Title: Comparative evaluation of the genomes of common bacterial members of the *Drosophila* intestinal community

Authors: Kristina Petkau, David Fast, Aashna Duggal, and Edan Foley.

Affiliation: University of Alberta, Department of Medical Microbiology and Immunology, Institute of Virology.

Running Title: Genomic evaluation of *Drosophila*-associated microbes

Correspondence: Edan Foley  
Department of Medical Microbiology and Immunology  
University of Alberta  
Edmonton  
Alberta  
T6G 2S2  
Canada.  
Tel.: (780) 492-0935  
Fax: (780) 492-7521  
Email: efoley@ualberta.ca

Key words: *Drosophila*, intestine, microflora, genomes
ABSTRACT

*Drosophila melanogaster* is an excellent model to explore the molecular exchanges that occur between an animal intestine and their microbial passengers. For example, groundbreaking studies in flies uncovered a sophisticated web of host responses to intestinal bacteria. The outcomes of these responses define critical events in the host, such as the establishment of immune responses, access to nutrients, and the rate of larval development. Despite our steady march towards illuminating the host machinery that responds to bacterial presence in the gut, we know remarkably little about the microbial products that influence bacterial association with a fly host. To address this deficiency, we sequenced and characterized the genomes of three common *Drosophila*-associated microbes: *Lactobacillus plantarum, Lactobacillus brevis* and *Acetobacter pasteurianus*. In each case, we compared the genomes of *Drosophila*-associated strains to the genomes of strains isolated from alternative sources. This approach allowed us to identify molecular functions common to *Drosophila*-associated microbes, and, in the case of *A. pasteurianus*, to identify genes that are essential for association with the host. Of note, many of the gene products unique to fly-associated strains have established roles in the stabilization of host-microbe interactions. We believe that these data provide a valuable starting point for a more thorough examination of the microbial perspective on host-microbe relationships.
INTRODUCTION

Environmental, microbial, and host factors act at mucosal barriers to establish a unique microclimate that shapes the lives of all participant species (1). For example, expression of a host genotype in gastrointestinal tissues works in concert with extrinsic factors to determine microbial associations (2). For their part, the metabolic outputs of the gastrointestinal microflora influence critical events in the host such as education of immune phenotypes (3, 4), development of intestinal structures (5), and access to essential micronutrients (6). Given the intertwined relationship between host phenotype and microbial genotype, it is of some surprise that hosts often tolerate extensive alterations to their microflora in response to environmental shifts such as changes in diet (7). However, alterations to the gastrointestinal microflora are not invariably without consequence, and intestinal dysbiosis may lead to chronic, debilitating, and occasionally deadly diseases within the host (8-12). Our appreciation of the holobiont as an intricate network of biochemical and genetic transactions between multiple participants mandates a thorough evaluation of the microbial genomes that shape host physiology. Unfortunately, such studies are tremendously complex in conventional mammalian models due to the size of the microbiome, and the lack of laboratory techniques for the isolation and manipulation of most intestinal microbes.

The simple invertebrate *Drosophila melanogaster* is an excellent model holobiont (13, 14). From a developmental perspective, the *Drosophila* posterior midgut shares a number of important similarities with the small intestine of more complex mammalian counterparts (15). Both organs are endodermal in origin, and are surrounded by a sheath of mesodermal visceral muscle (16, 17). The mammalian small intestine and *Drosophila* posterior midgut are maintained by regularly spaced, basal intestinal stem cells that generate transitory progenitor cells (18-20); the non-proliferative enteroblasts of *Drosophila*, and the transient-amplifying cells of mammals. In both systems, signals along the Notch-Delta axis promote
differentiation of transitory progenitors into secretory enteroendocrine cells or absorptive enterocytes (15, 21). In contrast to mammals, *Drosophila* lacks specialized basal paneth cells for the release of antimicrobial peptides. Nonetheless, the fly genome encodes antimicrobial peptides that actively contribute to the control of intestinal symbionts and pathogens (22), suggesting release of such factors into the *Drosophila* intestinal lumen. In both the mammalian small intestine and *Drosophila* midgut, host factors and biogeography favor association with members of the *Lactobacillaceae* family (2, 23). In return, metabolites from *Lactobacilli* activate host response pathways that promote intestinal stem cell proliferation (24). Combined with the genetic accessibility of flies and their suitability for longitudinal studies of large populations in carefully defined environments, *Drosophila* is an excellent system to decipher the forces that determine genetic interactions within a holobiont (13, 25, 26).

In contrast to conventional vertebrate models, the *Drosophila* microbiome consists of a small number of bacterial species that are easily isolated and cultured (27). The adult *Drosophila* intestine hosts little to no bacteria immediately after emergence from the pupal case, and the microfloral population grows in number over time (28). Several studies established that environmental factors and host genotype influence the diversity of the microflora (22, 29, 30). It is unclear if bacteria establish stable associations with the host gut, or if they cycle from the intestine to the environment and back (31). Nonetheless, lab-raised or wild *Drosophila* frequently associates with representatives of the genii *Lactobacillus* and *Acetobacter*. These data suggest that the intestinal lumen of an adult fly favors the survival of specific bacteria, and that such bacteria encode the necessary factors to survive or proliferate within a *Drosophila* intestine.

Consistent with a long-term association between the fly intestine and specific microbes, many *Drosophila* phenotypes are influenced by individual *Lactobacillus* or *Acetobacter* species. For example, *Lactobacillus plantarum*, a common *Drosophila*-associated microbe, promotes larval development via regulation of the TOR signal transduction pathway (32), while *Acetobacter pomorum* regulates host
Insulin Growth Factor signals to promote development and metabolic homeostasis (33). In addition, members of the *Acetobacter* and *Lactobacillus* populations regulate levels of essential nutrients in the host (34-36). Combined, these data present a compelling argument that *Lactobacilli* and *Acetobacter* are important members of the *Drosophila*-microbe holobiont.

Despite our advances in the elucidation of *Lactobacillus* and *Acetobacter* influences on their *Drosophila* host, we know remarkably little about the bacterial genomic elements that determine microbial responses to, and survival within, the *Drosophila* adult intestine. To address this issue, we prepared whole genome sequences of three bacterial species that regularly associate with *Drosophila* – *Lactobacillus brevis*, *Lactobacillus plantarum*, and *Acetobacter pasteurianus*. These sequences included those for a *Lactobacillus plantarum* strain isolated from our lab-raised flies, and a separate strain isolated from a wild *Drosophila*. For each species, we compared *Drosophila*-associated bacterial genomes, including ones reported previously, to the genomes of reference strains isolated from non-*Drosophila* sources. We focused on biochemical and signal transduction pathways that respond to environmental factors, as we reasoned that *Drosophila*-associated genomes will display signatures of adaptation to association with their fly host. Our studies identify several biochemical pathways that are enriched in the genomes of *Drosophila*-associated *Lactobacilli*, and uncover *Acetobacter* genomic elements that regulate viability within the host.
RESULTS

Intestinal bacteria face structural and chemical barriers that typically deter association with an animal intestine. In response, microbial genomes encode factors that overcome host defenses to permit bacterial growth or survival. Here, we examined the genomes of *L. brevis*, *L. plantarum* and *A. pasteurianus*, common members of the *Drosophila* intestinal community. For each species, we studied whole-genome sequences of bacterial strains isolated from *Drosophila* intestines, and compared them to related strains isolated from the environment. As it is unclear if *Lactobacilli* or *Acetobacter* establish autochtonous or allochtonous relationships with the adult intestine we will refrain from use of the term commensal throughout this study. Instead, we will identify bacterial strains isolated from flies as associated with *Drosophila*, and we will identify all other strains as environmental isolates. Details on the respective genomes characterized in this study are presented in table 1.

We processed each genome in a similar manner. Where necessary, we used genomic databases to identify the bacterial species of newly sequenced genomes. We then annotated each genome with RAST, used PHAST to scan each genome for intact prophages, and searched for possible CRISPR arrays in the respective genomes. We scrutinized the annotated genomes for functions that facilitate microbial survival within the intestinal lumen, with a focus on genes involved in signal transduction, transcriptional responses, orchestration of stress responses, or induction of virulence factors. Finally, we compared environmental and *Drosophila*-associated genomes for each species to identify bacterial factors that are unique to *Drosophila*-associated genomes. We present the results for each species below.
**Lactobacillus brevis**

**General Genomic Features**

*L. brevis* frequently associates with *Drosophila* intestines, and the whole genome sequence of a fly-associated strain, *L. brevis* EW is available (37). We prepared a whole-genome sequence of an additional *L. brevis* strain (*L. brevis* EF) that we isolated from the intestines of wild-type adult *Drosophila* in our lab. For comparative purposes, we extended our study to include the genome of the environmental ATCC 367 strain. We found that *L. brevis* ATCC 367 readily associated with wild-type adult *Drosophila*. Genome-to-genome distance calculations suggest that the *Drosophila*-associated EW and EF strains are more closely related to each other than to the ATCC 367 strain (Table 2). The genomes of *Drosophila*-associated strains are also larger than the environmental strain, with approximately 500,000 nucleotides more, and an additional 500 coding sequences (Table 2, and Figure 1A).

**Environmental Response Factors**

We then examined genetic regulatory networks within the individual *L. brevis* strains. We hypothesized that *Drosophila*-associated strains will encode distinct regulatory components adapted to the harsh environment of an adult intestine. For these studies, we paid particular attention to two-component systems, transcription factors and additional DNA-binding proteins within the respective genomes. To our surprise, we did not observe substantial differences between *Drosophila*-associated and environmental genomes (Table 3, and supplementary tables 1-3), with the exception of a greater number of transcription factors in the *Drosophila*-associated genomes. Likewise, we only observed slight differences between *Drosophila*-associated and environmental strains when we considered genes dedicated to signal transduction, stress responses, or virulence (Figure 1B-D and supplementary tables 4-6).
Prophages and CRISPR Responses

In contrast to the similarities between environmental and Drosophila associated strains noted above, we detected large differences when we searched for prophages within the respective genomes. The EW and EF genomes include four intact temperate prophages, compared to an absence of prophages from the environmental strain genome (Table 4). We found CRISPR sequences that target a common Lactobacillus phage within all three genomes, while the environmental strain encoded a separate CRISPR array that targets a Lactobacillus plasmid (Table 4). These results suggest an ongoing interaction between prophages and CRISPR defenses in the genomes of Drosophila-associated L. brevis strains.

Function-Based Comparisons of Commensal and Reference Strains of Lactobacillus brevis

The data above point to subtle differences between the genomes of Drosophila-associated and environmental strains of L. brevis. To characterize these differences in greater detail, we performed a function-based comparison of the 185 genes that are common to Drosophila-associated genomes, but absent from the environmental strain. This set of 185 genes describes thirteen distinct functional categories, with forty-seven unique roles (Table 5). Perhaps unsurprisingly, phage and CRISPR-associated gene products account for two of those categories, and cover eleven of the forty-seven unique roles.

Of the remaining gene products, the dominant functional categories are dedicated to roles suited for survival within an intestine. For example, both Drosophila-associated genomes encode a biochemical cascade that converts α-D-glucose-1-phosphate to dTDP-4-dehydro-6-deoxy-L-mannose, an exopolysaccharide that contributes to biofilm formation, and prokaryotic survival within a host intestine (38, 39). Drosophila-associated strains of L. brevis also encode gene products that contribute to the formation of rhamnose-containing glycans, a cell membrane component of acid-fast bacteria that affects host-microbe interactions, such as adhesion, recognition, and biofilm formation (40).
We also identified gene products within Drosophila-associated genomes that facilitate nutrient acquisition from different sources (Table 5). Bacteria frequently respond to limitations in nutritional environments through activation of the cAMP Receptor Protein, a transcription factor that we did not identify in the environmental strain of L. brevis, but found in both Drosophila-associated strains. The cAMP Receptor Protein controls, among other things, the expression of gene products that coordinate metabolism of citrate (41), a function that is also enriched among associated Drosophila-associated L. brevis genomes. In lactic acid bacteria, citrate lyase is activated in acidic environments such as those found in the gut, and increases carbon utilization and energy generation by blocking the inhibitory effects of the Lactobacillus fermentation product lactate (42). Finally, we detected an enrichment of genes involved in the transport and degradation of pectin in Drosophila-associated L. brevis genomes. Pectin is an abundant source of energy and carbon for bacteria that grow on plant and vegetable surfaces, and microbial consumption of pectin accelerates the decay of organic matter. As Drosophila feed on bacteria and yeast that grow on decomposing fruit, we believe that pectin metabolism by the EW and EF strains of L. brevis increases the likelihood of their association with Drosophila in the wild.

**Acetobacter pasteurianus**

**General Genomic Features**

Although Acetobacter frequently associate with the intestines of Drosophila in the wild and in the lab, we are unaware of any whole-genome sequences of A. pasteurianus strains derived from the intestines of adult Drosophila. We consider this a significant deficit given the impacts of Acetobacter on Drosophila development and metabolism. To address this shortcoming, we completed a whole-genome sequence of an A. pasteurianus strain (A. pasteurianus AD) that we isolated from the intestines of wild-type Drosophila. For comparative purposes, we examined the available genomic sequences of the NBRC 101655 strain, and the ATCC 33445 strain. The NBRC 101655 strain was originally isolated from
pineapple in Thailand and successfully associates with adult *Drosophila* (Figure 2). In contrast, we found that the environmental ATCC 33445 isolate fails to survive within the intestine of adult *Drosophila* (Figure 2).

At first glance, we did not observe substantial differences between the respective genomes. Each genome is approximately 3 MB in length, with similar GC content, and similar numbers of RNA, and predicted coding sequences (Table 6). From an evolutionary perspective, *A. pasteurianus* AD appears more closely related to the NBRC 101655 strain than the ATCC 33445 strain (Table 6). Consistent with a greater evolutionary distance to the ATCC strain, we found that the ATCC 33445 genome encodes 112 unique proteins, while the NBRC 101655 and AD strains share 112 genes that are absent from the ATCC 33445 genome (Figure 3A).

**Environmental Response Factors**

As with *L. brevis*, we first compared the *Drosophila*-associated and environmental genomes for distinctions in gene products that control responses to environmental factors. Specifically, we looked at signaling factors (Figure 3B, supplementary table 7), stress response factors (Figure 3C, supplementary table 8), and virulence factors (Figure 3D, supplementary table 9). We also identified two-component systems, transcription factors, and other DNA binding factors in each genome (Table 7, and supplementary tables 10-12). Across this series of comparisons, the most pronounced differences were commensurate with a closer relationship of the AD strain to the NBRC 101655 strain than to the ATCC 33445 strain. Thus, this admittedly limited comparison does not appear to identify genomic components that readily distinguish *Drosophila*-associated *A. pasteurianus* genomes from environmental counterparts. Nonetheless, these functional characterizations uncover differences between the AD and NBRC 101655 *A. pasteurianus* genomes that associate with *Drosophila*, and the ATCC 33445 genome that fails to do so.
Function-Based Comparisons of Commensal and Reference Strains of *Acetobacter pasteurianus*

Our fortuitous identification of an environmental *A. pasteurianus* strain that fails to survive passage through the fly gut allowed us to explore *A. pasteurianus* genomes for factors that permit association with *Drosophila*. As the AD and NBRC 101655 strains successfully pass through the adult intestine, we reasoned that the AD and NBRC 101655 genomes encode biochemical functions absent from ATCC 33445 that permit association with *Drosophila*, or that the ATCC 33445 genome encodes biochemical functions absent from the other strains that prevent association with *Drosophila*. This prompted us to identify biological subsystems shared exclusively by the AD and NBRC 101655 (Table 9), or unique to the ATCC 33445 genome (Table 10).

In this comparative analysis, we noted four subsystems exclusive to the AD and NBCR 101655 genomes that may explain their ability to survive passage through the fly intestine. Both strains encode polyamine metabolism factors that are frequently associated with cellular growth and survival, and have established roles in the formation of biofilms (43). The association-competent genomes also encode factors necessary for the conversion of urea to ammonium and carbon dioxide. A similar system operates in *Helicobacter pylori* where it raises the gastric pH to generate a more hospitable environment for microbial survival (44). We also detected the redox-sensitive transcriptional activator SoxR in both association-competent genomes. SoxR promotes microbial survival within the intestine by countering the antibacterial actions of host-derived superoxide anions (45). Finally, we detected several genes that contribute to organic sulfur assimilation in association-competent genomes. These gene products allow *A. pasteurianus* AD and NBRC 101655 to use alkanesulfonates as a source of sulfur during sulfate or cysteine starvation and may provide both strains a competitive advantage in the intestine if sulfur is limiting.
The association-incompetent ATCC 33445 strain also encodes products that contribute to generation of ammonia. However, the ATCC 33445 strain relies on respiratory nitrate reductase and nitrite reductase to generate ammonia, as well as assimilatory nitrate reductase to access nitrate for metabolic growth. This represents an entirely different strategy to use nitrogen as a fuel for metabolic energy and growth. Thus, it is feasible that the ATCC strain fails to grow within the adult intestine due to unsatisfied nitrogen requirements. We also observed two toxin-antitoxin systems unique to the association-incompetent ATCC 33445 genome – an addiction module toxin that ensures propagation of plasmids to progeny cells (46), and a MazE/MazF type toxin-antitoxin (47). The MazE/MazF system induces programmed cell death in prokaryotic cells in response to stressful environments. It is intriguing to speculate that the adult intestine activates this response, thereby inducing microbial death and preventing ATCC 33445 survival within the gut.

*Lactobacillus plantarum*

**General Genomic Features**

We used the environmental ATCC 14917 strain as a reference for our comparative analyses of *L. plantarum* genomes. *L. plantarum* ATCC 14917 was isolated from pickled cabbage and we confirmed viable association of this strain with the adult *Drosophila* intestine. We compared the ATCC 14917 strain to four *Drosophila*-associated genomes – WJL, DMCS_001, DF and KP. WJL and DMCS_001 were isolated from *Drosophila* raised in geographically separate labs (48, 49). We isolated the KP strain from the intestines of our lab-raised wild-type strain, and the DF strain from an isofemale wild *Drosophila melanogaster* line that we captured in Edmonton, Canada in the summer of 2014. The DF and KP genomes encode one chromosome and three closely related plasmids each (Figure 4). While all five genomes are closely related, genome-to-genome distance calculators suggest a greater degree of identity among the *Drosophila*-associated KP, DF, WJL and DMCS_001 strains (Table 11). In general,
environmental genome is smaller than the *Drosophila*-associated genomes (Figure 5A), encodes fewer RNAs and coding sequences, and contains fewer phage-associated proteins (Table 11).

**Environmental Response Factors**

Contrary to our expectations, we did not observe pronounced differences between the genomes of environmental or *Drosophila*-associated strains in terms of modifiers of signal transduction, stress responses or virulence (Figure 5 B-D, and supplementary tables 13-15). Likewise, there is not a notable enrichment of two-component systems, transcription factors or other DNA binding proteins in *Drosophila*-associated strains of *L. plantarum* (Table 12, supplementary tables 16-18). In general, these findings suggest that *Drosophila*-associated strains of *L. plantarum* are not easily distinguished on the basis of functions that sense and respond to specific environmental cues.

**Prophages and CRISPR Responses**

Similar to our observations with *L. brevis* genomes, we observed a greater number of intact prophage genomes in *Drosophila*-associated *L. plantarum* strains than in the environmental strain (Table 13). Of all *Lactobacillus* genomes studied, we detected an average of four intact prophage genomes in *Lactobacillus* strains isolated from flies, and a maximum of two prophage genomes in environmental strains (Figure 6). The main difference between the *Drosophila*-associated *brevis* and *plantarum* strains is that the *plantarum* strains do not appear to encode CRISPR-dependent anti-phage defenses within their genomes.

**Function-Based Comparisons of Commensal and Reference Strains of *Lactobacillus plantarum***

We then used the whole-genome sequence data for environmental and *Drosophila*-associated strains to prepare an *L. plantarum* pangenome (Figure 7). Visual inspection indicated that three of the
Drosophila-associated genomes (DF, KP and WJL) display similar organization, while the DMCS_001 genome appears more akin to the environmental genome. When we looked for factors exclusively common to the DF, KP and WJL genomes we detected several gene products that derive from prophages. However, we also detected functional elements that are consistent with support for microbial growth in the intestinal lumen. This includes access to methionine, an essential amino acid in L. plantarum; enzymes for the regulation of sialic acid, a comparatively rare bacterial molecule that is typically found in prokaryotes that live in association with deuterostomes (50); and a regulator of the biosynthesis of lipoteichoic acid, a regulator of cell wall autolysis in gram-positive bacteria (51, 52).

Strikingly, we also detected uridine phosphorylase exclusively in the genomes of the Drosophila-associated DF, KP and WJL strains. Uridine phosphorylase catalyzes conversion of uridine to uracil, a secreted metabolite that causes microbial-dependent inflammation in the intestine of adult Drosophila (53). Combined, these genes constitute a comparatively small number of molecular functions that may favor the propagation and establishment of L. plantarum populations in the fly intestine.

Finally, we examined the thirty-five genes exclusively observed in the genomes of DF, KP, WJL and DMCS_001. The majority (nineteen) were prophage genes, and an additional five were hypothetical proteins of unknown function. Rather strikingly, several of the remaining genes encode products that actively suppress the growth of competing microbes. These include the PlnMNO operon that encodes a bacteriocin and cognate immunity protein (54), and 1,3-propanediol dehydrogenase, an enzyme that converts propane-1,3-diol to 3-hydroxypropanal. 3-hydroxypropanal is also known as reuterin, a Lactobacillus reuteri metabolite that exerts broad-spectrum microbicidal effects on intestinal microbes (55). Combined the actions of PlnMNO and reuterin give Drosophila-associated L. plantarum a competitive advantage against other microbes within the gut lumen.
DISCUSSION

The last decade witnessed a proliferation of elegant studies that uncovered critical host responses to microbial factors in the *Drosophila* intestine (Reviewed in (13)). Bacterial cues promote larval growth (32, 33), direct innate immune responses (56, 57), orchestrate the proliferation of intestinal stem cells (58, 59), and regulate the uptake and storage of nutrients (36). Despite the importance of the intestinal microflora for *Drosophila* health and development, we know remarkably little about the biochemical events that permit bacterial survival within the hostile terrain of a fly intestine. A recent study identified microbial metabolism and stress response pathways that mediate interactions between intestinal bacterial and their *Drosophila* host (49). In this study, we examined the genomes of *Drosophila*-associated strains of *L. brevis*, *L. plantarum*, and *A. pasteurianus*. We were particularly interested in the identification of candidate signal transduction, virulence and defense factors that permit bacterial survival within the intestines of adult flies. To this end, we compared fly-associated genomes to environmental strains of the same species. For each species, we observed a small number of genetic pathways that were exclusive to the *Drosophila*-associated genomes. Many of the *Drosophila*-associated pathways encode products with established roles in host-microbe interactions, suggesting these products facilitate association of *Drosophila* with the individual strains.

**Lactobacilli**

For our studies of *Lactobacillus* genomes, we prepared whole-genome sequences of *L. brevis* or *L. plantarum* strains that we isolated from lab-raised wild-type flies, and an *L. plantarum* strain that we isolated from a wild *Drosophila*. These genomes formed the cornerstones of a comparative study that included three previously reported *Drosophila*-associated genomes (37, 48, 49), and the genomes of environmental strains that successfully associate with the intestines of our wild-type *Drosophila*. In this manner, we identified bacterial functions that are unique to *Drosophila*-associated genomes of *L. brevis*
and *L. plantarum*. The functions fall into four broad categories – antibacterial, structural, metabolic, and phage-related.

The most striking feature common to all four *Drosophila*-associated *L. plantarum* genomes was the presence of broad-spectrum bactericidal factors. For example, the DF, KP, WJL and DMCS_001 genomes all encode a complete *PnlMNO* operon, which encodes a bacteriocin and a corresponding immunity protein (54). Bacteriocins are produced by many lactic acid bacteria to kill neighboring bacteria, while the immunity protein protects *L. plantarum* from collateral damage (60). In addition, the *Drosophila*-associated *L. plantarum* genomes encode the enzymatic capacity to generate 3-hydroxypropanal/reuterin, a bacterial toxin expressed by *L. reuteri* in the gut to suppress the growth of other commensals. Combined, these bactericidal molecules counter the growth of competing bacteria inside a *Drosophila* host, and favor expansion of *L. plantarum*. The competitive advantages conferred by the *PnlMNO* operon and 3-hydroxypropanal may explain why *L. plantarum* is frequently reported in studies that characterize the intestinal microflora of *Drosophila*.

The *Drosophila*-associated genomes of *L. plantarum* and *L. brevis* also encode structural components that stabilize associations with their fly host. In particular, we detected metabolic pathways for modifications to cell walls that permit host-microbe interactions and biofilm formation. These include the construction of exopolysaccharides by *L. brevis* and the regulation of sialic acid by *L. plantarum*. Sialic acid is a comparatively rare microbial metabolite, but has been observed on microbes that associate with deuterostomes. Bacteria use sialic acid as a nutrient, but they also use it to mask detection by host immune responses. While the role of sialic acid in *L. plantarum* association with *Drosophila* requires further investigation, we feel that these elements merit consideration as host-microbe interaction factors.

The *Drosophila*-associated genomes of *L. plantarum* and *brevis* also includes gene products that address nutritional requirements. These functions include access to limited resources such as
methionine by *L. plantarum*, and utilization of citrate as an energy source by *L. brevis*. We were particularly struck by the presence of pectin metabolism factors within the genomes of *Drosophila*-associated strains of *L. brevis*. Pectin is an excellent source of carbon for bacteria that grow on plants. However, bacterial utilization of pectin accelerates the ripening and decay of the same plants (61). Thus, *Drosophila*-associated *L. brevis* genomes express factors that contribute to the decay of organic substrates. This is particularly notable, as *Drosophila* preferably consumes decayed matter as a source of nutrients. The ability of *L. brevis* to generate meals for their *Drosophila* host provides a possible explanation for the fact that *Drosophila* frequently associate with *L. brevis*, even though intestinal bacteria do not appear to establish stable populations within the guts of adult *Drosophila*. As *L. brevis* generates palatable meals for fly hosts, we speculate that their chances of association with flies in the wild are greater than those for many other bacteria. This host-microbe relationship is similar to a proposed mechanism for association of *Erwinia carotovora* with *Drosophila* in the wild (62).

The final difference we noted between environmental and *Drosophila*-associated *Lactobacillus* genomes was an accumulation of temperate prophage genomes throughout *Drosophila*-associated *Lactobacilli*. Intestinal stresses such as high levels of reactive oxygen species are known to trigger lysogenic induction of temperate prophages (63). Thus, it is feasible that bacterial strains that pass though the fly intestine will release and integrate greater numbers of lytic prophages, explaining the increased numbers of prophage genomes in *Lactobacillus* strains that associate with adult *Drosophila*.

**Acetobacter**

In this study, we report the first genome of a *Drosophila*-associated strain of *A. pasteurianus*, and identified an *A. pasteurianus* strain that fails to survive passage through the fly gut. A comparison of association-competent and association-incompetent genomes is limited by the fact that only one *Drosophila*-associated genome is available for study. Nonetheless, our identification of an
environmental *A. pasteurianus* strain that fails to associate with *Drosophila* yields a comparatively short list of candidate functions that regulate association of *Acteobacter* with adult *Drosophila*. This list includes gene products that process nitrogen, and gene products that directly control the induction of cell death in *A. pasteurianus*. To our knowledge, ours is the first study to identify a set of bacterial factors that regulate association of an *Acetobacter* with *Drosophila*. We believe that future characterization of mutations in the respective gene products will identify the bacterial factors that control viability of *A. pasteurianus* in the adult intestine. This approach has considerable value given the relationship between *Acetobacter* and *Drosophila* development.

In summary, our comparative study of bacterial genomes uncovers a number of genetic signatures of association with *Drosophila*. As many of the gene products have established roles in host-microbe interactions, we propose that these genes include factors that promote the frequent association of *Drosophila* with *Lactobacillus* and *Acetobacter* strains. Future characterization of mutations in the individual products will reveal the relationship between the individual factors and host physiology.
CONCLUSION

The intestinal microflora influences a number of physiological outcomes in their animal host. Microbial factors shape the establishment and execution of immune responses, influence the availability of critical nutrients and often determine the progression of debilitating inflammatory illnesses. Studies in *Drosophila melanogaster* have provided valuable insights into the host cellular machinery that coordinates responses to microbial presence. However, we know remarkably little about the microbial factors that permit long-term association with the host. In this study, we characterized the genomes of common *Drosophila*-associated microbes and identified putative bacterial products that favor association with *Drosophila*. We believe that products constitute valuable starting points to examine the microbial half of the host-microbe relationship.
MATERIALS AND METHODS

Drosophila husbandry

All Drosophila assays were performed with virgin *w^1118* male and female flies raised on standard corn-meal medium (Nutri-Fly Bloomington Formulation, Genesee Scientific) in a humidified incubator at 29°C. To generate germ-free flies, we transferred freshly eclosed (0 – 16 hr old) adult flies to standard medium that we supplemented with an antibiotic (100 μg/ml ampicillin, 50 μg/ml vancomycin, 100 μg/ml neomycin and 100 μg/ml metronidazole dissolved in ethanol). This cocktail has been described previously (22). To generate gnotobiotic flies, we raised adult flies on the antibiotic cocktail for five days, starved flies for two hours, and transferred the flies to a vial containing an autoclaved fly vial cotton plug soaked with the respective bacteria resuspended in a sterile-filtered solution of 10% sucrose in PBS. We fed the flies the bacterial meal for 16 hours and then transferred the flies to vials of freshly autoclaved food. To confirm association, we periodically plated fly homogenates on bacterial medium selective for *Acetobacter* or *Lactobacilli*.

Bacterial Isolation and Sequencing

We plated homogenates of 15 day old adult *Drosophila* on Acetobacter and MRS culture plates. We found that *L. brevis* colonies are easily distinguished from *L. plantarum* colonies on MRS-agar medium. We isolated individual colonies of *A. pasteurianus*, *L. brevis* and the KP strain of *L. plantarum* and grew them statically at 29°C in liquid MRS (*L. brevis* and *L. plantarum*), or shaking in liquid (*A. pasteurianus*). The DF strain of *L. plantarum* was isolated from a wild, mated isofemale *Drosophila melanogaster* captured on a rotting strawberry in the kitchen of EF in Edmonton, Canada. Bacterial DNA was isolated with the Microbial DNA Isolation kit from MO BIO Laboratories Inc. (catalog number: 12224-250) according to their instructions. The genomes of *L. plantarum* strains DF and KP were sequenced and
assembled at the McGill University and Génome Québec Innovation Centre on the PacBio platform. The genomes of *A. pasteurianus* (strain AD) and *L. brevis* (strain EF) were sequenced at The Applied Genomics Core of the University of Alberta. For the latter genomes, we prepared Nextera XT libraries from the isolated microbial DNA according to Illumina’s protocol and sequenced the libraries with using the V3-600 cycle Kit (Illumina). Whole genome sequences were then assembled using Lasergene software (DNASTAR). This Whole Genome Shotgun project for *L. brevis* EF has been deposited at DDBJ/EMBL/GenBank under the accession LPXV00000000. The version described in this paper is version LPXV01000000. The Whole Genome Shotgun project for *A. pasteurianus* AD has been deposited at DDBJ/EMBL/GenBank under the accession LPWU00000000. The version described in this paper is version LPWU01000000. Chromosome1 of *L. plantarum* KP has been deposited at DDBJ/EMBL/GenBank under the accession CP013749 and plasmids 1-3 for the same strain have been deposited under the accession numbers CP013750, CP013751 and CP013752, respectively. Chromosome1 of *L. plantarum* DF has been deposited at DDBJ/EMBL/GenBank under the accession CP013753 and plasmids 1-3 for the same strain have been deposited under the accession numbers CP013754, CP013755 and CP013756, respectively.

**Genome Assembly and Annotation**

For each sequencing project, we confirmed the individual species with the SpeciesFinder 1.2 algorithm (64) and calculated genome to genome distances with the genome to genome distance calculator of the Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures (65). We annotated each genome with RAST (66), identified intact prophage genomes with the PHAST server (67) and identified CRISPR arrays with CRISPRFinder (68). We used the CRISPRTarget algorithm to predict CRISP targets for the individual CRISPR arrays (69). To identify regulatory proteins within the respective
genomes, we used the P2RP identifier (70). We used the GView tool to generate graphical representations of bacterial genomes (71).
ACKNOWLEDGEMENTS

This research was funded by a grant from the Canadian Institutes of Health Research to EF (MOP 77746). Next Generation Sequencing services were performed by The Applied Genomics Core (TAGC) at the Faculty of Medicine & Dentistry, University of Alberta.
**TABLES AND FIGURES**

<table>
<thead>
<tr>
<th>Bacteria</th>
<th>Strain</th>
<th>Isolated from <em>Drosophila</em></th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Lactobacillus brevis</em></td>
<td>ATCC 367</td>
<td>No</td>
<td>(72)</td>
</tr>
<tr>
<td><em>Lactobacillus brevis</em></td>
<td>EW</td>
<td>Yes</td>
<td>(37)</td>
</tr>
<tr>
<td><em>Lactobacillus brevis</em></td>
<td>EF</td>
<td>Yes</td>
<td>This study</td>
</tr>
<tr>
<td><em>Acetobacter pasteurianus</em></td>
<td>NBRC 101655</td>
<td>No</td>
<td>(73)</td>
</tr>
<tr>
<td><em>Acetobacter pasteurianus</em></td>
<td>ATCC 33445</td>
<td>No</td>
<td>(73)</td>
</tr>
<tr>
<td><em>Acetobacter pasteurianus</em></td>
<td>AD</td>
<td>Yes</td>
<td>This study</td>
</tr>
<tr>
<td><em>Lactobacillus plantarum</em></td>
<td>JOJTO1.1</td>
<td>No</td>
<td>(49)</td>
</tr>
<tr>
<td><em>Lactobacillus plantarum</em></td>
<td>ATCC 14917</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td><em>Lactobacillus plantarum</em></td>
<td>WJL</td>
<td>Yes</td>
<td>(48)</td>
</tr>
<tr>
<td><em>Lactobacillus plantarum</em></td>
<td>KP</td>
<td>Yes</td>
<td>This study</td>
</tr>
<tr>
<td><em>Lactobacillus plantarum</em></td>
<td>DF</td>
<td>Yes</td>
<td>This study</td>
</tr>
</tbody>
</table>

**Table 1:** Details on the eleven bacterial strains used in this study.
<table>
<thead>
<tr>
<th></th>
<th>L. brevis EF</th>
<th>L. brevis EW</th>
<th>L. brevis ATCC #367</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of contigs or scaffolds</td>
<td>32</td>
<td>38</td>
<td>3</td>
</tr>
<tr>
<td>Genome to genome distance (Prob. DDH &gt;= 70%)</td>
<td>100</td>
<td>98.3</td>
<td>89.92</td>
</tr>
<tr>
<td>Genome size</td>
<td>2,864,530</td>
<td>2,885,101</td>
<td>2,340,228</td>
</tr>
<tr>
<td>CG content (%)</td>
<td>45.3</td>
<td>45.3</td>
<td>46.1</td>
</tr>
<tr>
<td>Number of RNAs</td>
<td>80</td>
<td>82</td>
<td>79</td>
</tr>
<tr>
<td>Predicted CDS</td>
<td>2808</td>
<td>2830</td>
<td>2284</td>
</tr>
<tr>
<td>Assigned function</td>
<td>1961</td>
<td>1969</td>
<td>1786</td>
</tr>
<tr>
<td>Uncharacterized</td>
<td>13</td>
<td>13</td>
<td>8</td>
</tr>
<tr>
<td>Conserved hypothetical</td>
<td>11</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>Unknown function</td>
<td>31</td>
<td>31</td>
<td>32</td>
</tr>
<tr>
<td>Hypothetical</td>
<td>673</td>
<td>682</td>
<td>431</td>
</tr>
<tr>
<td>Phage-associated proteins</td>
<td>119</td>
<td>126</td>
<td>22</td>
</tr>
</tbody>
</table>

Table 2: Details on *Lactobacillus brevis* genomes described in this study

<table>
<thead>
<tr>
<th></th>
<th>L. brevis EF</th>
<th>L. brevis EW</th>
<th>L. brevis ATCC #367</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two-component systems</td>
<td>22</td>
<td>22</td>
<td>20</td>
</tr>
<tr>
<td>Transcription Factors</td>
<td>203</td>
<td>204</td>
<td>158</td>
</tr>
<tr>
<td>Other DNA binding proteins</td>
<td>14</td>
<td>16</td>
<td>18</td>
</tr>
</tbody>
</table>

Table 3: Organization of regulatory proteins within *Lactobacillus brevis* genomes described in this study

<table>
<thead>
<tr>
<th></th>
<th>L. brevis EF</th>
<th>L. brevis EW</th>
<th>L. brevis ATCC #367</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Prophages</td>
<td>4</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Identity of Prophages</td>
<td>3 X NC_019489 (Lactobacillus phage Sha1)</td>
<td>3 X NC_019489 (Lactobacillus phage Sha1)</td>
<td>NC_009813 (Listeria phage B054)</td>
</tr>
<tr>
<td></td>
<td>NC_027990 (Lactobacillus phage LBR48)</td>
<td>NC_027990 (Lactobacillus phage LBR48)</td>
<td>NC_009813 (Listeria phage B054)</td>
</tr>
<tr>
<td>Number of CRISPRs</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>CRISPR targets</td>
<td><em>Lactobacillus</em> phage LBR48</td>
<td><em>Lactobacillus</em> phage LBR48</td>
<td><em>Lactobacillus</em> plasmid sequences</td>
</tr>
</tbody>
</table>

Table 4: Prophage and CRISPR identification in *Lactobacillus brevis* genomes described in this study
<table>
<thead>
<tr>
<th>Category</th>
<th>Subcategory</th>
<th>Subsystem</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino Acids and Derivatives</td>
<td>Aromatic amino acids and derivatives</td>
<td>Chorismate: Intermediate for synthesis of Tryptophan, PAPA antibiotics, PABA, 3-hydroxyanthranilate and more.</td>
<td>Phosphoribosylantranilate isomerase (EC 5.3.1.24)</td>
</tr>
<tr>
<td>Amino Acids and Derivatives</td>
<td>Aromatic amino acids and derivatives</td>
<td>Common Pathway For Synthesis of Aromatic Compounds (DAHP synthase to chorismate)</td>
<td>Shikimate/quinate 5-dehydrogenase I beta (EC 1.1.1.282)</td>
</tr>
<tr>
<td>Carbohydrates</td>
<td>Aminosugars</td>
<td>Chitin and N-acetylglicosamine utilization</td>
<td>Chitinase (EC 3.2.1.14)</td>
</tr>
<tr>
<td>Carbohydrates</td>
<td>Monosaccharides</td>
<td>D-Galacturionate and D-Glucuronate Utilization</td>
<td>2-keto-3-deoxygluconate permease (KDG permease)</td>
</tr>
<tr>
<td>Carbohydrates</td>
<td>Monosaccharides</td>
<td>D-Galacturionate and D-Glucuronate Utilization</td>
<td>Polygalacturonase (EC 3.2.1.15)</td>
</tr>
<tr>
<td>Carbohydrates</td>
<td>Monosaccharides</td>
<td>D-Galacturionate and D-Glucuronate Utilization</td>
<td>Rhamnogalacturonides degradation protein RhiN</td>
</tr>
<tr>
<td>Carbohydrates</td>
<td>Organic acids</td>
<td>Citrate Metabolism, Transport, and Regulation</td>
<td>2-(5&amp;39;&amp;39; triphosphoribosyl)-3&amp;39; dephosphoocenzyme-A synthase (EC 2.7.8.25)</td>
</tr>
<tr>
<td>Carbohydrates</td>
<td>Organic acids</td>
<td>Citrate Metabolism, Transport, and Regulation</td>
<td>Apo-citrate lyase phosphoribosyl-dephospho-CoA transferase (EC 2.7.7.61)</td>
</tr>
<tr>
<td>Carbohydrates</td>
<td>Organic acids</td>
<td>Citrate Metabolism, Transport, and Regulation</td>
<td>Citrate lyase alpha chain (EC 4.1.3.6)</td>
</tr>
<tr>
<td>Carbohydrates</td>
<td>Organic acids</td>
<td>Citrate Metabolism, Transport, and Regulation</td>
<td>Citrate lyase beta chain (EC 4.1.3.6)</td>
</tr>
<tr>
<td>Carbohydrates</td>
<td>Organic acids</td>
<td>Citrate Metabolism, Transport, and Regulation</td>
<td>Citrate lyase gamma chain, acyl carrier protein (EC 4.1.3.6)</td>
</tr>
<tr>
<td>Carbohydrates</td>
<td>Organic acids</td>
<td>Citrate Metabolism, Transport, and Regulation</td>
<td>Citrate lyase transcriptional regulator Citl</td>
</tr>
<tr>
<td>Carbohydrates</td>
<td>Organic acids</td>
<td>Citrate Metabolism, Transport, and Regulation</td>
<td>Oxaloacetate decarboxylase involved in citrate fermentation (EC 4.1.1.3)</td>
</tr>
<tr>
<td>Carbohydrates</td>
<td>Organic acids</td>
<td>Citrate Metabolism, Transport, and Regulation</td>
<td>[Citrate [pro-35]-lyase] ligase (EC 6.2.1.22)</td>
</tr>
<tr>
<td>Carbohydrates</td>
<td>Sugar alcohols</td>
<td>Glycerol and Glycerol-3-phosphate Uptake and Utilization</td>
<td>GipG protein (membrane protein of glp regulon)</td>
</tr>
<tr>
<td>Cell Wall and Capsule</td>
<td>Capsular and extracellular polysaccharides</td>
<td>Exopolysaccharide Biosynthesis</td>
<td>Manganese-dependent protein-tyrosine phosphatase (EC 3.1.3.48)</td>
</tr>
<tr>
<td>Cell Wall and Capsule</td>
<td>Capsular and extracellular polysaccharides</td>
<td>Exopolysaccharide Biosynthesis</td>
<td>Tyrosine-protein kinase EpsD (EC 2.7.10.2)</td>
</tr>
<tr>
<td>Cell Wall and Capsule</td>
<td>Capsular and Exopolysaccharide</td>
<td>Exopolysaccharide Biosynthesis</td>
<td>Tyrosine-protein kinase</td>
</tr>
<tr>
<td>Category</td>
<td>Subcategory</td>
<td>Description</td>
<td>Enzyme or Protein</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>-------------</td>
<td>------------------------------------------------------------------------------</td>
<td>------------------------------------------------</td>
</tr>
<tr>
<td>Capsule</td>
<td>extracellular polysaccharides</td>
<td>Biosynthesis</td>
<td>transmembrane modulator EpsC</td>
</tr>
<tr>
<td>Cell Wall and Capsule</td>
<td>Capsular and extracellular polysaccharides</td>
<td>Exopolysaccharide Biosynthesis</td>
<td>Undecaprenyl-phosphate galactosephosphotransferase (EC 2.7.8.6)</td>
</tr>
<tr>
<td>Cell Wall and Capsule</td>
<td>Capsular and extracellular polysaccharides</td>
<td>Rhamnose containing glycan</td>
<td>Glucose-1-phosphate thymidylyltransferase (EC 2.7.7.24)</td>
</tr>
<tr>
<td>Cell Wall and Capsule</td>
<td>Capsular and extracellular polysaccharides</td>
<td>Rhamnose containing glycan</td>
<td>dTDP-4-dehydrorhamnose 3,5-epimerase (EC 5.1.3.13)</td>
</tr>
<tr>
<td>Cell Wall and Capsule</td>
<td>Capsular and extracellular polysaccharides</td>
<td>Rhamnose containing glycan</td>
<td>dTDP-4-dehydrorhamnos reductase (EC 1.1.1.133)</td>
</tr>
<tr>
<td>Cell Wall and Capsule</td>
<td>Capsular and extracellular polysaccharides</td>
<td>Rhamnose containing glycan</td>
<td>dTDP-glucose 4,6-dehydratase (EC 4.2.1.46)</td>
</tr>
<tr>
<td>Cell Wall and Capsule</td>
<td>Capsular and extracellular polysaccharides</td>
<td>dTDP-rhamnose synthesis</td>
<td>dTDP-rhamnosyl transferase RfbF (EC 2.--.----)</td>
</tr>
<tr>
<td>Cell Wall and Capsule</td>
<td>Gram-Positive cell wall components</td>
<td>Teichoic and lipoteichoic acids biosynthesis</td>
<td>Minor teichoic acid biosynthesis protein GgaB</td>
</tr>
<tr>
<td>Clustering-based subsystems</td>
<td>no subcategory</td>
<td>CBSS-479436.4.peg.472</td>
<td>ATP-utilizing enzyme of the PP-loop superfamily</td>
</tr>
<tr>
<td>Clustering-based subsystems</td>
<td>no subcategory</td>
<td>CBSS-479436.4.peg.472</td>
<td>FIG099352: hypothetical protein</td>
</tr>
<tr>
<td>Cofactors, Vitamins, Prosthetic Groups, Pigments</td>
<td>Quinone cofactors</td>
<td>Menaquinone and Phylloquinone Biosynthesis</td>
<td>2-heptaprenyl-1,4-naphthoquinone methyltransferase (EC 2.1.1.163)</td>
</tr>
<tr>
<td>DNA Metabolism</td>
<td>CRISPs</td>
<td>CRISPRs</td>
<td>CRISPR-associated protein Cas1</td>
</tr>
<tr>
<td>DNA Metabolism</td>
<td>CRISPs</td>
<td>CRISPRs</td>
<td>CRISPR-associated protein Cas2</td>
</tr>
<tr>
<td>DNA Metabolism</td>
<td>CRISPs</td>
<td>CRISPRs</td>
<td>CRISPR-associated protein, Csn1 family</td>
</tr>
<tr>
<td>DNA Metabolism</td>
<td>DNA uptake, competence</td>
<td>DNA processing cluster</td>
<td>DNA topoisomerase III (EC 5.99.1.2)</td>
</tr>
<tr>
<td>Fatty Acids, Lipids, and Isoprenoids</td>
<td>Phospholipids</td>
<td>Glycerolipid and Glycerophospholipid Metabolism in Bacteria</td>
<td>Aldehyde dehydrogenase (EC 1.2.1.3)</td>
</tr>
<tr>
<td>Nucleosides and Nucleotides</td>
<td>Detoxification</td>
<td>Nudix proteins (nucleoside triphosphate hydrolases)</td>
<td>NADH pyrophosphatase (EC 3.6.1.22)</td>
</tr>
<tr>
<td>Phages, Prophages, Transposable elements, Plasmids</td>
<td>Phages, Prophages</td>
<td>Phage capsid proteins</td>
<td>Phage capsid protein</td>
</tr>
<tr>
<td>--------------------------------------------------</td>
<td>-----------------</td>
<td>-----------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Phages, Prophages, Transposable elements, Plasmids</td>
<td>Phages, Prophages</td>
<td>Phage packaging machinery</td>
<td>Phage DNA packaging</td>
</tr>
<tr>
<td>Phages, Prophages, Transposable elements, Plasmids</td>
<td>Phages, Prophages</td>
<td>Phage packaging machinery</td>
<td>Phage portal protein</td>
</tr>
<tr>
<td>Phages, Prophages, Transposable elements, Plasmids</td>
<td>Phages, Prophages</td>
<td>Phage replication</td>
<td>Single stranded DNA-binding protein, phage-associated</td>
</tr>
<tr>
<td>Phages, Prophages, Transposable elements, Plasmids</td>
<td>Phages, Prophages</td>
<td>Phage tail fiber proteins</td>
<td>Phage tail fiber protein</td>
</tr>
<tr>
<td>Phages, Prophages, Transposable elements, Plasmids</td>
<td>Phages, Prophages</td>
<td>Phage tail proteins</td>
<td>Phage major tail protein</td>
</tr>
<tr>
<td>Phages, Prophages, Transposable elements, Plasmids</td>
<td>Phages, Prophages</td>
<td>Phage tail proteins</td>
<td>Phage tail assembly</td>
</tr>
<tr>
<td>Phages, Prophages, Transposable elements, Plasmids</td>
<td>Phages, Prophages</td>
<td>Phage tail proteins</td>
<td>Phage tail length tape-measure protein</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RNA Metabolism</th>
<th>no subcategory</th>
<th>Group II intron-associated genes</th>
<th>Retron-type RNA-directed DNA polymerase (EC 2.7.7.49)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regulation and Cell signaling</td>
<td>no subcategory</td>
<td>cAMP signaling in bacteria</td>
<td>Prophage Clp protease-like protein</td>
</tr>
<tr>
<td>Regulation and Cell signaling</td>
<td>no subcategory</td>
<td>cAMP signaling in bacteria</td>
<td>cAMP-binding proteins - catabolite gene activator and regulatory subunit of cAMP-</td>
</tr>
</tbody>
</table>
Table 5: Function-based identification of systems unique to commensal strains of *Lactobacillus brevis*.

<table>
<thead>
<tr>
<th>Dependent protein kinases</th>
</tr>
</thead>
</table>
| **Sulfur Metabolism**     | no subcategory  
|                           | Galactosylceramide and  
|                           | Sulfatide metabolism  
| **Virulence, Disease and Defense** |  
|                           | Resistance to  
|                           | antibiotics and  
|                           | toxic compounds  
|                           | Beta-galactosidase (EC 3.2.1.23)  
|                           | Beta-lactamase (EC 3.5.2.6)  

The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY 4.0 International license.
### Table 6: Details on *Acetobacter pasteurianus* genomes described in this study

<table>
<thead>
<tr>
<th></th>
<th><em>A. pasteurianus AD</em></th>
<th><em>A. pasteurianus ATCC 33445</em></th>
<th><em>A. pasteurianus NBRC 101655</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of contigs or scaffolds</td>
<td>161</td>
<td>306</td>
<td>294</td>
</tr>
<tr>
<td>Genome to genome distance (Prob. DDH &gt;= 70%)</td>
<td>100</td>
<td>90.67</td>
<td>98.3</td>
</tr>
<tr>
<td>Genome size</td>
<td>2,830,055</td>
<td>2,888,200</td>
<td>3,018,312</td>
</tr>
<tr>
<td>CG content (%)</td>
<td>52.7</td>
<td>53.1</td>
<td>52.7</td>
</tr>
<tr>
<td>Number of RNAs</td>
<td>42</td>
<td>44</td>
<td>42</td>
</tr>
<tr>
<td>Predicted CDS</td>
<td>2673</td>
<td>2797</td>
<td>2834</td>
</tr>
<tr>
<td>Assigned function</td>
<td>1845</td>
<td>1866</td>
<td>1923</td>
</tr>
<tr>
<td>Uncharacterized</td>
<td>10</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>Conserved hypothetical</td>
<td>8</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>Unknown function</td>
<td>12</td>
<td>15</td>
<td>11</td>
</tr>
<tr>
<td>Hypothetical</td>
<td>770</td>
<td>854</td>
<td>852</td>
</tr>
<tr>
<td>Phage-associated proteins</td>
<td>28</td>
<td>44</td>
<td>30</td>
</tr>
</tbody>
</table>

### Table 7: Organization of regulatory proteins within *Acetobacter pasteurianus* genomes described in this study

<table>
<thead>
<tr>
<th></th>
<th><em>A. pasteurianus AD</em></th>
<th><em>A. pasteurianus ATCC 33445</em></th>
<th><em>A. pasteurianus NBRC 101655</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Two-component systems</td>
<td>28</td>
<td>26</td>
<td>28</td>
</tr>
<tr>
<td>Transcription Factors</td>
<td>105</td>
<td>100</td>
<td>127</td>
</tr>
<tr>
<td>Other DNA binding proteins</td>
<td>19</td>
<td>15</td>
<td>20</td>
</tr>
</tbody>
</table>

### Table 8: Prophage and CRISPR identification in *Acetobacter pasteurianus* genomes described in this study

<table>
<thead>
<tr>
<th></th>
<th><em>A. pasteurianus AD</em></th>
<th><em>A. pasteurianus ATCC 33445</em></th>
<th><em>A. pasteurianus NBRC 101655</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Prophages</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Number of CRISPRs</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

### Table 9: Phenotypic and metabolic properties

<table>
<thead>
<tr>
<th>Category</th>
<th>Subcategory</th>
<th>Subsystem</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino Acids and Derivatives</td>
<td>Alanine, serine, and glycine</td>
<td>Glycine Biosynthesis</td>
<td>Low-specificity L-threonine aldolase (EC 4.1.2.5)</td>
</tr>
<tr>
<td>Category</td>
<td>Example</td>
<td>Process/Function</td>
<td>EC Number</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>Amino Acids and Derivatives</td>
<td>Arginine; urea cycle, polyamines</td>
<td>Polyamine Metabolism</td>
<td>4-aminobutyraldehyde dehydrogenase (EC 1.2.1.19)</td>
</tr>
<tr>
<td>Amino Acids and Derivatives</td>
<td>Arginine; urea cycle, polyamines</td>
<td>Polyamine Metabolism</td>
<td>Spermidine Putrescine ABC transporter permease component PotB (TC 3.A.1.11.1)</td>
</tr>
<tr>
<td>Amino Acids and Derivatives</td>
<td>Arginine; urea cycle, polyamines</td>
<td>Polyamine Metabolism</td>
<td>Spermidine Putrescine ABC transporter permease component potC (TC_3.A.1.11.1)</td>
</tr>
<tr>
<td>Clustering-based subsystems</td>
<td>Clustering-based subsystems</td>
<td>CBSS-292415.3.peg.2341</td>
<td>Major facilitator superfamily (MFS) transporter</td>
</tr>
<tr>
<td>Clustering-based subsystems</td>
<td>no subcategory</td>
<td>Conserved gene cluster associated with Met-tRNA formyltransferase</td>
<td>16S rRNA (cytosine(967)-C(5))-methyltransferase (EC 2.1.1.176)</td>
</tr>
<tr>
<td>Cofactors, Vitamins, Prosthetic Groups, Pigments</td>
<td>Biotin</td>
<td>Biotin biosynthesis</td>
<td>Long-chain-fatty-acid--CoA ligase (EC 6.2.1.3)</td>
</tr>
<tr>
<td>Cofactors, Vitamins, Prosthetic Groups, Pigments</td>
<td>Folate and pterines</td>
<td>5-FCL-like protein</td>
<td>Butyryl-CoA dehydrogenase (EC 1.3.8.1)</td>
</tr>
<tr>
<td>Cofactors, Vitamins, Prosthetic Groups, Pigments</td>
<td>Tetrapyrroles</td>
<td>Cobalamin synthesis</td>
<td>Cobalamin synthase (EC 2.7.8.26)</td>
</tr>
<tr>
<td>DNA Metabolism</td>
<td>DNA repair</td>
<td>DNA repair, bacterial UvrD and related helicases</td>
<td>ATP-dependent DNA helicase UvrD/PcrA, proteobacterial paralog</td>
</tr>
<tr>
<td>DNA Metabolism</td>
<td>no subcategory</td>
<td>DNA structural proteins, bacterial</td>
<td>DNA-binding protein HBsu</td>
</tr>
<tr>
<td>Iron acquisition and metabolism</td>
<td>no subcategory</td>
<td>Hemin transport system</td>
<td>Outer membrane receptor proteins, mostly Fe transport</td>
</tr>
<tr>
<td>Membrane Transport</td>
<td>Cation transporters</td>
<td>Magnesium transport</td>
<td>Mg(2+) transport ATPase protein C</td>
</tr>
<tr>
<td>Membrane Transport</td>
<td>Cation transporters</td>
<td>Transport of Nickel and Cobalt</td>
<td>HoxN/HupN/NixA family nickel/cobalt transporter</td>
</tr>
<tr>
<td>Metabolism of Aromatic Compounds</td>
<td>Peripheral pathways for catabolism of aromatic compounds</td>
<td>Benzoate degradation</td>
<td>Benzoate 1,2-dioxygenase alpha subunit (EC 1.14.12.10)</td>
</tr>
<tr>
<td>Metabolism of Aromatic Compounds</td>
<td>Peripheral pathways for catabolism of aromatic compounds</td>
<td>Benzoate degradation</td>
<td>Benzoate 1,2-dioxygenase beta subunit (EC 1.14.12.10)</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>-------------------------------------------------------</td>
<td>---------------------</td>
<td>------------------------------------------------------</td>
</tr>
<tr>
<td>Metabolism of Aromatic Compounds</td>
<td>no subcategory</td>
<td>Aromatic Amin Catabolism</td>
<td>Nitrilotriacetate monooxygenase component B (EC 1.14.13.-)</td>
</tr>
<tr>
<td>Metabolism of Aromatic Compounds</td>
<td>no subcategory</td>
<td>Aromatic Amin Catabolism</td>
<td>Phenylacetaldehyde dehydrogenase (EC 1.2.1.39)</td>
</tr>
<tr>
<td>Phosphorus Metabolism</td>
<td>no subcategory</td>
<td>Phosphate metabolism</td>
<td>Soluble pyridine nucleotide transhydrogenase (EC 1.6.1.1)</td>
</tr>
<tr>
<td>Protein Metabolism</td>
<td>Protein biosynthesis</td>
<td>tRNA aminoacylation, Pro</td>
<td>tRNA proofreading protein STM4549</td>
</tr>
<tr>
<td>Protein Metabolism</td>
<td>Protein processing and modification</td>
<td>G3E family of P-loop GTPases (metallocenter biosynthesis)</td>
<td>Urease accessory protein UreD</td>
</tr>
<tr>
<td>Protein Metabolism</td>
<td>Protein processing and modification</td>
<td>G3E family of P-loop GTPases (metallocenter biosynthesis)</td>
<td>Urease accessory protein UreE</td>
</tr>
<tr>
<td>Protein Metabolism</td>
<td>Protein processing and modification</td>
<td>G3E family of P-loop GTPases (metallocenter biosynthesis)</td>
<td>Urease accessory protein UreF</td>
</tr>
<tr>
<td>Protein Metabolism</td>
<td>Protein processing and modification</td>
<td>G3E family of P-loop GTPases (metallocenter biosynthesis)</td>
<td>Urease accessory protein UreG</td>
</tr>
<tr>
<td>Protein Metabolism</td>
<td>Protein processing and modification</td>
<td>G3E family of P-loop GTPases (metallocenter biosynthesis)</td>
<td>Urease alpha subunit (EC 3.5.1.5)</td>
</tr>
<tr>
<td>Protein Metabolism</td>
<td>Protein processing and modification</td>
<td>G3E family of P-loop GTPases (metallocenter biosynthesis)</td>
<td>Urease beta subunit (EC 3.5.1.5)</td>
</tr>
<tr>
<td>Protein Metabolism</td>
<td>Protein processing and modification</td>
<td>G3E family of P-loop GTPases (metallocenter biosynthesis)</td>
<td>Urease gamma subunit (EC 3.5.1.5)</td>
</tr>
<tr>
<td>Respiration</td>
<td>Electron accepting reactions</td>
<td>Anaerobic respiratory reductases</td>
<td>Vanillate O-demethylase oxidoreductase (EC 1.14.13.-)</td>
</tr>
<tr>
<td>Category</td>
<td>Subcategory</td>
<td>Subsystem</td>
<td>Role</td>
</tr>
<tr>
<td>---------------------------</td>
<td>-----------------</td>
<td>----------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Respiration</td>
<td>no subcategory</td>
<td>Carbon monoxide dehydrogenase maturation factors</td>
<td>Aerobic carbon monoxide dehydrogenase molybdenum cofactor insertion protein CoxF</td>
</tr>
<tr>
<td>RNA Metabolism</td>
<td>no subcategory</td>
<td>Group II intron-associated genes</td>
<td>Retron-type RNA-directed DNA polymerase (EC 2.7.7.49)</td>
</tr>
<tr>
<td>Stress Response</td>
<td>Oxidative stress</td>
<td>Oxidative stress</td>
<td>Redox-sensitive transcriptional activator SoxR</td>
</tr>
<tr>
<td>Sulfur Metabolism</td>
<td>Organic sulfur</td>
<td>Alkanesulfonate assimilation</td>
<td>Alkanesulfonate monoxygenase (EC 1.14.14.5)</td>
</tr>
<tr>
<td>Sulfur Metabolism</td>
<td>Organic sulfur</td>
<td>Alkanesulfonate assimilation</td>
<td>Alkanesulfonates ABC transporter ATP-binding protein</td>
</tr>
<tr>
<td>Sulfur Metabolism</td>
<td>Organic sulfur</td>
<td>Alkanesulfonate assimilation</td>
<td>Alkanesulfonates-binding protein</td>
</tr>
</tbody>
</table>

**Table 9:** Identification of RAST Subsystems absent from the ATCC 33445 strain of *Acetobacter pasteurianus.*
<table>
<thead>
<tr>
<th>subsystems</th>
<th>subcategory</th>
<th>Description</th>
<th>Protein or Reaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cofactors, Vitamins, Prosthetic Groups, Pigments</td>
<td>Riboflavin, FMN, FAD</td>
<td>Riboflavin, FMN and FAD metabolism</td>
<td>3,4-dihydroxy-2-butanoic acid 4-phosphate synthase (EC 4.1.99.12)</td>
</tr>
<tr>
<td>Cofactors, Vitamins, Prosthetic Groups, Pigments</td>
<td>Tetrapyrroles</td>
<td>Cobalamin synthesis</td>
<td>Cobalamin biosynthesis protein CbiG</td>
</tr>
<tr>
<td>DNA Metabolism</td>
<td>no subcategory</td>
<td>DNA structural proteins, bacterial</td>
<td>DNA-binding protein HU</td>
</tr>
<tr>
<td>DNA Metabolism</td>
<td>no subcategory</td>
<td>Restriction-Modification System</td>
<td>Type I restriction-modification system, specificity subunit S (EC 3.121.3)</td>
</tr>
<tr>
<td>Fatty Acids, Lipids, and Isoprenoids</td>
<td>Fatty acids</td>
<td>Fatty Acid Biosynthesis FASII</td>
<td>Enoyl-[acyl-carrier-protein] reductase [NADPH] (EC 1.3.1.10)</td>
</tr>
<tr>
<td>Membrane Transport</td>
<td>no subcategory</td>
<td>Ton and Tol transport systems</td>
<td>TolA protein</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>no subcategory</td>
<td>Broadly distributed proteins not in subsystems</td>
<td>Putative oxidoreductase YncB</td>
</tr>
<tr>
<td>Nitrogen Metabolism</td>
<td>Denitrification</td>
<td>Denitrifying reductase gene clusters</td>
<td>Respiratory nitrate reductase alpha chain (EC 1.7.99.4)</td>
</tr>
<tr>
<td>Nitrogen Metabolism</td>
<td>Denitrification</td>
<td>Denitrifying reductase gene clusters</td>
<td>Respiratory nitrate reductase beta chain (EC 1.7.99.4)</td>
</tr>
<tr>
<td>Nitrogen Metabolism</td>
<td>Denitrification</td>
<td>Denitrifying reductase gene clusters</td>
<td>Respiratory nitrate reductase delta chain (EC 1.7.99.4)</td>
</tr>
<tr>
<td>Nitrogen Metabolism</td>
<td>Denitrification</td>
<td>Denitrifying reductase gene clusters</td>
<td>Respiratory nitrate reductase gamma chain (EC 1.7.99.4)</td>
</tr>
<tr>
<td>Nitrogen Metabolism</td>
<td>no subcategory</td>
<td>Nitrate and nitrite ammonification</td>
<td>Assimilatory nitrate reductase large subunit (EC:1.7.99.4)</td>
</tr>
<tr>
<td>Nitrogen Metabolism</td>
<td>no subcategory</td>
<td>Nitrate and nitrite ammonification</td>
<td>Nitrate ABC transporter, nitrate-binding protein</td>
</tr>
<tr>
<td>Nitrogen Metabolism</td>
<td>no subcategory</td>
<td>Nitrate and nitrite ammonification</td>
<td>Nitrate/nitrite transporter</td>
</tr>
<tr>
<td>Nitrogen Metabolism</td>
<td>no subcategory</td>
<td>Nitrate and nitrite ammonification</td>
<td>Nitrite reductase [NAD(P)H] large subunit (EC 1.7.1.4)</td>
</tr>
<tr>
<td>Nitrogen Metabolism</td>
<td>no subcategory</td>
<td>Nitrate and nitrite ammonification</td>
<td>Response regulator NasT</td>
</tr>
<tr>
<td>Phages, Prophages, Transposable elements, Plasmids</td>
<td>Phages, Prophages</td>
<td>Phage tail proteins</td>
<td>Phage tail length tape-measure protein</td>
</tr>
<tr>
<td>Phages, Prophages</td>
<td>Phages, Prophages</td>
<td>Phage tail proteins</td>
<td>Phage tail tube protein</td>
</tr>
<tr>
<td>Section</td>
<td>Protein Metabolism</td>
<td>Regulation and Cell Signaling</td>
<td>RNA Metabolism</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>--------------------</td>
<td>--------------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Transposable elements, Plasmids</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phages, Prophages, Transposable elements, Plasmids</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein Metabolism</td>
<td>tRNAs</td>
<td>Phd-Doc, YdcE-YdcD toxin-antitoxin (programmed cell death) systems</td>
<td></td>
</tr>
<tr>
<td>Protein Metabolism</td>
<td>tRNAs</td>
<td>Phd-Doc, YdcE-YdcD toxin-antitoxin (programmed cell death) systems</td>
<td></td>
</tr>
<tr>
<td>Protein Metabolism</td>
<td>N-linked Glycosylation in Bacteria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein Metabolism</td>
<td></td>
<td>Prevent host death protein, Phd antitoxin</td>
<td></td>
</tr>
<tr>
<td>Protein Metabolism</td>
<td></td>
<td>Programmed cell death antitoxin MazE like</td>
<td></td>
</tr>
<tr>
<td>Protein Metabolism</td>
<td></td>
<td>Programmed cell death toxin MazF like</td>
<td></td>
</tr>
<tr>
<td>Protein Metabolism</td>
<td></td>
<td>Toxin-antitoxin replicon stabilization systems</td>
<td></td>
</tr>
<tr>
<td>Protein Metabolism</td>
<td>ATP-dependent RNA helicases, bacterial</td>
<td>RelE/StbE replicon stabilization toxin</td>
<td></td>
</tr>
<tr>
<td>Protein Metabolism</td>
<td></td>
<td>Cold-shock DEAD-box protein A</td>
<td></td>
</tr>
<tr>
<td>Protein Metabolism</td>
<td>Nitric-oxide reductase (EC 1.7.99.7), quinol-dependent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein Metabolism</td>
<td>Organic sulfur assimilation</td>
<td>Alkanesulfonate assimilation</td>
<td></td>
</tr>
<tr>
<td>Protein Metabolism</td>
<td>Resistance to antibiotics and toxic compounds</td>
<td>Arsenic resistance</td>
<td></td>
</tr>
<tr>
<td>Protein Metabolism</td>
<td>Arsenic efflux pump protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Virulence, Disease and Defense</td>
<td>Resistance to antibiotics and toxic compounds</td>
<td>Arsenic resistance</td>
<td>Arsenic resistance protein ArsH</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----------------------------------------------</td>
<td>-------------------</td>
<td>-------------------------------</td>
</tr>
</tbody>
</table>

**Table 10**: Identification of RAST Subsystems exclusive to the ATCC 33445 strain of *Acetobacter pasteurianus*. 
<table>
<thead>
<tr>
<th></th>
<th>L. plant KP</th>
<th>L. plant DF</th>
<th>L. plant WJL</th>
<th>L. plant JOJT01.1</th>
<th>L. plant ATCC 14917</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of contigs or scaffolds</td>
<td>4</td>
<td>4</td>
<td>102</td>
<td>83</td>
<td>36</td>
</tr>
<tr>
<td>Genome to genome distance (Prob. DDH &gt;= 70%)</td>
<td>100</td>
<td>98.22</td>
<td>97.74</td>
<td>97.04</td>
<td>96.99</td>
</tr>
<tr>
<td>Genome size</td>
<td>3,692,742</td>
<td>3,697,306</td>
<td>3,477,495</td>
<td>3,194,687</td>
<td>3,198,761</td>
</tr>
<tr>
<td>CG content (%)</td>
<td>44</td>
<td>44.5</td>
<td>44.2</td>
<td>44.5</td>
<td>44.5</td>
</tr>
<tr>
<td>Number of RNAs</td>
<td>104</td>
<td>104</td>
<td>85</td>
<td>65</td>
<td>65</td>
</tr>
<tr>
<td>Predicted CDS</td>
<td>3569</td>
<td>3574</td>
<td>3365</td>
<td>3063</td>
<td>3061</td>
</tr>
<tr>
<td>Assigned function</td>
<td>2403</td>
<td>2400</td>
<td>2344</td>
<td>2227</td>
<td>2246</td>
</tr>
<tr>
<td>Uncharacterized</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Conserved hypothetical</td>
<td>13</td>
<td>12</td>
<td>10</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>Unknown function</td>
<td>72</td>
<td>72</td>
<td>74</td>
<td>69</td>
<td>70</td>
</tr>
<tr>
<td>Hypothetical</td>
<td>884</td>
<td>894</td>
<td>797</td>
<td>694</td>
<td>668</td>
</tr>
<tr>
<td>Phage-associated proteins</td>
<td>180</td>
<td>188</td>
<td>131</td>
<td>54</td>
<td>57</td>
</tr>
</tbody>
</table>

**Table 11:** Details on *Lactobacillus plantarum* genomes described in this study

<table>
<thead>
<tr>
<th></th>
<th>L. plant KP</th>
<th>L. plant DF</th>
<th>L. plant WJL</th>
<th>L. plant JOJT01.1</th>
<th>L. plant ATCC 14917</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two-component systems</td>
<td>27</td>
<td>27</td>
<td>29</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>Transcription Factors</td>
<td>252</td>
<td>252</td>
<td>244</td>
<td>230</td>
<td>230</td>
</tr>
<tr>
<td>Other DNA binding proteins</td>
<td>20</td>
<td>20</td>
<td>18</td>
<td>13</td>
<td>13</td>
</tr>
</tbody>
</table>

**Table 12:** Organization of regulatory proteins within *Lactobacillus plantarum* genomes described in this study
<table>
<thead>
<tr>
<th></th>
<th>L. plant KP</th>
<th>L. plant DF</th>
<th>L. plant WJL</th>
<th>L. plant JOJT01.1</th>
<th>L. plant ATCC 14917</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of Prophages</strong></td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td><strong>Identity of Prophages</strong></td>
<td>2 X NC_00435 (<em>Lactobacillus</em> phage)</td>
<td>2 X NC_00435 (<em>Lactobacillus</em> phage)</td>
<td>NC_00435 (<em>Lactobacillus</em> phage)</td>
<td>NC_009812 (<em>Lactobacillus</em> phage)</td>
<td>NC_005355 (<em>Lactobacillus</em> phage)</td>
</tr>
<tr>
<td></td>
<td>NC_023559 (<em>Oenococcus</em> phage)</td>
<td>NC_023559 (<em>Oenococcus</em> phage)</td>
<td>NC_023560 (<em>Oenococcus</em> phage)</td>
<td>NC_019489 (<em>Lactobacillus</em> phage)</td>
<td>NC_019489 (<em>Lactobacillus</em> phage)</td>
</tr>
<tr>
<td></td>
<td>2 X NC_019489 (<em>Lactobacillus</em> phage)</td>
<td>3 X NC_019489 (<em>Lactobacillus</em> phage)</td>
<td>NC_019489 (<em>Lactobacillus</em> phage)</td>
<td></td>
<td>NC_024384 (<em>Lactobacillus</em> phage)</td>
</tr>
</tbody>
</table>

| **Number of CRISPRs**        | 0           | 0           | 0            | 0                  | 0                   |
| **CRISPR targets**           | 0           | 0           | 0            | 0                  | 0                   |

**Table 13:** Prophage and CRISPR identification in *Lactobacillus plantarum* genomes described in this study

<table>
<thead>
<tr>
<th><strong>GENE</strong></th>
<th><strong>FUNCTION</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>prophage Lp2 protein 24</td>
<td>Phage</td>
</tr>
<tr>
<td>Prophage Lp1 protein 6</td>
<td>Phage</td>
</tr>
<tr>
<td>prophage Lp1 protein 23</td>
<td>Phage</td>
</tr>
<tr>
<td>prophage Lp1 protein 10</td>
<td>Phage</td>
</tr>
<tr>
<td>Integrase, phage associated</td>
<td>Phage</td>
</tr>
<tr>
<td>prophage Lp1 protein 17</td>
<td>Phage</td>
</tr>
<tr>
<td>prophage Lp2 protein 58</td>
<td>Phage</td>
</tr>
<tr>
<td>prophage Lp2 protein 13</td>
<td>Phage</td>
</tr>
<tr>
<td>Phage tail fibers</td>
<td>Phage</td>
</tr>
<tr>
<td>Lj965 prophage protein</td>
<td>Phage</td>
</tr>
<tr>
<td>prophage Lp2 protein 57</td>
<td>Phage</td>
</tr>
<tr>
<td>Phage tail fiber protein</td>
<td>Phage</td>
</tr>
<tr>
<td>prophage Lp1 protein 20</td>
<td>Phage</td>
</tr>
<tr>
<td>prophage Lp2 protein 12</td>
<td>Phage</td>
</tr>
<tr>
<td>prophage Lp1 protein 21</td>
<td>Phage</td>
</tr>
<tr>
<td>prophage Lp1 protein 19</td>
<td>Phage</td>
</tr>
<tr>
<td>Phage transcriptional activator RinA</td>
<td>Phage</td>
</tr>
<tr>
<td>prophage Lp2 protein 31</td>
<td>Phage</td>
</tr>
<tr>
<td><strong>Recombinational DNA repair protein</strong> RecT (prophage associated)</td>
<td>Phage</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>transposase, fragment (putative)</td>
<td></td>
</tr>
<tr>
<td>plantaricin biosynthesis protein PlnO</td>
<td>Antibacterial activity to kill inhibit growth of other bacteria</td>
</tr>
<tr>
<td>immunity protein PlnM</td>
<td>Antibacterial activity to kill inhibit growth of other bacteria</td>
</tr>
<tr>
<td>bacteriocin precursor peptide PlnN (putative)</td>
<td>Antibacterial activity to kill inhibit growth of other bacteria</td>
</tr>
<tr>
<td>1,3-propanediol dehydrogenase (EC 1.1.1.202)</td>
<td>Converts propane-1,3-diol to 3 hydroxypropanal, an antimicrobial produced by <em>L. reuteri</em></td>
</tr>
<tr>
<td>tRNA-Sup-TTA</td>
<td>Suppressor tRNA blocks termination of translation at UUA codon</td>
</tr>
<tr>
<td>tRNA-SeC-TCA</td>
<td>Incorporation of selenocysteine into nascent polypeptide chains at UCA codon</td>
</tr>
<tr>
<td>DNA replication protein</td>
<td></td>
</tr>
<tr>
<td>Stage V sporulation protein whose disruption leads to the production of immature spores (SpoVK)</td>
<td></td>
</tr>
<tr>
<td>phosphoglycerate mutase family protein</td>
<td>Eight enzyme in glycolysis</td>
</tr>
<tr>
<td>Biotin--protein ligase (EC 6.3.4.9, EC 6.3.4.10, EC 6.3.4.11, EC 6.3.4.15) / Biotin operon repressor</td>
<td></td>
</tr>
<tr>
<td><strong>FIG00748876: hypothetical protein</strong></td>
<td>Unknown</td>
</tr>
<tr>
<td><strong>FIG00754769: hypothetical protein</strong></td>
<td>Unknown</td>
</tr>
<tr>
<td><strong>FIG00743554: hypothetical protein</strong></td>
<td>Unknown</td>
</tr>
<tr>
<td><strong>FIG00745097: hypothetical protein</strong></td>
<td>Unknown</td>
</tr>
<tr>
<td><strong>FIG00750191: hypothetical protein</strong></td>
<td>Unknown</td>
</tr>
</tbody>
</table>

**Table 14:** Identification of RAST Subsystems unique to *Drosophila*-associated strains of *Lactobacillus plantarum*. 
**Figure 1:** Distribution of unique gene functions in the genomes of *L. brevis* strains EW, EF and ATCC 367. All data are based on gene function annotations within RAST and exclude gene products with unknown functions.
**Figure 2:** Quantification of *A. pasteurianus* association with conventionally reared (column 1) flies, germ-free (column 2) flies, gnotobiotic flies that were fed *A. pasteurianus* strain AD at OD600 of 50 and 200, respectively (columns 3 and 4), or gnotobiotic flies that were fed *A. pasteurianus* strain ATCC 33445 at OD600 of 50 and 200, respectively (columns 5 and 6). Each column shows the results of three separate measurements, and association was measure as bacterial colony-forming units per fly.
Figure 3: Distribution of unique gene functions in the genomes of *A. pasteurianus* strains AD, ATCC 33445 and NBRC 101655. All data are based on gene function annotations within RAST and exclude gene products with unknown functions.
Figure 4: Distribution of unique gene functions in the genomes of *L. plantarum* strains KP, DF, WJL, JOJT01 and ATCC 14917. All data are based on gene function annotations within RAST and exclude gene products with unknown functions.
Figure 5: Illustrations of the genomes for *L. plantarum* strains KP and DF. GC skew is indicated in purple, and GC content is indicated in black. All positive strand ORFs are shown in blue, and negative strand ORFs are shown in yellow.
Figure 6: Numbers of intact prophage genomes associated with *L. brevis* genomes (open circles) and *L. plantarum* genomes (closed circles) described in this study.
Figure 7: The inner wheel shows a pangenome for all *L. plantarum* strains described in this study, with the corresponding genomes aligned in outer wheels.
REFERENCES


