MICROBIOME

A conserved bacterial genetic basis for commensal-host specificity

Karina Gutiérrez-García[1]†, Kevin Aumiller[1,2]†, Ren Dodge[1], Benjamin Obadia[1], Ann Deng[2], Sneha Agrawal[2], Xincheng Yuan[2], Richard Wolff[3], Haolong Zhu[1,2], Ru-Ching Hsia[1], Nandita Garud[3,4], William B. Ludington[1,2]*

Animals selectively acquire specific symbiotic gut bacteria from their environments that aid host fitness. To colonize, a symbiont must locate its niche and sustain growth within the gut. Adhesins are bacterial cell surface proteins that facilitate attachment to host tissues and are often virulence factors for opportunistic pathogens. However, the attachments are often transient and nonspecific, and additional mechanisms are required to sustain infection. In this work, we use live imaging of individual symbiotic bacterial cells colonizing the gut of living *Drosophila melanogaster* to show that *Lactiplantibacillus plantarum* specifically recognizes the fruit fly foregut as a distinct physical niche. *L. plantarum* establishes stably within its niche through host-specific adhesins encoded by genes carried on a colonization island. The adhesin binding domains are conserved throughout the Lactobacillales, and the island also encodes a secretion system widely conserved among commensal and pathogenic bacteria.

Bacterial colonization of the animal gut assembles into a microbiome that aids in digestion (*1*), vitamin production (*2*), host immune modulation (*3*), and exclusion of pathogens (*4*). Colonization depends in part on bacterial requirements for nutrients, pH, and O_2; resistance to host factors, including stomach acid, bile salts, and immune effectors; as well as successful competition with other bacterial strains (*5*).

Along the length of the animal gut, there is physiological gradation of host digestive processes, such as protein decomposition and lipid absorption. Correspondingly, the mouth, stomach, jejunum, and colon have distinct microbial compositions (*6*, *7*). Hosts can control the localization of bacteria by providing a niche that sequesters and maintains specific bacteria (*8*–*11*). Bacteria have cell wall proteins called adhesins, which attach to host tissues and reduce washout by peristaltic flow (*12*). Adhesins do not have a conserved structure; they often bind nonspecifically by electrostatic interactions between the cell wall and host tissues (*12*) and, as a result, have a spectrum of affinities for different parts of the gut. The substrates of adhesin binding are often part of the extracellular matrix or the mucus, the composition of which varies along the digestive tract (*13*). Although specific binding by means of terminal mannose glycosylation in the urinary tract is known to occur in an Enterobacterial pathogen (*14*), conserved mechanisms of host specificity are not known to regulate symbiont colonization. However, cophylogenetic patterns of evolution between symbionts and their hosts (*15*–*17*) suggest that some conservation does exist. How hosts and symbionts recognize one another and maintain site-specific colonization remains poorly understood, even in the best-studied models.

Genomic islands that promote host association are best studied in opportunistic pathogens, including *Streptococcus* and *Staphylococcus*, which are lactic acid bacteria related to the lactobacilli (*18*–*20*). Symbiosis islands are also known from symbiotic bacteria (*21*), including the *Vibrio fischeri* strains that colonize the light organs of squid (*22*). These islands carry clusters of genes that are likely involved in bacterial association with their host and often encode adhesins (*23*).

Differentiating transient, nonspecific electrostatic attachments—often observed in opportunistic pathogens—from host-specific adhesion to a symbiotic tissue is complicated. Niche locations occur within parts of the host's body that are difficult to access and interrogate, and other selective pressures, such as nutrient availability, act simultaneously. Live imaging studies, for instance in zebrafish larvae, can be used to directly observe bacterial cells as they migrate through the gut (*24*). In this work, we present a single–bacterial cell tracking method in living *Drosophila* to differentiate between specific adhesion to a precise physical location in the gut and transient pathogen-like adhesion. *Drosophila melanogaster* fruit flies are an emerging model system for the gut microbiome, hosting a relatively low diversity of 5 to 20 commensal species, including Firmicutes and Proteobacteria (*25*). *Lactiplantibacillus plantarum*, a widely used human and animal probiotic (*26*, *27*), has consistently been found in *Drosophila* spp. (*28*, *29*), in which it colonizes a precise, spatially defined physical niche in the foregut of the adult fly, particularly the crop and proventriculus (*9*). A host-specific adhesion mechanism has never been identified for *L. plantarum*, despite its capacity to colonize many different metazoan hosts (*30*, *31*). In this work, we investigated the physical and genetic basis of specificity for fly gut colonization by *L. plantarum* and found broad conservation of genes across the Firmicutes—one of the most frequently host-associated bacterial phyla but for which host specificity genes are largely unknown.

Lactiplantibacillus colonizes by stable binding in the niche

L. plantarum WF (*LpWF*), isolated from the gut of a wild *D. melanogaster*, colonizes the gut with higher efficiency compared with related strains of *L. plantarum* isolated from humans, laboratory flies, or silage (Fig. 1, A to F, and fig. S1) (*9*, *32*). To investigate the stability of colonization at the single–bacterial cell level, we directly visualized individual mCherry-expressing bacterial cells colonizing the niche in living flies by optimizing the Bellymount video microscopy technique (*33*) at a focal plane in close proximity to the inner surface of the fly's crop (Fig. 1, G and H, and fig. S2). As a control, we imaged the silage isolate, *LpATCC8014* (Fig. 1I and movie S1), which moved with a median diffusion coefficient of 0.10 $\mu m^2/s$ (Fig. 1J), consistent with the theoretical prediction from the Stokes-Einstein equation (fig. S2). *L. plantarum* lacks cellular machinery for motility, such as flagella (*31*), so the observed Brownian motion is expected. In *LpATCC8014*, cases of transient binding were observed followed by detachment, consistent with observed probabilistic colonization (*32*). By contrast, *LpWF* cells adhered to the inner surface of the crop (Fig. 1, H and K, and movie S2) with a median diffusion coefficient of 0.001 $\mu m^2/s$ and did not detach. Attachment by *LpWF* cells and lack of attachment by *LpATCC8014* cells were independent of whether the cells were alive or dead (fig. S3), which indicates that the stable localization in the crop is the result of a physiochemical binding interaction. Thus, *LpWF* cells colonize by a form of high-affinity adhesion rather than by transient adhesion and rebinding. We note that high-affinity binding occurs only in the foregut and not in any other part of the gut, including the midgut (*9*).

Isolation of mutants with loss of colonization specificity

We first undertook a genomics analysis between *LpATCC8014* and *LpWF* to identify the genetic basis for *LpWF* adhesion in bacteria. We observed 596 separate genetic differences (table S1), none of which was a clear candidate for colonization (supplementary materials).

We next screened for colonization genes using an evolve and resequence approach (Fig. 1L). We used continuous growth of *LpWF* in a

[1]Biosphere Sciences and Engineering Division, Carnegie Institution for Science, Baltimore, MD, USA. [2]Department of Biology, Johns Hopkins University, Baltimore, MD, USA. [3]Department of Ecology and Evolutionary Biology, University of California Los Angeles, Los Angeles, CA, USA. [4]Department of Human Genetics, University of California Los Angeles, Los Angeles, CA, USA.
*Corresponding author. Email: ludington@carnegiescience.edu
†These authors contributed equally to this work.

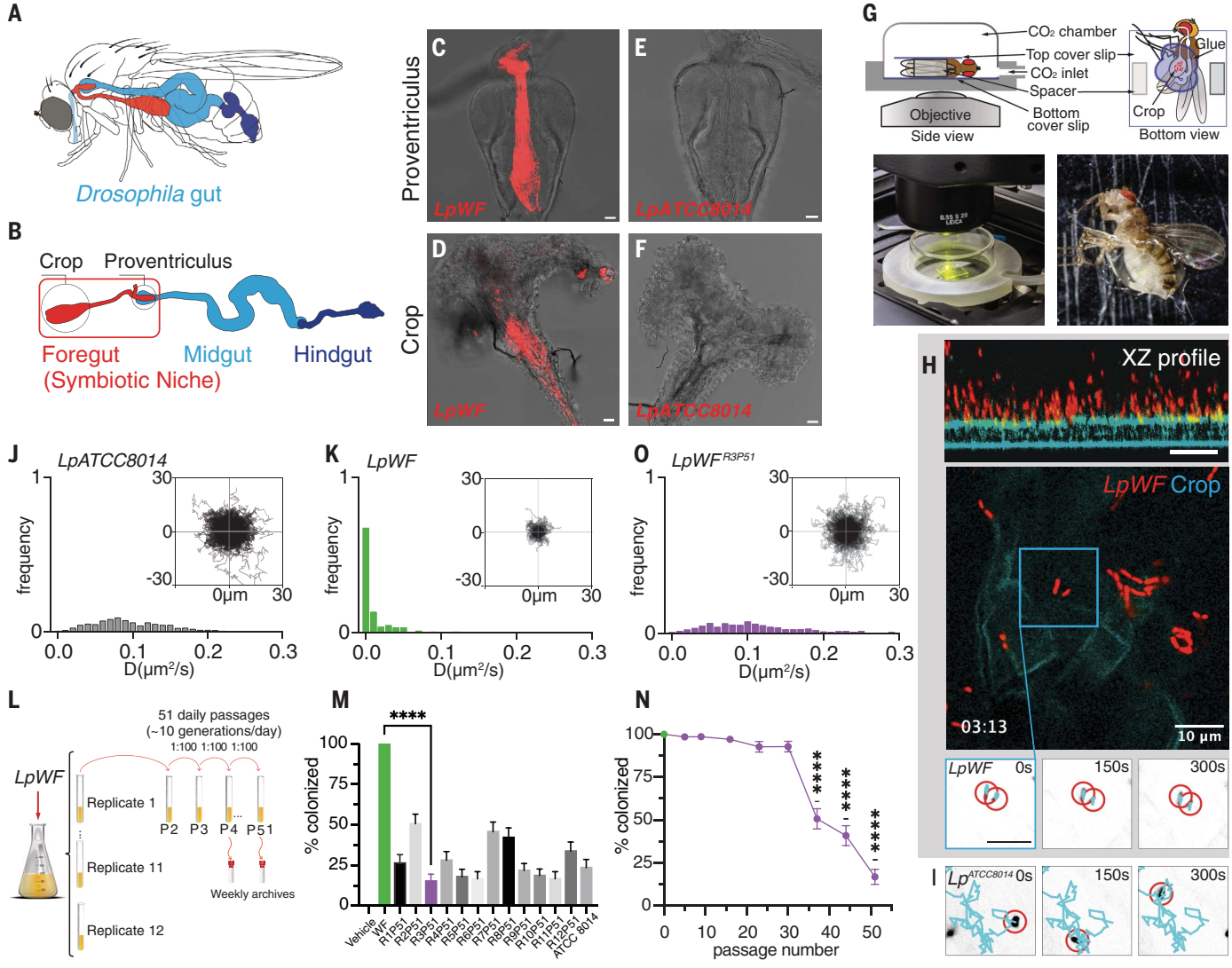

Fig. 1. Live imaging of bacterial cells in the *Drosophila* gut shows that *LpWF* attaches specifically to a physical niche in the host foregut. (**A** and **B**) Schematics showing the anatomy of the fruit fly gut in the fly (A) and dissected (B), with the symbiotic niche highlighted in red. (**C** and **D**) Confocal micrographs of proventriculus (C) and crop (D) colonized by mCherry-labeled *LpWF*. (**E** and **F**) Confocal micrographs of proventriculus (E) and crop (F) colonized by mCherry-labeled *LpATCC8014*. Scale bars, 20 µm. (**G**) Schematic and photos showing the Bellymount apparatus for live imaging. (**H**) X-Z (top), X-Y (middle), and X-Y (bottom) time series images of *LpWF* by Bellymount. (**I**) X-Y time series and particle tracking of *LpATCC8014*. (**J** and **K**) Histogram of coefficients of diffusion for individual *LpATCC8014* (J) and *LpWF* (K) cells. n = 500 tracks from five biological replicates. (**L**) Schematic describing the experimental evolution screen. P, passage. (**M**) Percentage of flies colonized by each of the 12 experimental replicates at passage 51. $n \geq 72$ individual flies from three biological replicates per strain. (**N**) Percentage of flies colonized by evolved strains diminished with passage number. $n \geq 72$ individual flies from three biological replicates per time point. (**O**) Histogram of coefficients of diffusion for *LpWF*^{R3P51}. n = 500 tracks from five biological replicates. Statistics by Kruskal-Wallis test; ****$P < 0.0001$.

rich liquid medium to eliminate selection pressure for attachment to the fly while maintaining selection for robust growth. After 51 passages, we observed a maintenance of growth rate (fig. S4 and table S2) and a decrease in colonization efficiency for all 12 replicates (Fig. 1M, fig. S4, and table S3). Replicate three at passage 51 (*LpWF*^{R3P51}) was the weakest colonizer between biological replicates (Fig. 1M and fig. S5). During the passage experiment, *LpWF*^{R3P51} decreased in colonization efficiency measurably at passage 37 (Fig. 1N and fig. S5), indicating that a mutant with loss of function in one or more colonization genes took over the population at this time point. Bellymount imaging revealed a sharp decrease in gut adhesion and an increase in diffusion rate to 0.122 µm^2/s for *LpWF*^{R3P51} versus 0.001 µm^2/s for *LpWF* (Fig. 1O and movie S3).

Deep sequencing of mutants reveals a putative colonization island

To identify genomic changes associated with the colonization defects observed in the evolved populations, we resequenced the evolved strains and mapped the reads to the *LpWF* assembly, finding only 11 single-nucleotide polymorphisms (SNPs) with no obvious linkage to colonization (fig. S6 and table S4). The *LpWF* Illumina assembly was highly fragmented, with 316 contigs (table S5). To search for structural variants, we resequenced both *LpWF* and *LpWF*^{R3P51} using PacBio HiFi (table S5), obtaining an assembly for *LpWF* composed of a single circular chromosome [3.23 mega–base pairs (Mbp)] plus five plasmids [plasmid 1, 185.5 kilo–base pairs (kbp); plasmid 2, 60.0 kbp; plasmid 3, 41.6 kbp;

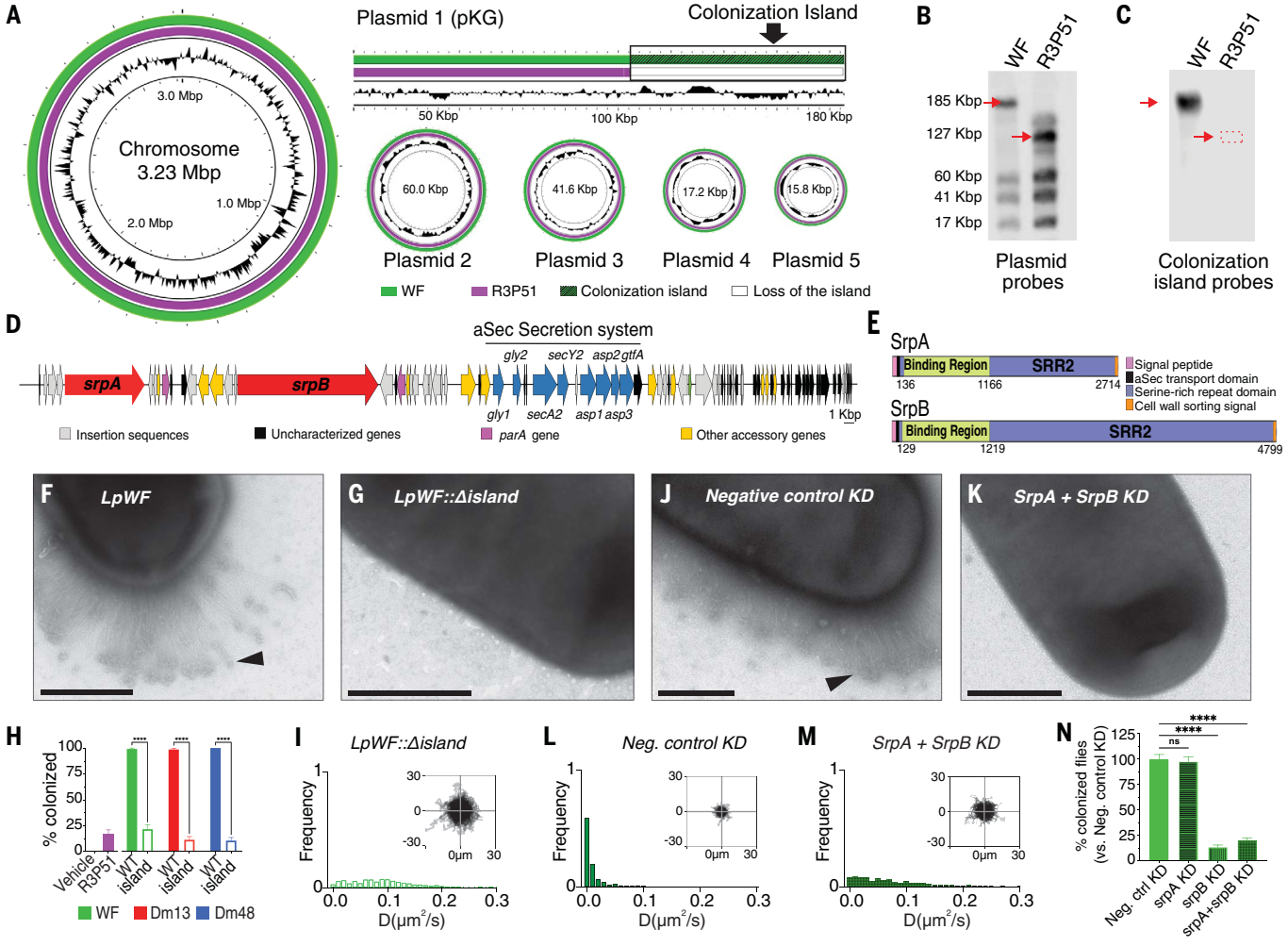

Fig. 2. A colonization island containing SRRPs drives *L. plantarum* adhesion to the gut. (**A**) Long-read sequencing of *LpWF* and the experimentally evolved mutants indicates a colonization island. (**B** and **C**) PFGE and Southern blotting of plasmid partition genes shows a reduction in length of the linear pKG plasmid that contains the colonization island (B) and loss of the colonization island (C) using probes for the island-specific genes *srpA* and *srpB*. (**D**) Map of the island with functional annotations. See fig. S13 for additional annotations. (**E**) SRRP protein schematics. (**F** and **G**) Negative stain EM of *LpWF* (F) and *LpWF::Δisland* (G). Arrowhead indicates fibril structures. Scale bars, 500 nm. (**H**) Quantification of colonization of flies by *LpWF*, *LpDm13*, *LpDm48*, and mutants. $n \geq 100$ flies per strain. (**I**) Bellymount quantification for *LpWF::Δisland*. (**J** and **K**) Negative stain EM of *LpWF* carrying CRISPRi with single guide RNA (sgRNA) negative control (J) and sgRNA targeting *srpA* and *srpB* (K). (**L** and **M**) Bellymount of *LpWF* carrying CRISPRi with sgRNA negative control (L) and sgRNA targeting *srpA* and *srpB* (M). (**N**) Colonization quantification for CRISPRi knockdowns (KDs) in *LpWF*. Statistics by Fisher's exact test; ****$P < 0.0001$; ns, not significant.

plasmid 4, 17.2 kbp; and plasmid 5, 15.8 kbp] (Fig. 2A). Analysis by polymerase chain reaction (PCR) and read mapping to our reference genome for *LpWF* showed that plasmid 1, which we named pKG-WF, is linear, whereas the four remaining plasmids are circular (figs. S7 and S8 and table S6). Linear plasmids are known to occur in several lactic acid bacteria, including *L. plantarum* (*34*).

In $LpWF^{R3P51}$, the linear pKG-WF plasmid was truncated by loss of a contiguous 82.8-kbp region from one end (Fig. 2A). Alignments of short reads from passage 51 of each of the 12 evolved population replicates showed significant reduction in coverage of this region (fig. S9), indicating a consistent genetic basis for the loss of colonization. On the basis of short-read resequencing of each time point in replicate 3, we found that loss of the region occurred in the second week of passaging, with a negative selection coefficient of 4% per generation (fig. S10), consistent with the faster growth rate of the mutants (fig. S4). We confirmed loss of the region by pulsed-field gel electrophoresis (PFGE) and Southern blot analysis, which identified a reduction in the molecular weight of the linear plasmid pKG, consistent with the PacBio genome assembly (Fig. 2, B and C, and fig. S11).

Functional annotation of the lost region using InterProScan (*35*) revealed 80 open reading frames (ORFs) (Fig. 2D), including two putative large adhesins belonging to the group of serine-rich repeat proteins (SRRPs), which are known from *Streptococcus*, *Staphylococcus*, and one lactobacillus species, yet their specific role in host association remains little understood (*20*, *36*). The two *LpWF* SRRPs, which we call *srpA* and *srpB*, are 2714 amino acids and 4799 amino acids, respectively, and make up 26.7% of the region's sequence, which is large for bacterial proteins but average for SRRPs (Fig. 2E). The region also contained the auxiliary SRRP secretion genes called the aSec system (Fig. 2D) (*19*, *20*, *37*, *38*). SRRPs typically form fibrils on the cell wall (*39*, *40*). Using negative stain electron microscopy (EM), we observed filamentous structures that were roughly 500 nm in length on the surface of *LpWF* cells (Fig. 2F).

The colonization island is necessary for colonization specificity

To determine the role that deletion of the region played in the evolved mutant colonization

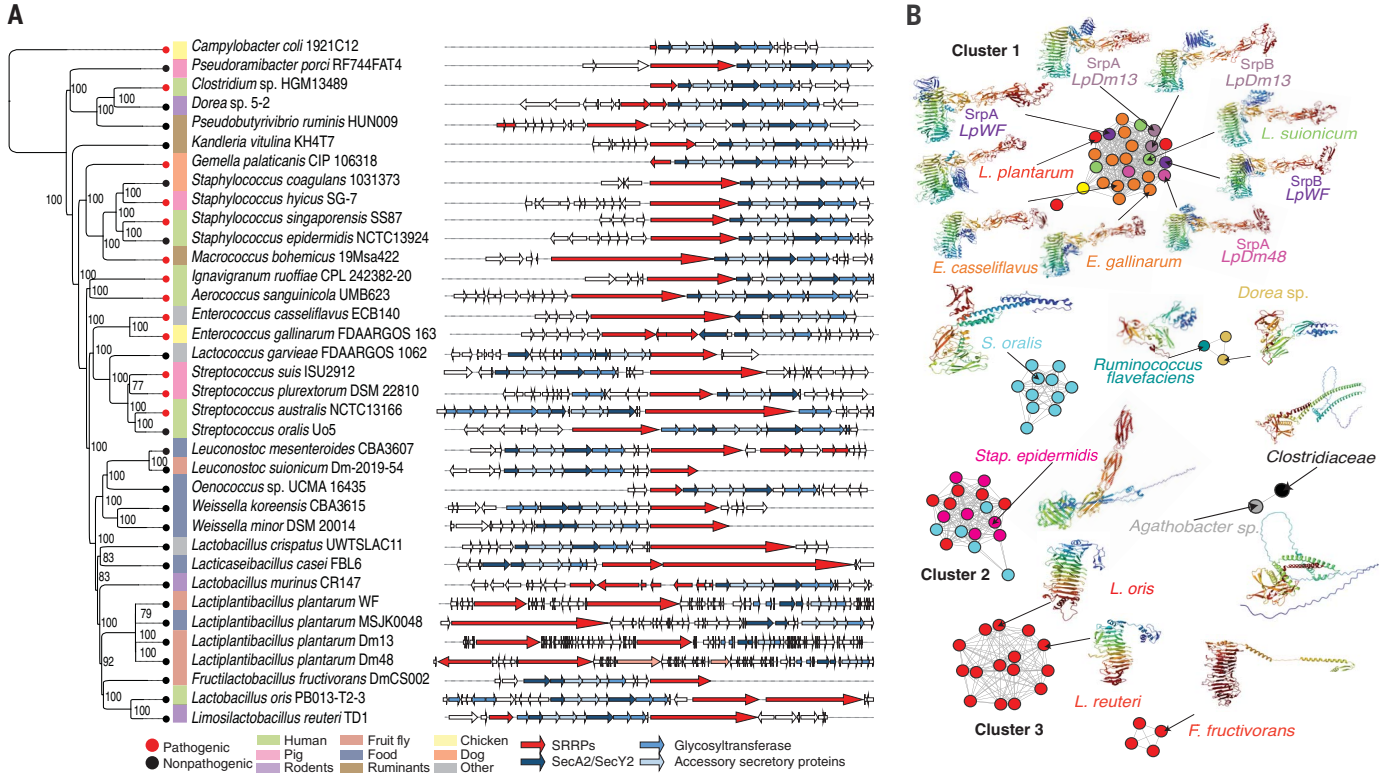

Fig. 3. The colonization island is widely conserved in host-associated bacteria. (**A**) Core genome phylogeny of 36 species and maps of their corresponding colonization islands. (**B**) Clustering analysis of the binding regions of diverse bacterial colonization islands and representative predicted structures for each of the clusters. See fig. S24 for complete results.

phenotypes compared with the effect of the 11 point mutations, we developed a Cas9-based approach to remove the complete region by truncating the pKG plasmid in the *LpWF* background (fig. S12). Negative stain EM confirmed loss of the surface filaments in this mutant, *LpWF::Δisland* (Fig. 2G). The mutant exhibited an ≈85% reduction in colonization efficiency relative to the wild type, indistinguishable from the $LpWF^{R3P51}$ mutant, which indicates that the region is necessary for strong colonization (Fig. 2H). We therefore call this region a colonization island. Bellymount of *LpWF::Δisland* confirmed that the island is necessary for single-cell stable attachment to the crop luminal surface (median diffusion coefficient 0.073 $\mu m^2/s$ versus 0.0010 $\mu m^2/s$ for *LpWF*) (Fig. 2I and movie S4).

Instability of the island during 51 days of evolution appeared to arise from recombination of repeated insertion sequences (ISs), which are small transposable elements that often occur in genomic islands. We used them to predict 16 recombination sites within the island, dividing it into 13 different blocks of genes flanked by ISs (I to XIII) (fig. S13 and supplementary materials).

To identify mutants with partial deletions of the island, we used quantitative PCR directed at specific IS blocks to screen the early time points in the evolution experiment when the populations were heterogeneous (fig. S14 and table S7). We isolated two main genotypes, which we further confirmed using long-read sequencing (fig. S15): (i) One mutant, *LpWF::ΔsrpB+aSec*, had deletions of blocks III to XIII, which included *srpB* and the aSec system and carried the same colonization defects as *LpWF::Δisland*, indicating that *srpB* and the aSec system were necessary for the function of the island (fig. S14 and movie S5). (ii) The other mutant, *LpWF::ΔsrpB+25*, had deletions of blocks II to IV, including *srpB*, and an additional 25 kbp of unknown function at the plasmid terminus. *LpWF::ΔsrpB+25* retained *srpA* and the aSec secretion machinery. Accordingly, *LpWF::ΔsrpB+25* colonized a greater proportion of flies compared with *LpWF::Δisland* or *LpWF::ΔsrpB+aSec*, which indicates that *srpA* and the aSec system aid colonization (fig. S16 and movie S6). To confirm the role of the adhesins in colonization, we adapted the CRISPRi system (*41*, *42*) for single knockdowns of *srpA* and *srpB* and a double *srpA* and *srpB* knockdown (fig. S17). Negative stain EM confirmed loss of the surface fibers in the double knockdown (Fig. 2, J and K) and loss of the longer fibers in the *srpB* knockdown (fig. S17). The negative control knockdown had the same characteristics as *LpWF* (fig. S17 and movie S7). The *srpA* knockdown showed wild-type colonization and slightly increased diffusion in the crop (fig. S17 and movie S8), indicating that it is not necessary for colonization. The *srpB* knockdown showed diminished colonization and adhesion, similar to *LpWF::ΔsrpB+25*, indicating the greater importance of *srpB* in colonization (fig. S17 and movie S9). The double knockdown did not adhere to the crop or colonize, indicating that the adhesins are necessary for colonization by *LpWF* (Fig. 2, L to N, and movie S10).

Adhesins commonly bind host glycans (*43*). We previously had found that several glycan-binding lectins localize to the foregut niche (*9*). We fed flies with lectins before bacterial inoculation to test whether they interfered with *LpWF* attachment to the niche. Wheat germ agglutinin (WGA) and succinylated wheat germ agglutinin (S-WGA) both blocked *LpWF* binding to the niche tissue (fig. S18), which indicates that the lectins' primary target, *N*-acetyl glucosamine, is bound by the adhesins.

The colonization island is found in other *L. plantarum* strains

The colonization island is not present in short-read genomes for the technical reason that extensive sequence repeats in the SRRPs and ISs cause fragmented assemblies with very short contigs. To directly assess the prevalence of the colonization island in other *L. plantarum*

strains without assembly, we downloaded the raw reads for the National Center for Biotechnology Information genomes from the Sequence Read Archive (SRA) and aligned them to the *LpWF* island. We first constructed a whole-genome shotgun (WGS) database with 1247 raw read datasets sequenced by Illumina (96.1%), PacBio (2.4%), Oxford Nanopore (0.89%), and other long-read technologies (0.61%) (table S8). Next, we mapped the reads against the nucleotide sequence of the island and identified two *L. plantarum* strains, *LpDm13* and *LpDm48*, which were isolated from wild *D. melanogaster* flies collected in Ithaca, New York (fig. S19) (*44*). We confirmed the presence of the colonization islands by long-read sequencing. The *LpDm13* and *LpDm48* islands are comparable in size to the *LpWF* island (97.3 kbp and 113.8 kbp, respectively) and are also contained within a linear plasmid (fig. S20). The island in *LpDm13* consists of 79 ORFs, whereas the *LpDm48* island comprises 108 ORFs. Both islands harbor two SRRP alleles (SrpA 513 kD and SrpB 442 kD for *LpDm13*; and SrpA 352 kD and SrpB 493 kD for *LpDm48*) and a corresponding aSec secretion system, with numerous adjacent ISs.

LpDm13 and *LpDm48* both colonized nearly all flies when inoculated at a low dose (Fig. 2H) with stable populations averaging ~20,000 to 50,000 colony-forming units within the fly gut, consistent with *LpWF* (fig. S21), and showed similar spatial localization and diffusion (fig. S22 and movies S11 and S12). We next constructed mutants in *LpDm13* and *LpDm48* with a knockout of the complete island. These mutants exhibited an acute colonization deficiency (Fig. 2H and fig. S21), loss of spatial localization, and adhesion deficiency by Bellymount (fig. S22 and movies S13 and S14) compared with the corresponding wild-type strains. Overall, our results indicate a conserved colonization specificity island within *L. plantarum*.

The colonization island is conserved across the Firmicutes

SRRPs with aSec systems were previously known from just three genera, all within the Lactobacillales (*36*). To explore conservation of the colonization island across other bacterial genera, we first used BLASTP to search for the complete aSec system in the Pathosystems Resource Integration Center (PATRIC) database of bacterial genomes, focused on human bacteria (*45*). Most bacterial genome assemblies were highly fragmented as a result of the high proportion of repeat elements, which were also abundant in the *LpWF* colonization island. As a conservative approach to identify homologous colonization islands, we excluded BLAST hits where the aSec genes were not in a semicontiguous block because some aSec genes have homologs in the genome (*20*). We first considered cases with at least one complete SRRP gene within the aSec block. Owing to the lack of amino acid sequence conservation in SRRP genes, we used a hidden Markov model–based search to detect them (materials and methods). Despite the preponderance of highly fragmented genome assemblies, we discovered complete colonization islands in 207 bacterial genomes, spanning the entire Firmicutes phylum, including 134 Lactobacillales and 25 Clostridia, as well as in three Actinomycetes and one ε-Proteobacterium (Fig. 3A, fig. S23, and table S9). We found the aSec system with a non-SRRP, large adhesin in an additional 42 genomes, including seven Clostridia and one γ-Proteobacterium (fig. S23). Although there was broad synteny of the aSec genes across the phylogeny (Fig. 3A), there were differences between the adhesins, consistent with these being host specific.

To investigate the conservation of the SRRPs, we constructed a similarity network on the basis of structural motifs in the binding regions, revealing 15 clusters and two singletons (Fig. 3B and fig. S24). We simulated their structures using RosettaFold (*46*). Three of these clusters, including the *LpWF* cluster, contain a beta-solenoid motif, which was present in genera including *Enterococcus*, *Fructilactibacillus*, *Lactobacillus*, *Lactoplantibacillus*, *Leuconostoc*, and *Weisella*. Another structural motif was a chain of immunoglobulin-like domains, which was present in genera including *Oenococcus*, *Pediococcus*, and *Staphylococcus* as well as in combination with the beta-solenoid domain in the *LpWF* cluster. Additional distinct binding regions were present in the Clostridiales, including in *Dorea* and *Pseudoramibacter*. We analyzed the conservation of amino acids within each of the three largest clusters of binding regions (clusters 1 to 3 in Fig. 3B) and found a strong signature of positive selection (fig. S25), consistent with the role of these proteins in host specificity. Nonsynonymous nucleotide changes were concentrated in the beta-solenoid motifs of clusters 1 and 3 and in the first immunoglobulin-like domain of cluster 2 (fig. S25).

On the basis of the conservation of this genomic region, we defined the minimum colonization island gene set composed of seven core genes with at least one *srp*; the tandem transporters *secA2* and *secY2*; the glycosyltransferase *gtfA*; and three accessory secretory protein genes, *asp1*, *asp2*, and *asp3*. The order of the genes was conserved at the genus level but differed between some families, suggesting rearrangements within the island on longer evolutionary timescales (Fig. 3A).

To assess whether the island is predominantly passed between strains by vertical transfer or horizontal gene transfer, we performed cophylogenetic analysis. Vertical transfer would produce parallel phylogenies for the core genome and the colonization island genes, whereas horizontal transfers of the island would produce discrepancies (*47*). We first reconstructed the phylogeny of the host species using the core genome, which was composed of 81 orthologous proteins (table S10), and the island using the six conserved aSec proteins (fig. S26). We then compared the two phylogenies for each well-resolved clade using a cutoff bootstrap value of >70 and estimated the Patristic distance (*48*, *49*). The branching pattern of the aSec phylogeny overall matched the genome phylogeny (fig. S26; mean ParaFit Global index, 0.021; $P < 0.05$), which indicates that the colonization island was likely present in the ancestor of the Firmicutes, a largely host-associated phylum. Some cases of horizontal gene transfer were apparent, such as the acquisition of the island by the Gram-negative *Campylobacter coli* from the Gram-positive *Staphylococcus epidermidis*, indicating spread of the island beyond the Firmicutes. The bacterial genomes that potentially acquired the island by horizontal gene transfer show no significant difference in GC (guanine and cytosine) content compared with those bacterial genomes where the genomic region was acquired by vertical transfer, which suggests that the horizontal transfers are not recent (fig. S26; Brunner-Munzel test, $P < 0.005$).

Conclusions

Our study identified a set of genes for symbiont gut colonization through specific binding of a gut niche. The ability to directly observe the colonization process at the single–bacterial cell level allows a clear differentiation between a dynamically stable population where individual cells transiently bind the gut wall and maintain the population by growing faster than they are shed versus a population where the individual cells form long-term stable attachments to the gut wall. The molecular mechanistic basis for the binding specificity remains unknown. Future studies to identify the role of receptor-ligand binding specificity for symbiont attachment in mammals will require methodological developments and the identification of symbiont strains, their specificity genes, their physical niches, and the ligands that define them. Technical challenges include not only the rapid loss of colonization genes in culture in rich media (fig. S10) but also fragmented genome assemblies resulting from short-read sequencing of repetitive elements, including SRRPs. The study of wild strains using current long-read technologies can overcome both obstacles.

The genes to enable host colonization include the aSec system, which is known for secretion of virulence factors in *Streptococcus* and *Staphylococcus* species that are predominantly opportunistic pathogens (*20*). The evolutionary selection for molecules that lead opportunistic pathogens to cross an epithelial barrier and cause a systemic infection (*14*) is distinct

from the long-term selective pressures on nonpathogenic commensals (*50*). To colonize a new host, a symbiont must selectively bind the symbiotic niche and not attach to other lower-quality sites in the host (*8*, *51*). The evolution of the host digestive tract has also shaped the selection of beneficial bacterial symbionts through the availability of different binding substrates, including glycans (*52*). Much remains to be discovered about how hosts regulate the development and maintenance of physical niches for symbiotic bacteria.

REFERENCES AND NOTES

1. W. H. Karasov, C. Martínez del Rio, E. Caviedes-Vidal, *Annu. Rev. Physiol.* **73**, 69–93 (2011).
2. O. M. Sokolovskaya, A. N. Shelton, M. E. Taga, *Science* **369**, eaba0165 (2020).
3. A. J. McDermott, G. B. Huffnagle, *Immunology* **142**, 24–31 (2014).
4. J. Walter, M. X. Maldonado-Gómez, I. Martínez, *Curr. Opin. Biotechnol.* **49**, 129–139 (2018).
5. M. Wolter *et al*., *Nat. Rev. Gastroenterol. Hepatol.* **18**, 885–902 (2021).
6. P. J. Turnbaugh *et al*., *Sci. Transl. Med.* **1**, 6ra14 (2009).
7. G. McCallum, C. Tropini, *Nat. Rev. Microbiol.* **22**, 105–118 (2024).
8. K. Takeshita, Y. Kikuchi, *Res. Microbiol.* **168**, 175–187 (2017).
9. R. Dodge *et al*., *Nat. Commun.* **14**, 1557 (2023).
10. C. Fung *et al*., *PLOS Biol.* **17**, e3000231 (2019).
11. S. V. Nyholm, E. V. Stabb, E. G. Ruby, M. J. McFall-Ngai, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 10231–10235 (2000).
12. K. A. Kline, S. Fälker, S. Dahlberg, S. Normark, B. Henriques-Normark, *Cell Host Microbe* **5**, 580–592 (2009).
13. S. Etzold, N. Juge, *Curr. Opin. Struct. Biol.* **28**, 23–31 (2014).
14. E. H. Beachey, *J. Infect. Dis.* **143**, 325–345 (1981).
15. J. G. Sanders *et al*., *Nat. Microbiol.* **8**, 1039–1050 (2023).
16. T. A. Suzuki *et al*., *Science* **377**, 1328–1332 (2022).
17. M. K. Nishiguchi, E. G. Ruby, M. J. McFall-Ngai, *Appl. Environ. Microbiol.* **64**, 3209–3213 (1998).
18. M. Juhas *et al*., *FEMS Microbiol. Rev.* **33**, 376–393 (2009).
19. C. J. Sanchez *et al*., *PLOS Pathog.* **6**, e1001044 (2010).
20. M. Braunstein, B. A. Bensing, P. M. Sullam, *Microbiol. Spectr.* **7**, 10.1128/microbiolspec.psib-0025-2018 (2019).
21. H. Ochman, N. A. Moran, *Science* **292**, 1096–1099 (2001).
22. E. G. Ruby *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **102**, 3004–3009 (2005).
23. P. J. Planet, S. C. Kachlany, D. H. Fine, R. DeSalle, D. H. Figurski, *Nat. Genet.* **34**, 193–198 (2003).
24. T. J. Wiles *et al*., *PLOS Biol.* **14**, e1002517 (2016).
25. N. A. Broderick, B. Lemaitre, *Gut Microbes* **3**, 307–321 (2012).
26. H. A. Seddik *et al*., *Probiotics Antimicrob. Proteins* **9**, 111–122 (2017).
27. M. Schwarzer *et al*., *Science* **351**, 854–857 (2016).
28. J.-H. Ryu *et al*., *Mol. Cell. Biol.* **24**, 172–185 (2004).
29. G. Storelli *et al*., *Cell Metab.* **27**, 362–377.e8 (2018).
30. M. E. Martino *et al*., *Environ. Microbiol.* **18**, 4974–4989 (2016).
31. N. Buntin, W. M. de Vos, T. Hongpattarakere, *Appl. Microbiol. Biotechnol.* **101**, 7663–7674 (2017).
32. B. Obadia *et al*., *Curr. Biol.* **27**, 1999–2006.e8 (2017).
33. L. A. J. Koyama *et al*., *PLOS Biol.* **18**, e3000567 (2020).
34. D. Davray, H. Bawane, R. Kulkarni, *Food Microbiol.* **109**, 104153 (2023).
35. M. Blum *et al*., *Nucleic Acids Res.* **49**, D344–D354 (2021).
36. M. S. Cinar, A. Niyas, F. Y. Avci, *J. Bacteriol.* **206**, e00241-23 (2024).
37. S. Sequeira *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, E2706–E2715 (2018).
38. A. K. Singh, S. A. Woodiga, M. A. Grau, S. J. King, *Infect. Immun.* **85**, e00774-16 (2017).
39. A. Ronis *et al*., *Infect. Immun.* **87**, e00406-19 (2019).
40. S. Ramboarina *et al*., *J. Biol. Chem.* **285**, 32446–32457 (2010).
41. J. M. Peters *et al*., *Nat. Microbiol.* **4**, 244–250 (2019).
42. I. S. Myrbråten *et al*., *MSphere* **4**, e00007-19 (2019).
43. K. Moonens, H. Remaut, *Curr. Opin. Struct. Biol.* **44**, 48–58 (2017).
44. J. G. McMullen II, E. Bueno, F. Blow, A. E. Douglas, *Genome Biol. Evol.* **13**, evab127 (2021).
45. J. J. Gillespie *et al*., *Infect. Immun.* **79**, 4286–4298 (2011).
46. M. Baek *et al*., *Science* **373**, 871–876 (2021).
47. S. M. Soucy, J. Huang, J. P. Gogarten, *Nat. Rev. Genet.* **16**, 472–482 (2015).
48. P. Legendre, Y. Desdevises, E. Bazin, *Syst. Biol.* **51**, 217–234 (2002).
49. E. Paradis, *Analysis of Phylogenetics and Evolution with R* (Springer, ed. 2, 2012).
50. S. P. Brown, D. M. Cornforth, N. Mideo, *Trends Microbiol.* **20**, 336–342 (2012).
51. S. V. Nyholm, M. J. McFall-Ngai, *Nat. Rev. Microbiol.* **19**, 666–679 (2021).
52. K. Nakashima *et al*., *Nat. Commun.* **9**, 3402 (2018).
53. K. Gutierrez-Garcia *et al*., A conserved bacterial genetic basis for commensal-host specificity, Zenodo (2024); https://doi.org/10.5281/zenodo.13736826.
54. K. Gutierrez-Garcia *et al*., A conserved genetic basis for commensal-host specificity through live imaging of colonization dynamics, Zenodo (2024); https://doi.org/10.5281/zenodo.13115916.

ACKNOWLEDGMENTS

J. Chaston (Brigham Young University) provided the *LpDm13* and *LpDm48* strains isolated from wild *D. melanogaster*. R. Grabherr and S. Heinl (BOKU Vienna) provided the pCD256-mCherry plasmid. L. O'Brien, K. C. Huang, A. Aranda-Diaz, and M. Siddiqi provided Bellymount support. **Funding:** This study was supported by National Institutes of Health grant R01 DK128454 (W.B.L.); National Institutes of Health grant R21 AI173779 (W.B.L.); National Science Foundation grant IOS 2032985 (W.B.L.); National Science Foundation CAREER Award, grant IOS 2144342 (W.B.L.); the Carnegie Institution for Science Endowment; National Institutes of Health T32 training grant to JHU CMDB program; and NIGMS NIH award R35GM151023 and NSF CAREER award 2240098 (N.G.).
Author contributions: Conceptualization: K.G.-G., K.A., and W.B.L. Investigation: K.G.-G., K.A., R.D., R.-C.H., H.Z., B.O., A.D., S.A., X.Y., and R.W. Supervision: W.B.L. and N.G. Writing: K.G.-G., K.A., R.D., and W.B.L. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data used in this analysis are free to access and can be found in the SRA under BioProject accession nos. PRJNA1094941, PRJNA1094948, and PRJNA1093452; on Zenodo (*53*, *54*); and on GitHub (https://github.com/K-Gutierrez/LpColonizationIsland/). All strains are available upon request.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adp7748
Materials and Methods
Figs. S1 to S26
Tables S1 to S15
References (*55*–*94*)
MDAR Reproducibility Checklist
Movies S1 to S14

Submitted 10 April 2024; accepted 7 October 2024
10.1126/science.adp7748