Genomic Comparisons of *Lactobacillus crispatus* and *Lactobacillus iners* Reveal Potential Ecological Drivers of Community Composition in the Vagina

Michael T. France, Helena Mendes-Soares,* Larry J. Forney

Institute for Bioinformatics and Evolutionary Studies and Department of Biological Sciences, University of Idaho, Moscow, Idaho, USA

ABSTRACT

Lactobacillus crispatus and *Lactobacillus iners* are common inhabitants of the healthy human vagina. These two species are closely related and are thought to perform similar ecological functions in the vaginal environment. Temporal data on the vaginal microbiome have shown that nontransient instances of cooccurrence are uncommon, while transitions from an *L. iners*-dominated community to one dominated by *L. crispatus*, and vice versa, occur often. This suggests that there is substantial overlap in the fundamental niches of these species. Given this apparent niche overlap, it is unclear how they have been maintained as common inhabitants of the human vagina. In this study, we characterized and compared the genomes of *L. iners* and *L. crispatus* to gain insight into possible mechanisms driving the maintenance of this species diversity. Our results highlight differences in the genomes of these two species that may facilitate the partitioning of their shared niche space. Many of the identified differences may impact the protective benefits provided to the host by these two species.

IMPORTANCE

The microbial communities that inhabit the human vagina play a critical role in the maintenance of vaginal health through the production of lactic acid and lowering the environmental pH. This precludes the growth of nonindigenous organisms and protects against infectious disease. The two most common types of vaginal communities are dominated by either *Lactobacillus iners* or *Lactobacillus crispatus*, while some communities alternate between the two over time. We combined ecological theory with state-of-the-art genome analyses to characterize how these two species might partition their shared niche space in the vagina. We show that the genomes of *L. iners* and *L. crispatus* differ in many respects, several of which may drive differences in their competitive abilities in the vagina. Our results provide insight into factors that drive the complicated temporal dynamics of the vaginal microbiome and demonstrate how closely related microbial species partition shared fundamental niche space.

The microbial communities that inhabit the vaginas of healthy reproductive age women commonly contain high proportions of *Lactobacillus* species (1). Studies have shown that these communities can be divided into five different types, four of which are dominated by either *Lactobacillus iners*, *Lactobacillus crispatus*, *Lactobacillus gasseri*, or *Lactobacillus jensenii* (2). These four species are closely related and are thought to perform similar ecological functions in the vaginal environment (namely, the production of lactic acid) (3, 4). Instances of cooccurrence among these species are rare, and temporal data have demonstrated that shifts in the dominant *Lactobacillus* are common (5), suggesting that the species compete for shared niche space in the vagina. Ecological theory predicts that multiple species cannot occupy the same niche indefinitely, as one will eventually outcompete the others (6). Therefore, it is unclear how the four *Lactobacillus* species have been maintained as common inhabitants of the vaginal niche.

Previous studies have shown that competing species can partition their shared niche space through a variety of mechanisms. One such mechanism, termed resource partitioning, occurs when competing species specialize in the use of different subsets of resources, thereby dividing the niche into multiple niches and allowing them to cooccur (7–9). However, we argue that it is unlikely that the vaginal lactobacilli are dividing their shared niche space in this way because they rarely cooccur (2). Species can also partition shared niche space temporally through a mechanism termed conditional differentiation. This occurs when the species differ in competitive ability across the niche's range of environ-

mental conditions (10–12). The abundance of the species is then determined by the abiotic and biotic factors that influence their competitive interactions. For example, in their 2014 work, Mamola and Isaia showed that variation in the temperature and humidity levels in caves allowed two competing spider species to partition their shared niche space (12). We argue that given the complex temporal fluctuations exhibited by vaginal *Lactobacillus* species, these species likely partition their shared niche via this mechanism. In the present study, we characterized and compared the genomes of two of the four prominent vaginal *Lactobacillus* species, *L. crispatus* and *L. iners*, to identify possible ecological factors that might drive these temporal fluctuations in the dominant *Lactobacillus* species.

Received 17 August 2016 Accepted 22 September 2016

Accepted manuscript posted online 30 September 2016

Citation France MT, Mendes-Soares H, Forney LJ. 2016. Genomic comparisons of *Lactobacillus crispatus* and *Lactobacillus iners* reveal potential ecological drivers of community composition in the vagina. *Appl Environ Microbiol* 82:7063–7073. doi:10.1128/AEM.02385-16.

Editor: P. D. Schloss, University of Michigan

Address correspondence to Larry J. Forney, lforney@uidaho.edu.

* Present address: Helena Mendes-Soares, Center for Individualized Medicine, Department of Surgery, Mayo Clinic, Rochester, Minnesota, USA.

Copyright © 2016, American Society for Microbiology. All Rights Reserved.

Little is known about the abiotic and biotic factors that might be relevant to competitive interactions between vaginal *Lactobacillus* species. We speculate that the host's physiology plays a critical role in shaping the vaginal environment through at least two different mechanisms. First, the host is the exclusive source of nutrients available in the environment. These nutrients originate both from the mucus produced by the cervix, which contains a rich mixture of carbohydrates, fatty acids, and trace elements (13), and from vaginal epithelial cells, which in reproductive-age women are loaded with glycogen (14). The amount and composition of cervical mucus, as well as the amount of glycogen available, vary among women and through time in a single woman. Temporal variation in these characteristics occurs both on the scale of the menstrual cycle as well as through the lifetime of an individual (15–19). Additionally, the host can effect change in the vaginal communities via the immune system. The vagina contains various components of the innate and adaptive immune systems that protect this important interface from infection (20). The activity of the host's immune system also exhibits variation between women and through time within women (21), and some of this variation results from interactions with the microbes inhabiting the vagina (22). While it is clear that host physiology must play some role in shaping the microbial communities inhabiting the vagina, the magnitude and nature of its effect are not known.

Although there are four dominant vaginal *Lactobacillus* species, we chose to focus our efforts on genomic comparisons of *L. crispatus* and *L. iners* for two reasons. First, of the five community types, the two most commonly identified in reproductive-age women (>75% relative abundance) are dominated by one of these two species (2, 5). Second, although both of these species produce lactic acid as the end product of fermentation, they differ in many other respects, including their relationship to host health. While the presence of *L. crispatus* in the vagina has always been associated with good health, some studies have hinted that communities dominated by *L. iners* may provide the host with fewer protective benefits (23–25). For example, *L. iners*, but not *L. crispatus*, commonly cooccurs with many of the bacterial species that colonize the vagina during incidences of bacterial vaginosis (23, 26, 27). Communities dominated by *L. iners* have also been associated with a higher vaginal pH than that in communities dominated by *L. crispatus* (2). In addition, these two species also differ in their specificity for the vaginal habitat. While *L. iners* has almost exclusively been isolated from human vaginal secretions, *L. crispatus* has also been identified in other habitats, like the vertebrate gastrointestinal tract (28), although it is unclear whether *L. crispatus* is a frequent colonizer of these other habitats. We argue that by focusing our efforts on these two common but markedly different vaginal colonizers, we are likely to identify both ecologically and medically relevant factors that govern competitive interactions between the vaginal *Lactobacillus* species.

We employed comparative genomics techniques to investigate niche partitioning by *L. crispatus* and *L. iners* and to identify putative ecological factors that govern competitive interactions between these species. The two previous comparative genomic studies of the vaginal *Lactobacillus* species have either focused on *L. crispatus* only (29), or on broad functional differences between vaginal versus nonvaginal *Lactobacillus* (30). In this study, we characterized and compared the genomes of 15 *L. iners* and 15 *L. crispatus* strains. We report differences in the size, functional makeup, and evolution of the genomes of these two species, con-

TABLE 1 Bacterial genome data

Strain	Source	Genome size (Mbp)	No. of contigs	No. of ORFs
<i>L. crispatus</i>				
JV-V01	Human vagina	2.22	86	2,151
MV-1A-US	Human vagina	2.25	7	2,383
125-2-CHN	Human vagina	2.30	30	2,196
MV-3A-US	Human vagina	2.44	76	2,458
CTV-05	Human vagina	2.36	25	2,425
FB049-03	Human vagina	2.46	5	2,474
FB077-07	Human vagina	2.70	10	2,688
SJ-3C-US	Human vagina	2.09	201	2,199
214-1	Human vagina	2.07	187	2,100
2029	Human vagina	2.19	295	2,545
EM-LC1	Human vagina	1.83	63	1,839
ST1	Chicken crop	2.04	1	2,060
VCM6	Human vagina	2.34	253	2,397
VCM7	Human vagina	2.10	247	2,108
VCM8	Human vagina	2.33	255	2,427
<i>L. iners</i>				
UPII 60-B	Human vagina	1.32	31	1,288
UPII 143-D	Human vagina	1.26	21	1,213
LactinV 01V1-a	Human vagina	1.29	92	1,506
LactinV 03V1-b	Human vagina	1.30	67	1,454
LactinV 09V1-c	Human vagina	1.31	35	1,384
LactinV 11V1-d	Human vagina	1.31	27	1,362
LEAF 2052A-d	Human vagina	1.32	28	1,268
LEAF 2053A-b	Human vagina	1.37	37	1,288
LEAF 2062A-h1	Human vagina	1.30	24	1,278
LEAF 3008A-a	Human vagina	1.27	25	1,216
SPIN 1401G	Human vagina	1.28	52	1,224
SPIN 2503 V10-D	Human vagina	1.28	31	1,293
ATCC 55195	Human vagina	1.24	7	1,152
AB-1	Human vagina	1.29	7	1,230
DSM 13335	Human urine	1.28	12	1,212

sistent with niche partitioning via conditional differentiation. Our analysis indicates that *L. iners* has a genome whose size is drastically reduced, likely making the species dependent on exogenous sources (e.g., cervical mucus or other vaginal species) of vital nutrients. In comparison, *L. crispatus* has a larger genome that contains a broader array of metabolic machinery, likely allowing it to function under a more diverse subset of environmental conditions. These genotypic differences may provide the species with differential competitive abilities across the range of conditions common to the vaginal environment, facilitating the partitioning of their shared niche space.

MATERIALS AND METHODS

Sequences. All available complete and draft genome sequences for *L. iners* and *L. crispatus* were obtained from the National Center for Biotechnology Information (NCBI) FTP site (<ftp://ftp.ncbi.nih.gov/genomes/Bacteria/>) in March 2016. Of the 30 *L. crispatus* and *L. iners* strains used in this study, only two were not isolated from the human vagina (*L. iners* DSM 13335 was isolated from urine, and *L. crispatus* ST1 was isolated from the crop of a chicken). A complete list of the strains with accompanying genome quality statistics can be found in Table 1. The number of contigs for each genome ranged from 7 to 97 contigs for *L. iners* and 1 to 295 contigs for *L. crispatus*. Additionally, sequence data for *Lactobacillus acidophilus* NCFM, *Lactobacillus casei* BD-II, *Lactobacillus delbrueckii* ATCC 11842, *Lactobacillus helveticus* CNRZ32, *Lactobacillus johnsonii* NCC 533, *Lacto-*

bacillus plantarum WCFS1, *Lactobacillus rhamnosus* GG, *Bacillus subtilis* 168, *Enterococcus faecalis* ATCC 29212, and *Streptococcus pneumoniae* R6 were obtained from NCBI for use as outgroups in the construction of the phylogenetic tree.

Gene prediction and functional annotation. Open reading frames (ORFs) were predicted in all genomes using Glimmer version 3.0, with a maximum overlap of 50 bp between ORFs and a minimum ORF length of 110 bp (31–33). Contigs from draft genomes were concatenated with 25-bp spacers of ambiguous DNA sequences. Any ORFs containing spacer sequences were later removed from the analysis. The identified ORFs were translated to protein sequences, and orthologous groups were identified among them using OrthoMCL version 2.0, with a percent identity threshold of 50% (34). Pangenome, core-genome, and accessory-genome accumulation curves were generated from the OrthoMCL output using a python script. This python script and all others used in the analysis are available at github.com/michaelfrance/lactobacillus_genomics.

Core- and accessory-gene sets were assigned functional annotations using Kyoto Encyclopedia of Genes and Genomes (KEGG) BlastKOALA searches against the prokaryotic species database (35). However, this method proved incapable of assigning functions to the genes that are found in only a single strain. We therefore assigned this gene set to functions based on results from Blast searches against the RefSeq nonredundant protein database (36). The top 20 matches for each Blast search with >50% identity (calculated as the number of matches divided by the total length of the two aligned sequences) were examined and, when in agreement, used to assign a broad functional category.

Phylogenetic analysis. OrthoMCL was also used to identify the genes that *L. crispatus*, *L. iners*, and the outgroup strains have in common (50% identity, 90% coverage thresholds). The 39 strains were found to have 242 genes in common spanning more than 250 kb. The orthologous gene clusters were aligned with ClustalW version 2 (37) and concatenated into a single alignment using Phyutility version 2.2.6 (38). The rCluster hierarchical clustering algorithm implemented by PartitionFinder version 1.1.1 (39) was used to identify the optimal partitioning scheme for the concatenated alignment based on codon positions; a total of 301 partitions were used. RAXML version 8 was used to construct a maximum likelihood phylogenetic tree from the partitioned concatenated multiple-sequence alignment (40). Each partition was modeled separately under the general time-reversible gamma substitution model (41). One thousand bootstrap replicates were performed, and convergence occurred after 700 bootstrap replicates. A separate phylogeny was constructed in the same manner from an alignment of the *L. iners* cytolysin gene and matching sequences from the NCBI nonredundant nucleotide database.

Evolutionary analysis. We used the program TimeZone version 1.0 to characterize the evolution of each species' core genome separately (42). This program executes a pipeline designed to identify genes exhibiting signatures of positive selection and recombination. TimeZone relies on the use of a reference genome and is therefore only suitable for the analysis of genes shared among all strains (i.e., core genes). In both cases, the most complete genome was used as the reference sequence (*L. crispatus* ST1 and *L. iners* AB-1). A percent identity and coverage cutoff of 75% were used for both within-species analyses. We used TimeZone to calculate the ratio of nonsynonymous to synonymous nucleotide changes (dN/dS) for the core genome of each species. In general, dN/dS ratios below 1 are indicative of purifying selection, while dN/dS ratios above 1 are indicative of positive selection.

Evolution can also occur through the horizontal acquisition of foreign DNA. To identify these horizontally acquired sequences, we analyzed two different sets of genes. The first included the genes uniquely conserved in the core genomes of either *L. crispatus* or *L. iners*. The second set included those genes present in only a single strain of *L. crispatus* or *L. iners*. To determine whether these sequences were more likely to have been inherited or horizontally acquired, we used Blast searches against the RefSeq nonredundant protein database (as described in "Gene prediction and

functional annotation" above) to identify the genus of the closest matching homologous sequence in the database. If the closest matching sequences in the database were identified in other *Lactobacillus* species, we inferred that the gene is more likely to have been inherited, whereas if the gene matched only non-*Lactobacillus* species, we considered it to have been horizontally acquired.

RESULTS

Phylogenetic analysis. To understand the phylogenetic relationships among the strains used in our study, we constructed a maximum likelihood phylogenetic tree using a partitioned concatenated multiple-sequence alignment of the 242 genes that were present in all of the *L. crispatus* and *L. iners* strains, as well as the outgroup species. In the resultant phylogenetic tree, the 15 strains of *L. crispatus* and 15 strains of *L. iners* clustered together into single clades (Fig. 1). Branching patterns within the *L. crispatus* clade, but not the *L. iners* clade, are generally well supported. Branch lengths within the two species are orders of magnitude shorter than those between species, indicating that the majority of the observed diversity in these 242 genes occurred between rather than within species. Additionally, our analysis indicates that these two vaginal species are not sisters of one another; rather, *Lactobacillus johnsonii* is sister to *L. iners*, and both *Lactobacillus helveticus* and *Lactobacillus acidophilus* are sisters to *L. crispatus* (Fig. 1). In the following analyses, we used this tree to determine whether specific traits of *L. crispatus* and *L. iners* are more likely to be derived characteristics of the species or are ancestral.

Differences in genome size. Perhaps the most obvious difference between the genomes of *L. iners* and *L. crispatus* are a matter of scale. While the average size of the *L. crispatus* genome was 2.25 Mbp, the *L. iners* genome was only 1.28 Mbp on average (Table 1; Welch's t test, $t = 17.8$, $P < 0.001$). *L. crispatus* was also found to have roughly twice as many open reading frames (ORFs) as *L. iners* (Table 1; Welch's t test, $t = 15.9$, $P < 0.001$). Based on the phylogeny presented in Fig. 1, the reduced genome of *L. iners* appears to be a derived characteristic of the species. In comparison, *L. crispatus* has maintained a larger genome, more similar to that of other vagina-associated *Lactobacillus* species (*L. delbrueckii*, *L. acidophilus*, and *L. johnsonii*).

Next, we sorted the predicted open reading frames into orthologous gene sets using OrthoMCL. These orthologous gene sets were then categorized as either core genes, meaning those present in all strains of a species, or accessory genes, whose presence varies across the strains (43). Additionally, the union of core and accessory genes is defined as the pangenome and contains all orthologous gene sets identified in these species. Consistent with the genome size data, we found that the pangenome of *L. crispatus* had almost twice as many genes as *L. iners* (4,300 versus 2,300 genes; Fig. 2A). Furthermore, the pangenome accumulation curves indicate that this difference is likely to be maintained as more strains are sequenced for each species (Fig. 2A). Similarly, the core genome of *L. crispatus* was larger than that of *L. iners* (1,442 genes versus 993 genes; Fig. 2C). Finally, the accessory genome of *L. crispatus* contained 2,884 genes, of which 45% were present in only one strain (singletons), while the accessory genome of *L. iners* contained only 1,233 genes, of which 56% were singletons (Fig. 2B). These differences in genome size are consistent with *L. crispatus* having access to a broader array of metabolic pathways.

Functional differences and similarities. Niche partitioning by

FIG 1 Maximum likelihood tree of the phylogenetic relationships between the strains of *L. iners* and *L. crispatus* used in this study. The phylogeny was constructed from a partitioned concatenated alignment of the 242 genes shared between the included *L. crispatus* and *L. iners* strains, as well as several outgroup species. Genome size in megabase pairs is mapped onto the tips of the tree to give an idea of how this trait has evolved along the phylogeny.

L. crispatus and *L. iners* through conditional differentiation could be driven by differences in the functional makeup of the two species genomes. To investigate this possibility, we used the Blast-KOALA function from the Kyoto Encyclopedia of Genes and Genomes (KEGG) to assign the core genes of both species to metabolic pathways and functions (Table 2). One might expect that given the larger genome size of *L. crispatus*, this species might have access to a broader array of metabolic functions. In many respects, our functional analysis confirms this expectation. While both *L. crispatus* and *L. iners* rely heavily on fermentation to generate energy, we found that they may differ in respects to the carbon sources they are capable of fermenting. In total, *L. crispatus* has 85 enzymes related to carbohydrate metabolism, whereas *L. iners* has only 59 enzymes (Table 2). Both species have the genetic

capability to metabolize glucose, mannose, maltose, and trehalose. However, only *L. crispatus* has the genetic capability to ferment lactose, galactose, sucrose, and fructose (Fig. 3). Our analysis also indicates that the two species differ in regard to the isomers of lactic acid that they can produce as end products of fermentation: *L. iners* can only produce L-lactic acid, while *L. crispatus* can produce L- and D-lactic acid. Furthermore, we found that the core genome of *L. crispatus* also contains the gene pyruvate oxidase which converts pyruvate into acetate, generating hydrogen peroxide in the process. These differences in the genetic potential for carbon metabolism may influence competitive interactions between these two species.

We found that *L. crispatus* and *L. iners* also differ in their repertoire of enzymes related to the biosynthesis and metabolism of amino acids. The core genome of *L. crispatus* encodes 54 different amino acid-related enzymes, while that of *L. iners* encodes only 43 enzymes (Table 2). More specifically, the core genome of *L. crispatus* has a complete pathway for the biosynthesis of lysine, while the *L. iners* core genome is almost completely devoid of these genes. *L. crispatus* also has more genes related to cysteine and methionine biosynthesis and glycine, serine, and threonine biosynthesis. However, we also found that the two species have similar numbers of genes related to alanine, aspartate, and glutamine metabolism (Table 2). In addition to the genes related to the biosynthesis of the essential amino acids, *L. crispatus*, but not *L. iners*, also has the genetic capability to transport and break down putrescine, a product of ornithine catabolism. These differences are consistent with *L. iners* being more reliant on exogenous sources of amino acids than *L. crispatus*.

Bacterial cells rely on transport proteins to import extracellular supplies of carbohydrates, amino acids, nucleic acids, and inorganic ions. If *L. iners* truly relies more heavily on exogenous sources of nutrients, one might expect that its core genome would

FIG 2 Pangenome, accessory-genome, and core-genome accumulation curves for *Lactobacillus crispatus* (red) and *Lactobacillus iners* (blue). Line thickness represents the 95% confidence interval around the mean.

TABLE 2 Functional category and metabolic pathways encoded in the core genome

Functional category/pathway ^a	No. of core genes	
	<i>L. crispatus</i>	<i>L. iners</i>
Carbohydrate metabolism	85	59
Glycolysis	17	14
Citric acid cycle	3	1
Pentose phosphate pathway	14	12
Fructose and mannose metabolism	18	14
Galactose metabolism	11	8
Starch and sucrose metabolism	16	10
Amino acid metabolism	54	43
Ala, Asp, and Glu metabolism	11	10
Gly, Ser, and Thr metabolism	9	3
Cys and Met metabolism	8	5
Lysine biosynthesis	12	4
Arginine biosynthesis	3	1
Lipid metabolism	21	17
Nucleic acid metabolism	51	56
Metabolism of cofactors and vitamins	33	27
Thiamine metabolism	5	3
Riboflavin metabolism	5	1
Vitamin B ₆ metabolism	2	1
Nicotinate metabolism	5	5
CoA biosynthesis	5	5
Folate biosynthesis	2	5
Membrane transporter	70	54
ABC transporter	39	31
Phosphate transport system	23	15
Bacterial secretion system	8	8
Replication and repair	41	36
DNA replication	14	14
Base excision repair	9	7
Nucleotide excision repair	7	7
Mismatch repair	16	15
Homologous recombination	19	19
Transcription	4	5
Translation	78	79
Peptidoglycan biosynthesis	14	14

^a Entries in bold font represent functional categories while indented entries are specific metabolic pathways within each category. Enzymes can appear in multiple pathways but are only counted once in the functional category total. CoA, coenzyme A.

contain more transport proteins. However, our analysis indicates that the reverse is true. The core genome of *L. crispatus* contains eight more ABC transporter genes and eight more phosphate transport system genes (Table 2). These additional ABC transport genes include a complete phosphonate transporter, an oligopeptide transporter, and an iron transporter. The unique presence of this iron transport system in *L. crispatus* may allow the species to more effectively sequester iron. Interestingly, we found that the core genome of *L. iners* has an ABC-type zinc transporter and an osmoprotectant transporter that are not found in *L. crispatus*. Additionally, the core genome of these two species also contain a number of phosphotransferase system (PTS) genes. *L. crispatus* and *L. iners* share fructose, galactose, glucose, maltose, mannose, and trehalose PTS genes, while only *L. crispatus* has the sucrose

PTS gene. This result is consistent with the unique capability of *L. crispatus* to ferment sucrose.

Our functional analysis also indicates that *L. crispatus* and *L. iners* are functionally similar in many respects (Table 2). This is not surprising given that the two species are closely related (Fig. 1) and that many metabolic pathways are necessary to life. The two species have similar numbers of genes related to the metabolism of lipids, nucleic acids, and cofactors. However, the two species vary somewhat in cofactor pathways: *L. crispatus* has more genes related to the metabolism of riboflavin, while *L. iners* has more genes related to the metabolism of folate. The two species also have similar numbers of genes related to peptidoglycan biosynthesis, transcription, and translation, as well as replication and repair of their chromosome. Included in these replication and repair proteins are complete sets of base excision repair, nucleotide repair excision, mismatch repair, and homologous recombination proteins.

In addition to a core genome, bacterial species also have an accessory genome that contains genes that are present in some but not all strains of a given species. We first split the accessory genome into unique genes, which are those present in only one strain, and variable genes, which are those present in multiple strains. We then characterized the function of the variable genes using the same BlastKOALA approach described above. Unfortunately, this approach only annotated about 16% of this gene set, likely due to its focus on well-annotated metabolic pathways. However, our analysis indicated that the two species differ in the functional makeup of their variable genome (Fig. 4). We found that *L. crispatus* has more variable genes in every functional category except translation, where *L. iners* has slightly more, and replication and repair, where the two species have similar numbers. Specifically, we found that two *L. crispatus* strains contain a complete pathway for the metabolism of L-rhamnose, a deoxy-sugar commonly found in the outer membrane of some bacterial species. Another seven *L. crispatus* strains have the genetic capability to break down raffinose, an oligosaccharide common in plant matter. Additionally, four of the *L. crispatus* strains contained

FIG 3 Graphical representation of the fermentation pathways encoded by the core genome of *L. crispatus* and *L. iners*. Included are those shared by *L. crispatus* and *L. iners* (black) and those unique to *L. crispatus* (red). G6P, glucose-6-phosphate; F6P, fructose 6-phosphate; DHAP, dihydroxyacetone phosphate; GAP, glyceraldehyde 3-phosphate.

FIG 4 Broad functional characteristics of the variable accessory genes found in *L. crispatus* (red) and *L. iners* (blue).

both genes needed for a functional multidrug transport protein. In comparison, our analysis of the *L. iners* variable genes revealed lactose and galactose transporters in only four and six strains, respectively. However, we did not detect the additional enzymes required to route these sugars into central metabolism.

We also functionally characterized the genes that are unique to individual *L. crispatus* and *L. iners* strains, although we found that the BlastKOALA approach we previously used to annotate the core and variable genes was not well suited for this task. Instead, we used Blast searches against NCBI's nonredundant protein database to predict the function of these unique genes. Our analysis indicated that the set of unique *L. crispatus* genes is enriched for transposable elements (Fig. 5A, $P < 0.001$). On average, each *L. crispatus* strain contained approximately 12 unique transposons or integrases in their genome compared to <1 per *L. iners* genome. In contrast, we found that the set of unique genes for *L. iners* was enriched for restriction-modification enzymes (Fig. 5B, $P < 0.001$). While each *L. iners* genome contained approximately three unique restriction-modification enzymes, each *L. crispatus* genome contained only one. Other notable differences between the unique proteins of *L. crispatus* and *L. iners* include that *L. crispatus* has twice as many amino acid metabolism genes and three times as many carbohydrate metabolism genes, while *L. iners* has twice as many replication and repair genes.

Evolutionary differences. Thus far, we have focused on describing differences between these two species based on the presence or absence of specific genes and metabolic pathways. However, further analysis of the sequence diversity within a gene can be used to gain insight into its evolution. These evolutionary analyses can predict the strength of selection acting on the gene and can identify whether a gene has been horizontally transferred into the

FIG 5 Boxplot comparisons of the number of transposable elements (A) and restriction-modification (restr. mod.) enzymes (B) per strain in the set of unique *L. crispatus* and *L. iners* genes. Boxes span the first and third quartiles, with the inner line representing the median value. Whiskers represent $1.5 \times$ the length of the inner quartile range, and points are outliers. Significance was assessed using a two-sided Poisson comparison (***, $P < 0.001$).

genome. We characterized the genetic diversity present in the core genes of *L. crispatus* and *L. iners* using the program TimeZone. Our analysis identified 35,992 single nucleotide polymorphisms (SNPs) across 890 *L. crispatus* core genes and 32,089 SNPs across 791 *L. iners* core genes. Using these SNPs, we estimated the nucleotide diversity (the average number of polymorphisms per base pair of a gene between any two randomly selected sequences) for each gene. Comparisons of the mean nucleotide diversity for each species indicated that the core genome of *L. crispatus* is more diverse than the core genome of *L. iners* (Fig. 6, 0.0173 versus 0.0114, respectively; Welch's t test; $t = 7.0$; $P < 0.001$). However, closer analysis of these data revealed that this result was primarily driven by a series of highly diverse *L. crispatus* genes, including several uncharacterized proteins and transposable elements. A nonparametric Wilcoxon rank sum test demonstrated that the median nucleotide diversity is greater for *L. iners* than *L. crispatus* (0.010 versus 0.006, respectively; $w = 26,218$; $P < 0.001$).

The strength of selection acting on a gene can be inferred based

FIG 6 Scatterplot of nucleotide diversity (x axis) versus dN/dS ratio (y axis) for the core genes of *L. crispatus* (A) and *L. iners* (B). Nucleotide diversity was estimated as the average number of pairwise nucleotide differences per site (Π). Stars in the plot represent the average values for these two parameters, while the colored lines on the axes represent the median values.

TABLE 3 Core genes exhibiting dN/dS ratio >2

Species	Gene name	Annotation	Length (nt)	No. of synonymous changes	No. of nonsynonymous changes	dN/dS^a
<i>L. crispatus</i>	<i>comEB</i>	dCMP deaminase	159	0	4	inf
	LCRIS_RS01075	β -Propeller of dehydrogenase	194	0	3	inf
	<i>purR</i>	PUR operon repressor	276	0	9	inf
	<i>psiE</i>	Phosphate starvation	135	0	7	inf
	<i>ribF</i>	Riboflavin biosynthesis	316	0	6	inf
	LCRIS_RS09970	Haloacid dehydrogenase	257	0	5	inf
	LCRIS_RS09110	Enterocin A immunity	107	0	4	inf
	LCRIS_RS05415	Homocysteine methyltransferase	76	0	3	inf
	LCRIS_RS09270	<i>tet</i> (R) transcription regulator	174	1	9	3.307
	LCRIS_RS03325	α/β -Superfamily hydrolase	306	1	10	3.072
	LCRIS_RS06755	α/β -Superfamily hydrolase	291	1	8	2.871
	LCRIS_RS05305	Lactocepine S-layer protein	166	1	9	2.504
	LCRIS_00991	D-Gluconic acid reductase	278	1	11	2.209
<i>L. iners</i>	RS0100505	Uncharacterized protein	94	0	4	inf
	RS06445	Uncharacterized protein	75	1	9	3.25

^a inf, infinity.

on the ratio of nonsynonymous to synonymous nucleotide changes (dN/dS). Under neutral evolution conditions, the rate of nonsynonymous changes is expected to be equal to the rate of synonymous changes, and their ratio is expected to be one. Values above one indicate an overabundance of nonsynonymous changes, which is usually interpreted as a signature of positive selection, while those under 1 indicate a lack of such changes and usually interpreted as a signature of purifying selection. Of the identified SNPs in the core genes, only 8,946 SNPs (24.8%) and 7,646 SNPs (23.8%) were nonsynonymous mutations in *L. crispatus* and *L. iners*, respectively. The vast majority of the core genes for both species have dN/dS ratios below the neutral expectation of one, indicating that these genes are likely under strong purifying selection (Fig. 6). This result is not unexpected given that the core genes typically carry out essential cellular processes (e.g., transcription, translation, and central metabolism). Our analysis indicates that the median dN/dS ratio for the *L. crispatus* core genome was 0.126, roughly twice the observed median for *L. iners* (Fig. 4; 0.069; Wilcoxon rank sum test, $w = 265,218$; $P < 0.001$). This result is consistent with purifying selection being stronger on the core genome of *L. iners* than on the core genome of *L. crispatus*.

Although the majority of the core genes from both species bear the signature of strong purifying selection (low dN/dS), there is a number which shows signatures of positive selection ($dN/dS > 1$; Fig. 6). We argue that insight into the ecology of the two species in the vaginal environment can be drawn from the products these candidate positively selected genes encode (Table 3). For example, our analysis indicates that the riboflavin synthesis protein RibF may be subject to positive selection in *L. crispatus*. This is interesting because *L. crispatus*, but not *L. iners*, has the riboflavin biosynthesis pathway encoded in its core genome. Another *L. crispatus* core gene with a high dN/dS ratio encodes an enterocin A immunity protein that protects its carrier against the activity of a bacteriocin (44). Our analysis also indicated that the gene encoding the lactocepine S-layer protein (LCRIS_RS05305) had nine nonsynonymous mutations, with only one synonymous (Table 3, $dN/dS \sim 2.5$). This protein has been shown to degrade human proinflammatory chemokines in other *Lactobacillus* species (45). Other can-

didate positively selected genes in *L. crispatus* can be found in Table 3. In comparison, we found that *L. iners* had only two genes, with dN/dS values of >2 (Table 3). Unfortunately, we were unable to annotate either gene beyond “uncharacterized protein.”

Bacterial species can also evolve through the acquisition of foreign DNA through horizontal gene transfer events. The acquisition of foreign DNA can have a dramatic impact on the ecology of bacterial strains and species (46). We used Blast searches against the nonredundant protein database to identify candidate horizontally acquired genes in the core and accessory genomes of *L. iners* and *L. crispatus*. If the Blast searches demonstrated that the gene is present in other *Lactobacillus* species, we inferred that it is likely to have been inherited, whereas if the gene is present only in non-*Lactobacillus* species, we considered it to have been horizontally acquired. Our analysis indicates that only one of the unique core genes of *L. crispatus* is likely to have been horizontally acquired, while we found *L. iners* to have 14 such genes. The single *L. crispatus* unique core gene matched glycosyltransferase found in the *Chlamydia trachomatis* genome. The 14 core genes in *L. iners* that were potentially acquired by horizontal gene transfer matched genes in *Gardnerella vaginalis* ($n = 4$), *Chlamydia trachomatis* ($n = 2$), *Aerococcus christensenii* ($n = 2$), *Parvimonas micra*, *Facklamia hominis*, *Finnegoldia magna*, *Streptococcus* sp., and *Enterococcus faecium*. Most of these species are commonly identified in the human vagina, further reinforcing the notion that they may have been horizontally acquired. These 14 genes include several toxin-antitoxin proteins, a zinc and a phosphate transporter, two DNA repair proteins, and several uncharacterized proteins. Furthermore, our analysis indicated that the cytolysin gene of *L. iners* is also likely to have been horizontally acquired. We found that the *L. iners* sequence for this gene most closely matches cytolysins identified in *G. vaginalis* and various *Streptococcus* species. We extracted these matching sequences from the database and constructed a maximum likelihood tree to identify their phylogenetic relationships (Fig. 7). Our analysis indicated that the *L. iners* cytolysin is most closely related to the *G. vaginalis* cytolysin but has diverged substantially in sequence since being acquired by *L. iners*.

Genes that appear in only a single strain of a species are also

FIG 7 Maximum likelihood tree of the phylogenetic relationships between the *L. iners* cytolysin gene and the closest matching sequences from the NCBI nonredundant nucleotide database. Phylogeny is rooted with the single *Gemella* sp. sequence as the outgroup. sub., substitutions.

likely to have been horizontally acquired. We used the same Blast approach to identify these genes and to determine the closest matching sequence in the database. Our analysis indicated that the majority of the unique genes for both species mostly closely matched genes from other *Lactobacillus* species (*L. crispatus*, 68.7%; *L. iners*, 50.1%). However, we also found that some of these unique accessory genes, both from *L. crispatus* and *L. iners*, did not match genes of other *Lactobacillus* species but instead matched genes from other genera. Notably, we found *L. iners* to have twice as many such genes as *L. crispatus* ($n = 156$ versus $n = 78$, respectively), as well as a more diverse pool of genera to which these sequences matched. These 78 *L. crispatus* genes matched sequences from three different genera: *Chlamydia* ($n = 59$), *Mycoplasma* ($n = 6$), and *Streptococcus* ($n = 3$). The other 10 *L. crispatus* genes matched sequences identified in genomes of multiple genera. In comparison, the 156 *L. iners* genes matched sequences from six different genera: *Streptococcus* ($n = 25$), *Chlamydia* ($n = 20$), *Anaerococcus* ($n = 14$), *Gardnerella* ($n = 13$), *Peptoniphilus* ($n = 13$), and *Atopobium* ($n = 7$). Another 84 of these *L. iners* singleton genes matched sequences identified in genomes of multiple genera. Species from all of the identified genera are routinely found in vaginal samples, supporting the notion that these genes have been acquired through horizontal transfer events.

DISCUSSION

Lactobacillus crispatus and *L. iners* are both common inhabitants of the healthy human vagina. These two species are closely related and perform similar ecological functions, namely, the production of lactic acid. They are rarely found to coexist for extended periods of time, and transitions between an *L. crispatus*-dominated community and one dominated by *L. iners* are common, making it likely that there is substantial overlap in their fundamental niches. Ecological theory predicts that two species cannot occupy the same niche indefinitely (6), making it unclear how these two species have been maintained as common inhabitants of the human vagina. In this study, we characterized and compared the genomes of *L. crispatus* and *L. iners* to identify possible ecological factors that drive niche partitioning by these species. Our results highlight several key differences in the genomes of these two species that we believe may influence their ecology in the vaginal environment.

The typical *L. iners* genome is almost half the size of the *L. crispatus* genome. We have shown that this is likely a derived trait unique to *L. iners* (Fig. 1). The reduced genome size of *L. iners* limits the number of proteins encoded by both its core and accessory genomes, which corresponds to reduced metabolic capabilities.

Our analysis demonstrated that *L. iners* is likely capable of fermenting fewer carbon sources than *L. crispatus* and lacks more of the machinery necessary to synthesize essential amino acids. With fewer metabolic pathways available, *L. iners* likely relies more heavily on exogenous sources for essential resources than *L. crispatus*. Their dependence on nutrients derived from the host or other community members is likely facilitated by the species' ability to bind to human fibronectin, which allows it to maintain close contact with host tissues (47). The limited genetic repertoire of *L. iners* likely makes the species more sensitive to environmental fluctuations. Accordingly, temporal studies on the vaginal microbiome have indicated that communities dominated by this species may be more unstable than those dominated by *L. crispatus* (5, 48).

The process of genome reduction is widespread among host-associated bacterial species and can lead to genomes that are less than 250 kb in size (49–51). We do not completely understand the driving force behind this process, although several mechanisms have been proposed. While we cannot definitively determine what drove the reduction of the *L. iners* genome, we can discuss the consistency of our data with the various proposed mechanisms. First, Muller's Ratchet posits that frequent bottlenecks and a lack of recombination allow slightly deleterious mutations to accumulate in genes, leading to their deterioration and subsequent loss (51, 52). However, we argue that our results are not consistent with this hypothesis, as almost all of the remaining core genes of *L. iners* exhibit signatures consistent with strong purifying selection. Furthermore, we showed that *L. iners* has retained the genes required for several DNA repair pathways, including homologous recombination. Instead, we argue that our results are more consistent with the Black Queen hypothesis, which argues that selection for streamlined genomes and a loss of functional redundancy may be common for host-associated and free-living bacteria that experience relatively constant environments (53, 54). Additional support for this hypothesis is provided by the species' limited accessory genomes, which are enriched for genes encoding mechanisms by which bacteria can resist the integration of foreign DNA into their genome (e.g., restriction-modification enzymes) (55, 56).

While the current *L. crispatus* genome is larger than that of *L. iners*, it is smaller than those of other *Lactobacillus* species, particularly those that are not host associated. We argue that both species have reduced their genome size and differ only in the magnitude of this effect. Some of our results even suggest that *L. crispatus* may still be undergoing this process of genome reduction. We have shown that the accessory genome of this species is bloated with transposable elements. Other studies have demonstrated that the accumulation of mobile genetic elements often precedes genome reduction (51, 57, 58). This effect is thought to be driven by the reduction in the effective population size that often accompanies the transition from being free living to becoming host associated. A smaller effective population size can allow for the fixation of mildly deleterious mutations, like the incorporation of selfish DNA into the genome (52). Later, when these selfish genes are purged from the genome, they may take with them adjacent stretches of the genome. The abundance of transposable elements in the genomes of *L. crispatus* suggest that strains of this species might be currently undergoing genome reduction. This hypothesis is further supported by the genome of *L. crispatus* EM-C1, which is already 400 kb smaller than the others included in this study. In contrast with this evidence, our results also indicate that most of the genes in the *L. crispatus* core genome are experiencing

strong purifying selection, suggesting that selection is still acting efficiently on this species. Only time and further observation will reveal whether *L. crispatus* is evolving along this trajectory.

Lactobacillus crispatus and *L. iners* rely on the fermentation of carbon-containing substrates to produce energy for the cell (59). The primary source of carbon and energy in the vaginal ecosystem is thought to be glycogen produced by host epithelial tissues (14, 15). However, the core genome of neither *L. crispatus* nor *L. iners* contains the enzymes necessary to degrade this polysaccharide. The two species must therefore rely on either the host (60, 61) or other bacterial species for the initial breakdown of glycogen. Our analysis indicated that the two species do have the genetic capability to ferment a number of other sugars, including several glycogen breakdown products. The two species share the ability to ferment glucose, trehalose, maltose, and mannose, of which glucose and maltose are common glycogen breakdown products. In addition to these shared functions, the core genome of *L. crispatus* also includes the enzymes necessary to ferment lactose, galactose, fructose, and sucrose. A previous study showed that the application of a sucrose gel can select for *L. crispatus* in the *Rhesus macaque* vagina (62), lending credence to this result. Furthermore, fructose can be found in high abundance in male ejaculate (63). We speculate that the abundance of these sugars in the vaginal environment may influence competitive interactions between *L. crispatus* and *L. iners*.

One unique trait that may influence the ecology of *L. iners* is its ability to produce inerolysin, a pore-forming cytolysin. Previous studies have demonstrated that this cytolysin is similar in sequence to those produced by *Gardnerella vaginalis* and *Streptococcus intermedius* (64). Our analysis confirms this result and is consistent with this gene having been horizontally acquired by *L. iners*. The gene was not identified in any other *Lactobacillus* species included in this study, and our Blast searches against the nonredundant database did not reveal any matching *Lactobacillus* sequences. Our phylogeny shows that the inerolysin gene is most closely related to that found in *Gardnerella vaginalis*, although its sequence is heavily divergent. This derived trait of *L. iners* may allow it to liberate resources from host cells (64). We speculate that this may give *L. iners* a competitive advantage in the vaginal environment when nutrients are scarce and the ability to liberate them directly from host tissue is favored. Indeed, microbial surveys have suggested that *L. iners* is capable of persisting under other potentially adverse conditions in the vagina (27, 65, 66). Additionally, because the glycogen content of the vaginal epithelium is linked to circulating estrogen levels (14), the abundance of nutrients in the vagina may vary across the female reproductive cycle as well as through a woman's lifetime. If *L. iners* does indeed have a competitive advantage in times of low nutrient abundance, it may also be selected for during times of low circulating estrogen.

We, along with others (64), have demonstrated that the cytolysin of *L. iners* is likely to have been horizontally acquired from a species outside *Lactobacillus*. Our analysis also identified several other core and accessory *L. iners* genes that are likely to have been horizontally acquired. These candidate horizontally acquired genes were found to most closely match those of species in other genera, including *Chlamydia*, *Streptococcus*, *Parvimonas*, *Gardnerella*, and *Atopobium*. In particular, several of these genes, including the cytolysin, match most closely to genes found in *Gardnerella vaginalis*. Historically, this species has been associated with bacterial vaginosis, a condition that increases the risk of women to

preterm birth, sexually transmitted infections, and other adverse sequelae. However, several recent studies have highlighted that *G. vaginalis* is frequently found in healthy asymptomatic vaginal samples from women of all ages (2, 67, 68). In comparison, we identified fewer candidate horizontally acquired genes in *L. crispatus*, and those that we did find matched sequences in a smaller subset of genera. We argue that while there is ongoing gene flow between both *L. iners* and *L. crispatus* and species from these other genera, the rate at which this occurs may be higher for *L. iners*. Furthermore, more of the genes acquired by *L. iners* are conserved across all of the strains analyzed, indicating that they may have been selectively favored in the evolution of the species. We speculate that the acquisition of these genes (e.g., cytolysin, zinc and phosphate transporters, and toxin antitoxin system genes) have shaped the ecology of *L. iners* in the vaginal environment.

Clinicians have long considered the dominance of *L. crispatus* in the vagina to be associated with good health, while only recently have researchers also demonstrated that healthy women can be colonized by *L. iners* (47). Our analysis pointed out several differences in the genomes of *L. crispatus* and *L. iners* that may influence how the species influence host health. Recent studies have demonstrated that the isomers of lactic acid have differential effects on the host immune system (69). We have shown that while *L. crispatus* can produce D- and L-lactic acid, *L. iners* has the capability to produce L-lactic acid only. *L. iners* also does not have the genetic capability to produce hydrogen peroxide via oxidation of pyruvate, a pathway which we and others (29) have demonstrated is available to *L. crispatus*. The production of hydrogen peroxide is thought to be one mechanism by which *Lactobacillus* species can prevent anaerobes from colonizing the vagina (70). *L. crispatus* also has the capability to breakdown putrescine, a malodorous amino acid commonly found in vaginal secretions during episodes of bacterial vaginosis (71). We found the core genome of *L. crispatus* to contain an iron transport system that is absent in the core genome of *L. iners*. This transport system may allow *L. crispatus* to sequester the iron released by the host during menses (72), thereby preventing other species, including vaginal pathogens, from acquiring this vital resource. We also found that the core genome of *L. crispatus* contains a gene encoding lactocepin, a serine protease that has been shown to degrade the proinflammatory chemokine interferon-gamma-inducible protein 10 (IP-10) (45). Our analysis of this gene indicated that it may be experiencing positive selection in *L. crispatus*, which could reflect adaptation via changes in this function. *In vitro* tests have demonstrated that colonization of vaginal epithelial cells with *L. iners* resulted in a more-proinflammatory signaling response from the host tissue than colonization by *L. crispatus* (22). However, it has yet to be determined if the lactocepin produced by *L. crispatus* is capable of preventing or reducing inflammation in the vagina. Based on these two results, we conclude that the presence of *L. crispatus* in the vagina may offer more protective benefits to the host than *L. iners*.

As closely related frequent colonizers of the human vagina, *L. crispatus* and *L. iners* are likely to compete for shared niche space. We have shown that the two species share the key fermentation pathways needed to metabolize glycogen breakdown products (glucose and maltose). However, their genomes differ in many other respects, likely providing them with differential competitive abilities across the range of environmental conditions common to the human vagina. For example, should sucrose or fructose be introduced into the vagina, *L. crispatus* may be selected for given

its unique capability to ferment these sugars, whereas *L. iners* may be favored when nutrients are rare and its cytolysin is needed to liberate them from host tissues. More study is needed to detail the biotic and abiotic factors (e.g., pH, nutrient availability, and co-inhabiting microbial species) under which *L. iners* and *L. crispatus* are favored. Additionally, this study is limited in scope given that it only focused on two of the four common vaginal lactobacilli. Analysis of *L. gasseri* and *L. jensenii* is needed to gain a more complete picture of niche partitioning and competitive interactions in the human vagina. Identifying the conditions under which particular vaginal *Lactobacillus* species flourish will enable a deeper understanding of the complicated temporal dynamics of the vaginal microbiome.

ACKNOWLEDGMENTS

We thank three anonymous reviewers for their insightful suggestions.

This research was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under grant U19AI084044 and by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under grant P30 GM103324.

FUNDING INFORMATION

This work, including the efforts of Larry J. Forney, was funded by HHS | NIH | National Institute of Allergy and Infectious Diseases (NIAID) (U19AI084044). This work, including the efforts of Larry J. Forney, was funded by HHS | NIH | National Institute of General Medical Sciences (NIGMS) (P30 GM103324).

REFERENCES

- Vásquez A, Jakobsson T, Ahrné S, Forsum U, Molin G. 2002. Vaginal *Lactobacillus* flora of healthy Swedish women. *J Clin Microbiol* 40:2746–2749. <http://dx.doi.org/10.1128/JCM.40.8.2746-2749.2002>.
- Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SSK, Mcculle SL, Karlebach S, Gorle R, Russell J, Tacket CO, Brotman RM, Davis CC, Ault K, Peralta L, Forney LJ. 2011. Vaginal microbiome of reproductive-age women. *Proc Natl Acad Sci U S A* 108:4680–4687. <http://dx.doi.org/10.1073/pnas.1002611107>.
- Linhares IM, Summers PR, Larsen B, Giraldo PC, Witkin SS. 2011. Contemporary perspectives on vaginal pH and lactobacilli. *Am J Obstet Gynecol* 204:1–6. <http://dx.doi.org/10.1016/j.ajog.2010.11.002>.
- O'Hanlon DE, Moench TR, Cone RA. 2011. In vaginal fluid, bacteria associated with bacterial vaginosis can be suppressed with lactic acid but not hydrogen peroxide. *BMC Infect Dis* 11:200. <http://dx.doi.org/10.1186/1471-2334-11-200>.
- Gajer P, Brotman RM, Bai G, Sakamoto J, Schutte UME, Zhong X, Koenig SSK, Fu L, Ma Z, Zhou X, Abdo Z, Forney LJ, Ravel J. 2012. Temporal dynamics of the human vaginal microbiota. *Sci Transl Med* 4:132ra52.
- Gause GF. 1934. The struggle for existence. Williams & Wilkins, Baltimore, MD.
- Schoener TW. 1974. Resource partitioning in ecological communities. *Science* 185:27–39. <http://dx.doi.org/10.1126/science.185.4145.27>.
- Hunt DE, David LA, Gevers D, Preheim SP, Alm EJ, Polz MF. 2008. Resource partitioning and sympatric differentiation among closely related bacterioplankton. *Science* 320:1081–1085. <http://dx.doi.org/10.1126/science.1157890>.
- Mendes-Soares H, Rychlik L. 2009. Differences in swimming and diving abilities between two sympatric species of water shrews: *Neomys anomalus* and *Neomys fodiens* (Soricidae). *J Ethol* 27:317–325. <http://dx.doi.org/10.1007/s10164-008-0122-z>.
- Chesson P. 2000. Mechanisms of maintenance of species diversity. *Annu Rev Ecol Syst* 31:343–366. <http://dx.doi.org/10.1146/annurev.ecolsys.31.1.343>.
- Amarasekare P. 2003. Competitive coexistence in spatially structured environments: a synthesis. *Ecol Lett* 6:1109–1122. <http://dx.doi.org/10.1046/j.1461-0248.2003.00530.x>.
- Mammola S, Isaia M. 2014. Niche differentiation in *Meta bournetii* and *M. menardi* (Araneae, Tetragnathidae) with notes on the life history. *Int J Speleol* 43:343–353.
- Wolf DP, Sokoloski JE, Litt M. 1980. Composition and function of human cervical mucus. *Biochim Biophys Acta* 630:545–558. [http://dx.doi.org/10.1016/0304-4165\(80\)90009-4](http://dx.doi.org/10.1016/0304-4165(80)90009-4).
- Cruikshank R, Sharman A. 1934. The biology of the vagina in the human subject. *BJOG* 41:208–226.
- Milwidsky A, Palti Z, Gutman A. 1980. Glycogen metabolism of the human endometrium. *J Clin Endocrinol Metab* 51:765–770. <http://dx.doi.org/10.1210/jcem-51-4-765>.
- Bigelow JL, Dunson DB, Stanford JB, Ecohard R, Gnoth C, Colombo B. 2004. Mucus observations in the fertile window: a better predictor of conception than timing of intercourse. *Hum Reprod* 19:889–892. <http://dx.doi.org/10.1093/humrep/deh173>.
- Farage M, Maibach H. 2006. Lifetime changes in the vulva and vagina. *Arch Gynecol Obstet* 273:195–202. <http://dx.doi.org/10.1007/s00404-005-0079-x>.
- Grande G, Milardi D, Vincenzoni F, Pompa G, Biscione A, Astorri AL, Fruscella E, De Luca A, Messina I, Castagnola M, Marana R. 2015. Proteomic characterization of the qualitative and quantitative differences in cervical mucus composition during the menstrual cycle. *Mol Biosyst* 11:1717–1725. <http://dx.doi.org/10.1039/C5MB00071H>.
- Mirmonsef P, Modur S, Burgad D, Gilbert D, Golub ET, French AL, McCotter K, Landay AL, Spear GT. 2015. Exploratory comparison of vaginal glycogen and *Lactobacillus* levels in premenopausal and postmenopausal women. *Menopause* 22:702–709. <http://dx.doi.org/10.1097/GME.0000000000000397>.
- Wira CR, Fahey JV, Sentman CL, Pioli PA, Shen L. 2005. Innate and adaptive immunity in female genital tract: cellular responses and interactions. *Immunol Rev* 206:306–335. <http://dx.doi.org/10.1111/j.0105-2896.2005.00287.x>.
- Wira CR, Fahey JV, White HD, Yeaman GR, Givan AL, Howell AL. 2002. The mucosal immune system in the human female reproductive tract: influence of stage of the menstrual cycle and menopause on mucosal immunity in the uterus, p 371–404. In Glass S, Aplin J, Guidice L, Tabibzadeh S (ed), *The endometrium*. Francis & Taylor, New York, NY.
- Doerflinger SY, Throop AL, Herbst-Kralovetz MM. 2014. Bacteria in the vaginal microbiome alter the innate immune response and barrier properties of the human vaginal epithelia in a species-specific manner. *J Infect Dis* 209:1989–1999. <http://dx.doi.org/10.1093/infdis/jiu004>.
- Tamrakar R, Yamada T, Furuta I, Cho K, Morikawa M, Yamada H, Sakuragi N, Minakami H. 2007. Association between *Lactobacillus* species and bacterial vaginosis-related bacteria, and bacterial vaginosis scores in pregnant Japanese women. *BMC Infect Dis* 7:128. <http://dx.doi.org/10.1186/1471-2334-7-128>.
- Petricic L, Domig KJ, Nierscher FJ, Sandhofer MJ, Fidesser M, Krondorfer I, Husslein P, Kneifel W, Kiss H. 2014. Characterisation of the vaginal *Lactobacillus* microbiota associated with preterm delivery. *Sci Rep* 4:5136.
- Wertz J, Isaacs-Cosgrove N, Holzman C, Marsh TL. 2008. Temporal shifts in microbial communities in nonpregnant African-American women with and without bacterial vaginosis. *Interdiscip Perspect Infect Dis* 2008:181253.
- Fredricks DN, Fiedler TL, Marrazzo JM. 2005. Molecular identification of bacteria associated with bacterial vaginosis. *N Engl J Med* 353:1899–1911. <http://dx.doi.org/10.1056/NEJMoa043802>.
- Ferris MJ, Norori J, Zozaya-Hinchliffe M, Martin DH. 2007. Cultivation-independent analysis of changes in bacterial vaginosis flora following metronidazole treatment. *J Clin Microbiol* 45:1016–1018. <http://dx.doi.org/10.1128/JCM.02085-06>.
- Ojala T, Kuparinen V, Koskinen JP, Alatalo E, Holm L, Auvinen P, Edelman S, Westerlund-Wikström B, Korhonen TK, Paulin L, Kankainen M. 2010. Genome sequence of *Lactobacillus crispatus* ST1. *J Bacteriol* 192:3547–3548. <http://dx.doi.org/10.1128/JB.00399-10>.
- Ojala T, Kankainen M, Castro J, Cerca N, Edelman S, Westerlund-Wikström B, Paulin BL, Holm L, Auvinen P. 2014. Comparative genomics of *Lactobacillus crispatus* suggests novel mechanisms for the competitive exclusion of *Gardnerella vaginalis*. *BMC Genomics* 15:1–21. <http://dx.doi.org/10.1186/1471-2164-15-1070>.
- Mendes-Soares H, Suzuki H, Hickey RJ, Forney LJ. 2014. Comparative functional genomics of *Lactobacillus* spp. reveals possible mechanisms for specialization of vaginal lactobacilli to their environment. *J Bacteriol* 196:1458–1470. <http://dx.doi.org/10.1128/JB.01439-13>.

31. Salzberg SL, Delcher AL, Kasif S, White O. 1998. Microbial gene identification using interpolated Markov models. *Nucleic Acids Res* 26:544–548. <http://dx.doi.org/10.1093/nar/26.2.544>.
32. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. 1999. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* 27:4636–4641. <http://dx.doi.org/10.1093/nar/27.23.4636>.
33. Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23:673–679. <http://dx.doi.org/10.1093/bioinformatics/btm009>.
34. Li L, Stoekert CJ, Jr, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 13:2178–2189. <http://dx.doi.org/10.1101/gr.1224503>.
35. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and Ghost-KOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* 428:726–731. <http://dx.doi.org/10.1016/j.jmb.2015.11.006>.
36. Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402. <http://dx.doi.org/10.1093/nar/25.17.3389>.
37. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948. <http://dx.doi.org/10.1093/bioinformatics/btm404>.
38. Smith SA, Dunn CW. 2008. Phyutility: a phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics* 24:715–716. <http://dx.doi.org/10.1093/bioinformatics/btm619>.
39. Lanfear R, Calcott B, Ho SYW, Guindon S. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol Biol Evol* 29:1695–1701. <http://dx.doi.org/10.1093/molbev/mss020>.
40. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313. <http://dx.doi.org/10.1093/bioinformatics/btu033>.
41. Tavaré S. 1986. Some probabilities and statistical problems in the analysis of DNA sequences. *Lect Math Life Sci* 17:57–86.
42. Chattopadhyay S, Paul S, Dykhuizen DE, Sokurenko EV. 2013. Tracking recent adaptive evolution in microbial species using TimeZone. *Nat Protoc* 8:652–665. <http://dx.doi.org/10.1038/nprot.2013.031>.
43. Medini D, Donati C, Tettelin H, Massignani V, Rappuoli R. 2005. The microbial pan-genome. *Curr Opin Genet Dev* 15:589–594. <http://dx.doi.org/10.1016/j.gde.2005.09.006>.
44. O’Keeffe T, Hill C, Ross RP. 1999. Characterization and heterologous expression of the genes encoding enterocin A production, immunity, and regulation in *Enterococcus faecium* DPC1146. *Appl Environ Microbiol* 65:1506–1515.
45. von Schillde MA, Hörmannspurger G, Weiher M, Alpert CA, Hahne H, Bäuerl C, Van Huynegem K, Steidler L, Hrnčir T, Pérez-Martínez G, Kuster B, Haller D. 2012. Lactocin secreted by *Lactobacillus* exerts anti-inflammatory effects by selectively degrading proinflammatory chemokines. *Cell Host Microbe* 11:387–396. <http://dx.doi.org/10.1016/j.chom.2012.02.006>.
46. Gyles C, Boerlin P. 2014. Horizontally transferred genetic elements and their role in pathogenesis of bacterial disease. *Vet Pathol* 51:328–340. <http://dx.doi.org/10.1177/0300985813511131>.
47. McMillan A, Macklaim JM, Burton JP, Reid G. 2012. Adhesion of *Lactobacillus iners* AB-1 to human fibronectin: a key mediator for persistence in the vagina? *Reprod Sci* 20:791–796.
48. Jakobsson T, Forsum U. 2007. *Lactobacillus iners*: a marker of changes in the vaginal flora? *J Clin Microbiol* 45:3145. <http://dx.doi.org/10.1128/JCM.00558-07>.
49. Andersson SG, Kurland CG. 1998. Reductive evolution of resident genomes. *Trends Microbiol* 6:263–268. [http://dx.doi.org/10.1016/S0966-842X\(98\)01312-2](http://dx.doi.org/10.1016/S0966-842X(98)01312-2).
50. Moran NA, Wernegreen JJ. 2000. Lifestyle evolution in symbiotic bacteria: insights from genomics. *Trends Ecol Evol* 15:321–326. [http://dx.doi.org/10.1016/S0169-5347\(00\)01902-9](http://dx.doi.org/10.1016/S0169-5347(00)01902-9).
51. McCutcheon JP, Moran NA. 2012. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* 10:13–26.
52. Moran NA. 1996. Accelerated evolution and Muller’s ratchet in endosymbiotic bacteria. *Proc Natl Acad Sci U S A* 93:2873–2878. <http://dx.doi.org/10.1073/pnas.93.7.2873>.
53. Bliven KA, Maurelli AT. 2012. Antivirulence genes: insights into pathogen evolution through gene loss. *Infect Immun* 80:4061–4070. <http://dx.doi.org/10.1128/IAI.00740-12>.
54. Morris JJ, Lenski RE, Zinser ER. 2012. The Black Queen hypothesis: evolution of dependencies through adaptive gene loss. *mBio* 3(2):e00036-12. <http://dx.doi.org/10.1128/mBio.00036-12>.
55. Bickle TA, Krüger DH. 1993. Biology of DNA restriction. *Microbiol Rev* 57:434–450.
56. Snyder L. 1995. Phage-exclusion enzymes: a bonanza of biochemical and cell biology reagents? *Mol Microbiol* 15:415–420. <http://dx.doi.org/10.1111/j.1365-2958.1995.tb02255.x>.
57. Toh H, Weiss BL, Perkin SAH, Yamashita A, Oshima K, Hattori M, Aksoy S. 2006. Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. *Genome Res* 16:149–156.
58. Burke GR, Moran NA. 2011. Massive genomic decay in *Serratia symbiotica*, a recently evolved symbiont of aphids. *Genome Biol Evol* 3:195–208. <http://dx.doi.org/10.1093/gbe/evr002>.
59. Carr FJ, Chill D, Maida N. 2002. The lactic acid bacteria: a literature survey. *Crit Rev Microbiol* 28:281–370. <http://dx.doi.org/10.1080/1040-840291046759>.
60. Nasioudis D, Beghini J, Bongiovanni A, Giraldo P, Linhares I, Witkin SS. 2015. α -Amylase in vaginal fluid: association with conditions favorable to dominance of *Lactobacillus*. *Reprod Sci* 22:1393–1398. <http://dx.doi.org/10.1177/1933719115581000>.
61. Spear GT, French AL, Gilbert D, Zariffard MR, Mirmonsef P, Sullivan TH, Spear WW, Landay A, Micci S, Lee BH, Hamaker BR. 2014. Human α -amylase present in lower-genital-tract mucosal fluid processes glycogen to support vaginal colonization by *Lactobacillus*. *J Infect Dis* 210:1019–1028. <http://dx.doi.org/10.1093/infdis/jiu231>.
62. Hu KT, Zheng JX, Yu ZJ, Chen Z, Cheng H, Pan WG, Yang WZ, Wang HY, Deng QW, Zeng ZM. 2015. Directed shift of vaginal microbiota induced by vaginal application of sucrose gel in *Rhesus macaques*. *Int J Infect Dis* 33:32–36. <http://dx.doi.org/10.1016/j.ijid.2014.12.040>.
63. Owed DH, Katz DF. 2005. A review of the physical and chemical properties of human semen and the formulation of a semen simulant. *J Androl* 26:459–469. <http://dx.doi.org/10.2164/jandrol.04104>.
64. Rampersaud R, Planet PJ, Randis TM, Kulkarni R, Aguilar JL, Lehrer RI, Ratner AJ. 2011. Inerolysin, a cholesterol-dependent cytolysin produced by *Lactobacillus iners*. *J Bacteriol* 193:1034–1041. <http://dx.doi.org/10.1128/JB.00694-10>.
65. Zozaya-Hinchliffe M, Lillis R, Martin DH, Ferris MJ. 2010. Quantitative PCR assessments of bacterial species in women with and without bacterial vaginosis. *J Clin Microbiol* 48:1812–1819. <http://dx.doi.org/10.1128/JCM.00851-09>.
66. Mayer BT, Srinivasan S, Fiedler TL, Marrazzo JM, Fredricks DN, Schiffer JT. 2015. Rapid and profound shifts in the vaginal microbiota following antibiotic treatment for bacterial vaginosis. *J Infect Dis* 212:793–802. <http://dx.doi.org/10.1093/infdis/jiv079>.
67. Hillier SL, Lau RJ. 1997. Vaginal microflora in postmenopausal women who have not received estrogen replacement therapy. *Clin Infect Dis* 25:123–126.
68. Hickey RJ, Zhou X, Settles ML, Erb J, Malone K, Hansmann MA, Shew ML, Van Der Pol B, Fortenberry JD, Forney LJ. 2015. Vaginal microbiota of adolescent girls prior to the onset of menarche resemble those of reproductive-age women. *mBio* 6(2):e00097-15. <http://dx.doi.org/10.1128/mBio.00097-15>.
69. Witkin SS, Mendes-Soares H, Linhares IM, Jayaram A, Ledger WJ, Forney LJ. 2013. Influence of vaginal bacteria and D- and L-lactic acid isomers on vaginal extracellular matrix metalloproteinase inducer: implications for protection against upper genital tract infections. *mBio* 4(4):e00460-13. <http://dx.doi.org/10.1128/mBio.00450-13>.
70. Felten A, Barreau C, Bizet C, Lagrange PH, Philippon A. 1999. *Lactobacillus* species identification, H₂O₂ production, and antibiotic resistance and correlation with human clinical status. *J Clin Microbiol* 37:729–733.
71. Wolrath H, Forsum U, Larsson PG, Borén H. 2001. Analysis of bacterial vaginosis-related amines in vaginal fluid by gas chromatography and mass spectrometry. *J Clin Microbiol* 39:4026–4031. <http://dx.doi.org/10.1128/JCM.39.11.4026-4031.2001>.
72. Hallberg L, Hogdahl A-M, Nilsson L, Rybo G. 1966. Menstrual blood loss and iron deficiency. *Acta Med Scand* 180:639–650.