Winter 11-27-2012

The Evolution of Host Specificity in the Vertebrate Gut Symbiont *Lactobacillus reuteri*

Steven Frese
University of Nebraska-Lincoln, steve.frese@gmail.com

THE EVOLUTION OF HOST SPECIFICITY IN THE VERTEBRATE GUT SYMBIONT *LACTOBACILLUS REUTERI*

By

Steven A. Frese

A DISSERTATION

Presented to the Faculty of
The Graduate College at the University of Nebraska
in Partial Fulfillment of Requirements
for the Degree of Doctor of Philosophy

Major: Food Science and Technology

Under the Supervision of Professor Jens Walter

Lincoln, Nebraska

December, 2012

THE EVOLUTION OF HOST SPECIFICITY IN THE VERTEBRATE GUT SYMBIONT *LACTOBACILLUS REUTERI*

Steven A. Frese, Ph.D

University of Nebraska, 2012

Advisor: Dr. Jens Walter

The vertebrate gut is home to one of the densest populations of life on Earth. This microbial community has a profound effect on host health, nutrition, development, behavior, and evolution. However, very little is known about how these microbes have evolved with their vertebrate hosts, how and whether they select hosts or how they remain associated with their hosts. Recent work identified *Lactobacillus reuteri* as an organism that is composed of host-specific subpopulations, each population associated with a different host animal.

Representatives from each host-associated population were tested for their ability to colonize gnotobiotic mice, which only rodent strains could accomplish. To determine the genetic source of these differences, genome sequences from representative strains (from humans, pigs, chickens, and rodents) and genomic microarrays were used to identify host-specific genes tied to ecological success *in vivo*. Many of the rodent-specific genes were essential to the ecological success of this strain. Overall, the types of genes which were essential suggested that they were related to biofilm formation. Using a novel method of biofilm quantification and a germ-free mouse model, the ability of *L. reuteri* strains to form biofilms on the forestomach of germ-free mice was examined. *In vivo* biofilm formation was exclusive to the rodent associated strains. Using genome comparisons, *in vitro* biofilm formation, and related models of biofilm

formation, genes essential to biofilm formation were identified. Genes responsible for adherence, aggregation, and gene regulation were all critical to biofilm formation.

To determine whether the observed host specificity of rodent associated *L. reuteri* in mice extended to other host-associated lineages, a crossover human experimental trial was conducted. Three species of *Lactobacillus* were tested for their ability to persist in the human gastrointestinal tract. A human-associated strain of *L. reuteri*, as well as a related species isolated from humans reached higher populations than a non-associated strain, suggesting that host-specificity of *L. reuteri* extends to other lineages as well.

Together, these experiments demonstrated the host-specificity of the species and identified the molecular mechanisms by which rodent-associated *L. reuteri* colonize their host.

Contents

List of Figures

List of Tables

Acknowledgments

First and foremost, I would like to thank my family, who have supported me through these years without hesitation. Their support and love has been beyond value.

I would also like to thank my 'Nebraska Family,' who have been excellent friends throughout my time here in Lincoln. I am so very thankful to the kindness and support you have all given me and the encouragement to continue. Whether it was ballads about life in Lincoln, Nebraska football, rugby games, concerts, or just sitting down to dinner, you all gave me the balance and friendship in my life that helped make all the PCR reactions, growth curves, mouse experiments, microscopy, and writing possible.

I would like to thank my colleagues in the Walter Lab, especially Phaik Lyn Oh, Kenzi Clark, and Ines Martinez. Your help, advice, commiseration, and support have always be greatly appreciated.

I would especially like to thank my advisor, Dr. Jens Walter for his patience, guidance, advice, and direction throughout this project. He has been invaluable in his contribution to my scientific, personal, and professional development. I would also like to thank the rest of my my advisory committee, Dr. Andrew Benson, Dr. Paul Blum, Dr. Robert Hutkins, and Dr. Daniel Peterson. You have been excellent teachers and advisors and have given me indispensible advice.

Preface

This dissertation is composed of five chapters. In **Chapter 1**, a discussion of the current understanding of the the vertebrate gut microbiota, rules governing its assembly, the evolution of symbionts, models which explain their evolution, case studies of the evolution of animal-microbe symbioses and the case of *Lactobacillus reuteri* are discussed.

Chapter 2 describes the evidence which suggests that *Lactobacillus reuteri* is a host-adapted symbiont, and the experiments which confirmed this observation. This work was published in 2011 (Frese et al., 2011) and concluded with a hypothesis that the host-distinct phenotype observed in this organism (**Chapters 2**) is a result of different ecological strategies adopted during adaptation to different hosts.

To identify the mechanisms for host-specific colonization, adherence and biofilm formation on the rodent squamous forestomach epithelium as a biofilm, was investigated (**Chapter 3**). Biofilm growth mirrored the host restriction observed in **Chapter 2**. The molecular mechanisms which contributed to biofilm formation were also identified. These genes, involved in aggregation, adhesion, and gene regulation, included genes which were rodent specific and genes which were common to all *L. reuteri*. This suggested that biofilm growth is a common feature to the species, but the different host-associated subpopulations demonstrate host-specific adherence.

Another *L. reuteri*-host interaction was also studied. Using human volunteers, the ability of a human-lineage strain of *L. reuteri* to persist in the human gastrointestinal tract was compared to isolates from two other species with differing phylogenetic or ecological histories. One organism autochthonous to the human gut and another which was not were tested against a human-associated *L. reuteri* strain (ATCC PTA 6475). The results suggested that the human *L. reuteri* strains are indeed autochthonous to

the human gut, in the same way the rodent strains are likewise adapted to their host environment (**Chapter 4**).

Finally, a concluding chapter (**Chapter 5**) summarizes the findings, poses open questions, and suggests future studies which can elaborate on these results.

Chapter 1

Vertebrate Hosts and their Gut Bacteria:

A Review

1.1 Introduction

The gastrointestinal (GI) tract of humans and animals is home to one of the densest microbial communities found on Earth. The population density exceeds 10^{11} cells/gram (Gill et al., 2006) representing a collective genome two orders of magnitude larger than that of the host (Gill et al., 2006). The human microbial community is so important to the host that Round (2009) described humans simply as a ‘‘scaffold on which diverse microbial ecosystems are established.” The composition of this community also has profound functional and physiological consequences for the host, as it provides a critical source of essential nutrients (Kulwich et al., 1953, Sumi et al., 1977, Albert et al., 1980), shapes physical (Kikuchi et al., 2009, Tannock et al., 1988) and immunological development (Mazmanian et al., 2005), and is critical to disease resistance (Croswell et al., 2009, Stecher and Hardt, 2008, van der Waaij et al., 1971). Yet, vertebrate young must acquire this microbial consortia at each generation as they are virtually sterile at birth (Palmer et al., 2007). Although the benefits that this community confers to the host are well established, the ecological and evolutionary processes governing it are less clear (Herre et al., 1999).

The acquisition of the microbial community, and the ecology of specific constituents has been an active field of research even before Elie Metchnikoff first suggested that the bacteria found in the colon were responsible for aging (Metchnikoff, 1907). Early interest focused on how specific organisms could influence health and nutrition (Kulwich et al., 1953, Shirota et al., 1966), and later the advent of gnotobiotic techniques allowed for the isolation of the host vertebrate from the microbiota (Reyniers, 1959). These models helped scientists understand the role of the microbiota and its impact on the host. In recent years modern molecular and microscopic techniques have been used to understand how the microbes found within the GI tract interact

with their host. These developments, combined with new understandings of ecological and evolutionary theory, have helped the field begin to answer exciting questions regarding the development, acquisition, and ecology of the GI community.

This review will address the current understanding of the vertebrate GI microbiota as a community, its constituent parts, and as an ecological model. It will first describe the development of the microbiota from birth, when all niches are vacant, to the climax community in which all niches are occupied. Then, the ecological principles which govern it will be detailed. Next, the forces that shape the constituents of this community and how they are affected by host behaviors will be discussed. Model systems that offer well-characterized case studies of intimate animal-microbe relationships and how they can apply to the vertebrate gut microbiota will be described. In this context, *Lactobacillus reuteri* will be presented as a model for vertebrate gut symbionts. Finally, the rationale for the use of *L. reuteri* as a model organism to understand vertebrate host and gut symbiont relationships will be described and open questions will be posed.

1.2 The Assembly of the Gut Microbiota.

At birth, the gastrointestinal tract is sterile (Palmer et al., 2007) but it is rapidly colonized by microbes. Initially, the community composition differs greatly between individuals but as the community approaches a steady 'adult' state, it becomes more similar between individuals (Palmer et al., 2007, Trosvik et al., 2010). This is because of the ecological forces which shape these communities. Three main ecological theories are thought to govern the assembly of the microbiota. They are (i) the Niche-Related theory, (ii) the Neutral theory, and (iii) the Historical Processes theory (Cavender-Bares et al., 2009).

First, the niche-related theory, applied to the gut microbiota, suggests that host-defined physiological differences (eg. gut type, body temperature, or diet) create the major distinction between the host gut and the outside environment (Cavender-Bares et al., 2009, Ley et al., 2008a,b). For example, host body temperature in warm-blooded animals remains tightly regulated. This restricts microbial growth to organisms which thrive at a specific temperature. Second, the microbe must survive passage in an acidic stomach. In addition to a low stomach pH, these cells must survive digestive enzymes (eg. proteases), bile acids, and secretory immunoglobulins secreted into the GI tract (Ouwehand et al., 2002). In this way, the host's physiology selects for organisms which can survive this transit. According to the niche-related theory, these combined hurdles select for the specific bacteria which compose the final community.

Another governing theory of community assembly is the neutral theory (Cavender-Bares et al., 2009). This states that host animals gather their communities at random from the environment beginning at birth. According to this assumption, the resulting community would contain a cross-section of microbes common to the contemporary environment. Based on community sequencing studies, this does not appear to be the case. The microbial communities of vertebrates differs significantly from free-living communities (Ley et al., 2008b) and the neutral theory fails to predict the composition of fecal samples (Sloan et al., 2005). This theory also predicts that the resulting community would be open to newcomers, temporally unstable, and extremely individualized (Cavender-Bares et al., 2009). Even monozygotic twins differ in their GI communities (Turnbaugh et al., 2009), which appears to support the neutral theory. However, the adult microbiota is a stable, invader-resistant community (Stecher and Hardt, 2008, Costello et al., 2009), which contradicts the hypothesis of a neutral theory of assembly. Still, random events do have an impact on the community especially on the order of colonization (Dethlefsen et al., 2006), which may have important effects

through the historical theory of community assembly.

The historical theory of development posits that exposure to early colonizers shapes the developing community. This exposure is largely random, but early colonizers can shape the environmental conditions that are available to newcomers (Cavender-Bares et al., 2009). This theory, in practice, would allow host organisms to influence the probability of exposure to favored early colonizers, whether at birth or shortly thereafter. Early exposure could preferentially shape future niche availability, which impacts the ability of newcomers to colonize and ultimately shapes the final community (Cavender-Bares et al., 2009)

These three theories describe the patterns which govern the communities found in the vertebrate GI tract. They explain the stability of the ecosystem (Stecher and Hardt, 2008, Costello et al., 2009), its redundancy (Turnbaugh et al., 2009), the resistance to perturbation (Cavender-Bares et al., 2009, Tannock et al., 2012), and the benefit to hosts which influence early colonizers. When viewed through these theories, it is possible to understand how the host can influence the development of this community, and the benefits of doing so.

1.3 The Acquisition of Specific Microbial Partners

Hosts have a vested interest in the development of their microbial communities. The importance of the gut microbiota and its specific constituents, for host diet (Kulwich et al., 1953, Sumi et al., 1977, Albert et al., 1980), immune development (Mazmanian et al., 2005), and disease resistance (van der Waaij et al., 1971, Croswell et al., 2009) increases the need to acquire the correct symbionts at each generation. For vertebrates, these populations are acquired after birth and are a result of the exposure of naïve hosts to microbes from their environment, which includes the ecosystem,

parental contact, and social behaviors.

Exposure to microbes from the environment contributes the organisms which make up the microbiota. This stochastic exposure shapes the microbiota by routinely introducing potential colonizers which can fill vacant niches in the community. These microbes could originate from soil, water, or through the diet. While the climax community found in adult animals is very stable, the developing microbiota continuously encounters these new microbes from the environment. Surprisingly, despite the diversity of microbes in the environment, the dominant populations in the vertebrate gastrointestinal tract are dominated by only a few phyla (Ley et al., 2008b). This suggests that only a few of the microbes encountered from the environment actually become associated with the host. Host physiology is known to play a part in this (Ley et al., 2008a), and host behaviors are also likely to encourage colonization by some microbes and not others, ensuring transmission of key microbes to offspring.

For live-borne animals, the birth canal contains the first bacteria to which the young are exposed. The human vagina is colonized predominantly by lactobacilli, but other microbes are present as well (Ravel et al., 2011), and vaginal inoculation of infants results in early colonization. If infants are prevented from vaginal exposure during delivery, a 100-fold decrease in the microbial population is observed for first few days after birth (Tannock et al., 1990). Second, The inoculation of the infant with *Lactobacilli*, *Bifidobacterium*, and *Enterobacteriaceae* appears to result in the early colonization by *Bifidobacterium* and *Enterobacteriaceae* as a result of delivery (Tannock et al., 1990). These organisms do not persist in the long-term, but their early colonization could influence colonization by bacteria to which the infant is later exposed (Palmer et al., 2007).

After birth, parental contact can facilitate the transfer of microbes to offspring. Mother's milk contains lactobacilli (Martín et al., 2003, Sinkiewicz and Ljunggren,

2008) that can be transferred to offspring through nursing (Martín et al., 2003). In addition to the direct provision of these bacteria, human milk contains specific complex polysaccharides that encourage the growth of *Bifidobacterium* which selectively utilize these compounds (LoCascio et al., 2007).

Social contact can also transfer microbes between animals. Social contact is hypothesized to increase access to beneficial microbes (Lombardo, 2008). This could be especially helpful among herbivores, who depend on a functional rumen prior to weaning, and contact with other animals facilitates the acquisition of these microbes (Fonty et al., 1987). In birds, nestlings are rapidly colonzied as a result of fecal contamination of their nests (Mills et al., 1999) and colonization by microbes from parental saliva are critical for the survival young birds (Kyle and Kyle, 1993).

Coprophagy also provides access to communities of bacteria, allowing for transmission of gut microbes from parent or kin to offspring, and between individuals. Many animals are coprophagic and consume feces for nutritive benefit (Kulwich et al., 1953, Mameesh and Johnson, 1960), to absorb nutrients produced in the colon that would otherwise not be absorbed by the host or further digestion of plant material (Alexander, 1993). However, this behavior also provides an inoculum for transmission of microbes between individuals (Roberts, 2005). The transfer of a complete community from parent (or kin) to offspring transfers the beneficial relationships acquired by the parent (Lombardo, 2008).

1.3.1 Transmission Between Hosts

The transmission of gut microbes falls largely within two categories. These categories are transmission from parent to offspring, or vertical transmission, and transmission from the environment, known as horizontal transmission. While both mechanisms result in exposure to these microbes, they face different selective pressures

and produce different outcomes.

In horizontal transmission, microbes are passed across a contemporary host population. For the microbe, success by horizontal transmission relies on the ability to colonize new hosts. Highly transmissible organisms gain the selective advantage, rather than organisms which establish long-term associations with their host animal. Transmission by this method selects for organisms which spread, at the cost of host fitness, precisely because these long term relationships are not favored (Wickham et al., 2007). Instead of being subject to long-term reductions in host fitness, the colonizer needs only two features for success; a mechanism for transmission and a population of susceptible hosts. In experimental models where pathogens are experimentally transmitted from parent to offspring, rather than horizontally, virulence is reduced (Stewart et al., 2005). When the fitness of the pathogen is tied to survival and reproduction of the host, selection tempers virulence.

This tempering is the result of vertical transmission. When organisms align their fitness interests, behaviors which damage host fitness also damage the fitness of the microbe and are abandoned (Stewart et al., 2005). However, what may be damaging in the host may be critical for survival in a habitat outside the host animal (Woolhouse et al., 2001). The loss of this secondary habitat and transmission exclusively within a host population is thought to contribute to the intimacy of the host-microbe relationship and discourage negative behaviors (Stewart et al., 2005). In a symbiont population which can only be transferred from parent to offspring, phylogenetic signals begin to reflect host genetic diversity as a result of co-diversification, if the transmission does not allow for outside diversification (Baumann et al., 1995).

In microbes who have lost access to a secondary host habitat, genes responsible for colonizing that habitat are rapidly lost (Moran et al., 2001). While a single gene can change habitat specificity (Mandel et al., 2009), the decay of whole clusters of

relevant genes precludes a return to these habitats. Given a stable relationship with a host and a loss or abandonment of a secondary niche, microbes can become associated with a single host species, a result known as host specificity. The processes which drive this require an extremely stable relationship, so mechanisms of transfer between generations, and the incentives to maintain a close relationship must develop.

1.4 Partner Pressure can Strain or Maintain Host-microbe Relationships

In vertebrates, the gut microbiota and its constituent populations provide a net benefit to the host, and broad community-level associations between host and microbiota are observed (Ley et al., 2008a). General community effects such as colonization resistance do not require a specific interaction or association between host and microbe. This effect could be considered to be a by-product of the presence of a stable community (Stecher and Hardt, 2008). Yet, specific benefits are provided by constituent species as well. Behaviors by the host even actively work to benefit from these relationships (Kulwich et al., 1953, Mameesh and Johnson, 1960). These cooperative relationships are common throughout nature. Their origins and stability, however, present an apparent problem for evolutionary theorists. Given the potential cost of cooperation, how do these species initiate such cooperation? How do these partners cooperate given the incentives to 'cheat' in the relationship?

Initially, cooperation may be encouraged by a low cost-to-benefit ratio. The lowest possible costs exist in byproduct mutualisms. In these cases, a zero-cost benefit from one partner aids the other (Sachs et al., 2004). In the case of the gut microbiota, the metabolic products of gut microbes can provide a benefit to the host. For example, short-chain fatty acid production by gut microbes may be an example of byproduct

mutualism (Walter et al., 2011).

In cases where costs are involved, several models have been proposed to explain how cooperation can exist and continue. Well-known models of human interaction, such as the Prisoner's dilemma (PD), have been applied to the problem (Trivers, 1971). In the PD, cooperation maximizes the total payoff. Each partner can choose to ‘‘cheat" and maximize their own payoff at the cost of their partner. The cheater leaves the "sucker" with no benefit. A microbe which normally benefits its host, but loses a beneficial gene could be an example of such a cheater. If the host expels the cheater, there is no gain to either partner. When both host and microbe cooperate however, both parties benefit.

However, the application of the PD to natural systems is limited because it is a single instance, not the repeated interactions between populations which are more likely in natural settings. Taking the PD and accounting for repeated interactions, Alexander and Hamilton (1981) demonstrated that multiple repetitions of the PD, the iterated PD (IPD), allows for feedback between partners. The IPD allows partners to reward cooperation, punish cheating, and forgive a partner when they begin to cooperate again.

Two important points are present in Alexander and Hamilton's PD (1981) and its extension into the IPD (Doebeli and Knowlton, 1998, Foster and Wenseleers, 2006). The first is the notion of repeated interactions. Partners who cooperate repeatedly and build a relationship that aligns their fitness interests have a more stable relationship than one which is facultative among unrelated hosts (Trivers, 1971, Axelrod and Hamilton, 1981). The former resembles the vertical transmission of partners, from parent (or kin, pseudo-vertical transmission, Wilkinson and Sherratt 2001) to offspring. The latter resembles horizontal transmission across populations. While horizontal transmission of gut symbionts within populations is possible, it is thought to be less

likely to result in long-term associations. This is because there is less incentive for the two partners to align their shared interests (Trivers, 1971, Axelrod and Hamilton, 1981). In contrast, vertical (or, if by kin, 'pseudo-vertical') transmission is modeled to promote these relationships through feedback, and resembles the second important feature of the IPD (Foster and Wenseleers, 2006).

The second important feature of the IPD is the "tit-for-tat" which rewards cooperation and punishes defection (Foster and Wenseleers, 2006). Partners with these relationships preferentially interact with more helpful partners and punish the less helpful partner. These concepts are known respectively as "partner choice" and "partner sanction" (Noë and Hammerstein, 1995). Hosts which can punish non-cooperative partners (partner sanction) and encourage cooperative partners (partner choice) provide additional incentive for partners to maintain their relationship and discourage cheating.

Combining these themes, Foster and Wenseleers (2006) present a comprehensive model of the formation of stable beneficial relationships which applies broadly, but also to host-gut microbe interactions. Initially, they develop predicted to evolve by this model when cost-to-benefit ratios are low and it will continue when feedback encourages the maintenance of the relationship (Foster and Wenseleers, 2006). Vertical transmission, or pseudo-vertical transmission of microbes across generations provides for a reliable means by which selection can act on partnerships, and encourage partner choice and sanction (Foster and Wenseleers, 2006). Selection on relationships between vertebrates and gut microbe populations, as with other traits, results in genetic markers for associations between partners. These markers have been recently identified in rodent populations (Benson et al., 2010, Buhnik-Rosenblau et al., 2011), but future work will be needed to define the mechanisms which reflect host control of these populations.

1.5 Case Studies in Animal-Microbe Symbioses

Several well-studied examples of host-microbe relationships can be found among invertebrates. Though different, they offer insight into how vertebrate-symbiont interactions might behave. They may also offer the models and methods by which to study vertebrate-gut microbe symbioses. Presented here are three well characterized models of animal-microbe symbioses, each representing a different type of partnership. First, *Buchnera*, an ancient obligate endosymbiont transmitted from parent to offspring (vertical transmission) is discussed. This example demonstrates the consequences of tightly controlled vertical transmission. Second, is an ant which maintains a vertically transmitted fungal symbiont through partner choice. Third, a squid which acquires a specific constituent from its environment and uses partner choice and sanction to acquire its microbial symbiont.

1.5.1 *Buchnera* and aphids.

More than 200 million years ago, an aphid ancestor was infected by an ancestor of the γ-Proteobacterium *Buchnera*, setting the stage for a long term animal-microbe endosymbiosis that remains to this day (Moran et al., 1993). This relationship is so ancient and so intimate that *Buchnera* and host aphid speciation show co-divergent phylogenies (Baumann et al., 1995) as a result of strict vertical transmission from mother to offspring (Douglas, 1998). This association has persisted so long that it is now obligatory; neither the symbiont nor the host survives without the other (Douglas, 1998).

Aphids feed on phloem sap from plants, a diet rich in carbohydrates but extremely poor in nitrogen. *Buchnera* endosymbionts provide the necessary genetic material to synthesize essential amino acids from what few available amino acids are found in

the diet and from recycled cellular waste nitrogen to help the host survive (Douglas, 1998, Whitehead et al., 1992). Without these amino acids, the host would not survive. Aphids from whom *Buchnera* have been removed by antibiotic treatment suffer from delayed growth and reduced fecundity that can be partially rescued by growth on a chemically defined diet (Douglas, 1998).

Within the aphid, specialized organs known as bacteriocytes house the symbionts, and they transfer the bacteria to the ovum to prenatally infect all offspring (Douglas, 1998). It is this process which ensures that the fitness interests of the host and microbe are now aligned, as the offspring of both organisms depends on the success of the other. *Buchnera* now lacks most free-living biosynthetic necessities (Moran et al., 2008) and relies completely on the host to provide its key energy requirements and the host's diet restricts its ability to live without *Buchnera*.

The vertical transmission of a symbiont which has aligned its fitness interests with its host results in an exceptionally stable relationship. There is effectively no cheating on the part of *Buchnera* endosymbionts because of this strict transmission because a cheating *Buchnera* symbiont would soon find itself a victim of its own success. The dependence of the host on these symbionts also maintains the transmission of the symbiont (Thomas et al., 2009), and this stability has restricted these symbionts to the aphid host.

1.5.2 Attine Ants and Fungal Gargens

Attine ants in the Amazon basin have cultivated fungi for some 50 million years (Mueller et al., 2001). They grow this fungus within their nests, actively weed for contaminants, and provide plant material for the fungus to grow on, giving the popular moniker of the 'the leafcutter ants.' To prevent contamination by a fungal nest pathogen, *Escovopsis weberi*, the ants also culture bacterial communities in

specialized organs on their bodies (Currie et al., 2006). These communities produce antifungal compounds that inhibit this pathogen, keeping their fungal gardens clear of contaminants brought in by the workers (Currie et al., 2006). In related ants who do not cultivate these bacteria, highly developed cleaning behavior is observed to prevent contamination (Currie and Stuart, 2001, Little et al., 2006).

These ants preferentially cultivate their specific fungal partner, which is transmitted to each new colony by the queen (Currie et al., 1999). In experiments, ants given a choice between their natural partner and a closely related fungus choose their natural partner. When deprived of the normal species, the ants will switch to whatever strain is available, but abandon these new partners when a preferred option is provided or can be found (Mueller et al., 2004). In this way, the ants maintain a relationship with a productive fungus, and avoid less productive variants. This demonstrates an example of partner choice. By doing this, the ants can remove or punish any cheating fungi variants that emerge during cultivation and remove invading fungi from their gardens, stabilizing the relationship between the species. Maintainence of this relationship shows a co-divergent phylogeny, suggesting that this partnership has shaped both host and microbe (Hinkle et al., 1994)

1.5.3 Sepiolid Squid and *Vibrio fischeri*

One of the best characterized symbioses between animal and microbe is the relationship between the squid *Euprymna scolopes* and its bioluminescent symbiont, *Vibrio fischeri*. *E. scolopes* is a squid that lives in the shallow waters around the Hawaiian islands (Berry, 1912). By day, it burrows into the sandy bottom and at night it emerges to forage. To avoid predators, the squid employs a clever camouflage mechanism called 'counter-illumination' which masks its shadow, when seen from below (Ruby and Mcfall-Ngai, 1992), which is dependent upon the presence of *V.*

fischeri. The bioluminescent *V. fischeri* grow in specialized organs on the animal, where they aggregate and form biofilms within these organs on host mucus (Yip et al., 2006). As in vertebrates, this organ is sterile when the animal emerges from the egg, and must be rapidly colonized to confer protection for the animal. However, in the water column, many potential bioluminescent bacteria can be found. Yet, not all of them provide the requisite illumination necessary for this camouflage (Ruby and Mcfall-Ngai, 1992).

The host animal must select the correct specific constituents from seawater and experimental evidence demonstrates specific associations between *Vibrio* species and *E. scolopes* (Mcfall-Ngai and Ruby, 1991). *E. scolopes* manages this by secreting mucus to which *V. fischeri* adheres, then uses a selective pore through which only motile symbionts can pass. Potential symbionts must swim toward the light organs in the face of toxic levels of reactive oxygen species (ROS) and against flow generated by cilia lining the pore (Mcfall-Ngai and Ruby, 1998). This mechanism imposes a significant cost on all potential symbionts. Potential partners must bear the cost of resistance to ROS to survive. This cost is prohibitive for most organisms. Coincidentally, the luficerase light generating mechanism used by *V. fischeri* actively detoxifies ROS (Ruby and McFall-Ngai, 1999), allowing competent symbionts who produce light to colonize the light organ.

Despite acquiring these microbes from their environment, *E. scolopes* shows a phylogenetic codivergence with its *Vibrio* partners. Strains isolated from different species of squid showed a matching phylogenetic tree to the squid host. In addition to phylogenetic association, the strains also show host-specificity (Nishiguchi et al., 1998). Individually, these strains were able to colonize related squid, but when presented in competition with a strain of *V. fischeri* from another squid host, the native strain outcompeted the introduced strain (at 48 hr after introduction), despite a starting

(12hr after introduction) ratio of 1:1 within the light organ (Nishiguchi et al., 1998).

1.5.4 Evidence for Vertebrate Gut Symbioses

If stable, beneficial relationships between host and gut symbionts are enhanced by vertical or pseudo-vertical transmission (Wilkinson and Sherratt, 2001, Doebeli and Knowlton, 1998), then investigations using population genetics should help identify symbionts of vertebrates. At present, there are few studies which leverage this type of study to distinguish vertebrate host-gut symbiont evolutionary relationships. One of the only vertebrate gut microbes which has been studied extensively from a population genetics standpoint is *Escherichia coli*.

E. coli is a common gut commensal in vertebrates. It is found in humans, wild mammals, birds, and reptiles (Gordon and Cowling, 2003, Penders et al., 2006). Isolates from the environment can also be found, but this is assumed to be the secondary habitat of the species (Winfield and Groisman, 2003). Unlike *Buchnera*, attine ant fungi, or *V. fischeri*, which shows co-divergence with their hosts (Moran et al., 1993, Hinkle et al., 1994, Nishiguchi et al., 1998), *E. coli* does not show host specificity or co-diversification among lineages of the species, despite robust phylogenetic analyses (Tenaillion et al., 2010). This type of evolutionary pattern is consistent with a broad host range and horizontal transmission, suggesting that it is not stably associated gut symbiont.

Other organisms described as gut symbionts, such as *Bacteroidetes thetaiotaomicron*, present the functional evidence for host glycan foraging among gut isolates (Xu et al., 2003). Impressive experimental evidence for the role of these genes *in vivo* has also been shown (Goodman et al., 2009). However, there is little phylogenetic evidence tying this organism to specific hosts which complicates the use of this organism as a model of host-microbe evolution (Xu et al., 2003).

A better model is needed to address the current lack of knowledge of vertebrate gut microbe evolution. This model should demonstrate an evolutionary history with a host, it should also be tractable to study from both the host and the microbe side (ie culturable and amenable to genetic manipulation). Finally, this organism should show genetic adaptation to its niche. One such potential organism is *Lactobacillus reuteri*.

1.6 The Case of *Lactobacillus reuteri*.

Lactobacillus reuteri was first isolated in 1961 by Gerhard Reuter as "*Lactobacillus fermentum* Biotype II" from human feces (Lerche and Reuter, 1961) and was reclassified as *Lactobacillus reuteri* by Kandler et al as a new species of heterofermentative *Lactobacillus* (1980). In the following years, researchers continued to isolate additional strains of the species in humans, and chickens, pigs, and rodents. *L. reuteri* is often described as ''a commensal inhabitant" and it can be readily isolated from the gastrointestinal tract of many vertebrate hosts (Talarico et al., 1988, Axelsson et al., 1989, Chung et al., 1989, Sudenko et al., 1996). However, it does not appear to have an ecological niche other than that of the GI tract of vertebrate animals (Walter et al., 2011).

1.6.1 The Native Niche of *L. reuteri*

The proximal GI tract of pigs, chickens, and rodents is dominated by lactobacilli, including *L. reuteri*. *L. reuteri* is considered autochthonous to the forestomach of mice, and homologous proximal regions of other animals (Tannock, 2004). They can be isolated throughout the GI tract and from feces (Walter, 2008), but *L. reuteri* grow poorly below 34°C (Tobajas et al., 2007), suggesting that there is little growth of the organism outside the host.

In the proximal GI tract of rodents, pigs, birds, horses, and other animals there is a non-secretory region of the stomach covered by a keratinized squamous stratified epithelium. This epithelium is normally colonized by layers of lactobacilli and streptococci (Yuki et al., 2000, Fuller, 1973, Hilmi et al., 2007, Fuller et al., 1978, Suegara et al., 1975) (Figure 1.1), and *L. reuteri*. Here, lactobacilli have access to simple sugars, nucleotides, amino acids, vitamins, and other factors which they need for growth. In comparison to the colon, and other distal regions of the gastrointestinal tract, the low pH of this site precludes the growth of many other organisms, and lactobacilli may even contribute to the acidity of the environment (Ward and Coates, 1987).

In the forestomach, these thick layers of lactobacilli are characterized as a biofilm (Tannock, 2004, Walter et al., 2008, Tannock et al., 2005, Wilson et al., 2012), and may provide a persistent reservoir of cells that remain stably associated with the host throughout its lifetime. In addition, adherence to host tissue by the lactobacilli found in this environment has been repeatedly shown to be host-specific, as strains isolated from an animal only adhere to epithelial cells from that host animal (Lin and Savage, 1984, Wesney and Tannock, 1979, Fuller et al., 1978, Fuller, 1973). Experiments in gnotobiotic mice also showed that only strains from rodents could colonize the forestomach of these animals (Wesney and Tannock, 1979). This type of host specificity suggests that these organisms lack a secondary host. A secondary host would eliminate an ecological incentive for a particular environment and erase evidence of host specificity (Tenaillion et al., 2010).

The lack of an environmental reservoir or a secondary host necessitates a stable method for transmission between naïve hosts. *L. reuteri* can be isolated from the human vagina (Walter, 2008, Kiss et al., 2007), providing for inoculation during birth (Mandar and Mikelsaar, 1996). It can also be isolated from the breast milk of humans,

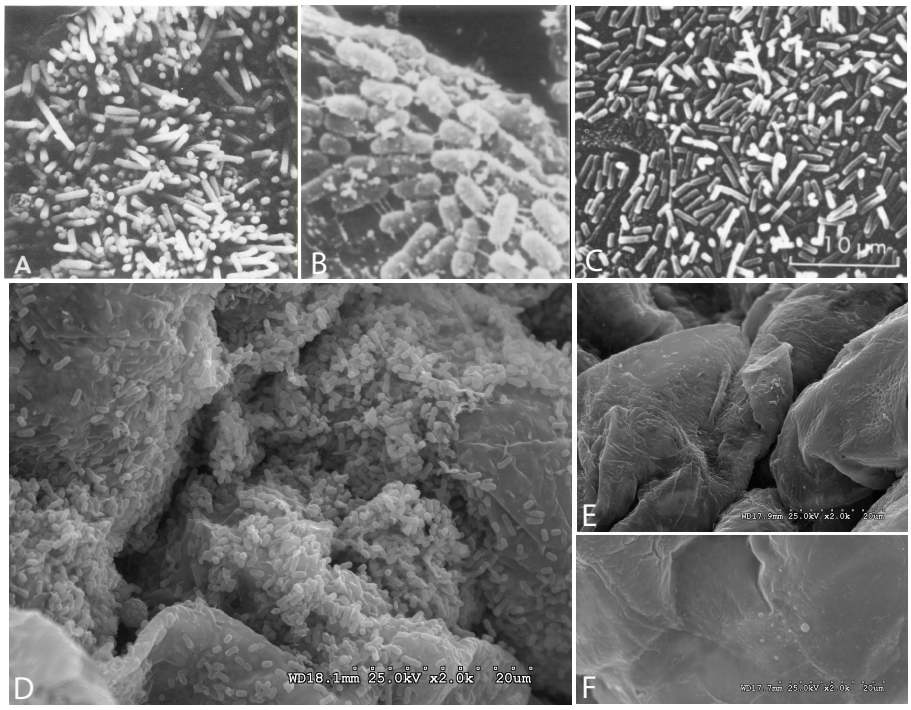

Figure 1.1 *Lactobacillus* biofilms found in conventional animals (A-C). (A) Forestomach from conventional mouse (Savage and Blumershine, 1974), (B) crop from conventional chicken (Fuller and Turvey, 1971), (C) *pars oesophagus* from conventional pig (Fuller et al., 1978). Biofilms formed in monoassociation with *L. reuteri* (D-E). (D) forestomach from germ-free mouse colonized with rodent-associated strain of *L. reuteri* (100-23), (E) forestomach from germ-free mouse colonized with human-associated strain of *L. reuteri* (DSM20016^T), (F) uncolonized forestomach from germ-free mouse.

monkeys, pigs, and dogs (Sinkiewicz and Ljunggren, 2008, Jin et al., 2011, Martín et al., 2009, 2010). This may also provide a mechanism for vertical transmission in these animals. In rodents, coprophagy provides a ready mechanism for the inoculation of offspring.

1.6.2 Evidence for *L. reuteri* host specificity

In 2010, Oh et al (2010) conducted a phylogenetic analysis of the species, and found that the isolates, obtained over the preceding fifty years from multiple continents and many locations, clustered by host of origin and not by location, based on multi-locus sequence analysis (MLSA). This study showed the first evidence for the host specific diversification of a vertebrate gut symbiont. Interestingly, isolates recovered from sourdough fermentations cluster with rodent strains and present similar phenotypes, which can be explained by the presence of rodents in grain processing facilities (Su et al., 2012). This is in stark contrast to more promiscuous organisms, such as *E. coli*, which can be found in environmental samples as well as in the GI tract, and utilizes environmental reservoirs as a secondary habitat (Tenaillion et al., 2010). The deeply diverging lineages of *L. reuteri* reflect an ancient host-microbe association, estimated to be on the order of ten million years (Oh et al., 2010).

The apparent phylogenetic divergence of populations of *L. reuteri* indicate that this organism has aligned its fitness interests with those of its specific host animals. First, isolates of *L. reuteri* from rodents persisted in animals challenged with Schaedler Lactobacilli, while isolates of *L. reuteri* from humans did not (Casas and Dobrogosz, 2000). Further evidence of this host specificity was reported by Oh et al (2010), who also demonstrated colonization discrepancies between host-associated lineages of *L. reuteri*. In this experiment, mice were colonized with a multi-strain mixture with representatives of different host-associated *L. reuteri* lineages. Eleven days after

colonization, two rodent-lineage *L. reuteri* strains dominated the gastrointestinal population. These studies provided key evidence that in addition to an evolutionary history with different hosts, this history has been accompanied by genetic differentiation reflecting the specialized adaptations to host ecosystems.

1.6.3 *Lactobacillus reuteri* as a model organism

L. reuteri presents a unique opportunity to study a vertebrate gut symbiont whose evolution is clearly tied to specific host animals. This model also provides the ability to study a host-restricted gut symbiont, whose success is tied to its host as a result of this ancient evolution. As a natural inhabitant of the rodent gastrointestinal tract, *L. reuteri* also provides an excellent model system for the study a vertebrate gut symbiont in its evolutionary niche.

L. reuteri is well characterized genetically and is a tractable model organism. Nine genome sequences have been reported for this organism, representing strains isolated from humans (4), rodents (3), pig (1), and chicken (1). These genomes have been analyzed and were used to probe more than 50 additional strains, assessing genetic content across the lineages of the species. This study revealed that in addition to host-association between lineages, specific genes were also associated with each host-associated lineage (Frese et al., 2011). To study these host specific genes, there are established genetic methods available for use in *L. reuteri* which produce labeled gene knockouts or subtle deletions (Walter et al., 2005, van Pijkeren and Britton, 2012). Plasmid based expression systems are also available (Lizier et al., 2010) as well as promoter trap systems to assess *in vivo* gene expression (Walter et al., 2003). These techniques have been used extensively to study L. reuteri in the rodent gut (Sims et al., 2011, Walter et al., 2008, 2007, 2005, Tannock et al., 2005, Walter et al., 2003).

Rodent populations have been studied for their power in mapping quantitative trait loci (QTL) to phenotypes. This has been successfully exploited to identify host genetic features which influence gut microbe populations (Benson et al., 2010), and even populations of *Lactobacillus* specifically (Buhnik-Rosenblau et al., 2011). Unlike other models of a vertebrate-gut microbe relationships, this relationship has a demonstrable past (Oh et al., 2010), rather than an incidental or pathogenic one, such as in *E. coli* (Tenaillion et al., 2010). This relationship is stable and likely persists throughout the host's life, begging the question of how this microbe persists in the gastrointestinal tract of its host and the impact that the microbe has on the host organism.

Rodent-associated strains of *L. reuteri* and their host mouse populations present a model system of vertebrate host-gut microbe symbiosis. First, both host and microbe populations can be studied from a genetic standpoint (Oh et al., 2010, Benson et al., 2010, Frese et al., 2011). Second, the ecological consequences of microbial genes can examined in their native ecosystem (eg. Walter et al. 2005, Tannock et al. 2005, Walter et al. 2007, 2008). Finally, host features can be manipulated to examine the impact of these genes on microbial colonization (Peterson et al., 2007). These methods present a powerful set of tools to examine this vertebrate-microbe partnership.

1.7 Remaining Questions: Why study a model of host-microbe symbiosis?

As studies continue to show the importance of the gut microbiota on host health and disease (eg. Turnbaugh et al. 2006, van der Waaij et al. 1971, Mazmanian et al. 2005), it becomes increasingly apparent that more information is needed to understand this system. We know very little about how the gut microbiota colonizes the host, how

individual microbes compete in the ecosystem, or which ecological principles govern the community. The careful study of well constructed models will help shed light on these important aspects of the community as a whole.

Invertebrate models of host-microbe symbiosis have greatly enhanced models of evolution and host-microbe interactions. Indirectly, these models have transformed such fields of study as pest control (Douglas, 2007) and medicine (Visick and McFall-Ngai, 2000). Extending models to the vertebrate gut also presents such potential. Individual species or strains can have a dramatic influence on host physiology (Mazmanian et al., 2005, Gaboriau-Routhiay et al., 2009), and the identification of the evolutionary and ecological forces which shape these organisms may help identify other key species and their role in the gut.

Studies are beginning to identify the potential for host genetic control of microbial populations (Benson et al., 2010, Buhnik-Rosenblau et al., 2011, Campbell et al., 2012). These studies have identified host genetic features responsible for the control of populations, specifically of *Lactobacillus* populations. The identification of a host-associated subpopulation of *L. reuteri* which can be experimentally manipulated in this context is an incredible opportunity. However, several questions about *L. reuteri* will have to be answered. First, how do the host-associated populations of *L. reuteri* differ? Are there genetic and phenotypic traits which characterize these populations and are these related to ecological performance? Second, the mechanism by which rodent associated strains colonize the rodent host must be determined. The answers to these questions will build the case for the study of the *L. reuteri* and its murine host as a model of vertebrate gut symbiont evolution.

The first steps have been made by identifying microbes which have evolutionary histories associated with specific hosts (Oh et al., 2010) and how hosts have evolved to control their microbial populations (Benson et al., 2010, Buhnik-Rosenblau et al., 2011).

These and future studies will allow researchers to examine how these interactions have evolved under natural selection. This also opens the possibility to test ecological principles that govern host-microbe and ecological interactions in a real, ecologically relevant system.

1.8 Specific Aims

The work presented in this dissertation presents a gut symbiont, *Lactobacillus reuteri*, to answer important questions about the evolution of the vertebrate host-gut microbe symbiosis. Specifically it asks;

- Have rodent-associated lineages of *L. reuteri* experienced host-specific evolution and if so, what are the genetic impacts?
- What mechanisms do these strains employ for ecological success *in vivo*?
- Do all lineages of this species exhibit comparative ecological success?

In **Chapter 2**, animal experiments provided an important mechanistic insight into the host specificity of *L. reuteri*, which revealed that only rodent-lineage strains could colonize a gnotobiotic mouse host. A population genetics approach was used to examine what genetic differences existed between these lineages. These experiments revealed that genes associated with adherence, pH resistance, carbohydrate utilization, sensing and regulation, and polysaccharide production were found in rodent-associated strains but absent in human-associated strains. Using mutant knock-out strains of eight host-specific genes, nearly all were found to be essential for strains to succeed ecologically.

In **Chapter 3**, the mechanism by which rodent-lineage *L. reuteri* persist in the murine host was examined. Biofilm formation by *L. reuteri* strains was measured using

a novel method of *in vivo* biofilm quantification. Wild-type strains from human, pig, and chicken associated lineages did not form biofilms on the forestomach epithelium in mice, where *Lactobacillus* species normally dominate. In comparison, rodent-associated strains did form these biofilms and genes responsible for aggregation and adherence were found to be critical for the formation of this structure.

To determine if host-specificity was present in other lineages, a human *L. reuteri* strain was tested for persistence in the human host (**Chapter 4**). In comparison with two other strains, a human-associated strain of *L. mucosae* and a non-associated strain or *L. acidophilus*, *L. reuteri* reached higher population densities than the non-associated strain. However, none of the strains tested persisted longer than the others.

Together, these experiments present a comprehensive examination of rodent-associated *L. reuteri* strains, their evolution, and their mechanisms for ecological success *in vivo*. These experiments established (i) host specificity for rodent strains, (ii) determined the genetic features which contribute to their ecological success, (iii) developed a method to examine *in vivo* biofilm formation, and (iv) investigated the ecological fitness of a human-associated *L. reuteri* strain.

Chapter 2

The Evolution of Host Specialization in the Vertebrate Gut Symbiont *Lactobacillus reuteri*

Steven A. Frese, Andrew K. Benson, Gerald W. Tannock, Diane M. Loach, Jaehyoung Kim, Min Zhang, Phaik Lyn Oh, Nicholas C. K. Heng, Prabhu B. Patil,, Nathalie Juge, Donald A. MacKenzie, Bruce M. Pearson, Alla Lapidus, Eileen Dalin, Hope Tice, Eugene Goltsman, Miriam Land, Loren Hauser, Natalia Ivanova, Nikos C. Kyrpides, Jens Walter

Preface

This chapter was published in 2011, in the journal *PLoS Genetics*. Portions of this research were presented in 2009, at the Sackler Colloquium on ‘‘Microbes and Health” at the National Academy of Sciences in Irvine, California.

2.1 Abstract

Recent research has provided mechanistic insight into the important contributions of the gut microbiota to vertebrate biology, but questions remain about the evolutionary processes that have shaped this symbiosis. In the present study, we showed in experiments with gnotobiotic mice that the evolution of *Lactobacillus reuteri* with rodents resulted in the emergence of host specialization. To identify genomic events marking adaptations to the murine host, we compared the genome of the rodent isolate *L. reuteri* 100-23 with that of the human isolate *L. reuteri* F275, and we identified hundreds of genes that were specific to each strain. In order to differentiate true host-specific genome content from strain-level differences, comparative genome hybridizations were performed to query 57 *L. reuteri* strains originating from six different vertebrate hosts in combination with genome sequence comparisons of nine strains encompassing five phylogenetic lineages of the species. This approach revealed that rodent strains, although showing a high degree of genomic plasticity, possessed a specific genome inventory that was rare or absent in strains from other vertebrate hosts. The distinct genome content of *L. reuteri* lineages reflected the niche characteristics in the gastrointestinal tracts of their respective hosts, and inactivation of seven out of eight representative rodent-specific genes in *L. reuteri* 100-23 resulted in impaired ecological performance in the gut of mice. The comparative genomic analyses suggested fundamentally different trends of genome evolution in rodent and human *L. reuteri* populations, with the former possessing a large and adaptable pan-genome while the latter being subjected to a process of reductive evolution. In conclusion, this study provided experimental evidence and a molecular basis for the evolution of host specificity in a vertebrate gut symbiont, and it identified genomic events that have shaped this process.

2.2 Author Summary

The gastrointestinal microbiota of vertebrates is important for nutrient utilization, resistance against pathogens, and immune maturation of its host, but little is known about the evolutionary relationships between vertebrates and individual bacterial members of these communities. Here we provide robust evidence that the evolution of the gut symbiont *Lactobacillus reuteri* with vertebrates resulted in the emergence of host specialization. Genomic approaches using a combination of genome sequence comparisons and microarray analysis were used to identify the host-specific genome content in rodent and human strains and the evolutionary events that resulted in host adaptation. The study revealed divergent patterns of genome evolution in rodent and human lineages and a distinct genome inventory in host-restricted sub-populations of *L. reuteri* that reflected the niche characteristics in the gut of their particular vertebrate hosts. The ecological significance of representative rodent-specific genes was demonstrated in gnotobiotic mice. In conclusion, this work provided evidence that the vertebrate gut symbiont *Lactobacillus reuteri*, despite the likelihood of horizontal transmission, has remained stably associated with related groups of vertebrate hosts over evolutionary time and has evolved a lifestyle specialized to these host animals.

2.3 Introduction

Vertebrates are associated with trillions of microbes, the majority of which inhabit the digestive tract (Ley et al., 2006). Research has led to an appreciation of the importance of these microbial communities, revealing substantial roles in development and performance of the host (Cerutti and Rescigno, 2008, Dethlefsen et al., 2007). As vertebrates evolved, they did so in association with microbes, and

these reciprocal interactions have shaped both the attributes of the microbiomes and the phenotypic complexity of the host species (Ley et al., 2008a). It is conceivable that the beneficial functions of the gut microbiota conferred important selective traits during vertebrate evolution (Dethlefsen et al., 2007, Walter et al., 2011). A joint evolutionary trajectory between host and microbes is evident in anatomical features of vertebrates (rumen, cecum) which allow bacterial fermentations that provide energy to the host and an intensive gut associated immune system that is in place to maintain beneficial microbial communities (McFall-Ngai, 2007, Stevens and Hume, 1998). These features serve as clear testimony that we cannot attempt to understand the evolution of vertebrates without considering their microbial partners (Ley et al., 2006, 2008b).

Comparative analysis of genomes of bacteria originating from human hosts, greatly facilitated through the Human Microbiome Project, provided important insight into the adaptations and ecological roles of different microbial species in the human gut (Sela et al., 2008, Turnbaugh et al., 2007). Despite these advances, very little is known about the evolutionary strategies of vertebrate gut symbionts. It is often postulated that the evolution of gut microbes involved coevolution of individual lineages with their host species, which is supported by the presence of phylotypes that are specific to particular vertebrate species (Dethlefsen et al., 2007). However, conclusive evidence for stable associations of specific lineages with vertebrate hosts over evolutionary time-scales has not been provided by 16S rRNA data. Patterns of community similarity provide evidence for codiversification of entire gut communities with their hosts, which suggests that there are host-specific evolutionary interactions between mammals and their microbiomes (Ley et al., 2008a). In addition, some gut microbes are highly host specific, such as *Helicobacter pylori*, which has been used to track human migrations over long-time spans (Linz et al., 2007). However, many microbial lineages in the mammalian gut are shared across host species (Ley et al.,

2008a), implying that some members of the gut microbiota are generalists that pursue promiscuous lifestyles. Such an evolutionary strategy is exemplified by commensal *Escherichia coli*, which have a broad host range and alternate between niches within the environment and their vertebrate hosts (Tenaillion et al., 2010, Touchon et al., 2009). To date, there are very few vertebrate gut symbionts for which host specificity has been clearly established. Furthermore, little is known about the mechanisms by which gut microbes, for whom symbiotic life is facultative and which have ample opportunities for horizontal transmission, can evolve stable associations with their host species that would allow for reciprocal evolutionary interactions between bacterial lineages and host genotypes.

The Gram-positive bacterium *Lactobacillus reuteri* is an excellent model organism to study the evolutionary strategy of a vertebrate gut symbiont as this species inhabits the gastrointestinal tract (GIT) of mammals as diverse as humans, pigs, mice, and rats as well as different species of birds. In rodents, pigs, and chickens, it is one of the dominant species in the GIT and forms biofilm-like associations with the stratified squamous epithelial lining of the proximal regions of the digestive tract (Brooks et al., 2003, Walter, 2008). We recently observed that strains of *L. reuteri* from global sources comprised distinct phylogenetic clusters that can be detected with Multilocus Sequence Analysis (MLSA) and Amplified Fragment Length Polymorphism (AFLP), and these clades show significant association with host origin (Oh et al., 2010). The population structure suggests a stable association of *L. reuteri* with particular vertebrates over evolutionary time and the emergence of host adapted subpopulations. In addition to the genotypic patterns, an adaptive evolutionary process is also reflected by the phenotypic characteristics of *L. reuteri* strains in terms of ecological performance in the gut and adhesion to epithelial cells (Oh et al., 2010, Suegara et al., 1975). However, the molecular basis for these host adaptations is still unknown, and it is unclear

to what degree the lifestyle and evolution of *L. reuteri* have remained restricted to particular hosts.

Genomic approaches in combination with experiments in animal models offer mechanistic insight into the evolution and ecology of microbial symbionts of vertebrates. In this study, we used such an approach and showed that only rodent isolates of *L. reuteri* colonize the gut of reconstituted *Lactobacillus*-free (LF) mice in high numbers, while isolates from humans, swine, and chicken form either lower populations or fail to colonize. We determined the genome sequence of the rodent isolate *L. reuteri* 100-23 and performed a comparative genomic analysis with the genome of the human isolate F275. A microarray analysis using genes representative of both strains was used to probe 57 *L. reuteri* strains, revealing specific gene combinations in host-adapted lineages of *L. reuteri*. Further genomic comparisons of nine isolates across five MLSA lineages confirmed the microarray data and further allowed the identification of the evolutionary processes that resulted in host-specific genomic features. The ecological significance of rodent-specific genes was demonstrated in gnotobiotic mice, where perturbations in 7 out of 8 genes unique to the rodent lineage resulted in impaired ability to propagate in the murine host.

2.4 Results and Discussion

2.4.1 Identification of Host-Specific Genome Content in *L. reuteri* Strains

Genome sequence of the rodent *L. reuteri* strain 100-23.

The evolution of host specialization in bacteria, which is very well understood for symbionts of invertebrates, can follow diverse paths that can include large scale

gene acquisitions and loss and more subtle modifications such as modifications of gene (protein) sequences and regulatory pathways (Mandel et al., 2009, Moran, 2007, Moran and Plague, 2004, Moya et al., 2008, Dale and Moran, 2006). The availability of *L. reuteri* isolates that differed in their ability to colonize the GIT of LF mice paved the way for a genomic analysis to identify the molecular basis of host specificity in a vertebrate symbiont. The sequence of the human isolate F275 was previously completed (Morita et al., 2008), and four human isolates (CF48-3a, MM4-1, MM2-3, ATCC 55730) were sequenced in the course of the Human Microbiome Project (Nelson et al., 2010). We produced a high quality genome sequence of *L. reuteri* 100-23, an isolate from the stomach of a rat that served as a model organism for mechanistic studies on gut microbial ecology, *in vivo* biofilm formation, and immunology in combination with LF mice (Wesney and Tannock, 1979, Tannock et al., 2005, Walter et al., 2005, 2003, 2007, Hoffmann et al., 2008, Livingston et al., 2010). This strain groups with the rodent-associated MLSA lineage III (Oh et al., 2010). Genome sequencing resulted in two scaffolds of 729,351 bp and 1,576,206 bp with a total of 2375 detected genes, consisting of 2269 protein-coding genes, and 106 RNA genes (including six rRNA operons). The general genome features are listed in Table A.1. *L. reuteri* 100-23 harbors two indigenous plasmids, pGT231 (5,254 bp; GenBank accession no. GU108604) and pGT232 (5,123 bp, GenBank accession no. NC_001757), with pGT232 belonging to the pC194/pUB110 family of rolling-circle plasmids (Heng et al., 1999). The significance of these plasmids for the biology of *L. reuteri* has not yet been determined, and the plasmid sequences were not included in the comparative genomic analysis below.

Strain	Origin	Number of Animals	Fecal Sample (Day 2)	Forestomach (Day 14)	Cecum (Day 14)
N2D	Rat	7	7.5	8.7 (0.1)[1]	8.6 (0.3)
100-23	Rat	5	ND[2]	8.0 (0.2)	7.4 (0.2)
6799-JM1	Mouse	7	6.5	8.7 (0.1)	7.8 (0.3)
#20	Mouse	6	ND	8.6 (0.1)	7.7 (0.2)
JW2015	Pig	7	5.1	6.0 (0.8)	4.7 (0.6)
LPA1	Pig	6	4.4	< 2.0	< 2.0
DSM20016^T	Human	5	4.2	< 2.0	< 2.0
CF4-6G	Human	7	< 2.0	< 2.0	< 2.0
ATCC55730	Human	7	3.5	< 2.0	< 2.0
M27U15	Human	6	2.8	< 2.0	< 2.0
CF48-3A1	Human	Females 3	4.3	5.5 (0.8)	4.0 (0.8)
		Males 4	3.4	< 2.0	< 2.0
1366	Chicken	6	4.3	< 2.0	< 2.0
CSF8	Chicken	6	< 2.0	< 2.0	< 2.0

Table 2.1 Log_{10} lactobacilli per gram of organ or fecal samples of ex-*Lactobacillus*-free mice inoculated with a single strain of *Lactobacillus reuteri*.
(1) Mean (standard error of the mean); (2) ND = not done.

Comparison of the genomes of the rodent strain 100-23 and the human isolate F275.

L. reuteri F275 is a fecal isolate from a healthy adult human and a member of the human-associated MLSA lineage II (Clonal complex CC-47) (Oh et al., 2010). The strain is unable to colonize LF mice (Table 2.1), and its genome is around 270 kb smaller than 100-23, containing around 290 fewer genes (Table A.1). F275 does not contain any plasmids, and in contrast to 100-23, it contains 35 pseudogenes. A whole genome BLASTP comparison revealed that *L. reuteri* 100-23 contains 633 genes with no orthologues in F275, while the latter has 352 genes without an orthologue in 100-23. A summary of the unique genes is given in Table A.2. Both genomes contained more than a hundred genes annotated as transposases (most with homologies to described IS elements), integrases, and phage related proteins, many of which were strain specific. Genes with assigned functions that are unique to the two genomes included genes coding for cell wall and membrane bound proteins, transport proteins, regulatory proteins, enzymes, and glycosyltransferases. An auxiliary protein secretion system (SecA2 cluster) and a urease gene cluster were unique to 100-23, while only F275 contained the pdu-cbi-cob-hem cluster (Morita et al., 2008, Sriramulu et al., 2008).

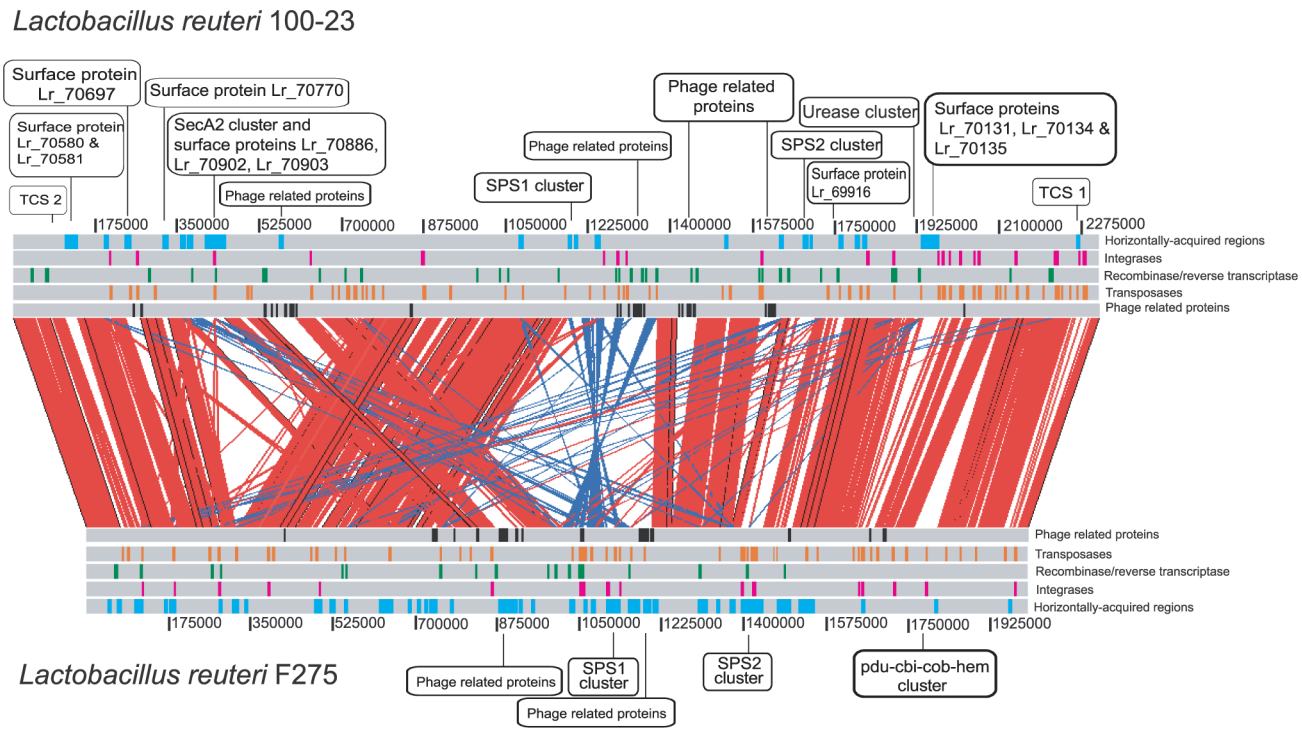

Figure 2.1 Linear genomic comparison of the chromosomes of 100-23 and F275 (using the sequence of JCM1112^T). Both sequences are read left to right from the predicted origin of replication. Homologous regions within the two genomes identified by reciprocal BLASTN are indicated by red (same orientation) and blue (reverse orientation) bars. Putative horizontally acquired islands as identified by Alien_hunter (blue boxes), phage proteins (black boxes), transposases (orange boxes), and integrases (pink boxes) are indicated.

We used the Artemis Comparison Tool to localize strain specific genomic regions in *L. reuteri* 100-23 and F275 (Figure B.1). The two genomes contained many regions of synteny, especially around the origin of replication. However, one major rearrangement and a major inversion were also present. The rearrangement was likely to have occurred in 100-23 as the F275 genome sequence shows greater synteny with the genome of the related species *Lactobacillus fermentum* (Figure B.1), and it is therefore likely to reflect the ancestral structure. In addition, the inversion within the genome of 100-23 was rich in genetic elements (e.g. transposases), which may have caused the rearrangement through a recombination event (Figure 2.1). Many of the genes that were unique to 100-23 or F275 were clustered in genomic regions that were completely absent in the other strain. Several of these regions showed characteristics of genomic islands as they were associated with unusual sequence features such as low %GC content, atypical codon bias, mobile genetic elements (prophage related genes or putative IS elements/transposons), and they were predicted to be transferred by lateral gene transfer (LGT) using the software Alien_hunter. Several other regions were identified to be present in both genomes but differed significantly in terms of gene content. These regions coded for genes involved in the production of surface polysaccharides (SPS1 and SPS2) or contained putative prophages.

Genomic survey of 57 *L. reuteri* strains from different vertebrate hosts.

To differentiate true host-specific gene content from strain-level differences, the genomes of 100-23 and F275 were used to design representative spotted genomic microarrays. These were used to interrogate the genome content of 55 additional *L. reuteri* strains (>99% homology on 16S rRNA sequence to the type strain F275) by comparative genomic hybridization (CGH). The strain collection was composed of isolates from six different vertebrate hosts that belong to five distinct MLSA lineages

(Table S3). MLSA previously revealed that human isolates belong to two separate lineages (II and VI) (Oh et al., 2010), and representatives from both lineages were included in CGH. MLSA lineage II is mainly composed of strains isolated from human fecal samples, and strain F275, which falls within this cluster, has been reported to be detectable in fecal samples of the same individual for around six months (Reuter, 2001). The strains from MLSA lineage II are therefore likely autochthonous to the human digestive tract. In contrast, human isolates from lineage VI primarily originate from other body parts (vagina, mouth, breast-milk), and they group tightly with strains from poultry, indicating that they are allochthonous to the human GIT (Oh et al., 2010).

CGH patterns were analyzed using the MARKFIND program (Zhang et al., 2003), which performs a cluster analysis based on genome polymorphisms by the unweighted pair-group method with arithmetic means (UPGMA). As shown in Figure 2, the phylogeny inferred from the genome polymorphisms reflected both host origin and MLSA typing. The rodent strains formed one cluster comprised of three sub-groups (i, ii, and iii), with i and ii corresponding to MLSA lineages III and I, respectively. The human and poultry isolates that belong to the MLSA lineage VI also formed one separate group with CGH. The human strains from MLSA cluster II formed an isolated cluster that grouped distantly from all other *L. reuteri* strains, which is indicative of markedly different genome content (Figure 2.2). Thus, although the topologies of the dendrograms inferred from gene polymorphisms and MLSA sequences were different, both methods resulted in trees with host-specific phylogenetic clusters that were congruent. This indicates that *L. reuteri* has diverged into genetically and ecologically cohesive subpopulations (ecotypes) whose gene content reflects particular host niches.

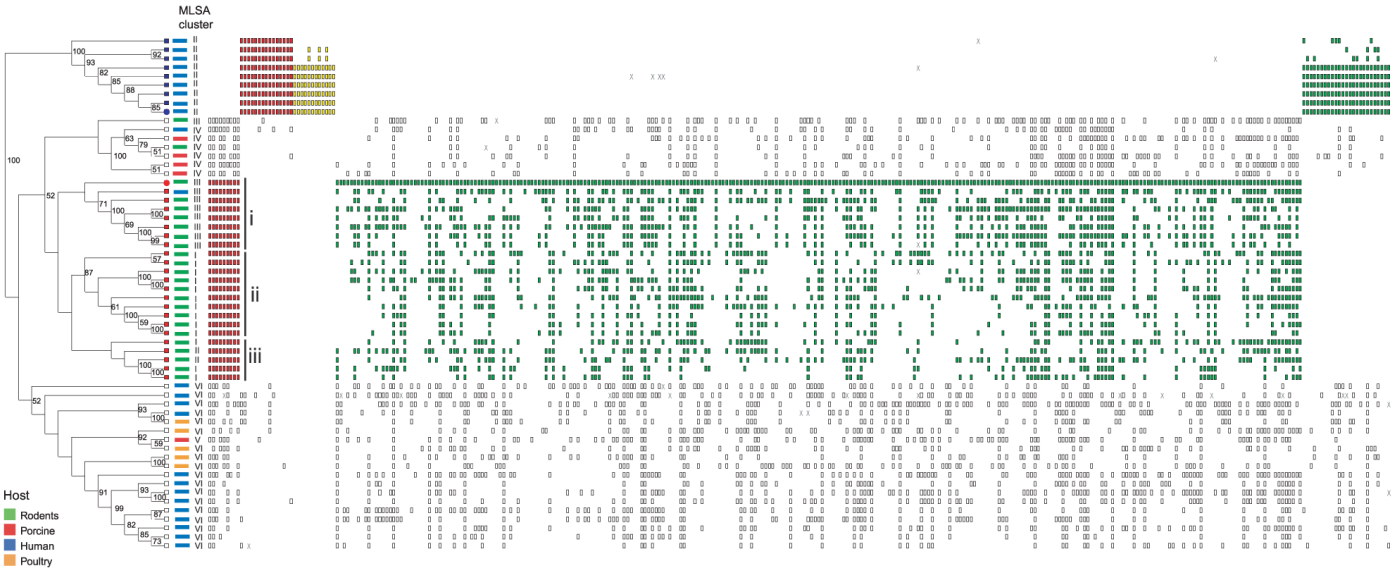

Figure 2.2 The dendrogram is derived from UPGMA analysis of binary data generated from 57 strains of *Lactobacillus reuteri*. Bootstrap scores, using 10,000 repetitions of a UPGMA search, are shown on nodes scoring >50%. Each strain is color coded by host, and affiliation to MLSA lineages of each strain is indicated. MARKFIND was used to identify gene polymorphisms specific to human (MLSA lineage II) and rodent strains (mostly MLSA lineages I and III), which clustered in the dendrogram. The leaves of human (lineage II) and rodent strains, which have been used for comparisons by MARKFIND, are colored blue and red, respectively. The two reference strains, 100-23 and F275, are labeled by a red and a blue circle, respectively. Vertical rectangles to the right depict polymorphisms present in a given strain sorted by the MARKFIND program. Those polymorphisms conserved within all members of a lineage and absent in the other lineage are colored red. Genes that are lineage-specific but non conserved are colored green (polyphyletic) or yellow (monophyletic). Polymorphisms that are not lineage specific are not shown.

2.4.2 Evolutionary genomics of *L. reuteri*.

CGH with multiple-strain comparisons both within and across *L. reuteri* lineages allowed us to identify evolutionarily- and ecologically-relevant patterns of genome variation. We used MARKFIND to identify genes that were unique to rodent strains (including 100-23) when compared to strains of the human MLSA cluster II (including F275) and vice versa, and these polymorphisms are represented by red, green, and yellow rectangles in Figure 2.2. We focused this analysis on the differences of rodent and human lineage II strains as the spotted microarray was based on the genomes of strains that belong to these groups, assuring more reliable hybridizations (see Materials and Methods). MARKFIND identified eight genes that are conserved in all rodent strains but absent in human strains and 256 genes that were specific to rodent strains but non-conserved (Table A.3). 15 genes were identified to be conserved among the human cluster II and absent in rodent strains, while 37 genes were identified to be specific to human strains but not conserved (Table A.5). It is of note that all genes identified as 'rodent-specific' when compared to the human lineage II were also detected in at least some strains associated with pigs and poultry.

A summary of the host-specific genes detected by MARKFIND in comparisons between rodent and human MLSA lineage II strains is presented in Table 2.2. Many of these genes are mobile genetic elements. In addition, almost half of the rodent-specific genes encoded hypothetical proteins with unknown function that showed a very low conservation even among rodent strains. Only 10 genes with a functional annotation other than transposition were specific to the human lineage. Most of these genes were glycosyl-transferases from the SPS2 cluster and several enzymes (histidine decarboxylase, histidyl-tRNA synthetase, dextransucrase, two lipolytic proteins). Rodent strains possessed 93 host-specific genes with assigned functions

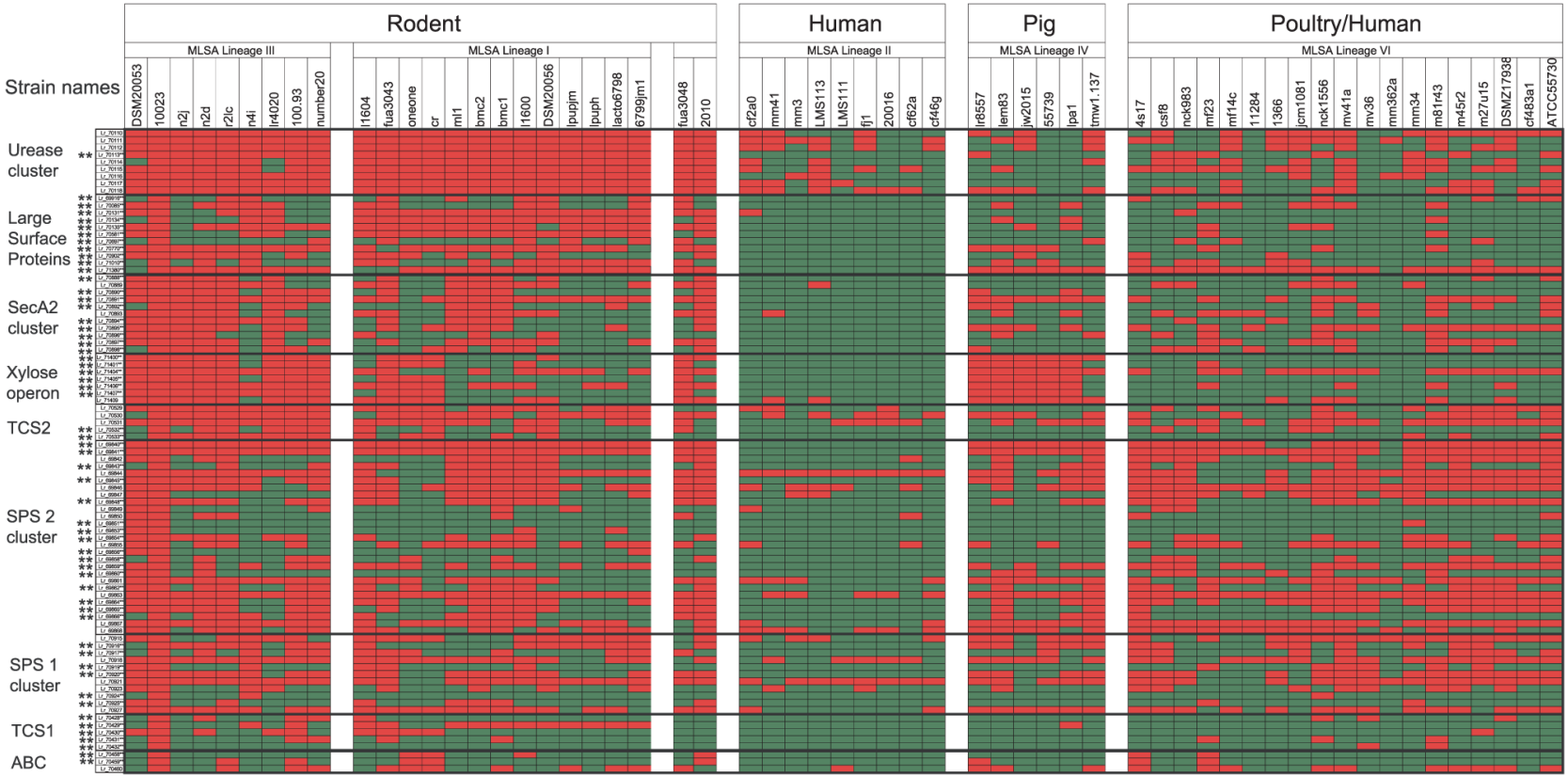

Figure 2.3The strains are arranged by host and MLSA lineage (Oh et al., 2010). Red and green colors indicate presence or absence, respectively. Genes that have been identified as rodent-specific when compared to human lineage II strains by MARKFIND are indicated by asterisks (**).

other than DNA transposition. Most of these genes fell within the putative genomic islands identified above (Figure 2.1). The distribution of the genes within these islands among all strains included in the CGH analysis is shown in Figure 2.3. The urease cluster was the only feature that was both conserved across rodent strains and absent in isolates from other hosts. All other rodent-specific clusters showed different degrees of strain-to-strain variation. Genes encoding eleven large surface proteins and the Two-Component Regulatory System (TCS) TCS2 were rare in isolates from non-rodent hosts, while the xylose cluster and the asp3 gene of the SecA2 cluster were also detectable in isolates from pigs. The two SPS clusters, though to a large degree host-specific, showed a very high variability in gene composition among rodent strains and were also detectable in strains of lineages IV (pig) and VI (poultry/human). A second regulatory system (TCS1) and the Multidrug efflux cluster (ABC) were only detectable in a small number of rodent strains.

2.4.3 Validation of Host-Specific Gene Content by Genome Comparisons and PCR

The species *L. reuteri* shows a significant degree of genetic variation, especially between strains from different MLSA lineages (Oh et al., 2010). Sequence divergence can confound the CGH data as it impairs hybridizations. This was apparent because even though hybridizations were very reliable for the genomes of the reference strains 100-23 and F275 (>96% accuracy), the error rate was approximately 18.5% for strain CF48-3A of lineage VI. Therefore, to confirm the findings obtained with the CGH analysis and to gain further insight into the distribution of host-specific gene content throughout the entire *L. reuteri* population, we performed additional genomic comparisons in combination with PCR. First, we generated draft genome sequences

(>15x coverage) of two additional rodent strains (lpuph1 and MLC3) and one pig strain (ATCC 53608). We then determined the presence of the host-specific genes identified by CGH and the pdu-cbi-cob-hem cluster in all available *L. reuteri* genomes (100-23, lpuph1, MLC3, ATCC 53608, F275, MM4-1a, MM2-3, ATCC55730, and CF48-3A). These genomes represent five MLSA lineages, lineages I and III (rodent), lineage II (human), lineage IV (pig), and lineage VI (poultry/human), and the genome characteristics are shown in Table A.6. The average nucleotide identity (ANI) of a core set of genes within these *L. reuteri* genomes and *L. vaginalis* is shown in Table A.7. An ANI of >95% was determined in all the *L. reuteri* genome comparisons, providing additional evidence that these strains, despite their considerable genomic differences, fall within what is currently considered to be one prokaryotic species (Goris et al., 2007).

As shown in Figure 2.4A, the genomic comparisons confirmed the findings obtained with the CGH analysis. The pdu-cbi-cob-hem cluster was detected in all human isolates (MLSA lineage II and VI) and the pig isolate ATCC 53608 (MLSA lineage IV), but it was only present in one of the three rodent strains. The urease cluster was strictly conserved among the three rodent strains and absent in all other

Genes	**Rodent**	**Human**
Transposases/Integrases	21	2
Phage-related proteins	32	16
DNA Binding and restriction endonucleases	11	8
Urease gene cluster	1	0
Cell Wall/Membrane bound proteins	12	0
Transport proteins (ion, peptide, sugar)	19	0
Regulatory proteins	11	1
Enzymes (peptidases, hydrolases, dehydrogenases, kinases, amylases, reductases)	25	5
Glycosyl transferases and sugar isomerases/epimerases	25	4
Hypothetical and unknown proteins	107	16
Total	264	52

Table 2.2 Host-speific genes in rodent and human lineage II strains as identified by CGH

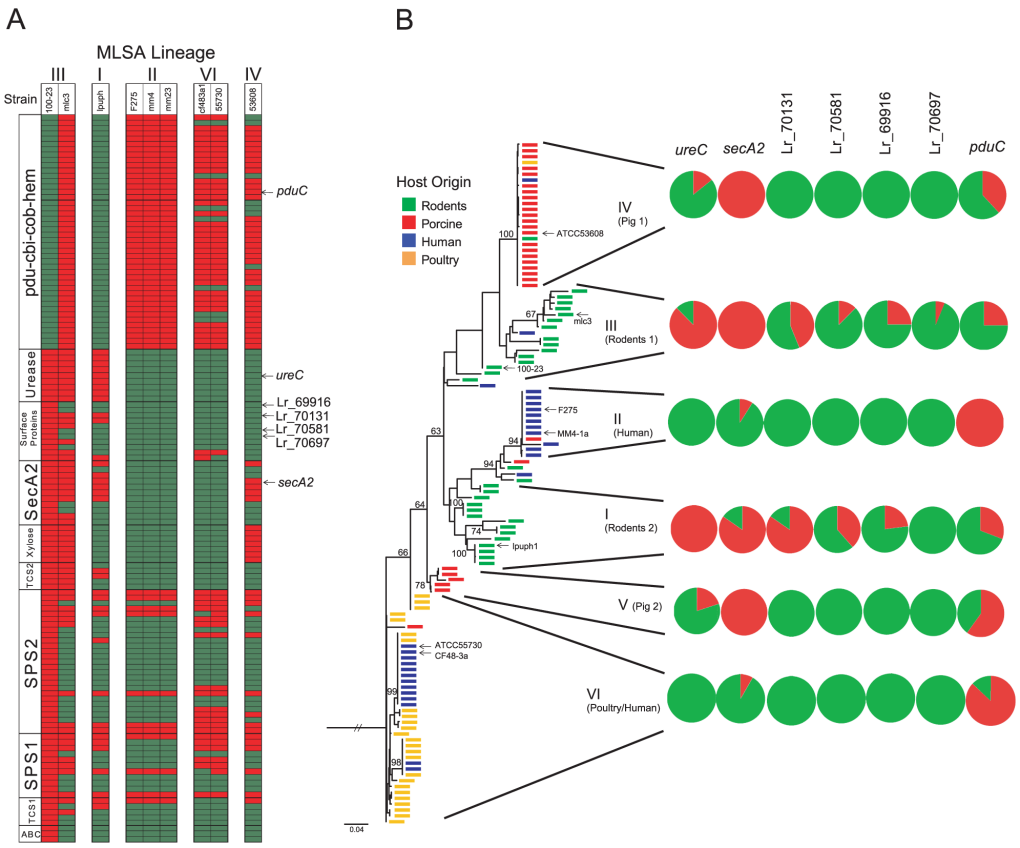

Figure 2.4(A) A heatmap representation of BLASTN comparison of rodent-specific genes against sequenced *L. reuteri* genomes. Genes whose distribution throughout the species *L. reuteri* was determined by PCR are marked by arrows. (B) Distribution of host-specific genes throughout the phylogenetic spectrum of *L. reuteri*. A maximum likelihood tree of MLSA data from 116 *L. reuteri* strains is shown [20], and the strains for which genome sequences were included are marked by arrows. Pie charts showing the proportion of queried strains within lineages possessing the targeted genes (ureC, urease; Lr_70892, secA2, surface proteins Lr_70131, Lr_70581, Lr_70697, and pduC of the pdu-cbi-cob-hem cluster) as detected by PCR.

genomes, while the surface proteins and the TCS2 cluster were to a large degree specific to rodents but more variable. The SecA2 and xylose clusters were detectable in rodent and porcine strains but completely absent in strains from lineage II and VI. The SPS and TCS1 clusters showed a much higher variability among rodent strains and several of the genes were datable in the lineage VI and IV strains, while most of the genes were absent in human lineage II strains. Consistent with CGH, the ABC transporter was specific to strain 100-23. To study the distribution of host-specific genomic features throughout the *L. reuteri* population, PCR was used to determine the presence of genes encoding SecA2, several surface proteins (Lr_70131, Lr_70581, Lr_70697, Lr_69916), UreC (the urease alpha subunit), and PduC (diol/glycerol dehydratase encoded by the pdu-cbi-cob-hem cluster) in 88 *L. reuteri* strains (Table A.3). The results are shown in Figure 2.4B in a phylogenetic context. This analysis confirmed that several of the key genetic determinants identified by CGH are to a large degree associated with specific MLSA lineages and vertebrate hosts.

2.4.4 Genomic Features Associated with Host Origin

The urease cluster.

This cluster (genes Lr_70110–Lr_70118 in the genome of 100-23) is highly conserved among rodent strains and highly host specific (Figure 2.3 and Figure 2.4). This is in accordance to previous phenotypic characterizations which showed that urease activity can be detected in rodent *L. reuteri* isolates, while the activity is rare in porcine isolates and absent in human and poultry isolates (Walter et al., 2011). Genes for urease production are absent in all currently available genomes of other Lactobacillus species, but orthologs (42–75% amino acid identity) are present in Streptococcus species, suggesting that the cluster was acquired by *L. reuteri* through

LGT. Urease has been shown to be an important component of survival in acidic conditions as well as in biofilm communities by ameliorating the buildup of acidic metabolic end-products (Li et al., 2000, Sjostrom and Larsson, 1996), by which it could contribute to the survival of *L. reuteri* in the forestomach of rodents.

Large surface proteins.

Eleven large (>750 aa) surface proteins were detected in rodent strains that were very rare in isolates of pigs and poultry and absent in human isolates of MLSA lineage II (Figure 2.3). The characteristics of these proteins are described in Table 2.3 and their schematic representation is shown in Figure 2.5. Most of the surface proteins are predicted to be involved in epithelial adhesion and biofilm formation. Six proteins (Lr_69656, Lr_70131, Lr_70134, Lr_70135, Lr_70581, Lr_71380) contained putative mucin-binding MucBP domains (Pfam PF06458) and other domains involved in extracellular matrix binding. Additional domains detected included a glycosyl-transferase (family 68) domain in a predicted levansucrase (Lr_71010), and a putative IgA-specific protease (Lr_69916).

The accessory Sec (SecA2) system.

The SecA2 cluster was detected by PCR in most strains from rodents and pigs (MLSA lineages I, III, IV, and V), while it is rare in isolates from human and poultry hosts (MLSA lineages II and VI). This auxiliary protein secretion system is present in a limited number of gram-positive bacteria and mycobacteria in addition to the canonical SecA system (Rigel and Braunstein, 2008). Conservation of the SecA2 cluster with other members of the Class Bacilli and sparse distribution among different species of lactobacilli implies that this system was horizontally acquired by only a few *Lactobacillus* lineages. LGT of this cluster in *L. reuteri* is supported by the presence

Protein(AA)	**Similar proteins (name and/or GenBank entry)**	**Organisms (%Identity/no amino acids)**	**Features (Interpro)**
Lr_70770 (3787)	NP_964984.1	*L. johnsonii* NCC533 (55/3971)	YSIRK signal peptide, LPXTG anchor, 28 DUF1542 motifs (Pfam07564), 3 Ebh (extracellular matrix-binding) motifs (Pfam07554)
Lr_70697 (2129)	ZP_04009761	*L. salivarius* ATCC 11741 (27/1040	LPXTG motif, signal peptide
Lr_70131 (1828)	Lr_70581	*L. reuteri* 100-23 (85/807)	YSIRK signal peptide, LPXTG anchor, 6 MucBP motifs (Pfam 06458)
	Mlp	*L. reuteri* BR11 (78/802)	
Lr_70134 (4634	YP_194248	*L. acidophilus* NCFM 75(3130)	YSIRK signal peptide, LPXTG anchor, 19 MucBP motifs (Pfam 06458)
Lr_70135 (4141)	LJ0484/NP_964510	*L. johnsonii* NCC 533 (80/2495)	LPXTG motif, 6 MucBP motifs (Pfam 06458)
Lr_70581 (4454)	Mlp/AAP41738	*L. reuteri* BR11 (73/2094)	YSIRK signal peptide, LPXTG anchor, 18 MucBP motifs (Pfam 06458)
	YP_193900	*L. acidophilus* NCFM (58/1166)	
Lr_70902 (2180)	LJ1711	*L. johnsonii* NCC 533	LPXTG motif, shares similarities (around 15%) with trophinin from mice and rat, a host protein.
Lr_69916 (1235)	LJ1680	*L. johnsonii* NC 533 (54/1234)	LPXTG motif, IgA protease conserved among streptococci
Lr_70085 (1034)	EEW53535.1	*L. antri* DSM16041 (30/483)	Rib-binding domain, LPXTG anchor
Lr_71010 (796)	ZP_03974578	*L. reuteri* CF483A1 (82/591)	Levansucrase Ftf,
		L. sanfranciscensis (51/451)	KxYKxGKxW signal peptide, LPXTG anchor
Lr_71380 (1036)	ZP_03975092.1	*L. reuteri* CF483A1 (90/516)	2 MucBP domains
		L. gasseri JV-V03 (24/103)	

Table 2.3 Large surface proteins (> 750 amino aids) specific to rodent strains

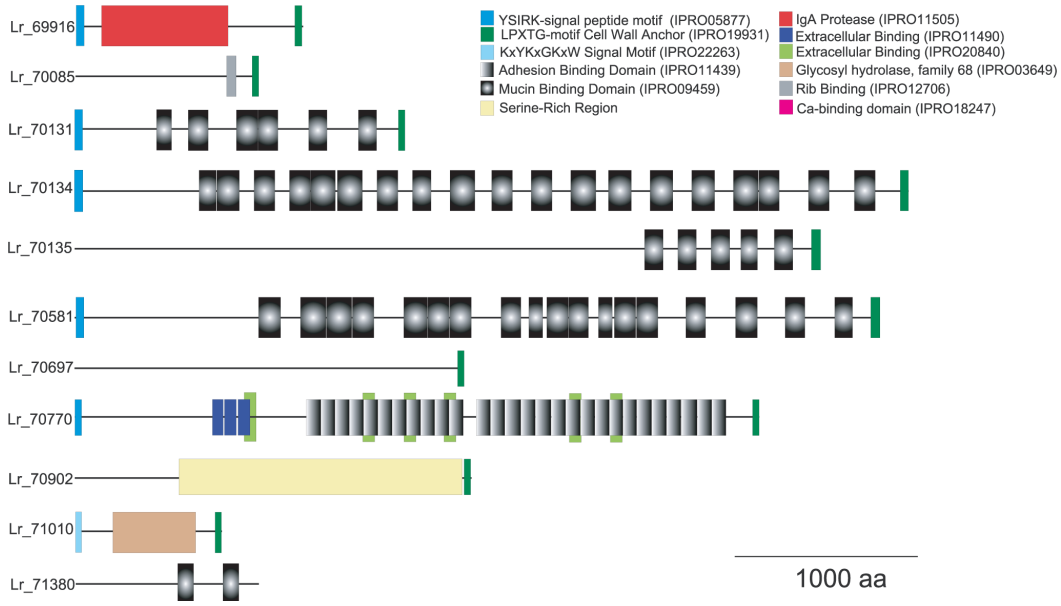

Figure 2.5 The architecture of large surface proteins (>750 aa) specific to rodent strains when compared to human isolates of the MLSA lineage II. Functional domains of each protein are shown as indicated and to scale.

of mobile genetic elements (Lr_70899 and Lr_70901) within the cluster, a low GC content (Figure 2.6), and by analysis with Alien_hunter (Figure 2.1). As shown in Figure B.2, gene content within the accessory Sec cluster is conserved in L. gasseri, Streptococcus gordonii, and *L. reuteri* 100-23. In streptococci, the accessory SecA2 system facilitates the selective export of glycosylated serine-rich proteins that often function as adhesins (Rigel and Braunstein, 2008, Bensing and Sullam, 2009). Though we do not yet know which proteins are secreted through this pathway in *L. reuteri*, the surface proteins Lr_70886, Lr_70902, and Lr_70903 are adjacent to the cluster in the genome of 100-23. Of those, Lr_70902 is unusually serine rich (35% serine), and the serine residues may be glycosylated by glycosyltransferases associated with the SecA2 cluster (Lr_70896–Lr_70898) analogous to serine rich surface protein in streptococci, such as GspB (Rigel and Braunstein, 2008, Bensing and Sullam, 2009, 2002).

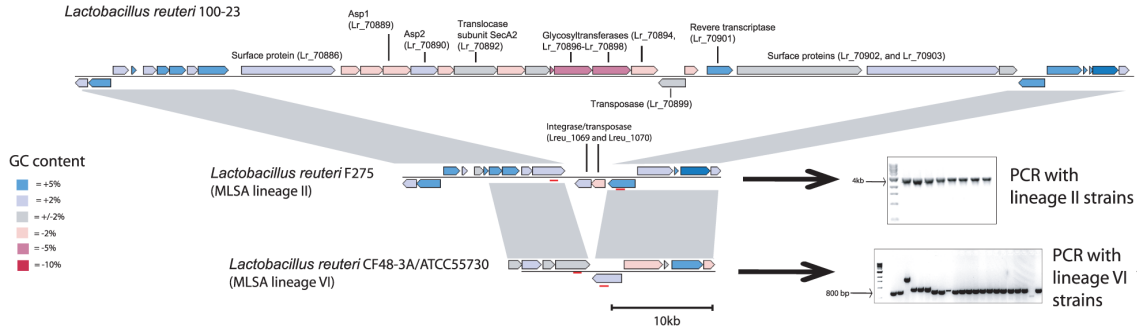

Figure 2.6 Gene map of the accessory SecA2 cluster in Lactobacillus reuteri 100-23 and the same genomic region in strains F275 (MLSA lineage II) and CF48-3A and ATCC55730 (MLSA lineage VI). Genes are colored according to differences in GC content when compared to the genome background (39%). The PCR products shown were generated with primers that targeted conserved genes that flank the location of the SecA2 cluster (primer sites are shown by red bars).

The xylose operon.

A xylose operon is highly conserved in rodent (especially in lineage III) and porcine strains (MLSA lineage IV), while it was absent in all human and poultry strains (lineage II and VI). Xylose could be an important substrate for gut bacteria as it is a plant-derived sugar commonly found in straw and bran, and the xylA promoter of strain 100-23 was previously identified by *in vivo* expression technology to specifically induced in the gut of mice (Walter et al., 2003).

A two-component regulatory system.

TCS2 was detected by CGH in more than half of the rodent strains, and the system contains a putative histidine kinase (Lr_70529), a response regulator of the LytR/AlgR family (Lr_70530), a bacteroicin-like peptide (Lr_70531), an ABC-type bacteriocin transporter (Lr_70532), and an ABC-type bacteriocin/lantibiotic exporter, containing an N-terminal double-glycine peptidase domain (Lr_70533). Lr_70532 showed high similarity (55%) to AbpT of *Lactobacillus salivarius* UCC118, which was

shown to be important for bacteriocin activity (Flynn et al., 2002). The Lr_70531 peptide shares no common sequence homology to other proteins in the NCBI database, but contains a double-glycine motif. Cleavage at this site (at amino acid position 19) would produce a 35-aa peptide that shows characteristics described for extracellular bacterial signaling peptides (Kleerebezem et al., 1997). Since strain 100-23 does not produce a bacteriocin, it is possible that this regulatory system is involved in quorum sensing (QS). The specificity of TCS2 to rodent *L. reuteri* strains suggests that it might affect the transcriptome facilitating host adaptation. In this respect, it is of note that a single two-component sensor kinase can alter the host range of *Vibrio fischeri* (Mandel et al., 2009).

2.4.5 Genetic features specific to rodent strains that show high inter-strain variability.

Several genes were identified by CGH to be rodent specific but were detected in only a small number of strains. These included a second two-component system (TCS1) that was comprised of a histidine kinase (Lr_70430), a LytR/AlgR family response regulator (Lr_70431), and a bacteriocin processing peptidase (Lr_70432). This system meets the criteria established by Sturme and coworkers for a peptide-based QS two-component regulatory system (Sturme et al., 2007). In addition, three genes (Lr_70458, Lr_70459, Lr_70460) that comprise a putative ABC-type Multidrug Efflux System were detected by CGH in three rodent strains. Also, several genes present in the two SPS clusters (SPS1 and SPS2), encoding predicted glycosyltransferses, epimerases, and capsular polysaccharide biosynthesis proteins (Figure B.3), were identified by MARKFIND to be host specific. These clusters showed a very high variability among rodent strains (Figure 2.3). Several genes were identified by CGH to

be rodent specific but were detected in only a small number of strains. These included a second two-component system (TCS1) that was comprised of a histidine kinase (Lr_70430), a LytR/AlgR family response regulator (Lr_70431), and a bacteriocin processing peptidase (Lr_70432). This system meets the criteria established by Sturme and coworkers for a peptide-based QS two-component regulatory system (Sturme et al., 2007). In addition, three genes (Lr_70458, Lr_70459, Lr_70460) that comprise a putative ABC-type Multidrug Efflux System were detected by CGH in three rodent strains. Also, several genes present in the two SPS clusters (SPS1 and SPS2), encoding predicted glycosyltransferses, epimerases, and capsular polysaccharide biosynthesis proteins (Figure S3), were identified by MARKFIND to be host specific. These clusters showed a very high variability among rodent strains (Figure 2.3).

2.4.6 Host-specific gene content in human strains.

Around half of the genes specific to the human MLSA lineage II were related to mobile elements (transposases/integrases, phage proteins, restriction endonucleases) and hypothetical proteins with unknown functions (Table 2.2). The pdu-cbi-cob-hem cluster was conserved within human strains, and the cluster was absent in rodent strains of the CGH sub-groups i and ii, while it was present in 4 out of the 5 strains in the rodent sub-lineage iii (Figure 2.2). This cluster codes for cobalamin (vitamin B12) biosynthesis, glycerol utilization, propanediol fermentation, and production of the antimicrobial compound reuterin (Morita et al., 2008, Sriramulu et al., 2008, Talarico et al., 1990, 1988).

2.4.7 Host-Specific Gene Content Reflects Niche Characteristics in Different Hosts

The functions of the genetic features associated with *L. reuteri* ecotypes are reflective of their lifestyle in respective hosts. In rodents, *L. reuteri* adheres directly to the stratified squamous epithelium present in the murine forestomach and forms thick cell layers that show characteristics of biofilms (Lin and Savage, 1984, Wesney and Tannock, 1979, Tannock et al., 2005, Walter et al., 2007). Accordingly, several of the rodent-specific surface proteins are predicted to function as adhesins or mediators of biofilm formation, and the SecA2 system is likely involved in the secretion of some of these proteins (e.g. Lr_70902). Other factors, such as the TCS2, fructosyltransferase (Ftf), IgA specific metallopeptidase, and the urease cluster are likely to play roles in biofilm formation, cell aggregation, and the mitigation of low pH and exposure to IgA, respectively. It is striking that several of the genes identified as rodent-specific by CGH were also detectable in at least some strains that originate from pigs and poultry (Figure 2.2 and Figure 2.3), reflecting the similar lifestyle of rodent, porcine, and poultry lactobacilli which all form biofilm-like associations with epithelial surfaces in the proximal GIT (Tannock, 1992, Fuller et al., 1978, Fuller and Turvey, 1971).

The genome content of strains within the human MLSA lineage II is strikingly different when compared to other *L. reuteri* lineages. The absence of many genetic features involved in biofilm formation and adhesion reflects the lifestyle of *L. reuteri* in the human gut. Squamous stratified epithelia are absent, and epithelial cell layers rich in lactobacilli equivalent to those found in animals have not been described in the human GIT (Walter, 2008). The genome content of strain F275 suggests a planktonic lifestyle in more distal regions of the human gut and limited, if any, interactions with the gut epithelium. This lifestyle would require fast multiplication rates, which could

explain the absence of the large surface proteins in lineage II strains, which are likely to be a significant energetic burden. In addition, easily accessible nutrients are in low supply in the human colon having been absorbed in the small intestine, and the ability of *L. reuteri* to use 1,2-propanediol as an energy source through the pdu-cbi-cob-hem cluster might therefore constitute an important colonization factor in the human gut. The production of reuterin, which is also conferred by this cluster, might contribute to the fitness of *L. reuteri* in the human gut through inhibition of competitors in the same niche (as reviewed in Stevens and Hume 1998). Enzymes involved in 1,2-propanediol utilization and reuterin formation require Vitamin B12 as a co-factor (Sriramulu et al., 2008, Talarico et al., 1990). The synthesis of Vitamin B12 is also encoded by the pdu-cbi-cob-hem cluster, and it appears to be an important colonization factor for colonic bacteria, as demonstrated for *Bacteroides thetaiotaomicron* (Goodman et al., 2009).

2.4.8 Rodent-Specific Genes Contribute to Fitness in the Mouse GIT

Although it is striking how gene content of *L. reuteri* lineages reflects niche characteristics in particular hosts, differences in gene frequencies within populations can arise not only through natural selection but also random genetic drift. In order to test whether the rodent specific genes were of ecological significance in the GIT of mice, we investigated the fitness of isogenic mutants of strain 100-23C in the gut of LF mice in competition with the parental strain. Eight genes representing major groups of genetic functions among the lineage-specific genes were selected for these experiments: Lr_70902 (serine-rich surface protein), Lr_70770 (putative adhesin), Lr_70892 (SecA2 translocase), Lr_70890 (Asp2, involved in SecA2 transport system),

Lr_70894 (SecY2, involved in SecA2 transport system), Lr_70430 (two-component system histidine kinase), Lr_70458 (ABC-type multidrug transport system), Lr_70532 (ABC-type transporter of TCS2). This selection included sets of genes with high (Lr_70902, Lr_70770, Lr_70892, Lr_70890, Lr_70894, Lr_70532) and low conservation (Lr_70430, Lr_70458) among rodent strains. Further, it included genes with a variety of functions, such as adherence, secretion of surface proteins, and environmental sensing. As shown in Figure 2.7, when the parental strains and their mutant derivatives were introduced into LF mice, seven out of the eight mutants had impaired ecological fitness. The most significant defect in competitive fitness was caused through the inactivation of Lr_70890, Lr_70894, and Lr_70902, which are all associated with the secA2 operon. The only gene that did not contribute to ecological performance was Lr_70770, which encoded a putative adhesin. Given the large number of putative adhesins in the genome of *L. reuteri* 100-23 (Table 2.3), it is possible that redundancy exists in mechanisms that confer adherence.

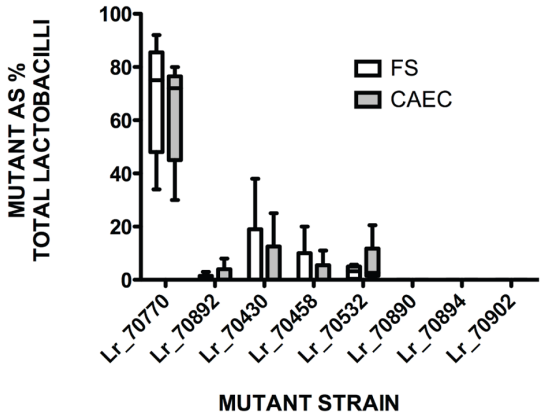

Figure 2.7 Competition between wild-type and mutant *L. reuteri* 100-23C strains in *Lactobacillus*-free mice. Mixtures of wild-type and mutant (1:1) were used to inoculate mice, and the percentage of mutants in the total *Lactobacillus* population was determined after 7 days. The whiskers show the range of values obtained with different animals (n = 5–6), with the box indicating the 25th and 75th percentile. The bars in the boxes indicate median values. Mutants are listed by gene names; Lr_70770 (surface protein), Lr_70892 (SecA2 translocase), Lr_70430 (two-component system histidine kinase), Lr_70458 (ABC-type multidrug transport system), Lr_70532 (ABC-type transporter of TCS2), Lr_70890 (Asp2, involved in SecA2 transport system), Lr_70894 (SecY2), Lr_70902 (serine-rich surface protein).

2.4.9 Identification of Genetic Mechanisms That Led to Differences in *L. reuteri* Genomes

The genetic architecture reflected in the genomes of the rodent and human-adapted *L. reuteri* strains 100-23 and F275 provides insight into the evolutionary processes that underlie host specialization. First, it is clear that that LGT played an important role in the evolution of *L. reuteri*, as many of the host-specific functions were found to be encoded on putative genomic islands or on regions with lost synteny between the two related strains (Figure 2.1). In addition, the pdu-cbi-cob-hem cluster, which is absent in most rodent strains, has previously been identified to be a horizontal acquisition of *L. reuteri* (Morita et al., 2008, Santos et al., 2008). Therefore, the acquisition of novel genetic material could have led to phenotypic innovations in *L. reuteri* and might have allowed lineages to become associated with vertebrates, radiate

among vertebrate hosts, or to switch hosts during evolution.

However, closer scrutiny of the gene organizations at the loci of genomic difference between *L. reuteri* strains 100-23 and F275 suggested an additional mechanism of genome evolution. As shown in Figure 2.6 and Figures 7.3, 7.4, and 7.5, the pdu-cbi-cob-hem, SecA2, urease, and SPS clusters as well as the xylose operon and most of the surface proteins (Lr_70770, Lr_70131–Lr_70137 cluster, Lr_69916, Lr_70580/Lr_70581 cluster, and Lr_71380) are all replaced or interrupted by mobile genetic elements (e.g. putative IS elements and phage related genes) in the genomes of strains 100-23 and F275, respectively. These findings indicate that most of the lineage-specific genes in rodent and human lineage II strains were ancestral and appeared to be jettisoned after divergence of the two lineages. This means that genome evolution of *L. reuteri* strains is, in many cases, a process associated with gene deletions, possibly caused by mobile genetic elements that mediated rearrangements through recombination. Functional gene loss is a common mechanism that underlies host specialization in both pathogenic and symbiotic bacteria from various phylogenetic groups (Moran, 2007, Moran and Plague, 2004, Moya et al., 2008, Kikuchi et al., 2009). Our findings indicate that it also plays an important role for host specialization in *L. reuteri*, especially in the human lineage II.

2.4.10 Important Genomic Events in the Evolutionary History of *L. reuteri*

Given the long time periods involved and the lack of intermediate steps, it is currently difficult to reconstruct the evolutionary processes that have shaped *L. reuteri* subpopulations. However, the genomic comparisons of strains spanning several MLSA lineages allowed us to pinpoint some specific key events in the evolution of the species.

The pdu-cbi-cob-hem cluster appears to be an ancient acquisition of *L. reuteri* as it is distributed through the entire phylogenetic spectrum of the species (Figure 2.4). This is in accordance with conclusions based on codon adaptation index and GC content (Santos et al., 2008). The cluster is absent in most rodent strains, and the analysis of the loci in strain 100-23 indicated that the cluster was deleted through the action of mobile elements (Figure B.4A). It is one of only very few examples of gene loss exclusive to this lineage, making it interesting to speculate as to why its function may be obsolete for the success of *L. reuteri* in the rodent forestomach.

The SecA2 cluster, which is highly conserved in rodent and porcine strains (Figure 2.4), appears to be a later acquisition of *L. reuteri*, as all but one strain from the lineage VI lack this cluster. As shown in Figure 2.6, there is no evidence for deletion of the cluster in lineage VI strains, while strains of MLSA lineage II showed evidence for deletion through mobile genetic elements. This indicates that the cluster was acquired after diversification of more recent lineages from lineage VI. The acquisition of the SecA2 cluster might have been a pivotal innovation of *L. reuteri* strains to colonize the gut of mammals. The biological significance of the SecA2 cluster for life in the rodent gut was clearly demonstrated in our competition experiments in LF mice, in which inactivation of four different genes in strain 100-23C associated with this cluster (Lr_70890, Lr_70892, Lr_70894, and the surface protein Lr_70902) had the most detrimental effects when compared to the other mutants tested (Figure 2.7).

The comparison of the genomes of *L. reuteri* 100-23 and F275 revealed evidence for only one event of LGT since the split of the two lineages. The surface protein Lr_70697 is arranged in an island with two transposases and two phage integrases next to a transfer RNA gene (tRNA-Val) in the genome of 100-23. This locus is intact in the genomes of F275, CF48-3a, and ATCC55730. Therefore, this gene cluster was likely acquired by a recent ancestor of 100-23 and inserted into a tRNA-Val gene,

as described for islands in meseorhizobia and several pathogenic bacteria (Sullivan and Ronson, 1998). As with mesorhizobia, insertion of the cluster in *L. reuteri* left the entire tRNA gene (a Thr-tRNA) intact upon integration, whereas a small part (22 nucleotides in *L. reuteri*) became duplicated as a direct repeat (see Figure B.5B). Both CGH (Figure 2.3) and PCR (Figure 2.4B) analyses showed that Lr_70697 was to a large degree specific to strain 100-23, supporting the hypothesis that this cluster was a recent genomic acquisition.

2.4.11 A Rodent-Specific Accessory Genome of *L. reuteri*

A recent study on the genomes of human *L. reuteri* strains revealed a closed pan-genome, with individual strains contributing to a very small number of new genes (Nelson et al., 2010). Our CGH analysis supported these observations, showing similar genome content and little genetic diversity among strains belonging to the human MLSA II lineage (Figure 2.2). However, strains from other hosts, and especially rodents, possessed a more variable gene content, and the majority of the rodent-specific genes detected by CGH were not conserved among rodent strains (Figure 2.2). Comparisons of the genomes of the three rodent *L. reuteri* strains 100-23, MLC3, and lpuph1 confirmed that rodent strains possess a larger pan-genome with a gene repertoire that extends beyond that of individual strains. Open pan-genomes have been described for many bacterial species, and they consist of a 'core genome' (genes present in all strains) and an 'accessory' genome (genes variable among strains) (Touchon et al., 2009, Medini et al., 2005, Tettelin et al., 2005).

As shown in Figure 2.8, the three rodent strains shared around 1463 of the predicted protein coding genes. Of this core genome, only 25 genes were unique to rodent strains (Figure 2.8), confirming the CGH analysis in that only a small

number of rodent specific genes are conserved among strains. Each strain possessed a significant proportion of genes that were absent in the other strains (528 proteins in 100-23; 235 in MLC3; and 309 in lpuph1), confirming the more variable gene pool among rodent strains. Of note, a large proportion of these genes were not found in the genomes of *L. reuteri* strains from non-rodent hosts (Figure 2.8). This rodent-specific accessory genome was comprised, apart from a large portion of mobile genetic elements, of the same functional groups as the genes identified by the CGH analysis to be rodent-specific (Figure 2.8 and Table A.8). Thus, many of the rodent-specific surface proteins, glycosyltransferases involved in SPS synthesis, transport proteins, and regulatory proteins that are present in the genome of strain 100-23 are substituted by genes that are predicted to perform similar functions in strains MLC3 and lpuph1. The genomic comparisons revealed only one group of rodent-specific genes that were absent in the genome of 100-23 and were therefore not detected by CGH. These genes were all CRISPR-related and are likely to be involved in phage resistance.

It is important to point out that the seven rodent-specific genes that contributed to ecological fitness in colonization experiments in LF mice (Figure 2.7) were not conserved among rodent strains. A key conclusion of this study is therefore that adaptive traits that allow life in the murine gut are encoded by a rodent-specific accessory genome and that different combinations of these genes promote successful colonization. This of course begs the question of why plasticity is favored in the rodent *L. reuteri* population but not the human lineage II. It has been suggested that bacterial accessory genomes encode special ecological adaptations in genes that remain unbounded and can be more rapidly incorporated where and when they become advantageous (Reno et al., 2009, Prosser et al., 2007). Thus, the larger gene pool within in the rodent *L. reuteri* population might be sampled by individual cells through LGT to form the basis for adaption to environmental fluctuations. The population

genetic structure of *L. reuteri* (Oh et al., 2010) and the colonization phenotypes in LF mice imply that lineages maintained a broader host range and evolved with at least two diverse host genera (*Mus* and *Rattus*), and probably many species (around 40% of the world's mammalian species are rodents). Such an evolutionary strategy would require individual cells to adapt not only to physiological and immunological differences of individual animals but different host genera, and the larger accessory genome of the rodent *L. reuteri* population might reflect a higher diversity among the host population.

2.4.12 Reductive Evolution and a Population Bottleneck in Human *L. reuteri* Strains

The ecological forces that have shaped the autochthonous *L. reuteri* population in the human GIT appear fundamentally different than those in other hosts. Strains within the human-specific MLST lineage II, although obtained from world-wide locations, are highly conserved genetically and are clonally related (Walter et al., 2011), (Oh et al., 2010), suggesting a recent population bottleneck, founder effect, or clonal expansion. The genomic comparison of strain 100-23 and F275 further revealed that human strains underwent a process of reductive genome evolution. These evolutionary patterns resemble to some degree those found for genetically monomorphic pathogens, such as *Yersinia pestis* and *Mycobacterium leprae* (Achtman, 2008, Holt et al., 2008, Chain et al., 2004), which show high clonality and genome evolution characterized by functional gene loss.

We can only speculate on what caused the specific genetic features of the human *L. reuteri* population. It has been suggested that the evolution of monomorphic pathogens was influenced by an expansion of the human population within the last

10,000–20,000 years, which possibly led to a significant increase of the available niche and a restriction to the human host (Achtman, 2008). The population bottleneck might also have been caused through altered transmission dynamics and changes in the human environment, which could have reduced the effective population size (Walter et al., 2011). Low population sizes favor genetic drift and can lead to both decreased genetic variability (Bright and Bulgheresi, 2010) and the loss of genes (even if slightly beneficial) (Moran, 2003). Alternatively, *L. reuteri* might have been acquired by humans more recently. Restriction to particular hosts or host changes have both been accompanied with a clonal population structure and functional gene loss, especially those associated with the cell envelope (Eppinger et al., 2006, Holden et al., 2009, Parkhill et al., 2003, Lowder et al., 2009). As described above, the genome of F275 shows clear evidence for pseudogene formation, gene deletions, and genome reduction, and although we do not yet know the causes of these patterns, the dramatic removal of surface proteins *L. reuteri* F275 suggests a process by which to bypass deleterious responses from the human immune system.

2.5 Concluding Remarks

The gut of vertebrates provides a multitude of nutrient rich habitats inhabited by complex microbial communities, whose composition is remarkably host specific and stable (Ley et al., 2008a, Costello et al., 2009). These communities are important for normal development and growth of the host, but must be acquired during each generation as most vertebrates are essentially germ-free at birth. This process is poorly understood but relevant as benefits to the host are increased by the correct selection of true mutualists and their stable maintenance over evolutionary time (Ley et al., 2008a, Mandel et al., 2009). This study clearly established host specificity within

the species *L. reuteri* through a combination of animal experiments and evolutionary genomics, and it revealed a first insight into the genomic changes that underlie host adaptation. Host specificity of *L. reuteri* in the mouse gut appears to be mediated to a large degree by specific adhesins. However, other factors are likely to contribute to host specificity and include adaptations to the environmental conditions (the urease cluster, Ig-A protease, factors for biofilm formation) and their regulation (possibly through TCS involved in quorum sensing).

In the last decade, our understanding of genome evolution in host-associated bacteria has advanced dramatically due to the availability of hundreds of sequenced genomes (Medina and Sachs, 2010). Common trends have been identified and range from those observed in obligate bacterial symbionts, who show extensive reductive genome evolution, to those of facultative symbionts with free-living stages, who have expanded genomes and high levels of LGT (Medina and Sachs, 2010, Ruby et al., 2005, Normand et al., 2007). Genome evolution of *L. reuteri* shares some patterns that have been observed in other host associated bacteria, and the findings suggest an evolutionary intermediate transitioning from a facultative to an obligate, mutualisitic lifestyle, which concurs with the observed degree of host specialization. Accordingly, the high amount of mobile elements (e.g. IS elements) in *L. reuteri* genomes is a characteristic that is often associated with recent obligate host associations in bacteria (Moran and Plague, 2004). Although mobile elements are common in all *L. reuteri* genomes, there are distinct trends of genome evolution in the rodent and human lineages, with the former possessing a large and adaptable pan-genome while the latter being subjected to a process of reductive evolution. These distinctions are likely related to differences in the microbe's host range and the ecology and genetic diversity of the host population.

Taken together, the results of this study revealed host adapted subpopulations

among the species *L. reuteri* whose genome content reflected niche characteristics in their respective hosts. Although physiological and immunological differences of vertebrates were likely to constitute important selective forces that drove this specialization, the distinct patterns of genome evolution in rodent and human lineages suggest that the evolutionary trajectories of a vertebrate gut symbiont are not only determined by microbial competition but also by the ecology and evolutionary history of the host.

2.6 Materials and Methods

2.6.1 Ethics Statement

All animal experiments were approved by the Otago University Animal Ethics Committee (approval number 2/09).

2.6.2 Strains, Media, and Growth Conditions

Lactobacillus reuteri strains used in this study are listed in Table S3 and were grown anaerobically on MRS (Difco) plus 5g/L Fructose and 10g/L Maltose at 37°C or 45°C (where indicated). *Escherichia coli* EC1000, which was used for cloning vectors for gene inactivation in *L. reuteri*, was grown aerobically in LB media at 37°C. Erythromycin (200 μg/mL for *E. coli*, 5 μg/mL for lactobacilli), kanamycin (40 μg/mL for *E. coli*), and chloramphenicol (7.5 μg/mL for lactobacilli) were used for the propagation of recombinant strains. *L. reuteri* 100-23C, which is a plasmid-free derivative of strain 100-23, was used to test the ecological relevance of selected genes (see below).

2.6.3 Determination of Colonization Phenotype in LF mice

LF mice were raised under gnotobiotic conditions, and the absence of lactobacilli was regularly tested by anaerobic culture on Rogosa SL agar for 48 hours. Mice (around 6 weeks of age) were inoculated by gavage on a single occasion with 10^6 *Lactobacillus* cells that had been cultured anaerobically in MRS medium overnight. Cell numbers of lactobacilli in fecal samples, the forestomach, and the cecum were determined by quantitative culture on Rogosa SL agar as described previously (Walter et al., 2007).

2.6.4 Genome Sequencing

Sequencing of *L. reuteri* 100-23 (rodent isolate) and DSM20016T (human isolate F275) genomes were accomplished through the Community Sequencing Program of the Joint Genome Institute (Walnut Creek, CA), using a combination of whole-genome shotgun sequencing of three libraries with 3-Kb, 8-Kb, and 40-Kb DNA inserts. The genomes were further sequenced using a Roche Genome Sequencer (FLX-GS) to reduce the amount of contigs, and gaps were closed manually by sequencing PCR products generated from the ends of contigs. This process resulted in a circular genome for DSM20016^T and two scaffolds for 100-23 (729,351 bp and 1,576,206 bp). PCR reactions to amplify the DNA between these scaffolds failed on several attempts, probably due to the highly repetitive nature of the termini. Genomes were annotated using the JGI annotation pipeline, and the genome sequences have been deposited in GenBank under the accession numbers NC_009513 (strain DSM20016^T) and NZ_AAPZ00000000 (strain 100-23).

The genomes of *L. reuteri* lpuph1 and MLC3 (rodent isolates) were sequenced to draft status at the Core for Applied Genomics and Ecology (CAGE, University

of Nebraska, Lincoln, USA) with a standard shotgun library prep kit of the Roche GS FLX Titanium series. The genome of *L. reuteri* ATCC53608 (pig isolate) was sequenced at the Biotechnology and Biological Research Council's TGAC (The Genome Analysis Centre, Norwich Research Park, UK). Sequencing resulted in 185,905 (lpuph1), 115,542 (MLC3), and 617,241 (ATCC53608) reads that were assembled de novo using the gsAssembler (Newbler) module of the GS-FLX Off-Instrument Sofware Suite. This resulted in draft sequences of 127, 126, and 142 contigs, for lpuph1, MLC3, and ATCC53608 respectively. The draft sequencing resulted in a final coverage of around 30 fold (lpuph1), 20 fold (MLC3), and 100 fold (ATCC53608). The genome characteristics are listed in Table A.6. Genome sequences for mlc3 and lpuph are available at DDBJ/EMBL/GenBank under the accession numbers AEAW00000000 and AEAX00000000, respectively. Genome sequences for ATCC 53608 are available at EMBL under the accession numbers CACS01000001 to CACS01000142.

2.6.5 Genome Analysis and Comparison

L. reuteri F275 was isolated in the 1960s and later deposited in both the Japanese and German culture collections. The genome sequence of the strain deposited in the Japan Collection of Microorganisms (JCM1112^T) was recently published (Morita et al., 2008). In the present study, the strain deposited in the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSM20016^T) was sequenced. As described by Morita and coworkers, strain DSM20016^T has undergone genomic modifications of its genome during in vitro propagation (Morita et al., 2008), and as a consequence, JCM1112^T contains two additional regions (a total of 40 kb) compared to the genome of DSM20016^T. For the analysis and comparisons of gene content in 100-23 and F275, we used the genome annotations of strains 100-23 and DSM20016^T, as they were

both done with the JGI annotation pipeline. The genes that were encoded by the extra sequence identified in JCM112^T were added to the gene set of DSM20016^T and considered in all comparisons.

The Integrated Microbial Genomes (IMG) system of the JGI was used to analyze genome characteristics and compare genomes (Markowitz et al., 2008). Unique and conserved genes between strains 100-23 and F275 were determined by the BLASTP algorithm implemented in the IMG Phylogenetic Profiler with a maximum E value of 1e-5 and a minimum amino acid identity of 70%. Whole genome comparisons were completed using the Artemis Comparison Tool (ACT) (Carver et al., 2006). For this analysis, the two remaining scaffolds of the 100-23 genome were combined and the chromosome replication initiation site was identified. Visual genome comparisons of the genomes of strains 100-23 and JCM1112^T were prepared by using ACT (BLASTN with a score cutoff of 1900). Alien_hunter was used to identify areas affected by LGT (Vernikos and Parkhill, 2006). This program utilizes interpolated variable order motifs to identify regions of the genome with atypical sequence composition and thus integrates codon, and nucleotide compositional changes into its predictions.

BLASTP was used to identify homologous genes (>70% identity, >70% coverage) found in all *L. reuteri* strains and the closely-related *L. vaginalis*. Nucleotide sequences for these 169 orthologous genes were individually aligned in MUSCLE and concatenated and used to calculate the average nucleotide identity (ANI) as described by Konstantinidis and Tiedje (Konstantinidis and Tiejde, 2005). The same BLASTP criteria were applied to determine the core and accessory genomes of *L. reuteri* strains (Figure 2.8).

2.6.6 Comparative Genome Hybridization Using Spotted Microarrays

Spotted microarrays were designed to contain probes representing all detected open reading frames (ORFs) of the rodent strain 100-23 and ORF unique to strain F275 when compared to 100-23. The phylogenetic profiler tool of the IMG platform was used to identify unique genes of F275 (using the sequence of strain DSM20016^T) with a maximum E value of 1e-10 and an amino acid percentage of less than 90%. This analysis revealed 403 unique genes for F275. Probes (60 bp) were designed for all ORFs of sufficient size by using Oligo Array 2.1 (Rouillard et al., 2003). Multiple probes were designed for genes of 100-23 larger than 4.5 kb (3 per gene). In total, the probe set comprised 2192 probes representing 2170 genes of strain 100-23 and 320 probes representing 320 genes of F275. Oligomers were synthesized by Invitrogen (Carlsbad, CA USA) and spotted in duplicate using an Omnigrid arrayer (Gene Machines, San Carlos, California).

L. reuteri strains used in microarray typing are listed in Table A.3, which include 24 isolates from humans (including DSM20016^T), 24 from rodents (including 100-23), 5 pig isolates, and 5 chicken isolates. Chromosomal DNA of bacteria was prepared as described by Oh and coworkers (Oh et al., 2010). DNA of strains 100-23 and DSM20016^T was mixed at a 1:1 ratio, and 2 μg was amplified by random priming using Cy5 dye-labeled nucleotides and the BioPrime DNA labeling kit (Life Technologies, Rockville, Md.) to generate the reference DNA. Test DNA was generated by random priming PCR from all strains with Cy3 dye-labeled nucleotides. Concentrated labeled products from each reference test pair were hybridized in formamide-containing buffer (Array Hyb Low Temp; Sigma, St. Louis, Mo.) for 4 h at 47°C. Slides were washed once each in 1x SSC (0.15 M NaCl plus 0.015 M sodium citrate)-0.03% sodium dodecyl

sulfate, 0.2x SSC, and finally 0.05x SSC. Fluorescence intensities of the array addresses were determined using a GenePix4000 multicolor microarray scanner and GenePix software (Axon Instruments, Union City, CA USA).

Genome content comparisons were performed using MARKFIND, as described by Zhang et al (Zhang et al., 2003). MARKFIND performs a cluster analysis based on genome polymorphisms implementing the unweighted pair-group method with arithmetic means (UPGMA). The program also uses an algorithm for sorting polymorphic characters in the binary strings relative to user-specified groups of taxa. For those genes being represented by three probes (i.e. large surface proteins), the gene was marked as present if at least two probes showed hybridization.

The accuracy of the microarray analysis was tested by comparing the results obtained by hybridizations with whole genome BLASTN comparisons. BLASTN was performed by comparing all gene sequences that are represented on the microarray slide with the genome sequences of *L. reuteri* 100-23 (rodent III cluster), lpuph1 (rodent I), F275 (human cluster II), and CF48-3A1 (human/chicken cluster IV). Genes were considered present if BLASTN resulted in alignments with more than 70% identity and at least 50% coverage to the query sequence. This analysis revealed that the microarray analysis had a very high accuracy for the two reference strains, showing >96.7% and 96.3% accuracy for 100-23 and DSM20016^T, respectively. The accuracy dropped for strain lpuph1 to 92.5%, and it was lowest for strain CF48-3a (81.5%). So as expected, the accuracy of the microarray analysis decreased as gene divergence between the test and reference strains increased (see Table A.7 for ANIs).

2.6.7 Confirmation of Rodent-Specific Gene Polymorphisms by PCR

Eighty-eight *L. reuteri* strains from all known MLSA lineages of the species (Table A.3) were tested by PCR for the presence of representative rodent-specific genes: surface proteins (Lr_70131, Lr_70581, Lr_70697, Lr_69916), secA2 (Lr_70892), pduC (encoding a subunit of diol/glycerol dehydratase, the first enzyme in the propanediol fermentation/reuterin formation pathway), and ureC (encoding the urease alpha subunit). Primers were constructed based on the sequences of all strains that possessed the gene to first amplify an internal region of the gene, and second, to target the flanking genes and amplify the loci in which the gene was located in strain 100-23. The PCRs were carried out in 25 μl volumes containing 20 pmol of each primer and 0.5 units of Taq polymerase (Takara). After an initial denaturation for 3 min at 94°C, the reaction mixtures were cycled 30 times at 94°C for 30 s, 30 s at appropriate annealing temp, and 72°C for 3 min, followed by a 7-min extension at 72°C. Primer sequences and annealing temperatures are listed in Table S9.

2.6.8 Determination of the Ecological Relevance of Genes in *L. reuteri* 100-23C

The contribution of genes for ecological performance was determined as described previously (Walter et al., 2005). Briefly, genes were inactivated in strain 100-23C by insertional mutagenesis by inserting the plasmid pORI28 into the target sites, which renders the mutant erythromycin-resistant. 1:1 mixtures of mutant and wild type were administered by intragastric gavage to anesthetized LF mice. The mice were killed 7 days after inoculation, and lactobacilli were cultured quantitatively from the forestomach and cecum. To determine the proportion of the mutant strain, lactobacilli

were quantified on agar plates with and without erythromycin.

Chapter 3

Identification and functional characterization of *Lactobacillus reuteri* genes involved in host-specific biofilm formation

Steven A. Frese, Daniel A. Peterson, Chaomei Zhang, Andrew K. Benson, Teresa Fangman, Robert Schmaltz, You Zhou, Jens Walter

Preface

This chapter is in preparation for publication. Portions of this research were presented at the 2012 Annual Meeting of the International Scientific Association for Probiotics and Prebiotics (ISAPP) in Cork, Ireland.

Abstract

Although vertebrates harbor bacterial populations in their gastrointestinal tract whose composition is host specific and, to some degree, phylogenetically conserved, little is known on how specific bacterial lineages become acquired. We have recently shown that the vertebrate gut symbiont *Lactobacillus reuteri* evolved to be host specific. The goal of this study was to gain insight into the mechanisms that mediate host specific colonization of *Lactobacillus reuteri* in the murine gut. Experiments with mono-associated mice revealed that the ability of *L. reuteri* strains to form biofilms in the mouse forestomach, but not colonization per se, is strictly dependent on the origin of the strain. To unravel the molecular basis for this host-specific biofilm formation, we applied a combination of transcriptome analysis and comparative genomics to identify genes of *L. reuteri* 100-23 whose transcription was differentially expressed during textitin vitrop biofilm formation and/or that were predicted to be important in biofilm formation. Eleven candidate genes were selected and their ecological significance and transcriptional expression during *in vivo* biofilm formation were determined. The findings revealed that several genes involved in epithelial adherence, cell aggregation, environmental sensing, and cell lysis were important for biofilm formation and over-expressed during gut colonization. These findings provided a first insight into the molecular processes that underlie biofilm formation of *L. reuteri*, which are likely to constitute a key mechanism by which a vertebrate symbiont becomes specifically associated with its host.

3.1 Introduction

In conventional animals, *Lactobacillus* spp. dominate the proximal gastrointestinal (GI) tract of animals such as rodents, pigs, chickens, and horses (Savage and Blumershine, 1974, Kohl et al., 2011, Fuller et al., 1978, Fuller and Turvey, 1971, Yuki et al., 2000). Populations of these organisms are highest in the stomach of the host (Roach et al., 1977, Dubos et al., 1965), where lactobacilli are considered to be 'autochthonous' and they adhere to the keratinized nonsecretory stomach epithelium in a biofilm (Suegara et al., 1975). Studies in chickens, mice, and pigs have found that the lactobacilli isolated from this structure uniquely bind to the host's forestomach, and that strains from other hosts fail to bind (Suegara et al., 1975, Wesney and Tannock, 1979, Fuller and Turvey, 1971).

One of the species that is normally found in the proximal GI tract of these animals is *Lactobacillus reuteri*. *L. reuteri* is found in the gastrointestinal (GI) tract of humans, rodents, pigs, poultry, and other vertebrates (Walter, 2008). This species is composed of ancient host-associated subpopulations which also show host specificity. Only strains from rodent associated subpopulations were able to successfully colonize gnotobiotic mice (Oh et al., 2010, Frese et al., 2011) and this reflects the host-specific genetic content of these strains (Frese et al., 2011). These genes included surface proteins involved in adherence and mechanisms for the export of these surface proteins. Previous studies identified adherence to the epithelium as a relevant trait with regards to ecological fitness and also identified large surface proteins and biofilm formation as part of this fitness (Walter et al., 2005).

Bacteria are found in nature as free-living microbes and as sessile members of adherent communities known as biofilms. These structures impart resistance to antimicrobials (Stewart and William Costerton, 2001), co-localize cooperative

microbes (Brockhurst et al., 2006), and create a persistent reservoir for inoculation of downstream ecosystems (Parsek and Singh, 2003). They also play an important role in establishing and maintaining relationships between microbes and animal hosts (Yip et al., 2006). *Lactobacillus* populations have been reported in many animals and microscopic evidence for these biofilms is extensive (Savage and Blumershine, 1974, Fuller et al., 1978, Fuller and Turvey, 1971, Walter et al., 2008, Savage et al., 1968, Yuki et al., 2000). Yet, there are few studies indicating how these biofilms form, what mediates their growth, or whether they are ecologically relevant. Further, while studies have shown that isolates tend to be host-specific, they were conducted *in vitro* and may not reflect conditions encountered *in vivo*.

Given that the predominant niche of *L. reuteri* lies within the vertebrate host, remaining stably associated with the host is a critical factor for its success. Rapid colonization of naïve hosts also helps to ensure the transmission of *L. reuteri*. Biofilms may provide an effective mechanism for the colonization of new hosts. Therefore, these biofilms may play an important ecological role in the life cycle of *L. reuteri* and may be crucial to its ecological success.

Only rodent-associated strains of *L. reuteri* colonize gnotobiotic mice (Frese et al., 2011), so the ability to form a biofilm *in vivo* was examined using strains from different host-associated lineages. Of the strains tested, only rodent-lineage strains of *L. reuteri* were able to form a biofilm on this surface, despite all strains reaching similar population densities in the forestomach. To identify genes important in the formation of this structure, gene expression microarrays and an *in vitro* model of biofilm growth were used to identify gene expression differences between free-living (batch culture) growth and growth as a biofilm. Focusing on differentially up-regulated genes, rodent-specific genes identified in a previous study (Frese et al., 2011), and other well-characterized biofilm systems (Brunskill and Bayles, 1996, Bayles, 2007), selected

genes were tracked during colonization of germ-free mice by the rodent-associated strain *L. reuteri* 100-23. These genes were also inactivated to assess their impact on biofilm formation in the rodent host.

3.2 Methods

3.2.1 Strains and Media used in the study.

Strains used in this study are described in Table 3.1. *Lactobacillus reuteri* 100-23 is a strain belonging to the rodent subpopulation of the species (Oh et al., 2010). The genome sequence for this organism has been determined (NZ_AAPZ00000000.2) (Frese et al., 2011). This strain has also been used in previous experiments examining biofilm formation *in vivo* in the rodent host (Tannock et al., 2005, Walter et al., 2005, 2008). Lactobacilli were cultured anaerobically on modified MRS (mMRS) medium (MRS supplemented with 10g/L maltose and 5g/L fructose) at 37°C, unless otherwise noted.

3.2.2 Mouse experiments.

Groups of 6-9 week old germ-free (GF) Swiss Webster mice were used in colonization experiments. Animals were housed in individually ventilated sterile biocontainment isolator cages (Allentown Inc, Allentown, NJ) by treatment group. Mice in a treatment group (n = 3) were gavaged with 100 μL (10^7 CFU) of *L. reuteri* or phosphate-buffered saline (PBS, pH 7.0) as a negative control. Prior to gavage, *L. reuteri* cells were washed twice and re-suspended in PBS. After two days, the mice were sacrificed and the forestomachs were removed. Tissue was immediately transferred to fixative for microscopy or snap-frozen in liquid nitrogen for RNA extraction. Forestomach and cecal contents were diluted 1:10 in phosphate-buffered saline (pH

Strain	Relevant Characteristics	Source or Reference
L. reuteri 100-23C	Plasmid-cured derivative of 100-23, isolated from Rat	McConnell et al. 1991
L. reuteri 100-23C KO cgl	Cystathionine γ-lyase inactivated.	This study
L. reuteri 100-23C KO ureC	Urease α-subunit inactivated.	This study
L. reuteri 100-23C KO lrgA	lrgA homolog inactivated.	This study
L. reuteri 100-23C KO lytS	lytS homolog inactivated.	This study
L. reuteri 100-23C KO lysM2	lysM-domain protein inactivated.	This study
L. reuteri 100-23C KO lysM3	lysM-domain protein inactivated.	This study
L. reuteri 100-23C KO 70430	Two-component system inactivated.	Frese et al. 2011
L. reuteri 100-23C KO secA2	secA2-transport system inactivated.	Frese et al. 2011
L. reuteri 100-23C KO 70902	Large surface protein inactivated.	Frese et al. 2011
L. reuteri 100-23C KO 70532	Two-component system inactivated.	Frese et al. 2011
L. reuteri 100-23C KO lsp	Large surface protein inactivated.	Walter et al. 2005
L. reuteri lpuph	Isolated from mouse.	Frese et al. 2011
L. reuteri mlc3	Isolated from mouse.	Frese et al. 2011
L. reuteri DSM20016^T (ATCC 23272)	Isolated from human feces.	Frese et al. 2011
L. reuteri cf4-6g	Isolated from human feces.	Oh et al. 2010
L. reuteri ATCC53608	Isolated from pig.	Mackenzie et al. 2010, Heavens et al. 2011
L. reuteri LPA1	Isolated from pig.	Oh et al. 2010
L. reuteri CSF8	Isolated from chicken.	Hammons et al. 2010
L. reuteri 1366	Isolated from chicken.	Oh et al. 2010

Table 3.1 Strains used in this study

7.0) and serially diluted in 0.9% saline for enumeration by plate count on mMRS. From each cage, contents from one forestomach and one cecum were also plated on BHI as a sterility control.

For time course colonization experiments, eighteen 6-9 week old GF Swiss Webster mice housed in a sterile isolator were gavaged with 100 μL (10^7 CFU) of a 14 hr culture of *L. reuteri* 100-23. Mice were co-housed in three cages within the isolator and one was removed from each cage at each time point to mitigate any potential cage effects. Forestomachs were removed from mice 6, 12, 24, 48, 72, and 96 hours after gavage. Tissue was fixed for confocal microscopy or snap frozen in liquid nitrogen for RNA extraction.

3.2.3 Scanning Electron Microscopy

Tissue from the forestomach was fixed immediately in 0.1M Sorenson's phosphate buffer containing 2.5% EM-grade gluteraldehyde (Electron Microscopy Sciences, Hatfield, PA USA) then critical-point dried and palladium-sputter coated and visualized using a Hitachi S3000N scanning electron microscope.

3.2.4 Confocal Microscopy

Forestomach tissue was fixed immediately in 3% formalin/phosphate buffered saline (PBS, pH 7.0) for 30 min and then transferred to fresh 3% formalin/PBS pH 7.0 and stored at 4°C until the tissue was stained and mounted. For staining, samples were transferred to PBS pH 7.0 to remove methanol, for 60 minutes. PBS was replaced after the first 30 minutes. Tissue was then stained in 5 μg/mL propidium iodide in PBS pH 7.0 for ten minutes. Samples were then washed twice in PBS (pH 7.0). After de-staining, samples are mounted on glass coverslips in Fluorogel (Electron

Microscopy Sciences, Hatfield, PA USA) suspended by a CultureWell^{TM} chambered coverglass (Grace Biolabs Bend, OR USA) and imaged. Images were taken for confocal analysis with an Olympus Ix81 inverted microscope. Confocal images captured using red and green channels, with Z-stacks captured at three random sites by a blinded technician. Three stacks, each for separate fields of view for each sample were grouped for analysis. *L. reuteri* cells were stained by propidium iodide (excitation/emission at 536nm/617nm) while auto-fluorescence by mouse forestomach tissue, a result of the fixative process, was captured as background (350nm/470nm). Confocal images were analyzed by previous methods (Berberov et al., 2004), with some modifications. To measure the amount of adherent *L. reuteri*, Cy3-channel pixel area was measured in the images captured from three separate fields of view for each mouse, (totaling 138.675 mm^2). Measurements are expressed as percentage of wild-type *L. reuteri* 100-23C and treatments are compared to the wild-type by one way ANOVA with Dunnett's Multiple Comparison Test.

For 3D rendering, tissue was fixed in the same manner, but imaged using a Nikon 90i upright scanning confocal microscope at 1 μM slices and rendered using the Nikon Analysis software.

3.2.5 *In vitro* Biofilm Model.

L. reuteri 100-23, was grown in MRS supplemented with 1% maltose, 0.5% fructose and 0.5% sucrose (suMRS), adjusted to pH 5.5, for 14 hours. 2.5 mL of this culture was injected into a disposable glass bottomed flow cell chamber (IBI Scientific, Peosta IA USA) which had been pre-conditioned with 0.5X suMRS (pH 5.0) and pre-warmed to 37°C as described previously (Walter et al., 2008). Flow from a sterile reservoir of 0.5X suMRS, pH 5.0, began 30 minutes after inoculation

and maintained at a rate of 12 mL/hr for 24 hours. After 24 hours, the biofilm was harvested immediately for RNA extraction. In parallel, three biological replicates of the biofilm were prepared. 100mL batch cultures (three biological replicates) of pre-warmed (37°C) suMRS (pH 5.5), was inoculated 1:100 from 14 hr cultures of *L. reuteri* 100-23 grown in the same medium. These batch cultures were incubated for 4 hrs at 37°C and 50 mL was harvested by centrifugation at 4°C and immediately used for RNA extraction. At harvest, the average batch culture pH was 5.0, identical to the biofilm culture medium.

3.2.6 RNA extraction and purification.

Three different sample types were extracted using this method; *In vitro* batch cultures (for expression microarray analysis), forestomach biofilm/tissue samples, and triplicate *in vitro* cultures (8hr, late logarithmic) as reference samples for qRT-PCR. To isolate RNA from forestomach tissue, forestomachs from colonized mice were removed (described above), cut open, the contents were removed, and the tissue was processed for RNA isolation.

Samples were homogenized using zirconia/silica beads for three one minute intervals with a Mini-Bead Beater (BioSpec Products, Inc. Bartlesville, OK USA) in TRI-REAGENT (Molecular Research Center, Inc Cincinnati, OH USA) and kept on ice for one minute between intervals. Total RNA was extracted from the disrupted cells according to the TRI-REAGENT instructions and genomic DNA was removed using the TURBO DNA-free^{TM}kit (Applied Biosystems/Ambion Austin, TX USA) followed by on-column DNase-treatment using the Qiagen RNeasy Kit (Qiagen Valencia, CA USA). DNase-treated RNA was quantified using the Nanodrop-1000 (NanoDrop Technologies, Wilmington, DE USA) and overall RNA integrity was determined in a

RNase-Free 1.2% agarose gel.

DNase-treated RNA was reverse transcribed using the Superscript VILO RT kit according to the manufacturer's instructions (Invitrogen CA USA). Briefly, 20 μL reactions, containing approximately 1 μg of total RNA, of the Superscript VILO RT reaction were incubated for 10 minutes at 25°C, 60 minutes at 42°C and the reaction was terminated by heating to 85°C for 5 minutes.

3.2.7 Microarray transcriptome analysis

For microarray analyses, the quality and concentration of RNA was determined using an Agilent 2100 Bioanalyzer (Agilent, Palo Alto, CA USA) and a NanoDrop ND-1000 Spectrophotomoter (ThermoScientific Wilmington, DE USA). Microarrays synthesized previously (Frese et al., 2011) were used for the experiment. Total RNA was directly labeled by reversed transcription using SuperScript II (Invitrogen, Carlsbad, CA) according to the manufacturer's instruction. The 20 μL reaction mix included 20 μg total RNA, random hexamers (200 ng/μL), 0.01 M dithiothreitol, 0.05 mM dATP, 0.05mM dTTP, 0.05 mM dGTP and 0.02 mM dCTP, SUPERase (2 U/μL), 3.75 nM Cy3-dCTP dye or Cy5-dCTP (GE Healthcare UK limited, Little Chalfont Buckinghamshire, UK), and Superscript II reverse transcriptase (30U/μL). The reaction was incubated at 42 ° C for 2 hr. The reaction was terminated by adding 3 μL of 0.2 μm-filtered 0.5 M EDTA (final concentration 0.05 M) and incubate for 2 min at RT. The RNA was removed by adding 3 μL 0.2 μm-filtered 1 M NaOH (final concentration 0.1 M) and incubating at 65°C for 30 min. The solution was neutralized by adding 3 μL 0.2 μm-filtered 1 M HCl. The labeling concentration was measured using a Nanodrop, and equal amounts of Cy3 and Cy5 labeled cDNAs were mixed together and purified using QIAquick PCR purification kit according

to the manufacturer's instruction (Qiagen Valencia, CA USA). 22 μL of LowTemp Hybridization buffer (ArrayIt Corporation, Sunnyvale, CA , USA) was used for elution. The final hybridization solution was prepared by mixing the 22.0 μL labeling mix, 3.5 μL Salmon sperm DNA (5 mg/mL) and 2.0 μL yeast tRNA (9.2 mg/mL). The hybridization was incubated at 43°C in dark overnight (approximately 16-20 hrs). The hybridized chips were washed using 1X SSC buffer plus 0.03% SDS, followed by 0.2X SSC, then 0.05X SSC for 5 min at room temperature sequentially with gentle agitation. Slides were immediately scanned with an Axon GenePix 4000 scanner (Axon, Union City, CA). Images were subsequently analyzed using Axon GenePix 4.0 software (Axon, Union City, CA). The experiment was performed in triplicate with biologically independent samples. The statistical analysis was carried out using R/Bioconductor and the LIMMA analysis package (Gentleman et al., 2004).

3.2.8 Quantitative Reverse Transcription PCR (qRT-PCR).

qRT-PCR was carried out using the Quanti-Fast SYBR Green PCR kit on an Eppendorf Mastercycler Realplex2 machine (Eppendorf AG, Hamburg, Germany) using primers designed with Primer3 (Rozen and Skaletsky, 1999) and verified for specificity using the NCBI nr database (Table A.11) targeting genes identified by microarray analysis and genomic comparisons from previous work (Frese et al., 2011). Primers were validated using serial tenfold dilutions of pooled cDNA to confirm specificity and efficiency. Tenfold dilutions of pooled cDNA were also used as standard curves for each primer set in experimental reactions as efficiency controls. Standard curves were carried out in triplicate and experimental samples were performed in duplicate. For RT-PCR of target genes, 12.5 μL of 2X SYBR Mastermix, 1 μL of cDNA, and 25 picomol of each primer was used per 25μL reaction. A 5 minute

denaturation step at 95°C was followed by 40 2-step cycles of 10 seconds at 95°C, then 30 s at 60°C. Melting curves were also performed, consisting of a denaturation step of 15 s at 95°C, an increase from 60°C-95°C over a 20-min period, and a final step of 15 s at 95°C. Finally, products from each reaction were visualized in an agarose gel to confirm specificity of the products. Gene transcripts were quantified relative to the glucose-3-phosphate dehydrogenase housekeeping gene, whose expression did not differ between biofilm and batch culture growth (Table S1). qRT-PCR results were analyzed by the method of Pfaffl (Pfaffl, 1999) and compared using one-way ANOVA.

3.2.9 Gene Inactivation

Insertional inactivation of target genes was carried out as described previously (Schwab et al., 2007). Briefly, genes were inactivated by directed insertional mutagenesis using a temperature sensitive helper plasmid, pVE6007. pORI28 bearing a 300-500bp target sequence and an in-frame stop codon is selected for integration when grown at the non-permissive temperature in the presence of erythromycin. Insertion confers erythromycin resistance and all mutants were confirmed by PCR and no growth deficits were found *in vitro* (data not shown).

3.2.10 Statistics

Biofilm formation values are shown as a percentage of the wild-type mean ± the standard error of the mean. Comparisons between strains (WT or mutant) were performed by ANOVA with Dunnett's multiple comparison test. Significance, $p < 0.05$ is denoted by a single asterisk (*), $p < 0.01$ as two asterisks (**), and $p < 0.001$ by three asterisks (***). Numbers of *L. reuteri* per gram organ contents are shown as individual points for each mouse and as the mean ± the standard error of the mean.

3.3 Results

3.3.1 Biofilm Formation by *L. reuteri* on the Mouse Forestomach.

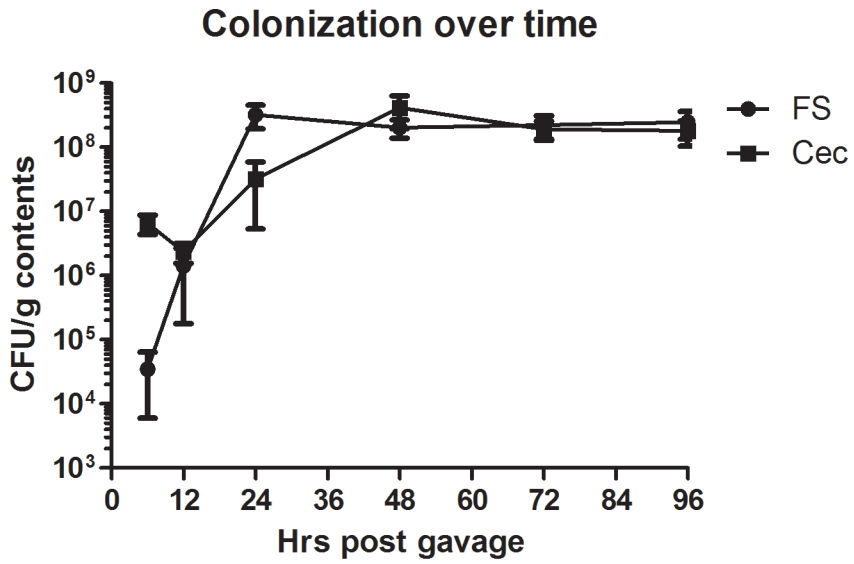

Figure 3.1 *Lactobacillus reuteri* populations after gavage with a single dose of 10^7 cells. *L. reuteri* 100-23 increases post-gavage to a final stable population $>10^8$ CFU/g contents in the forestomach and cecal contents.

Six hours after gavage, forestomach lumen populations are low ($4x10^5$ CFU/g) but increase and after 48 hours, they reach a steady state of 10^8 CFU/g contents (Figure 3.1). Cecal populations follow the same trend and reach 10^8 CFU/g after 48 hours as well. Six hours after gavage, few cells were found attached to the forestomach epithelium (Fig 3.2E). After 24 hours, individual cells were adherent, and small patchy distributions of cells could be seen but did not yet resemble a mature biofilm (Fig 3.2E). By 48 hours after gavage, the mature biofilm was present, which did not change through 96 hours after gavage. In contrast, in uncolonized animals (Figure 3.2B,D) the epithelium remained smooth had no adherent bacteria. Both confocal scanning

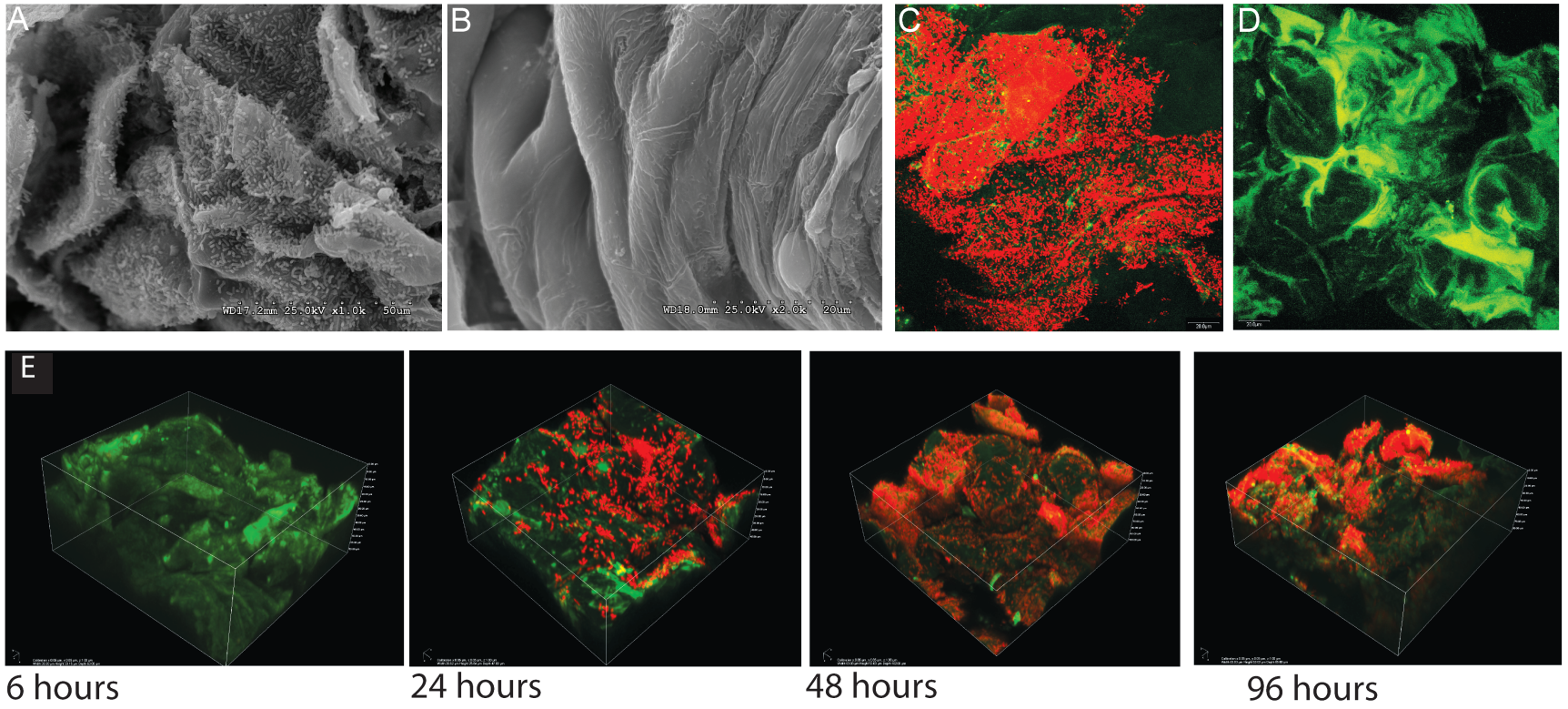

Figure 3.2 *Lactobacillus reuteri* 100-23 forms a biofilm (A,C) on the forestomach of ex-germ mice two days after a single gavage of 10^7 CFU *L. reuteri* 100-23. For comparison, uncolonized forestomachs of germ-free mice are shown (B,D). Both scanning electron microscopy (A, B) and confocal microscopy (C, D), after staining with propidium iodide, show thick layers of cells covering the keratinized squamous forestomach epithelium. (E) Confocal images of biofilm formation post-inoculation. Biofilms reach their approximate maximum growth after 48 hours, which continues through 96 hours after inoculation.

laser microscopy and scanning electron microscopy (Figure 3.2A-D, S1A-C) confirmed these observations.

In the biofilm, many cells had some direct contact with the forestomach epithelium (Figure 3.2A,C,E, 3.S1A-C). In some places, recent losses of outer epithelial layers removed adherent cells, revealing vacant areas. Presumably, these are colonized with cells from the lumen and undergo biofilm development, as we observed from 6-48 hours.

3.3.2 Biofilm formation is Host Specific.

In germ-free mice colonized by *L. reuteri*, forestomach and cecal *L. reuteri* populations were enumerated by plate count, which showed that forestomach lumen populations of each strain all reached high levels (10^8 CFU/g) among strains from human, pig, and rodent-associated lineages (Figure 3.3A). Strains from chicken associated lineages did not reach such high levels ($P < 0.05$), but were comparable to the other strains ($> 10^7$ CFU/g).

Tissue samples obtained from colonized mice were fixed in 3% formalin in PBS and examined using confocal scanning laser microscopy (CSLM) after staining with propidium iodide. Confocal microsopy indicated that the rodent strains (100-23, lpuph, and mlc3) adhered to the mouse forestomach epithelium and formed biofilms. In comparison, strains which originated from non-rodent hosts (human strains DSM20016, ATCC PTA 6475, and cf4-6g; pig strains ATCC53608 and LPA1; chicken strains CSF8 and 1366;) did not adhere to the forestomach (Figure 3.3B). Despite comparable populations in the lumen, these cells were completely absent from the epithelium (Figure 3.3).

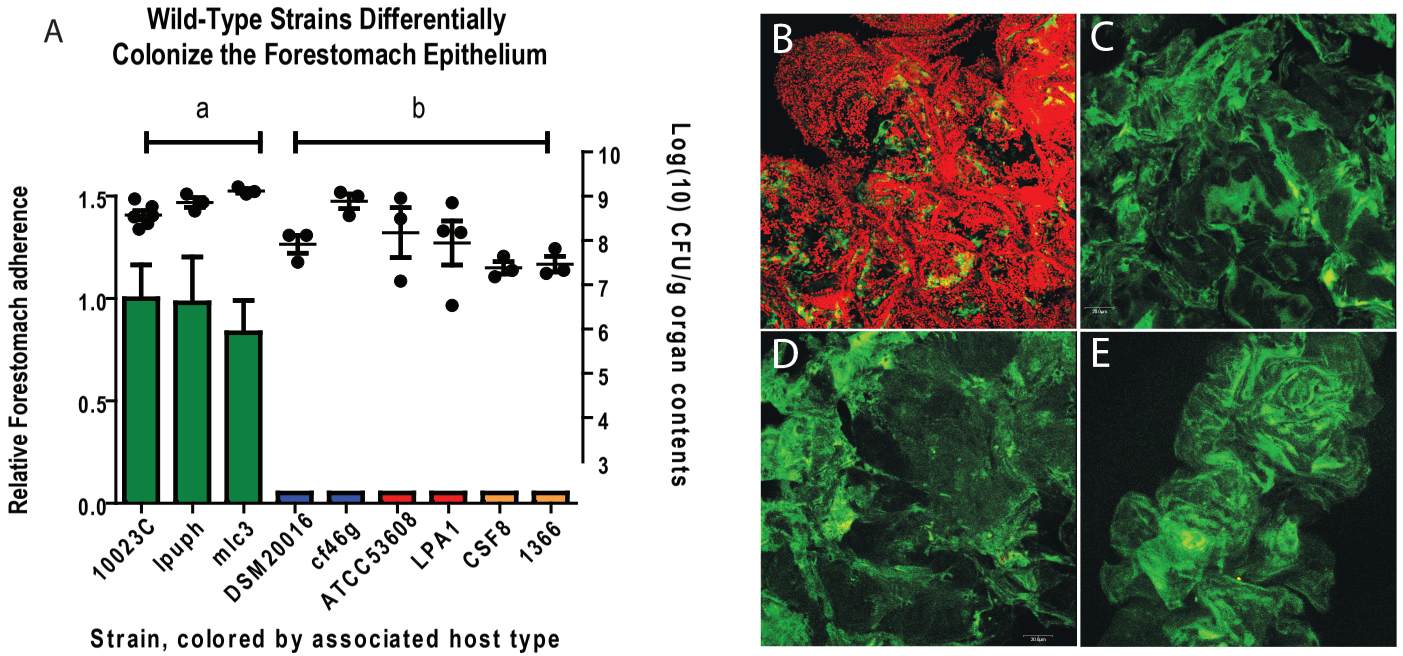

Figure 3.3 (A)Two days after gavage with a single dose of 10^7 CFU of *Lactobacillus reuteri*, total *L. reuteri* populations (points) reach $>10^7$ CFU/g forestomach contents, regardless of the strain's host origin. However, when the forestomach epithelium is examined for adherent lactobacilli, only strains 100-23 (B), lpuph, and mlc3, all of rodent origin, adhere and form biofilms. Strains from humans (DSM20016 (C), cf4-6g), pigs (ATCC53608 (D), LPA1), and chickens (CS-F8 (E), 1366) fail to adhere to the forestomach epithelium (bars). (ANOVA with Dunnett's multiple comparison test, $P < 0.001$).

3.3.3 Transcriptome Analysis in Batch and Biofilm Culture.

To determine which genes were affected during biofilm formation, an *in vitro* model of biofilm growth was used to measure gene expression differences. Microarray analysis indicated that 1130 genes were differentially ($p < 0.05$) expressed between the two conditions tested, batch culture and biofilm growth. Many of these genes showed a small (less than one-fold) change. However 91 genes were up-regulated in the biofilm condition (greater than 2-fold) and 37 genes were down regulated in these conditions (greater than 2-fold; Table S3.1). Of the genes most up-regulated in the biofilm, several classes of genes were noted. A cystathionine-γ-lyase cluster was significantly up-regulated, as well as four LysM-domain proteins, and homologs to the LrgAB regulatory system. In the batch culture, surface proteins were among the most highly expressed genes (Lr_70131, Lr_70134) as well as genes involved in metabolism, such as a glucose-6-phosphate isomerase (Lr_71411), an amino acid transport protein (Lr_70032), and genes involved in amino acid synthesis (Lr_69051).

3.3.4 The Molecular Basis of Biofilm Formation.

To identify genes responsible for biofilm formation *in vivo*, we used the information describing differentially regulated genes in biofilm growth (Table A.10). In addition, rodent-lineage specific or ecologically relevant genes from other studies (Table 3.1, 3.2), and homologs to other models of biofilm growth were studied to determine the molecular basis of adherence and biofilm formation *in vivo* (Table 3.2). To test for their ecological significance, eleven genes were inactivated (Table 3.2). The resulting mutants did not demonstrate a growth deficit *in vitro* (data not shown) and they all reached comparable final population densities *in vivo*, as determined by plate count on mMRS containing erythromycin (5μg/mL) (Figure 3.4). Using the

Gene	Function	Described In	Reason for Study
Lr_71416	LysM/YG Domain protein	*L. johnsonii, L. gasseri, L. acidophilus*	Upregulated in Biofilm
Lr_70152	LysM/YG Domain protein	*L. johnsonii, L. gasseri, L. acidophilus*	Upregulated in Biofilm
Lr_69271	LrgA regulator	*Staphylococcus aureus*	Upregulated in Biofilm
Lr_69360	Cystathionine γ Lyase	*L. reuteri* BR11	Upregulated in Biofilm
Lr_70892	secA2 protein translocase	*Streptococcus gordonii*	Critical for ecological success
Lr_70902	Large surface protein	*L. reuteri* 100-23	Critical for ecological success
Lr_70532	Putative Bacteriocin transporter	*L. reuteri* 100-23	Critical for ecological success
Lr_70430	Two-component Histidine Kinase	*L. reuteri* 100-23	Critical for ecological success
Lr_70114	Urease enzyme, α subunit	*L. reuteri* 100-23, *H. pylori*	Absent in Human *L. reuteri strains*
Lr_69269	LytS regulator	*Staphylococcus aureus*	Biofilm regulator
Lr_70580	lsp large surface protein	*L. reuteri* 100-23	Important for competitive colonization

Table 3.2 Genes predicted to be involved in biofilm formation in *L. reuteri* 100-23, as determined by ecological studies, biofilm transcriptome analysis, or other Gram-positive models of biofilm growth

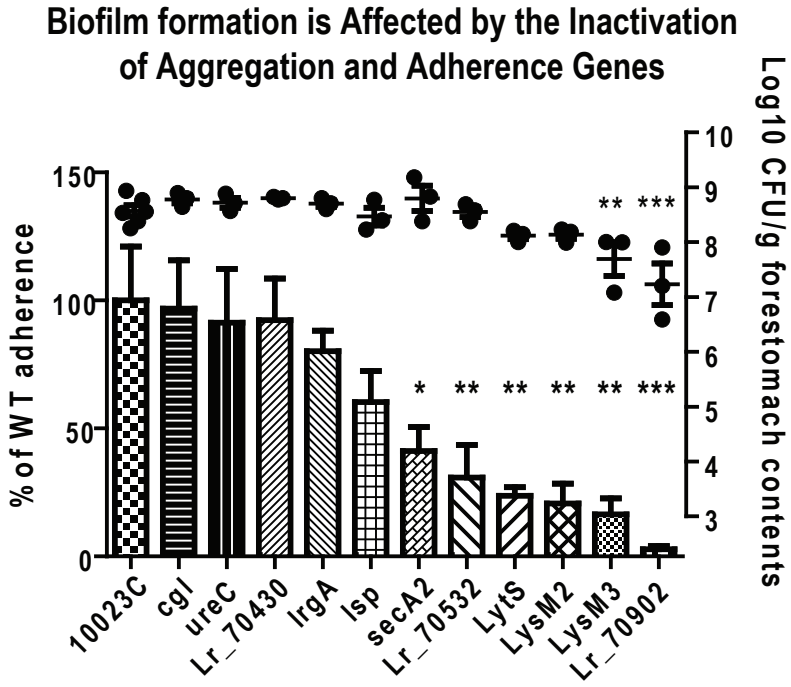

Figure 3.4 Two days after gavage with a single dose of 10^7 CFU of mutant *Lactobacillus reuteri* 100-23C, total populations (points above bars) reach 10^8 CFU/g forestomach contents. However, strains with mutations at the secA2, Lr_70532, LytS, LysM2, LysM3, and Lr_70902 loci show significant reductions in biofilm growth on the forestomach. (ANOVA with Dunnett's multiple comparison test, *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$)

same experimental conditions as those used to test adherence and population densities in wild-type *L. reuteri*, we examined the ability of these mutant strains to colonize ex-germ free Swiss Webster mice. In these animals, the mutant strains reached cell densities which mimicked the wild-type parent strain, *L. reuteri* 100-23C. However, six mutants showed significant reductions in biofilm formation.

First, several genes were critical for biofilm formation. These genes included a secA2 protein secretion system, a large surface-anchored protein (Lr_70902), several proteins which contain a LysM peptidoglycan binding domain and a YG protein motif (Lr_71416 and Lr_70152; LysM 2 and LysM3, respectively), as well as a rodent specific

regulatory system (Lr_70532) the LytS (Lr_69269) regulatory system, (Fig 3.4).

For other loci, such as the urease catalytic subunit (ureC; Lr_70114) and the cystathionine γ-lyase (Lr_69360), inactivation did not result in a reduction in biofilm formation. Two regulatory systems (Lr_70430 and Lr_69269, LrgA) were also not tied to biofilm formation. Other genes, such as the lsp large surface protein (Lr_70580) were reduced in their ability to form biofilms, but this reduction failed to reach significance (Figure 3.4).

3.3.5 Characterization of Genes Critical to Biofilm Formation

LysM-domain/YG-motif proteins. Four lysM-domain containing proteins were found in the genome of *L. reuteri* 100-23 whose inactivation impaired biofilm formation. These proteins possess a lysM-peptidoglycan binding domain. In addition to this domain, the C-terminal end of the protein contains a YG carbohydrate binding motif, which is predicted to bind carbohydrate moieties. Lr_69719, Lr_69721, Lr_71416 share common LysM and YG domains which share high homology (>80% homology) from amino acids 1 to 90 and 190 to the end of the predicted 203 to 356 amino acid sequence. Between Lr_69719 and Lr_69721, the YG domains are nearly identical, at 91% homology, suggesting that their target ligands are very similar (Figure B.7).

Lr_70532 - A putative bacteriocin-transporter. Despite having a predicted bacteroicin transporter and a colocalized bacteriocin-like peptide (Lr_70531), *L. reuteri* 100-23 has not been reported to produce a peptide with antimicrobial activity and there are no identified immunity proteins, which would be associated with such a system. This transporter may be responsible for the export of a quorum-sensing molecule instead. The Lr_70532 transporter and surrounding proteins bear similarity

to genes found in *Lactobacillus salivarius* strains and *Lactobacillus crispatus*, but each gene shares homology with a different organism (Table A.12). This gene is specific to rodent-associated strains of the species.

Lr_70902 and secA2 (Lr_70892) – A surface protein and a transport system. Gene sequence analysis of Lr_70902 indicates that this 2180 aa protein lacks a predicted signal peptide, but contains an LPXTG cell wall anchor and a serine-rich repeating motif of 'SVSMSESLSN.' This motif is repeated identically and in succession 74 times in Lr_70902. This motif is also found repeated in a similar fashion (at 99% similarity) in a homolog in *L. johnsonii* NCC533. A similar repeating pattern of serine residues is found in a serine-rich adhesin in *Streptococcus* bears 50% homology to Lr_70902. This protein, Fap1, is a fimbrial adhesion in *Streptococcus parasanguinis* and is exported to the cell surface by *S. parasanguinis'* homologous secA2/secY2 system. The secA2 cluster is colocalized with Lr_70902 in *L. reuteri* 100-23 and consists of the secA2 translocase (Lr_70892), a secY2 translocase (Lr_70888), accessory secretion proteins (asp1, asp2, asp3; Lr_70889, 70890, Lr_70891, respectively), and glycosyl transferases (Lr_70894, Lr_70896, Lr_70897, Lr_70898), which are typical of the secA2/secY2 protein transport system. Both secA2 and Lr_70902 are also specific to rodent-associated strains of the species.

LytS regulatory system. The lytS homolog Lr_69269 (Figure B.7) is a phylogenetically conserved histidine kinase, and can be found in other *L. reuteri* strains and other species, with highest similarity in the closely related species *L. vaginalis* ATCC 49540 (67% amino acid identity) and *L. mucosae* LM1 (51% amino acid identity) (Figure B.7). LytS has a signal peptide and five predicted transmembrane domains predicted by TMHMM (pfam07694). The 380 C-terminal amino acids of the protein contain a GAF-family domain (pfam 13492), a histidine kinase domain (pfam 06580), and an HATPase_C domain (pfam 02518), all of which are typical of a histidine kinase.

Co-located with LytS is the lytR gene (Lr_69270), which is a response regulator, as indicated by the signal receiver domain (pfam 00072) and a LytTr DNA-binding domain (pfam04397). These proteins constitute a two-component regulatory system. These genes bear sequence homology to the sensor histidine kinase and the response regulator described in *Staphylococcus aureus*, with amino acid identity at 49% and 37%, respectively (Figure B.7).

In *S. aureus* these genes are found directly upstream of the related genes, lrgAB. Homologs to lrgAB are also found in *L. reuteri* 100-23, colocalized with LytSR as Lr_69720 and Lr_69721 (Figure B.7).

3.3.6 Temporal examination of *L. reuteri* 100-23 Gene Expression during formation of an *in vivo* biofilm

The expression of selected genes was tracked during colonization of the rodent host. Gene expression differed between growth conditions (*in vitro* versus *in vivo*), rather than over time during colonization. Instead, their expression appeared to be driven by colonization of the host, and showed marked increases in relative expression, compared to *in vitro* growth (Figure 3.5B, C, D). One exception was LytS (Lr_69269). Expression of the lytS regulator increased, beginning at 48 hours and reaching significantly different levels by 96 hours (Figure 3.5A). This differential expression reflected the expression differences found in the *in vitro* biofilm experiment (Table A.10).

Despite not affecting biofilm formation per se, the expression of the alpha subunit of the urease enzyme (ureC, Lr_70114) was among the most highly unregulated genes studied by qRT-PCR (Figure 3.5B). Also, while only slightly reducing biofilm formation, the lsp protein (Lr_70580) was another highly expressed protein (Figure

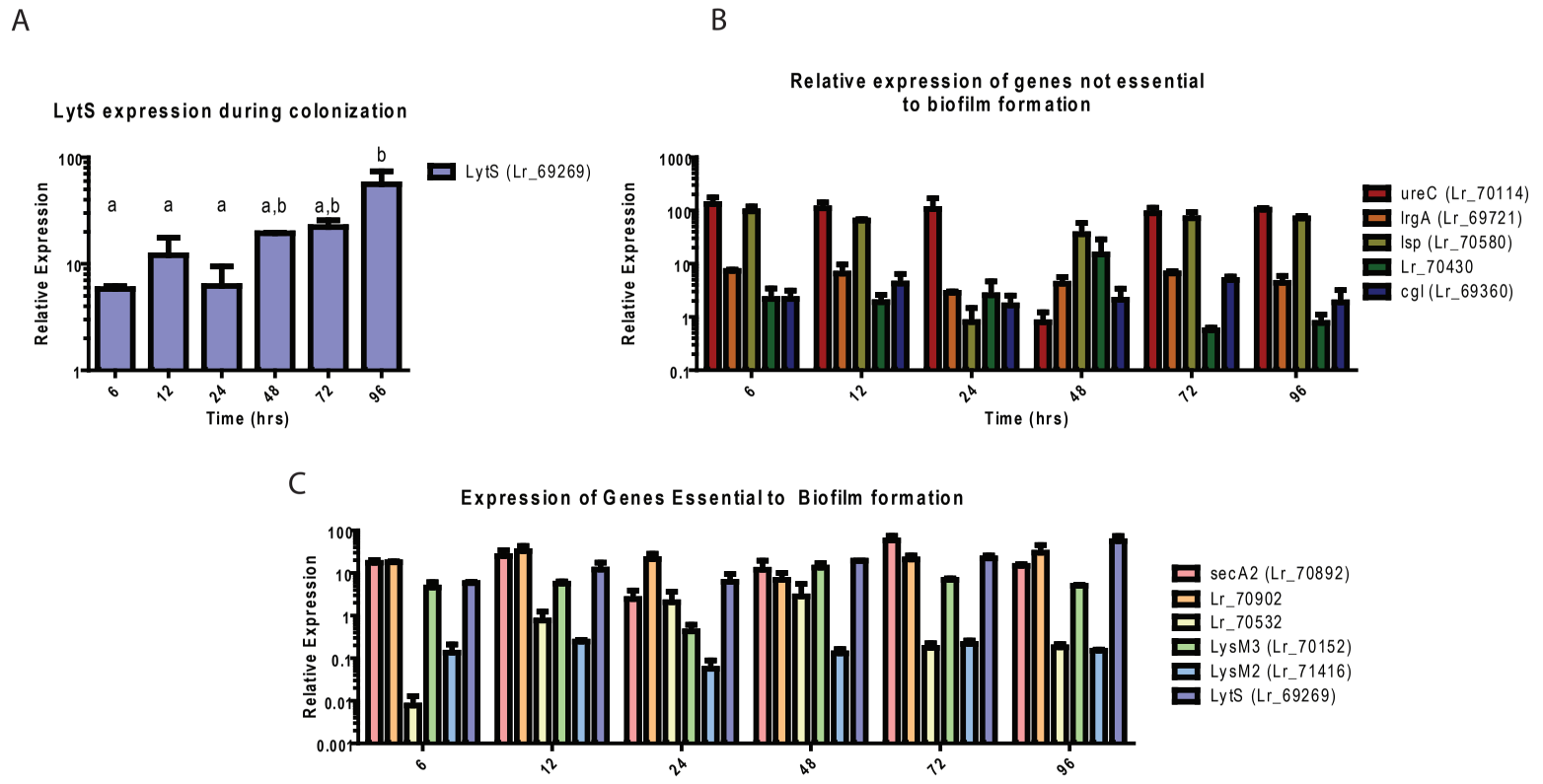

Figure 3.5 *Lactobacillus reuteri* 100-23 colonization of the rodent forestomach over time shows significantly different patterns of gene expression. Relative expression of the LytS regulator increases after 48 hours and reaches a significant difference at 96 hours after gavage (A), while genes that are not responsible for biofilm formation (B) and genes essential to biofilm formation (B) do not show significant differences over time. However, many of these genes showed an increased expression level compared to *in vitro* growth in late logarithmic phase.

3.5B). Other genes which were not responsible for biofilm formation were not highly up-regulated during colonization, such as the rodent-lineage specific histidine kinase, Lr_70430, or the cystathionine γ-lyase (Lr_69360) (Figure 3.5B).

In contrast, genes which were essential to biofilm formation were not upregulated during biofilm formation. For example, the Lr_70532 bacteriocin transporter was not differentially expressed, compared to *in vitro* and did not change over time (Figure 3.5C). The secA2 surface protein transporter and the large surface protein Lr_70902 were, however, expressed roughly 10 fold more *in vivo* than in the late logarithmic phase of batch culture. Interestingly, the LysM2 gene (Lr_71416) was not differentially expressed, compared to *in vitro*, but the LysM3 (Lr_70152) gene was expressed nearly ten-fold more (Figure 3.5C).

3.4 Discussion

3.4.1 The Structure formed by Rodent *L. reuteri* on the Forestomach Epithelium of Mice is a Biofilm.

Microbial life can exist in a free living or adherent state. Adherent cells often form assemblies that are referred to as biofilms, the specific characteristics of which vary between monocultures, mixtures of species, and environmental conditions. The definition of a biofilm as a ‘‘surface accumulation” (Characklis and Marshall, 1990) has since been expanded to account for the staggering diversity of microbial biofilms which may be several microns thick or just a few cells. As such, Stoodley et al (Stoodley et al., 1997) suggested that definitions of systems refer to ‘‘variously shaped aggregates in a slime matrix” or a ‘‘thin base film” or ‘‘a sparse monolayer or up to a few cells thick,” to accommodate the broad diversity. By these definitions, it is clear that the

L. reuteri adherence on the rodent forestomach after two days qualifies as a biofilm (Figure 3.1).

3.4.2 Biofilm Formation is Host-Specific.

Adherence by lactobacilli to the non-secretory stomach epithelium has been reported in pigs, birds, horses, and rodents (Savage and Blumershine, 1974, Kohl et al., 2011, Fuller et al., 1978, Fuller and Turvey, 1971, Yuki et al., 2000). Isolates taken from these organs subsequently demonstrate an inability to colonize the same tissue type in other animals (Wesney and Tannock, 1979, Fuller, 1975), showing indications of host-specificity. However, these studies do not discriminate between species and these differences may be a result of such species differences. The recently identified subpopulations of *L. reuteri* (Oh et al., 2010) however, provided an opportunity to test host-specific sub-populations of a single species for the ability to form biofilms in a single host, under germ-free conditions, and demonstrate that the ability to compete ecologically is related to adherence and biofilm formation, rather than an inability to reach sufficient population densities.

Strains from *L. reuteri* lineages associated with hosts other than rodents were unable to adhere to the rodent forestomach epithelium and form biofilms. In contrast, germ-free mice colonized with rodent-lineage *L. reuteri* (eg. 100-23, lpuph, mlc3) produced robust biofilms on the forestomach epithelium. In all the animals mono-associated with wild-type *L. reuteri*, lactobacilli populations reached those which mimicked populations found in conventional animals. In contrast, our previous experiments found that in the ecologically competitive model found in Lactobacilli-Free (LF) mice, these strains were eliminated from the animals after 14 days (Frese et al., 2011). Taken together, these results indicate that the inability of these strains

to persist in gnotobiotic mice is directly related to their inability to form biofilms in these animals. The non-adherent strains still reached high population levels but only in germ-free mice, suggesting that fitness deficits observed previously were a result of their inability to adhere, and not some other intrinsic deficit.

3.4.3 Identification of Genes that are Differentially Expressed in *in vitro* Biofilms when Compared to Batch Culture

To identify key molecular traits that may be responsible for biofilm formation and determine which genes are involved in maintaining the biofilm structure, microarray gene expression profiles were compared between these two growth phases *in vitro*. Interestingly, many of these genes were not rodent specific. While the surface proteins are up-regulated during batch culture are largely rodent-specific, the regulatory proteins (lrgAB), lysM-domain proteins (eg. Lr_71416 and Lr_69721), and a cystathionine γ-lyase gene are found in many other strains of *L. reuteri* (Frese et al., 2011).

This suggests two important points. First, that the ability to form a biofilm is not actually unique to rodent associated lineages of *L. reuteri*. Given that these structures are also reported by lactobacilli in organisms where *L. reuteri* can be found, this is very likely (Savage and Blumershine, 1974, Fuller et al., 1978, Fuller and Turvey, 1971). Second, the rodent-specific surface proteins mediate adhesion to a rodent-specific receptor. Indeed, rodent-specific adhesion genes have been described elsewhere as being relevant to ecological persistence of *L. reuteri* (Walter et al., 2005, Frese et al., 2011). The phylogenetic distribution of these other genes (LrgAB, lysM-domain proteins, cystathionine γ-lyase) may be a reflection of how common biofilms are in bacteria (Characklis and Marshall, 1990) and part of a general response to life in a biofilm instead of life in a rodent forestomach, specifically. Thus, the limiting

steps involved in adherence contribute to the host specificity of biofilms in rodents, rather than a general ability to form biofilms.

3.4.4 Targeted gene inactivation reveals that genes involved in aggregation and adhesion are key to biofilm formation *in vivo*.

The formation of a biofilm occurs in several steps. First, cells must be able to adhere to the surface on which the biofilm is to be formed. This initial attachment can be transient and cells must quickly establish themselves before they can be removed. The attachment step can be limiting for biofilm growth as all other steps in biofilm formation rely on secure attachment (Characklis and Marshall, 1990). Once established, these cells must remain attached, and aggregation factors are known to promote cell-cell adhesion. As such, the genes which impacted biofilm formation most significantly were genes involved in adherence and aggregation. Large surface proteins (Lr_70902), their export mechanisms (secA2), and aggregation factors (LysM2 and LysM3) were critical for biofilm formation.

Lr_70902, a large surface protein, has a repetitive serine-rich motif and an LPXTG cell-surface anchor. Across the serine-rich repeat region, Lr_70902 is >45% homologous to the serine-rich region in the fimbirial adhesin, Fap1, which is a surface adhesion protein and has been described in *Streptococcus parasanguinis*. Fap1 is a protein which is critical for adhesion *in vivo* and without this protein *S. parasanguinis* is unable to attach and colonize surfaces (Froeliger and Fives-Taylor, 2001).

The Fap1 protein is so critical to adherence that an entire protein transport system, the secA2 system, is maintained to selectively export this protein. In *L. reuteri* 100-23, this cluster is predicted to export Lr_70902, and the inactivation of

the secA2 translocase reduced biofilm formation. While it is yet uncertain whether other proteins are exported via the secA2 translocase, or whether Lr_70902 itself is transported through this cluster, there is some evidence that at least Lr_70902 is translocated through this system. The colocation of Lr_70902 with the secA2 cluster (Frese et al., 2011), the homology to other experimentally described proteins and translocases (Chen et al., 2007) and the matching effect on biofilm formation suggest that the Lr_70902 surface protein is one of the proteins secreted by secA2. Further, the impact on biofilm formation demonstrate that these proteins are key to adherence to the forestomach epithelium.

In addition to adherence, aggregation is a critical feature of biofilm formation *in vivo*. If bacteria could only adhere to a surface, they would only be able to form a single monolayer of cells. However, this is not the case with many bacterial biofilms, including *L. reuteri*. Instead, these cells must be able to adhere to other bacteria. LysM domain/YG-motif proteins, also called ‘‘aggregation promoting factors” or *apf* proteins, contribute to aggregation. They contribute to cell aggregation (Reniero et al., 1992) and are typified by a lysM domain, which binds to peptidoglycan on the surface of Gram-positive bacteria, and a YG-motif, which binds to carbohydrate epitopes presumably on the surface of other cells. These domains are located on the N and C-terminus of the protein, respectively, giving two adhesive ends to the protein and a mechanism for cell-cell adherence (Goh and Klaenhammer, 2010).

Two key regulatory systems were also found to be important to biofilm formation, Lr_70532 and LytS (Lr_69269). First, the peptide transporter Lr_70532 is co-localized with a putative bacteriocin-like peptide (Lr_70531), which could be involved in quorum sensing, a histidine kinase (Lr_70529), and a response regulator (Lr_70530) which could detect and mediate gene expression differences related to rodent colonization by this strain. However, these other genes were not found to be rodent specific (Frese

et al., 2011) or differentially regulated during biofilm growth and their exact function has yet to be determined.

The other regulatory system, LytS (Lr_69269) is a homolog to the LytSR regulator of autolysis in *Staphylococcus aureus* and bears sequence and structural homology to this system (Brunskill and Bayles, 1996, Sharma-Kuinkel et al., 2009). In *S. aureus*, this regulator controls cell autolysis to create an extracellular matrix of DNA and to regulate cell death in the biofilm (Mann et al., 2009, Bayles, 2007). In *S. aureus*, these processes are vital to biofilm growth (Sharma-Kuinkel et al., 2009). However, in *L. reuteri* 100-23, inactivation of the LytS and LrgA regulators did not affect biofilm formation as they did in *S. aureus* (Brunskill and Bayles, 1996, Sharma-Kuinkel et al., 2009). In *S. aureus*, inactivation of LytS led to a decrease in autolysis and a buildup of biofilm growth, the opposite effect of what was observed here. Thus, while these systems seem to have sequence and structural similarity, the ultimate action of these regulators is very different.

3.4.5 A Model for Biofilm Formation by *L. reuteri* 100-23 *in vivo*

Together, these results provide evidence for a hypothetical model of biofilm formation by rodent-lineage *Lactobacillus reuteri in vivo* (Figure 3.6). First, *L. reuteri* cells growing in batch culture, express surface proteins associated with the rodent *L. reuteri* lineage (eg. Lr_70131, Lr_70134, Lr_70902). These proteins contribute to initial adhesion to the forestomach epithelium. This is a key step in biofilm formation, as adhesion to the surface is critical for all subsequent steps.

Concurrently, environmental cues from the host gastrointestinal tract (eg. the low pH of the stomach) stimulate sensing histidine kinases to activate response regulators,

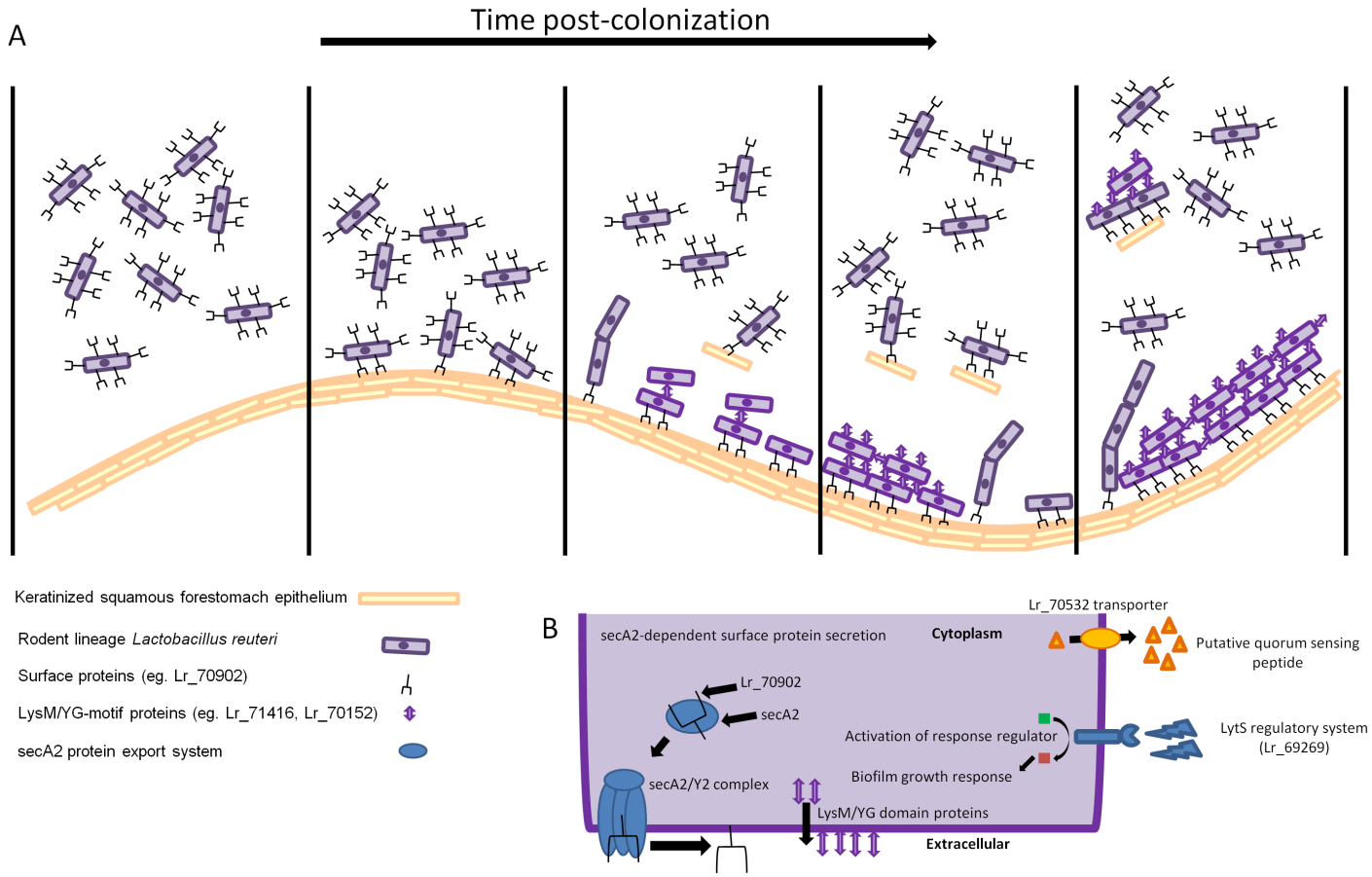

Figure 3.6(A) Hypothetical model of biofilm formation by rodent lineage *L. reuteri* after inoculation. First, cells in the lumen (first panel) adhere to the surface (second panel) using large surface proteins and begin to multiply (third panel) while they begin to shift gene expression to produce LysM protein binding epitopes (fourth panel) which could include modified teichoic acids or other cell surface modifications. Finally, a stable population on the forestomach epithelium is present, despite occasional shedding of the epithelium layer (fifth panel). The mechanisms by which large surface proteins are exported and key biofilm regulatory genes affect cell growth is also shown (B).

leading to downstream gene expression differences (Figure 3.6B). Responding to a low pH, in particular, is a physiologically relevant response that has been studied in *L. reuteri* for its impact on global gene expression shifts (Wall et al., 2007) and the induced responses have been shown to be relevant for survival *in vivo* (Walter et al., 2007). The gene expression shifts as a result of growth in the rodent host, rather than colonization status, appeared to drive gene expression changes, and these regulatory networks are likely to be responsible for this. How these sensing histidine kinases such as LytS (Lr_69269) recognize these changes and induce gene expression shifts has yet to be determined.

Finally, cells express lysM/YG proteins such as lysM2 and lysM3 (Lr_71416 and Lr_70152, respectively) which are predicted to facilitate aggregation by binding to both peptidoglycan (with the lysM domain) and carbohydrate moieties (with the YG-motif) as LysM/YG proteins can induce aggregation in competent lactobacilli (Goh and Klaenhammer, 2010, Turner et al., 2004). However, this aggregation is not observed in *in vitro* assays with *L. reuteri* 100-23 (Walter et al., 2008), likely for two reasons. First, gene expression analysis (Table A.10) indicates that these proteins are highly expressed in biofilms relative to batch culture (Figure 3.5C). Thus the predicted aggregation phenotype may only be observed under specific conditions. As the inactivation of these proteins results in reduced biofilm formation, they are clearly linked to biofilm growth. Second, the YG motif in these proteins binds specific carbohydrate moieties, which may include teichoic acids (Reniero et al., 1992), an ecologically relevant component of the *L. reuteri* cell surface (Walter et al., 2007), or other differentially expressed surface moieties (Wall et al., 2007, Hüfner et al., 2008, Sims et al., 2011) to aggregate these cells on the stomach epithelium (Figure 3.6).

3.5 Conclusions

This study identified host-specific biofilm formation among separate host-associated lineages of *L. reuteri*. Using *in vitro* models of biofilm formation and transcriptomics, bioinformatic comparisons, and other models of Gram positive biofilms, we identified the molecular basis for this biofilm formation and found that cell aggregation and adherence were key to the formation of this structure. Future studies which determine the cues of these regulatory networks (especially Lr_69269 and Lr_70532) will provide further important evidence as to how rodent-lineage *L. reuteri* regulate adherence and aggregation genes which are important to biofilm formation *in vivo*.

3.6 Acknowledgments, Contributions, and Funding

We graciously thank the Core for Applied Genomics and Ecology (University of Nebraska Lincoln) for their assistance in conducting the microarray experiments and conducting the statistical analysis; Dr. You Zhou, Teresa Fangman, Han Chen, and Christian Elowsky at the University of Nebraska Lincoln Microscopy Core for their assistance with confocal and electron microscopy; and Robert Schmaltz for assistance with animal experiments. SAF, DAP, AKB, and JW designed experiments; YZ and HC provided methodological and experimental support; SAF, JW, and CZ performed the experiments; SAF, TF, and CZ collected data; SAF and JW wrote the paper. We also thank Biogaia AB (Stockholm, Sweden) for their financial support.

Chapter 4

Comparison of the Colonization Ability of Autochthonous and Allochthonous Strains of Lactobacilli in the Human Gastrointestinal Tract

Steven A. Frese, Robert W. Hutkins, Jens Walter

Preface

This chapter was published in September 2012 edition of the journal *Advances in Microbiology* .

Keywords: *Lactobacillus*, probiotic, *Lactobacillus reuteri*, *Lactobacillus mucosae*, *Lactobacillus acidophilus*, gut microbiology, microbial ecology, colonization resistance

Abbreviations and Acronyms:GI, gastrointestinal; RAPD, random amplification of polymorphic DNA; qPCR, quantitative real-time PCR; PCR, polymerase chain reaction; MRS, Mann-Rogosa-Sharpe Media; NCBI, National Center for

Biotechnology Information; MLSA, multi-locus sequencing analysis; IRB, Institutional Review Board; CFU, colony-forming units;

4.1 Abstract

Bacteria of the genus *Lactobacillus* are widely used as oral probiotics due to their putative health benefits. In this study, we compared the colonization ability of two *Lactobacillus* strains that were identified as autochthonous to the human gastrointestinal tract (*Lactobacillus reuteri* ATCC PTA 6475 (MM4-1a) and *Lactobacillus mucosae* FSL-04) with that of an allochthonous strain (*Lactobacillus acidophilus* DDS-1). Colonization ability was tested in a single-blinded, cross-over study, with twelve human subjects. The test strains were quantified in fecal samples by two independent methods, selective plating and molecular typing and quantitative real time PCR. The study revealed that the two autochthonous strains (*L. reuteri* ATCC PTA 6475 and *L. mucosae* FSL-04) reached higher population levels in fecal samples and were recovered more frequently from subjects compared to the allochthonous strain (*L. acidophilus* DDS-1). All three strains became undetectable 8 days after the end of consumption with one exception, showing that persistence of all three strains remains short term in most individuals. In conclusion, this study showed that autochthonous Lactobacillus strains can be established more efficiently, albeit temporarily, in the human gastrointestinal tract, suggesting that evolutionary and ecological characteristics could be valuable criteria for the selection of probiotic strains.

4.2 Introduction

Probiotic bacteria are defined by the FAO/WHO ‘‘as live organisms which when administered in adequate amounts confer a health benefit to the host”. Implied in this definition is the expectation that these orally ingested organisms would reach the intestinal tract alive and at physiologically relevant levels (Ouwehand et al., 2002). Probiotic microorganisms are hypothesized to be functionally active in the human gut, such that they can influence the host through a variety of physiological mechanisms, including direct effects on the host immune system, *in situ* production of bioactive compounds, and competition with the resident microbiota and pathogens (O'Toole and Cooney, 2008). Some probiotic functions, such as *in vivo* production of bioactive compounds and competition with pathogens, also require that probiotic bacteria are metabolically active in the human gut and competitive under the prevailing conditions. Accordingly, it is often considered an important prerequisite of probiotic cultures that they originate from humans in order to ensure adaptation and persistence in the human gut (Ouwehand et al., 2002, Dunne et al., 1999, Sanders, 2003).

Although many of the commercially available probiotic strains are of human fecal origin and are capable of surviving gastrointestinal passage, they are still rapidly eliminated after administration has ended (Jacobsen et al., 1999, Vesa et al., 2000, de Champs et al., 2003, Valeur et al., 2004, Klingberg and Budde, 2006, Oozeer et al., 2006, Tuohy et al., 2007, Dommels et al., 2009). The inability of probiotic bacteria to persist in the human intestinal tract has been attributed to the phenomenon of colonization resistance, whereby the resident gut microbiota restricts access of allochthonous organisms (Stecher and Hardt, 2011). Moreover, even strains that are true autochthonous members of the microbiota of specific human subjects may not be able to colonize the gastrointestinal tract (GIT) of other humans due to individual

differences (Walter, 2008).

However, it is also important to recognize that many strains currently used as probiotics belong to species which are considered allochthonous to the human intestinal tract (Reuter, 2001, Walter, 2008). Species often used as probiotics such as *Lactobacillus acidophilus, Lactobacillus casei, Lactobacillus paracasei, Lactobacillus rhamnosus, Lactobacillus delbrueckii, Lactobacillus brevis, Lactobacillus johnsonii, Lactobacillus plantarum*, and *Lactobacillus fermentum*, although commonly found in fecal samples, have never been shown to form stable populations in the human gut and are likely to originate from food or the oral cavity (Reuter, 2001, Walter, 2008). In contrast, other species of *Lactobacillus*, in particular, *Lactobacillus reuteri, Lactobacillus ruminis, Lactobacillus gasseri*, and *Lactobacillus salivarius*, have been reported to be autochthonous to the human gastrointestinal tract (Tannock et al., 2000, Reuter, 2001). As autochthonous members of the gut microbiota, these species are likely to occupy specific niches that allow their replication and the establishment of stable population over long periods (Savage, 1977). Recent comparative genome studies have begun to characterize the molecular basis of autochthony in the species *L. gasseri*, *L. reuteri*, and *L. ruminis*, and these studies identified adaptive traits that might contribute to the ecological success of lactobacilli in the human GIT (Azcarate-Peril et al., 2008, Forde et al., 2011, Frese et al., 2011).

In this study, our goal was to compare survival and persistence rates of orally-consumed autochthonous and allochthonous *Lactobacillus* strains in human subjects. Three strains were included in this study, *L. reuteri* ATCC PTA 6475 (MM4-1a), *L. mucosae* FSL-04, and *L. acidophilus* DDS1. *L. reuteri* is a species long considered to be an autochthonous member of the human gut microbiota (Reuter, 2001). Strain ATCC PTA 6475 is a member of the MLSA lineage II of *L. reuteri*, which is almost completely composed of strains of human fecal origin, indicating that this subpopulation is adapted

to the human GIT (Oh et al., 2010). The second strain, *L. mucosae* FSL-04, was continuously isolated from fecal samples from a single healthy adult subject in high numbers during a 15-week period (see below). The third strain, *L. acidophilus* DDS-1, is a commonly-consumed probiotic (Sanders, 2003, Rabot et al., 2010) belonging to the species *L. acidophilus* which is considered allochthonous to the human GIT (Tannock et al., 2000, Reuter, 2001, Sanders and Klaenhammer, 2001, Walter et al., 2001, Walter, 2008). To compare establishment and persistence of these three strains in the human gastrointestinal tract, we performed a human cross-over study in which healthy individuals consumed the test strain and provided fecal samples that were then analyzed by cultural enumeration, molecular typing of isolates, and quantitative real-time PCR (qPCR).

4.3 Materials and Methods

4.3.1 Use of Human Subjects

The human trial of this study was approved by the Institutional Review Board of the University of Nebraska (IRB Approval Number: 2009079919FB), and written informed consent has been obtained from all subjects.

4.3.2 Bacterial Strains and Culture Conditions

Strains used in this study are listed in Table 1. All strains were grown in Mann-Rogosa-Sharpe (MRS) (Difco) supplemented with 1.0% Maltose and 0.5% Fructose under anaerobic conditions at 37°C. *L. acidophilus* DDS-1 was provided by Nebraska Cultures (Walnut Creek, CA USA). *L. reuteri* ATCC PTA 6475 (MM4-1a) was obtained from BioGaia (Stockholm, Sweden). *L. mucosae* FSL-04 was isolated from

Strain	Origin	Reference or Source
L. reuteri ATCC PTA 6475 (MM4-1A)	Human	BioGaia AB (Stockholm, Sweden)
L. reuteri 100-23	Rat	Wesney and Tannock 1979
L. reuteri mlc3	Mouse	Oh et al. 2010
L. reuteri lpuph	Mouse	Oh et al. 2010
L. mucosae FSL-04	Human, fecal sample	This Study
L. mucosae S5 (DSM13346)	Porcine, small intestine	Roos et al. 2000
L. mucosae 1028	Porcine, small intestine	Axelsson and Lindgren 1987
L. mucosae 1031	Porcine, small intestine	Axelsson and Lindgren 1987
L. acidophilus DDS-1	Milk	Nebraska Cultures (Walnut Creek, CA)
L. acidophilus ATCC 4356	Human	Efthymiou and Hansen 1962

Table 4.1 Strains used in this study.

the feces of a healthy adult human during a previous human trial as described above (Martinez et al., 2010).

4.3.3 Molecular Typing and Identification of Isolates

The molecular typing of strains was performed by Random Amplification of Polymorphic DNA (RAPD) using the primer M13V (Table 2) as described by Meroth et al. (Meroth et al., 2003). Isolates were assigned to species by 16S rRNA gene comparisons as described by Hammons et al. (Hammons et al., 2010). To obtain around 1300 bp of the 16S rRNA gene for exact classification, PCR products were generated and sequenced with primers 8F (5'-AGAGTTTGATCCTGGCTCAG-3') and 1391R (5'-GACGGGCGGTGWGTRCA-3').

4.3.4 Preparation of *Lactobacillus* Dosages

Food-grade freeze-dried powders of *L. reuteri* ATCC PTA 6475, *L. mucosae* FSL-04, and *L. acidophilus* DDS-1 were prepared by Culture Systems, Inc (Mishawaka, Indiana USA) and stored at -20°C throughout the study. Upon receipt, viable cell counts and purity of all three preparations were determined and strain identify was verified by RAPD. Prior to each feeding period, viable cell counts were determined in the powders to adjust the daily dose (10^9 cells) for each strain, where necessary.

Subjects were provided with pre-dosed freeze-dried preparations of lactobacilli and instructed to store the preparations at 4°C and reconstitute the powder in 40 mL cold or room-temperature milk and consume at their own convenience, with a meal.

4.3.5 Human Trial

A blinded, crossover study was performed with twelve healthy adult humans (six male and six female, age 21-27). Subjects were selected to be tolerant of milk products, free of chronic gastrointestinal disorders, non-vegetarians, and had not consumed antibiotics in the two months prior to the study. No dietary restrictions were placed on participants, except to avoid probiotic supplements, cultured dairy products, or products advertised as having live and active cultures. The study was conducted over three separate 8-week periods (each period separated by 3 to 4 weeks) where the individual strains were tested in succession. Each feeding period began with a two-week baseline period (no change in diet). The subjects then consumed a daily dose of 10^9 viable bacterial cells for 7 days, followed by a 5 week wash out period. Fecal samples were collected weekly, resulting in 2 fecal samples during the baseline period, and fecal samples taken at day 1, 8, and 15 of the wash out periods. The human trial of this study was approved by the Institutional Review Board of the University of Nebraska (IRB Approval Number: 2009079919FB), and written informed consent was obtained from all subjects.

Subjects completed a symptoms diary to assess the potential side effects of experimental strain administration. The symptoms included were bowel movement, stool consistency, discomfort, flatulence, abdominal pain, and bloating, and subjects were asked to score them on a scale from 1 (none, normal, good well-being) to 5 (severe symptoms and discomfort). All twelve subjects completed the trials, and self-reported

compliance with the experimental treatments was 100%.

4.3.6 Microbial Analysis of Fecal Samples

Subjects provided fresh fecal samples in sterile fecal sample collection containers, and samples were processed within four hours of defecation. A ten-fold dilution of each sample in sterile phosphate buffered saline (PBS) (pH 7.0) was immediately frozen at -80°C for later DNA extraction (see below). Furthermore, a 10-fold dilution series was made with sterile saline (0.9% NaCl), and aliquots were plated on Rogosa SL Agar, which is selective for lactobacilli. Plates were incubated anaerobically for 48 hours at 37°C before enumeration. To analyze the total *Lactobacillus* population, 10 colonies were picked at random from a dilution agar plate containing about 100 colonies. Selection of colonies was randomized by drawing intersecting lines across the plate, and picking colonies along the lines until ten had been recovered, in order to remove operator bias. The isolates were differentiated after subculture by RAPD analysis through direct comparison with the molecular fingerprint obtained with the test strains (Figure 1), a strategy which has been effectively used to differentiate *Lactobacillus* isolates from oral and fecal samples (Hammons et al., 2010, Bello et al., 2003). To determine the bacterial population of the test strains in fecal samples, total counts of lactobacilli (on Rogosa SL Agar) were multiplied by the percentage of the strain among the 10 typed colonies. If none of the ten colonies corresponded to the test strain, *Lactobacillus* counts were multiplied by 0.09. Although this analysis would over-estimate the numbers of the test strains in fecal samples from subjects with background *Lactobacillus* counts, the analysis still allowed the detection of a significant increase due to the administration of the strain during the test period.

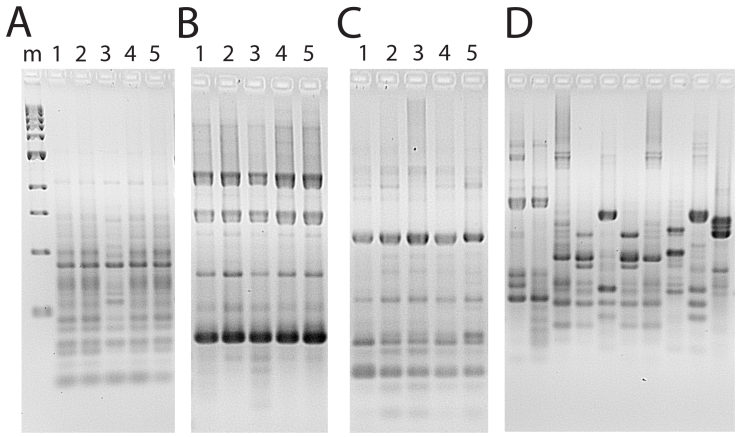

Figure 4.1 Identification of *Lactobacillus* strains by RAPD typing. RAPD patterns from *L. reuteri* MM4-1a (a), *L. mucosae* FSL-04(b), and *L. acidophilus* DDS-1 (c). Each lane represents a RAPD-typed culture obtained from the original stock culture (Lane 1), the freeze-dried poweders used in the human trial (Lane 2), and from colonies isolated from subject fecal samples after consumption of the strain (Lanes 3-5). Examples of isolates obtained by fecal culture during this study from multiple subjects indicating banding patterns are distinct between strains (d). Lane m: 1kb DNA ladder (New England Biolabs, Massachusetts, USA).

4.3.7 Quantitative real-time PCR (qPCR)

Lactobacillus strains were quantified in human fecal samples by qPCR. Strain specific PCR systems could not be designed for the human *L. reuteri* strains of the MLSA lineage II and for *L. acidophilus* strains, as these strains are clonal at all loci that were tested (Frese et al., 2011). Therefore, species-specific primers that targeted the 16S-23S rDNA intergenic spacer region (Table 2) were used. Specific primers for *L. acidophilus* targeting this region were previously described by Haarman and Knol (Haarman and Knol, 2006). Primers for *L. reuteri* and *L. mucosae* were designed to also target the same region within the 16S-23S rDNA spacer. For *L. mucosae*, the 16S-23S rDNA spacer region was amplified from four strains (Table 1) by primers 16S/p2 and 23S/p10 (Kabadjova et al., 2002) (Table 2) and sequenced (Table 3). Sequences for *L. reuteri* were obtained from available genome sequences (*L. reuteri* ATCC PTA

Primer Name	Sequence 5' to 3'	Target	Reference
16S p2	CTTGTACACACCGCCCGTC	1388-1406 16S rRNA	Kabadjova et al. 2002
23s p10	CCTTTCCCTCACGGTACTG	546-474 23S rRNA	Kabadjova et al. 2002
F_acidophilus_IS	GAAAGAGCCCAAACCAAGTGATT	16S-23S rDNA spacer	Haarman and Knol 2006
R_acidophilus_IS	CTTCCCAGATAATTCAACTATCGCTTA	16S-23S rDNA spacer	Haarman and Knol 2006
L.mucosae_For	CACAATTAAACCGAGAACACC	16S-23S rDNA spacer	This Study
L.mucosae_Rev	ATGATCTTACGATCACCTCAGTTA	16S-23S rDNA spacer	This Study
L.reuteri_For	AACAATAAACCGAGAACACC	16S-23S rDNA spacer	This Study
L.reuteri_Rev	CCTTCATAACTTAACCTAAACAA	16S-23S rDNA spacer	This Study
M13V	GTTTCCCCAGTCACGAC	-	-

Table 4.2 Primers used in this study.

6475, 100-23, mlc3, and lpuph), and sequences were aligned using CLUSTALW to generate consensus sequences from which species-specific primers were constructed. Primers were validated for target specificity using DNA from lactobacilli-negative fecal samples ($< 10^2$ CFU/g lactobacilli) and DNA isolated from strains of each of the species of Lactobacillus used (Table 1). Specificity for all primers was further validated in silicio using the NCBI database.

Genomic DNA was extracted from fecal samples as described previously (Walter et al., 2001). qPCR was performed in a Mastercycler Realplex2 (Eppendorf AG, Hamburg, Germany) using the Quanti-Fast SYBR Green PCR system (Qiagen, Düsseldorf, Germany) as directed by the manufacturer. The PCR program consisted of a single 95°C step for 5 mins, followed by 40 cycles of a two-step PCR reaction, beginning with a 10 second 95°C denaturation step and a 30 second 60°C annealing/extension step. Melting curves were also performed, consisting of a denaturation step of 15 s at 95°C, an increase from 60°C-95°C over a 20-min period, and a final step of 15 s at 95°C. Reactions were performed in 25 μL volumes containing 0.5 μm each primer and 1 μL of extracted DNA.

Standard curves for absolute quantification were prepared from overnight cultures (14 h) of each organism (*L. reuteri* ATCC PTA 6475, *L. mucosae* FSL-04, and *L. acidophilus* DDS-1) that were plated in triplicate and enumerated, in parallel to DNA

extraction from 1 ml of the culture. A tenfold dilution series was generated with this DNA in 10 mM Tris-Cl pH 8.0 to generate a range of DNA concentrations representing 10^9 cells to 10^4 cells, based on culture-based enumeration from the original culture. Reactions were performed in triplicate for the standard curves and in duplicate for fecal DNA samples. The correlation coefficient, r^2, for the three standard curves were 0.99 and PCR efficiencies were 0.93 (for *L. mucosae*), 0.98 (for *L. reuteri*) and 1.00 (for *L. acidophilus*.

The qPCR systems were validated by spiking a lactobacilli-negative fecal sample (less than 10^2 CFU per gram lactobacilli as determined by selective culture on Rogosa SL Agar) with 10 to 10^8 cells per gram of each of the three test strains. DNA was extracted and qPCR was performed in duplicate as described above. Predicted numbers of lactobacilli demonstrated a linear dynamic range closely reflecting the number of known added cells ($r^2 > 0.98$ for MM4-1a, $r^2 > 0.99$ for FSL-04, and $r^2 > 0.93$ for DDS-1) from 10^4 cells per g to 10^8 CFU per gram. Below 10^4 cells per gram fecal samples, background signal prevented accurate quantification. Within the dynamic range, cell numbers determined by qPCR were in agreement with the cell numbers that were spiked to the samples as indicated by the slopes of the trend line in linear regression (b = 0.91 for ATCC PTA 6475, b = 0.86 for FSL-04 and b = 1.01 for DDS-1) when predicted cell numbers were plotted against cells added.

4.3.8 Statistical Analysis

Results are presented as means ± standard deviations. Statistical tests for treatment effects of test strain administration on the abundance of individual strains or abundance of *Lactobacillus* species were performed either by one-way analysis of

Strain	Sequence Determined	Accession Number
L. mucosae FSL-04	16S/23S intergenic spacer region	JN368428
L. mucosae FSL-04	Partial 16S rRNA gene	JN092131
L. mucosae S5	16S/23S intergenic spacer region	JN592586
L. mucosae 1028	16S/23S intergenic spacer region	JN592585
L. mucosae 1031	16S/23S intergenic spacer region	JN592584
L. acidophilus DDS-1	16S/23S intergenic spacer region	JN368427
L. acidophilus DDS-1	Partial 16S rRNA gene	JN368429

Table 4.3 Sequences determined during this study.

variance (ANOVA) with repeated measure or by one-way ANOVA, followed by Tukey post-hoc test.

4.4 Results

4.4.1 Selection of *Lactobacillus mucosae* FSL-04 as an autochthonous member of the human intestinal tract

To identify human autochthonous lactobacilli, we screened fecal samples from 11 human subjects that had participated in a previous study (Martinez et al., 2010). Fecal *Lactobacillus* populations were quantified over the duration of four months by plating serial dilutions on Rogosa SL Agar plates. As observed in previous studies (Walter, 2008, Tannock et al., 2000, Walter et al., 2001, Bello et al., 2003), most of the subjects harbored low numbers of lactobacilli in fecal samples ($< 10^6$ CFU/g). However, one subject consistently shed high levels of lactobacilli in fecal samples (Figure 4.2A) over the entire 4 month study. Two dominant colony types were identified, and 28 isolates were picked randomly during the 15 week duration of the trial representing both colony morphologies. Molecular typing of these isolates by RAPD revealed the presence of two unique strains. One strain was detected throughout the entire 15

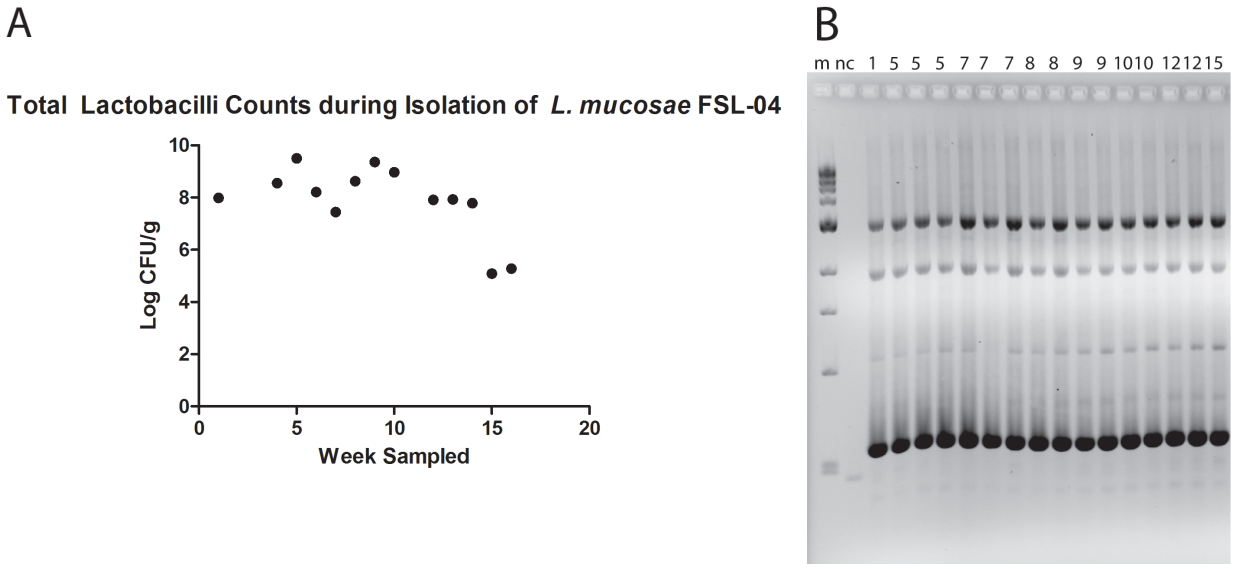

Figure 4.2 Identification of *Lactobacillus mucosae* FSL-04 as an autochthonous member of the human gut microbiota. (a) Total lactobacilli (CFU/g feces) enumerated from Rogosa SL agar over several weeks during isolation of FSL-04 from a single healthy adult human. (b) RAPD patterns of isolates identified as *Lactobacillus mucosae* obtained from these fecal samples. Lane m: 1kb DNA ladder (New England Biolabs, Massachusetts, USA). Lane nc: negative control. Numbered lanes correspond to the respective weeks were obtained.

week time period by RAPD-typing (Figure 4.2B) and classified as *L. mucosae* by 16S rRNA sequence analysis (>99.9% homology to 16S rRNA gene from *L. mucosae* CCUG 43179T). Given that this strain was found in the fecal samples of this human subject in high numbers, we concluded that it represented an autochthonous member of the gut microbiota. Importantly, no lactobacilli were cultured from saliva samples obtained from this subject at three different time points (data not shown), excluding the oral cavity as a potential origin of this strain. A single isolate (*L. mucosae* FSL-04) was selected for use in this study.

4.4.2 Human trial to compare persistence of *Lactobacillus* strains

We performed a human cross-over feeding study to compare the intestinal establishment and persistence of *L. reuteri* ATCC PTA 6475, *L. mucosae* FSL-04, and *L. acidophilus* DDS-1 after oral administration of a daily dose containing 10^9 viable cells for 7 days to 12 human subjects. No adverse effects, as assessed by symptoms diaries querying bowel movement, stool consistency, general discomfort, flatulence, abdominal pain, and bloating, were detected for any of the strains during the feeding period (data not shown). The *Lactobacillus* population was characterized in fecal samples two weeks before subjects received the probiotic strains. Persistence was tested at day 1, 8 and 15 of the post-test period by quantitative culture and qPCR. The proportion of each probiotic strain as a percent of the total cultivable *Lactobacillus* population was determined by RAPD typing of 10 random isolates at each time-point, a number of isolates that has been shown to give a sufficient overview about the *Lactobacillus* strain distribution in human fecal samples (Tannock et al., 2000). As shown in Figure 1, each strain had a distinct RAPD-pattern and could easily be distinguished from other isolates obtained throughout the study.

4.4.3 Transient recovery of lactobacilli from humans after administration of three test strains

The number of fecal lactobacilli varied markedly among the 12 subjects during the baseline period, ranging from $< 10^2$ to 10^9 CFU/gram, with an average of around 10^4 CFU/gram (Table 4). RAPD analysis revealed that none of the three test strains were detectable during the baseline. After subjects had consumed the test strains for 7 days, culture analysis of day 1 fecal samples showed an increase in total *Lactobacillus*

numbers to about 10^5 CFU/gram. The RAPD analysis of random isolates revealed that the test strains could be detected in a majority of the subjects (Table 4), indicating that the increase in total lactobacilli detected soon after consumption was a result of administration of the respective experimental strain.

We next estimated the fecal populations of the test strains by multiplying total lactobacilli counts by the relative proportions of each strain, as determined by RAPD analysis (see Materials and Methods). Administration of *L. reuteri* ATCC PTA 6475 and L. mucosae FSL-04 led to a significant increase in the numbers of the respective strains at day 1 after consumption when compared to baseline, while administration of DDS-1 did not result in a significant increase (Figure 3A). The populations of *L. reuteri* ATCC PTA 6475 (5.22 Log_{10} CFU/gram) and *L. mucosae* (4.94 Log_{10} CFU/gram) were higher than that of *L. acidiophilus* (4.26 Log_{10} CFU/gram), but differences did not reach statistical significance.

The number of total lactobacilli returned to baseline at day 8 after consumption (Table 4 and Figure 3A), and the probiotic strains were not detectable anymore with one exception (*L. reuteri* ATCC PTA 6475 was detected at day 8 and 15 of the washout period in one subject).

4.4.4 *L. reuteri* ATCC PTA 6475 and *L. mucosae* FSL-4 can be established more efficiently and in higher numbers than *L. acidophilus* DDS-1

All three test strains could be detected by the culture-RAPD method in a subset of subjects at day 1, but the rate of recovery differed between strains (Table 4). At day 1 after administration, 84% of the isolates were identified as *L. reuteri* ATCC PTA 6475 at day 1, compared to 59% and 24% for *L. mucosae* FSL-04 and *L. acidophilus*

Total Lactobacilli in Log_{10} CFU/g Feces (Percent of isolates typed as the probiotic strain)													
	Subject												
Strain	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Mean**
Baseline 1													
ATCC PTA 6475	8.64(nd)	<2 (nd)	<2 (nd)	<2 (nd)	<2 (nd)	2.77 (nd)	3.29 (nd)	3.47 (nd)	<2 (nd)	<2 (nd)	3.57 (nd)	4.21 (nd)	3.16 (nd)
FSL-04	4.30(nd)	3.20(nd)	3.00(nd)	5.85(nd)	5.30(nd)	5.76(nd)	<2(nd)	6.79(nd)	<2(nd)	<2(nd)	4.43(nd)	2.84(nd)	3.96(nd)
DDS-1	2.60(nd)	2.30(nd)	5.95(nd)	7.59(nd)	<2(nd)	5.68(nd)	3.55(nd)	3.00(nd)	<2(nd)	<2(nd)	4.65(nd)	<2(nd)	3.61(nd)
Baseline 2													
ATCC PTA 6475	<2(0)	6.39(0)	<2(0)	6.94(0)	<2(0)	3.96(0)	3.55(0)	<2(0)	<2(0)	<2(0)	<2(0)	3.30(0)	3.18(0)
FSL-04	5.84(0)	<2(0)	6.82(0)	7.92(0)	<(0)	4.70(0)	4.75(0)	3.41(0)	4.63(0)	<2(0)	4.19(0)	<2(0)	4.19(0)
DDS-1	5.40(0)	4.74(0)	6.30(0)	7.16(0)	6.48(0)	2.30(0)	3.30(0)	2.30(0)	<2(0)	<2(0)	3.58(0)	4.77(0)	4.19(0)
Day 1													
ATCC PTA 6475	7.37(100)	6.79(90)	6.45(100)	7.46(100)	5.09(100)	4.34(90)	3.51(80)	6.21(0)	3.57(100)	3.90(100)	4.44(90)	5.03(60)	5.35(84)
FSL-04	4.81(100)	6.14(40)	3.49(0)	8.01 (10)	5.72(70)	6.11(40)	5.34(90)	4.97(80)	5.16(90)	3.43(100)	5.20(0)	5.21(90)	5.30(59)
DDS-1	3.83(0)	4.87(100)	5.01(20)	4.38(0)	5.09(90)	7.09(0)	5.06(70)	5.56(10)	<2(0)	4.52(0)	5.13(0)	5.76(0)	4.86(24)
Day 8													
ATCC PTA 6475	4.90(0)	<2(0)	4.34(100)	5.77(0)	<2(0)	<2(0)	3.27(0)	5.30(0)	<2(0)	<2(0)	<2(0)	4.95(0)	3.38(8)
FSL-04	3.94(0)	2.90(0)	<2(0)	7.15(0)	2.90(0)	3.77(0)	2.84(0)	4.89(0)	5.05(0)	2.47(0)	4.41(0)	4.05(0)	3.86(0)
DDS-1	4.22(0)	<2(0)	<2(0)	8.18(0)	2.47(0)	5.28(0)	<2(0)	2.60(0)	4.43(0)	<2(0)	4.03(0)	<2(0)	3.43(0)
Day 15													
ATCC PTA 6475	4.54(0)	<2(0)	6.41(40)	6.93(0)	3.30(0)	3.77(0)	<2(0)	3.82(0)	4.73(0)	<2(0)	<2(0)	4.03(0)	3.79(3)
FSL-04	<2(0)	3.98(0)	3.64(0)	7.58(0)	<2(0)	5.38(0)	<2(0)	3.68(0)	4.47(0)	<2(0)	6.93(0)	4.13(0)	3.98(0)
DDS-1	2.84(0)	2.47(0)	4.03(0)	7.51(0)	<2(0)	5.38(0)	7.19(0)	<2(0)	4.25(0)	<2(0)	3.21(0)	6.03(0)	4.08(0)

Table 4.4 Log_{10} total lactobacilli populations in subject fecal samples before (Baseline 1 and 2) and post-test (Day 1, 8, and 15). Percentages of isolates typed as the respective probiotic strain by RAPD are shown in parentheses (nd, not determined).

DDS-1, respectively (Figure 3B). The difference in recovery rate of L. reuteri ATCC PTA 6475 was significantly higher than that of *L. acidophilus* DDS-1 (Figure 3B). In addition, the autochonous strains were detectable in fecal samples of more subjects. *L. reuteri* ATCC PTA 6475 was recovered from 11 out of the 12 subjects, *L. mucosae* FSL-04 and was present in 10 subjects, while *L. acidiophilus* DDS-1 was present in 5 subjects.

Although the background *Lactobacillus* populations in human fecal samples were low, they still confounded the culture-based analyses as they prohibited an exact quantification of the test strains. To determine the establishment of the test strains in the human gastrointestinal tract, samples taken during baseline, day 1 post-test, and day 8 post-test were analyzed by species-specific qPCR. We used a species and not strain-specific primers for the qPCR, as human *L. reuteri* and *L. acidophilus* isolates are highly clonal, making the development of strain specific primers impractical. Furthermore, we chose to target the same gene (the 16S-23S rRNA spacer region) in all three strains used during this study to avoid PCR bias. Population levels were determined by absolute quantification using standard curves obtained from bacterial cells. While strain-specific primers would be advantageous, the analysis of baseline fecal samples indicated that no confounding *Lactobacillus* populations were present in any of the subjects. As shown in Figure 3C, the levels of *L. reuteri, L. mucosae,* and *L. acidophilus* were generally below the detection limit (10^4 cell/gram) in most of the samples during baseline and at day 8 of wash-out, indicating that background levels of the species did not confound this technique.

At day 1 of wash-out, there was a statistically significant increase in each of the three species when compared to the baseline and the day 8 wash-out, with *L. reuteri* ATCC PTA 6475 reaching 5.14 Log_{10} cells/gram, *L. mucosae* FSL-04 reaching 5.03 Log_{10} cells/gram, and *L. acidophilus* DDS-1 4.32 Log_{10} cells/gram feces. The qPCR

analysis revealed that the two autochthonous strains *L. reuteri* ATCC PTA 6475 and *L. mucosae* FSL-04 reached significantly higher populations ($p < 0.01$) in fecal samples when compared to *L. acidophilus* DDS-1 (Figure 3C).

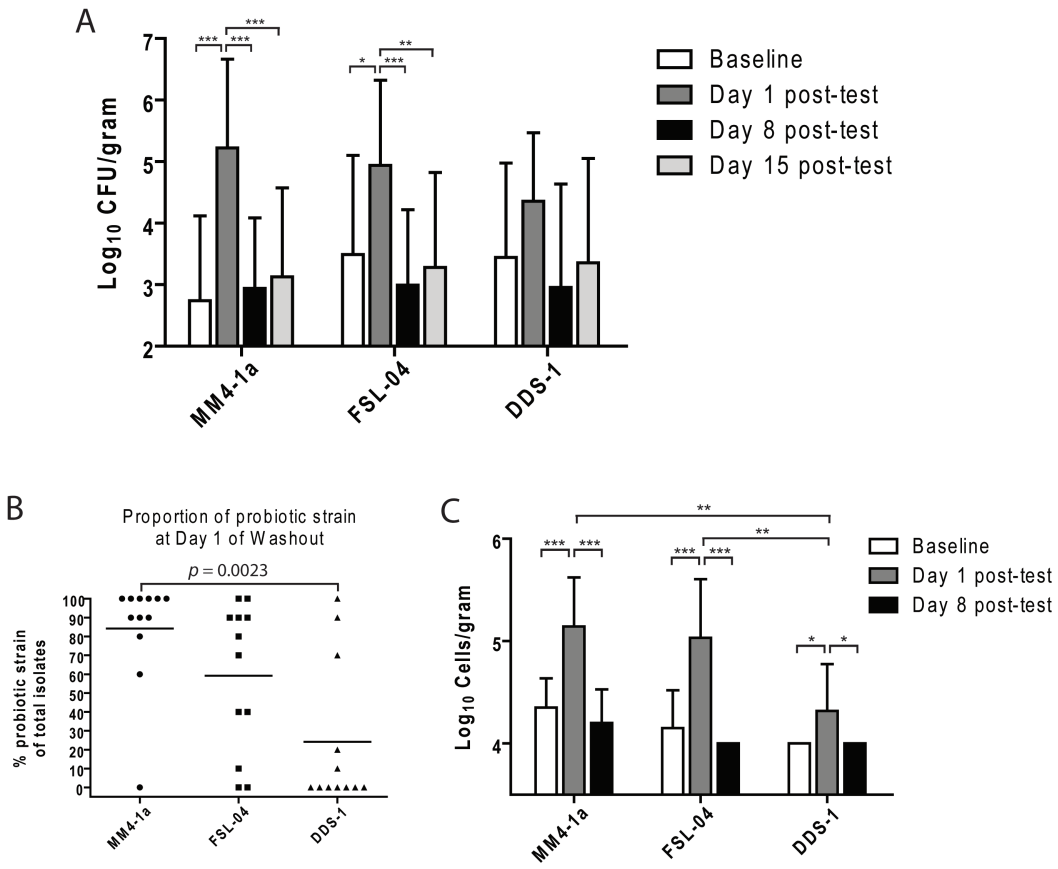

Figure 4.3 Quantification of probiotic strains in fecal samples of subjects as determined by culture- and molecular-based methods. (a) Culturable counts of *L. reuteri* MM4-1a, *L. mucosae* FSL-04, and *L. acidophilus* DDS-1 during baseline (second sample) and day 1, 8, and 15 post-test. Populations were determined by multiplying total number of lactobacilli with the proportion of the probiotic strain as determined by RAPD typing of 10 random colonies. Statistical analysis was performed with One-way ANOVA with repeated measures and Turkey post-hoc tests. (b) Percentage of the probiotic strains to the total *Lactobacillus* population at day 1 post-test as determined by RAPD typing of 10 random colonies. Statistical analysis was performed with One-way ANOVA with repeated measures and Turkey post-hoc tests. (c) Quantification of the *Lactobacillus* species *L. reuteri* (during treatment with strain MM41-1a), *L. mucosae* (during treatment with strain FSL-04), and *L. acidophilus* (during treatment with strain DDS-1) by qRT-PCR. Statistical analysis was performed with One-way ANOVA and Turkey post-hoc tests.

4.5 Discussion

The objective of this study was to test whether autochthonous *Lactobacillus* strains (*L. mucosae* FSL-04 and *L. reuteri* ATCC PTA 6475) can be established more efficiently in the human gastrointestinal tract than an allochthonous strain (*L. acidophilus* DDS-1) following a 7-day feeding period. Accordingly, we characterized the *Lactobacillus* populations in fecal samples obtained from humans that had consumed standardized, freeze dried cell preparations of three test strains using both culture-based and molecular (qPCR) methods. As observed in previous probiotic trials, all three *Lactobacillus* strains survived gastric passage and were temporarily detectable in the fecal samples (Jacobsen et al., 1999, Valeur et al., 2004, Dommels et al., 2009, Tannock et al., 2000, Sui et al., 2002). Most importantly, our findings indicate that the two autochthonous strains (*L. reuteri* ATCC PTA 6475 and *L. mucosae* FSL-04) reached higher populations and were generally more persistent than the allochthonous strain. In addition, the autochthonous strains, and *L. reuteri* ATCC PTA 6475 in particular, could be detected in a larger proportion of subjects and represented a larger percentage of the total *Lactobacillus* population (Table 4).

We suggest here that the more efficient establishment of autochthonous probiotic strains is due to their adaptation to the human gastrointestinal tract. The *L. reuteri* strain that was used during this study, which reached the highest levels of colonization and recovery, belongs to a subpopulation of the species that has been shown to be highly specific to the human gastrointestinal tract (Oh et al., 2010), and which possesses a genome content that reflects its adaptation to the human gastrointestinal tract (Frese et al., 2011). *L. mucosae* is a species that is ordinarily able to bind to both mucus and human blood group antigens (Roos et al., 2000, Watanabe et al., 2010), and the strain used in our study was routinely detected in a human subject in high

numbers over 4 months (Figure 2). It is unlikely that the population differences of*L. reuteri* and *L. mucosae*, compared to *L. acidophilus* are due to comparatively lower survival rate of *L. acidophilus* during gastric transit, as the latter species has been shown to have high rates of tolerance towards low acidity and bile acids (Jacobsen et al., 1999). Instead, the restricted ability of *L. acidophilus* to form stable populations in the human, as shown previously (Tannock et al., 2000, Reuter, 2001, Sanders and Klaenhammer, 2001, Walter et al., 2001) is likely due to the absence of specific adaptive features that are evidently present in *L. reuteri* and *L. mucosae*.

Although our data indicated that the two autochthonous *Lactobacillus* strains could be established in higher numbers in the human gut, the strains were no more persistent than the allochthonous strain. That probiotic bacteria can only be transiently established in the gut has been shown in many previous studies, and duration of persistence does not seem to depend on differences in inoculum dose, strain or species, carrier medium, and even duration of consumption (Jacobsen et al., 1999, Klingberg and Budde, 2006, Tannock et al., 2000). The data obtained during our study indicates that autochthony does not increase the duration of persistence. This finding suggests that it may be difficult, if not impossible, to establish a probiotic strain long term in the gastrointestinal tract of most human subjects, even if that microbe was derived from a stable component of one individual's gut microbiota. Therefore, the ability of the microbiota to prevent establishment of foreign organisms (colonization resistance), applies not only to pathogens, but to other gut microbes as well.

The ecological principles that govern community assembly of the gut microbiota and determine which lineages can become established are not completely understood. Modern concepts of community ecology suggest that a combination of niche-related and historic processes (e.g. in situ evolution of early colonizers) govern the process (Dethlefsen et al., 2006, Cavender-Bares et al., 2009, Walter and Ley, 2011). Historic

factors such as colonization order, transmission, and niche construction are inherently stochastic, and therefore cause variation in individual microbiomes, while in situ evolution of members ensures that members are highly adapted to available niches (Walter and Ley, 2011). These concepts have important implications if the goal is to establish a probiotic organism in the gut. First, the microbiome is individualized and composed of microbes that are adapted to occupy specific niches within each person's particular community, and thus, members of microbiomes are not necessarily interchangeable. Second, establishment of new microbes is only possible as long as niches are still open (early during assembly) or if niches become available (when community members are intentionally or accidentally removed from the community). If community structure becomes disrupted, for example through a high dose of antibiotics, microbes from another individual can be successfully and more permanently established (Khoruts et al., 2010). Therefore, probiotic bacteria might be more permanently established when administered after antibiotic treatments or early in life.

Although our findings indicate the autochthony of probiotic lactobacilli does not increase the duration of their persistence in the human gastrointestinal tract, the observation that autochthonous strains can be more efficiently established is clearly of practical importance. We not only detected *L. reuteri* ATCC PTA 6475 and *L. mucosae* FSL-04 in more subjects after administration, but these strains also reached about ten times higher cell numbers in fecal samples when compared to *L. acidophilus* DDS-1. These findings indicate that the effective dose of a probiotic is likely to be higher when an autochthonous strain versus an allochthonous strain is used. If probiotic action is dependent on the metabolic functionality of the organisms in the gut, such as direct antagonism or the production of bioactive substances, then a higher dose of viable organisms present the gut is likely to increase the success of a probiotic strategy. In conclusion, the results described in this study support the

notion that evolutionary and ecological characteristics are valuable criteria for the selection of probiotic strains (Walter, 2008). Future studies should be aimed to test persistence of other autochthonous strains, especially *L. ruminis*, which appears to be the dominant *Lactobacillus* species in the human gut (Tannock et al., 2000, Reuter, 2001, O'Callaghan and O'Toole, 2012).

4.6 Acknowledgments

The dedication of the participants in the human trial is gratefully acknowledged. We are thankful to Stefan Roos (Swedish University of Agricultural Sciences) for providing *L. mucosae* S5, 1028, and 1031 and Dr. David Mills (University of California - Davis) for providing *L. acidophilus* ATCC 4356. We thank Nebraska Cultures (Walnut Creek, CA) for providing *L. acidophilus* DDS-1 and funding this study.

Chapter 5

Conclusions, Remaining Questions, and Future Directions.

5.1 Conclusions

The gut microbiota is a complex, powerful, and important community with regards to host health and physiology (Bäckhead et al., 2004, Turnbaugh et al., 2006, Mazmanian et al., 2005). However, little is known about how microbes associate with their hosts or the ecological and evolutionary forces which shape both partner and host (Herre et al., 1999). In order to fill this knowledge gap, vertebrate gut host-microbe relationships must be identified to begin to characterize these relationships, understand how they are maintained, and identify the ecological and evolutionary pressures they face. Using established relationships to characterize host-microbe symbioses has proven to be a successful strategy in invertebrates (Ruby and Mcfall-Ngai, 1992, Mueller et al., 2001, Baumann et al., 1995). This work has resulted in surprising implications for fields as diverse as medicine and pest control (Stewart and William Costerton, 2001, Douglas, 2007) and such wide-ranging impacts could also result from similar research with vertebrates. However, better models, using vertebrates and their gut symbionts, will be necessary to answer these questions about this ecosystem and its constituents.

Chapter 1 provided an exposition of the origins of the microbiota, its assembly from a sterile community at birth to a robust and diverse ecosystem. It also described the ecological and evolutionary principles which shape assembly and maintenance of the community. Three case studies of invertebrate-microbe symbioses were presented and they provided a framework for studying host-microbe evolution. In that context, *L. reuteri* was presented as a model to adapt this framework in a vertebrate model of gut symbiosis. *L. reuteri* can be regularly isolated from multiple vertebrate host animals such as humans, mice, rats, pigs, chickens, and turkeys (Walter, 2008) and isolates of *L. reuteri* from these animals compose phylogenetic lineages associated

with different host animal groups (Oh et al., 2010).

Different host-associated lineages appeared to show different tactics for colonization of their respective hosts. In a gnotobiotic model mouse system, only strains from the rodent-associated lineages were able to colonize the mouse host at levels which mimicked conventional populations. Using genome sequencing and genomic microarrays we identified lineage-specific differences associated with different host animals (**Chapter 2**; Frese et al., 2011). Most notably, the biggest differences seen were a result of gene loss among human-associated strains, while rodent strains maintained a large, open, pan-genome. Among the genes which were lost among the human strains were large surface proteins predicted to be involved in host epithelial adherence. The rodent-associated strains also had a number of host-specific features such as a urease gene cluster, a xylose utilization operon, and several regulatory networks which were absent in human strains. These differences illuminated the evolutionary strategies which this species has taken to adapt to life in the gastrointestinal tract of different vertebrates.

The differences observed at the genome level between human-associated strains and rodent associated strains provided consistent themes which alluded to the strategy pursued by rodent-associated strains *in vivo*. Others have repeatedly noted the formation of a biofilm by lactobacilli in the forestomach of mice where *L. reuteri* is autochthonous (native) (Roach et al., 1977, Dubos et al., 1965, Suegara et al., 1975, Lin and Savage, 1984, Fuller and Turvey, 1971). The themes of adherence and adhesion characterized much of the genomic differences between human and rodent-associated *L. reuteri* strains (**Chapter 2**; Frese et al., 2011), suggesting that biofilm formation was important for these strains *in vivo*.

To test this hypothesis, *L. reuteri* strains from each host-associated lineage were compared in their ability to colonize germ free mice. Despite the ecological

differences that we found previously (**Chapter 2**; Frese et al., 2011), these strains reached comparable population densities in the forestomach contents of mice. However, only rodent-associated were able to form biofilms *in vivo*. To identify the molecular mechanisms responsible for biofilm formation, a rodent-associated strain, *L. reuteri* 100-23, was studied (**Chapter 3**). Using expression microarrays, genome comparisons, and homology to other models of biofilm formation, the genes responsible for biofilm formation were identified. These genes included genes related to aggregation and adhesion, a key step in biofilm formation and growth, and many of the genes which were identified previously as affecting ecological success (**Chapter 2**; Frese et al., 2011) also affected biofilm formation (**Chapter 3**).

These discoveries led to the development of a potential model for biofilm formation in rodents (Figure 1). In rodents, pigs, horses and chickens, a biofilm predominated by Gram positive bacilli has been observed on the non-secretory epithelium found in the stomach of these animals (Roach et al., 1977, Dubos et al., 1965, Suegara et al., 1975, Lin and Savage, 1984, Fuller and Turvey, 1971). *L. reuteri* forms this biofilm in germ-free mice as well (**Chapter 3**). Biofilm attachment begins initially with adherence, as a result of large surface proteins found among strains isolated from hosts where this biofilm is observed (Frese et al., 2011, Heavens et al., 2011, Mackenzie et al., 2010). After adherence, LysM-domain/YG-motif proteins serve as aggregative factors and promote aggregation of these cells. The regulation of these steps may be a result of lineage-specific quorum sensing mechanisms or from sensory regulatory systems (Figure 1). Ultimately, *L. reuteri* forms a stable biofilm shortly after inoculation and likely persists in the rodent host directly as a result of this structure.

In **Chapter 2** and **Chapter 3**, gnotobiotic mouse models were used to test the ecological success and biofilm formation of *L. reuteri*. To determine if host-specificity extended to other lineages, namely the human-associated lineage, a human-associatd

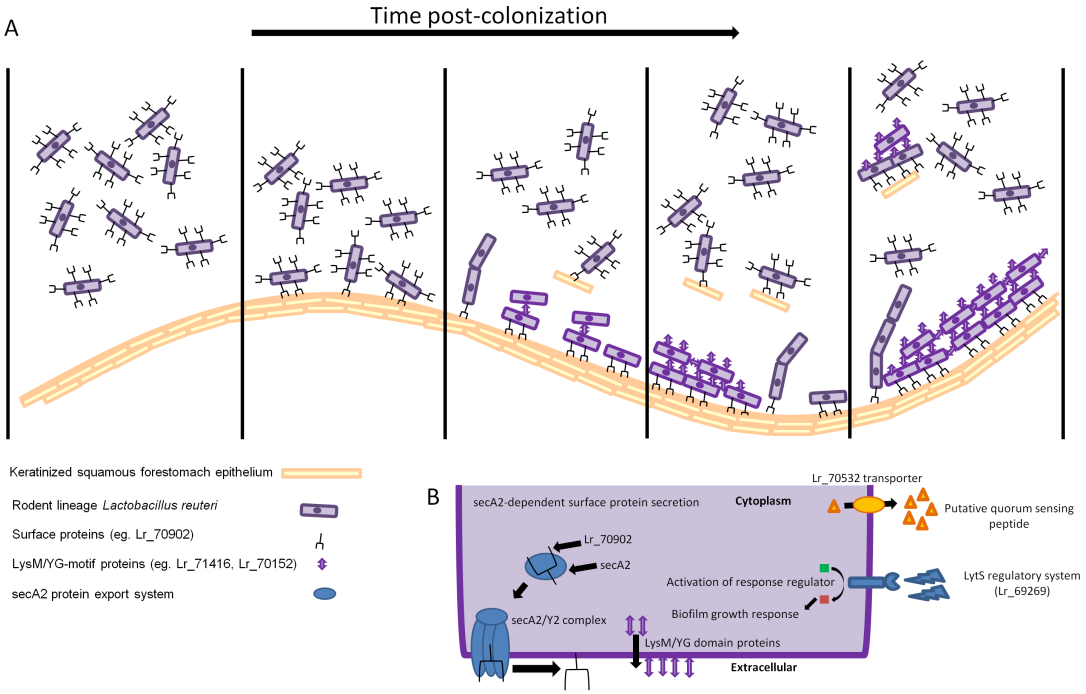

Figure 5.1(A) Hypothetical model of biofilm formation by rodent lineage *L. reuteri* after inoculation. First, cells in the lumen (first panel) adhere to the surface (second panel) using large surface proteins and begin to multiply (third panel) while they begin to shift gene expression to produce LysM protein binding epitopes (fourth panel) which could include modified teichoic acids or other cell surface modifications. Finally, a stable population on the forestomach epithelium is present, despite occasional shedding of the epithelium layer (fifth panel). The mechanisms by which large surface proteins are exported and key biofilm regulatory genes affect cell growth is also shown (B). This image is also found in **Chapter 3**.

strain *L. reuteri* ATCC PTA 6475 (MM4-1a) was compared to two other *Lactobacillus* species in a crossover human feeding trial (**Chapter 4**; Frese et al., 2012). Of these other strains, one strain, *Lactobacillus mucosae* FSL-04, was isolated continuously from a human subject and the other strain, *Lactobacillus acidophilus* DDS-1, is not associated with humans. As predicted, the autochthonous (native) *L. reuteri* and *L. mucosae* strains reached higher population densities than the allochthonous (non-native) *L. acidophilus* strain. However, none of the three species tested could be re-isolated from the subjects seven days later (**Chapter 4**; Frese et al., 2012). Together, these experiments (**Chapters 2, 3, 4**) present a comprehensive model of host-specific vertebrate gut symbiont, which has evolved to live in different host gut ecosystems.

5.2 Future Directions and Remaining Questions

First, the role of the host in this relationship will be a critical question moving forward with the mouse-*L. reuteri* model of vertebrate gut symbiosis. The ability to genetically manipulate both host and microbe provides an excellent opportunity to determine the effect of the immune system, specifically the adaptive immune system, on biofilm formation, a factor which plays an important role in oral biofilms (Williams and Gibbons, 1972). This model also provides the possibility to map traits to specific genes associated with *Lactobacillus* populations (Benson et al., 2010, Buhnik-Rosenblau et al., 2011) and inactivate them to experimentally demonstrate their relevance. Second, the biofilms formed by rodent-associated strains of *L. reuteri* (**Chapter 3**) are not likely to exist in nature as monocultures. Other species are likely to be present and it will be of interest to examine how these multi-species biofilms develop, whether the presence of other species enhances or suppresses biofilm

formation, and what effects this has on the host.

While the ecological success of rodent strains in the rodent host is clear (**Chapter 2, 3**), and rodent associated strains out-compete strains from other lineages of the species (Oh et al., 2010), it remains to be shown whether this behavior can be extended to other host-associations. In the human trial (**Chapter 4**) clear differences were seen between different species of *Lactobacillus* with different ecological and evolutionary histories, but it has not been shown that this is not simply a general feature of this species. This leaves open the question of whether the human-associated lineage's adaptations to the human gastrointestinal tract contribute to its apparent fitness *in vivo*, or whether this can be attributed to other differences between *L. reuteri* and *L. acidophilus*.

Appendix A

Supplementary Tables

Table S1. General genome features of *Lactobacillus reuteri* strains 100-23 and F275

Strain	100-23	F275
Size (bp)	2 301 232	2 025 102
Coding capacity	89%	87%
G+C content	38.7%	38.9%
Open reading frames	2375	2086
Protein-coding (CDS)	2269	1963
rRNA operons	6	6
tRNA	76	68
Pseudogenes	0	35
Two component regulatory networks		
Histidine Kinases	9	8
Response regulators	17	4
Putative transposases	63	44
Phage associated proteins	47	34
Integrase/Recombinase	4/10	24/1
Proteins with signal peptides	347	308
Proteins with transmembrane regions	544	454

Figure A.1 General Genome features of *Lactobacillus reuteri* strains 100-23 and F275

Table S2. Genes unique to Rodent (100-23) and Human (F275) Strain by functional class

Gene class	100-23	F275
Transposases/Integrases	21	25
Phage-related proteins	52	37
Restriction/Modification Systems and DNA-binding proteins	16	15
Vitamin B12 Synthesis/Glycerol/propanediol utlization/Reuterin production	1	52
Urease cluster	6	0
Cell Wall/Membrane bound proteins	25	4
Transport proteins (Ion, peptide, sugar)	26	5
Regulatory proteins	30	17
Enzymes (peptidases, hydrolases, dehydrogenases, kinases, amylases, reductases)	22	17
Metabolic proteins (replication, carbohydrate utilization, stress-protection)	28	42
Glycosyl transferases and sugar isomerases/epimerases	17	5
Hypothetical proteins	389	133
Total	**633**	**352**

Figure A.2 Unique genes in *L. reuteri* 100-23 and F275 in pair-wise comparisons

Table S3. Strains used in this study[1].

Strain Name	Analysis[2]	Host[3]	Origin of Isolation	MLSA Lineage
100-23	M, P, C, S	Rat	Australasia	III
DSM20016^T	M, P, C, S	Human (F)	Europe	II
6799jm1	M, P, C	Mouse	North America	I
bmc1	M, P	Rat	Europe	I
bmc2	M, P	Rat	Europe	I
Cr	M, P	Rat	North America	I
DSM20056	M, P	Rat	ND	I
fua3043	M, P	Rat	North America	I
l16001	M, P	Mouse	North America	I
lacto6798jm1	M, P	Mouse	North America	I
lpuph1	M, P, S	Mouse	North America	I
lpupjm1	M, P	Mouse	North America	I
ml1	M, P	Mouse	Europe	I
Oneone	M, P	Rat	North America	I
2010	M	Rat	North America	II
cf2a0	M, P	Human (F)	Europe	II
cf62a	M, P	Human (F)	Europe	II
fj1	M, P	Human (O)	Asia	II
fua3048	M	Rat	North America	II
lms11.1	M	Human (F)	North America	II
lms11.3	M, P	Human (F)	North America	II
mm41a	M, P	Human (B)	Europe	II
100.93	M, P	Mouse	Australasia	III

Figure A.3 Strains used in this study.

dsm20053	M, P	Human (F)	ND	III
l16041	M, P	Mouse	North America	III
lr4020	M, P	Mouse	North America	III
n2d	M, P, C	Rat	Europe	III
n2j	M, P	Rat	Europe	III
n4i	M, P	Rat	Europe	III
number20	M, P, C	Mouse	Australasia	III
r2lc	M, P	Rat	Europe	III
atcc55739	M, P	Pat	ND	IV
jw2015	M, P, C	Pig	North America	IV
lem83	M, P	Pig	Europe	IV
lpa1	M, P, C	Pig	North America	IV
lr85573	M, P	Human (U)	Europe	IV
tmw1137	M, P	Pig	Europe	IV
4s17	M, P	Pig	Australasia	V
1366	M, P, C	Chicken	Europe	VI
11284	M, P	Chicken	North America	VI
atcc55730	M, P, C	Human (B)	South America	VI
cf483a1	M, P	Human (F)	Europe	VI
csf8	M, P, C	Chicken	North America	VI
cf4-6g	M, P, C	Human (F)	Europe	II
dsm17938	M	Human (B)	ND	VI
jcm1081	M, P	Chicken	Asia	VI
m27u15	M, P, C	Human (B)	Africa	VI
m45r2	M, P	Human (B)	Europe	VI

m81r43	M, P	Human (B)	Asia	VI
mf14c	M, P	Human (F)	Europe	VI
mf23	M, P	Human (F)	Europe	VI
mm344a	M, P	Human (B)	Europe	VI
mm361a	M, P	Human (B)	Europe	VI
mv362a	M, P	Human (V)	Europe	VI
mv41a	M, P	Human (V)	Europe	VI
nck1556	M	Human (U)	South America	VI
nck983	M, P	Chicken	North America	VI
1048	P	Pig	Europe	IV
676	P	Pig	South America	IV
tmw1137	P	Pig	Europe	IV
ks6	P	Chicken	Europe	IV
173.5	P	Pig	Europe	IV
Lp167.67	P	Pig	North America	IV
ATCC53608	P, S	Pig	Europe	IV
10c2	P	Pig	Australasia	IV
P97	P	Pig	Europe	IV
6S15	P	Pig	Australasia	IV
32	P	Pig	South America	IV
Dbc2	P	Mouse	North America	III
Ilc4	P	Mouse	North America	III
1063	P	Pig	Europe	IV
1073	P	Pig	Europe	IV
Ad23	P	Rat	Europe	III

27.4	P	Pig	Europe	IV
JW2019	P	Pig	North America	IV
Mouse2	P	Mouse	Asia	III
Tmw11294	P	Pig	Europe	IV
Cp395	P	Pig	Europe	V
Mouse56	P	Mouse	Asia	III
Sr11	P	Human (S)	Europe	II
Me261	P	Human (F)	Asia	II
Mlc3	P, S	Mouse	North America	III
Lk94	P	Chicken	Europe	VI
Mm31a	P	Human (B)	Europe	II
3c6	P	Pig	Australasia	V
Tu160	P	Turkey	Europe	VI
20.2	P	Pig	Europe	V
L461	P	Pig	Europe	V
Rat19	P	Rat	Asia	I
L722	P	Human (F)	Europe	II
Lk20	P	Chicken	Europe	VI
11283	P	Chicken	North America	VI
1013	P	Pig	Europe	VI
T2	P	Turkey	North America	VI

[1] Derivative strains (100-23C) and mutants are not listed, as they are derived from strain 100-23.

[2]Analysis: M (CGH microarray), P (PCR confirmation of polymorphisms), C (Colonization of LF-Mice), S (genome sequencing in this study).

[3]Isolation source for human strains: feces (F), stomach (S), breast milk (B), vagina (V), oral cavity (O) and unknown (U).

Table S4. Rodent-lineage Specific Genes, as detected by MARKFIND. Genes in bold were conserved among rodent strains (Type I genes).

Gene ID	Product Name	AA Length	Enzymes	COGs	Pfams	TIGRfams
Lr_69049	Transposase and inactivated derivatives	406		COG3547	pfam01548/pfam02371	
Lr_69082	Predicted Zn peptidase	148		COG2856	pfam06114	
Lr_69088	hypothetical protein	78				
Lr_69095	hypothetical protein	70				
Lr_69096	hypothetical protein	203				
Lr_69099	hypothetical protein	78				
Lr_69101	hypothetical protein	70				
Lr_69104	hypothetical protein	67				
Lr_69105	Helix-turn-helix.	200			pfam01381	
Lr_69108	hypothetical protein	41				
Lr_69114	hypothetical protein	44				
Lr_69117	hypothetical protein	99				
Lr_69119	Deoxynucleoside kinases	233	EC:2.7.1.113	COG1428	pfam01712	
Lr_69124	phage transcriptional activator, RinA family	170				TIGR01636
Lr_69128	Uncharacterized conserved protein, COG1704	181		COG1704	pfam04011	
Lr_69131	hypothetical protein	148				
Lr_69132	hypothetical protein	113				
Lr_69134	hypothetical protein	102				
Lr_69137	hypothetical protein	57				
Lr_69138	hypothetical protein	128				
Lr_69140	nicotinamide mononucleotide transporter PnuC	247		COG3201	pfam04973	TIGR01528
Lr_69142	hypothetical protein	86				
Lr_69145	hypothetical protein	316				
Lr_69146	hypothetical protein	105				
Lr_69150	hypothetical protein	82				
Lr_69151	hypothetical protein	77				
Lr_69152	hypothetical protein	57				
Lr_69155	hypothetical protein	136				
Lr_69158	hypothetical protein	132				
Lr_69159	hypothetical protein	74				
Lr_69163	phage terminase, small subunit, putative, P27 family	152		COG3747	pfam05119	TIGR01558
Lr_69165	group I intron endonuclease	243			pfam01541	TIGR01453
Lr_69166	Phage terminase-like protein, large subunit	446		COG4626	pfam03354	
Lr_69169	phage major capsid protein, HK97 family	389			pfam05065	TIGR01554
Lr_69171	uncharacterized phage protein (possible DNA packaging)	144			pfam05135	TIGR01560
Lr_69173	hypothetical protein	160				
Lr_69174	hypothetical protein	127			pfam05657	
Lr_69175	Phage major tail protein.	234			pfam04630	
Lr_69178	phage tail tape measure protein, TP901 family, core region	1941		COG3953	pfam10145	TIGR01760
Lr_69180	Cell wall-associated hydrolases (invasion-associated proteins)	617		COG0791	pfam00877/pfam06605	
Lr_69181	hypothetical protein	78				
Lr_69186	hypothetical protein	96				
Lr_69189	hypothetical protein	241				
Lr_69263	Transposase and inactivated derivatives	406		COG3547	pfam01548/pfam02371	
Lr_69307	hypothetical protein	124				
Lr_69312	Transposase and inactivated derivatives	406		COG3547	pfam01548/pfam02371	
Lr_69327	thiol peroxidase (atypical 2-Cys peroxiredoxin) (EC 1.11.1.5) (IMGterm)	166	EC:1.11.1.15	COG2077	pfam00578	
Lr_69328	Predicted transcriptional regulators	122		COG1733	pfam01638	
Lr_69356	Transposase and inactivated derivatives	406		COG3547	pfam01548/pfam02371	
Lr_69390	Predicted membrane protein	203		COG4709	pfam08006	
Lr_69445	2[4Fe-4S] protein, putative	495		COG1139	pfam02589	TIGR00273
Lr_69466	hypothetical protein	143		COG3344		
Lr_69509	death-on-curing family protein	213		COG3654	pfam02661	TIGR01550
Lr_69510	Lyzozyme M1 (1,4-beta-N-acetylmuramidase)	309		COG3757	pfam01183/pfam08460	
Lr_69512	hypothetical protein	108				
Lr_69513	phage uncharacterized protein, XkdX family	50			pfam09693	TIGR01669
Lr_69516	hypothetical protein	134				
Lr_69526	hypothetical protein	160				
Lr_69528	hypothetical protein	118			pfam05135	

Figure A.4 Rodent-specific genes, as detected by MARKFIND

Table S5. Genes specific to the human MLSA lineage II identified by MARKFIND. Genes in bold are conserved among strains (Type I genes).

Gene ID	Locus Tag	Product Name	AA length	Enzymes	COGs	Pfams	TIGRfams
640590246	Lreu_0439	hypothetical phage protein	319				
640590496	Lreu_0663	hypothetical protein	90				
640590497	Lreu_0664	hypothetical protein	166				
640590499	Lreu_0666	phage head-tail adaptor, putative	119			pfam05521	TIGR01563
640590500	Lreu_0667	hypothetical protein	135				
640590501	Lreu_0668	HNH endonuclease	174			pfam01844	
640590502	Lreu_0669	phage terminase, small subunit, putative, P27 family	154		COG3747	pfam05119	TIGR01558
640590503	Lreu_0670	phage Terminase	569		COG4626	pfam03354	
640590505	Lreu_0672	phage major capsid protein, HK97 family	506		COG3740	pfam04586/pfam05065	TIGR01543/TIGR01554
640590506	Lreu_0673	phage transcriptional regulator, RinA family	140				TIGR01636
640590642	Lreu_0810	putative transcriptional regulator, XRE family	90		COG3655		
640590643	Lreu_0811	RecT protein	309		COG3723	pfam03837	TIGR00616
640590647	Lreu_0815	hypothetical protein	123				
640590648	Lreu_0816	hypothetical protein	92				
640590653	Lreu_0821	hypothetical protein	113				
640590656	Lreu_0824	hypothetical protein	103				
640590663	Lreu_0831	ParB domain protein nuclease	259			pfam02195	
640590667	Lreu_0835	phage putative head morphogenesis protein	316			pfam04233	TIGR01641
640590670	Lreu_0838	hypothetical phage protein	358				
640590674	Lreu_0842	hypothetical phage protein	136				
640590678	Lreu_0846	Lytic transglycosylase, catalytic	1216		COG5412	pfam01464	TIGR02675
640590679	Lreu_0847	protein of unknown function DUF1306	295			pfam05709	TIGR01633
640590684	Lreu_0852	lipolytic protein, G-D-S-L family	381			pfam00657	
640590685	Lreu_0853	hypothetical protein	369				
640590686	Lreu_0854	hypothetical protein	134				
640590885	**Lreu_1043**	**hypothetical protein**	**141**			**pfam02556**	
640590887	**Lreu_1045**	**hypothetical protein**	**289**				
640590890	**Lreu_1048**	**Superfamily I DNA and RNA helicase-like protein**	**363**		**COG0210**	**pfam00580/pfam01443**	
640590937	Lreu_1098	hypothetical protein	148				
640590938	Lreu_1099	hypothetical protein	134				
640590940	Lreu_1101	lipolytic protein, G-D-S-L family	381			pfam00657	
640590945	Lreu_1106	protein of unknown function DUF1306	281			pfam05709	
640590974	Lreu_1135	hypothetical protein	123				
640590977	Lreu_1138	hypothetical phage protein	274				
640590981	Lreu_1142	phage antirepressor protein	267		COG3645	pfam02498/pfam03374	
640590983	Lreu_1144	helix-turn-helix domain protein	145			pfam01381	
640590985	Lreu_1146	hypothetical phage protein	169			pfam07553	
640591217	**Lreu_1369**	**galactofuranosyltransferase**	**338**				
640591220	**Lreu_1372**	**lipopolysaccharide biosynthesis protein**	**209**		**COG3944**	**pfam02706**	
640591221	**Lreu_1373**	**oligosaccharide repeat unit polymerase Wzy**	**398**				
640591279	**Lreu_1433**	**restriction modification system DNA specificity dmn**	**372**	**EC:3.1.21.3**	**COG0732**	**pfam01420**	
640591280	**Lreu_1434**	**phage integrase family protein**	**322**		**COG4974**	**pfam00589**	
640591281	**Lreu_1435**	**restriction modification system DNA specificity dmn**	**340**	**EC:3.1.21.3**	**COG0732**	**pfam01420**	
640591283	**Lreu_1437**	**hypothetical protein**	**402**				
640591441	Lreu_1595	Transposase and inactivated derivatives-like protein	156		COG3464		
640591453	Lreu_1607	phage integrase family protein	402		COG4974	pfam00589	
640591675	**Lreu_1830**	**histidyl-tRNA synthetase (EC 6.1.1.21) (IMGterm)**	**433**	**EC:6.1.1.21**	**COG0124**	**pfam00587/pfam03129**	**TIGR00442**
640591676	**Lreu_1831**	**hypothetical protein**	**174**				
640591677	**Lreu_1832**	**histidine decarboxylase, pyruvoyl type**	**311**	**EC:4.1.1.22**		**pfam02329**	**TIGR00541**
640725732	**Lreu_1052**	**reverse transcriptase**	**193**				
640725737	Lreu_1338	transposase	206		COG2801	pfam00665	
640725741	**Lreu_1347**	**glucosyltransferase, dextransucrase**	**127**				

Figure A.5 Genes specific to the human MLSA lineage II

Table S6. Genome characteristics of *Lactobacillus reuteri* strains

	100-23	mlc3	F275	MM4-1	MM2-3	lpuph1	CF48-3a	ATCC 55730	ATCC 53608
Origin (Host)	rat	mouse	human	human	human	mouse	human	human	pig
MLSA Lineage	III	III	II	II	II	I	VI	VI	IV
Contigs	2	126	1	131	167	127	244	339	142
Genome Size, bp	2 301 232	2 018 627	2 025 102	1 935 151	1 943 466	2 116 617	2 032 595	2 036 411	1969869
G + C, %	39	39	39	39	39	38	39	39	39
Number of protein-coding genes, predicted	2269	1859	1963	2045	2045	1902	2164	2235	1855

Figure A.6 Characteristics of the nine *Lactobacillus reuteri* genomes included in this study

Table S7. Average nucleotide identity (ANI) in percent between *L. reuteri* genomes and *L. vaginalis* ATCC49540, based on 169 conserved orthologs. Strain colors reflect host MLSA lineage (blue, II (human); green III (rodent); light green, I (rodent); red, IV (pig); yellow, VI (poultry/human)).

Strain	100-23	MLC3	ATCC53608	MM4	mm23	F275	lpuph	CF483a	ATCC55730	*L. vaginalis*
100-23	100.00									
MLC3	97.07	100.00								
ATCC53608	96.55	96.80	100.00							
MM4	96.61	95.93	96.78	100.00						
MM23	96.51	96.34	97.06	99.54	100.00					
F275	96.13	96.15	96.72	99.63	99.98	100.00				
Lpuph	96.11	95.93	96.47	96.89	96.91	96.81	100.00			
CF483a	95.83	95.75	96.09	96.93	97.18	97.22	96.92	100.00		
ATCC55730	95.81	95.69	95.78	96.70	97.04	97.10	96.73	99.99	100.00	
L. vaginalis	77.87	77.56	77.36	77.61	78.36	78.07	77.64	77.55	77.27	100.00

Figure A.7 Average Nucleotide Identity (ANI) between genomes

Table S8. Genes unique among rodent strains 100-23, mlc3, lpuph

Gene class	100-23	mlc3	lpuph
Transposases/Integrases	13	7	8
Phage-related proteins	111	28	47
Restriction/Modification Systems and DNA-binding proteins	28	2	19
Vitamin B12 Synthesis/Glycerol/Reuterin production	0	55	0
Cell Wall/Membrane bound proteins	20	10	8
Transport proteins (Ion, peptide, sugar)	15	12	8
Regulatory proteins	23	6	15
Ribosomal proteins (LSU/SSU)	5	1	0
Metabolic Enzymes (peptidases, hydrolases, dehydrogenases, kinases, amylases, reductases)	31	37	31
Glycosyl transferases and sugar isomerases/epimerases	15	10	12
Hypothetical proteins	267	67	161
Total	**528**	**235**	**309**

Figure A.8 Unique genes in the genomes of the rodent strains 100-23, mlc3, and lpuph.

Table S9. PCR primers used to confirm rodent-specific gene acquisitions.

Application	Target	Primer Name	Primer sequence 5' to 3'	Annealing Temperature (°C)
Amplification of Target Gene	*Lr_70131*	70131for	GTTAAGGCTGTTGATGTAACAAC	64
		70131rev	ATAGGTTGTCCGTAAAGTGTAAC	
	Lr_70581	70581for	TAGCTCAAGTGCTCAGACAAG	62
		70581rev	CATTAGTAGTAGACTTAGCAAC	
	secA2	secA2for	CATTCGGGAAGCGGACTACC	60
		secA2rev	AATTGCTCTTTATCACGTTGG	
	Lr_70697	70697for	CCACGGCTATGCAAGCTG	62
		70697rev	CGTTAGCAGCACCGACGC	
	Lr_69916	69916for	GTACATTGACCTTAAATATCGCA	64
		69916rev	CACTAATCGGTCCCGAAACAAC	
	16S rRNA	8F	AGAGTTTGATCCTGGCTCAG	60
		1391R	GACGGGCGGTGWGTRCA	
	ureC	ureCF	CATTGGTAAGGGTGGTAACC	60
		ureCR	CAAACCATTGTCATGTAGAAC	
	pduC	pduCF	CCTGAAGTAAAYCGCATCTT	55
		pduCR	GAAACYATTTCAGTTTATGG	
Amplification of Flanking genes	*Lr_70131*	70131flankfor	GATGATACCACCGGACCAC	60
		70131flankrev	GATAGGCCACTTCCTTGAC	
	Lr_70581	70581flankfor	GATCCTCGGACTGGATACC	60
		70581flankrev	TTACACCACGCGTAGCTAAC	
	Lr_70697	70697flankfor	GATTAGCAGAATTAGCGCCG	60
		70697flankrev	GGTAGCAGCTAYYTGTTGG	
	secA2	secA2flankfor	CACCAGAAATGGTAGAACGAA	62
		secA2flankrev	GTACCTATGTGCCATTCTTCAAC	

Figure A.9 PCR primers used in this study to confirm host-specificity of genes

Genes Up-regulated in biofilm growth

Gene ID	Annotation	Fold Change
2500069000	Group II intron, maturase-specific domain.	2.233550695
2500069049	Transposase and inactivated derivatives	3.439791214
2500069072	tyrosine recombinase XerD subunit (IMGterm)	2.121066716
2500069106	hypothetical protein	2.344038068
2500069134	hypothetical protein	2.041693069
2500069136	hypothetical protein	2.495763102
2500069151	hypothetical protein	2.142309616
2500069206	Uncharacterized protein conserved in bacteria, COG4698	2.162212025
2500069225	hypothetical protein	2.019642581
2500069272	Putative effector of murein hydrolase	2.453907725
2500069274	fructokinase (EC 2.7.1.4) (IMGterm)	2.011600755
2500069290	recombination protein U	2.065346728
2500069304	degV family protein	2.628360198
2500069343	Uncharacterised protein family (UPF0236).	2.227181295
2500069346	aspartate racemase (EC:5.1.1.13)	2.271137417
2500069361	amino acid ABC transporter membrane protein, PAAT family (TC 3.A.1.3.-) (IMGterm)	6.766041378
2500069362	amino acid ABC transporter ATP-binding protein, PAAT family (TC 3.A.1.3.-) (IMGterm)	3.149646681
2500069367	deoxyuridine 5'-triphosphate nucleotidohydrolase (EC 3.6.1.23) (IMGterm)	2.239422904
2500069389	transcriptional regulator, PadR family (IMGterm)	2.371506791
2500069392	alcohol dehydrogenase AdhE (EC 1.1.1.1) / acetaldehyde dehydrogenase (EC 1.2.1.10) (IMGterm)	2.695307733
2500069427	amino acid/polyamine/organocation transporter, APC superfamily (TC 2.A.3) (IMGterm)	2.236546882
2500069428	hypothetical protein	2.590348126
2500069455	Excinuclease ABC subunit A (IMGterm)	2.046456499
2500069458	conserved hypothetical protein, cofD-related TIGR01826	2.517509112
2500069512	hypothetical protein	2.064027224
2500069520	tape measure domain	2.015570156
2500069548	hypothetical protein	2.003868726
2500069555	hypothetical protein	2.517099069
2500069556	hypothetical protein	2.618540648
2500069557	DnaD and phage-associated domain	2.463010879
2500069625	Predicted metal-binding, possibly nucleic acid-binding protein	2.223966413
2500069686	hypothetical protein	2.9281856
2500069687	(tRNA)	3.243750824
2500069692	ribokinase (EC:2.7.1.15)	2.052567974
2500069713	Nucleotidyltransferase/DNA polymerase involved in DNA repair	2.313106994
2500069717	NADPH-dependent glutamate synthase beta chain and related oxidoreductases	2.092358412
2500069751	hypothetical protein	2.056514152
2500069752	phage protein, HK97 gp10 family	3.046553341
2500069753	hypothetical protein	2.106311776
2500069754	hypothetical protein	2.075702954
2500069755	hypothetical protein	2.093894856
2500069759	phage portal protein, SPP1 family	2.005969958
2500069769	hypothetical protein	2.143585556
2500069770	hypothetical protein	3.345922432
2500069781	hypothetical protein	2.144150242

2500069782	hypothetical protein	2.631960796
2500069792	Site-specific recombinase XerD	2.409279884
2500069816	hypothetical protein	4.150203273
2500069848	Muramidase (flagellum-specific)	2.40971579
2500069862	putative glutamate--cysteine ligase/putative amino acid ligase	2.000457529
2500069886	hypothetical protein	2.005486146
2500069890	hypothetical protein	3.780340014
2500069920	hypothetical protein	3.113846873
2500069933	drug resistance transporter, EmrB/QacA subfamily	2.184779981
2500069998	DNA-directed RNA polymerase subunit beta' (EC 2.7.7.6) (IMGterm)	2.076838452
2500070078	LPXTG-motif cell wall anchor domain/Gram-positive signal peptide, YSIRK family	2.054857002
2500070079	NADPH-dependent FMN reductase.	3.166785169
2500070081	cystathionine beta-synthase (acetylserine-dependent) (EC 4.2.1.-) (IMGterm)	4.505302574
2500070146	hypothetical protein	2.088204495
2500070151	hypothetical protein	2.166227853
2500070153	Arabinose efflux permease	3.509980462
2500070215	Transposase and inactivated derivatives, IS30 family	2.053621066
2500070220	Thioredoxin reductase (EC:1.8.1.9)	2.428150762
2500070221	Inosine-uridine nucleoside N-ribohydrolase (EC:3.2.2.1)	2.229238539
2500070222	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)	3.610565963
2500070344	hypothetical protein	3.421777614
2500070347	glycerol 2-dehydrogenase (NAD+) (EC 1.1.1.6) (IMGterm)	2.057305439
2500070423	Flavodoxins	2.085340543
2500070566	hypothetical protein	2.045310675
2500070635	(rRNA 16S)	2.02023343
2500070804	(tRNA)	2.407043142
2500070899	Transposase and inactivated derivatives	2.111338651
2500071012	hypothetical protein	2.288301262
2500071062	Uncharacterized phage-associated protein, COG3600	2.083680385
2500071063	CotH protein.	2.157946354
2500071103	hypothetical protein	2.019168068
2500071104	hypothetical protein	2.287765213
2500071139	Uncharacterized conserved protein, COG0398	2.213419646
2500071144	Transposase and inactivated derivatives	2.823093674
2500071212	aspartate kinase (EC 2.7.2.4) (IMGterm)	2.171182504
2500071213	diaminopimelate epimerase (EC 5.1.1.7) (IMGterm)	2.003432635
2500071277	glutamate racemase (EC 5.1.1.3) (IMGterm)	2.00889617
2500071306	amino acid ABC transporter membrane protein, PAAT family (TC 3.A.1.3.-) (IMGterm)	2.036054247
2500071417	Cof subfamily of IIB subfamily of haloacid dehalogenase superfamily/HAD-superfamily hydrolase, subfamily IIB	2.063343475
2500069271*	Putative effector of murein hydrolase LrgA	2.09127241
2500069360*	cystathionine gamma-lyase (EC 4.4.1.1) (IMGterm)	3.529211766
2500069719*	LysM domain.	5.517623437
2500069721*	LysM domain.	2.90212166
2500070152*	LysM domain.	2.467015876
2500070770*	LPXTG-motif cell wall anchor domain/Gram-positive signal peptide, YSIRK family	5.766372356
2500071416*	LysM domain.	5.425454501

Genes downregulated in biofilm growth		
Gene ID	Annotation	Log(FC)
2500069046	amino acid ABC transporter substrate-binding protein, PAAT family (TC 3.A.1.3.-) / amino acid ABC transporter membrane protein, PAAT family (TC 3.A.1.3.-) (IMGterm)	-2.414241503
2500069051	argininosuccinate synthase (EC 6.3.4.5) (IMGterm)	-2.115487264
2500069836	putative glutamate--cysteine ligase/putative amino acid ligase/bacterial surface protein 26-residue repeat	-2.107876
2500070032	amino acid ABC transporter membrane protein 1, PAAT family (TC 3.A.1.3.-) (IMGterm)	-2.264872756
2500070095	hypothetical protein	-2.131575269
2500070131	LPXTG-motif cell wall anchor domain/Gram-positive signal peptide, YSIRK family	-3.325589274
2500070134	LPXTG-motif cell wall anchor domain/Gram-positive signal peptide, YSIRK family	-3.286563819
2500070588	nucleoside transporter	-2.207984432
2500070589	Inosine-uridine nucleoside N-ribohydrolase (EC:3.2.-)	-2.067059619
2500070591	16S rRNA m(7)G-527 methyltransferase (EC 2.1.1.170) (IMGterm)	-2.016145838
2500070593	chromosome segregation ATPase (IMGterm)	-2.284914212
2500070594	chromosome segregation DNA-binding protein (IMGterm)	-2.298894086
2500070595	Uncharacterized protein conserved in bacteria, COG4481	-2.054589248
2500070598	deoxyribose-phosphate aldolase (EC:4.1.2.4)	-3.832431485
2500070599	phosphopentomutase (EC:5.4.2.7)	-2.962756376
2500070656	Predicted esterase of the alpha/beta hydrolase fold	-2.072453885
2500070664	DNA or RNA helicases of superfamily II	-2.107452333
2500070666	Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide dehydrogenase (E3) component, and related enzymes	-2.057311143
2500070670	heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase	-3.631183908
2500070671	phosphopentomutase (EC:5.4.2.7)	-2.419904309
2500070672	hypothetical protein	-2.163639288
2500070744	ABC-type Mn2+/Zn2+ transport systems, permease components	-2.247841014
2500070749	Cof subfamily of IIB subfamily of haloacid dehalogenase superfamily/HAD-superfamily hydrolase, subfamily IIB	-2.395306047
2500070750	HD superfamily phosphohydrolases	-2.304701647
2500070751	Lipoate-protein ligase A	-2.35109038
2500070812	(tRNA)	-2.039930499
2500070823	(tRNA)	-2.157065522
2500070924	Glycosyltransferases involved in cell wall biogenesis (EC:2.4.1.-)	-2.037273963
2500070925	hypothetical protein	-2.342241769
2500070955	Protein involved in cell division	-2.165814976
2500070998	molybdopterin-guanine dinucleotide biosynthesis protein MobB	-2.002330331
2500071088	hypothetical protein	-2.216709934
2500071294	DNA mismatch repair protein MutL (IMGterm)	-2.155824893
2500071320	Negative regulator of septation ring formation	-2.380888403
2500071333	Uncharacterized conserved protein, COG5506	-2.025011358
2500071394	Predicted permease, DMT superfamily	-2.043341019
2500071411	glucose-6-phosphate isomerase (EC 5.3.1.9) (IMGterm)	-2.204299107

Supplementary Table 2. RTPCR Primers used in this study.

Gene	F/R	Oligonucleotide sequence (5' -> 3')	Role	Relevant Citations
Glucose-3-phosphate dehydrogenase	F	CGGATTCACGAACTTAACACAA	House-keeping reference gene	--
	R	CCTTACCATCAACAACAATAC		
Lr_69360 cystathionine γ-lyase	F	CGGAACTTTGGAGCAATGAT	ROS resistance	Lo et al 2009
	R	GGTCCTCAATTCCCACTGAA		
Lr_70902 LPXTG-surface protein	F	ACACGGTCCAAGCTGAAGAT	Adherence	Frese et al 2011
	R	AAATTCCACCGTTTCCATCA		
Lr_70580 lsp surface protein	F	ATACACCAGATTGGGCTTCG	Adherence	Walter et al 2005
	R	CGCCCCAAGTTACAGTTTGT		
Lr_70892 secA2 transport protein	F	GCAACCTCACCTTTTGTGGT	Export of large surface proteins	Frese et al 2011
	R	CCCGTGTTTCGCTAACAAAT		
Lr_70114 urease α-subunit	F	AACTAACCCATATTGTAAGAACA	pH resistance	Frese et al 2011
	R	AGAAGTCATCATACTAAGGGCA		
Lr_71416 LysM2	F	TGCTACTCTTGGTGCAGTGG	Unknown	Buist et al 2008; Goh et al 2010
	R	TTTCACCATCGCTCTTGATG		
Lr_70152 LysM3	F	AATAACATGGCGATGCAACA	Unknown	Buist et al 2008; Goh et al 2010
	R	CGCTTCCTTCACGGTTTAAG		
Lr_69271 LrgA antiholin	F	GCCCATTTTAGTTCAAAT	Autolysis	Brunskill and Bayles 1996
	R	TGGGACAAACAGAAAAGCAA		
Lr_69269 LytS two-component regulator	F	CAAATAACCAGTTGCATAACA	Autolysis	Brunskill and Bayles 1996
	R	ACCCCTGCCACTAACCCTAC		
Lr_70532 two-component regulator	F	TTTTCGCCTGCTAAGGGATA	Rodent-lineage specific	Frese et al 2011
	R	CGAGGGAATAAACCGATCAA		
Lr_70430 two-component regulator	F	TCGAATAATTGGGATCGCTTA	Rodent-lineage specific	Frese et al 2011
	R	TTTTGAGCCGTTGCTAATCC		

Table S4. Bioinformatic analysis of the Lr_70532 putative Bacteriocin cluster.

Gene	Predicted function	Top match organism	Match %	Coverage %
Lr_70529	histidine kinase	*L. salivarius* GJ-24	38	49
Lr_70530	LytTr Response regulator	*L. salivarius* GJ-24	34	88
Lr_70531	bacteriocin-type signal sequence	No match	--	--
Lr_70532	ABC-type bacteriocin transporter	*L. salivarius* ACS-116-V-Col5a	55	98
Lr_70533	ABC- transporter-related protein (AbpT, Lactacin F)	*L. crispatus* 135-2-CHN	56	98

Figure A.14

Appendix B

Supplementary Figures

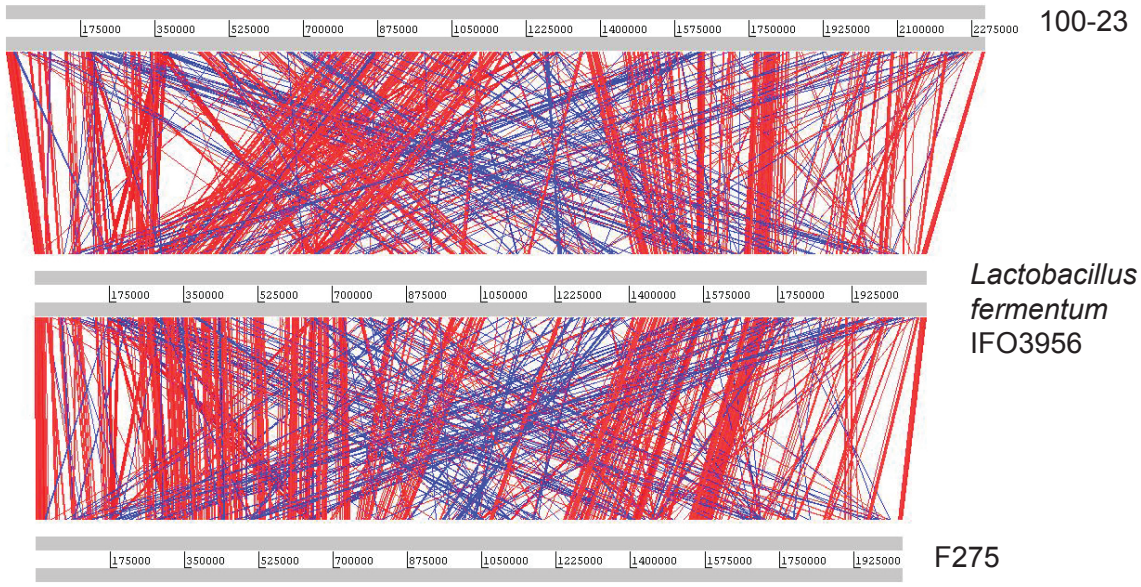

Figure B.1 Genome comparisons between *L. reuteri* strains and *Lactobacillus fermentum*. Comparisons of the genomes of *Lactobacillus reuteri* 100-23 (top) and F275 (bottom) with that of *Lactobacillus fermentum* IFO3956 (center). The BLASTN comparison using the Artemis Comparison Tool revealed an inversion in 100-23 when compared to the other two genomes.

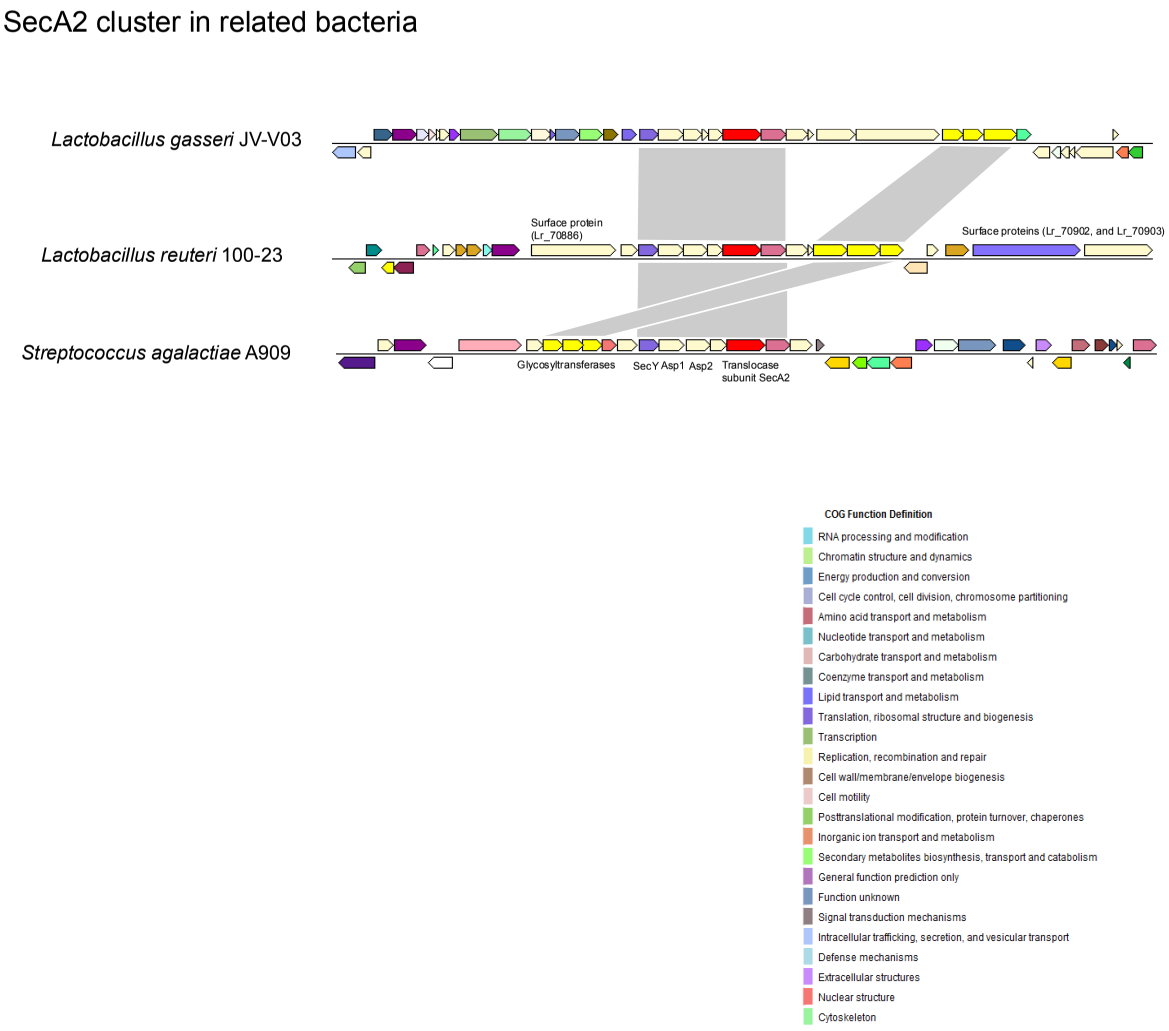

Figure B.2 Accessory SecA2 cluster in *L. reuteri* and related bacteria. SecA2 cluster in *L. reuteri* 100-23, *Lactobacillus gasseri* JV-V03, and *Streptococcus agalactiae* A909. Conserved regions are labeled with grey boxes.

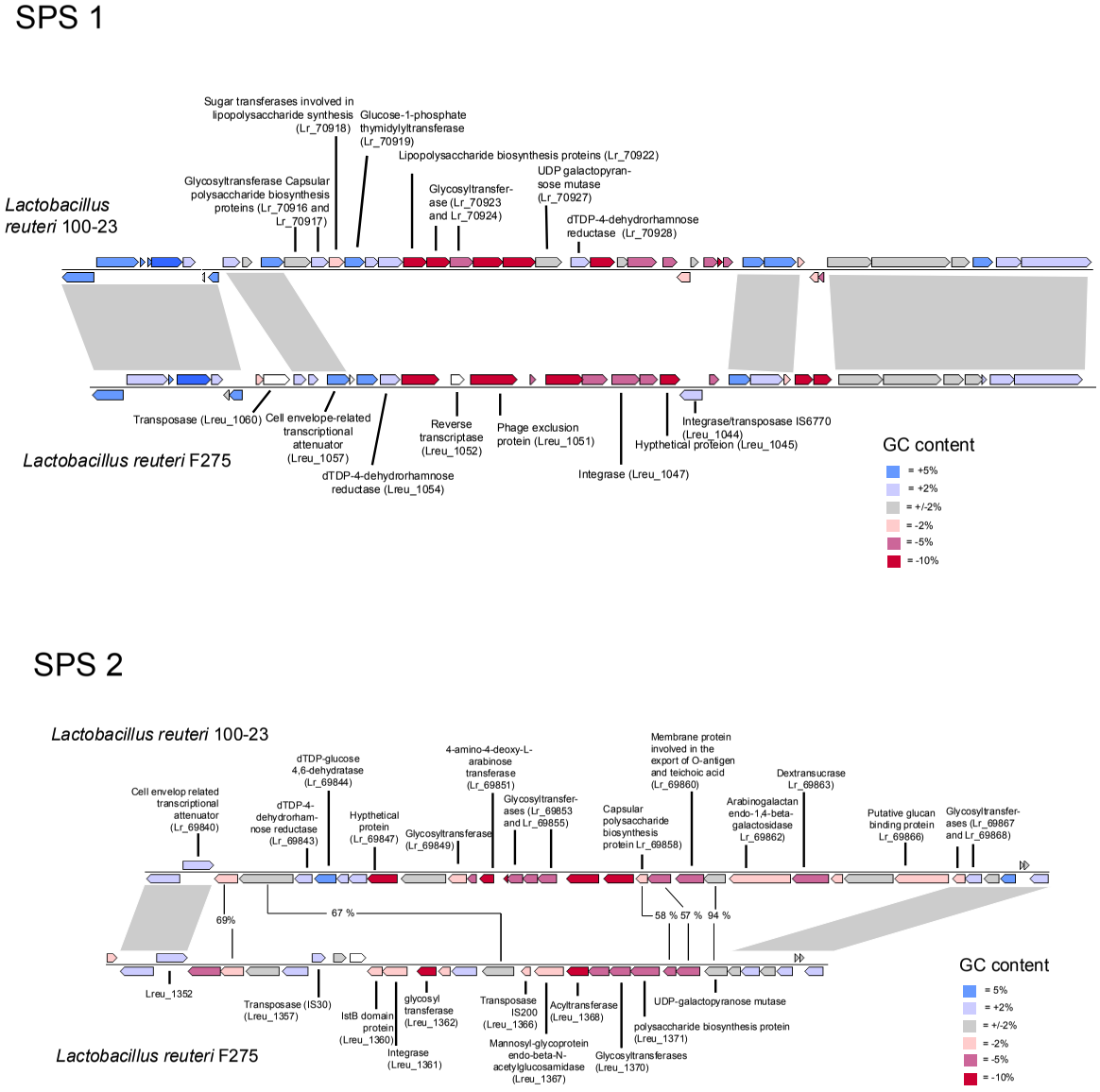

Figure B.3 Gene maps of SPS clusters in *L. reuteri* 100-23 and F275. Genomic regions that encode for proteins involved in surface polysaccharide synthesis in *L. reuteri* 100-23 and F275. Flanking regions that are conserved in both genomes are indicated by boxes

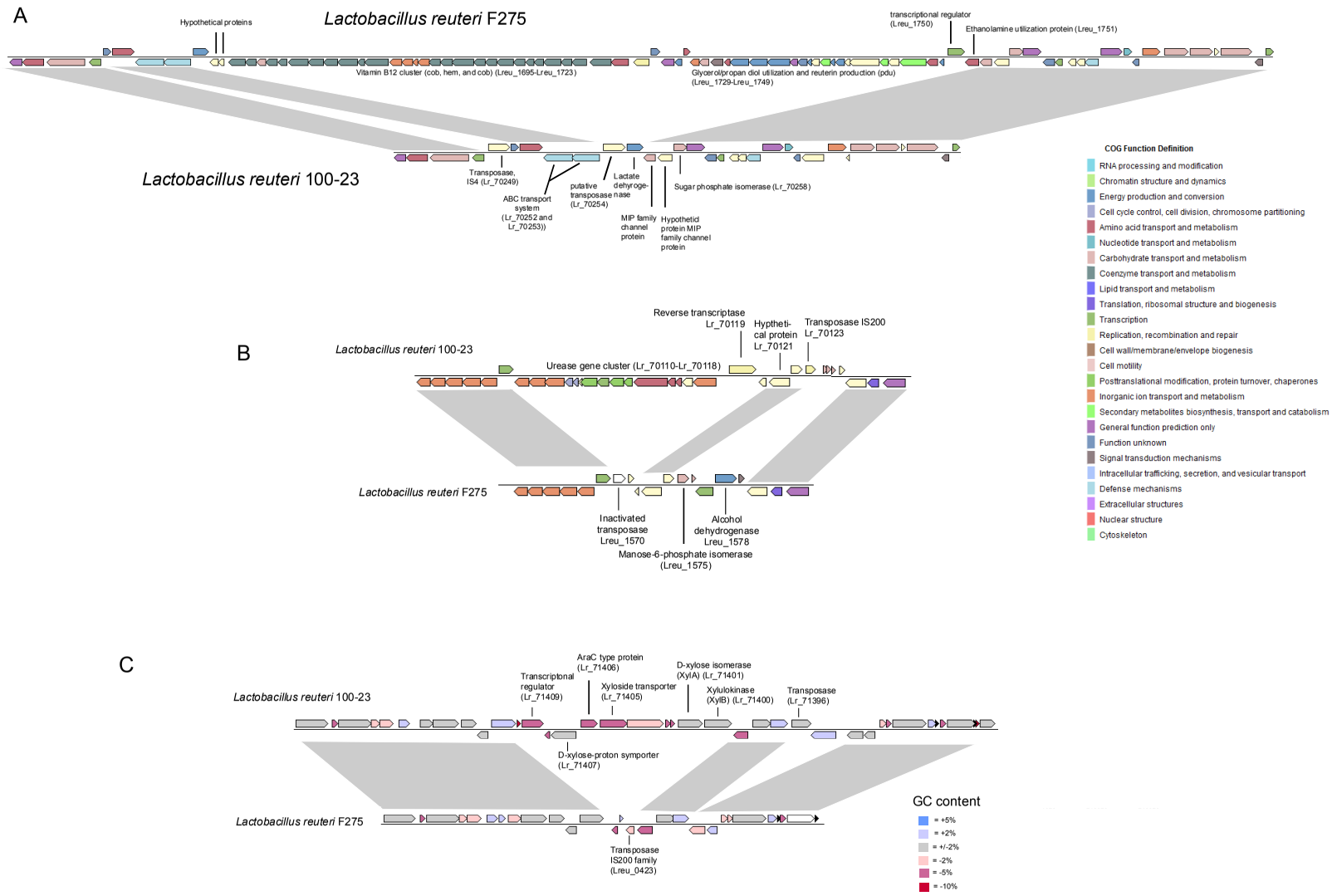

Figure B.4 Comparison of genomic locations that contain host-specific genes. (A) Genome region with the pdu-cbi-cob-hem cluster in F275 and the same genomic region in 100-23. (B) Genome schematic showing the urease cluster in strain 100-23, and the same loci in F275. (C) Genomic region containing the xylose operon in 100-23 and the same loci, without xylose operon, in F275

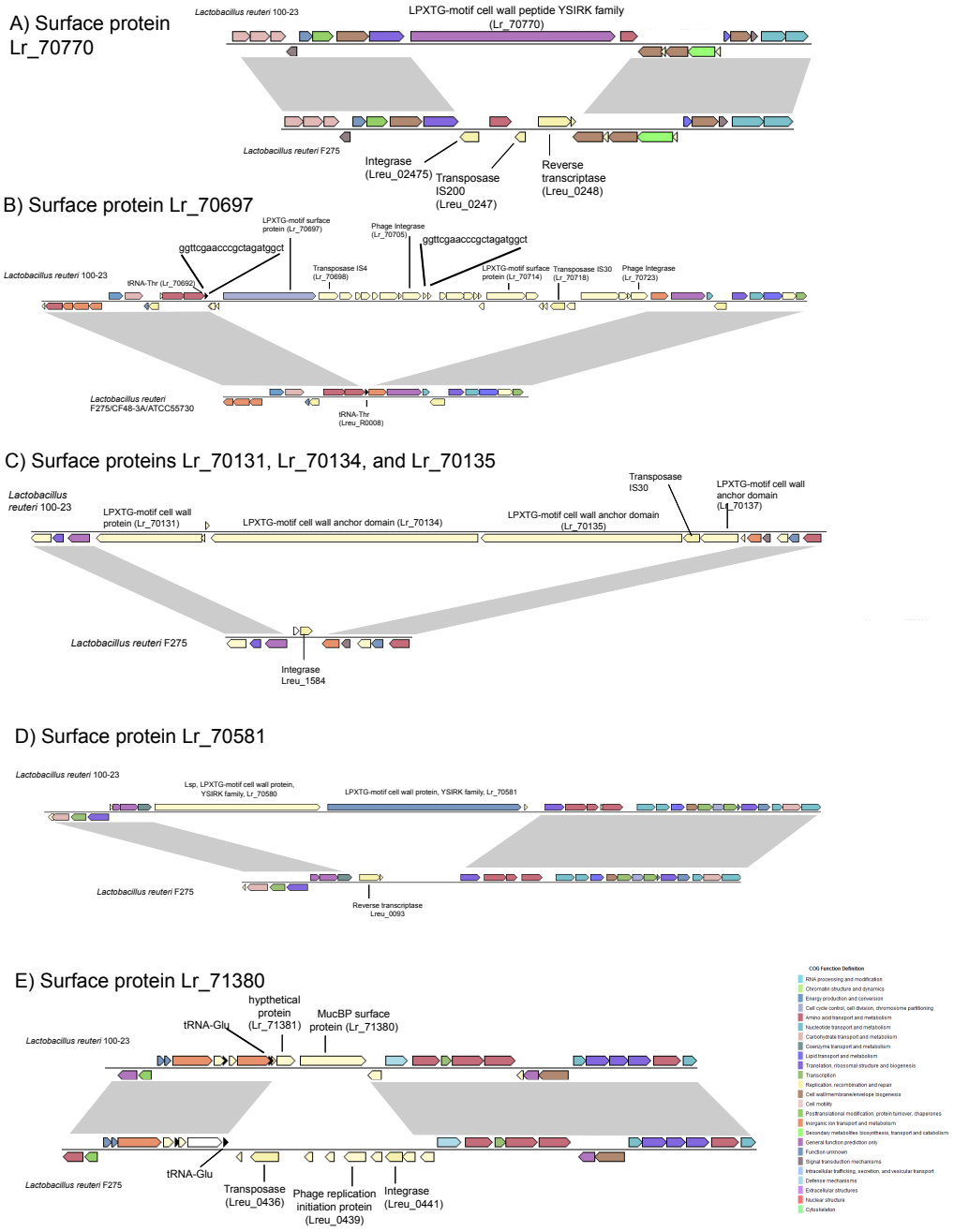

Figure B.5 Comparison of the genomic locations that contain large surface proteins in L. *reuteri* 100-23. Comparison of the sites in strains 100-23 and F275 that contain rodent-specific large surface proteins. (A) Lr_70770, (B) Lr_70697, (C) Lr_70131, Lr_70134, and Lr_70135, (D) Lr_70581, and (E) Lr_71380.

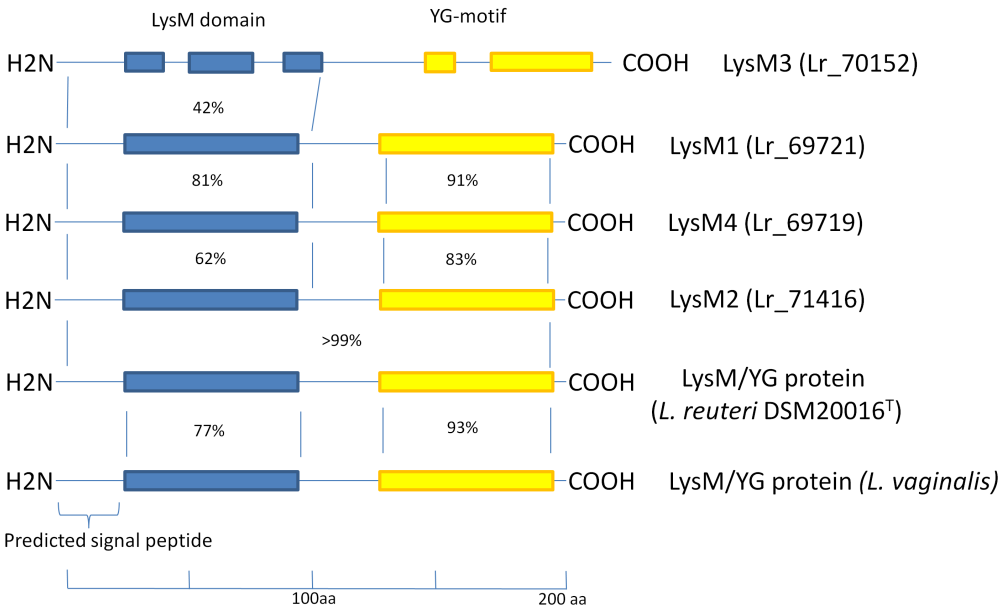

Figure B.6 LysM domain/YG-motif proteins in L. *reuteri* 100-23 compared to those found in *L. reuteri* DSM20016^T and *L. vaginalis*

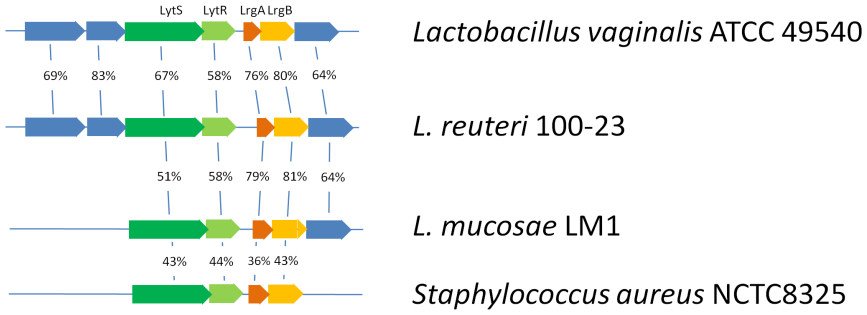

Figure B.7 LytSR gene cluster in *L. reuteri* 100-23 and related organisms.

Bibliography

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

MJ. Albert, VI. Mathan, and SJ. Baker. Vitamin b12 synthesis by human small intestinal bacteria. *Nature*, 283:781--82, 1980.

R. Alexander. The energetics of coprophagy: a theoretical analysis. *Journal of Zoology*, 230(4):629--637, August 1993.

R Axelrod and WD. Hamilton. The evolution of cooperation. *Science*, 211:1390--1396, 1981.

L. Axelsson and S. Lindgren. Characterization and dna homology of lactobacillus strains isolated from pig intestine. *J Appl Bacteriol*, 62:433--440, 1987.

LT Axelsson, TC Chung, WJ Dobrogosz, and SE Lindgren. Production of a broad spectrum antimicrobial substance by lactobacillus reuteri. *Microbial Ecology in Health and Disease*, 2(2):131--136, 1989.

M. A. Azcarate-Peril, E. Altermann, Y. J. Goh, R. Tallon, R. B. Sanozky-Dawes, E. A. Pfeiler, S. O'Flaherty, B. L. Buck, A. Dobson, T. Duong, M. J. Miller, R. Barrangou, and T. R. Klaenhammer. 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, 2008.

F. Bäckhead, H. Ding, T. Wang, LV. Hooper, GY. Kohm, A. Nagy, CF. Semenkovich, and JI. Gordon. The gut microbiota as an environmental factor that regulates fat storage. *PNAS*, 101(44):15718--15723, 2004.

P. Baumann, L. Baumann, CY. Lai, and D. Rouhbakhsh. Genetics, physiology, and evolutionary relationships of the genus buchnera: Intracellular symbionts of aphids. *Annu Rev. Microbiol.*, 49:55--94, 1995.

KW. Bayles. The biological role of death and lysis in biofilm development. *Nature Reviews Microbiology*, 5(9):721--726, 2007.

Dal Bello, Walter F., W. J., Hammes, and C. Hertel. Increased complexity of the species composition of lactic acid bacteria in human feces revealed by alternative incubation condition. *Microb Ecol*, 45:455--463, 2003.

BA. Bensing and Pm Sullam. An accessory sec locus of streptococcus gordonii is required for export of the surface protein gspb and for normal levels of binding to human platelets. *Molecular Microbiology*, 44(4):1081--1094, 2002.

BA. Bensing and Pm Sullam. Characterization of streptococcus gordonii seca2 as a paralogue of seca. *Journal Bacteriology*, 191(11):3482--3491, 2009.

AK. Benson, SA. Kelly, R. Legge, F. Ma, SJ. Low, J. Kim, M. Zhang, PL. Oh, D. Nehrenberg, K. Hua, SD. Kachman, EN. Moriyama, J. Walter, DA. Peterson, and D. Pomp. Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *PNAS*, 107(44): 18933--18938, 2010.

EM Berberov, Y Zhou, DH. Francis, MA Scott, Kachman SD, and Moxley RA. Relative importance of heat-labile enterotoxin in the causation of severe diarrheal disease in the gnotobiotic piglet model by a strain of enterotoxigenic escherichia coli that produces multiple enterotoxins. *Infection and Immunity*, 72(7):124--130, 2004.

SS Berry. *The Cephalopoda of the Hawaiian Islands*. Bulletin of the Bureau of Fisheries 32, 1912.

M Bright and S Bulgheresi. A complex journey: transmission of microbial symbionts. *Nat Rev Microbiol*, 8:218--230, 2010.

M.A. Brockhurst, M.E. Hochberg, T. Bell, and A. Buckling. Character displacement promotes cooperation in bacterial biofilms. *Current Biology*, 16(20):2030--2034, 2006.

SPJ Brooks, M McAllister, M Sandoz, and ML Kalmokoff. Culture-independent phylogenetic analysis of the faecal flora of the rat. *Canadian Journal of Microbiology*, 49:589--601, 2003.

EW Brunskill and Kw Bayles. Identification of lytsr-regulated genes from staphylococcus aureus. *J. Bacteriol*, 178(19):5810--5812, 1996.

K Buhnik-Rosenblau, Y Danin-Poleg, and Y Kashi. Predominant effect of host genetics on levels of lactobacillus johnsonii bacteria in the mouse gut. *Appl. Environ. Microbiol77*, 77:6531--6538, September 2011.

J.H. Campbell, C.M. Foster, T. Vishnivetskaya, A.G. Campbell, Z.K. Yang, A. Wymore, A.V. Palumbo, E.J. Chesler, and M. Podar. Host genetic and environmental effects on mouse intestinal microbiota. *The ISME Journal*, 6:2033--2044, 2012.

TJ Carver, KM Rutherford, M Berriman, MA Rajandream, and BG Barrell. et al. (2005) act: the artemis comparison tool. *Bioinformatics*, 21(3422-3423):2196--2203, 2006.

I.A. Casas and W.J. Dobrogosz. Validation of the probiotic concept: Lactobacillus reuteri confers broad-spectrum protection against disease in humans and animals. *Microbial ecology in health and disease*, 12(4):247--285, 2000.

J. Cavender-Bares, K. H. Kozak, P. V. Fine, and S. W. Kembel. The merging of community ecology and phylogenetic biology. *Ecol Lett*, 12:693--715, 2009.

A. Cerutti and M. Rescigno. The biology of intestinal immunoglobulin a responses. *Immunity*, 28(6):740--750, 2008.

PS Chain, E Carniel, FW Larimer, J Lamerdin, and et al. Stoutland, PO and. Insights into the evolution of yersinia pestis through whole-genome comparison with yersinia pseudotuberculosis. *Proc Natl Acad Sci U S A*, 101:13826--13831, 2004.

W. G. Characklis and K. C. Marshall. *Biofilms*. Wiley, 1990.

Q. Chen, B. Sun, H. Wu, Z. Peng, and P.M. Fives-Taylor. Differential roles of individual domains in selection of secretion route of a streptococcus parasanguinis serine-rich adhesin, fap1. *Journal of bacteriology*, 189(21):7610--7617, 2007.

L. Chung, S. E. Axelsson, S Lindgren, and WJ Dobrogosz. In vitro studies on reuterin synthesis by lactobacillus reuteri. *Microbial Ecology in Health and Disease*, 2(2): 137--144, 1989.

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

A. Croswell, E. Amir, P. Teggatz, M. Barman, and NH. Salzman. Prolonged impact of anitbiotics on intestinal microbial ecology and susceptibility to enteric salmonella infection. *Infect. Immun*, 77:2741--2743, 2009.

CR. Currie and AE. Stuart. Weeding and grooming of pathogens in agriculture by ants. *Proc. R Soc. London. B*, 268(1471):1033--1039, 2001.

CR. Currie, UG. Mueller, and D. Malloch. The agricultural pathology of ant fungus gardens. *PNAS*, 19(14):7998--8002, 1999.

CR. Currie, M. Poulsen, J. Mendenhall, JJ. Boomsma, and J. Billen. Coevolved crypts and exocrine glands support mutualistic bacteria in fungus-growing ants. *Science*, 31(5757):81--83, 2006.

C Dale and NA Moran. Molecular interactions between bacterial symbionts and their hosts. *Cell*, 126:453--465, 2006.

de Champs, Maroncle C., Balestrino N., C. D., Rich, and C. Forestier. Persistence of colonization of intestinal mucosa by a probiotic strain, lactobacillus casei subsp. rhamnosus lcr35, after oral consumption. *J Clin Microbiol*, 41:1270--1273, 2003.

L. Dethlefsen, PB. Eckburg, EM. Bik, and DA. Relman. Assembly of the human intestinal microbiota. *Trends in Ecology and Evolution*, 21(9):517--523, 2006.

L. Dethlefsen, M. McFall-Ngai, and DA. Relman. An ecological and evolutionary perspective on human-microbe mutualism and disease. *Nature*, 449(7164):811--818, 2007.

M. Doebeli and N. Knowlton. The evolution of interspecific mutualisms. *PNAS*, 95 (15):8676--8680, 1998.

Y. E. Dommels, R. A. Kemperman, Y. E. Zebregs, R. B. Draaisma, A. Jol, D. A. Wolvers, E. E. Vaughan, and R. Albers. Survival of lactobacillus reuteri dsm 17938 and lactobacillus rhamnosus gg in the human gastrointestinal tract with daily consumption of a low-fat probiotic spread. *Appl Environ Microbiol*, 75:6198--6204, 2009.

AE. Douglas. Nutritional interactions in insect-microbial symbioses: Aphids and their symbiotic bacteria buchnera. *Annu Rev. Entomol.*, 43:17--37, 1998.

A.E. Douglas. Symbiotic microorganisms: untapped resources for insect pest control. *TRENDS in Biotechnology*, 25(8):338--342, 2007.

R. Dubos, RW. Schaedler, R. Costello, and P. Hoet. Indigenous, normal, and autochthonous flora of the gastrointestinal tract. *Journal of Experimental Medicine*, 67(1):67--76, 1965.

C. Dunne, L. Murphy, S. Flynn, L. O'Mahony, S. O'Halloran, M. Feeney, D. Morrissey, G. Thornton, G. Fitzgerald, C. Daly, B. Kiely, E. M. Quigley, G. C. O'Sullivan, F. Shanahan, and J. K. Collins. Probiotics: from myth to reality demonstration of functionality in animal models of disease and in human clinical trials. *Antonie Van Leeuwenhoek*, 76:279--292, 1999.

C. Efthymiou and P. A. Hansen. An antigenic analysis of lactobacillus acidophilus. *J Infect Dis*, 110:258--267, 1962.

M Eppinger, C Baar, B Linz, G Raddatz, C Lanz, and et al. Who ate whom? adaptive helicobacter genomic changes that accompanied a host jump from early humans to large felines. *PLoS Genetics*, 2(e120):e120, 2006.

S Flynn, D van Sinderen, GM Thornton, H Holo, IF Nes, and et al. 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, 2002.

G. Fonty, P. Gouet, JP. Jouany, and J. Senaud. Establishment of the microflora and anaerobic fungi in the rumen of lambs. *J Gen Microbiol*, 133:1835--1843, July 1987.

B. M. Forde, B. A. Neville, M. M. O'Donnell, E. Riboulet-Bisson, M. J. Claesson, A. Coghlan, R. P. Ross, and P. W. O'Toole. Genome sequences and comparative genomics of two lactobacillus ruminis strains from the bovine and human intestinal tracts. *Microb Cell Fact*, 10 Suppl 1:S13, 2011.

KR. Foster and T. Wenseleers. A general model for the evolution of mutualisms. *Journal of Evolutionary Biology*, 19(4):1283--1293, 2006.

SA. Frese, AK. Benson, GW. Tannock, DM. Loach, J. Kim, M. Zhang, PL. Oh, NC. Heng, PB. Patil, N. Juge, DA. Mackenzie, BM. Pearson, A. Lapidus, E. Dalin, H. Tice, E. Goltsman, M. Land, L. Hauser, N. Ivanova, NC. Kyrpides, and J. Walter. The evolution of host specialization in the vertebrate gut symbiont lactobacillus reuteri. *PLoS Genetics*, 7(2):e1001314, 2011.

SA. Frese, RW. Hutkins, and J Walter. Comparison of the colonization ability of autochthonous and allochthonous strains of lactobacilli in the human gastrointestinal tract. *Advances in Microbiology*, 2:e, 2012.

E.H. Froeliger and P. Fives-Taylor. Streptococcus parasanguisfimbria-associated adhesin fap1 is required for biofilm formation. *Infection and immunity*, 69(4): 2512--2519, 2001.

R. Fuller. Ecological studies on the lactobacillus flora associated with the crop epithelium of the fowl. *Journal of Applied Microbiology*, 36(1):131--139, 1973.

R. Fuller. Nature of the determinant responsible for the adhesion of lactobacilli to chicken crop epithelial cells. *Journal of general microbiology*, 87(2):245--250, 1975.

R. Fuller, PA. Barrow, and BE. Brooker. Bacteria associated with the gastric epithelium of neonatal pigs. *Applied and Environmental Microbiology*, 35(3):582--591, 1978.

RA. Fuller and A. Turvey. Bacteria associated with the intestinal wall of the fowl (gallus domesticus). *Journal of Applied Bacteriology*, 34(3):617--622, 1971.

V. Gaboriau-Routhiay, S. Rakotobe, E. Lecuyer, I. Mulder, A. Lan, C. Bridonneau, V. Rochet, A. Pisi, De Paepe, Brandi M., Eberl G., Snel G., Kelly J., and Cerf-Bensussan D. The key role of segmeneted filamentous bacteria in the coordinate maturation of gut helper t cell responses. *immunity*, 31(4):677--689, 2009.

RC. Gentleman, VJ. Carey, BJ. Bates, BM. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothom, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, AJ. Rossini, G. Sawitzki, C. Smith, GK. Smyth, L. Tierney, YH. Yang, and J. Zhang. Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biology*, 5(10):1--22, 2004.

SR. Gill, M. Pop, RT. DeBoy, and PB Eckburg. Metagenomic analysis of the human distal gut microbiome. *Science*, 312(5778):1355--1359, 2006.

J Goh and TR. Klaenhammer. Functional roles of aggregation-promoting-like factor in stress tolerance and adherence of lactobacillus acidophilus ncfm. *Applied and Environmental Microbiology*, 76(15):5005--5012, 2010.

A.L. Goodman, N.P. McNulty, Y. Zhao, D. Leip, R.D. Mitra, C.A. Lozupone, R. Knight, and J.I. Gordon. Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell host & microbe*, 6(3):279--289, 2009.

DM. Gordon and A. Cowling. The distribution and genetic structure of escherichia coli in australian vertebrates: host and geographic effects. *Microbiology*, 149(12):3575--3586, 2003. doi: 10.1099/mic.0.26486-0. URL `http://mic.sgmjournals.org/content/149/12/3575.abstract`.

J Goris, KT Konstantinidis, JA Klappenbach, T Coenye, P Vandamme, and et al. Dna-dna hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol*, 57:81--91, 2007.

M. Haarman and J. Knol. Quantitative real-time pcr analysis of fecal lactobacillus species in infants receiving a prebiotic infant formula. *Appl Environ Microbiol*, 72: 2359--2365, 2006.

S Hammons, PL. Oh, I Martinez, K. Clark, V. Schlegel, E. Sitorius, SE. Scheideler, and J Walter. A small variation in diet influences the¡ i¿ lactobacillus¡/i¿ strain composition in the crop of broiler chickens. *Systematic and Applied Microbiology*, 33:275--281, 2010.

D Heavens, LE. Tailford, L. crossman, F. jeffers, D. Mackenzie, M. Caccamo, and N Juge. Genome sequence of the vertebrate gut symbiont lactobacillus reuteri atcc 53608. *Journal of Bacteriology*, 193:4015--4016, 2011.

NC Heng, JM Bateup, DM Loach, X Wu, HF Jenkinson, and et al. Influence of different functional elements of plasmid pgt232 on maintenance of recombinant

plasmids in lactobacillus reuteri populations in vitro and in vivo. *Appl Environ Microbiol*, 65:5378--5385, 1999.

EA. Herre, N. Knowlton, UG Mueller, and SA. Rehner. The evolution of mutualisms: exploring the paths between conflict and cooperation. *Trends in Ecology and Evolution*, 14(2):49--53, 1999.

Abbas Hilmi, Surakka HT., Apajalahti A., and Saris PEJ. J. Identification of the most abundant lactobacillus species in the crop of 1- and 5-week-old broiler chickens. *Applied and Environmental Microbiology*, 73(24):7867--7873, 2007.

G. Hinkle, J.K. Wetterer, T.R. Schultz, M.L. Sogin, et al. Phylogeny of the attine ant fungi based on analysis of small subunit ribosomal rna gene sequences. *Science-AAAS-Weekly Paper Edition*, 266(5191):1695--1696, 1994.

M Hoffmann, E Rath, G Holzlwimmer, L Quintanilla-Martinez, D Loach, and et al. Lactobacillus reuteri 100-23 transiently activates intestinal epithelial cells of mice that have a complex microbiota during early stages of colonization. *J Nutr*, 138: 1684--1691, 2008.

MT Holden, Z Heather, R Paillot, KF Steward, K Webb, and et al. Genomic evidence for the evolution of streptococcus equi: host restriction, increased virulence, and genetic exchange with human pathogens. *PLoS Pathog*, 5:e1000346, 2009.

KE Holt, J Parkhill, CJ Mazzoni, P Roumagnac, FX Weill, and et al. High-throughput sequencing provides insights into genome variation and evolution in salmonella typhi. *Nat Genet*, 40:987--993, 2008.

E. Hüfner, R.A. Britton, S. Roos, H. Jonsson, and C. Hertel. Global transcriptional

response of¡ i¿ lactobacillus reuteri¡/i¿ to the sourdough environment. *Systematic and applied microbiology*, 31(5):323--338, 2008.

C. N. Jacobsen, Rosenfeldt Nielsen, Hayford V., Moller A. E., Michaelsen P. L., Paerregaard K. F., Sandstrom A., M. B., Tvede, and M. Jakobsen. Screening of probiotic activities of forty-seven strains of lactobacillus spp. by in vitro techniques and evaluation of the colonization ability of five selected strains in humans. *Appl Environ Microbiol*, 65:4949--4956, 1999.

L. Jin, K. Hinde, and L. Tao. Species diversity and relative abundance of lactic acid bacteria in the milk of rhesus monkeys (macaca mulatta). *Journal of medical primatology*, 40(1):52--58, 2011.

P. Kabadjova, X. Dousset, V. Le Cam, and H. Prevost. Differentiation of closely related carnobacterium food isolates based on 16s-23s ribosomal dna intergenic spacer region polymorphism. *Appl Environ Microbiol*, 68:5358--5366, 2002.

Otto Kandler, Karl-Otto Stetter, and Ruth Kohl. Lactobacillus reuteri sp. nov., a new species of heterofermentative lactobacilli. *Allgemeine, angewandte und Oekologische Mikrobiologie*, 1:264--269, September 1980.

A. Khoruts, J. Dicksved, J. K. Jansson, and M. J. Sadowsky. Changes in the composition of the human fecal microbiome after bacteriotherapy for recurrent clostridium difficile-associated diarrhea. *J Clin Gastroenterol*, 44:354--360, 2010.

Y. Kikuchi, T. Hosokawa, N. Nikoh, XY. Meng, Y. Kamagata, and T. Fukatsu. Host-symbiont co-speciation and reductive genome evolution in gut symbiotic bacteria of acanthosomatid stinkbugs. *BMC Biol.*, 7(2):e2, 2009.

H. Kiss, B. Kögler, L. Petricevic, I. Sauerzapf, S. Klayraung, K. Domig, H. Viernstein,

and W. Kneifel. Vaginal lactobacillus microbiota of healthy women in the late first trimester of pregnancy. *BJOG: An International Journal of Obstetrics & Gynaecology*, 114(11):1402--1407, 2007.

M Kleerebezem, LE Quadri, OP Kuipers, and WM de Vos. Quorum sensing by peptide pheromones and two-component signal-transduction systems in gram-positive bacteria. *Mol Microbiol*, 24:895--904, 1997.

T. D. Klingberg and B. B. Budde. The survival and persistence in the human gastrointestinal tract of five potential probiotic lactobacilli consumed as freeze-dried cultures or as probiotic sausage. *Int J Food Microbiol*, 109:157--159, 2006.

KD. Kohl, RB. Weiss, C. Dale, and MD. Dearing. Diversity and novelty of the gut microbial community of an herbivorous rodent (neotoma bryanti). *Symbiosis*, 54: 47--54, 2011.

KT Konstantinidis and JM Tiejde. Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci U S A*, 102:2567--2572, 2005.

R. Kulwich, L. Struglia, and PB. Pearson. The effect of coprophagy on the excretion of b vitamins by the rabbit. *Journal of Nutrition*, 49(4):639--645, 1953.

P.D. Kyle and G.Z. Kyle. An evaluation of the role of microbial flora in the salivary transfer technique for hand-rearing chimney swifts. *Wildlife Rehabilitation8*, 8: 65--71, 1993.

M. Lerche and G. Reuter. Isolierung und differenzierung anaerober lactobacillaceae aus dem darm erwachsener menschen (beitrag zum lactobacillus bifidus-problem). *Zbl. Bakt I Orig.*, 180:324--356, 1961.

RE Ley, DA Peterson, and JI. Gordon. Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell*, 124:837--848, 2006.

RE. Ley, M. Hamady, C. Lozupone, PJ. Turnbaugh, RR. Ramey, JS. Bircher, ML. Schlegel, TA. Tucker, MD. Schrenzel, R. Knight, and JI. Gordon. Evolution of mammals and their gut microbes. *Science*, 320(5883):1647--1651, 2008a.

RE. Ley, CA. Lozupone, M. Hamady, R. Knight, and JI. Gordon. Worlds within worlds: evolution of the vertebrate gut microbiota. *Nature Reviews Microbiology*, 6: 776--788, 2008b.

YH Li, YY Chen, and RA Burne. Regulation of urease gene expression by streptococcus salivarius growing in biofilms. *Environ Microbiol*, 2:169--177, 2000.

J HC. Lin and DC. Savage. Host specificity of the colonization of murine gastric epithelium by lactobacilli. *FEMS Microbiology Letters*, 24(1):67--71, 1984.

B. Linz, F. Balloux, Y. Moodley, A. Manica, H. Liu, P. Roumagnac, D. Falush, C. Stamer, F. Prugnolle, S.W. Van Der Merwe, et al. An african origin for the intimate association between humans and helicobacter pylori. *Nature*, 445(7130): 915--918, 2007.

AEF. Little, T. Murakami, UG. Mueller, and CR. Currie. Defending against parasites: fungus-growing ants combine specialized behaviours and microbial symbionts to protect their fungus gardens. *Biol Letters*, 2(1):12--16, 2006.

M Livingston, D Loach, M Wilson, GW Tannock, and M Baird. Gut commensal lactobacillus reuteri 100-23 stimulates an immunoregulatory response. *Immunol Cell Biol*, 88:99--102, 2010.

M Lizier, PG. Sarra, R. Cauda, and F. Lucchini. Comparison of expression vectors in lactobacillus reuteri strains. *FEMS Microb Letters*, 308(1):8--15, 2010.

RG. LoCascio, MR. Ninonuevo, SL. Freeman, DA. Sela, R. Grimm, Lebrilla CB, Mills DA, and German JB. Glycoprofiling of bifidobacterial consumption of human milk oligosaccharides demonstrates strain specific, preferential consumption of small chain glycans secreted in early human lactation. *J Agric Food Chem*, 55(22):8914--8919, 2007.

MP. Lombardo. Access to mutualistic endosymbiotic microbes: an under-appreciated benefit of group living. *Behav. Ecol. Sociobiol.*, 62:479--497., 2008.

BV Lowder, CM Guinane, NL Ben Zakour, LA Weinert, A Conway-Morris, and et al. Recent human-to-poultry host jump, adaptation, and pandemic spread of staphylococcus aureus. *Proc Natl Acad Sci U S A*, 106:19545--19550, 2009.

D. Mackenzie, F. Jeffers, ML. Parker, A. Vibert-Vallet, RJ. Bongaerts, S. Roos, J Walter, and N. Juge. Strain-specific diversity of mucus-binding proteins in the adhesion and aggregation properties of lactobacillus reuteri. *microbiology*, 156: 3368--3378, 2010.

MS. Mameesh and BC. Johnson. Dietary vitamin k requirement of the rat. *Exp Biol Med*, 103:378--380, 1960.

R. Mandar and M. Mikelsaar. Transmission of mother's microflora to the newborn at birth. *Neonatology*, 69(1):30--35, 1996.

M.J. Mandel, M.S. Wollenberg, E.V. Stabb, K.L. Visick, and E.G. Ruby. A single regulatory gene is sufficient to alter bacterial host range. *Nature*, 458(7235):215--218, 2009.

E.E. Mann, K.C. Rice, B.R. Boles, J.L. Endres, D. Ranjit, L. Chandramohan, L.H. Tsang, M.S. Smeltzer, A.R. Horswill, and K.W. Bayles. Modulation of edna release and degradation affects staphylococcus aureus biofilm maturation. *PLoS One*, 4(6): e5822, 2009.

VM Markowitz, E Szeto, K Palaniappan, Y Grechkin, and K Chu. The integrated microbial genomes (img) system in 2007: data content and analysis tool extensions. *Nucleic Acids Res*, 36:suppl 1, 2008.

R. Martín, S. Langa, C. Reviriego, E. Jimínez, M.L. Marín, J. Xaus, L. Fernández, and J.M. Rodríguez. Human milk is a source of lactic acid bacteria for the infant gut. *The Journal of pediatrics*, 143(6):754--758, 2003.

R. Martín, S. Delgado, A. Maldonado, E. Jiménez, M. Olivares, L. Fernández, O.J. Sobrino, and J.M. Rodríguez. Isolation of lactobacilli from sow milk and evaluation of their probiotic potential. *Journal of Dairy Research*, 76(04):418--425, 2009.

R. Martín, M. Olivares, M. Pérez, J. Xaus, C. Torre, L. Fernández, and J.M. Rodríguez. Identification and evaluation of the probiotic potential of lactobacilli isolated from canine milk. *The Veterinary Journal*, 185(2):193--198, 2010.

I. Martinez, J. Kim, P. R. Duffy, V. L. Schlegel, and J. Walter. Resistant starches types 2 and 4 have differential effects on the composition of the fecal microbiota in human subjects. *PLoS One*, 5:e15046, 2010.

Sk Mazmanian, CH Liu, AO Tzianabos, and DL Kasper. An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system cell. *Volume*, 1(15):107--118, July 2005.

MA McConnell, AA Mercer, and GW Tannock1. Transfer of plasmid pambl between

members of the normal microflora inhabiting the murine digestive tract and modification of the plasmid in a lactobacillus reuteri host. *Microbial Ecology in Health and Disease*, 4:343--355, 1991.

M. McFall-Ngai. Adaptive immunity: care for the community. *Nature*, 445(7124): 153--153, 2007.

MJ. Mcfall-Ngai and EG. Ruby. Symbiont recognition and subsequent morphogenesis as early events in an animal-bacterial mutualism. *Science*, 254(5037):1491--1494, 1991.

MJ. Mcfall-Ngai and EG. Ruby. Sepiolids and vibrios: When first they meet. *Bioscience*, 48(4):257--265, 1998.

M Medina and JL Sachs. Symbiont genomics, our new tangled bank. *Genomics*, 95: 129--137, 2010.

D Medini, C Donati, H Tettelin, V Masignani, and R Rappuoli. The microbial pan-genome. *Curr Opin Genet Dev*, 15:589--594, 2005.

C. B. Meroth, J. Walter, C. Hertel, M. J. Brandt, and W. P. Hammes. Monitoring the bacterial population dynamics in sourdough fermentation processes by using pcr-denaturing gradient gel electrophoresis. *Appl Environ Microbiol*, 69:475--482, 2003.

Elie Metchnikoff. *The Prolongation of Life: Optimistic Studies*. Putnam's Sons, New York, 1907.

Tamara K. Mills, Michael P. Lombardo, and Patrick A. Thorpe. Microbial colonization of the cloacae of nestling tree swallows. *The Auk*, 116(4):pp. 947--956, 1999. ISSN 00048038. URL `http://www.jstor.org/stable/4089674`.

NA Moran. Tracing the evolution of gene loss in obligate bacterial symbionts. *Curr Opin Microbiol*, 6:512--518, 2003.

NA Moran. Symbiosis as an adaptive process and source of phenotypic complexity. *Proc Natl Acad Sci U S A*, 104:18627--18633, 2007.

N.A. Moran and G.R. Plague. Genomic changes following host restriction in bacteria. *Current opinion in genetics & development*, 14(6):627--633, 2004.

N.A. Moran, A. Mira, et al. The process of genome shrinkage in the obligate symbiont buchnera aphidicola. *Genome Biol*, 2(0054):1--0054, 2001.

NA. Moran, JP. McCutcheon, and A. Nakabachi. Genomics and evolution of heritable bacterial symbionts. *Annu Rev. Genet*, 42:165--90, 2008.

NA. Munson Moran, Baumann MA., and Ishikawa H. P. A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. *Proc. R. Soc. Lond. B*, 253(1337):167--171, 1993.

H Morita, H. Toh, S. Fukuda, and H et al Horikawa. Comparative genome analysis of lactobacillus reuteri and lactobacillus fermentum reveal a genomic island for reuterin and cobalamin production. *DNA Research*, 15(3):151--161, 2008.

A Moya, J Pereto, R Gil, and A Latorre. Learning how to live together: genomic insights into prokaryote-animal symbioses. *Nat Rev Genet*, 9:218--229, 2008.

UG. Mueller, TR. Schultz, CR. Currie, RMM. Adams, and D. MAlloch. The origin of the attine ant-fungus mutualism. *The Quarterly Review of Biology*, 76(2):169--197, 2001.

Ulrich G. Mueller, Jessica Poulin, and Rachelle M. M. Adams. Symbiont choice in a fungus-growing ant (attini, formicidae). *Behavioral*

Ecology, 15(2):357--364, 2004. doi: 10.1093/beheco/arh020. URL `http://beheco.oxfordjournals.org/content/15/2/357.abstract`.

KE Nelson, GM Weinstock, SK Highlander, KC Worley, HH Creasy, and et al. A catalog of reference genomes from the human microbiome. *Science*, 328:994--999, 2010.

MK. Nishiguchi, EG. Ruby, and M. McFall-Ngai. Competitive dominance among strains of luminous bacteria provides an unusual form of evidence for parallel eovlution in sepiolid squid-vibrio symbioses. *Applied and Environmental Microbiology*, 64:3209--3213, 1998.

R. Noë and P. Hammerstein. Biological markets. *Tree*, 10(8):336--339, 1995.

P Normand, P Lapierre, LS Tisa, JP Gogarten, and N Alloisio. Genome characteristics of facultatively symbiotic frankia sp. strains reflect host range and host plant biogeography. *Genome Res*, 17:7--15, 2007.

J. O'Callaghan and P. W. O'Toole. *Lactobacillus: Host-Microbe Relationships*. Curr Top Microbiol Immunol, In press, 2012.

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

R. Oozeer, A. Leplingard, D. D. Mater, A. Mogenet, R. Michelin, I. Seksek, P. Marteau, J. Dore, J. L. Bresson, and G. Corthier. Survival of lactobacillus casei in the human digestive tract after consumption of fermented milk. *Appl Environ Microbiol*, 72: 5615--5617, 2006.

P. W. O'Toole and J. C. Cooney. Probiotic bacteria influence the composition and function of the intestinal microbiota. *Interdiscip Perspect Infect Dis*, 85(1752):1--9, 2008.

A. C. Ouwehand, S. Salminen, and E. Isolauri. Probiotics: an overview of beneficial effects. *Antonie Van Leeuwenhoek*, 82:279--289, 2002.

C. Palmer, EM Bik, DB DiGiulio, DA Relman, and PO Brown. Development of the human infant intestinal microbiota. *PLoS Biol*, 5(7):e177, 2007.

J Parkhill, M Sebaihia, A Preston, LD Murphy, N Thomson, and et al. Comparative analysis of the genome sequences of bordetella pertussis, bordetella parapertussis and bordetella bronchiseptica. *Nat Genet*, 35:32--40, 2003.

MR Parsek and PK. Singh. Bacterial biofilms: An emerging link to disease pathogenesis. *Annu. Rev. Microbiol*, 57:677--701, 2003.

John Penders, Carel Thijs, Cornelis Vink, Foekje F. Stelma, Bianca Snijders, Ischa Kummeling, Piet A. van den Brandt, and Ellen E. Stobberingh. Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics*, 118(2):511--521, 2006. doi: 10.1542/peds.2005-2824. URL `http://pediatrics.aappublications.org/content/118/2/511.abstract`.

D.A. Peterson, N.P. McNulty, J.L. Guruge, and J.I. Gordon. Iga response to symbiotic bacteria as a mediator of gut homeostasis. *Cell host & microbe*, 2(5):328--339, 2007.

MW Pfaffl. A new mathematical model for relative quantification in real-time rt-pcr. *Nucleic Acid Res*, 29(9):e45, 1999.

JI Prosser, BJ Bohannan, TP Curtis, RJ Ellis, MK Firestone, and et al. The role of ecological theory in microbial ecology. *Nat Rev Microbiol*, 5:384--392, 2007.

S. Rabot, J. Rafter, G. T. Rijkers, B. Watzl, and J. M. Antoine. Guidance for substantiating the evidence for beneficial effects of probiotics: impact of probiotics on digestive system metabolism. *J Nutr*, 140:677S--689S, 2010.

J. Ravel, P. Gajer, Z. Abdo, G.M. Schneider, S.S.K. Koenig, S.L. McCulle, S. Karlebach, R. Gorle, J. Russell, C.O. Tacket, et al. Vaginal microbiome of reproductive-age women. *Proceedings of the National Academy of Sciences*, 108(Supplement 1): 4680--4687, 2011.

R Reniero, P Cocconcelli, V Bottazzi, and L Morelli. High frequency of conjugation in lactobacillus mediated by an aggregation-promoting factor. *Microbiology*, 138: 763--768, 1992.

ML Reno, NL Held, CJ Fields, PV Burke, and RJ Whitaker. Biogeography of the sulfolobus islandicus pan-genome. *Proc Natl Acad Sci U S A*, 106:8605--8610, 2009.

G. Reuter. The lactobacillus and bifidobacterium microflora of the human intestine: Composition and succession. *Current Issues in Intestinal Microbiology*, 2(2):43--53, 2001.

JA. Reyniers. Germfree vertebrates: Present status. *Annals of the NY Academy of Sciences*, 78(1):1--400, 1959.

NW. Rigel and M Braunstein. A new twist on an old pathway, accessory sec systems. *M Mol Microbiol*, 69:291 -- 302., 2008.

S. Roach, DC. Savage, and GW. Tannock. Lactobacilli isolated from the stomach of conventional mice. *Applied and Envbironmental microbiology*, 33(5):1197--1203, 1977.

J.R.C. Roberts. *Evaluation of Helicobacter hepaticus Bacterial Shedding in Fostered and Strategically Housed C57BL/6 Mice*. PhD thesis, Louisiana State University, 2005.

S. Roos, F. Karner, L. Axelsson, and H. Jonsson. Lactobacillus mucosae sp. nov., a new species with in vitro mucus-binding activity isolated from pig intestine. *Int J Syst Evol Microbiol*, 50:251--258, 2000.

JM Rouillard, M Zuker, and E Gulari. Oligoarray 2.0: design of oligonucleotide probes for dna microarrays using a thermodynamic approach. *Nucleic Acids Res*, 31:3057--3062, 2003.

JL. Round and SK. Mazmanian. The gut microbiota shapes intestinal immune responses during health and disease. *Nature reviews immunology*, 9:313--323, 2009.

S Rozen and H Skaletsky. Lactobacillus reuteri (american type culture collection strain 55730) versus simethicone in the treatment of infantile colic: A prospective randomized study. *Methods in Molecular Biology*, 132(3):365--386, 1999.

EG. Ruby and MJ. Mcfall-Ngai. A squid that glows in the night: development of an animal-bacterial mutualism. *J. of Bacteriol*, 174(15):4865--4870, 1992.

EG. Ruby and MJ. McFall-Ngai. Oxygen-utilizing reactions and symbiotic colonization of the squid light organ by vibrio fischeri. *Trends in Microbiology*, 7(10):414--420, 1999.

EG Ruby, M Urbanowski, J Campbell, A Dunn, and M Faini. Complete genome sequence of vibrio fischeri: a symbiotic bacterium with pathogenic congeners. *Proc Natl Acad Sci U S A*, 102:3004--3009, 2005.

JL Sachs, UG Mueller, TP Wilcox, and JJ Bull. The evolution of cooperation. *The Quarterly Review of Biology*, 79(2):135--160, June 2004.

M. E. Sanders. Probiotics: considerations for human health. *Nutr Rev*, 61:91--99, 2003.

M. E. Sanders and T. R. Klaenhammer. Invited review: the scientific basis of lactobacillus acidophilus ncfm functionality as a probiotic. *J Dairy Sci*, 84:319--331, 2001.

F. Santos, J. Vera, R. van der Heijden, G. Valdez, WM. de Vos, F. Sesma, and J. Hugenholtz. The complete coenzyme b12 biosynthesis gene cluster of lactobacillus reuteri crl1098. *Microbiology*, 154:81--93, January 2008.

D. C. Savage. Microbial ecology of the gastrointestinal tract. *Annu Rev Microbiol*, 31: 107--133, 1977.

D.C. Savage and R.V.H. Blumershine. Surface-surface associations in microbial communities populating epithelial habitats in the murine gastrointestinal ecosystem: scanning electron microscopy. *Infection and immunity*, 10(1):240--250, 1974.

DC. Savage, R. Dubos., and RW. Schaedler. The gastrointestinal epithelium and its autochthonous bacterial flora. *Journal of Experimental Medicine*, 127(1):67--76, 1968.

C. Schwab, J. Walter, G.W. Tannock, R.F. Vogel, and M.G. Gänzle. Sucrose utilization and impact of sucrose on glycosyltransferase expression in¡ i¿ lactobacillus reuteri¡/i¿. *Systematic and applied microbiology*, 30(6):433--443, 2007.

DA. Sela, J. Chapman, A. Adeuya, and JH et al. Kim. The genome sequence of

bifidobacterium longum subsp infantis reveals adaptations for milk utilization within the infant microbiome. *PNAS*, 105(48):18946--18969, 2008.

B.K. Sharma-Kuinkel, E.E. Mann, J.S. Ahn, L.J. Kuechenmeister, P.M. Dunman, and K.W. Bayles. The staphylococcus aureus lytsr two-component regulatory system affects biofilm formation. *Journal of bacteriology*, 191(15):4767--4775, 2009.

M. Shirota, K. Aso, and A. Iwabuchi. Studies on intestinal microflora 1. its constitution in healthy infants and the effect of oral administration of l. acidophilus strain shirota. *Nihon Saikingaku Zasshi*, 21(5):274--283, 1966.

I. Sims, SA. Frese, J. Walter, D. Loach, M. Wilson, K. Appleyard, J. Eason, M. Livingston, M. Baird, G. Gook, and GW. Tannock. Structure and functions of exopolysaccharide produced by gut commensal lactobacillus reuteri 100-23. *The ISME Journal*, 5:1115--1124, 2011.

G. Sinkiewicz and Lennart Ljunggren. Occurrence of lactobacillus reuteri in human breast milk. *Microbial Ecology in Health and Disease*, 20(3):122--126, 2008.

JE Sjostrom and H Larsson. Factors affecting growth and antibiotic susceptibility of helicobacter pylori: effect of ph and urea on the survival of a wild-type strain and a urease-deficient mutant. *J Med Microbiol*, 44:425--433, 1996.

W.T. Sloan, M. Lunn, S. Woodcock, I.M. Head, S. Nee, and T.P. Curtis. Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environmental microbiology*, 8(4):732--740, 2005.

DD Sriramulu, M Liang, D Hernandez-Romero, E Raux-Deery, H Lunsdorf, and et al. Lactobacillus reuteri dsm 20016 produces cobalamin-dependent diol dehydratase in

metabolosomes and metabolizes 1,2-propanediol by disproportionation. *J Bacteriol*, 190:4559--4567, 2008.

B. Stecher and W. D. Hardt. Mechanisms controlling pathogen colonization of the gut. *Curr Opin Microbiol*, 14:82--91, 2011.

B. Stecher and WD. Hardt. The role of microbiota in infectious disease. *Trends in Microbiology*, 16(3):107--114, 2008.

C.E. Stevens and I.D. Hume. Contributions of microbes in vertebrate gastrointestinal tract to production and conservation of nutrients. *Physiological reviews*, 78(2): 393--427, 1998.

AD. Stewart, JM. Jogsdon, and SE. Kelley. An empirical study of the evolution of virulence under both horizontal and vertical transmission. *Evolution*, 59:730--739., 2005.

P.S. Stewart and J. William Costerton. Antibiotic resistance of bacteria in biofilms. *The Lancet*, 358(9276):135--138, 2001.

P. Stoodley, J. D. Boyle, I. Dodds, and H. M. Lappin-Scott. Consensus model of biofilm structure. In *: Biofilms: community interactions and control*, pages 1--9. Edited by Wimpenny, J.W.T., Handley, P.S., Gilbert, P., Lappin-Scott, H.M., and Jones, M. BioLine, Cardiff, UK. Edited by Wimpenny, J.W.T., Handley, P.S., Gilbert, P., Lappin-Scott, H.M., and Jones, M. BioLine, Cardiff, UK, 1997.

MH Sturme, C Francke, RJ Siezen, WM de Vos, and M Kleerebezem. Making sense of quorum sensing in lactobacilli: a special focus on lactobacillus plantarum wcfs1. *Microbiology*, 153:3939--3947, 2007.

MSW Su, PL Oh, J Walter, and MG Gaenzle. Intestinal origin of sourdough lactobacillus reuteri isolates as revealed by phylogenetic, genetic, and physiological analysis. *Appl. Environ. Microbiol*, 78:6777--6780, September 2012.

VI Sudenko, LI Groma, and VS Podgorskii. The antagonistic properties of microaerophilic bacteria isolated from the human and mink digestive tracts. *Mikrobiol Z*, 5:58--66, 1996.

N. Suegara, M. Morotomi, T. Watanabe, Y. Kawal, and M. Mutai. Behavior of the microflora in the rat stomach: adhesion of lactobacilli to the keratinized epithelial cells of the rat stomach in vitro. *Infection and Immunity*, 12(1):173--179, 1975.

J. Sui, S. Leighton, F. Busta, and L. Brady. 16s ribosomal dna analysis of the faecal lactobacilli composition of human subjects consuming a probiotic strain lactobacillus acidophilus ncfm. *J Appl Microbiol*, 93:907--912, 2002.

JT Sullivan and CW Ronson. Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-trna gene. *Proc Natl Acad Sci U S A*, 95:5145--5149, 1998.

Y. Sumi, M. Miyakawa, M. Kanzaki, and Y. Kotake. Vitamin b6 deficiency in germ free rats. *J. Nutr*, 107:1707--14, 1977.

T. Talarico, L. Axelsson, J. Novotny, M Fiuzat, and W. Dobrogosz. Utilization of glycerol as a hydrogen acceptor by lactobacillus reuteri: Purification of 1,3-propanediol:nad+ oxidoreductase. *Appl Environ Microbiol*, 56:943--948, 1990.

TL. Talarico, IA Casas, TC Chung, and WJ Dobrogosz. Production and isolation of reuterin, a growth inhibitor produced by lactobacillus reuteri. *Antimicrob. Agents Chemother*, 32:1854--1858, 1988.

G. W. Tannock, K. Munro, H. J. Harmsen, G. W. Welling, J. Smart, and P. K. Gopal. Analysis of the fecal microflora of human subjects consuming a probiotic product containing lactobacillus rhamnosus dr20. *Appl Environ Microbiol*, 66:2578--2588, 2000.

GW Tannock. Lactic microbiota of pigs, mice and rats. In Wood BJB, editor, *The Lactic Acid Bacteria in Health and Disease:*, pages 21--48. Elsevier, London, 1992.

G.W. Tannock. A special fondness for lactobacilli. *Applied and environmental microbiology*, 70(6):3189--3194, 2004.

GW. Tannock, C. Crichton, GW. Welling, JP. Koopman, and T. Midtvedt. Reconstitution of the gastrointestinal microflora of lactobacillus-free mice. *Applied and Environmental Microbiology*, 54(12):2971--2975, 1988.

G.W. Tannock, R. Fuller, SL Smith, and MA Hall. Plasmid profiling of members of the family enterobacteriaceae, lactobacilli, and bifidobacteria to study the transmission of bacteria from mother to infant. *Journal of clinical microbiology*, 28(6):1225--1228, 1990.

GW Tannock, S Ghazally, J Walter, D Loach, and H Brooks. Ecological behavior of lactobacillus reuteri 100-23 is affected by mutation of the luxs gene. *Appl Environ Microbiol*, 71:8419--8425, 2005.

GW. Tannock, CM. Wilson, D. Loach, GM. Cook, J. Eason, PW. O'Toole, G. Holtrop, and B. Lawley. Resource partitioning in relation to cohabitation of lactobacillus species in the mouse forestomach. *The ISME Journal*, 6:927--938, 2012.

O. Tenaillion, D. Skurnik, B. Picard, and E. Denamur. The population genetics of commensal escherichia coli. *Nature Reviews Microbiology*, 8(3):207--217, 2010.

H Tettelin, V Masignani, MJ Cieslewicz, C Donati, D Medini, and et al. Genome analysis of multiple pathogenic isolates of streptococcus agalactiae: implications for the microbial ''pan-genome''. *Proc Natl Acad Sci U S A*, 102:13950--13955, 2005.

G.H. Thomas, J. Zucker, S.J. Macdonald, A. Sorokin, I. Goryanin, and A.E. Douglas. A fragile metabolic network adapted for cooperation in the symbiotic bacterium buchnera aphidicola. *BMC systems biology*, 3(1):24, 2009.

M. Tobajas, AF Mohedano, JA Casas, and JJ Rodríguez. A kinetic study of reuterin production by¡ i¿ lactobacillus reuteri¡/i¿ pro 137 in resting cells. *Biochemical engineering journal*, 35(2):218--225, 2007.

M Touchon, C Hoede, O Tenaillon, V Barbe, S Baeriswyl, and et al. Organised genome dynamics in the escherichia coli species results in highly diverse adaptive paths. *PLoS Genet*, 5:e1000344, 2009.

RL. Trivers. The evolution of reciprocal altruism. *Quarterly Review of Biology*, 46: 35--57, 1971.

P Trosvik, NC. Stenseth, and K. Rudi. Convergent temporal dynamics of the human infant gut microbiota. *The ISME Journal*, 4:151--158, 2010.

K. M. Tuohy, M. Pinart-Gilberga, M. Jones, L. Hoyles, A. L. McCartney, and G. R. Gibson. Survivability of a probiotic lactobacillus casei in the gastrointestinal tract of healthy human volunteers and its impact on the faecal microflora. *J Appl Microbiol*, 102:1026--1032, 2007.

PJ. Turnbaugh, RE. Ley, MA. Mahowald, V. Magrini, ER. Mardis, and JI. Gordon. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*, 444:1027--1031, 2006.

P.J. Turnbaugh, R.E. Ley, M. Hamady, C.M. Fraser-Liggett, R. Knight, and J.I. Gordon. The human microbiome project. *Nature*, 449(7164):804--810, 2007.

PJ. Turnbaugh, M. Hamady, T. Yatsunenko, BL. Cantarel, A. Duncan, RE. Ley, ML. Sogin, WJ. Jones, BA. Roe, JP. Affourtit, M. Egholm, B. Henrissat, AC Heath, R. Knight, and JI. Gordon. A core gut microbiome in obese and lean twins. *Nature*, 457:480--484, 2009.

MS. Turner, LM. Hafner, T. Walsh, and PM. Giffard. Identification and characterization of the novel lysm domain-containing surface protein sep from lactobacillus fermentum br11 and its use as a peptide fusion partner in lactobacillus and lactococcus. *Applied and Environmental Microbiology*, 70(6):3673--3680, 2004.

N. Valeur, P. Engel, N. Carbajal, E. Connolly, and K. Ladefoged. Colonization and immunomodulation by lactobacillus reuteri atcc 55730 in the human gastrointestinal tract. *Appl Environ Microbiol*, 70:1176--1181, 2004.

D. van der Waaij, JM. Berghuis-de Vries, and JEC. Lekkerkerk-van der Wees. Colonization resistance of the digestive tract in conventional and antibiotic-treated mice. *The Journal of Hygiene*, 69(3):405--411, 1971.

J.P. van Pijkeren and R.A. Britton. High efficiency recombineering in lactic acid bacteria. *Nucleic acids research*, 40(10):e76--e76, 2012.

G.S. Vernikos and J. Parkhill. Interpolated variable order motifs for identification of horizontally acquired dna: revisiting the salmonella pathogenicity islands. *Bioinformatics*, 22(18):2196--2203, 2006.

T. Vesa, P. Pochart, and P. Marteau. Pharmacokinetics of lactobacillus plantarum

ncimb 8826, lactobacillus fermentum kld, and lactococcus lactis mg 1363 in the human gastrointestinal tract. *Aliment Pharmacol Ther*, 14:823--828, 2000.

KL. Visick and MJ. McFall-Ngai. An exclusive contract: Specificity in he vibrio fischeri-euprymna scolopes partnership. *Journal of Bacteriology*, 182(7):1779--1787, 2000.

T. Wall, K. Bath, RA. Britton, A. Jonsson, J. Versalovic, and S. Roos. The early response to acid shock in lactobacillus reuteri involves the clpl chaperone and a putative cell wall-altering esterase. *Appl. Environ. Microbiol.*, 73(12):3924--3635, 2007.

J. Walter. Ecological role of lactobacilli in the gastrointestinal tract: implications for fundamental and biomedical research. *Appl Environ Microbiol*, 74:4985--4996, 2008.

J. Walter and R. Ley. The human gut microbiome: ecology and recent evolutionary changes. *Annu Rev Microbiol*, 65:411--429, 2011.

J. Walter, C. Hertel, G. W. Tannock, C. M. Lis, K. Munro, and W. P. Hammes. Detection of lactobacillus, pediococcus, leuconostoc, and weissella species in human feces by using group-specific pcr primers and denaturing gradient gel electrophoresis. *Appl Environ Microbiol*, 67:2578--2585, 2001.

J. Walter, N.C.K. Heng, W.P. Hammes, D.M. Loach, G.W. Tannock, and C. Hertel. Identification of lactobacillus reuteri genes specifically induced in the mouse gastrointestinal tract. *Applied and environmental microbiology*, 69(4):2044--2051, 2003.

J. Walter, P. Chagnaud, GW. Tannock, DM. Loach, Dal Bello, Jenkinson F., Hammes HF., and Hertel WP. 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, 2005.

J Walter, DM Loach, M Alqumber, C Rockel, and C Hermann. D-alanyl ester depletion of teichoic acids in lactobacillus reuteri 100-23 results in impaired colonization of the mouse gastrointestinal tract. *Environ Microbiol*, 9:1750--1760, 2007.

J. Walter, C. Schwab, DM. Loach, MG. Gaenzle, and GW. Tannock. 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(1):72--80, 2008.

J. Walter, RA. Britton, and S. Roos. Host-microbial symbiosis in the vertebrate gastrointestinal tract and the lactobacillus reuteri paradigm. *PNAS*, 108(S1):4645--4652, 2011.

FW. Ward and ME. Coates. Gastrointestinal ph measurement in rats: influence of hte microbial flora, diet, and fasting. *Laboratory Animals*, 21(3):216--222, 1987.

M. Watanabe, H. Kinoshita, M. Nitta, R. Yukishita, Y. Kawai, K. Kimura, N. Taketomo, Y. Yamazaki, Y. Tateno, K. Miura, A. Horii, H. Kitazawa, and T. Saito. Identification of a new adhesin-like protein from lactobacillus mucosae me-340 with specific affinity to the human blood group a and b antigens. *J Appl Microbiol*, 109: 927--935, 2010.

E. Wesney and G.W. Tannock. Association of rat, pig, and fowl biotypes of lactobacilli with the stomach of gnotobiotic mice. *Microbial Ecology*, 5(1):35--42, 1979.

LF. Whitehead, TL. Wilkinson, and AE. Douglas. Nitrogen recycling in the pea aphid (acyrthosiphon pisum) symbiosis. *Proc. R Soc. Lond. B*, 250(1328):115--117, 1992.

M.E. Wickham, N.F. Brown, E.C. Boyle, B.K. Coombes, and B.B. Finlay. Virulence is positively selected by transmission success between mammalian hosts. *Current biology*, 17(9):783--788, 2007.

D.M. Wilkinson and T.N. Sherratt. Horizontally acquired mutualisms, an unsolved problem in ecology? *Oikos*, 92(2):377--384, 2001.

RC. Williams and RJ. Gibbons. Inhibition of bacterial adherence by secretory immunoglobulin a: a mechanism of antigen disposal. *Science*, 177:697--699, 1972.

CM. Wilson, RBM. Aggio, P. O'Toole, S. Villas-Boas, and Tannock GW. Transcriptional and metabolomic consequences of luxs inactivation reveal a metabolic rather than quorum-sensing role for luxs in lactobacillus reuteri 100-23. *Journal of Bacteriology*, 194:1743--1746, 2012.

Mollie D. Winfield and Eduardo A. Groisman. Role of nonhost environments in the lifestyles of salmonella and escherichia coli. *Applied and Environmental Microbiology*, 69(7):3687--3694, 2003. doi: 10.1128/AEM.69.7.3687-3694.2003. URL `http://aem.asm.org/content/69/7/3687.short`.

M.E.J. Woolhouse, L.H. Taylor, and D.T. Haydon. Population biology of multihost pathogens. *Science*, 292(5519):1109--1112, 2001.

J. Xu, MK. Bjursell, J. Himrod, S. Deng, LK. Carmichael, HC. Chiang, LV. Hooper, and JI. Gordon. A genomic view of the human-bacteroides thetaiotaomicron symbiosis. *Science*, 299(5615):2074--2076, 2003.

ES. Yip, K. Geszvain, CR. Deloney-Marino, and KL. Visick. The symbiosis regulator rscs controls the syp gene locus, biofilm formation and symbiotic aggregation by vibrio fischeri. *Molecular Microbiology*, 62(6):1586--1600, 2006.

N. Yuki, T. Shimazaki, A. Kushiro, K. Watanabe, K. Uchida, T. Yuyama, and M. Morotomi. Colonization of the stratified squamous epithelium of the nonsecreting area of horse stomach by lactobacilli. *Applied and environmental microbiology*, 66 (11):5030--5034, 2000.

C Zhang, M Zhang, J Ju, J Nietfeldt, J Wise, and et al. Genome diversification in phylogenetic lineages i and ii of listeria monocytogenes: identification of segments unique to lineage ii populations. *J Bacteriol*, 185:5573--5584, 2003.