains which are representative of the groups illustrated in Fig. 2.3 and the origins from which the strains were isolated. Strains were cultured in three growth media, gal-SDM, glu-SDM, and suc-SDM. NCIMB8816 did not grow in gal-SDM and therefore analysis of this strain, in this media was not included in Fig. 2.4. All cells were incubated at 30^{o}C until early stationary phase. EPS is expressed in µg / ml culture. Error bars represent the standard deviation values of three replicate experiments.

2.3.5 Effect of EPS production on biofilm formation

It has been suggested that EPS production might confer adhesion ability on commensal bacteria (63). In contrast, it has also been proposed that EPS may have a shielding effect on surface components that are responsible for the adhesive properties of some strains. Indeed, a recent study has demonstrated that an EPS deficient mutant of *L. rhamnosus* GG has increased capacity to produce biofilms (35). We therefore tested biofilm formation in *L. salivarius* strain UCC118 and strain CCUG44481, the highest EPS producer. We benchmarked against *L. rhamnosus* GG (Fig. 2.5, panel A) which has been previously shown to produce monospecies biofilms (36). We tested the three strains in: suc-SDM, a major modifier of EPS production levels in *L. salivarius* (this study); glu-SDM, which contains the carbon source which is regularly used to culture *L. salivarius*; and AOAC, which allowed direct comparison with the published biofilm production data for strain LGG (36). In conjunction to being a strain-dependant trait within the *L. salivarius* species, the capacity to form biofilms was also highly medium-dependant. Of the three strains examined, UCC118 consistently formed the thickest biofilms in the three media tested. A significant reduction (*p*-value =0.001) in biofilm thickness was seen when UCC118 was grown in AOAC medium (4.13 +/- 0.75 μm) in comparison to suc-SDM (8.2 +/- 0.97 μm) or glu-SDM (6.75 +/- 1.2 μm). In contrast to UCC118, CCUG44481 preferentially formed biofilms in AOAC medium (3.8 +/- 0.52 μm) in comparison to either suc-SDM (2.0 +/- 0.7 μm) or glu-SDM (2.25 +/- 0.65 μm) in which it formed thin biofilms containing many voids and hollows. *L. rhamnosus* GG also formed thin biofilms on the glass substrate, forming biofilms of similar depths as CCUG 44481 on both suc-SDM (2.5 +/- 0.7 μm) and glu-SDM (2.0 +/- 0 μm). There was no significant difference in the depths of the biofilms formed by strain LGG in the three media tested, and similarly to CCUG 44481, voids and hollows were present throughout the LGG biofilms. Prominent 3-D structures (Fig. 2.5, panel B) were visible within the architecture of the *L. salivarius* biofilm. These features were evident for both *L. salivarius* strains when the medium used was capable of supporting biofilms of a minimum depth of 3.8 μm. These structures, which resembled flat-topped mushrooms were completely absent from the LGG images. The biofilms formed by strain LGG were more easily dislodged by the washing steps employed, in comparison to *L. salivarius* biofilms. This may indicate that strain

LGG has a reduced affinity for the glass substrate employed in this study, in comparison to the *L. salivarius* strains. EPS production levels of the strains tested could not be quantified accurately when grown in the AOAC medium. However, biofilm formation by the two *L. salivarius* strains was inversely proportional to released EPS levels, when grown on sucrose and glucose.

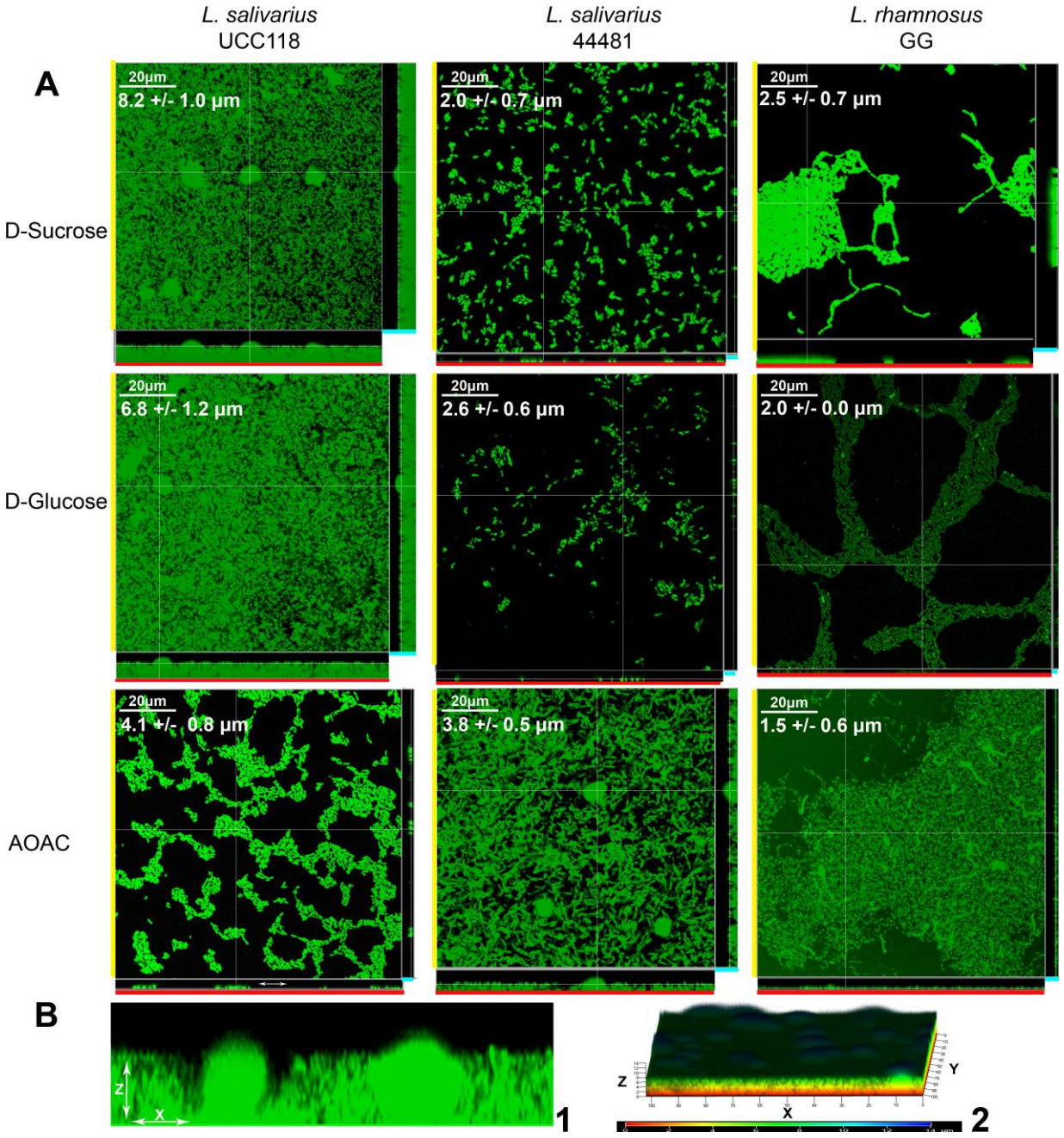


Fig. 2.5. *L. salivarius* and *L. rhamnosus* GG biofilms examined by confocal microscopy. (A) All strains were cultured statically in three media, AOAC, glu-SDM, and suc-SDM for 72 hours at 30°C, indicated at the left hand side of the rows of images. Bacterial cells (green) were stained with SYTO 9. The coloured bars on the edges of each image represent the orientation of the view point; red indicates the X-axis, yellow indicates the Y-axis and turquoise indicates the Z-axis. Representative images from two independent experiments are shown as single and stacked optical sections of each strain in each media at 630x magnification. Biofilm thickness measurements are represented for each condition tested in the top left hand corner of each image. The symbol +/- refers to the standard deviation among replicated measurements. (B. 1) 200x zoomed architecture of a UCC118 biofilm when grown in suc-SDM. (B. 2) A 3-D reconstitution of the surface view and cross-section of a UCC118 biofilm grown in suc-SDM. The color legend corresponds to the depth of the biofilm in µm.

2.4 DISCUSSION

The current study shows that up to 23.6 % of the gene content of *L. salivarius* strains is variable when compared to the UCC118 genome. The most conserved strains were human GIT isolates, while greatest divergence occurred in animal-associated isolates. The extent of diversity in *L. salivarius* is higher than that revealed by CGH in *Lactobacillus plantarum* (up to 20% gene divergence; (43)) and *L. casei* (up to 19% gene divergence; (9)), despite these two species inhabiting a wider range of environments and having considerably larger genomes than *L. salivarius* (31, 40). *L. johnsonii* (47) is primarily a GIT-associated organism and it also displayed a higher level of gene conservation (17% divergence; (5)) than *L. salivarius* using the gene conservation parameters applied in this study. The most conserved genes in *L. salivarius* were those associated with "Information storage and processing", while most divergence was noted in hypothetical proteins, mobile elements and pseudogenes. This high proportion of pseudogenes may indicate that *L. salivarius* is also subject to an ongoing process of genome degradation for niche adaptation, as reported for other LAB (40).

Two *L. salivarius* strains (AH43310 and AH43324) of human intestinal origin had near-identical gene content to UCC118, indicating their potential as probiotic candidates. Also included in group A was one strain isolated from a diseased subject, and a food isolate. The next closest group (Group B) contained the remaining AH strains, and surprisingly, five septicaemia isolates. Some isolates from human infections were also shown by MLST to be closely related to the Group B AH strains (Fig 2.1), and clinical isolates did not form a separate MLST cluster. This suggests that the rare cases of septicaemia caused by *L. salivarius* are due to compromised host barriers or defences, rather than specialized "pathogenic strains", as has been shown for other *Lactobacillus* species (52). We are also currently exploring the possibility that disease isolates harbour additional genes required for enhanced pathogenic potential, which would not be identified in the CGH data.

Group C defined by the CGH analysis included the majority of the animal isolates, five salivary isolates, and three human intestinal isolates. Although the primary CGH separation of group C is due to the small plasmids, this group was also evident from the MLST data, indicating that it reflects the evolutionary history of the whole genome. The proximity of some strains of human origin to animal-derived

strains emphasizes the fact that *L. salivarius* strains of human origin cannot be universally expected to exert probiotic effects in humans.

It is not clear if *L. salivarius* is truly autochthonous in the human GIT (i.e. if it colonizes and forms a self-sustaining population (48)). The anti-infective (51) and bacteriocin producing abilities (37) of *L. salivarius* are strain-specific. Together with the variable presence of genes that encode and regulate bacteriocin production, CGH also revealed considerable divergence in genes for survival in the harsh physical environment of the human and animal GIT. The widespread presence of a bile salt hydrolase gene (LSL_1801) homolog in all but one of the *L. salivarius* strains tested indicates a positive biological selection for this gene function, as described by Fang *et al* 2009 (21). The presence of such a gene in strains from non-intestinal sources suggests that all or most *L. salivarius* strains are adapted for periods of GIT transit, as some part of their long-term ecological life-styles. Reinforcing this idea, the genes associated with mucus-binding proteins in UCC118 were similarly not exclusively found in GIT isolates. A recent study by Oh *et al* found that, with few exceptions *L. reuteri* populations are composed of host-specific ecotypes which have co-evolved with specific vertebrates (45). In contrast, the grouping of *L. salivarius* strains by niche-association was not unequivocally supported by the gene content of the strains as revealed by CGH analysis, although a cleaner separation of e.g. animal strains was provided by MLST. This does not exclude the possibility that additional niche specific genes may be present in *L. salivarius* strains, which are not detectable by a CGH approach.

Despite the lack of grouping of strains based on their overall genomic profiles, most of the animal-associated isolates clustered by hierarchical analysis of EPS cluster 1 genes, but not cluster 2. However, for both clusters 1 and 2, a limited number of human derived strains showed significant conservation of the UCC118 EPS gene complement. Berger *et al.* (5) demonstrated a complex patchwork pattern of EPS genes in *L. johnsonii* strains, with well conserved regions alternating with regions of sequence diversity (5). The pattern of EPS gene diversity in *L. salivarius* is more systematic, with two readily visible blocks of diversity in EPS gene cluster 1 (Fig. 2.3). These genes are predicted by annotation to be responsible for the sugar decoration of the EPS molecule. Their variation, against a backdrop of conservation of the core genes for EPS production, suggests environmental selection or adaptation in EPS sugar content. The pattern of conservation of EPS genes and loci in *L.*

salivarius warrants further investigation from a functional perspective, and partly motivated our phenotypic analysis.

Until a recent report (39), the EPS production level of *L. salivarius* had not been described. *L. salivarius* BRC 14759 (DSM20555 / JCM1231) produced a combined EPS concentration of 45.3 mg/l, in chemically defined medium supplemented with 5g/l lactose at 40^{o}C (39). This is significantly lower than the levels observed for strain DSM20555 (121.48 mg/l) in the current study. This probably reflects the importance of culture conditions used for of EPS production, but may also be a feature of the screening method used in this study for handling a large number of strains. The amount of released EPS produced by CCUG44481 (8.15 g/l) when grown in sucrose was significantly higher than all other strains tested, on any other carbon source. EPS-production levels across multiple strains did not appear to be uniformly related to either source of isolation or the conservation of EPS cluster 1 or EPS cluster 2, but were highly dependent on the culture conditions. The production of EPS in *L. salivarius* strains cannot currently be attributed to either EPS cluster without functional characterization of these regions, informed by chemical analysis of purified EPS. In addition, the EPS-producing phenotype could also be the result of an additional functional EPS-related cluster of genes which are present in a subset of the *L. salivarius* strains but which are absent in UCC118 and therefore absent from the current CGH data.

Adherence and biofilm formation can increase the gut residence time of commensal strains as well as promote pathogen exclusion, host-cell interaction and immune stimulation (53). Biofilm formation has previously been shown to be a strain-specific trait in lactobacilli (36) and biofilm formation in *L. salivarius* was clearly strain dependent and in addition, medium dependent (Fig. 2.5, panel A). It has previously been reported that the *L. salivarius* type-strain JCM1231 (DSM20555 in this study) does not form substantial biofilms (33). It was therefore significant that the probiotic strain UCC118 showed the highest capacity for biofilm formation among the strains tested in this study. In suc-SDM, UCC118 formed biofilms over three times thicker than LGG. CCUG 44481 was generally incapable of forming substantial biofilms. Biofilm formation was inversely related to EPS-R, and the theory that cell surface molecules other than EPS are responsible for UCC118 biofilm formation warrants further investigation. The observation of mushroom-like protrusions from the surface of the *L. salivarius* biofilm mat is the first description of

its kind for a *Lactobacillus* biofilm and thus provides insight into the architecture of biofilms of commensal lactobacilli. The measured depths of the biofilms formed by UCC118 reported here are similar to those of the probiotic strain *L. reuteri* ATCC 55730 (7 +/- 2 μm) (27). It is interesting to note that UCC118 also formed thicker biofilms *in vitro* than those of *L. reuteri* 100-23 (ca. 2.5 μm after 32 hours *in vitro*), although the variation in experimental procedure may have plausibly caused the observed difference (58). This strain has been shown to form biofilms (ca. 20μm) on the epithelial surface of the fore-stomach of ex-*Lactobacillus*-free mice (60). It is possible that the substantial biofilms formed by UCC118 may contribute to the desirable host-interaction properties of this strain.

This study has established a phylogenomic framework for *L. salivarius*, and has shown limited clustering of strains of human intestinal, animal or blood sources. Correlation with complex phenotypes such as EPS production is challenging. We are currently over-laying this genomic framework with additional genome sequences and with data from a number of traits related to host interaction, to facilitate biodiversity-based screening of the relevant genes.

2.5 ACKNOWLEDGEMENTS

This research was supported by Science Foundation Ireland through a Research Frontiers Programme award to PWOT (05/RFP/GEN047), and by a Centre for Science, Engineering and Technology award to the Alimentary Pharmabiotic Centre. ES was supported by "Borsa di mobilità Socrates-Erasmus 2007-08" of the University of Verona. We are grateful to Alimentary Health Ltd., Kinsale Ireland, for providing *L. salivarius* AH strains; to Dr. Aldert Zomer for valuable discussions and advice on microarray processing and data analysis; to Dr. Marlies Mooij for advice and help related to CLSM of biofilms and analysis of biofilm images and to Dr. Michael Mangan for design of the microarray used in this study.

2.6 REFERENCES

1. **Abbas, H., T. Hanan, A. Surakka, J. Apajalahti, and P. E. J. Saris. 2007. Identification of the most abundant *Lactobacillus* species in the crop of 1- and 5-week-old broiler chickens. Appl Environ Microbiol 73:7867-7873.**
2. **Achtman, M. 2008. Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. Annu Rev Microbiol 62:53-70.**
3. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J Mol Biol 215:403-10.**
4. **Baldi, P., and A. D. Long. 2001. A Bayesian framework for the analysis of microarray expression data: regularized T-test and statistical inferences of gene changes. Bioinformatics 17:509-519.**
5. **Berger, B., R. D. Pridmore, C. Barretto, F. Delmas-Julien, K. Schreiber, F. Arigoni, and H. Brussow. 2007. Similarity and differences in the *Lactobacillus acidophilus* group identified by polyphasic analysis and comparative genomics. J. Bacteriol. 189:1311-1321.**
6. **Boekhorst, J., O. Helmer, M. Kleerebezem, and R. J. Siezen. 2006. Comparative analysis of proteins with a mucus-binding domain found exclusively in lactic acid bacteria. Microbiology 152:273-280.**
7. **Buck, B. L., E. Altermann, T. Svingerud, and T. R. Klaenhammer. 2005. Functional analysis of putative adhesion factors in *Lactobacillus acidophilus* NCFM. Appl. Environ. Microbiol 71:8344–8351.**
8. **Cai, H., B. T. Rodriguez, W. Zhang, J. R. Broadbent, and J. L. Steele. 2007. Genotypic and phenotypic characterization of *Lactobacillus casei* strains isolated from different ecological niches suggests frequent recombination and niche specificity. Microbiology 153:2655-2665.**
9. **Cai, H., R. Thompson, M. F. Budinich, J. R. Broadbent, and J. L. Steele. 2009. Genome sequence and comparative genome analysis of *Lactobacillus casei*: insights into their niche-associated evolution. Genome Biol Evol 2009:239-257.**
10. **Casey, P. G., G. D. Casey, G. E. Gardiner, M. Tangney, C. Stanton, R. P. Ross, C. Hill, and G. F. Fitzgerald. 2004. Isolation and characterization of anti-*Salmonella* lactic acid bacteria from the porcine gastrointestinal tract. Lett Appl Microbiol:431-438.**
11. **Chen, H.-I. H., F.-H. Hsu, Y. Jiang, M.-H. Tsai, P.-C. Yang, P. S. Meltzer, E. Y. Chuang, and Y. Chen. 2008. A probe-density-based analysis method for array CGH data: simulation, normalization and centralization. Bioinformatics 24:1749-1756.**
12. **Claesson, M. J., Y. Li, S. Leahy, C. Canchaya, J. P. van Pijkeren, J. Parkhill, S. Flynn, G. C. O'Sullivan, D. Higgins, F. Shanahan, G. F. Fitzgerald, D. van Sinderen, and P. W. O'Toole. 2006. Multireplicon genome architecture of *Lactobacillus salivarius*. Proc. Natl Acad. Sci. USA 103:6718-6723.**
13. **Claesson, M. J., D. van Sinderen, and P. W. O'Toole. 2007. The genus *Lactobacillus* - a genomic basis for understanding its diversity. FEMS Microbiology letters 269:22-28.**
14. **Davey, R., G. Savva, J. Dicks, and I. N. Roberts. 2007. MPP: a microarray-to-phylogeny pipeline for analysis of gene and marker content datasets. Bioinformatics 23:1023-1025.**
15. **de las Rivas, B., A. Marcobal, and R. Munoz. 2004. Allelic diversity and population structure in *Oenococcus oeni* as determined from sequence analysis of housekeeping genes. Appl. Environ. Microbiol. 70:7210-7219.**

16. de las Rivas, B., A. Marcobal, and R. Munoz. 2006. Development of a multilocus sequence typing method for analysis of *Lactobacillus plantarum* strains. Microbiology 152:85-93.
17. Denou, E., R. D. Pridmore, B. Berger, J.-M. Panoff, F. Arigoni, and H. Brussow. 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.
18. Diancourt, L., V. Passet, C. Chervaux, P. Garault, T. Smokvina, and S. Brisse. 2007. Multilocus sequence typing of *Lactobacillus casei* reveals a clonal population structure with low levels of homologous recombination. Appl. Environ. Microbiol. 73:6601-6611.
19. Drevinek, P., A. Baldwin, C. Dowson, and E. Mahenthiralingam. 2008. Diversity of the *parB* and *repA* genes of the *Burkholderia cepacia* complex and their utility for rapid identification of *Burkholderia cenocepacia*. BMC Microbiology 8:44-54.
20. DuBois, M., K. A. Gilles, J. K. Hamilton, P. A. Rebers, and F. Smith. 2002. Colorimetric method for determination of sugars and related substances. Anal Chem 28:350-356.
21. Fang, F., Y. Li, M. Bumann, E. J. Raftis, P. G. Casey, J. C. Cooney, M. A. Walsh, and P. W. O'Toole. 2009. Allelic variation of bile salt hydrolase genes in *Lactobacillus salivarius* does not determine bile resistance levels. J. Bacteriol. 191:5743-5757.
22. Grubbs, F. E. 1969. Procedures for detecting outlying observations in samples. Technometrics 11:1-21.
23. Heydorn, A., A. T. Nielsen, M. Hentzer, C. Sternberg, M. Givskov, B. K. Ersboll, and S. Molin. 2000. Quantification of biofilm structures by the novel computer program COMSTAT. Microbiology 146:2395-2407.
24. Huson, D. H., and D. Bryant. 2006. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23:254-267.
25. Jeltsch, A. 2003. Maintenance of species identity and controlling speciation of bacteria: a new function for restriction/modification systems? Gene 317:13-16.
26. Jolly, L., S. Vincent, P. Duboc, and J.-R. Neeser. 2002. Exploiting exopolysaccharides from lactic acid bacteria. Antonie Van Leeuwenhoek 82:367-374.
27. Jones, S. E., and J. Versalovic. 2009. Probiotic *Lactobacillus reuteri* biofilms produce antimicrobial and anti-inflammatory factors. BMC Microbiol 9.
28. Kimmel, S. A., and R. F. Roberts. 1998. Development of a growth medium suitable for exopolysaccharide production by *Lactobacillus delbrueckii* ssp. *bulgaricus* RR. Int J Food Microbiol 40:87-92.
29. Kimura, M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J Mol Evol:111-120.
30. Klaenhammer, T. R. 2000. Probiotic bacteria: today and tomorrow. J. Nutr. 130:415-416.
31. Kleerebezem, M., J. Boekhorst, R. van Kranenburg, D. Molenaar, O. P. Kuipers, R. Leer, R. Tarchini, S. A. Peters, H. M. Sandbrink, M. W. E. J. Fiers, W. Stiekema, R. M. K. Lankhorst, P. A. Bron, S. M. Hoffer, M. N. N. Groot, R. Kerkhoven, M. de Vries, B. Ursing, W. M. de Vos, and R. J. Siezen.

2003. Complete genome sequence of *Lactobacillus plantarum* WCFS1. Proc Natl Acad Sci U S A 100:1990-1995.

32. Koressaar, T., and M. Remm. 2007. Enhancements and modifications of primer design program Primer3. Bioinformatics 23:1289-1291.

33. Kumada, M., H. Senpuku, M. Motegi, R. Nakao, H. Yonezawa, H. Yamamura, H. Watanabe, and J. Tagami. 2008. Effects of *Enterococcus faecium* on *Streptococcus mutans* biofilm formation using flow cell system. Journal of Oral Biosciences 50:68-76.

34. Lebeer, S., J. Vanderleyden, and S. C. J. De Keersmaecker. 2008. Genes and molecules of lactobacilli supporting probiotic action. Microbiol Mol Biol Rev 72:728-764.

35. Lebeer, S., T. L. A. Verhoeven, G. Francius, G. Schoofs, I. Lambrichts, Y. Dufrene, J. Vanderleyden, and S. C. J. De Keersmaecker. 2009. Identification of a gene cluster for the biosynthesis of a long, galactose-rich exopolysaccharide in *Lactobacillus rhamnosus* GG and functional analysis of the priming glycosyltransferase. Appl Environ Microbiol 75:3554-3563.

36. Lebeer, S., T. L. A. Verhoeven, M. Perea Velez, J. Vanderleyden, and S. C. J. De Keersmaecker. 2007. Impact of environmental and genetic factors on biofilm formation by the probiotic strain *Lactobacillus rhamnosus* GG. Appl Environ Microbiol 73:6768-6775.

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

38. Li, Y., E. Raftis, C. Canchaya, G. F. Fitzgerald, D. van Sinderen, and P. W. O'Toole. 2006. Polyphasic analysis indicates that *Lactobacillus salivarius* subsp. *salivarius* and *Lactobacillus salivarius* subsp. *salicinius* do not merit separate subspecies status. Int J Syst Evol Microbiol:2397-2403.

39. Liu, C.-T., I. T. Hsu, C.-C. Chou, P.-R. Lo, and R.-C. Yu. 2009. Exopolysaccharide production of *Lactobacillus salivarius* BCRC 14759 and *Bifidobacterium bifidum* BCRC 14615. World J Microbiol Biotechnol 25:883-890.

40. Makarova, K., A. Slesarev, Y. Wolf, A. Sorokin, B. Mirkin, E. Koonin, A. Pavlov, N. Pavlova, V. Karamychev, N. Polouchine, V. Shakhova, I. Grigoriev, Y. Lou, D. Rohksar, S. Lucas, K. Huang, D. M. Goodstein, T. Hawkins, V. Plengvidhya, D. Welker, J. Hughes, Y. Goh, A. Benson, K. Baldwin, J. H. Lee, I. Diaz-Muniz, B. Dosti, V. Smeianov, W. Wechter, R. Barabote, G. Lorca, E. Altermann, R. Barrangou, B. Ganesan, Y. Xie, H. Rawsthorne, D. Tamir, C. Parker, F. Breidt, J. Broadbent, R. Hutkins, D. O'Sullivan, J. Steele, G. Unlu, M. Saier, T. Klaenhammer, P. Richardson, S. Kozyavkin, B. Weimer, and D. Mills. 2006. Comparative genomics of the lactic acid bacteria. Proc Natl Acad Sci USA 103:15611-15616.

41. Martín, R., E. Jiménez, M. Olivares, M. L. Marín, L. Fernández, J. Xaus, and J. M. Rodríguez. 2006. *Lactobacillus salivarius* CECT 5713, a potential probiotic strain isolated from infant feces and breast milk of a mother-child pair. Int J Food Microbiol 112:35-43.

42. Mitsuoka, T. 1969. Vergleichende untersuchungen über die *Laktobazillen* aus den faeces von menschen, schweinen und hühnern. Bakteriol 210:32-51.

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

44. Neville, B. A., and P. W. O'Toole. 2010. Probiotic properties of *Lactobacillus salivarius* and closely related *Lactobacillus* species. Future Microbiology 5:759-774.

45. Oh, P. L., A. K. Benson, D. A. Peterson, P. B. Patil, E. N. Moriyama, S. Roos, and J. Walter. 2010. Diversification of the gut symbiont *Lactobacillus reuteri* as a result of host-driven evolution. ISME 4:377-387.

46. Picozzi, C., G. Bonacina, I. Vigentini, and R. Foschino. 2010. Genetic diversity in Italian Lactobacillus sanfranciscensis strains assessed by Multilocus Sequence Typing and Pulsed Field Gel Electrophoresis analyses. Microbiology:mic.0.037341-0.

47. Pridmore, R. D., B. Berger, F. Desiere, D. Vilanova, C. Barretto, A. C. Pittet, M. C. Zwahlen, M. Rouvet, E. Altermann, R. Barrangou, B. Mollet, A. Mercenier, T. Klaenhammer, F. Arigoni, and M. A. Schell. 2004. The genome sequence of the probiotic intestinal bacterium *Lactobacillus johnsonii* NCC 533. Proc Natl Acad Sci U S A 101:2512 - 2517.

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

49. Rogosa, M., R. F. Wiseman, J. A. Mitchell, M. N. Disraely, and A. J. Beaman. 1953. Species differentiation of oral lactobacilli from man including descriptions of *Lactobacillus salivarius* nov spec and *Lactobacillus cellobiosus* nov spec. J Bacteriol 65:681-699.

50. Ruiz-Garbajosa, P., M. J. M. Bonten, D. A. Robinson, J. Top, S. R. Nallapareddy, C. Torres, T. M. Coque, R. Canton, F. Baquero, B. E. Murray, R. del Campo, and R. J. L. Willems. 2006. Multilocus sequence typing scheme for *Enterococcus faecalis* reveals hospital-adapted genetic complexes in a background of high rates of recombination. J Clin Microbiol 44:2220-2228.

51. Ryan, K. A., P. Daly, Y. Li, C. Hooton, and P. W. O'Toole. 2008. Strain-specific inhibition of *Helicobacter pylori* by *Lactobacillus salivarius* and other lactobacilli. J. Antimicrob. Chemother. 61:831-834.

52. Salminen, M., H. Rautelin, S. Tynkkynen, T. Poussa, M. Saxelin, V. Valtonen, and A. Järvinen. 2004. *Lactobacillus* bacteremia, clinical significance, and patient outcome, with special focus on probiotic *L. rhamnosus* GG. Clinical infectious diseases:an official publication of the Infectious Diseases Society of America 38:62-69.

53. Servin, A. L. 2004. Antagonistic activities of lactobacilli and bifidobacteria against microbial pathogens. FEMS Microbiol Rev 28:405-440.

54. Silva, M., N. V. Jacobus, C. Deneke, and S. L. Gorbach. 1987. Antimicrobial substance from a human *Lactobacillus* strain. Antimicrob. Agents Chemother. 31:1231-1233.

55. Sturn, A., J. Quackenbush, and Z. Trajanoski. 2002. Genesis: cluster analysis of microarray data. Bioinformatics 18:207-208.

56. Tallon, R., P. Bressollier, and M. C. Urdaci. 2003. Isolation and characterization of two exopolysaccharides produced by *Lactobacillus plantarum* EP56. Res Microbiol 154:705-712.

57. Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. . Mol Biol Evol 24:1596-1599.

58. Tannock, G. W. 2005. New perceptions of the gut microbiota: implications for future research. Gastroenterol Clin North Am 34:361-382.

59. Tannock, G. W. 2004. A special fondness for lactobacilli. Appl. Environ. Microbiol. 70:3189-3194.

60. Tannock, G. W., S. Ghazally, J. Walter, D. Loach, H. Brooks, G. Cook, M. Surette, C. Simmers, P. Bremer, F. Dal Bello, and C. Hertel. 2005. Ecological behavior of *Lactobacillus reuteri* 100-23 Is affected by mutation of the luxS gene. Appl. Environ. Microbiol. 71:8419-8425.

61. Thornton, G. M. 1996. Probiotic bacteria: selection of *Lactobacillus* and *Bifidobacterium* strains from the healthy human gastrointestinal tract; characterisation of a novel *Lactobacillus*-derived antibacterial protein. Ph.D. University College Cork, Cork, Ireland.

62. van Pijkeren, J. P., C. Canchaya, K. A. Ryan, Y. Li, M. J. Claesson, B. Sheil, L. Steidler, L. O'Mahony, G. F. Fitzgerald, D. van Sinderen, and P. W. O'Toole. 2006. Comparative and functional analysis of sortase-dependent proteins in the predicted secretome of *Lactobacillus salivarius* UCC118. Appl Environ Microbiol 72:4143–4153.

63. Walter, J. 2008. Ecological role of lactobacilli in the gastrointestinal tract: implications for fundamental and biomedical research. Appl. Environ. Microbiol. 74:4985-4996.

64. Woo, W. C. Y., A. M. Y. Fung, S. K. P. Lau, and K.-Y. Yuen. 2002. Identification by 16S rRNA gene sequencing of *Lactobacillus saivarius* bacteremic cholecystitis. J Clin Microbiol 40:265-267.

2.7 SUPPLEMENTARY MATERIAL

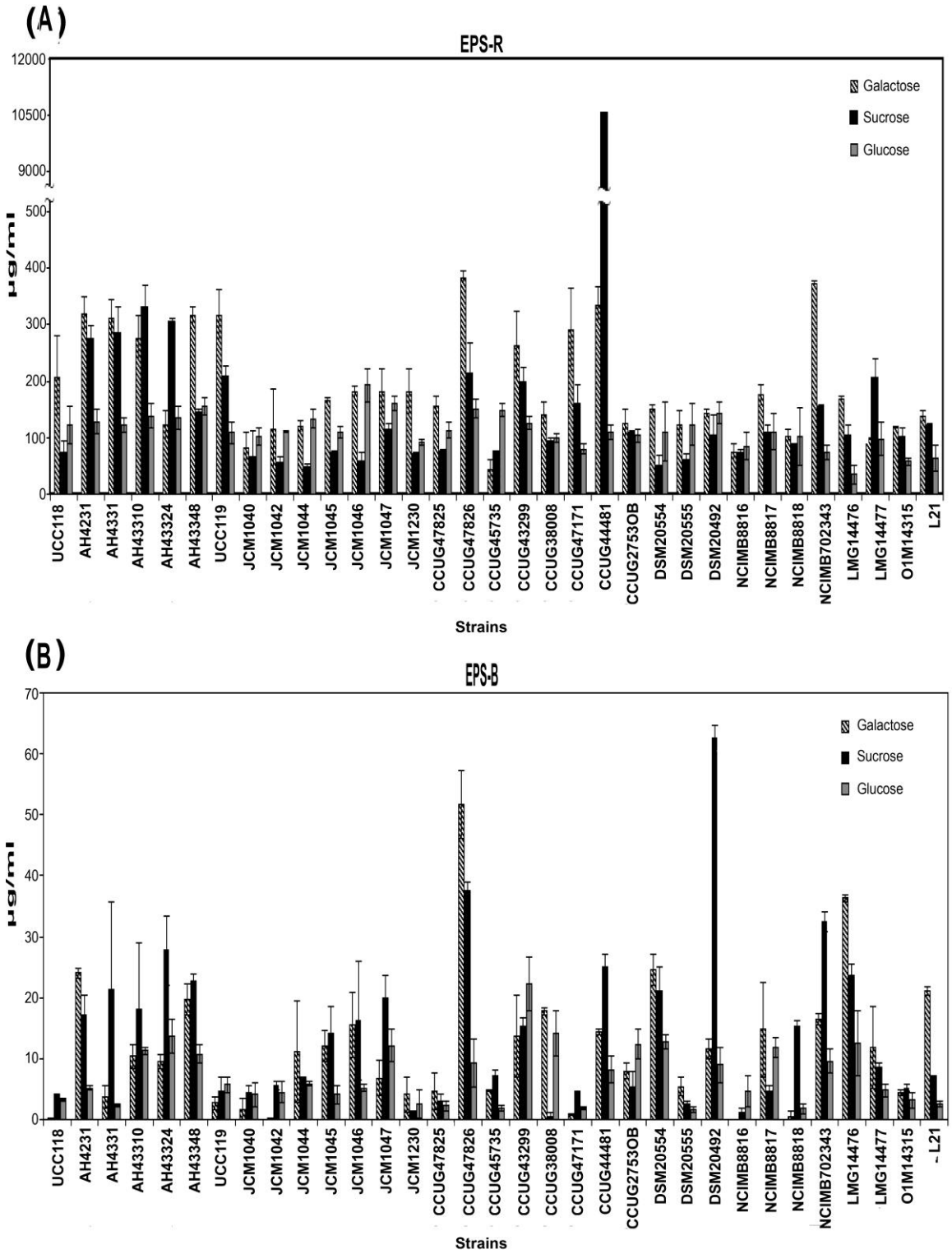
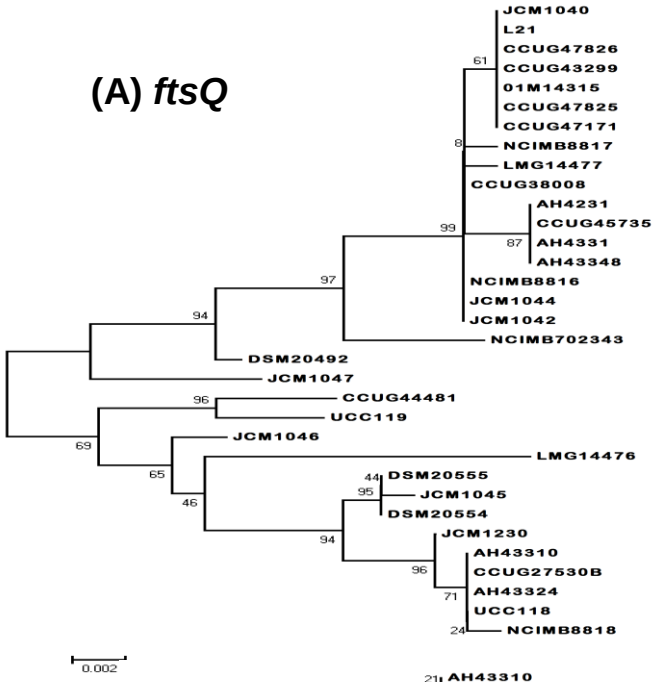
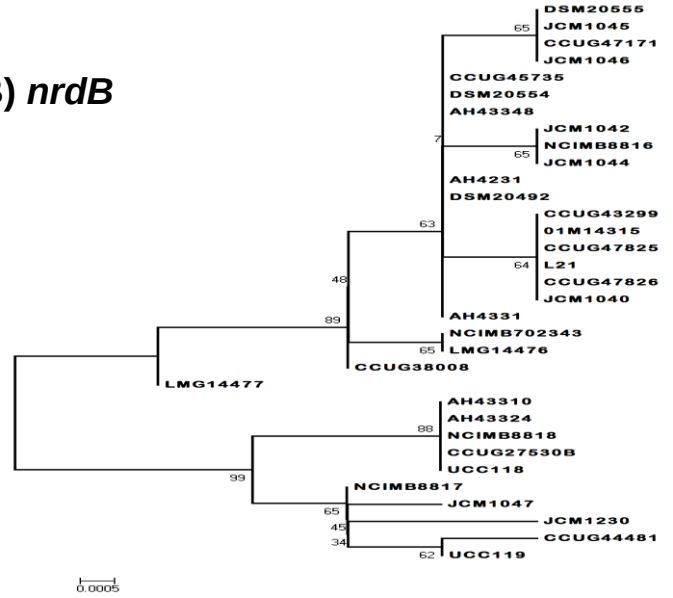
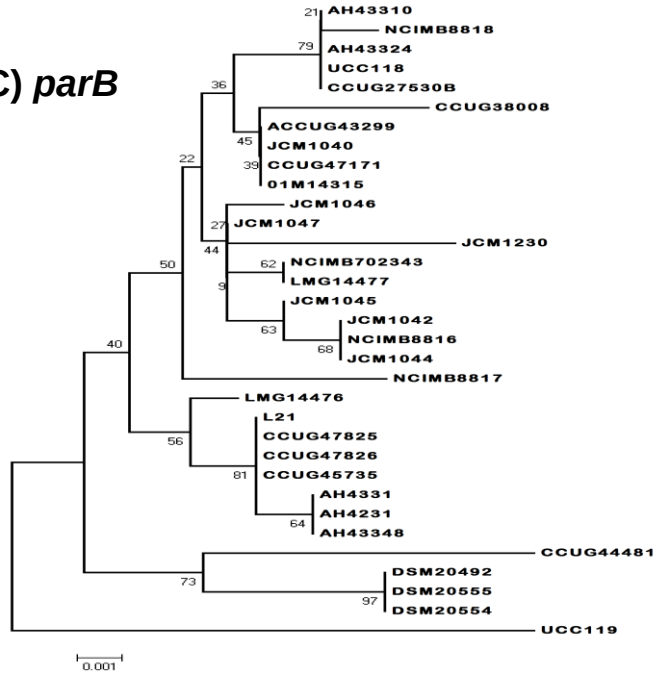


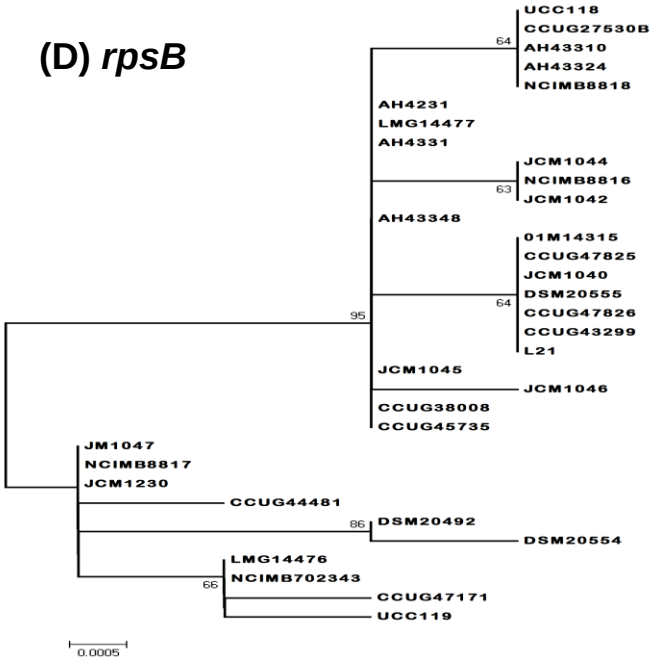
Fig. S2.1 Released exopolysaccharide (EPS-r) and (B) cell-bound exopolysaccharide (EPS-b) production levels of 33 *L. salivarius* strains cultured three growth media, gal-SDM, glu-SDM, and suc-SDM. NCIMB8816 did not grow in gal-SDM and therefore analysis of this strain, in this media was not included in Fig. 2.4. All cultures were incubated at 30^{o}C until early stationary phase. EPS is expressed in µg / ml culture

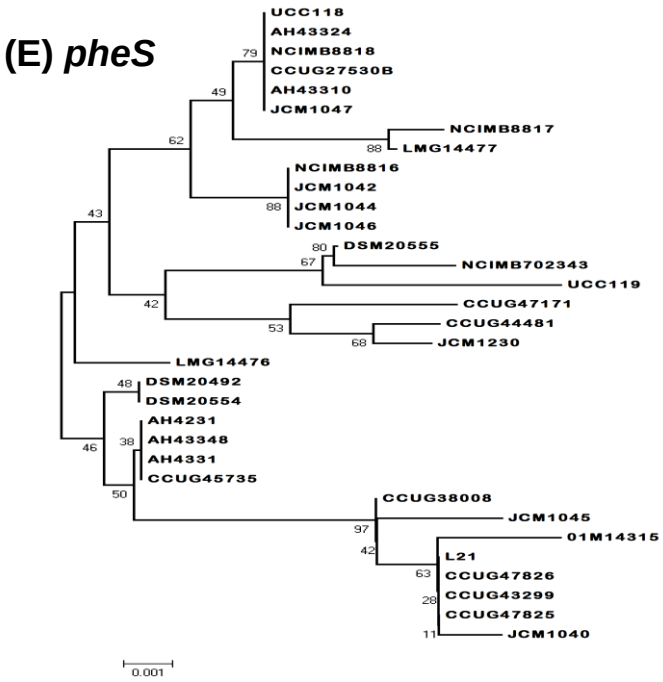
.


(A) ftsQ
JCM1040
L21
CCUG47826
CCUG43299
01M14315
CCUG47825
CCUG47171
NCIMB8817
LMG14477
CCUG38008
AH4231
CCUG45735
AH4331
AH43348
NCIMB8816
JCM1044
JCM1042
NCIMB702343
DSM20492
JCM1047
CCUG44481
UCC119
JCM1046
LMG14476
DSM20555
JCM1045
DSM20554
JCM1230
AH43310
CCUG27530B
AH43324
UCC118
NCIMB8818
0.002


(B) nrdB
DSM20555
JCM1045
CCUG47171
JCM1046
CCUG45735
DSM20554
AH43348
JCM1042
NCIMB8816
JCM1044
AH4231
DSM20492
CCUG43299
01M14315
CCUG47825
L21
CCUG47826
JCM1040
AH4331
NCIMB702343
LMG14476
CCUG38008
LMG14477
AH43310
AH43324
NCIMB8818
CCUG27530B
UCC118
NCIMB8817
JCM1047
JCM1230
CCUG44481
UCC119
0.0005


(C) parB
AH43310
NCIMB8818
AH43324
UCC118
CCUG27530B
CCUG38008
ACCUG43299
JCM1040
CCUG47171
01M14315
JCM1046
JCM1047
JCM1230
NCIMB702343
LMG14477
JCM1045
JCM1042
NCIMB8816
JCM1044
NCIMB8817
LMG14476
L21
CCUG47825
CCUG47826
CCUG45735
AH4331
AH4231
AH43348
CCUG44481
DSM20492
DSM20555
DSM20554
UCC119
0.001


(D) rpsB
UCC118
CCUG27530B
AH43310
AH43324
NCIMB8818
AH4231
LMG14477
AH4331
JCM1044
NCIMB8816
JCM1042
AH43348
01M14315
CCUG47825
JCM1040
DSM20555
CCUG47826
CCUG43299
L21
JCM1045
JCM1046
CCUG38008
CCUG45735
JM1047
NCIMB8817
JCM1230
CCUG44481
DSM20492
DSM20554
LMG14476
NCIMB702343
CCUG47171
UCC119
0.0005


(E) pheS
UCC118
AH43324
NCIMB8818
CCUG27530B
AH43310
JCM1047
NCIMB8817
LMG14477
NCIMB8816
JCM1042
JCM1044
JCM1046
DSM20555
NCIMB702343
UCC119
CCUG47171
CCUG44481
JCM1230
LMG14476
DSM20492
DSM20554
AH4231
AH43348
AH4331
CCUG45735
CCUG38008
JCM1045
01M14315
L21
CCUG47826
CCUG43299
CCUG47825
JCM1040
0.001


(F) pstB
DSM20555
JCM1045
CCUG27530B
AH4331
AH4231
JCM1047
AH43348
JCM1040
AH43310
UCC118
AH43324
CCUG43299
CCUG47826
CCUG45735
L21
CCUG47825
UCC119
01M14315
NCIMB8818
CCUG47171
NCIMB8816
JCM1042
JCM1044
CCUG44481
NCIMB8817
JCM1046
LMG14477
JCM1230
NCIMB702343
LMG14476
DSM20554
DSM20492
CCUG38008
0.001

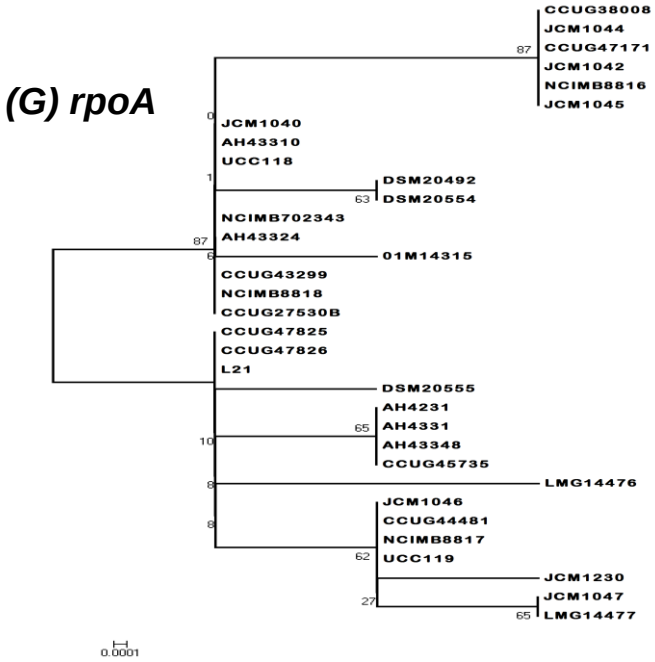

(G) rpoA
CCUG38008
JCM1044
CCUG47171
JCM1042
NCIMB8816
JCM1045
JCM1040
AH43310
UCC118
DSM20492
DSM20554
NCIMB702343
AH43324
01M14315
CCUG43299
NCIMB8818
CCUG27530B
CCUG47825
CCUG47826
L21
DSM20555
AH4231
AH4331
AH43348
CCUG45735
LMG14476
JCM1046
CCUG44481
NCIMB8817
UCC119
JCM1230
JCM1047
LMG14477
0.0001

Fig. S2.2. (A-G) Neighbour joining trees of the individual loci (*parB, pheS, rspB, ftsQ, rpoA, pstB, nrdB*) used for multilocus sequence typing of the *L. salivarius* strains. Trees were generated using the Kimura two parameter method and neigbour-joining algorithm. Bootstrap values (1000, replicates) over 60% are shown at the nodes. The scale bar represents the number of substitutions per site.

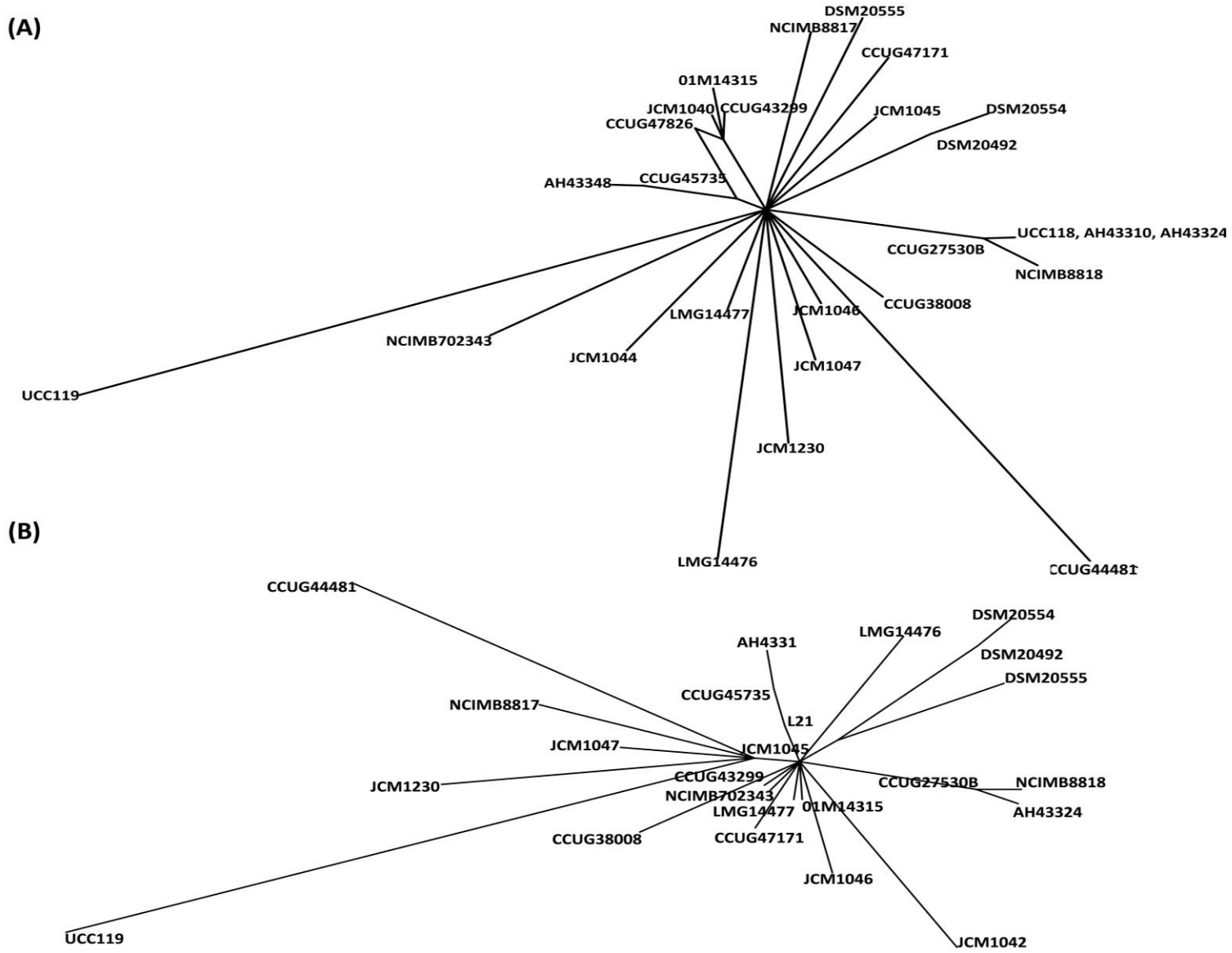

(A)
DSM20555
NCIMB8817
CCUG47171
01M14315
JCM1040
CCUG43299
JCM1045
DSM20554
CCUG47826
DSM20492
AH43348
CCUG45735
UCC118, AH43310, AH43324
CCUG27530B
NCIMB8818
CCUG38008
LMG14477
JCM1046
NCIMB702343
JCM1044
JCM1047
UCC119
JCM1230
LMG14476
CCUG44481
(B)
CCUG44481
DSM20554
LMG14476
AH4331
DSM20492
DSM20555
CCUG45735
NCIMB8817
L21
JCM1047
JCM1045
CCUG43299
CCUG27530B
NCIMB8818
JCM1230
NCIMB702343
LMG14477
01M14315
AH43324
CCUG38008
CCUG47171
JCM1046
UCC119
JCM1042

Fig. S2.3. (A) Split graph of genes (*parB, pheS, rspB, ftsQ, rpoA, pstB, nrdB*). Parallel edges represent phylogenetic incompatibilities in the data set, which are indicative of parallel gene gain/loss by multiple transduction events. (B) Split graph for genes (*parB, rspB, rpoA, pstB, nrdB*). Parallel edges represent phylogenetic incompatibilities in the data set, which are indicative of parallel gene gain/loss by multiple transduction events. All splits were generated using SplitsTree 4.8 (24)

Table S2.1. Locus position, gene identifier and gene function of the loci used for multilocus sequence typing analysis.

Gene	Locus identifier	Function	Genome position (strand)
pstB	LSL_0403	Phosphate transport, ATP-binding protein	450052..450846 (+)
rpsB	LSL_0511	SSU ribosomal protein S2P	557234..558031 (+)
pheS	LSL_0813	Phenylalanyl-tRNA synthetase α chain	829184..830230 (+)
ftsQ	LSL_1049	Cell division protein	1073425..1074279 (-)
nrdB	LSL_1233	Ribonucleoside-diphosphate reductase ß chain	1266599..1267579 (+)
rpoA	LSL_1409	RNA polymerase α chain	1482888..1483832 (-)
parB	LSL_1598	Chromosome partitioning protein	167779..1678637 (-)

Table S2.2. Names and nucleotide sequences of primers used for multilocus sequence typing.

Primer name	Nucleotide sequence
pstB-F	GGCATTAGAAACAAGGGATTTG
pstB-R	AGTTCACCTATCCAAAGTTTCCTG
rpsB-F	ACGTCGTTGGAACCCTAAGA
rpsB- R	CGCCTTCTACAACTTCAACG
pheS- F	GGTCCAATCACTGAAGTTTTACG
pheS- R	TCTGGTCCTAAACCGAATGC
ftsQ- F	GATAAGCATATGCATATGCGCCCACT
ftsQ- R	AAAGCCATGAGAGTAATGAAGAA
nrdB- F	AGGCTACGTGGGAAAAACTT
nrdB- R	CCATTACCAACTTGGGAGAA
rpoA- F	CCAAGATCTGCCAATTTAGCC
rpoA- R	CGCTTGAACGTGGTTATGGT
parB- F	TTCGATTCCATTTTCACTTGC
parB- R	CGATACACAAGAAAAGGTTCAAGA

Table S2.3. Composition of the semi-defined medium used to test for EPS production in *L. salivarius*.

Media components	g/L
*Carbon source (glucose, galactose or sucrose)	30
Tween 80	1
ammonium citrate	2
Sodium acetate	5
$MgSO_4.7H_2O$	0.01
$MnSO_4$	0.05
β-Glycerophosphate	2
Yeast nitrogen base	5
Bacto casitone	10
*FeSO4	0.02
*Uracil	0.01
*Inosine	0.01
*L-cysteine	0.5
*Sterilized separately and added post-autoclaving	
Media was adjusted to pH 6 prior to autoclaving	

Table S2.4. Genomic conservation of *L. salivarius* *

Clusters of Orthologous Genes (COG) category	**H%	**L%	**A%
Information storage and processing			
[J] Translation, ribosomal structure and biogenesis	100	97.9	0
[K] Transcription	100	83.1	9
[L] DNA replication, recombination and repair	100	78.3	13
Cellular processes			
[D] Cell division and chromosome partitioning	100	88	5
[O] Posttranslational modification, protein turnover, chaperones	100	84.9	8
[M] Cell envelope biogenesis, outer membrane	100	71.6	13
[N] Cell motility and secretion	100	87.5	8
[P] Inorganic ion transport and metabolism	100	92.7	1
[T] Signal transduction mechanisms	100	86.3	3
[U] Intracellular trafficking and secretion	100	77.8	13
[V] Defense mechanisms	100	48.4	23
Metabolism			
[C] Energy production and conversion	100	93.5	2
[G] Carbohydrate transport and metabolism	100	76.1	10
[E] Amino acid transport and metabolism	100	89.7	3
[F] Nucleotide transport and metabolism	100	97.1	0
[H] Coenzyme metabolism	100	95.2	1
[I] Lipid metabolism	100	89.7	7
[Q] Secondary metabolites biosynthesis, transport and catabolism	100	75	16
Poorly characterised			
[R] General function prediction only	100	84.3	9
[S] Function unknown	100	84.9	9
[-] Not in COGs	100	45.3	35
Pseudogenes ***	100	53.8	28
% of Present ORFs	100	78.3	87.3

***Calculated for strains tested using highly divergent and absent cutoff values, excluding genes associated with pSF118-20 and pSF118-44**

****Highest percentage conservation (H%), Lowest percentage conservation % (L%), Average percentage absent (A%). Genes that were assigned to multiple COG categories are included in each.**

*****Pseudogenes are included for completeness but are not assigned a COG category.**

Table S2.5. Ranking of *L. salivarius* strains from most to least conserved; calculated using highly divergent and absent cutoff values, excluding genes associated with pSF118-20 and pSF118-44.

Strain	% conservation of genes
UCC118	100
AH43310	100
AH43324	100
AH4331	96.1
CCUG2753OB	94.2
NCIMB8817	93.4
NCIMB8818	90.7
AH43348	90
JCM1045	88.7
CCUG47825	87.9
AH4231	87.6
CCUG45735	86
JCM1046	85.8
CCUG47826	85
DSM20555	83.1
LMG14476	82.9
L21	82.8
DSM20554	82.8
CCUG47171	82.4
DSM20492	81.7
CCUG43299	81.7
O1M14315	81.6
JCM1040	81.3
CCUG38008	81.2
LMG14477	80.8
JCM1044	80.7
JCM1042	80.5
NCIMB8816	80.2
JCM1047	80
NCIMB 702343	79
CCUG44481	77.8
UCC119	77.2
JCM1230	76.4

Table S2.6. Sequence characteristics of the internal gene fragments used for multilocus sequence typing analysis.

Gene	Fragment analysed (nt)	% GC of fragment	% GC of complete gene	Number of	
				Polymorphic sites	Alleles
pstB	731	28.6	33.2	24	16
rpsB	786	37.0	37	14	12
pheS	864	35.2	35.5	32	19
ftsQ	821	31.7	32.0	26	17
nrdB	734	30.7	31.6	19	13
rpoA	816	35.5	35.5	13	10
parB	795	34.2	33.2	42	17

Table S2.7. Distribution of *L. salivarius* strains according to sequence types.

Strain	ST*†	Allele number at each locus						
		ftsQ	nrdB	parB	pheS	pstB	rpoA	rspB
UCC118	ST-1	1	1	1	1	1	1	1
AH43310	ST-1	1	1	1	1	1	1	1
AH43324	ST-1	1	1	1	1	1	1	1
AH4231	ST-2	2	2	2	2	2	2	2
AH4331	ST-2	2	2	2	2	2	2	2
AH43348	ST-2	2	2	2	2	2	2	2
DSM20492	ST-3	3	3	3	3	3	3	3
DSM20555T	ST-4	4	3	3	4	4	4	4
DSM20554T	ST-5	4	2	3	5	3	3	5
NCIMB8816	ST-6	5	4	4	6	5	5	6
JCM1042	ST-6	5	4	4	6	5	5	6
JCM1044	ST-6	5	4	4	6	5	5	6
NCIMB8817	ST-7	6	5	5	7	6	6	7
NCIMB8818	ST-8	7	1	6	1	2	1	1
CCUG27530B	ST-9	1	1	1	1	2	1	1
CCUG38008	ST-10	5	6	7	8	3	5	2
CCUG43299	ST-11	8	7	8	9	7	1	4
CCUG45735	ST-12	2	2	9	2	2	2	2
CCUG47825	ST-13	8	7	9	9	2	7	4
CCUG47826	ST-13	8	7	9	9	2	7	4
L21	ST-13	8	7	9	9	2	7	4
NCIMB702343	ST-14	9	8	10	10	8	1	8
CCUG44481	ST-15	10	9	11	11	9	6	9
CCUG47171	ST-16	8	3	8	12	2	5	10
JCM1040	ST-17	8	7	8	13	2	1	4
JCM1045	ST-18	11	3	12	14	10	5	2
JCM1046	ST-19	12	3	13	6	11	6	11
JCM1047	ST-20	13	10	14	1	12	8	7
JCM1230	ST-21	14	11	15	15	13	9	7
UCC119	ST-22	15	12	16	16	14	6	12
01M14315	ST-23	8	7	8	17	2	10	4
LMG14476	ST-24	16	8	17	18	15	11	8
LMG14477	ST-25	17	13	10	19	16	8	2

* Sequence type

† Removal of genes *ftsQ* and *pheS* did not effect the ST grouping of the strains

A combination of seven alleles defines an allelic profile and each unique allelic profile represents a sequence type.

CHAPTER III

Unusual Genome Complexity
in
Lactobacillus salivarius JCM1046

This chapter has been published as a research article:

Unusual genome complexity in *Lactobacillus salivarius* JCM1046

Emma J. Raftis [1, 2], Brian M. Forde [1, 2,] Marcus J. Claesson [1, 2] and Paul W. O'Toole [1, 2*]

[1] *School of Microbiology University College Cork, Cork, Ireland.*

[2] *Alimentary Pharmabiotic Centre, University College Cork, Cork, Ireland.*

***BMC Genomics* 2014, 15:771. doi:10.1186/1471-2164-15-771**

*Corresponding author.

Mailing address: School of Microbiology, University College Cork, Ireland.

Phone: +353 21 490 3997.

E-mail:pwotoole@ucc.ie

Chapter III Table of Contents

ABSTRACT **182**
3.1 BACKGROUND **183**
3.2 RESULTS AND DISCUSSION **185**
3.2.1 Discovery of circular and linear extrachromosomal elements in *L. salivarius* JCM1046. 185
3.2.2 A conjugative transposon in *L. salivarius* JCM1046. 190
3.2.3 General genome features of *L. salivarius* JCM1046. 191
3.2.4 Phage, transposases and CRISPR regions. 193
3.2.5 Replication of extrachromosomal elements. 193
3.2.6 Plasmid maintenance. 196
3.2.7 Comparative L. salivarius genomics and relationship to phenotype 196
3.2.7.1 Chromosome 196
3.2.7.2 pMP1046A 197
3.2.7.3 pCTN1046 203
3.2.7.4 pMP1046B and pLMP1046 205
3.3 CONCLUSION **206**
3.4 METHODS **207**
3.4.1 Bacterial strains and culture conditions. 207
3.4.2 PFGE plug preparations. 207
3.4.3 S1-nuclease treatment. 207
3.4.4 Restriction of PFGE plugs. 207
3.4.5 Pulsed field gel electrophoresis. 207
3.4.6 Probe preparation and Southern hybridization. 207
3.4.7 Genome sequencing. 207
3.4.8 Genome annotation. 209
3.4.10 Genome comparisons. 209
3.4.11 Identification of novel genetic regions. 209
3.5 COMPETING INTERESTS **210**
3.6 AUTHORS' CONTRIBUTIONS **210**
3.7 ACKNOWLEDGEMENTS **210**
3.8 REFERENCES **211**
3.9 ADDITIONAL FILES **216**
3.10 LIST OF ABBREVIATIONS USED **237**

ABSTRACT

BACKGROUND

Lactobacillus salivarius strains are increasingly being exploited for their probiotic properties in humans and animals. Dissemination of antibiotic resistance genes among species with food or probiotic-association is undesirable and is often mediated by plasmids or integrative and conjugative elements. *L. salivarius* strains typically have multireplicon genomes including circular megaplasmids that encode strain-specific traits for intestinal survival and probiotic activity. Linear plasmids are less common in lactobacilli and show a very limited distribution in *L. salivarius*. Here we present experimental evidence that supports an unusually complex multireplicon genome structure in the porcine isolate *L. salivarius* JCM1046.

RESULTS

JCM1046 harbours a 1.83 Mb chromosome, and four plasmids which constitute 20% of the genome. In addition to the known 219 kb *repA*-type megaplasmid pMP1046A, we identified and experimentally validated the topology of three additional replicons, the circular pMP1046B (129 kb), a linear plasmid pLMP1046 (101 kb) and pCTN1046 (33 kb) harbouring a conjugative transposon. pMP1046B harbours both plasmid-associated replication genes and paralogues of chromosomally encoded housekeeping and information-processing related genes, thus qualifying it as a putative chromid. pLMP1046 shares limited sequence homology or gene synteny with other *L. salivarius* plasmids, and its putative replication-associated protein is homologous to the RepA/E proteins found in the large circular megaplasmids of *L. salivarius.* Plasmid pCTN1046 harbours a single copy of an integrated conjugative transposon (Tn*6224*) which appears to be functionally intact and includes the tetracycline resistance gene *tetM*.

CONCLUSION

Experimental validation of sequence assemblies and plasmid topology resolved the complex genome architecture of *L. salivarius* JCM1046. A high-coverage draft genome sequence would not have elucidated the genome complexity in this strain. Given the expanding use of *L. salivarius* as a probiotic, it is important to determine the genotypic and phenotypic organization of *L. salivarius* strains. The identification of Tn*6224*-like elements in this species has implications for strain selection for probiotic applications.

Keywords: *Lactobacillus salivarius*, megaplasmid, multireplicon, linear plasmid, Tn*6224*, conjugative transposon.

3.1 BACKGROUND

Lactobacillus salivarius [1] is a member of the indigenous microbiota of the oral cavity and the gastrointestinal tract (GIT) of both humans and animals [2-3], and has also been isolated from human breast milk [4]. The probiotic and immunomodulatory activities of *L. salivarius* strains have been recently reviewed [5] and are considered to be strain-specific traits [6]. Strains of *L. salivarius* are genetically diverse [7] and harbour distinctive multireplicon genomes. The first genome of this species to be published [8-9] was that of the well-characterised strain *L. salivarius* UCC118 [1, 10-13] whose megaplasmid pMP118 (242 kb) encodes genes involved in GI tract survival, fitness and probiotic activity [8, 10-11]. *L. salivarius* strains from a range of environmental sources harbour diverse circular megaplasmids [7, 12]. At least 10 additional *L. salivarius* genomes have been sequenced since that of strain UCC118; three of these have been completed (strains CECT 5713 [14] NIAS840 [15] and SMXD51 [16]) with two being finished to a draft quality status [17-18].

Unlike circular plasmids, linear plasmids are rarely observed in lactobacilli [12] but often confer advantageous phenotypes to their hosts [19-20] and have been extensively studied in *Streptomyces* [21-22], *Borrelia* [23] and *Bacillus* [24]. Linear phage genomes are also harboured by strains of *Escherichia coli* [25], *Yersinia enterocolitica* [26], *Klebsiella oxytoca* [27] as well as the probiotic cheese strain *Lactobacillus paracasei* NFBC 338 [28]. Prior to the discovery of linear megaplasmids in *L. salivarius* [12], a 150 kb linear plasmid was identified in *Lactobacillus gasseri* CNRZ222 [29]; but no characterization of the plasmid was performed. We previously identified linear megaplasmids in two porcine *L. salivarius* isolates, JCM1046 and JCM1047, and one human intestinal isolate AH43348 [12].

The conjugative transposon (CTs) Tn916 (18.5 kb) [30] and other Tn916-like elements are highly promiscuous [31], both in the lab and in natural environments [32]. They have demonstrated intra- and interspecies transfer from *L. lactis* [33] and *Lactobacillus paracasei* [34] food strains; and between streptococcal species in dental biofilms [35]. There is a growing concern that commensal bacteria may act as natural reservoirs for antibiotic resistance determinants [36] and may be responsible for transfer of antibiotic resistance to pathogens and opportunistic pathogens [37]. In addition to the introduction of additional functional modules to the host cell, CTs

have further potential to influence natural selection within a bacterial population [38]. There is therefore a growing need to characterize these mobile elements, particularly in species used in food or as probiotics.

Here we present experimental evidence for a highly unusual genome architecture in *L. salivarius* JCM1046, a strain that harbours multiple extrachromosomal replicons of varying sizes and topologies and which has an enhanced ability to withstand the stresses associated with GIT survival [10]. The present study describes an unprecedented level of genome complexity in *L. salivarius*.

3.2 RESULTS AND DISCUSSION

3.2.1 Discovery of circular and linear extrachromosomal elements in *L. salivarius* JCM1046.

Sequencing revealed that *L. salivarius* JCM1046 contains five replicons (Table 3.1): a 1.836 Mb chromosome, two circular megaplasmids of 219 and 129 kb, a linear megaplasmid of 101 kb, and a 33 kb plasmid harbouring an integrated conjugative transposon (Figure 3.1). The complexity of this genome configuration presented extraordinary challenges for genome assembly, described below. Experimental validation of the genome structure is presented in Figure 3.2. *L. salivariu*s strains JCM1047 and AH43348 were known to harbour linear megaplasmids that were presumed to be related to pLMP1046 [12] and were therefore included in these experiments.

Table 3.1. General genome features of *L. salivarius* JCM1046

Feature	Chromosome	pMP1046A	pMP1046B	pLMP1046	pCTN1046
Replicon size (bp)	1,836,297	219,748	129,218	101,883	33,315
GC Content (%)	33.1	32.04	33.87	30.91	34.89
Topology	Circular	Circular	Circular	Linear	Circular
% of genome size	79.1	9.4	5.5	4.3	1.4
Coding genes	1705	214	159	112	40
Coding density (%)	83.3%	80.7%	83.6%	82.6%	76%
rRNA operons	7	0	0	0	0
tRNAs	75	0	2	0	0
Pseudogenes	60	15	2	0	1

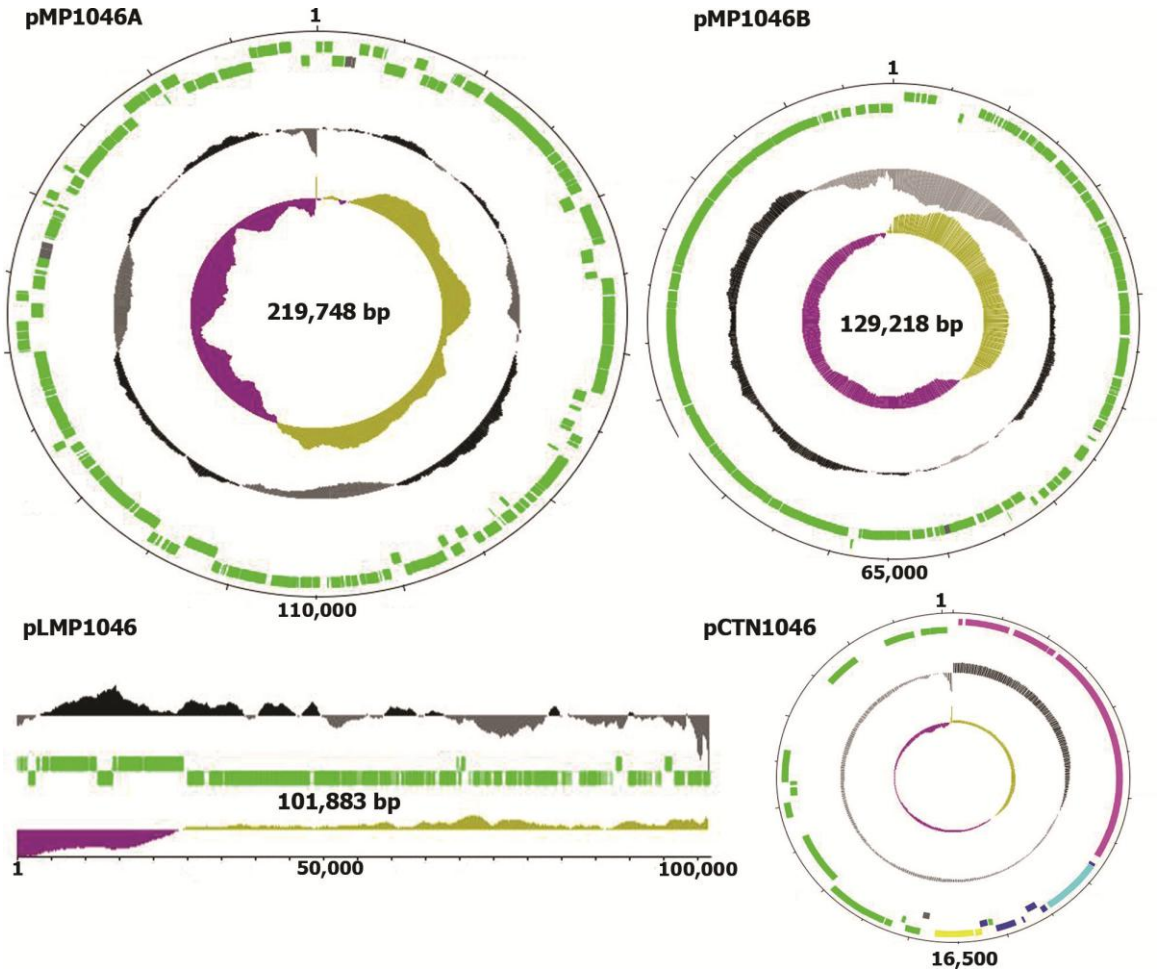


Figure 3.1. Genome atlas of the plasmids of *L. salivarius* JCM1046. A graphical representation of each plasmid in the *L. salivarius* JCM1046 genome was generated using DNAPLOTTER [39]. Genes on the forward and reverse strands (green); pseudogenes (grey blocks); GC % (black above mean and grey below mean); GC skew (mustard above mean and purple below mean) are illustrated for each replicon. Genes encoded by the plasmid backbone of pCTN1046 are also green, the genes present on the integrated conjugative transposon Tn*6224* are represented as follows: conjugative transfer (pink), accessory genes (turquoise), transcriptional regulation (dark blue) and recombination (yellow).

Our original study that identified pMP1046A (then designated pMP1046 [12]) in strain JCM1046 estimated its size as 230 kb, based on Pulsed Field Gel Electrophoresis (PFGE) [12]. However, the assembled sequence data revealed pMP1046A as closer to 220 kb in size. A combination of restriction digestion, PFGE and Southern hybridisation was used to validate the size of pMP1046A. *ApaI* was used to linearise the replicon prior to PFGE and Southern Blot analysis. Probes associated with the replication origin of pMP1046A hybridised to a band that migrated to a constant position between the 194 kb and 242.5 kb linear λ DNA markers, which was in keeping with the expected 219,748 bp size indicated by DNA sequencing.

We identified two novel plasmids pMP1046B and pCTN1046 from the genome sequence. A large contig (~130 kb) was assembled that could not be experimentally determined to form part of either the chromosome or previously described plasmid content of strain JCM1046 [12]. This contig harboured plasmid-associated replication and maintenance proteins. A PCR product off the ends of this contig was generated and subsequently sequenced (data not shown) which proved that the assembled contig was circular in the cell, and it was designated pMP1046B. Under the PFGE conditions that are routinely used to visualise the plasmid content of *L. salivarius* strains, pM1046B had previously gone undetected [8, 40] possibly because it was masked by the linear replicon pLMP1046 [12].

We employed restriction digestion and S1 nuclease treatment in conjunction with PFGE and Southern Blot analysis to confirm the sizes and topologies of the plasmids present in JCM1046. Figure 3.2 panels A and B illustrate the identification of a *repB*-type megaplasmid in JCM1046, panels C and D display the linear plasmids of JCM1046, JCM1047 and AH43348, and panels E and F illustrate the size and topolgy of pCTN1046. Chromosomal DNA bands of strains JCM1046, JCM1047 and AH43328 migrate to the equivalent of the 1 Mb marker (Figure 3.2 panels A, C and E). S1 Nuclease preferentially nicks and linearises megaplasmids due to inherent torsional stresses [41]. The linearised form of the *repA*-type circular megaplasmids of the *L. salivarius* strains are indicated by the open black arrows in Figure 3.2 panels A, C and E.

When an increased band intensity or band width is observed in a PFGE gel, it is often indicative of the presence of linear DNA, high copy number extrachromosomal elements or co-migrating bands of similarly sized DNA

fragments [42]. Strain JCM1046 gDNA revealed high-intensity bands in the S1-treated sample at a position just below the 145.5 kb lambda DNA marker. This band represents the overlapping linear forms of pMP1046B and pLMP1046. In the untreated sample of JCM1046, the circular form of pMP1046B is retained in the well; therefore the *repB* gene probe binds only to the well but not to the migrating linear plasmid pLMP1046 (Figure 3.2 panel B). However, in the S1-nuclease treated gDNA sample of JCM1046, the *repB* probe hybridised strongly to the overlapping pLMP1046/pMP1046B bands (Figure 3.2 panel B), thereby confirming that the discrete replicons pLMP1046 and pMP1046B appear as one overlapping 120 kb band in their linear forms (Figure 3.2 panel B). The *repB* probe did not hybridise to the lanes containing JCM1047 or AH43348 gDNA, indicating that these strains lack a second *repB*-type circular megaplasmid (Figure 3.2 panel B). The presence of a second circular megaplasmid has also been reported in strains NIAS840 and SMXD51, both of these strains being of animal origin [15-16].

Both S1-treated and untreated gDNA samples of JCM1046, JCM1047 and AH43348 show the presence of linear plasmids: pLMP1046 (140 kb), pLMP1047 (140 kb) and pLMP43348 (175 kb) respectively (Figure 3.2, panels A and C). Each of the linear plasmids hybridised to a gene probe derived from the pLMP1046 sequence (Figure 3.2D).

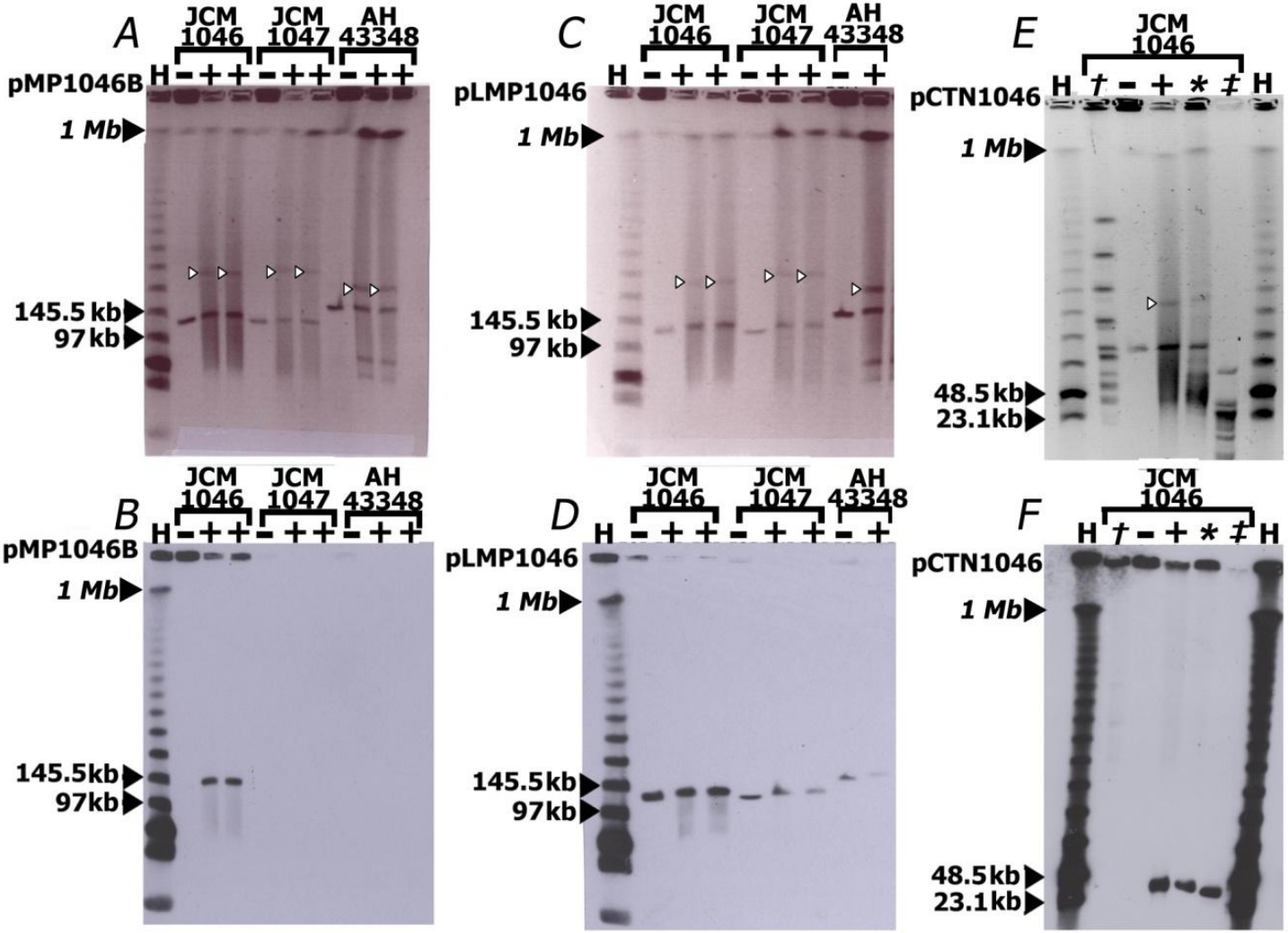

A
JCM 1046
JCM 1047
AH 43348
H - + + - + + - + +
pMP1046B
1 Mb
145.5 kb
97 kb
C
JCM 1046
JCM 1047
AH 43348
H - + + - + + - +
pLMP1046
1 Mb
145.5 kb
97 kb
E
JCM 1046
H † - + * ‡ H
pCTN1046
1 Mb
48.5 kb
23.1kb
B
JCM 1046
JCM 1047
AH 43348
H - + + - + + - + +
pMP1046B
1 Mb
145.5kb
97kb
D
JCM 1046
JCM 1047
AH 43348
H - + + - + + - +
pLMP1046
1 Mb
145.5kb
97kb
F
JCM 1046
H † - + * ‡ H
pCTN1046
1 Mb
48.5kb
23.1kb

Figure 3.2. Confirmation of the genome architecture of *L. salivarius* JCM1046. (A, C and D) PFGE gels of enzyme-treated gDNA of strains JCM1046, JCM1047 and AH43348. Corresponding Southern Hybridizations using replicon-specific probes are shown directly below each gel (B, D,and F). The probes used for the Southern Hybridizations targeted the following genes: the *repB* gene of pMP1046B (B), an endonuclease gene in pLMP1046 (D) and a region spanning the *int-xis* genes of pCTN1046 (F). None of the probes employed showed cross hybridisation with non-target replicons. S1 nuclease (+), SmaI (†), SphI (⁎), PstI (‡) were used individually or in combination to determine the plasmid profiles of each strain. Untreated samples of gDNA are denoted by (−). Closed-black arrowheads indicate λ DNA concatamers used as size standards (H) (A-F). Chromosomal DNA bands of each strain are seen migrating to the equivalent of the 1 Mb marker (A, C and E). Open-black arrows indicate the S1 nuclease-linearised *repA* megaplasmids in each strain examined (A, C and E). A *repB*-type megaplasmid was found to be present in strain JCM1046 but absent from strains JCM1047 and AH43348 (A and B). Both S1-treated and untreated gDNA samples of JCM1046, JCM1047 and AH43348 show the presence of linear plasmids of 140 kb, 140 kb and 175 kb respectively (C), each of which hybridise to a pLMP1046-derived probe (D). S1-nuclease, SphI and PstI were independently used to linearise pCTN1046 (33 kb) (E). A probe based on the *int* and *xis* genes of pCTN1046 binds to the linear form of pCTN1046 (F). pCTN1046 does not have a SmaI site and is retained in the well in its circular form in the SmaI-digested sample.

3.2.2 A conjugative transposon in *L. salivarius* JCM1046.

We further identified a 33 kb plasmid in strain JCM1046 that was not previously observed in the plasmid profile of strain JCM1046 [12, 40] and that was identified here by *de novo* scaffold assembly and designated pCTN1046. It harbours a Tn916-like element and was experimentally determined to have a circular topology. *In silico* analysis was first used to identify restriction enzymes whose use would resolve the chromosomal DNA of JCM1046 from that of pCTN1046. *SphI* and *PstI* each cut the chromosome multiple times, while linearising pCTN1046. Following treatment, pCTN1046 is visible as a band which migrates to a position between the 23.1 kb and 48.5 kb, in keeping with the assembled 33 kb size of pCTN1046 (Figure 3.2E). The chromosome of JCM1046 has multiple *SmaI* restriction sites, while pCTN1046 has none. The multiple DNA bands in the *SmaI*-treated gDNA sample (Figure 3.2E) are chromosomal fragments, while the uncut circular form of pCTN1046 was retained in the well. A probe spanning the *int* and *xis* genes of pCTN1046 hybridised strongly to the 33 kb bands in the S1-nuclease, *SphI* and *PstI* treated samples of JCM1046 (Figure 3.2F). Similarly, the same probe hybridised to the circular form of pCTN1046 retained in the well of the *SmaI*-treated

sample, but did not hybridise to the migrating chromosomal bands (Figure 3.2F). The same pattern of hybridisation was obtained when the experiment was repeated with a probe based on the *tetM* gene harboured by pCTN1046 (data not shown). Although Tn916-like elements have been shown to insert at a single site in some species, in almost all bacterial hosts they insert at multiple sites [43]. Our data indicate that the conjugative transposon in strain JCM1046 is integrated at a single site in pCTN1046 and is absent from the rest of the genome.

3.2.3 General genome features of *L. salivarius* JCM1046.

The unusual genome complexity of JCM1046 raised questions about gene distribution by replicon. Bioinformatic analysis identified 1,705 coding sequences in the chromosome, a coding density of 83.3% (Table 3.1). Biological functions could not be assigned to 360 of these protein coding sequences. The chromosome of *L. salivarius* JCM1046 contains 60 pseudogenes (Additional file 3.1). Seven rRNA operons were identified on the chromosome, as well as 76 tRNA genes for all 20 amino acids. The chromosome has an average GC content of 33.1%, with three regions displaying atypical GC content relative to the rest of the genome (see below).

The largest of the plasmids pMP1046A has a coding density of 80.7%. 214 coding sequences were identified, 79 of which were for hypothetical proteins. pMP1046A contains 15 pseudogenes (Additional file 3.1). The gene content of pMP1046A will be discussed in detail below.

We identified 159 coding regions in pMP1046B, though biological function could only be assigned to 29.7%, the vast majority (110/158) of genes remaining cryptic. The GC% content of pMP1046B (33.87%) correlates well with the 33.1% GC content of the JCM1046 chromosome (Table 3.1) suggesting long-term adaptation to the host cell, or acquisition from a bacterium with a similar % GC content. In addition to harbouring plasmid-associated replication machinery, pMP1046B harbours additional housekeeping and information-related genes, thus fulfilling the criteria for extrachromosomal elements known as chromids [44]. pMP1046B encodes two tRNA genes, tRNA (Gln) (LSJ_3064) and tRNA (Ser) (LSJ_3066) but these genes are not uniquely present on pMP1046B i.e. they are paralogs of chromosomally encoded genes. Gene duplication can offer a level of genomic redundancy to a strain that is adapting to a new environment [45], and the

tRNA genes encoded by pMP1046B may enable JCM1046 to respond more rapidly to changing environmental conditions .

pLMP1046 harbours 112 coding sequences, none of which were pseudogenes. However, 85 of the predicted coding sequence products were annotated as hypothetical proteins, some of which may represent remnants of functional genes. The average GC content of pLMP1046 (30.9 %) is significantly lower than that of the JCM1046 chromosome (33.1%), implying these replicons experienced distinct evolutionary histories and that pLMP1046 may be a recent acquisition.

PFGE analysis predicted the size of pLMP1046 to be approximately 130 kb (this study), but sequencing revealed a replicon that was 102 kb. It is reasonable to assume that this discrepancy and the lack of identifiable terminal inverted repeats (TIR) (discussed below) is an assembly artifact due to omission of the presumptive repeat sequences in the terminal regions of pLMP1046. The problems faced in the sequencing of the telomeres of linear elements are well recognised [46]

In keeping with the guidelines outlined by Roberts *et al* [47] the novel conjugative transposon contained within pCTN1046 was designated Tn*6224*. *In silico* analysis predicted a coding density of 76% for pCTN1046. Thirty-nine coding sequences were identified (Table 3.1), the majority of which (21/39) are linked to the integrated transposon. The sole pseudogene harboured by this replicon lies outside the Tn*6224* region and shows similarity to nitroreductase family proteins. The plasmid backbone of pCTN1046 has an average GC content of 30.8%, whereas Tn6224 has an average GC content of 38.6%. This suggests that Tn*6224* was most likely acquired *via* horizontal gene transfer (HGT). Insertion of Tn916-like elements is not random, with the insertion sites differing from species to species [38], but generally displaying a distinct preference for target sites which are A-T rich and that have a limited homology with the ends of the element [43]. As only one copy of Tn*6224* was found in the genome of JCM1046, a putative consensus of the target sequence in *L. salivarius* could not be determined. Accounting for the potential presence of coupling sequences, the 35 bp that flanked either end of Tn*6224* were examined to determine if the target sites in *L. salivarius* are in keeping with those generally described for these elements [38]. The AT content of the sequences upstream and downstream of Tn*6224* were found to be 97.1% and 85.7%

respectively, indicating that the target site for Tn*6224* is likely to be similar to those of other species [38]

3.2.4 Phage, transposases and CRISPR regions.

PHAST [48] identified two regions of bacteriophage-related DNA in the genome of JCM1046, both found on the chromosome of JCM1046. In addition to a 22.6 kb remnant prophage that spans residues 1378015- 1400296 bp, an intact 28,541 kb prophage was also identified on the chromosome which spans residues 1439831-1444300 bp. At 43.7%, the remnant prophage is one of the three regions of atypical GC content.

102 transposases (including 22 pseudogenes), representing eight IS families were found distributed across four of the five replicons of strain JCM1046. The distribution of transposases is detailed in Additional file 3.2.

Clusters of regularly interspersed short palindromic repeats (CRISPRs) and CRISPR-associated genes (*cas* genes) provide the host with acquired and heritable resistance against genetic transformation, phage and plasmid proliferation [49]. One CRISPR associated system (cas) was identified on the chromosome of JCM1046 at position 810173- 812140 bp, consisting of a 1059 bp repeat locus composed of a 36 bp direct repeat and 26 spacers. This CRISPR region is immediately upstream of the gene encoding Cas2 and immediately downstream of eight additional CRISPR-associated protein coding genes.

3.2.5 Replication of extrachromosomal elements.

The replication region of pMP1046A extends from LSJ_2000 to LSJ_2006 (6449 bp). The gene content and organisation of the replication region of pMP1046A (98% nt ID) is highly similar to that of pMP118 [8] and to those of other sequenced *L. salivarius* strains (Figure 3.3). pMP1046A is likely to replicate by theta-form replication [50].

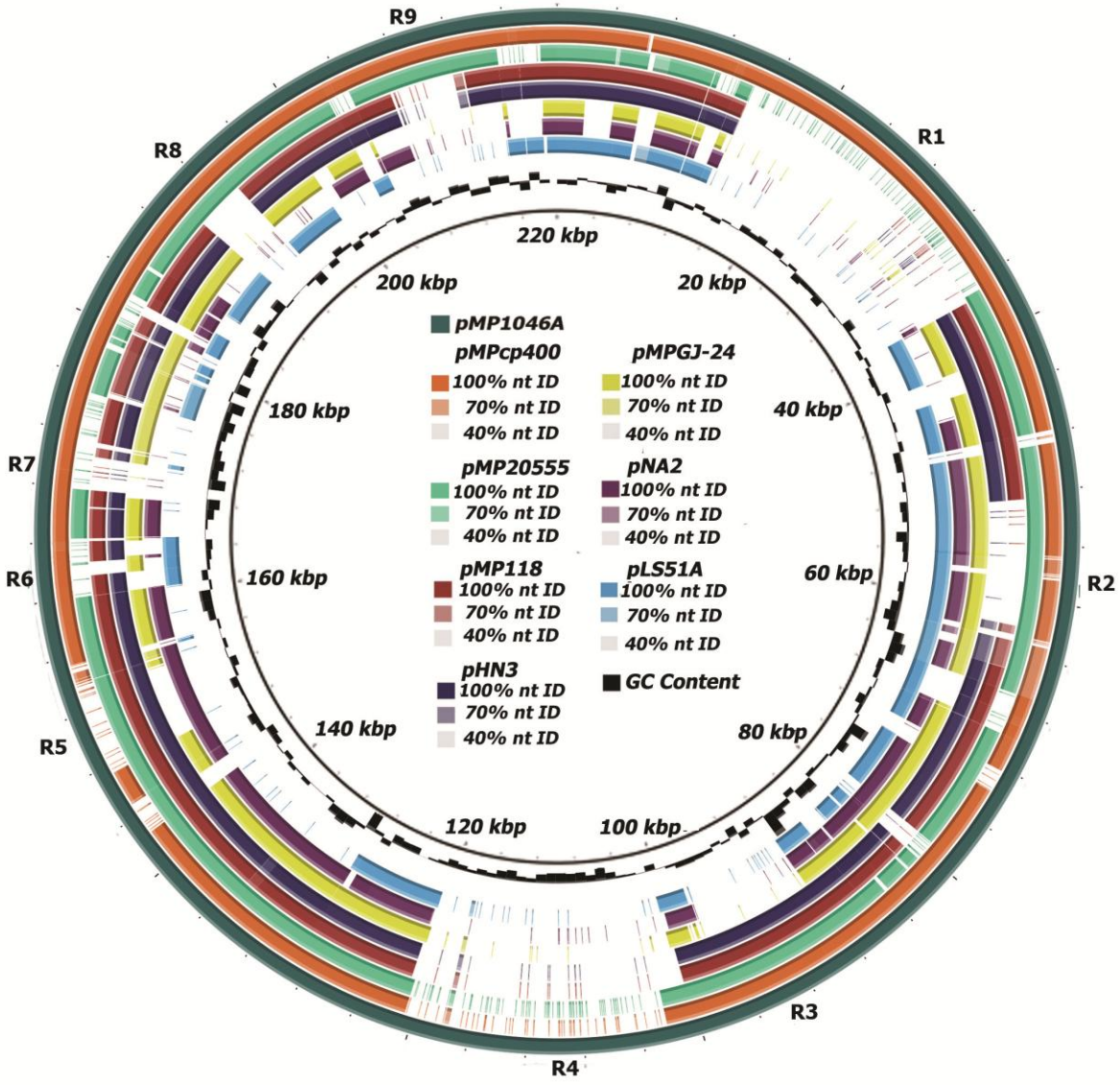


Figure 3.3. A comparison eight *repA*-type megaplasmids of *L. salivarius.* A BLAST atlas diagram of eight *repA*-type megaplasmids of *L. salivarius* was generated using BLAST Ring Image Generator (BRIG) [51], using pMP1046A as the reference replicon (the outer dark green ring). Working inwards from pMP1046A, the next seven rings represent query *repA*-type plasmids of *L. salivarius* strains: cp400, pMP20555, pMP118, pHN3, pMPGJ-24, pNA2, pLS51A. When the completed or circularised version of the *repA*-type megaplasmid was not available (*L. salivarius* cp400 [18] and *L. salivarius* DSM20555), all available sequence data for each strain was mapped to pMP1046A. Regions of diversity between the *repA*-type megaplasmids are indicated by the labels R1-R9. The GC % of pMP1046A was projected onto the mapped plasmid sequences (black ring) and sits outside the molecular clock (innermost ring) surrounding the figure legend at the centre of the figure.

The predicted replication region of pMP1046B spans residues 128175-1974 bp of the plasmid. This region includes a *repA* gene (LSJ_3160) at the position of a switch in GC skew that is characteristic of replication origins [52]. LSJ_3160 shares

36-56% aa identity (ID) with *L. salivarius* RepA protein sequences. The RepA protein of pMP1046B also displays 40% aa ID to the RepA protein of the pig isolate *Lactobacillus reuteri* ATCC 53608 [53]. The second gene in the pMP1046B *ori* region, LSJ_3000 encodes a predicted partitioning/copy control protein, RepB.

Analysis of pLMP1046 indicates that it shares limited sequence homology or gene synteny with linear replicons of other species. However, given the lack of sequenced counterparts in other lactobacilli, the absence of homologous genes in databases is unsurprising. Replication is commonly initiated from one or more internal *ori* sites in linear plasmids and proceeds bidirectionally towards the telomeres [54-56]. Our previous study indicated that the linear plasmids of *L. salivarius* did not harbour the *repA* and *repE* genes encoded by the circular *repA*-type megaplasmids of *L. salivarius* [12], and thus it was presumed that pLMP1046 utilised an alternate mode of replication to the circular plasmids of *L. salivarius* [12]. Sequence analysis identified two plasmid-associated replication genes encoded by pLMP1046, LSJ_4017 (nt 25084-26103) and LSJ_4096 (nt 89781-91007). LSJ_4017 exhibits 39-41% aa ID with proteins annotated as either RepE or RepA in the circular megaplasmids of *L. salivarius* plasmids. This level of sequence homology was not high enough to cause cross hybridisation between the replication genes of pMP118 and the *repA/E* gene identified in pLMP1046, thus accounting for the observations of our previous study [12]. LSJ_4096 encodes a putative replication initiator protein. The replication origins of *Streptomyces* LPs are comprised of helicase-like *rep* genes and interons [22], while the replication *ori* of N15 is located within the r*epA* gene, which acts as a multifunctional protein combining primase, helicase and origin-binding activities [57]. RepA boxes were not identified in the proximity of either the *repA* or *repB* genes of pLMP1046; however, the genomic region immediately upstream of the *repA* coincides with a switch in GC skew. This suggests that the *repA* gene lies within the putative *ori* region of pLMP1046.

The mechanism that pLMP1046 uses to prevent the progressive shortening of their telomeres after each cycle of replication is unknown. It is possible it employs a circular mode, as in some *Streptomyces* linear plasmids [58], but it is more plausible that the sequence of pLMP1046 is missing sections of its terminal regions due to a sequencing or assembly artefact. Further analysis of the terminal regions of pLMP1046 will be required to fully elucidate the mechanism involved in the replication of *L. salivarius* linear plasmids.

There are two replication associated genes harboured by the plasmid backbone of pCTN1046 which are separated by approximately 6 kb. LSJ_5030c shares 52% aa ID with a replication-associated protein in *Lactobacillus amylovorus* GRL 1112. LSJ_5035c encodes the plasmid associated replication protein, RepB, the gene for which coincides with the position of a switch in GC skew, and is therefore the presumed to be the replication origin of pCTN1046. LSJ_5035c shares 36% aa ID with the RepB protein of *L. lactis* subsp. *cremoris* TIFN1 and 100% aa ID to a replication initiation protein in the 30.6 kb plasmid pLS51C in *L. salivarius* SMXD51.

3.2.6 Plasmid maintenance.

Several of the JCM1046 plasmids encode genes implicated in plasmid incompatibility. Three of the plasmids (pMP1046B, pLMP1046 and pCTN1046) encode a *repB*-like gene, two (pMP1046A and pMP1046B) encode a *repE*-like genes and two (pMP1046A and pLMP1046) encode *repA*-like genes. However, the presumptive replication regions of the co-resident plasmids display low levels of sequence ID with the highest nt ID shared between the *repB* genes of pLMP1046 and pCTN1046 at 58.7%. The mosaic nature of the replication regions as well as the lack of nucleotide homology between the respective replication associated genes of the co-resident plasmids is a plausible explanation for the compatibility of the plasmids that co-exist in strain JCM1046. Several complete Toxin-Antitoxin (TA) systems were identified on plasmids pMP1046A and pLMP1046 and likely play a role in the stability and maintenance of the co-resident plasmids in JCM1046.

3.2.7 Comparative L. salivarius genomics and relationship to phenotype.

3.2.7.1 Chromosome

In contrast to the human probiotic strains *L. salivarius* UCC118 and *L. salivarius* CECT 5713 which share 98.5 % nt pairwise ID between their chromosomes and 98.6 % nt pairwise ID between their *repA*-type megaplasmids, the genome structure, and sequence of JCM1046 diverges significantly from the other published *L. salivarius* strains.

The chromosome of JCM1046 shares 68.4% nt pairwise ID with strain UCC118 and includes 55 regions (min 800 bp) [59], representing 16.5% of the chromosome, that are absent from strain UCC118 (Additional file 3.3). Indeed, a comparison of the chromosome of strain JCM1046 to that of the other published *L.*

salivarius genome sequences revealed 48 chromosomally encoded genes in JCM1046 that were absent in the other published *L. salivarius* genomes (Additional file 3.4). These genes primarily belong to categories of genes that have been shown to be hypervariable among *L. salivarius strains* [7] and other *Lactobacillus* species [60] and include transposases, phage-associated genes, and genes involved in carbohydrate metabolism and host interaction (Additional file 3.4). The GC % map of the JCM1046 chromosome identifies three regions with significantly deviating GC content, one of which is the remnant prophage that is resident on the chromosome. The smallest of these regions stretches from residues 782,449 to 793,883 bp. This 11.4 kb region has a GC % content of 43.6% and encodes a protein containing a mucin-binding MucBP domain (LSJ_0784), several transposases, hypothetical proteins and a choloylglycine hydrolase (BSH2, LSJ_0788). Although present in the porcine strains JCM1046 and cp400, this region is absent from other sequenced genomes of *L. salivarius* and may represent a niche specific adaptation.

BSH2 is one of two choloylglycine hydrolase genes encoded by the genome of JCM1046; the second (BSH1, LSJ_2111) is present on pMP1046A and is widespread among *L. salivarius* strains [10]. In contrast, BSH2 has only been identified in three isolates to date, JCM1046, LMG14476 and cp400, all of which are of animal origin. BSH2 confers on JCM1046 an ability to resist much higher concentrations of the major human conjugated bile acids when compared to strains that harbour BSH1 alone [10]. In addition, BSH2 has recently been shown to reduce weight gain and serum low density lipoprotein (LDL) cholesterol and liver triglycerides in mice fed normal or high-fat diets [61].

We have previously shown that exopolysaccharide (EPS) production levels and the presence of associated genes vary widely in *L. salivarius* [7]. JCM1046 harbours a single EPS gene cluster that spans 33 kb, containing 33 genes, including two pseudogenes (Additional file 3.5). The EPS locus exhibits an atypical GC content relative to the rest of the chromosome, 29.7% and 33.1%, respectively.

3.2.7.2 pMP1046A

Nine substantial regions of sequence diversity, ranging in size from 3.8-22.6 kb were identified between pMP1046A and the sequences of the other published *repA*-type megaplasmids (Figure 3.3; Table 3.2). Hypothetical proteins and transposases are abundant within these regions (Table 3.2). Indeed, region two and

region four primarily harbour hypothetical proteins, while region six harbours only IS elements (Table 3.2, R2, R4 and R6). Regions three and eight mostly encode solute transporters (Table 3.2 R3 and R8).

Table 3.2 Regions of diversity in pMP1046A

Region of diversity: base coordinates (*size bp*)	Gene	Start position	End position	Gene product
R1: 14543..37152 (*22609*)	LSJ_2012	14616	15683	Hypothetical membrane protein
	LSJ_2013	15827	16882	Hypothetical membrane protein
	LSJ_2014	17072	17242	Conserved hypothetical protein
	LSJ_2015	17257	18330	Putative thiosulfate sulfurtransferase
	LSJ_2016	18782	20686	D-proline reductase, *prdA*
	LSJ_2017	20688	21005	Conserved hypothetical protein
	LSJ_2018 (P)	20995	21720	Proline reductase, probable pseudogene
	LSJ_2019	21740	22480	Conserved hypothetical protein
	LSJ_2020	22501	22974	D-proline reductase
	LSJ_2021	22999	24018	Proline racemase
	LSJ_2022	24031	24909	Hypothetical membrane protein
	LSJ_2023	25005	26603	Amino acid permease
	LSJ_2024	26684	27322	Conserved hypothetical protein
	LSJ_2025	27434	27691	Conserved hypothetical protein
	LSJ_2026	27691	29583	Selenocysteine-specific elongation factor
	LSJ_2027	29573	30712	Cysteine desulfurase
	LSJ_2028	30713	32113	L-seryl-tRNA selenium transferase, *selA*
	LSJ_2029	32119	32448	Conserved hypothetical protein
	LSJ_2030	32560	33870	NADH dehydrogenase
	LSJ_2031	33963	34937	Selenophosphate synthase, *selD*
	LSJ_2032	35037	36209	Hypothetical membrane protein
	LSJ_2033	36228	36473	Conserved hypothetical protein
	LSJ_2034	36698	37129	Conserved hypothetical protein
R2: 52540..64667 (*12127*)	LSJ_2049	52523	54520	Conserved hypothetical protein
	LSJ_2050	54507	55370	Conserved hypothetical protein
	LSJ_2051	55360	56145	Conserved hypothetical protein
	LSJ_2052	56204	58906	DNA methylase
	LSJ_2053	58913	60871	DEAD/DEAH box helicase family protein
	LSJ_2054	60864	62066	Conserved hypothetical protein
	LSJ_2055	62137	64662	Conserved hypothetical protein
R3: 88322..98017 (*9695*)	LSJ_2078 (P)	88463	88739	Transposase, probable pseudogene
	LSJ_2079	88814	89641	Transposase ISLasa15, IS3 family
	LSJ_2080	90080	91279	MFS Transport protein
	LSJ_2081	91593	92813	MFS Transport protein
	LSJ_2082	92865	93761	Transcriptional regulators,

Region of diversity: base coordinates (*size bp*)	Gene	Start position	End position	Gene product
				LysR family
	LSJ_2083	93920	94756	Conserved hypothetical protein
	LSJ_2084	94785	95612	2-deoxy-D-gluconate 3-dehydrogenase
	LSJ_2085	95631	97346	Fumarate reductase flavoprotein subunit
	LSJ_2086	97367	98242	Shikimate 5-dehydrogenase
R4: 100291..121050 (*20759*)	LSJ_2089	100815	101621	Conserved hypothetical protein
	LSJ_2090	101614	101832	Hypothetical protein
	LSJ_2091	102071	102190	Hypothetical protein
	LSJ_2092	102310	103761	Plasmid replication protein-primase
	LSJ_2093	103865	104446	Hypothetical membrane protein
	LSJ_2094	104468	104620	Hypothetical membrane protein
	LSJ_2095	104818	105573	Hypothetical protein
	LSJ_2096	105746	106198	Hypothetical secreted protein
	LSJ_2097	106590	106853	Hypothetical protein
	LSJ_2098	106973	107965	Conserved hypothetical protein
	LSJ_2099	108439	108663	Hypothetical secreted protein, possible signal peptide
	LSJ_2100	109441	110277	Hypothetical protein
	LSJ_2101	110287	111057	Putative DNA-entry nuclease
	LSJ_2102	111064	111543	Conserved hypothetical protein
	LSJ_2103	111576	111743	Hypothetical secreted protein
	LSJ_2104	111993	112103	Hypothetical protein
	LSJ_2105	112160	112756	Conserved hypothetical protein
	LSJ_2106	112749	115094	Conserved hypothetical protein
	LSJ_2107	115914	118892	Hypothetical protein
	LSJ_2108	119031	120428	Conserved hypothetical protein
	LSJ_2109	120839	121078	Hypothetical protein
R5: 147401.. 153337 (*5936*)	LSJ_2136c	147401	148081	Fructose-6-phosphate aldolase
	LSJ_2137c	148148	148528	PTS system, glucitol/sorbitol-specific IIA component
	LSJ_2138c	148565	149590	PTS system, glucitol/sorbitol-specific IIBC component
	LSJ_2139c	149607	150149	PTS system, glucitol/sorbitol-specific IIC2 component
	LSJ_2140c	150161	150658	Sorbitol operon activator
	LSJ_2141c	150659	152518	Sorbitol operon transcription regulator
	LSJ_2142c	152534	153337	Sorbitol-6-phosphate 2-dehydrogenase
R6: 160003..164289 (*4286*)	LSJ_2150	160058	160575	Transposase ISLasa1a, IS1223 family
	LSJ_2151	160607	161473	IS1223 family transposase

Region of diversity: base coordinates (*size bp*)	Gene	Start position	End position	Gene product
	LSJ_2152	161537	162544	Transposase fragment
	LSJ_2153	162694	163983	ISL3 family transposase
R7: 167503..182637 (*15134*)	LSJ_2155	166716	167537	Integrase
	LSJ_2156	167573	167839	Hypothetical protein
	LSJ_2157	168343	169011	Hypothetical protein
	LSJ_2158	169087	169419	Hypothetical protein
	LSJ_2159	169424	170053	Conserved hypothetical protein
	LSJ_2160	170398	170802	Toxin antitoxin system, toxin component
	LSJ_2161	170802	171023	Toxin antitoxin system, antitoxin component
	LSJ_2162	171466	172614	AbpD bacteriocin export accessory protein
	LSJ_2163 (P)	1172630	174788	AbpT bacteriocin export accessory protein, probable pseudogene due to frameshift
	LSJ_2164	175441	175680	Hypothetical membrane spanning protein
	LSJ_2165	175717	176511	AbpR response regulator
	LSJ_2166	176525	177817	AbpK sensory Transduction Histidine Kinase
	LSJ_2167	177819	177938	AbpIP induction peptide
	LSJ_2168 (P)	178086	178232	AbpIM bacteriocin immunity protein
	LSJ_2169	178371	178577	Abp118 bacteriocin beta peptide
	LSJ_2170	178595	178789	Abp118 bacteriocin alpha peptide
	LSJ_2171	178795	179052	Bacteriocin-like prepeptide
	LSJ_2172	179182	179355	Nonfunctional salvaricin B precursor
	LSJ_2173	179588	179851	Hypothetical membrane spanning protein
	LSJ_2174	179890	180219	Hypothetical protein
	LSJ_2175	180441	181415	Hypothetical membrane associated protein
	LSJ_2176 (P)	1181578	182227	HAD-superfamily hydrolase, probable pseudogene due to frameshift
	LSJ_2177	182338	182679	Hypothetical protein
R8: 189782..193560 (*3778*)	LSJ_2187	189909	191414	Sodium solute symporter
	LSJ_2188	191432	192556	Na(+)/H(+) antiporter
	LSJ_2189	192623	193354	Xylose isomerase domain protein
R9: 204232..215364	LSJ_2201	204295	206286	Transketolase

Region of diversity: base coordinates (*size bp*)	Gene	Start position	End position	Gene product
(*11132*)	LSJ_2202	206304	206957	Transaldolase
	LSJ_2203	207466	208521	L-iditol 2-dehydrogenase
	LSJ_2204	208562	209611	Alcohol dehydrogenase
	LSJ_2205	209625	210896	Galacitol PTS, EIIC
	LSJ_2206	210923	211219	Galactitol PTS, EIIB
	LSJ_2207	211254	211706	Galacitol PTS, EIIA
	LSJ_2208	211888	212688	DeoR family transcriptional regulator
	LSJ_2209	212812	215178	Xylulose-5-phosphate/fructose-6-phosphate phosphoketolase

Genes associated with the regions of diversity (R1-R9) in pMP1046A, as illustrated in Figure 3.3. Genes present on the reverse strand are denoted by the suffix c following the locus tag (LSJ_XXX). Pseudogenes are denoted by (P). Numbers in italics represents the size of the region in bp.

The largest region of diversity among the strains examined is 22.6 kb (Figure 3.3, R1) and harbours several genes predicted to work synergistically with chromosomally encoded pathways to broaden the metabolic capabilities of strain JCM1046. Although present in strain cp400, this region is highly divergent in all other examined plasmids (Figure 3.3) and primarily encodes proteins involved in amino acid metabolism. JCM1046 is prototrophic for L-proline due to the presence of a chromosomally-encoded pathway. Three paralogous genes (LSJ_2016, LSJ_2020 and LSJ_2021) in this region are responsible for the interconversion of L-proline to D-proline. Also present in this region are two genes (LSJ_2031, *selD* and LSJ_2028, *selA*) which work in conjunction with the chromosomally encoded gene (LSJ_0220, *serS*) to synthesise L-selenocysteine. These increased biosynthetic capabilities are likely to enhance the ability of JCM1046 to thrive in the competitive porcine GIT.

The genes present in regions five and nine (Table 3.2) are primarily involved in the metabolism and transport of carbohydrates, and vary from strain to strain (Figure 3.3, R5 and R9). Similarly to pMP118, pMP1046A harbours both single copy and paralogous genes that complete a number of the carbohydrate fermentative pathways that are partially encoded by the chromosome of JCM1046A. These include the pentose phosphate and gluconeogenesis pathways as well as the fermentation pathways for sorbitol and rhamnose.

Bacteriocin production is a putative probiotic trait of *L. salivarius* strains (see review [62]). The genetic organisation of the 7.9 kb bacteriocin locus in pMP1046A is analogous to that of the Abp118 locus in the human isolate UCC118 (Figure 3.3 R7). The structural genes (LSJ_2170 and LSJ_2169) of the bacteriocin locus of pMP1046A, are identical to the genes (Sln1 and Sln2) which are responsible for the production of the two-component antilisterial bacteriocin Salivaricin P. This bacteriocin differs in sequence to Apb118 by two amino acids [63] and is produced by several other porcine isolates of *L. salivarius* [64]. However, a frame-shift in the *abpT* gene (LSJ_2163) of JCM1046 is likely responsible for the bacteriocin negative phenotype observed in this strain [12].

3.2.7.3 pCTN1046

The putative conjugative element Tn*6224* harboured by plasmid pCTN1046 shares 96.2 % nt sequence ID with the conjugative element Tn*916* and lacks only two genes which encode hypothetical proteins in the conjugative region of TN*916*. When comparing pCTN1046 to other sequenced *L. salivarius* genomes, pCTN1046 shares 64.6 % nt ID with the 30.4 kb plasmid pLS51C harboured by the probiotic avian isolate SMDX51 [16]. This plasmid shares sequence homology with both the plasmid backbone and conjugative element of pCTN1046 (Figure 3.4). Tn*6224* appears to be functionally intact, containing the conjugative, recombination, transcriptional regulation and accessory genes (Additional file 3.6) associated with Tn916. In contrast, the integrated conjugative element that is resident in pSL51C appears to be a remnant of a conjugative element as it lacks the recombination genes *xis* (LSJ_5019) and *int* (LSJ_5020). pLS51C harbours a limited number of the conjugative genes present in Tn*6224* and Tn*916* but lacks the *ardA* gene present in pCTN1046 which has been recently shown to aid the transfer of mobile genetic elements (MGE) between unrelated bacterial species [65]. A putative TnGBS1-like element (TnLsal1.1) was identified in *L. salivarius* strain DSM20555. However, our analysis suggests that the contig predicted to harbour TnLsal1.1 [66] forms part of the putative pMP20555 megaplasmid in the type-strain *L. salivarius* DSM20555. The weak homology between the proteins identified in TnLsal1.1 and those identified in other TnGBS1-like elements [66] may be due to their similar functional roles in their respective replicons.

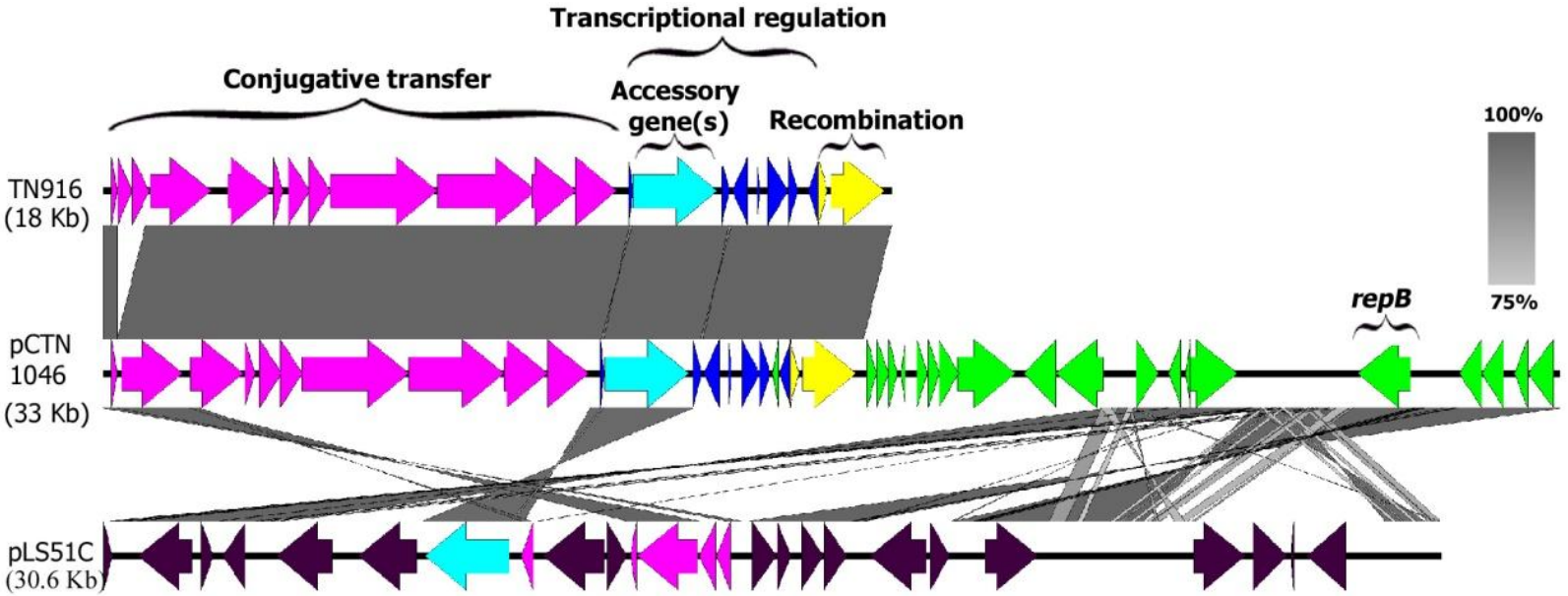


Figure 3.4. Sequence alignment of Tn*916*, pCTN1046 and pLS51C.
A linear comparison of the BLASTN matches between the extrachromosomal replicons pCTN1046 and pLS51C (harboured by *L. salivarius* strain SMXD51 [16]) and the conjugative transposon Tn916. Vertical grey-coloured blocks between sequences indicate regions of shared nt similarity ID. The gradient of the grey colour corresponds to the percentage of shared nt ID (dark grey (100%)-light grey (75%)). The genes in each element are coloured according to their function in the conjugative transposon Tn*916*: pink (conjugative transfer), turquoise (accessory genes and transcriptional regulation), dark blue (transcriptional regulation) and yellow (recombination). Genes encoded by the plasmid backbone of pCTN1046 are green, and those associated with the backbone of pLS51C are dark purple.

L. salivarius strains are increasingly being examined for their probiotic properties in both humans and animals [5]. Dissemination of antibiotic resistance genes *via* the food chain to either the resident microbiota of the human gut or pathogenic bacteria is likely to have far reaching effects on both human and animal health and present a major financial cost [67]. Thus, the identification of conjugative transposons carrying antibiotic resistance genes in the genomes of two animal isolates of *L. salivarius* may have repercussions for strain selection in future probiotic studies.

3.2.7.4 pMP1046B and pLMP1046

Plasmids pMP1046B and pLMP1046 share neither sequence homology nor gene synteny with the additional *L. salivarius* plasmids sequenced to date. Both of these replicons require further functional characterisation to determine whether or not they have an impact on the phenotype and ecological properties of JCM1046.

3.3 CONCLUSION

The porcine strain JCM1046 harbours the most structurally complex multipartite genome identified in *L. salivarius* to date. Through complete sequencing and assembly of the genome of JCM1046 we identified two additional replicons that were not previously known to form part of the plasmid complement of this strain, and that would probably not have been identified by the high-coverage draft genome sequencing commonly applied. We determined that one of these replicons, pMP1046B is a candidate chromid, though much of its gene function remains cryptic. The plasmids of *L. salivarius* probably confer on their host many of the genes associated with niche adaptation and which are known to modulate the phenotype of a strain significantly. JCM1046 was found to harbour both plasmid-encoded (pMP1046A) and chromosomally encoded genes associated with adaptation to the GIT environment. The putative replication *ori* of pLMP1046 was identified and the sequence of this linear plasmid will provide a genetic platform for the study of linear DNA replication in *Lactobacillus* sp. An integrated conjugative transposon (Tn*6224*), carrying tetracycline resistance was identified in plasmid pCTN1046, the first described in a sequenced *L. salivarius* genome. It will be interesting to see how prevalent Tn*6224*-like elements are within the *L. salivarius* population, as more genome sequences become available.

3.4 METHODS

3.4.1 Bacterial strains and culture conditions.

L. salivarius strains were routinely cultured at 37^{o}C under micro-aerophilic conditions (5% CO_2) in de Man-Rogosa-Sharpe (MRS) medium (Oxoid Ltd, Basingstoke, Hampshire, UK).

3.4.2 PFGE plug preparations.

Agarose gel plugs of high molecular weight DNA for PFGE were prepared according to a published protocol [12]

3.4.3 S1-nuclease treatment.

Single slices (2 mm × 2 mm) were treated with *Aspergillus oryzae* S1 nuclease (Roche, Mannheim, Germany) according to a published protocol [12].

3.4.4 Restriction of PFGE plugs.

Single slices (2 mm × 2 mm) were washed three times for 15 min in 1ml 10 mM Tris.Cl, 0.1 mM EDTA (pH 8.0) at room temperature. Each slice was pre-incubated with 250 μl of restriction buffer recommended for the enzyme for 30 min at 4^{o}C and then replaced with 250 μl of fresh buffer containing 20 units of restriction enzyme. Restriction digests were carried out overnight at temperatures recommended by the supplier.

3.4.5 Pulsed field gel electrophoresis.

Treated (S1-nuclease/restriction enzyme) and untreated plugs of genomic DNA were examined under conditions employed in a previously published protocol [12]. Gels were stained in distilled water containing 0.5 μg/ml ethidium bromide for 60 min in light-limited conditions and destained in water for 30 min.

3.4.6 Probe preparation and Southern hybridization.

Probe preparations and Southern blot hybridizations were carried out according to a published protocol [12]. The primers used to generate PCR amplicons that were used as probes are listed in Additional file 3.7.

3.4.7 Genome sequencing.

L. salivarius genomic DNA (gDNA) isolation was performed as described previously [1]. The genome of JCM1046 genome was sequenced using a combination of shotgun sequencing by the Sanger method (4-fold coverage), pyrosequencing (24-fold coverage) and Illumina (204-fold coverage). A large-insert

(~40 kb) fosmid library was constructed in the CopyControl™ pCCFOS™ vector system (Epicentre Technologies, USA).Corporation, USA) Insert ends (~800 bp/read) were sequenced generating mate pairs and 7.5 Mb sequencing data. Pyrosequencing generated approximately 217,000 unpaired reads (~ 250 nt); from a half plate on a 454 FLX instrument (Agencourt Biosciences, Beverly, MA). In addition to the shotgun and 454 data for the JCM1046 genome, an additional half lane of Illumina sequencing (23 Mb total sequence data) was obtained which consisted of a 3 kb mate-pair library and a 400 bp paired-end library (Fasteris, Geneva, Switzerland). Each Illumina library provided an average of 204-fold coverage. Illumina reads were assembled (default settings) into contigs using Velvet v 0.7 [68], which were then used to generate 300 bp pseudocontigs. A *de novo* genome assembly of the shotgun, 454 and Illumina (pseudocontigs) sequence data was performed using the Roche/454 Life Sciences Newbler (Gs) assembler v 2.3 [69], producing an initial assembly of 102 contigs (>500 bp) distributed over 32 scaffolds for the genome of JCM1046. The resulting 454 assembly was then used as a reference for the mapping of raw Illumina data. This mapping assembly was performed using Mira [70] and undertaken to extend contigs, close gaps and for error correction of the draft genome. Gap closure was achieved using a PCR-based strategy. Primers were designed at the end of contigs and Dreamtaq DNA polymerase (Fermentas, Ontario, Canada) was used to amplify products corresponding to contig-contig gaps. Scaffolds were ordered and oriented by PCR using primers were designed at the ends of the scaffolds and the inter-scaffold region was amplified using Extensor long PCR enzyme mix (Abgene, Epsom, UK). PCR products for both the sequencing gaps and the inter-scaffold gaps were sequenced by Eurofins MWG Operon (Ebersberg, Germany) and the sequences were integrated into the assembly using PHRAP [71]. Correct placement of the gap sequences was confirmed by observation using Tablet, a next generation sequencing graphical viewer [72].

3.4.8 Genome annotation.

Annotation was carried out as according to a published protocol [73] with minor modifications. Specifically, initial *a*nnotation was transferred from the related strain *L. salivarius UCC118* [74] and then manually curated in Artemis [75]**.** PHAST [48] was used to identify prophage regions within the genome sequence.

3.4.9 Data availability

The annotated genome sequence has been deposited in GenBank under accession numbers CP007646 (chromosome), CP007647 (pMP1046A), CP007648 (pMP1046B), CP007649 (pLMP1046), CP007650 (pCTN1046).

3.4.10 Genome comparisons.

Nucleotide alignments were generated using a local BLAST v 2.2.22 installation which were then visualized and analyzed for gene conservation and sequence synteny using the Artemis Comparison Tool (ACT) [76].

3.4.11 Identification of novel genetic regions.

The Novel Region Finder module of Pan seq v 2.0 [59] was used to identify novel genomic regions in strain JCM1046, compared to other *L. salivarius* genome sequences. A minimum novel region size of 800 bp was chosen and default Nucmer values were used.

3.5 COMPETING INTERESTS

The authors declare that they have no competing financial interests.

3.6 AUTHORS' CONTRIBUTIONS

EJR performed research, analyzed data and drafted the manuscript. BMF and MJC analyzed data and provided ongoing advice throughout the study. PWOT conceived the research, participated in its design and coordination, analyzed data and drafted the manuscript. All authors read and approved the final manuscript.

3.7 ACKNOWLEDGEMENTS

This research was supported by Science Foundation Ireland through a Research Frontiers 448 Programme award to PWOT (05/RFP/GEN047), and by a Centre for Science, Engineering and Technology award to the Alimentary Pharmabiotic Centre

3.8 REFERENCES

1. Li Y, Raftis E, Canchaya C, Fitzgerald GF, Sinderen Dv, O'Toole PW: **Polyphasic analysis indicates that *Lactobacillus salivarius* subsp. *salivarius* and *Lactobacillus salivarius* subsp. *salicinius* do not merit separate subspecies status**. *Int J of Syst Evol Microbiol* 2006(56):2397-2403.
2. Reuter G: **The *Lactobacillus* and *Bifidobacterium* microflora of the human intestine: composition and succession.** *Curr Issues Intest Microbiol* 2001, **2**(2):43–53.
3. Mitsuoka T: **Vergleichende untersuchungen über die *Laktobazillen* aus den faeces von menschen, schweinen und hühnern**. *Bakteriol* 1969, **210**:32-51.
4. Martín R, Jiménez E, Olivares M, Marín ML, Fernández L, Xaus J, Rodríguez JM: ***Lactobacillus salivarius* CECT 5713, a potential probiotic strain isolated from infant feces and breast milk of a mother-child pair.** *Int J Food Microbiol* 2006, **112**(1):35-43.
5. Neville BA, O'Toole PW: **Probiotic properties of *Lactobacillus salivarius* and closely related *Lactobacillus* species**. *Future Microbiol* 2010, **5**(5):759-774.
6. Ryan KA, Daly P, Li Y, Hooton C, O'Toole PW: **Strain-specific inhibition of *Helicobacter pylori* by *Lactobacillus salivarius* and other lactobacilli**. *J Antimicrob Chemother* 2008, **61**(4):831-834.
7. Raftis EJ, Salvetti E, Torriani S, Felis GE, O'Toole PW: **Genomic diversity of *Lactobacillus salivarius***. *Appl Environ Microbiol* 2011, **77**(3):954-965.
8. Claesson M, Li Y, Leahy S, Canchaya C, van Pijkeren J, Cerdeno-Tarraga A, Parkhill J, Flynn S, O'Sullivan G, Collins J: **Multireplicon genome architecture of *Lactobacillus salivarius***. *Proc Natl Acad Sci USA* 2006, **103**(17):6718 - 6723.
9. Flynn S, van Sinderen D, Thornton GM, Holo H, Nes IF, Collins JK: **Characterization of the genetic locus responsible for the production of ABP-118, a novel bacteriocin produced by the probiotic bacterium *Lactobacillus salivarius* subsp. *salivarius* UCC118**. *Microbiology* 2002, **148**(4):973-984.
10. Fang F, Li Y, Bumann M, Raftis EJ, Casey PG, Cooney JC, Walsh MA, O'Toole PW: **Allelic variation of bile salt hydrolase genes in *Lactobacillus salivarius* does not determine bile resistance levels**. *J Bacteriol* 2009, **191**(18):5743-5757.
11. Corr SC, Li Y, Riedel CU, O'Toole PW, Hill C, Gahan CGM: **Bacteriocin production as a mechanism for the antiinfective activity of *Lactobacillus salivarius* UCC118**. *Proc Natl Acad Sci USA* 2007:0700440104.
12. Li Y, Canchaya C, Fang F, Raftis E, Ryan KA, van Pijkeren J-P, van Sinderen D, O'Toole PW: **Distribution of megaplasmids in *Lactobacillus salivarius* and other lactobacilli.** *J Bacteriol* 2007, **189**(17):6128-6139.
13. van Pijkeren J, Canchaya C, Ryan K, Li Y, Claesson M, Sheil B, Steidler L, O'Mahony L, Fitzgerald G, van Sinderen D: **Comparative and functional analysis of sortase-dependent proteins in the predicted secretome of *Lactobacillus salivarius* UCC118**. *Appl Environ Microbiol* 2006, **72**(6):4143 - 4153.

14. Jimenez E, Martin R, Maldonado A, Martin V, Gomez de Segura A, Fernandez L, Rodriguez JM: **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.
15. Ham J-S, Kim H-W, Seol K-H, Jang A, Jeong S-G, Oh M-H, Kim D-H, Kang D-K, Kim G-B, Cha C-J: **Genome sequence of *Lactobacillus salivarius* NIAS840, isolated from chicken intestine**. *J Bacteriol* 2011, **193**(19):5551-5552.
16. Kergourlay G, Messaoudi S, Dousset X, Prévost H: **Genome sequence of *Lactobacillus salivarius* SMXD51, a potential probiotic strain isolated from chicken cecum, showing anti-*Campylobacter* activity**. *J Bacteriol* 2012, **194**(11):3008-3009.
17. Cho Y-J, Choi JK, Kim J-H, Lim Y-S, Ham J-S, Kang D-K, Chun J, Paik H-D, Kim G-B: **Genome sequence of *Lactobacillus salivarius* GJ-24, a probiotic strain isolated from healthy adult intestine**. *J Bacteriol* 2011, **193**(18):5021-5022.
18. MacKenzie DA, McLay K, Roos S, Walter J, Swarbreck D, Drou N, Crossman LC, Juge N: **Draft genome sequence of a novel *Lactobacillus salivarius* strain isolated from piglet**. *Genome Announc* 2014, **2**(1).
19. Overhage J, Sielker S, Homburg S, Parschat K, Fetzner S: **Identification of large linear plasmids in *Arthrobacter* spp. encoding the degradation of quinaldine to anthranilate**. *Microbiology* 2005, **151**(2):491-500.
20. Kinashi H: **Giant linear plasmids in *Streptomyces*: a treasure trove of antibiotic biosynthetic clusters**. *J Antibiot* 2011, **64**(1):19-25.
21. Chater K, Kinashi H: ***Streptomyces*; linear plasmids: their discovery, functions, interactions with other replicons, and evolutionary significance**. In: *Microbial Linear Plasmids.* Edited by Meinhardt F, Klassen R, vol. 7: Springer Berlin / Heidelberg; 2007: 1-31.
22. Chen C: ***Streptomyces*; linear plasmids: replication and telomeres**. In: *Microbial Linear Plasmids.* Edited by Meinhardt F, Klassen R, vol. 7: Springer Berlin / Heidelberg; 2007: 33-61.
23. Barbour A, Garon C: **Linear plasmids of the bacterium *Borrelia burgdorferi* have covalently closed ends**. *Science* 1987, **237**(4813):409-411.
24. Stromsten NJ, Benson SD, Burnett RM, Bamford DH, Bamford JKH: **The *Bacillus thuringiensis* linear double-stranded DNA phage Bam35, which is highly similar to the *Bacillus cereus* linear plasmid pBClin15, has a prophage state**. *J Bacteriol* 2003, **185**(23):6985-6989.
25. Ravin NV: **N15: The linear phage-plasmid**. *Plasmid* 2011, **65**(2):102-109.
26. Hertwig S, Klein I, Lurz R, Lanka E, Appel B: **PY54, a linear plasmid prophage of *Yersinia enterocolitica* with covalently closed ends**. *Mol Microbiol* 2003, **48**(4):989-1003.
27. Casjens SR, Gilcrease EB, Huang WM, Bunny KL, Pedulla ML, Ford ME, Houtz JM, Hatfull GF, Hendrix RW: **The pKO2 linear plasmid prophage of *Klebsiella oxytoca***. *J Bacteriol* 2004, **186**(6):1818-1832.
28. Alemayehu D, Ross RP, O'Sullivan O, Coffey A, Stanton C, Fitzgerald GF, McAuliffe O: **Genome of a virulent bacteriophage Lb338-1 that lyses the probiotic *Lactobacillus paracasei* cheese strain**. *Gene* 2009, **448**(1):29-39.
29. Roussel Y, C. Colmin, J. M. Simonet, Decaris B: **Strain characterization, genome size and plasmid content in the *Lactobacillus acidophilus* group (Hansen and Mocquot).** *J Appl Bacteriol* 1993, **74**:549–556.

30. Franke AE, Clewell DB: **Evidence for a chromosome-borne resistance transposon (Tn*916*) in *Streptococcus faecalis* that is capable of "conjugal" transfer in the absence of a conjugative plasmid**. *J Bacteriol* 1981, **145**(1):494-502.
31. Clewell DB, Flannagan SE, Jaworski DD: **Unconstrained bacterial promiscuity: the Tn*916*-Tn*1545* family of conjugative transposons**. *Trends Microbiol* 1995, **3**(6):229-236.
32. Bertram J, Stratz M, Durre P: **Natural transfer of conjugative transposon Tn*916* between gram-positive and gram-negative bacteria**. *J Bacteriol* 1991, **173**(2):443-448.
33. Boguslawska J, Zycka-Krzesinska J, Wilcks A, Bardowski J: **Intra- and interspecies conjugal transfer of Tn*916*-like elements from *Lactococcus lactis in vitro* and *in vivo***. *Appl Environ Microbiol* 2009, **75**(19):6352-6360.
34. Devirgiliis C, Coppola D, Barile S, Colonna B, Perozzi G: **Characterization of the Tn*916* conjugative transposon in a food-borne strain of *Lactobacillus paracasei***. *Appl Environ Microbiol* 2009, **75**(12):3866-3871.
35. Roberts AP, Cheah G, Ready D, Pratten J, Wilson M, Mullany P: **Transfer of Tn*916*-like elements in microcosm dental plaques**. *Antimicrob Agents Chemother* 2001, **45**(10):2943-2946.
36. Schjørring S, Krogfelt KA: **Assessment of bacterial antibiotic resistance transfer in the gut**. *Int J Microbiol* 2011, **2011**:10.
37. Wozniak RAF, Waldor MK: **Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow**. *Nat Rev Micro* 2010, **8**(8):552-563.
38. Roberts AP, Mullany P: **A modular master on the move: the Tn916 family of mobile genetic elements**. *Trends Microbiol* 2009, **17**(6):251-258.
39. Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J: **DNAPlotter: circular and linear interactive genome visualization**. *Bioinformatics* 2009, **25**:119 - 120.
40. Fang F, Flynn S, Li Y, Claesson MJ, van Pijkeren J-P, Collins JK, van Sinderen D, O'Toole PW: **Characterization of endogenous plasmids from *Lactobacillus salivarius* UCC118**. *Appl Environ Microbiol* 2008, **74**(10):3216-3228.
41. Barton BM, Harding GP, Zuccarelli AJ: **A general method for detecting and sizing large plasmids**. *Analyt Biochem* 1995, **226**(2):235-240.
42. Warner JE, Onderdonk AB: **Method for optimizing pulsed-field gel electrophoresis banding pattern data**. *JMD* 2003, **5**(1):21-27.
43. Wang H, Roberts AP, Mullany P: **DNA sequence of the insertional hot spot of Tn*916* in the *Clostridium difficile* genome and discovery of a Tn*916*-like element in an environmental isolate integrated in the same hot spot**. *FEMS Microbiol Lett* 2000, **192**(1):15-20.
44. Harrison PW, Lower RPJ, Kim NKD, Young JPW: **Introducing the bacterial `chromid': not a chromosome, not a plasmid**. *Trends Microbiol* 2010, **18**(4):141-148.
45. Sanchez-Perez G, Mira A, Nyirő G, Pašić L, Rodriguez-Valera F: **Adapting to environmental changes using specialized paralogs**. *Trends Genet* 2008, **24**(4):154-158.
46. Wagenknecht M, Dib J, Thürmer A, Daniel R, Farías M, Meinhardt F: **Structural peculiarities of linear megaplasmid, pLMA1, from**

***Micrococcus luteus*; interfere with pyrosequencing reads assembly**. *Biotechnol Lett* 2010, **32**(12):1853-1862.
47. Roberts AP, Chandler M, Courvalin P, Guédon G, Mullany P, Pembroke T, Rood JI, Jeffery Smith C, Summers AO, Tsuda M *et al*: **Revised nomenclature for transposable genetic elements**. *Plasmid* 2008, **60**(3):167-173.
48. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS: **PHAST: A Fast Phage Search Tool**. *Nucleic Acids Res* 2011, **39**(suppl 2):W347-W352.
49. Jore MM, Brouns SJJ, van der Oost J: **RNA in Defense: CRISPRs Protect Prokaryotes against Mobile Genetic Elements**. *Cold Spring Harb Perspect Biol* 2012, **4**(6).
50. Jannière L, Gruss A, Ehrlich. D: **Plasmids**. In: *Bacillus subtilis and other gram-positive bacteria.* Edited by A. L. Sonenshein JAH, and R. Losick. Washington, D.C: American Society for Microbiology; 1993: 625-644.
51. Alikhan N-F, Petty N, Ben Zakour N, Beatson S: **BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons**. *BMC Genomics* 2011, **12**(1):402.
52. Bentley SD, Parkhill J: **Comparative genomic structure of prokaryotes**. *Ann Rev Genet* 2004, **38**(1):771-791.
53. Heavens D, Tailford LE, Crossman L, Jeffers F, MacKenzie DA, Caccamo M, Juge N: **Genome sequence of the vertebrate gut symbiont *Lactobacillus reuteri* ATCC 53608**. *J Bacteriol* 2011, **193**(15):4015-4016.
54. Beaurepaire C, Chaconas G: **Mapping of essential replication functions of the linear plasmid lp17 of *B. burgdorferi* by targeted deletion walking**. *Mol Microbiol* 2005, **57**(1):132-142.
55. Chang P-C, Cohen SN: **Birdirectional replication from an internal origin in a linear *Streptomyces* plasmid**. *Science* 1994(265):952-954.
56. Ravin NV, Kuprianov VV, Gilcrease EB, Casjens SR: **Bidirectional replication from an internal ori site of the linear N15 plasmid prophage**. *Nucleic Acids Res* 2003, **31**(22):6552-6560.
57. Mardanov A, Ravin N: **Functional characterization of the *repA* replication gene of linear plasmid prophage N15**. *Res Microbiol* 2006, **157**(2):176-183.
58. Shiffman D, Cohen SN: **Reconstruction of *Streptomyces* linear plasmid replication from separately cloned DNA fragment: existence pf a cryptic origin of circular replication within the linear plasmid** *Proc Natl Acac Sci USA* 1992(89):6129-6133.
59. Laing C, Buchanan C, Taboada E, Zhang Y, Kropinski A, Villegas A, Thomas J, Gannon V: **Pan-genome sequence analysis using Panseq: an online tool for the rapid analysis of core and accessory genomic regions**. *BMC Bioinformatics* 2010, **11**(1):461.
60. Berger B, Pridmore R, Barretto C, Delmas-Julien F, Schreiber K, Arigoni F, Brüssow H: **Similarity and differences in the *Lactobacillus acidophilus* group identified by polyphasic analysis and comparative genomics**. *J Bacteriol* 2007, **189**(4):1311-1321.
61. Joyce SA, MacSharry J, Casey PG, Kinsella M, Murphy EF, Shanahan F, Hill C, Gahan CGM: **Regulation of host weight gain and lipid metabolism by bacterial bile acid modification in the gut**. *Proc Nat Acad Sci USA* 2014, **111**(20):7421-7426.

62. Messaoudi S, Manai M, Kergourlay G, Prévost H, Connil N, Chobert JM, Dousset X: ***Lactobacillus salivarius*: bacteriocin and probiotic activity**. *Food Microbiol* 2013, **36**(2):296-304.
63. Barrett E, Hayes M, O'Connor P, Gardiner G, Fitzgerald GF, Stanton C, Ross RP, Hill C: **Salivaricin P, one of a family of two-component antilisterial bacteriocins produced by intestinal isolates of *Lactobacillus salivarius***. *Appl Environ Microbiol* 2007, **73**(11):3719-3723.
64. O'Shea EF, O'Connor PM, Raftis EJ, O'Toole PW, Stanton C, Cotter PD, Ross RP, Hill C: **Production of multiple bacteriocins from a single locus by gastrointestinal strains of *Lactobacillus salivarius***. *J Bacteriol* 2011.
65. Chen K, Reuter M, Sanghvi B, Roberts GA, Cooper LP, Tilling M, Blakely GW, Dryden DTF: **ArdA proteins from different mobile genetic elements can bind to the EcoKI Type I DNA methyltransferase of *E. coli* K12**. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 2014, **1844**(3):505-511.
66. Guérillot R, Da Cunha V, Sauvage E, Bouchier C, Glaser P: **Modular evolution of TnGBSs, a new family of integrative and conjugative elements associating insertion sequence transposition, plasmid replication, and conjugation for their spreading**. *J Bacteriol* 2013, **195**(9):1979-1990.
67. Panel F: **Opinion of the Scientific Panel on additives and products or substances used in animal feed (FEEDAP) on the updating of the criteria used in the assessment of bacteria for resistance to antibiotics of human or veterinary importance** *EFSA Journal* 2005.
68. Zerbino DR, Birney E: **Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs**. *Genome Res* 2008, **18**(5):821-829.
69. Mardis ER: **Next-generation DNA sequencing methods**. *Ann Rev Genet Hum Genet* 2008, **9**(1):387-402.
70. Chevreux B, Wetter T, S S: **Genome sequence assembly using trace signals and additional sequence information.** In: *Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB).* 1999.
71. Green P: **PHRAP v 1.080812**. In. [http://phrap.org]; 1999.
72. Milne I, Bayer M, Cardle L, Shaw P, Stephen G, Wright F, Marshall D: **Tablet—next generation sequence assembly visualization**. *Bioinformatics* 2010, **26**(3):401-402.
73. Forde B, Neville B, O' Donnell M, Riboulet-Bisson E, Claesson M, Coghlan A, Ross R, O' Toole P: **Genome sequences and comparative genomics of two *Lactobacillus ruminis* strains from the bovine and human intestinal tracts**. *Microb Cell Fact* 2011, **10**(1):S13.
74. Otto TD, Dillon GP, Degrave WS, Berriman M: **RATT: Rapid Annotation Transfer Tool**. *Nucleic Acids Res* 2011.
75. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream M-A, Barrell B: **Artemis: sequence visualization and annotation**. *Bioinformatics* 2000, **16**(10):944-945.
76. Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J: **ACT: the Artemis comparison tool**. *Bioinformatics* 2005, **21**:3422 - 3423.

3.9 ADDITIONAL FILES

Additional file 3.1 -Pseudogenes in the *L. salivarius* JCM1046 genome.

	Gene	Start	Stop	Product
Chromosome	LSJ_0013	14166	14972	Membrane proteins related to metalloendopeptidase-like, M23, probable pseudogene
	LSJ_0015	17225	17786	Similar to hypothetical membrane spanning protein; nonfunctional due to frameshift
	LSJ_0035	39059	39375	Similar to hypothetical protein; nonfunctional due to frameshift and interrupting stop codon
	LSJ_0036c	39465	40856	Similar to conserved hypothetical protein; nonfunctional due to frameshift
	LSJ_0051	55863	56213	Similar to conserved hypothetical protein; nonfunctional due to frameshift
	LSJ_0082	88856	89753	IS1223 family transposase
	LSJ_0115	129766	130572	hydrolase, haloacid dehalogenase-like family
	LSJ_0120c	133839	134525	Hypothetical membrane spanning protein, putative pseudogene
	LSJ_0157	171119	171436	4-carboxymuconolactone decarboxylase, pseudogene
	LSJ_0175	192343	192639	Similar to transposase ISLasa3b, IS200 family; pseudogene due to frameshift
	LSJ_0206	223825	225257	O-succinylbenzoic acid--CoA ligase, putative pseudogene
	LSJ_0249	271441	272624	Aminotransferase class I and II, putative pseudogene
	LSJ_0258	283300	285123	Similar to ABC transporter, ATP-binding protein; nonfunctional due to interrupting stop codon
	LSJ_0259	285432	285866	Acetyltransferase, GNAT family, probable pseudogene due to a frameshift
	LSJ_0272	298623	299889	3-hydroxy-3-methylglutaryl-coenzyme A reductase, probable pseduogene
	LSJ_0284	311110	312330	IS605 OrfB family Transposase, putative pseudogene
	LSJ_0321c	350936	352215	Transposase ISLasa9b, IS200-IS605 family, putative pseudogene
	LSJ_0335c	362299	364666	DinG family ATP-dependent helicase, putative pseudogene
	LSJ_0398	426326	427974	Phosphoenolpyruvate carboxykinase (ATP)
	LSJ_0427	452773	453075	ComG operon protein 3, putative pseudogene
	LSJ_0552	555936	557079	Conserved hypothetical protein, putative pseudogene
	LSJ_0560c	562019	563218	Transposase ISLasa4k, ISL3 family, putative pseudogene
	LSJ_0691c	688846	690133	Transposase ISLasa4k, ISL3 family, pseudogene due to frameshift
	LSJ_0695	694844	696132	Transposase ISLasa4k, ISL3 family, putative pseudogene

Gene	Start	Stop	Product
LSJ_0736c	739258	739963	HAD family hydrolase, pseudogene due to frameshift
LSJ_0740	743377	745639	ComE operon protein 3, putative pseudogene
LSJ_0779	778470	779236	Ribonuclease HII (EC 3.1.26.4), pseudogene due to frameshift
LSJ_0780	779301	779780	DNA processing protein, putative due to truncation
LSJ_0781c	779932	780630	IS116/IS110/IS902 family Transposase, putative pseudogene
LSJ_0811c	817783	818919	Integrase
LSJ_0835	835603	836493	Ribosomal protein L11 methyltransferase
LSJ_0845	845183	847114	Carbamoyl-phosphate synthase large chain, probable pseudogene due to truncation
LSJ_0922c	924651	926097	ATP-dependent DNA helicase, putative pseudogene
LSJ_0964c	971520	972696	Cyclopropane-fatty-acyl-phospholipid synthase, putative pseudogene
LSJ_0982	982582	983553	ISL3 family transposase, putative pseudogene
LSJ_0983	983538	984008	ISL3 family transposase, putative pseudogene
LSJ_0996c	993104	993438	Hypothetical membrane spanning protein,putative pseudogene
LSJ_1011c	1007686	1009341	DNA/RNA helicase, DEAD/DEAH box family, pseudogene due to introduction of a premature stop codon
LSJ_1195c	1209113	1210145	O-sialoglycoprotein endopeptidase, putative pseudogene
LSJ_1260c	1269915	1271133	Transposase ISLasa12, IS607 family, putative pseudogene
LSJ_1261c	1271200	1271727	Hypothetical membrane spanning protein
LSJ_1321c	1339068	1340403	pyridine nucleotide-disulfide oxidoreductase, putative pseudogene
LSJ_1328c	1347472	1350273	ATPase of the AAA+ class, pseudogene due to truncation
LSJ_1444c	1486638	1487535	IS1223 family transposase
LSJ_1445	1488187	1489477	Transposase ISLasa4k, ISL3 family, putative pseudogene
LSJ_1491c	1517727	1518141	Conserved hypothetical protein
LSJ_1521	1549221	1550478	Transposase ISLasa4k, ISL3 family, putative pseudogene

	Gene	Start	Stop	Product
	LSJ_1558c	1583336	1585259	treB, PTS system, trehalose-specific IIBC component, putative pseudogene
	LSJ_1572c	1605141	1608064	Phosphoglycerol transferase and related proteins, alkaline phosphatase superfamily
	LSJ_1588c	1628373	1628973	Transposase ISLasa7, IS3 family, putative pseudogene
	LSJ_1589c	1629349	1631730	Hypothetical exported protein, putative pseudogene
	LSJ_1619	1664260	1664670	Transposase ISLasa1b, IS1223 family, putaive pseudogene
	LSJ_1626c	1671775	1672836	Beta-hexosaminidase, putative pseudogene
	LSJ_1634	1679721	1680362	Membrane-bound protease, CAAX family
	LSJ_1655c	1696618	1697527	L-2-hydroxyisocaproate dehydrogenase, putative pseudogene
	LSJ_1696	1743819	1744132	Nitroreductase family protein, pseudogene due to a truncation
	LSJ_1717	1762714	1764612	Hypothetical protein with PPE repeats, pseudogene due to introduction of a stop codon
	LSJ_1730c	1773854	1774099	COG0583 [K] Transcriptional regulator
	LSJ_1770	1814660	1815094	Transposase ISLasa4k, ISL3 family, putative pseudogene
	LSJ_1798	1831784	1832808	Inner membrane protein, putative pseudogene
pMP1046A	LSJ_2003c	3607	4482	PTS system, sucrose-specific IIABC component,possible pseudogene due to truncation
	LSJ_2004c	4611	4925	CAT domain transcriptional antiterminator, probablepseudogene due to truncation
	LSJ_2018	20995	21720	Proline reductase, probable pseudogene due to aninternal stop codon; Internal stop codon
	LSJ_2044	48004	48887	Hypothetical protein, probable pseudogene
	LSJ_2058c	66885	67781	Transposase IS1223 family; label: Transposase IS1223 family CDS
	LSJ_2060c	68396	69686	Transposase, probable pseudogene due to aframeshift
	LSJ_2065	72929	73882	Hypothetical protein, probable pseudogene due to an internal stop codon
	LSJ_2078	88463	88739	Transposase, probable pseudogene due to frameshift
	LSJ_2109	120839	121078	pseudogene due to the introduction of a premature stop codon; label: Hypothetical protein CDS

	Gene	Start	Stop	Product
	LSJ_2145c	154333	154724	Conserved hypothetical protein; Possibly phage related; label: Conserved hypothetical protein CDS
	LSJ_2150	160058	160575	Transposase ISLasa1a, IS1223 family; COG0675 [L] Transposase and inactivated derivatives
	LSJ_2163c	172630	174788	AbpT bacteriocin export accessory protein,probable pseudogene due to frameshift
	LSJ_2168c	178086	178232	AbpIM bacteriocin immunity protein; change in the start codon
	LSJ_2176c	181578	182227	HAD-superfamily hydrolase, probable pseudogene due to frameshift
	LSJ_2211c	217055	217952	ISL3 family transposase, probable pseudogene due toframeshift
pMP1046B	LSJ_3062	42447	42641	Glutatedoxin-like protein, probable pseudogene due to truncation
	LSJ_3092c	59181	59666	DNA methylase, probable pseudogene due to truncation
pCTN1046	LSJ_5021	17463	17687	Nitroreductase family protein

Pseudogenes were characterized as such due to the presence of in-sequence frame-shifts, deletions, or interruptions of the gene by insertion sequences (IS). Genes present on the reverse strand are denoted by the suffix c following the locus tag (LSJ_XXX).

Additional file 3.2 - Transposable elements and insertion sequence (IS) elements in the *L. salivarius* JCM1046 genome.

	Gene	Start	Finish	Product
Chromosome	LSJ_0010	10563	11852	Transposase ISLasa4k, ISL3 family
	LSJ_0025c	27909	29198	Transposase ISLasa4k, ISL3 family
	LSJ_0044	48438	49193	Transposase ISLasa4k, ISL3 family
	LSJ_0066c	73984	74721	Transposase ISLasa6a, IS3 family
	LSJ_0067c	74936	75517	Transposase ISLasa7, IS3 family
	LSJ_0079c	86840	87337	Transposase ISLasa4k, ISL3 family
	LSJ_0082 (P)	88856	89753	IS1223 family transposase
	LSJ_0093c	101872	102768	IS1223 family transposase
	LSJ_0094c	102771	103289	Transposase ISLasa1b, IS1223 family
	LSJ_0175 (P)	192343	192639	Similar to transposase ISLasa3b, IS200 family; pseudogene due to frameshift
	LSJ_0247c	269207	270496	Transposase ISLasa4k, ISL3 family
	LSJ_0275c	303091	304380	Transposase ISLasa4k, ISL3 family
	LSJ_0284 (P)	311110	312330	IS605 OrfB family Transposase, putative pseudogene
	LSJ_0297	323296	324168	Transposase ISLasa4k, ISL3 family
	LSJ_0298	324254	324583	Transposase ISLasa4k, ISL3 family
	LSJ_0321c (P)	350936	352215	Transposase ISLasa9b, IS200-IS605 family, putative pseudogene
	LSJ_0322	352438	352746	Transposase ISLasa8a, IS200-IS605 family
	LSJ_0332	360874	361569	Transposase ISLasa4k, ISL3 family
	LSJ_0333	361566	361718	Transposase ISLasa4k, ISL3 family
	LSJ_0347c	378554	379399	Transposase ISLasa4k, ISL3 family
	LSJ_0363	396716	398005	Transposase ISLasa4k, ISL3 family
	LSJ_0412c	440362	441258	Transposase ISLasa1b, IS1223 family
	LSJ_0413c	441261	441779	Transposase ISLasa1b, IS1223 family
	LSJ_0433	455607	456125	Transposase ISLasa1b, IS1223 family

	Gene	Start	Finish	Product
	LSJ_0434	456267	457025	IS1223 family transposase
	LSJ_0492c	501188	501943	Transposase ISLasa4k, ISL3 family
	LSJ_0494	502738	504081	Transposase, IS605 OrfB family
	LSJ_0495c	504266	504667	Transposase ISLasa4k, ISL3 family
	LSJ_0543c	550078	550407	Transposase ISLasa4k, ISL3 family
	LSJ_0544c	550529	551362	Transposase ISLasa4k, ISL3 family
	LSJ_0560c (P)	562019	563218	Transposase ISLasa4k, ISL3 family, putative pseudogene
	LSJ_0605	599616	600905	Transposase ISLasa4k, ISL3 family
	LSJ_0628c	626592	627881	Transposase ISLasa4k, ISL3 family
	LSJ_0649c	645848	647137	Transposase ISLasa4k, ISL3 family
	LSJ_0691c (P)	688846	690133	Transposase ISLasa4k, ISL3 family, pseudogene due to frameshift
	LSJ_0695	694844	696132	Transposase ISLasa4k, ISL3 family, putative pseudogene
	LSJ_0724	726450	727868	Transposase ISLasa1b, IS1223 family
	LSJ_0781c (P)	779932	780630	IS116/IS110/IS902 family Transposase, putative pseudogene
	LSJ_0782	781473	781796	IS1181 family Transposase
	LSJ_0783	781881	782237	IS1181 family Transposase
	LSJ_0787	787340	788629	Transposase ISLasa4k, ISL3 family
	LSJ_0969	976307	976906	Transposase ISLasa4k, ISL3 family
	LSJ_0970	976984	977598	Transposase ISLasa4k, ISL3 family
	LSJ_0982 (P)	982582	983553	ISL3 family transposase, putative pseudogene
	LSJ_0983 (P)	983538	984008	ISL3 family transposase, putative pseudogene
	LSJ_1029	1029714	1031003	Transposase ISLasa4k, ISL3 family
	LSJ_1093c	1098034	1098930	Transposase ISLasa2b, IS1223 family
	LSJ_1094c	1098933	1099451	Transposase ISLasa1b, IS1223 family

	Gene	Start	Finish	Product
	LSJ_1178	1188213	1188731	Transposase ISLasa1b, IS1223 family
	LSJ_1179	1188734	1188994	IS1223 family transposase
	LSJ_1180	1189014	1189631	IS1223 family transposase
	LSJ_1199	1212717	1214006	Transposase ISLasa4k, ISL3 family
	LSJ_1260c (P)	1269915	1271133	Transposase ISLasa12, IS607 family, putative pseudogene
	LSJ_1277	1289668	1290957	Transposase ISLasa4k, ISL3 family
	LSJ_1290	1308920	1310209	Transposase ISLasa4k, ISL3 family
	LSJ_1297c	1316960	1318249	Transposase ISLasa4k, ISL3 family
	LSJ_1319c	1337562	1338458	IS1223 family transposase
	LSJ_1320c	1338461	1338979	Transposase ISLasa1b, IS1223 family
	LSJ_1333	1357371	1358660	Transposase ISLasa4k, ISL3 family
	LSJ_1387c	1420292	1420492	Transposase
	LSJ_1405	1436416	1437126	Transposase and inactivated derivatives IS30 family-like protein
	LSJ_1407c	1438091	1439209	Transposase ISLasa4k, ISL3 family
	LSJ_1409c	1444687	1445808	Transposase
	LSJ_1410	1446072	1447361	Transposase ISLasa4k, ISL3 family
	LSJ_1434c	1474467	1475303	Transposase ISLasa4k, ISL3 family
	LSJ_1444c (P)	1486638	1487535	IS1223 family transposase
	LSJ_1445 (P)	1488187	1489477	Transposase ISLasa4k, ISL3 family, putative pseudogene
	LSJ_1521 (P)	1549221	1550478	Transposase ISLasa4k, ISL3 family, putative pseudogene
	LSJ_1587c	1627386	1628186	Transposase ISLasa6a, IS3 family
	LSJ_1588c (P)	1628373	1628973	Transposase ISLasa7, IS3 family, putative pseudogene
	LSJ_1619 (P)	1664260	1664670	Transposase ISLasa1b, IS1223 family, putaive pseudogene
	LSJ_1620	1664781	1665575	IS1223 family transposase

	Gene	Start	Finish	Product
	LSJ_1625	1671258	1671719	Transposase ISLasa6a, IS3 family
	LSJ_1719c	1764886	1765458	Transposase ISLasa4k, ISL3 family
	LSJ_1720c	1765458	1766111	Transposase ISLasa4k, ISL3 family
	LSJ_1723	1767870	1769159	Transposase ISLasa4k, ISL3 family
	LSJ_1731c	1774262	1775482	IS605 OrfB family transposase
	LSJ_1770 (P)	1814660	1815094	Transposase ISLasa4k, ISL3 family, putative pseudogene
	LSJ_1771	1815314	1815949	Transposase ISLasa4k, ISL3 family
pMP1046A	LSJ_2007	6909	8126	Transposase ISLasa12, IS607 family
	LSJ_2010	12159	13448	Transposase ISL3 family
	LSJ_2058c (P)	66885	67781	Transposase IS1223 family; label: Transposase IS1223 family CDS
	LSJ_2059c	67784	68302	Transposase
	LSJ_2060c (P)	68396	69686	Transposase, probable pseudogene due to a frameshift
	LSJ_2073c	83574	84005	Transposase ISLasa8a, IS200-IS605 family
	LSJ_2074	84105	85382	Transposase ISLasa9a, IS200-IS605 family
	LSJ_2078 (P)	88463	88739	Transposase, probable pseudogene due to frameshift
	LSJ_2079	88814	89641	Transposase ISLasa15, IS3 family
	LSJ_2150 (P)	160058	160575	Transposase ISLasa1a, IS1223 family
	LSJ_2151	160658	161473	IS1223 family transposase
	LSJ_2152	161537	162544	Transposase
	LSJ_2153	162694	163983	ISL3 family transposase
	LSJ_2179c	183327	184604	IS200/IS605 family transposase
	LSJ_2180	184704	185135	IS200/IS605 family transposase
	LSJ_2200c	202805	204094	ISL3 family transposase

	Gene	Start	Finish	Product
	LSJ_2210	215500	216789	ISL3 family transposase
	LSJ_2211c (P)	217055	217952	ISL3 family transposase, probable pseudogene due to a frameshift
	LSJ_2212c	217955	218473	Transposase ISLasa1b, IS1223 family
pMP1046B	LSJ_3032	21303	22646	Transposase
	LSJ_3065	44483	45739	Transposase of the family IS605
	LSJ_3097c	61580	62734	Transposase, of the family IS605
pCTN1046	LSJ_5034	24860	25939	Transposase IS607 family protein

Pseudogenes are denoted by (P). Genes present on the reverse strand are denoted by the suffix c following the locus tag (LSJ_XXX).

Additional file 3.3 - Genes harboured by the chromosome of strain JCM1046 that are absent from the chromosome of UCC118

Gene	Start	Stop	Product
LSJ_0040	45142	45441	Hypothetical protein
LSJ_0041	45478	45873	Hypothetical protein
LSJ_0042	45890	47674	Cell division protein FtsK
LSJ_0045c	49297	49926	Hypothetical protein
LSJ_0055c	59590	60981	Glycerol-3-phosphate transporter
LSJ_0056c	61136	62263	Putative endonuclease-exonuclease-phosphatase family protein
LSJ_0057	62409	63638	Preprotein translocase SecY2 subunit
LSJ_0058	63662	65263	Accessory secretory protein Asp1
LSJ_0059	65256	66812	Accessory secretory protein Asp2
LSJ_0060	66907	67725	Accessory secretory protein Asp3
LSJ_0061	67738	70089	Protein export cytoplasm protein SecA2 ATPase RNA helicase
LSJ_0062	70109	70273	Hypothetical protein
LSJ_0063	71649	72494	Hypothetical protein
LSJ_0064	72593	73255	LPXTG-motif cell wall anchor domain protein
LSJ_0065	73448	73876	Hypothetical protein
LSJ_0068	75739	77238	Poly(glycerol-phosphate) alpha-glucosyltransferase GftA
LSJ_0069	77449	78285	Putative glycosyltransferase protein
LSJ_0070	78315	79646	GftB: Glycosyl transferase, family 8
LSJ_0071	79749	80747	Nucleotide sugar synthetase-like protein, putative
LSJ_0072	80764	81294	Hypothetical protein
LSJ_0073	81732	82076	Putative bifunctional glycosyl transferase, family 8
LSJ_0074	82154	82882	Hypothetical protein
LSJ_0075	82901	83905	Nucleotide sugar synthetase-like protein
LSJ_0076	83923	84831	Glycosyltransferase
LSJ_0077	84834	85847	Nucleotide sugar synthetase-like protein
LSJ_0078	85876	86748	Glycosyltransferase
LSJ_0083	89864	90148	Hypothetical protein
LSJ_0084	90212	91969	Glycosyl transferase, family 8
LSJ_0112c	125886	127322	Permease of the major facilitator superfamily
LSJ_0168	183403	183969	Restriction endonuclease S subunits
LSJ_0169	183970	187182	Type I restriction-modification system, restriction subunit R
LSJ_0170	187199	188380	Type I restriction-modification system, specificity subunit S
LSJ_0171	188405	190000	Type I restriction-modification system, DNA-methyltransferase subunit M
LSJ_0172	190059	190775	Type I restriction-modification system, specificity subunit S
LSJ_0173	190959	191486	Integrase
LSJ_0174c	191478	191837	Type I restriction-modification system, specificity subunit S

Gene	Start	Stop	Product
LSJ_0293	321319	321543	NAD(P)H oxidoreductase YRKL
LSJ_0294	321616	321855	Hypothetical protein
LSJ_0295	321855	322127	Hypothetical protein
LSJ_0317	346707	347423	Hypothetical secreted protein
LSJ_0318c	347491	348420	Hypothetical membrane spanning protein
LSJ_0336	365173	367875	Predicted cell-wall-anchored protein (LPXTG motif)
LSJ_0337	367895	369871	Predicted cell-wall-anchored protein (LPXTG motif)
LSJ_0472	481564	482625	Type II restriction endonuclease
LSJ_0473	482576	483361	Type II restriction endonuclease
LSJ_0474	483363	486689	Type II restriction endonuclease
LSJ_0475	486697	488676	RloF
LSJ_0493	502140	502745	First ORF in transposon ISC1904
LSJ_0494	502738	504081	Transposase, IS605 OrfB family
LSJ_0495c	504266	504667	Transposase ISLasa4k, ISL3 family
LSJ_0515	523837	524490	tRNA binding domain protein
LSJ_0516c	526247	526438	Hypothetical protein
LSJ_0517c	527464	528369	Hypothetical phage associated protein
LSJ_0518c	528615	528872	Hypothetical protein
LSJ_0519c	529049	529168	Hypothetical protein
LSJ_0520c	529171	529509	Hypothetical protein
LSJ_0521c	529732	529929	Hypothetical protein
LSJ_0522c	529945	530274	Hypothetical protein
LSJ_0523c	530394	530801	Hypothetical protein
LSJ_0524c	530815	530997	Hypothetical protein
LSJ_0525c	531011	531649	Prophage ps3 protein 13
LSJ_0526c	531662	531856	Hypothetical protein
LSJ_0527	531983	532792	Transcriptional regulator, Cro/CI family
LSJ_0528	532894	534048	Phage integrase
LSJ_0722	724523	725338	Putative membrane protein
LSJ_0723	725490	726380	Hypothetical protein
LSJ_0724	726450	727868	Transposase ISLasa1b, IS1223 family
LSJ_0725	728026	728178	Hypothetical protein
LSJ_0726	728237	728983	Hypothetical protein
LSJ_0727	729287	729448	Hypothetical protein
LSJ_0728	730174	731616	Hypothetical protein
LSJ_0729	732316	732675	Hypothetical protein
LSJ_0730	735063	735734	Phage integrase
LSJ_0782	781473	781796	IS1181 family Transposase
LSJ_0783	781881	782237	IS1181 family Transposase
LSJ_0784	782449	785184	MucBP mucin binding domain protein with YSIRK signal peptide
LSJ_0785	785216	786133	Hypothetical protein
LSJ_0786	786111	787040	Hypothetical protein
LSJ_0788	789241	790218	Choloylglycine hydrolase

Gene	Start	Stop	Product
LSJ_0789c	790568	790720	Hypothetical protein
LSJ_0790	790754	791617	Argininosuccinate synthase
LSJ_0791	791607	793883	Argininosuccinate lyase
LSJ_0792c	794365	794859	Hypothetical protein
LSJ_0798	802770	803426	CRISPR repeat RNA endoribonuclease Cas6
LSJ_0799	803447	805738	CRISPR-associated protein, Csm1 family
LSJ_0800	805740	806180	CRISPR-associated protein, Csm2 family
LSJ_0801	806200	806865	CRISPR-associated RAMP Csm3
LSJ_0802	806869	807771	CRISPR-associated RAMP protein, Csm4 family
LSJ_0803	807737	808780	CRISPR-associated protein, Csm5 family
LSJ_0804	808805	809713	CRISPR-associated protein Cas1
LSJ_0805	809716	810021	CRISPR-associated protein Cas2
LSJ_0806	812222	813457	CRISPR-associated protein Csm6
LSJ_0811c	817783	818919	Integrase
LSJ_0812c	818980	819582	Hypothetical protein
LSJ_0813c	819634	820782	DNA polymerase III alpha subunit
LSJ_0814c	820784	820981	Hypothetical protein
LSJ_0815c	820987	821565	Hypothetical protein
LSJ_0816c	821597	822127	Phage protein
LSJ_0817c	822153	822566	Hypothetical protein
LSJ_0818	822565	822768	Hypothetical protein
LSJ_0819c	822819	823244	Hypothetical protein
LSJ_0820c	823258	823683	Transcriptional regulator, Cro/CI family
LSJ_0821	823869	824072	Hypothetical protein
LSJ_0862	861288	861542	Hypothetical protein
LSJ_0863	861526	861696	Hypothetical protein
LSJ_0864	861706	863631	Peptidase
LSJ_0976c	979278	979565	Hypothetical protein
LSJ_0977c	979566	980927	HNH endonuclease
LSJ_0978c	981171	981380	Hypothetical protein
LSJ_1308c	1326871	1327308	CopAB ATPases metal-fist type repressor
LSJ_1309c	1327703	1328032	LPXTG-motif cell wall anchor domain protein
LSJ_1310c	1328089	1328616	Hypothetical protein
LSJ_1326	1345344	1345901	Conserved hypothetical protein
LSJ_1327c	1345976	1347124	Oxidoreductase, FAD/FMN-binding protein
LSJ_1329c	1350681	1352627	ATP-dependent DNA helicase Rep
LSJ_1330c	1352612	1354066	Hypothetical protein
LSJ_1331c	1354230	1354745	Hypothetical protein
LSJ_1332c	1354936	1357230	Type I restriction-modification system, restriction subunit R
LSJ_1334c	1358657	1359472	Type I restriction-modification system, restriction subunit R
LSJ_1335c	1359474	1360652	Type I restriction-modification system, specificity subunit S

Gene	Start	Stop	Product
LSJ_1336c	1360665	1362752	Type I restriction-modification system, DNA-methyltransferase subunit M
LSJ_1337c	1363043	1363354	Hypothetical protein
LSJ_1338c	1363457	1364314	Mrr restriction system protein
LSJ_1339	1364656	1364817	Hypothetical protein
LSJ_1340c	1365017	1366291	Hypothetical protein
LSJ_1341c	1366407	1367594	Type I restriction-modification system, specificity subunit S
LSJ_1342c	1367598	1369226	Type I restriction-modification system, DNA-methyltransferase subunit M
LSJ_1343c	1369232	1372306	Putative HsdR,a host restriction; endonuclease R
LSJ_1344c	1372339	1372560	Hypothetical protein
LSJ_1345c	1372574	1375189	DEAD/DEAH box helicase domain protein
LSJ_1346c	1375194	1376501	Hypothetical protein
LSJ_1347	1376883	1377410	DNA-directed RNA polymerase specialized sigma subunit
LSJ_1348	1377699	1378031	Hypothetical protein
LSJ_1349	1378015	1379157	Phage protein
LSJ_1350	1379159	1379710	Phage protein
LSJ_1351	1379761	1381695	Phage encoded DNA polymerase I
LSJ_1352	1382057	1382188	Hypothetical protein
LSJ_1353	1382191	1384440	DNA primase/helicase, phage-associated
LSJ_1354	1384897	1386279	DNA helicase, phage-associated
LSJ_1355	1386248	1386718	Hypothetical protein
LSJ_1356	1387347	1387889	Hypothetical protein
LSJ_1357	1387889	1389118	Adenine-specific methyltransferase
LSJ_1358	1389190	1389822	Virulence-related protein
LSJ_1359	1390084	1391685	Phage terminase large subunit
LSJ_1360	1391713	1391874	Hypothetical protein
LSJ_1361	1392236	1393492	Phage portal protein
LSJ_1362	1393489	1394151	Prophage Clp protease-like protein
LSJ_1363	1394172	1395350	Phage major capsid protein
LSJ_1364	1395641	1396021	Bacteriophage head-tail adaptor
LSJ_1365	1396206	1396433	Holin, toxin secretion/phage lysis
LSJ_1366	1396531	1396719	Hypothetical protein
LSJ_1367	1396782	1398332	Site-specific recombinase, DNA invertase Pin related protein
LSJ_1368	1398319	1398723	Hypothetical protein
LSJ_1369	1398710	1400296	Site-specific recombinase
LSJ_1370c	1400540	1401913	tRNA (Uracil-5-)-methyltransferase
LSJ_1386c	1418989	1420221	Integrase
LSJ_1387c	1420292	1420492	Transposase
LSJ_1388c	1420543	1421109	Hypothetical protein
LSJ_1389c	1421122	1421355	Hypothetical protein
LSJ_1390c	1421522	1422967	Hypothetical protein

Gene	Start	Stop	Product
LSJ_1391c	1422960	1424462	Phage-like protein
LSJ_1392	1424748	1425617	Hypothetical protein
LSJ_1393c	1425886	1426833	Hypothetical protein
LSJ_1394c	1426975	1427373	Phage-like protein
LSJ_1395	1427531	1428106	Site-specific recombinase, DNA invertase Pin related protein
LSJ_1396	1428749	1429390	Resolvase
LSJ_1397	1429387	1429887	Hypothetical protein
LSJ_1398	1430311	1430535	Hypothetical protein
LSJ_1399	1430927	1431415	Mobile element protein
LSJ_1400	1431535	1432407	Mobile element protein
LSJ_1401c	1432489	1434318	Putative mucus binding protein
LSJ_1402c	1434381	1434974	Extracellular protein
LSJ_1403c	1435002	1435238	Extracellular protein
LSJ_1404	1435553	1435894	Putative transcriptional regulator
LSJ_1405	1436416	1437126	Transposase and inactivated derivatives IS30 family-like protein
LSJ_1406c	1437141	1437971	Mobile element protein
LSJ_1408c	1439831	1444300	Adhesion exoprotein
LSJ_1409c	1444687	1445808	Transposase
LSJ_1508c	1536385	1537395	Hypothetical protein
LSJ_1509c	1537562	1539538	Surface protein
LSJ_1510c	1539563	1540096	Hypothetical protein
LSJ_1511c	1540086	1540616	Hypothetical protein
LSJ_1541c	1564112	1567138	Hypothetical membrane spanning protein
LSJ_1542c	1567308	1568255	Conserved hypothetical protein
LSJ_1543c	1568292	1569131	Hypothetical protein
LSJ_1564c	1592762	1595578	N-acetylmuramoyl-L-alanine amidase
LSJ_1569	1600661	1603012	Hypothetical exported protein
LSJ_1570c	1603232	1604200	Hypothetical protein
LSJ_1571c	1604611	1605084	Hypothetical protein
LSJ_1574	1608610	1608870	Hypothetical protein
LSJ_1575	1608992	1609177	Hypothetical protein
LSJ_1576c	1609250	1610986	Hypothetical protein
LSJ_1577	1611485	1612540	O-acetyl transferase
LSJ_1585c	1622062	1625673	DNA-binding protein
LSJ_1586c	1626319	1627218	Hypothetical protein
LSJ_1600c	1643309	1646074	Hypothetical exported protein
LSJ_1601c	1646287	1647282	beta-1,6-galactofuranosyltransferase
LSJ_1602c	1647290	1648477	Hypothetical membrane spanning protein
LSJ_1603c	1648493	1649917	Oligosaccharide translocase
LSJ_1604c	1650000	1650992	Glycosyltransferase
LSJ_1605c	1651011	1652123	UDP-galactopyranose mutase
LSJ_1606c	1652145	1653029	Glycosyltransferase

Gene	Start	Stop	Product
LSJ_1611c	1656680	1657474	Chain length regulator
LSJ_1612c	1657574	1658824	Putative polysaccharide polymerase wzy
LSJ_1613c	1658844	1659812	dTDP-rhamnosyl transferase rfbF
LSJ_1614c	1659831	1660829	Capsular polysaccharide phosphotransferase wcwK
LSJ_1615c	1660814	1661890	Beta-1,3-glucosyltransferase
LSJ_1616c	1661932	1662432	Glycosyltransferase
LSJ_1617c	1662450	1662908	UDP-N-acetylglucosamine:LPS N-acetylglucosamine transferase
LSJ_1618c	1663384	1664148	Hypothetical protein
LSJ_1621c	1665697	1666575	Site-specific tyrosine recombinase/Phage integrase
LSJ_1622c	1667357	1668727	Hypothetical protein
LSJ_1623c	1668874	1670019	Transcriptional regulator, XRE family
LSJ_1624c	1670019	1670723	Hypothetical protein
LSJ_1625	1671258	1671719	Transposase ISLasa6a, IS3 family
LSJ_1631c	1676627	1677553	Alpha-L-Rha alpha-1,3-L-rhamnosyltransferase
LSJ_1632c	1677583	1678593	Glycosyltransferase
LSJ_1683c	1727476	1728180	Choline binding protein A
LSJ_1684	1728354	1729700	NADH oxidase
LSJ_1731c	1774262	1775482	IS605 OrfB family transposase
LSJ_1751	1795598	1797412	Glycerophosphodiester phosphodiesterase
LSJ_1754c	1799529	1801145	Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA
LSJ_1755c	1801172	1801783	Prolyl aminopeptidase
LSJ_1756c	1801774	1802070	Proline aminopeptidase
LSJ_1757c	1802526	1803221	Fumarate reductase flavoprotein subunit
LSJ_1758	1803244	1803519	LysR family transcriptional regulator
LSJ_1759	1803597	1804637	Alcohol dehydrogenase
LSJ_1760c	1804694	1805992	O-acetylhomoserine sulfhydrylase / O-succinylhomoserine sulfhydrylase
LSJ_1761c	1806008	1806859	Homoserine O-succinyltransferase
LSJ_1762c	1806889	1808022	5-methyltetrahydropteroyltriglutamate/homocysteine S-methyltransferase
LSJ_1763c	1808475	1809158	Oxygen-insensitive NADPH nitroreductase

Regions of diversity were determined using Panseq [59]. Pseudogenes are denoted by (P). Genes present on the reverse strand are denoted by the suffix c following the locus tag (LSJ_XXX)

Additional file 3.4 -Genes harboured by the chromosome of strain JCM1046 that are absent from other *L. salivarius* sequenced genomes.

Gene	Start	Stop	Product
LSJ_0517c	527464	528369	Hypothetical phage associated protein
LSJ_0526c	531662	531856	Hypothetical protein
LSJ_0725	728026	728178	Hypothetical protein
LSJ_0729	732316	732675	Hypothetical protein
LSJ_1329c	1350681	1352627	ATP-dependent DNA helicase Rep
LSJ_1330c	1352612	1354066	Hypothetical protein
LSJ_1331c	1354230	1354745	Hypothetical protein
LSJ_1340c	1365017	1366291	Hypothetical protein
LSJ_1341c	1366407	1367594	Type I restriction-modification system, specificity subunit S
LSJ_1342c	1367598	1369226	Type I restriction-modification system, DNA-methyltransferase subunit M
LSJ_1343c	1369232	1372306	Putative HsdR,a host restriction; endonuclease R
LSJ_1344c	1372339	1372560	Hypothetical protein
LSJ_1345c	1372574	1375189	DEAD/DEAH box helicase domain protein
LSJ_1346c	1375194	1376501	Hypothetical protein
LSJ_1347	1376883	1377410	DNA-directed RNA polymerase specialized sigma subunit
LSJ_1348	1377699	1378031	Hypothetical protein
LSJ_1349	1378015	1379157	Phage protein
LSJ_1350	1379159	1379710	Phage protein
LSJ_1351	1379761	1381695	Phage encoded DNA polymerase I
LSJ_1352	1382057	1382188	Hypothetical protein
LSJ_1353	1382191	1384440	DNA primase/helicase, phage-associated
LSJ_1354	1384897	1386279	DNA helicase, phage-associated
LSJ_1355	1386248	1386718	Hypothetical protein
LSJ_1356	1387347	1387889	Hypothetical protein
LSJ_1357	1387889	1389118	Adenine-specific methyltransferase
LSJ_1358	1389190	1389822	Virulence-related protein
LSJ_1359	1390084	1391685	Phage terminase large subunit
LSJ_1360	1391713	1391874	Hypothetical protein
LSJ_1361	1392236	1393492	Phage portal protein
LSJ_1362	1393489	1394151	Prophage Clp protease-like protein
LSJ_1363	1394172	1395350	Phage major capsid protein
LSJ_1364	1395641	1396021	Bacteriophage head-tail adaptor
LSJ_1365	1396206	1396433	Holin, toxin secretion/phage lysis
LSJ_1366	1396531	1396719	Hypothetical protein
LSJ_1367	1396782	1398332	Site-specific recombinase, DNA invertase Pin related protein
LSJ_1368	1398319	1398723	Hypothetical protein
LSJ_1369	1398710	1400296	Site-specific recombinase
LSJ_1508c	1536385	1537395	Hypothetical protein
LSJ_1570c	1603232	1604200	Hypothetical protein

Gene	Start	Stop	Product
LSJ_1571c	1604611	1605084	Hypothetical protein
LSJ_1585c	1622062	1625673	DNA-binding protein
LSJ_1612c	1657574	1658824	Putative polysaccharide polymerase wzy
LSJ_1613c	1658844	1659812	dTDP-rhamnosyl transferase rfbF
LSJ_1614c	1659831	1660829	Capsular polysaccharide phosphotransferase wcwK
LSJ_1615c	1660814	1661890	Beta-1,3-glucosyltransferase
LSJ_1621c	1665697	1666575	Site-specific tyrosine recombinase/Phage integrase
LSJ_1622c	1667357	1668727	Hypothetical protein
LSJ_1623c	1668874	1670019	Transcriptional regulator, XRE family

Gene presence/absence was determined by BLASTP sequence comparisons. Pseudogenes are denoted by (P). Genes present on the reverse strand are denoted by the suffix c following the locus tag (LSJ_XXX).

Additional file 3.5 -Exopolysaccharide gene cluster present on the chromosome of *Lactobacillus salivarius* JCM1046.

Gene	Start position	End position	Gene product
LSJ_1603c	1648493	1649917	Oligosaccharide translocase
LSJ_1604c	1650000	1650992	Glycosyltransferase
LSJ_1605c	1651011	1652123	UDP-galactopyranose mutase
LSJ_1606c	1652145	1653029	Glycosyltransferase
LSJ_1607c	1653073	1653738	Undecaprenyl-phosphate beta-glucosephosphotransferase
LSJ_1608c	1653818	1654789	Transcriptional regulator, LytR family
LSJ_1609c	1654809	1655585	Phosphotyrosine-protein phosphatase
LSJ_1610c	1655937	1656662	Tyrosine-protein kinase
LSJ_1611c	1656680	1657474	Chain length regulator
LSJ_1612c	1657574	1658824	putative polysaccharide polymerase wzy
LSJ_1613c	1658844	1659812	dTDP-rhamnosyl transferase rfbF/Glycosyltransferase
LSJ_1614c	1659831	1660829	Capsular polysaccharide phosphotransferase wcwK
LSJ_1615c	1660814	1661890	Beta-1,3-glucosyltransferase
LSJ_1616c	1661932	1662432	Glycosyltransferase
LSJ_1617c	1662450	1662908	UDP-N-acetylglucosamine:LPS N-acetylglucosamine transferase
LSJ_1618c	1663384	1664148	DNA helicase
LSJ_1619 (P)	1664260	1664670	Transposase ISLasa1b, IS1223 family, pseudogene due to a truncation
LSJ_1620	1664781	1665575	IS1223 family transposase
LSJ_1621c	1665697	1666575	Site-specific tyrosine recombinase/ Phage integrase
LSJ_1622c	1667357	1668727	hypothetical protein
LSJ_1623c	1668874	1670019	Transcriptional regulator, XRE family
LSJ_1624c	1670019	1670723	hypothetical protein
LSJ_1625	1671258	1671719	Transposase ISLasa6a, IS3 family
LSJ_1626c (P)	1671775	1672836	Beta-hexosaminidase, pseudogene due to an inserted stop codon
LSJ_1627c	1673164	1674000	dTDP-4-dehydrorhamnose reductase
LSJ_1628c	1674056	1675087	dTDP-glucose 4,6-dehydratase
LSJ_1629c	1675099	1675680	dTDP-4-dehydrorhamnose 3,5-epimerase
LSJ_1630c	1675683	1676552	Glucose-1-phosphate thymidylyltransferase
LSJ_1631c	1676627	1677553	Alpha-L-Rha alpha-1,3-L-rhamnosyltransferase
LSJ_1632c	1677583	1678593	Glycosyltransferase
LSJ_1633c	1678675	1679487	Glycosyltransferase

Pseudogenes are denoted by (P). Genes present on the reverse strand are denoted by the suffix c following the locus tag (LSJ_XXX)

Additional file 3.6 - Genes harboured by pCTN1046

Systematic ID	Gene name	Start-Stop nt position	Product
LSJ_5000		194-313	Tn916-related conserved hypothetical protein
LSJ_5001		432-1817	Conjugative transposon-related FtsK/SpoIIIE family protein
LSJ_5002		1995-3200	Replication initiation factor
LSJ_5003		3243-3464	Hypothetical protein
LSJ_5004		3581-4078	Antirestriction (ArdA) protein/ Conjugative transposon protein
LSJ_5005		4053-4559	Conjugative transposon membrane protein, TcpE family
LSJ_5006		4543-6990	Conjugative transposon ATP/GTP-binding protein
LSJ_5007		6993-9170	Conjugative transposon membrane protein
LSJ_5008		9167-10168	TN916, NLP/P60 family protein
LSJ_5009		10165-11097	Tn916-related hypothetical protein
LSJ_5010		11372-11458	Tet (M) leader peptide
LSJ_5011	*TetM*	11474-13393	Tetracycline resistance protein TetM
LSJ_5012		13491-13679	Tn916-related hypothetical protein
LSJ_5013c		13739-14092	Putative DNA binding protein,
LSJ_5014		14297-14368	Conserved Tn916-related hypothetical protein
LSJ_5015		14597-15019	RNA polymerase sigma factor, sigma-70 family, region4
LSJ_5016		15016-15246	Tn916 hypothetical protein containing a helix-turn-helix motif
LSJ_5017c		15310-15438	Hypothetical protein
LSJ_5018c		15472-15723	Tn916-related conserved hypothetical protein
LSJ_5019	*Xis*	15707-15910	Tn916 excisionase
LSJ_5020	*Int*	15992-17209	Tn916 integrase
LSJ_5021(P)		17463-17687	Nitroreductase family protein
LSJ_5022		17684-17962	Hypothetical protein
LSJ_5023		17967-18158	Major facilitator family protein
LSJ_5024c		18241-18345	Bacteriocin-like protein
LSJ_5025		18608-18832	Hypothetical protein
LSJ_5026		18880-19119	Hypothetical protein
LSJ_5027		19116-19559	Hypothetical membrane spanning protein
LSJ_5028		19540-20862	Hypothetical membrane-associated protein
LSJ_5029c		21048-21800	Autolysin/Muramidase
LSJ_5030c		21793-22878	Replication initiation factor family protein
LSJ_5031		23627-24142	Hypothetical membrane protein
LSJ_5032c		24368-24643	Hypothetical protein
LSJ_5033c		24746-24859	Hypothetical protein
LSJ_5034		24860-25939	Transposase IS607 family protein
LSJ_5035c	*repB*	28686-29894	Replication protein RepB

Systematic ID	Gene name	Start-Stop nt position	Product
LSJ_5036c		31004-31519	Hypothetical protein
LSJ_5037c		31524-32021	Hypothetical protein
LSJ_5038c		32268-32570	Hypothetical protein
LSJ_5039c		32584-33153	Resolvase/Site-specific recombinase

Pseudogenes are denoted by (P). Genes present on the reverse strand are denoted by the suffix c following the locus tag (LSJ_XXX)

Additional file 3.7 - Primers used to generate Southern Hybridization probes.

Primer	Sequence 5'-3'	Target Replicon
repB_F	cggctaagatggatgttgtacc	pMP1046B
repB_R	cttactgctgctattctcaattgtc	pMP1046B
C77_F	gctgaaataacattgtcgaaggg	pLMP1046
C32_R	ggagcggtagataatatggtg	pLMP1046
Xis_F	atgaagcagactgacattcctatttgg	pCTN1046
Int_R	tttcatttctgccattgtagaatcg	pCTN1046

3.10 LIST OF ABBREVIATIONS USED

aa: amino acid

ACT: Artemis comparison tool

BLAST: Basic Local Alignment Search Tool

bp: Base pairs

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats

CAS: CRISPR-associated sequence

DR: direct repeat

EPS: Exopolysaccharide

GIT: Gastrointestinal tract

ID: identity

IS: insertion sequence

LAB: Lactic Acid Bacteria

LDL: low-density lipoprotein

NCBI: National Center for Biotechnology Information

PCR: polymerase chain reaction

nr: Nonredundant protein database

nt: Nucleotides

sp: species

TIR: terminal inverted repeat

CHAPTER IV

General Discussion and Future Perspectives

Chapter IV Table of Contents

4.1. General Discussion and Future Perspectives 241
4.2 References 252

4.1. General Discussion and Future Perspectives

The aim of this thesis was to fully characterise the level of genomic diversity of *L. salivarius* and to provide a genetic overview of this species. Prior to the studies presented herein, the genomic diversity of *L. salivarius* had been poorly characterized and its implications for strain selection had not been examined. Chapter II provided an in-depth genotypic and phenotypic analysis of *L. salivarius* and examined the phylogenetic structure of this species in detail. Chapter III presents the genome sequence of the porcine *L. salivarius* isolate JCM1046, which includes the first sequence of a linear plasmid in a *Lactobacillus* species and the first description of a Tn*916*-like element (Tn*6224*) in a *L. salivarius* genome.

Central to these studies was a collection of strains that were assembled from a range of ecological sources and which were known to be diverse in their plasmid content and metabolic capabilities. As the most thoroughly characterised strain of the species, the probiotic strain *L. salivarius* UCC118 acted as a genomic reference for the studies carried out in Chapters II and III. In addition to providing the genomic basis for the design of *L. salivarius* microarrays employed in Chapter II, strain UCC118 was also used as a reference for the assembly of the scaffolds of the genome of strain JCM1046 in Chapter III.

It is widely recognised that the phenotypic characteristics of the lactobacilli are both species and strain specific traits (Bron *et al.*, 2013; Lebeer *et al.*, 2008; Salminen *et al.*, 2004b). It is important that species of lactobacilli that are used in food or health-promoting products are rigorously characterised. The increasing interest in *Lactobacillus salivarius* as a candidate probiotic (Messaoudi *et al.*, 2013; Neville & O'Toole, 2010) merited an in-depth analysis of the genetic basis of its strain specific properties (Li *et al.*, 2006; Ryan *et al.*, 2008).

The research described in Chapter II provided the first genome-wide comparative analysis of *L. salivarius* strains. The panel of strains studied were representative of the species as a whole and reflect the range of habitats from which *L. salivarius* strains are frequently recovered, including human, animal, and environmental sources.

Comparative genomic hybridization (CGH) and multilocus sequence typing (MLST) were employed in Chapter II to assess the level of genomic diversity in *L. salivarius*. Our previous study (Li *et al.*, 2006) which resulted in the unification of

the original *L. salivarius* subspecies demonstrated the suitability of the housekeeping gene *groES* as a means of resolving intra-specific differences within *L. salivarius*. The study described in Chapter II is the first to evaluate the discriminatory power of MLST when applied to this species. In comparison to the CGH analysis, MLST produced a more complete separation of the strains on the basis of their isolation origins. However, niche associations were not unequivocally supported by the cluster analysis of strains generated by either technique.

Up to 23.6 % of the gene content of *L. salivarius* strains was found to be variable when compared to the UCC118 genome. The most conserved strains were human GIT isolates, while the greatest level of divergence was identified in animal associated isolates. The extent of diversity observed in *L. salivarius* is higher than that revealed by CGH of *Lactobacillus plantarum* (up to 20% gene divergence; (Molenaar *et al.*, 2005)) and *L. casei* (up to 19% gene divergence; (Cai *et al.*, 2009)) and *L. johnsonii* (17% divergence; (Berger *et al.*, 2007)). As evidenced in Chapter II, plasmid-encoded genes contribute significantly to this increased level of sequence diversity among the *L. salivarius* strains, emphasising the potential capacity of the extrachromosomal replicons to influence the phenotypic characteristics of their hosts.

Additionally, the data in Chapter II suggests that *L. salivarius* is undergoing a process of genome degradation which is most likely the result of niche adaptation, as has been observed in other LAB species (Makarova *et al.*, 2006).

EPS has a number of biologically significant roles in commensal lactobacilli including stress resistance, adhesion, and interaction with the immune system (Lebeer *et al.*, 2008; Lebeer *et al.*, 2011). One of the most prominent areas of gene variation among the *L. salivarius* strains were the two gene clusters associated with exopolysaccharide (EPS) biosynthesis in the UCC118. Production of EPS has been well characterised for other commensal lactobacilli (Badel *et al.*, 2011), but not for *L. salivarius*. Therefore, the distribution and relatedness of genes responsible for EPS biosynthesis were examined in greater detail in Chapter II and then correlated to the EPS production capabilities of the *L. salivarius* strains. Production levels of bound or released EPS did not appear to be related to either the source of isolation or the conservation of EPS cluster 1 or EPS cluster 2, but were highly dependent on the culture conditions and was most significantly modulated by the available carbohydrate as is often the case in EPS-producing lactobacilli (Badel *et al.*, 2011).

In Chapter II we suggest that the EPS-producing phenotype is likely due to the presence of an additional functional EPS-related cluster of genes that are present in a subset of the *L. salivarius* strains but that are absent in UCC118 and therefore absent from the presented CGH data. This represents a significant limitation of using a microarray design based on the genome sequence of a single strain. This limitation is mitigated when using multi-strain arrays (Siezen *et al.*, 2011) or a single reference genome sequence with the inclusion of additional niche-specific genes. As the cost of whole genome sequencing continues to fall, the sequencing of multiple strains of individual species has become more feasible and, subsequent to the work in Chapters II and III, has been carried out for *L. salivarius* in the lab of Prof. O'Toole.

The highest EPS-r production levels of any strain were achieved when CCUG44481 was grown in the presence of sucrose (3%), a substrate which is known to be capable of supporting high EPS production levels in many *Lactobacillus* species (van Geel-Schutten *et al.*, 1998). It should be noted that the production yield of homopolysaccharides in lactobacilli is generally under 1 g/L when culture conditions are not optimized and less still for the majority of heteropolysaccharides (Badel *et al.*, 2011). The highest *Lactobacillus* EPS yields are generally attributed to *L. rhamnosus*, with strain RW-9595M reported to synthesize approximately 2.7 g/L of polysaccharide (Macedo *et al.*, 2002). The unusually high CCUG44481 EPS-r yield may be a feature of the screening method used for handling a large number of strains. Although sufficient for initial evaluation of EPS production levels within *L. salivarius*, it is recommended that the protocol used to assess EPS-r yields (Chapter II) be optimised further for strains of particular interest. In order to increase the purity of the EPS-r extracted, a revised protocol should include an increase in the number of ethanol precipitations (from 1 to 3) of the EPS-r sample in addition to an increase in the duration of dialysis (increase from 24hrs to 72hrs) of the EPS-r sample in water.

Given the potential health promoting properties and industrial applications associated with EPS-producing lactobacilli (Jolly *et al.*, 2002), it is possible that strain CCUG44481 or its purified EPS-r could be exploited for commercial purposes. The EPS-r production levels observed in this thesis (Chapter II) fall below the quantities that are generally thought to be economically feasible (10-15 g/L) (van den Berg *et al.*, 1995) for industrial applications. However, the culture conditions used to promote EPS production (Chapter II) were designed to support growth and

production of EPS in a large group of *L. salivarius* strains, rather than being tailored for maximal EPS-r production levels in CCUG44481 specifically. It is well recognised that the availability of carbon and nitrogen sources in addition to the physico-chemical conditions under which lactobacilli are grown have a significant influence on the levels of EPS production (Badel *et al.*, 2011). It is likely that further optimization of culture conditions may increase the CCUG44481 EPS-r production levels to industrially-viable levels. Functional analysis of *L. salivarius*-derived EPS would benefit from both structural and chemical analysis of its purified EPS, and is certainly worth pursuing.

Adherence and biofilm formation can increase the gut-residence time of commensal strains, as well as promote pathogen exclusion, host-cell interaction, and immune stimulation (Servin & Coconnier, 2003; Servin, 2004). Adhesion and biofilm formation has previously been shown to be a strain-specific trait in lactobacilli and is known to be influenced by EPS production (Vélez *et al.*, 2007). Indeed, exopolysaccharides act as important mediators of biofilm formation in the human oral cavity (Parisotto *et al.*, 2010) but it is not fully understood what role EPS plays in the colonisation of the GIT by lactobacilli. The variability in EPS-production levels described in Chapter II led us to examine the capacity of *L. salivarius* to form biofilms and to examine the relationship between EPS production and biofilm formation in *L. salivarius*. Strains UCC118 (human isolate, low-EPS producer, probiotic, reference strain) and CCUG44481, (animal isolate, high EPS producer) were examined for their ability to form biofilms, in medium which was shown to modify EPS production levels (Chapter II). Biofilm formation was benchmarked against a known biofilm-forming probiotic strain *L. rhamnosus* GG (LGG) and as is the case for LGG (Lebeer *et al.*, 2007) biofilm formation in *L. salivarius* was found to be strongly modulated by the available carbon source and the culture conditions employed in Chapter II. UCC118 consistently formed the thickest biofilms and CCUG44481 was generally incapable of forming substantial biofilms under the experimental conditions tested. It is plausible that the substantial biofilms formed by UCC118 contribute to host-interaction properties observed in this strain, as has previously been described in probiotic strain *L. reuteri* ATCC 55730 (Jones & Versalovic, 2009).

In vitro biofilm formation assays have been described as a simplified model system for the *in vivo* state due to the absence of a competing microbiota, the use of

abiotic supports, and the constant supply of a carbon source (Su & Gänzle, 2014). However, the formation of biofilms and the co-aggregation of *L. reuteri* TMW1.106, LTH5448 and 100-23 *in vitro* was generally representative of their competitiveness and ability to form biofilms *in vivo* (Frese *et al.*, 2011; Sims *et al.*, 2011; Su & Gänzle, 2014; Walter *et al.*, 2008). The ability of *L. reuteri* to colonize animals involves its capacity to form biofilms on non-secretory stratified squamous epithelia in the upper GIT, which has been demonstrated in the fore-stomachs of mice (Tannock *et al.*, 2005) and horses (Yuki *et al.*, 2000), and the crops of birds (Fuller & Brooker, 1974). It remains to be determined whether *L. salivarius* has the capacity to form biofilms in *vivo*. It would be interesting to examine the role that alternate growth conditions have on the ability of *L. salivarius* strains to form *in vivo* biofilms, including those which are particularly relevant for survival in the gastrointestinal environment, including low pH, high osmolarity and presence of bile, mucins and non-digestible polysaccharides. Additionally, the observation of mushroom-like protrusions from the surface of the *L. salivarius* biofilm mat is the first description of its kind for a *Lactobacillus* biofilm and provides insight into the architecture of biofilms of commensal lactobacilli.

LGG produces high-molecular-weight, galactose-rich heteropolymeric EPS. Mutation of the *welE* gene (priming glycosyltransferase) resulted in an EPS-deficient mutant strain of LGG, which was found to have increased adherence and biofilm formation capacity. Similarly, the capacity of *L. salivarius* strains to form a biofilm was found to be inversely related to production of EPS-r. This is consistent with the idea that EPS may have a shielding effect on surface components that are responsible for the adhesive properties of some strains of *L. salivarius*. The fructosyl transferase gene (*ftf*) is necessary for EPS production in *Lactobacillus reuteri* (Sims *et al.*, 2011). An EPS-negative mutant (Δ*ftf)* of *L. reuteri* 100-23 retains its ability to colonise the murine GIT in the absence of competition, but is impaired when in competition with the wild type (Sims *et al.*, 2011). Interestingly, the EPS-deficient mutant did not lose its ability to form a biofilm on the fore-stomach epithelial surface (Sims *et al.*, 2011). Similarly, colonisation of the fore-stomach of mice by *L. reuteri* TMW1.106 was unaffected when the *ftf* gene was mutated (Walter *et al.*, 2008). Together, these results indicate that despite the involvement of EPS in the colonisation of the murine GIT by *L. reuteri*, the capacity of *L. reuteri* to form biofilms *in vivo* is independent of the production of EPS (Sims *et al.*, 2011).

In natural ecosystems, bacterial biofilms form multispecies communities and it is possible that *L. salivarius* may exhibit alternate phenotypes under *in vivo* conditions. *L. salivarius* is frequently isolated from dental plaque and caries (Piwat *et al.*, 2010; Švec *et al.*, 2009) and also the titanium surfaces of dental implants (Godoy-Gallardo *et al.*, 2014). A study of the probiotic strain *L. salivarius* W24 indicated that it was incapable of forming a monospecies biofilm *in vitro*, but was able to establish itself into saliva-derived communities if co-inoculated with additional oral isolates, thus leading the authors to suggest that an interaction with other microorganisms or their products was necessary for biofilm formation by strain W24 (Pham *et al.*, 2009).

It has recently been shown that biofilm formation in *L. reuteri* 100-23 is regulated by the two-component regulatory systems encoded by the *bfrKRT* and *cemAKR* operons, the effects of which are dependent upon available carbon sources (Su & Gänzle, 2014). In order to determine the mechanisms by which *L. salivarius* interacts with the cells of the GIT and its resident microbiota, it is important that we gain a better understanding of how it interacts with its environment. It is therefore desirable to identify the genes and modes of regulation involved in EPS production and biofilm formation in *L. salivarius* strains, both *in vitro* and (potentially) *in vivo*.

The capacity of *L. salivarius* strains to form biofilms is not dependent on an EPS producer phenotype and it is likely that biofilm formation in UCC118 is mediated by alternative cell-surface molecules. For example, the adhesion of *L. reuteri* to the murine intestinal tract was found to be mediated by a large surface protein (Frese *et al.*, 2011; Walter *et al.*, 2005). Additionally, adherence of *L. reuteri* to porcine intestinal mucus or human epithelial cells was shown to be mediated by the mucus adhesion-promoting protein MapA, (Miyoshi *et al.*, 2006), and the mucus-binding protein (Mub) (Roos & Jonsson, 2002). An adhesin protein MabA has been shown to be a modulator of adhesion and biofilm production capacity in LGG. A *mabA* knockout mutant demonstrated a reduced biofilm formation capacity on abiotic surfaces and has shown a reduced ability to adhere to intestinal epithelial cells and tissues of the murine GIT (Vélez *et al.*, 2010). Subsequently, pili present on the cell surface of LGG have been shown to be key to the efficient adherence of LGG to the intestinal epithelia (Lebeer *et al.*, 2012). It will be important when moving this work forward to examine the role that additional surface molecules play in the formation of *L. salivarius* biofilms.

Collins *et al.* recently identified a megaplasmid encoded gene (CCUG_2371) in strain CCUG47825 which plays a crucial role in mediating biofilm formation and biofilm stability in this strain (Collins *et al.*, 2012). This gene encodes a fibrinogen binding surface protein and homologues of this gene are widespread (21/31 strains) among the *L. salivarius* isolates studied in Chapter II. Interestingly, despite lacking a copy of this gene, UCC118 is capable of forming robust biofilms (Chapter II). Collins *et al* suggest that *L. salivarius* strains are capable of utilizing alternative mechanisms for biofilm formation when a copy of CCUG_2371 is absent from the genome (Collins *et al.*, 2012).

Despite their long history of safe consumption, lactobacilli can cause rare cases of bacteraemia and endocarditis, usually in immunocompromised patients (Husni *et al.*, 1997). Cases of sepsis and endocarditis have also been associated with therapeutic administration of probiotic lactobacilli (Land *et al.*, 2005; Mackay *et al.*, 1999; Salminen *et al.*, 2004a). *L. salivarius* was reported to be the causative organism in a case of endocarditis in a 10 year old girl (Berger *et al.*, 1976) and a case of bacteremic cholecystitis in a 70 year old man (Woo *et al.*, 2002). Of the *L. salivarius* strains examined in Chapter II, nine were isolated from blood or tissue (e.g., gallbladder or pus). These clinical isolates did not form a separate MLST cluster, nor were they found to be genetically distinct from non-clinical isolates based on their gene content as determined by CGH. This suggested that the rare cases of septicaemia caused by *L. salivarius* are more likely to be due to compromised host defences, rather than specialized "pathogenic strains," as has been shown for other *Lactobacillus* species (Salminen *et al.*, 2004a). It is suggested in Chapter II that the disease isolates of *L. salivarius* may harbour additional genes required for enhanced pathogenic potential, which would not be identified in the CGH data described in Chapter II.

As mentioned above, gene CCUG_2371 was identified in the septicaemia isolate CCUG47825 and was found to confer a fibrinogen-binding phenotype and cause aggregation of human platelets under *in vitro* conditions (Collins *et al.*, 2012). Homologues of this gene were found in 8/9 clinical strains of *L. salivarius* tested (Collins *et al.*, 2012), all of which were test isolates in Chapter II. Although absent from UCC118, strains harbouring CCUG_2371 homologues were most frequently isolated from human sources (19 of 21 strains). Although gene expression was not observed under the conditions tested, the authors suggest that strains of *L. salivarius*

isolated from non-human sources may be safer for use as probiotic candidates due to the absence of potentially pathogenic genes (Collins *et al.*, 2012). Related to this proposal, it is clear from the data presented in Chapter II that CGH and MLST analyses can be used to inform selection of candidate probiotics. However, *L. salivarius* strains of human origin cannot be universally expected to exert probiotic effects in humans and each novel probiotic requires pragmatic experimental validation of desired phenotypic traits as well as genomic characterisation of the strain.

Although the use of microarrays has an on-going role to play in comparative genomics**,** as illustrated in Chapter II, the genomic diversity of bacterial species is increasingly being examined using whole genome sequencing, as recently demonstrated by the comparative genomic analysis of 100 *L. rhamnosus* strains using SOLiD sequencing (Life Technologies) (Douillard *et al.*, 2013). While assembled consensus sequences of the test strains are mapped to a reference strain (Douillard *et al.*, 2013), this approach may overcome some of the limitations of traditional microarray-based CGH (outlined above and in Chapter II). For example, the datasets of the test isolates can be mined for genes not present in the reference isolate. Additionally, should divergent phenotypes be identified in test strains, their complete genome sequence datasets are available for further analysis.

Chapter III describes the assembly and analysis of the genome sequence of the porcine isolate *L. salivarius* JCM1046, which has the most structurally complex multipartite genome identified in *L. salivarius* to date. Extensive experimental validation of the sequence assembly and plasmid topology was required to resolve the genome architecture of *L. salivarius* JCM1046. The generation of a draft or incomplete genome sequence can have certain advantages in terms of the speed and costs involved in generating data; however the production of a draft genome sequence would not have allowed resolution of the genome structure of the strain described herein. This serves to emphasise the importance of adapting the sequencing and assembly strategy to the strain being analysed and the question being posed. When considering the evolutionary impact that extrachromosomal elements can have on their host cells (Chapter I) the need to fully characterise these replicons is evident.

The chromosome of strain JCM1046 harbours an 11.4 kb sequence region which is thought to represent a niche specific adaptation, containing a protein with a

mucin-binding MucBP domain, several transposases, hypothetical proteins, and a choloylglycine hydrolase (BSH2). Although present in the porcine strains JCM1046 and cp400, this region is absent from other sequenced genomes of *L. salivarius* and may represent an animal-specific niche adaptation. BSH2 is thought to confer JCM1046 with an increased ability to withstand the stresses associated with GIT survival (Fang *et al.*, 2009) and has been identified in two additional animal isolates to date, LMG14476 and cp400. BSH2 confers JCM1046 with an ability to resist much higher concentrations of the major human conjugated bile acids when compared to strains that harbour BSH1 alone (Fang *et al.*, 2009). Additionally, BSH2 has recently been shown to reduce weight gain and serum LDL cholesterol and liver triglycerides in mice fed normal or high-fat diets (Joyce *et al.*, 2014). Any potential industrial applications of strain JCM1046 in the control of host obesity and hypercholesterolemia would have to be assessed in light of the potential hazards associated with the carriage of Tn*6224*.

The co-existence of the four plasmids within the cell is thought to be enabled by the maintenance systems outlined in Chapter III. These include several TA systems, similar to the TA-related genes which have recently been annotated in the genomes of other strains of *L. salivarius* (Cho *et al.*, 2011; Fang *et al.*, 2008; Kergourlay *et al.*, 2012). It is likely that these maintenance systems play at role in the stability and maintenance of the numerous extrachromosomal elements found in *L. salivarius*.

The largest (of nine) regions of sequence diversity identified between pMP1046A and the sequences of the other published *repA*-type megaplasmids encodes proteins which are predicted to work synergistically with chromosomally encoded metabolic pathways, thus expanding its metabolic capabilities and enhancing this strain's ability to thrive in the porcine GIT. The number and variety of co-resident replicons contained within a *L. salivarius* strain can be difficult to predict by sequencing alone, as evidenced by the genome project described in Chapter III. The assignation of specific genes content and their implied functionalities to an "unfinished" plasmid sequence should be accepted with caution until experimentally validated.

Two novel plasmids pMP1046B and pCTN1046 were identified by *de novo* scaffold assembly and subsequently experimentally determined to have a circular topology (Chapter III). These plasmids were not described in previous studies of the

plasmid content of strain JCM1046 (Fang *et al.*, 2008; Li *et al.*, 2007). pMP1046B will require further functional characterisation to determine whether or not it has an impact on the ecological properties of JCM1046 as it was found to share neither sequence homology nor gene synteny with the *L. salivarius* plasmids sequenced to date. Although much of its gene function remains cryptic, megaplasmid pMP1046B was determined to be a candidate chromid as it harbours both plasmid-associated replication genes and paralogues of chromosomally encoded housekeeping and information-processing related genes. This is the first such instance of a chromid being described in a *L. salivarius* strain and increases our understanding of the spectrum of replicons present in this species. Definitions relating to extrachromosomal elements present in microbial genomes are somewhat fluid within the literature and will continue to be revised as more of these elements are sequenced and their encoded functions are determined.

As described in Chapter III, PFGE analysis predicted the size of pLMP1046 to be approximately 130 kb, but sequencing of the replicon revealed a plasmid that was 102 kb in size. The problems faced when sequencing the telomeres of linear elements are well recognised (Wagenknecht *et al.*, 2010) and this size discrepancy together with a lack of identifiable terminal inverted repeats (TIR) (Chapter III) is thought to be due to an omission of the presumptive repeat sequences in the terminal regions of pLMP1046. This further emphasises the need to consider the nature and characteristics of the replicon to be examined when choosing a sequencing strategy. Despite identifying a candidate replication origin in pLMP1046 in Chapter II, further analysis and/or sequencing of the terminal regions of pLMP1046 will be required to fully elucidate the mechanism involved in the replication of *L. salivarius* linear plasmids. The sequence of pLMP1046 will provide a genetic reference for studying the features involved in the replication of linear DNA plasmids in *Lactobacillus* sp.

Plasmid pCTN1046 harbours a single copy of a Tn*916*-like integrative conjugative transposon (Tn*6224*), the first described in a sequenced *L. salivarius* genome. Tn*6224* was predicted to be functionally intact and includes the tetracycline resistance gene *tetM*. Dissemination of antibiotic resistance genes among species which have food or probiotic-associations is highly undesirable (Panel, 2005). pCTN1046 shares 64.6% nt identity with the 30.4 kb plasmid pLS51C harboured by the probiotic avian isolate SMDX51 (Kergourlay *et al.*, 2012). However in contrast to Tn*6224*, the integrated conjugative element that is resident in pSL51C appears to

be a remnant of a conjugative element. It will be interesting to see how prevalent Tn*6224*-like elements are within the *L. salivarius* species, as more genome sequences become available. The identification of conjugative transposons carrying antibiotic resistance genes in the genomes of two animal isolates of *L. salivarius* may have implications for future strain selection for probiotic applications. The prevalence of Tn*6224* within the wider *L. salivarius* population should be examined in further detail together with investigation of the intra- and inter-species transfer rates of Tn*6224*.

Given the expanding use of *L. salivarius* as a probiotic agent, it is important to understand the breadth of genotypic and phenotypic heterogeneity of this species. The study of intra-species diversity in *L. salivarius* is an on-going topic of interest within Prof. O'Toole's lab and the wider School of Microbiology in UCC. The group of strains studied in detail herein have been examined with regard to their antimicrobial activities (Ryan *et al.*, 2008), adhesion capabilities (Collins *et al.*, 2012), bacteriocin phenotypes (Li *et al.*, 2007; O'Shea *et al.*, 2011), bile salt hydrolase activities (Fang *et al.*, 2009), phylogenetic structure (Li *et al.*, 2006), carbohydrate fermentation (Li *et al.*, 2006) and plasmid profiles (Fang *et al.*, 2008; Li *et al.*, 2007). The work presented in Chapters II and III have benefited from the technical know-how provided by a number of these studies. In turn, the data presented in this thesis necessitated the development and advancement of a range of methods and protocols tailored to *L. salivarius*. These included the protocol design for the production and analysis of CGH data, development of a medium which promotes production and enables quantification of EPS in *L. salivarius*, in addition to a protocol which has been developed to quantify biofilm formation in *L. salivarius* strains. These methods are now publically available for the further study of *L. salivarius* (Raftis *et al.*, 2011) and have subsequently been employed in a study of the bacteriocin loci of bacteriocin-producing *L. salivarius* isolates (O'Shea *et al.*, 2011; O' Shea *et al.*, 2012). The work presented in this thesis provides insight into the gene content, phenotypic characteristics, genome structure and evolution of this important species.

4.2 References

Badel, S., Bernardi, T. & Michaud, P. (2011). New perspectives for lactobacilli exopolysaccharides. *Biotech Adv* 29, 54-66.

Berger, B., Pridmore, R., Barretto, C., Delmas-Julien, F., Schreiber, K., Arigoni, F. & Brüssow, H. (2007). Similarity and differences in the *Lactobacillus acidophilus* group identified by polyphasic analysis and comparative genomics. *J Bacteriol* 189, 1311-1321.

Berger, U., Lutz, P., Sievers, C. & Trefz, F. K. (1976). Endocarditis lenta durch *Lactobacillus salivarius* subsp. *salicinicus*. *Dtsch med Wochenschr* 101, 1349-1350.

Bron, P. A., Tomita, S., Mercenier, A. & Kleerebezem, M. (2013). Cell surface-associated compounds of probiotic lactobacilli sustain the strain-specificity dogma. *Curr Opin Microbiol* 16, 262-269.

Cai, H., Thompson, R., Budinich, M. F., Broadbent, J. R. & Steele, J. L. (2009). Genome sequence and comparative genome analysis of *Lactobacillus casei*: insights into their niche-associated evolution. *Genome Biol Evol* 2009, 239-257.

Cho, Y.-J., Choi, J. K., Kim, J.-H., Lim, Y.-S., Ham, J.-S., Kang, D.-K., Chun, J., Paik, H.-D. & Kim, G.-B. (2011). Genome sequence of *Lactobacillus salivarius* GJ-24, a probiotic strain isolated from healthy adult intestine. *J Bacteriol* 193, 5021-5022.

Collins, J., van Pijkeren, J.-P., Svensson, L. & other authors (2012). Fibrinogen-binding and platelet-aggregation activities of a *Lactobacillus salivarius* septicaemia isolate are mediated by a novel fibrinogen-binding protein. *Mol Microbiol* 85, 862-877.

Douillard, F. P., Ribbera, A., Kant, R. & other authors (2013). Comparative genomic and functional analysis of 100 *Lactobacillus rhamnosus* strains and their comparison with strain GG. *PLoS Genet* 9, e1003683.

Fang, F., Flynn, S., Li, Y., Claesson, M. J., van Pijkeren, J.-P., Collins, J. K., van Sinderen, D. & O'Toole, P. W. (2008). Characterization of endogenous plasmids from *Lactobacillus salivarius* UCC118. *Appl Environ Microbiol* 74, 3216-3228.

Fang, F., Li, Y., Bumann, M., Raftis, E. J., Casey, P. G., Cooney, J. C., Walsh, M. A. & O'Toole, P. W. (2009). Allelic variation of bile salt hydrolase genes in *Lactobacillus salivarius* does not determine bile resistance levels. *J Bacteriol* 191, 5743-5757.

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

Fuller, R. & Brooker, B. E. (1974). Lactobacilli which attach to the crop epithelium of the fowl. *Am J Clin Nutr* 27, 1305-1312.

Godoy-Gallardo, M., Rodríguez-Hernández, A. G., Delgado, L. M., Manero, J. M., Javier Gil, F. & Rodríguez, D. (2014). Silver deposition on titanium surface by electrochemical anodizing process reduces bacterial adhesion of *Streptococcus sanguinis* and *Lactobacillus salivarius*. *Clin Oral Implants Res*, 1-10.

Husni, R. N., Gordon, S. M., Washington, J. A. & Longworth, D. L. (1997). *Lactobacillus* bacteremia and endocarditis: review of 45 cases. *Clin Infect Dis* 25, 1048-1055.

Jolly, L., Vincent, S., Duboc, P. & Neeser, J.-R. (2002). Exploiting exopolysaccharides from lactic acid bacteria. *Antonie Van Leeuwenhoek* 82, 367-374.

Jones, S. E. & Versalovic, J. (2009). Probiotic *Lactobacillus reuteri* biofilms produce antimicrobial and anti-inflammatory factors. *BMC Microbiol* 9.

Joyce, S. A., MacSharry, J., Casey, P. G., Kinsella, M., Murphy, E. F., Shanahan, F., Hill, C. & Gahan, C. G. M. (2014). Regulation of host weight gain and lipid metabolism by bacterial bile acid modification in the gut. *Proc Nat Acad Sci USA* 111, 7421-7426.

Kergourlay, G., Messaoudi, S., Dousset, X. & Prévost, H. (2012). Genome sequence of *Lactobacillus salivarius* SMXD51, a potential probiotic strain isolated from chicken cecum, showing anti-*Campylobacter* activity. *J Bacteriol* 194, 3008-3009.

Land, M. H., Rouster-Stevens, K., Woods, C. R., Cannon, M. L., Cnota, J. & Shetty, A. K. (2005). *Lactobacillus* sepsis associated with probiotic therapy. *Pediatrics* 115, 178-181.

Lebeer, S., Verhoeven, T. L. A., Perea Velez, M., Vanderleyden, J. & De Keersmaecker, S. C. J. (2007). Impact of environmental and genetic factors on biofilm formation by the probiotic strain *Lactobacillus rhamnosus* GG. *Appl Environ Microbiol* 73, 6768-6775.

Lebeer, S., Vanderleyden, J. & De Keersmaecker, S. C. J. (2008). Genes and molecules of lactobacilli supporting probiotic action. *Microbiol Mol Biol Rev* 72, 728-764.

Lebeer, S., Claes, I. J. J., Verhoeven, T. L. A., Vanderleyden, J. & De Keersmaecker, S. C. J. (2011). Exopolysaccharides of *Lactobacillus rhamnosus* GG form a protective shield against innate immune factors in the intestine. *Microbial Biotech* 4, 368-374.

Lebeer, S., Claes, I., Tytgat, H. L. P. & other authors (2012). Functional analysis of *Lactobacillus rhamnosus* GG pili in relation to adhesion and

immunomodulatory interactions with intestinal epithelial cells. *Appl Environ Microbiol* 78, 185-193.

Li, Y., Raftis, E., Canchaya, C., Fitzgerald, G. F., van Sinderen, D. & O'Toole, P. W. (2006). Polyphasic analysis indicates that *Lactobacillus salivarius* subsp. *salivarius* and *Lactobacillus salivarius* subsp. *salicinius* do not merit separate subspecies status. *Int J Syst Evol Microbiol*, 2397-2403.

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.

Macedo, M. G., Lacroix, C., Gardner, N. J. & Champagne, C. P. (2002). Effect of medium supplementation on exopolysaccharide production by *Lactobacillus rhamnosus* RW-9595M in whey permeate. *Int Dairy J* 12, 419-426.

Mackay, A. D., Taylor, M. B., Kibbler, C. C. & Hamilton-Miller, J. M. T. (1999). *Lactobacillus* endocarditis caused by a probiotic organism. *Clinical Microbiology and Infection* 5, 290-292.

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

Messaoudi, S., Manai, M., Kergourlay, G., Prévost, H., Connil, N., Chobert, J. M. & Dousset, X. (2013). *Lactobacillus salivarius*: bacteriocin and probiotic activity. *Food Microbiol* 36, 296-304.

Miyoshi, Y., Okada, S., Uchimura, T. & Satoh, E. (2006). A mucus adhesion promoting protein, MapA, mediates the adhesion of *Lactobacillus reuteri* to Caco-2 human intestinal epithelial cells. *Biosci Biotech Biochem* 70, 1622-1628.

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.

Neville, B. A. & O'Toole, P. W. (2010). Probiotic properties of *Lactobacillus salivarius* and closely related *Lactobacillus* species. *Future Microbiol* 5, 759-774.

O'Shea, E. F., O'Connor, P. M., Raftis, E. J., O'Toole, P. W., Stanton, C., Cotter, P. D., Ross, R. P. & Hill, C. (2011). Production of multiple bacteriocins from a single locus by gastrointestinal strains of *Lactobacillus salivarius*. *J Bacteriol*.

O' Shea, E. F., O' Connor, P. M., Raftis, E. J., O' Toole, P. W., Stanton, C., Cotter, P. D., Ross, R. P. & Hill, C. (2012). Subspecies diversity in bacteriocin production by intestinal *Lactobacillus salivarius* strains. *Gut Microbes* 3, 468-473.

Panel, F. (2005). Opinion of the Scientific Panel on additives and products or substances used in animal feed (FEEDAP) on the updating of the criteria used

in the assessment of bacteria for resistance to antibiotics of human or veterinary importance *EFSA Journal*.

Parisotto, T. M., Steiner-Oliveira, C., Duque, C., Peres, R. C. R., Rodrigues, L. K. A. & Nobre-dos-Santos, M. (2010). Relationship among microbiological composition and presence of dental plaque, sugar exposure, social factors and different stages of early childhood caries. *Arch Oral Biol* 55, 365-373.

Pham, L. C., van Spanning, R. J. M., Röling, W. F. M., Prosperi, A. C., Terefework, Z., ten Cate, J. M., Crielaard, W. & Zaura, E. (2009). Effects of probiotic *Lactobacillus salivarius* W24 on the compositional stability of oral microbial communities. *Arch Oral Biol* 54, 132-137.

Piwat, S., Teanpaisan, R., Thitasomakul, S., Thearmontree, A. & Dahlén, G. (2010). *Lactobacillus* species and genotypes associated with dental caries in Thai preschool children. *Mol Oral Microbiol* 25, 157-164.

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.

Roos, S. & Jonsson, H. (2002). A high-molecular-mass cell-surface protein from *Lactobacillus reuteri* 1063 adheres to mucus components. *Microbiology* 148, 433 - 442.

Ryan, K. A., Daly, P., Li, Y., Hooton, C. & O'Toole, P. W. (2008). Strain-specific inhibition of *Helicobacter pylori* by *Lactobacillus salivarius* and other lactobacilli. *J Antimicrob Chemoth* 61, 831-834.

Salminen, M., Rautelin, H., Tynkkynen, S., Poussa, T., Saxelin, M., Valtonen, V. & Järvinen, A. (2004a). *Lactobacillus* bacteremia, clinical significance, and patient outcome, with special focus on probiotic *L. rhamnosus* GG. *Clinical infectious diseases:an official publication of the Infectious Diseases Society of America* 38, 62-69.

Salminen, S., von Wright, A. & Ouwehand, A. (2004b). Lactic Acid Bacteria: microbiological and functional aspects (3rd ed.) Revised and expanded edition.

Servin, A. L. & Coconnier, M.-H. (2003). Adhesion of probiotic strains to the intestinal mucosa and interaction with pathogens. *Best Pract Res Clin Gastroenterol* 17, 741-754.

Servin, A. L. (2004). Antagonistic activities of lactobacilli and bifidobacteria against microbial pathogens. *FEMS Microbiol Rev* 28, 405-440.

Siezen, R. J., Bayjanov, J. R., Felis, G. E., van der Sijde, M. R., Starrenburg, M., Molenaar, D., Wels, M., van Hijum, S. A. F. T. & van Hylckama Vlieg, J. E. T. (2011). Genome-scale diversity and niche adaptation analysis of *Lactococcus lactis* by comparative genome hybridization using multi-strain arrays. *Microbial Biotech* 4, 383-402.

Sims, I. M., Frese, S. A., Walter, J. & other authors (2011). Structure and functions of exopolysaccharide produced by gut commensal *Lactobacillus reuteri* 100-23. *ISME J* 5, 1115-1124.

Su, M. S.-W. & Gänzle, M. G. (2014). Novel two-component regulatory systems play a role in biofilm formation of *Lactobacillus reuteri* rodent isolate 100-23. *Microbiology* 160, 795-806.

Švec, P., Sedláček, I., Žáčková, L., Nováková, D. & Kukletová, M. (2009). *Lactobacillus* spp. associated with early childhood caries. *Folia Microbiol* 54, 53-58.

Tannock, G., Ghazally, S., Walter, J. & other authors (2005). Ecological behavior of *Lactobacillus reuteri* 100-23 is affected by mutation of the *luxS* gene. *Appl Environ Microbiol* 71, 8419-8425.

van den Berg, D., Robijn, G. W., Janssen, A. C., Giuseppin, M., Vreeker, R., Kamerling, J. P., Vliegenthart, J., Ledeboer, A. M. & Verrips, C. T. (1995). Production of a novel extracellular polysaccharide by *Lactobacillus sake* 0-1 and characterization of the polysaccharide. *Appl Environ Microbiol* 61, 2840-2844.

van Geel-Schutten, G. H., Flesch, F., ten Brink, B., Smith, M. R. & Dijkhuizen, L. (1998). Screening and characterization of *Lactobacillus* strains producing large amounts of exopolysaccharides. *Appl Microbiol Biotechnol* 50, 697-703.

Vélez, M. P., De Keersmaecker, S. C. J. & Vanderleyden, J. (2007). Adherence factors of *Lactobacillus* in the human gastrointestinal tract. *FEMS Microbiology Letters* 276, 140-148.

Vélez, M. P., Petrova, M. I., Lebeer, S., Verhoeven, T. L. A., Claes, I., Lambrichts, I., Tynkkynen, S., Vanderleyden, J. & De Keersmaecker, S. C. J. (2010). Characterization of MabA, a modulator of *Lactobacillus rhamnosus* GG adhesion and biofilm formation. *FEMS Immunol Med Microbiol* 59, 386-398.

Wagenknecht, M., Dib, J., Thürmer, A., Daniel, R., Farías, M. & Meinhardt, F. (2010). Structural peculiarities of linear megaplasmid, pLMA1, from *Micrococcus luteus*; interfere with pyrosequencing reads assembly. *Biotechnol Lett* 32, 1853-1862.

Walter, J., Chagnaud, P., Tannock, G. W., Loach, D. M., Dal Bello, F., Jenkinson, H. F., Hammes, W. P. & Hertel, C. (2005). A high-molecular-mass surface protein (Lsp) and methionine sulfoxide reductase B (MsrB) contribute to the ecological performance of *Lactobacillus reuteri* in the murine gut. *Appl Environ Microbiol* 71, 979-986.

Walter, J., Schwab, C., Loach, D. M., Ganzle, M. G. & Tannock, G. W. (2008). Glucosyltransferase A (*GtfA*) and inulosucrase (*Inu*) of *Lactobacillus reuteri* TMW1.106 contribute to cell aggregation, in vitro biofilm formation, and colonization of the mouse gastrointestinal tract. *Microbiology* 154, 72-80.

Woo, P. C. Y., Fung, A. M. Y., Lau, S. K. P. & Yuen, K.-Y. (2002). Identification by 16S rRNA gene sequencing of *Lactobacillus salivarius* bacteremic cholecystitis. *J Clin Microbiol* 40, 265-267.

Yuki, N., Shimazaki, T., Kushiro, A., Watanabe, K., Uchida, K., Yuyama, T. & Morotomi, M. (2000). Colonization of the stratified squamous epithelium of the nonsecreting area of horse stomach by lactobacilli. *Appl Environ Microbiol* 66, 5030-5034.

Acknowledgements

I would like to thank the many people who directly and indirectly contributed to my PhD research experience and made it possible to complete this thesis.

Sincere and heartfelt thanks to Prof. Paul O'Toole. Thank you for your patience, encouragement and wisdom. Thanks for putting up with me over the years. I really loved my time in lab339 and learned so much during my time in your lab.

To Prof. Douwe van van Sinderen, thank you for your advice, suggestions and feedbackon my work.

To all of the past and present members of FSB Lab 339 and BSI office 4.11, thank you so much for all of you encouragement and expertise over the years. In particular, I need to give a very big thank you to Delphine and Anne for all the Friday night dance parties in lab339 over the years. You put up with a lot of Dolly Parton and whinging about PFGEs/genome assemblies/problems of the day, I'm grateful that you were there for all of it!

I'd like to give a massive thank you to Marcus, Aldert and Brian for all of your help in the bioinformatics department. Marcus, you're a miracle worker/scripting ninja, you really did make me love Linux!

To the international collaborators and the individuals who contributed to the research included in this thesis, in particular a massive thank you to Elisa, you were a star!

A big thank you to the members of the UCC Microbiology department (past and present), in particular Margaret, Liam and Paddy. You kept everything running smoothly on a day-to-day basis, and were so generous with your time and knowledge.

Thank my family for their unfaltering belief in my ability to make it this far, and for your encouragement throughout my years in school and in UCC.

The final and biggest thanks are for my husband Seán (aka Excel Wizard Extraordinaire). You've been my biggest supporter and best friend and I couldn't have done this without you. You put up with late dinners, weekends in the lab and many many presentation-related freakouts. You always told me I was smart enough to do this, when I didn't believe it. I would never have happened without you. x