Genome Biology

RESEARCH

Insight into the ecology of vaginal bacteria through integrative analyses of metagenomic and metatranscriptomic data

Michael T. France[1], Li Fu[1], Lindsay Rutt[1], Hongqiu Yang[1], Michael S. Humphrys[1], Shilpa Narina[1], Pawel M. Gajer[1], Bing Ma[1], Larry J. Forney[2] and Jacques Ravel[1*]

*Correspondence:
jravel@som.umaryland.edu
[1] Institute for Genome
Sciences and Department
of Microbiology
and Immunology, University
of Maryland School
of Medicine, Baltimore, MD,
USA
Full list of author information
is available at the end of the
article

Abstract

Background: Vaginal bacterial communities dominated by *Lactobacillus* species are associated with a reduced risk of various adverse health outcomes. However, somewhat unexpectedly, many healthy women have microbiota that are not dominated by lactobacilli. To determine the factors that drive vaginal community composition we characterized the genetic composition and transcriptional activities of vaginal microbiota in healthy women.

Results: We demonstrate that the abundance of a species is not always indicative of its transcriptional activity and that impending changes in community composition can be predicted from metatranscriptomic data. Functional comparisons highlight differences in the metabolic activities of these communities, notably in their degradation of host produced mucin but not glycogen. Degradation of mucin by communities not dominated by *Lactobacillus* may play a role in their association with adverse health outcomes. Finally, we show that the transcriptional activities of *L. crispatus*, *L. iners*, and *Gardnerella vaginalis* vary with the taxonomic composition of the communities in which they reside. Notably, *L. iners* and *G. vaginalis* both demonstrate lower expression of their cholesterol-dependent cytolysins when co-resident with *Lactobacillus* spp. and higher expression when co-resident with other facultative and obligate anaerobes. The pathogenic potential of these species may depend on the communities in which they reside and thus could be modulated by interventional strategies.

Conclusions: Our results provide insight to the functional ecology of the vaginal microbiota, demonstrate the diagnostic potential of metatranscriptomic data, and reveal strategies for the management of these ecosystems.

Keywords: Vaginal microbiome, Microbial ecology, Metatranscriptome, Metagenome, Host-microbe interactions

France *et al. Genome Biology*   (2022) 23:66

Page 2 of 26

Background

The microbial communities that inhabit the human body are crucial determinants of health [1–3]. Of these, the microbiota that inhabit the human vagina are unique [4, 5]. They differ dramatically in taxonomic composition from the communities found in the vaginas of other mammals, including non-human primates [6], and often have a skewed structure in which the communities are dominated by a single species of *Lactobacillus* [7–11].

The vaginal microbiota in reproductive age women can be roughly divided into five distinct community state types (CSTs) based on species composition [7, 9]. Four of these are characterized by a high relative abundance of different *Lactobacillus* species (CST I-*L. crispatus* dominated, CST II-*L. gasseri* dominated, CST III-*L. iners* dominated, CST V-*L. jensenii* dominated), while the fifth (CST IV) is characterized by a more diverse assemblage of facultative and obligate anaerobes. Studies have concluded that women with communities dominated by *Lactobacillus* spp. are less likely to experience a number of adverse health outcomes, supporting a protective role for these species [12–20]. Lactobacilli are thought to provide protection through their production of lactic acid which acidifies the environment, although other factors are likely to contribute as well [21–25]. Women whose microbiota are largely depleted of lactobacilli are at greater risk to symptomatic bacterial vaginosis (BV), sexually transmitted diseases, and preterm birth [26–30]. Interestingly, the vaginal microbiota has also been shown to sometimes exhibit changes in composition through time, sometimes demonstrating dramatic and rapid changes from being dominated by *Lactobacillus* spp. to CST IV-like communities [31, 32]. The factors that drive these patterns are not currently understood.

Most studies of vaginal microbiota have relied on amplicon sequencing of variable regions of 16S rRNA genes to assess community composition, although a growing number of studies have employed shotgun sequencing of the vaginal metagenome [33–38]. Only a handful of studies, with few subjects, have examined the vaginal metatranscriptome. For example, a small study by Macklaim, Fernandes [39] focused on the transcriptional activity of *L. iners* in four reproductive age women and found variation in the species' transcriptional activity. A more recent study examined the metatranscriptome of 14 reproductive age women diagnosed with BV and treated with metronidazole [40]. The authors found a set of *Gardnerella vaginalis* CRISPR-associated Cas genes that were upregulated in patients who did not respond to treatment. Finally, a recent study examined the metatranscriptome of over 100 pregnant women in order to identify microbiota-driven causes of preterm birth (PTB, [36]). They found a number of genes whose expression was associated with either PTB or term birth, including a number of genes encoding bacterial secretion systems or predicted secreted proteins. A concurrent study from the same group described the vaginal composition and transcriptional activity in pregnant women from a number of racial-ethnic backgrounds [37] and reported a number of pathways associated with nucleotide metabolism whose expression was lower in women of African descent. Although the results of these studies have been illuminating, much remains to be learned from analyzing the vaginal metatranscriptome. Fine-scale, integrative analyses of these data have the potential to derive translational in vivo mechanistic
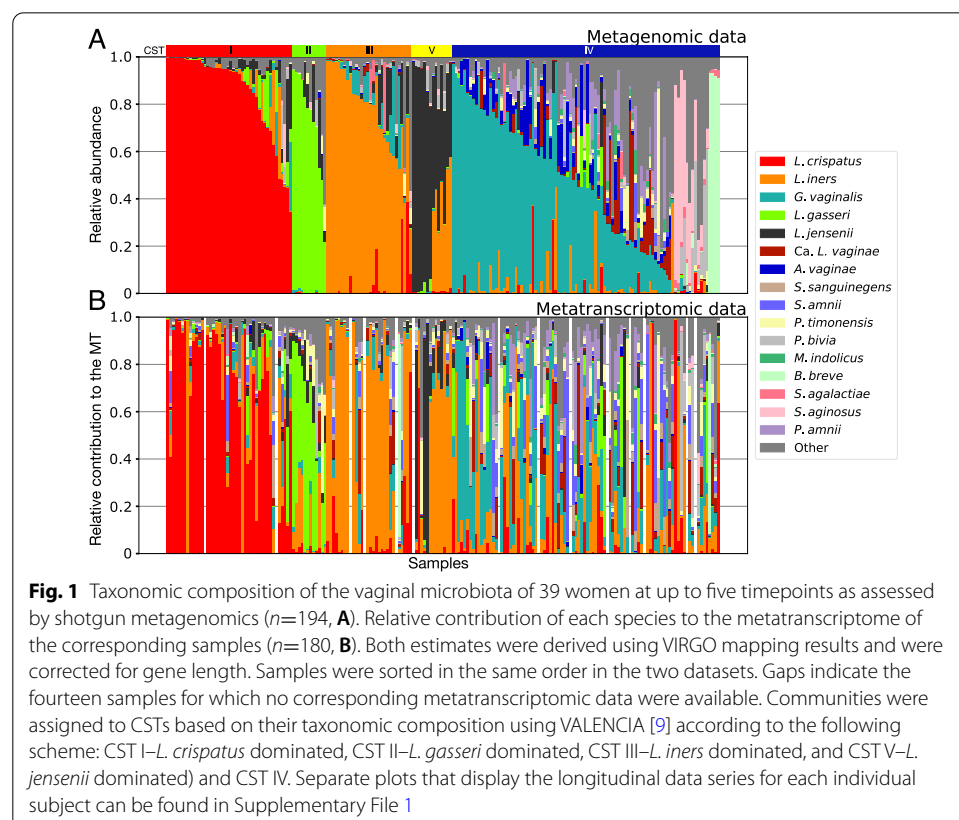
understanding of the role of the vaginal microbiome in health and disease by characterizing interactions between community members and between the microbiota and host tissues.

Here, we present an analysis of 194 metagenomic and 180 metatranscriptomic datasets from 39 non-pregnant reproductive age women who self-collected vaginal swabs daily over the course of 10 weeks [41]. These women self-identified as belonging to a number of different racial backgrounds and were found to have vaginal microbiota classified to different CST. We perform integrative analyses of these data in order gain insight into the functional ecology of the vaginal microbiota.

Results

The relative abundance of a species is not necessarily indicative of its contribution to the metatranscriptome
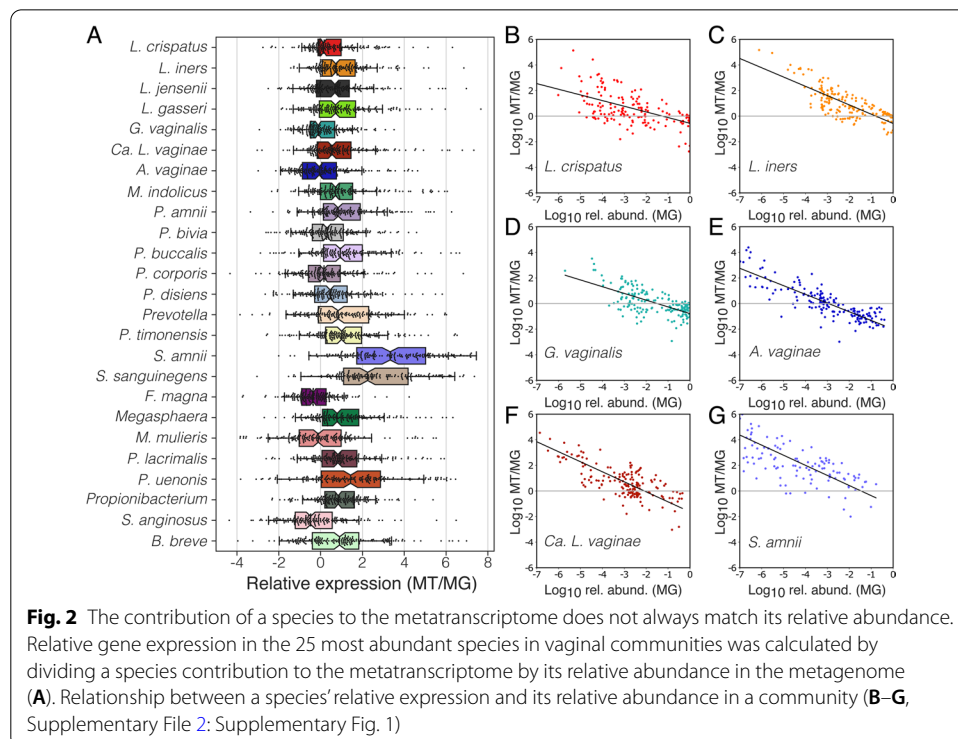
We characterized the vaginal microbiota of 39 North American reproductive age women (age range: 19–45) at five timepoints, spaced 2 weeks apart. Subjects included in the study self-identified as either Black or African American (n=24), White or Caucasian (n=10), Hispanic or Latino (n=4), and Asian (n=1). The metagenomes (n=194) and metatranscriptomes (n=180) of these samples were sequenced and the resulting reads were mapped to VIRGO, a non-redundant and comprehensive gene catalog of the vaginal microbiome [34] to determine the taxonomic composition of

Fig. 1 Taxonomic composition of the vaginal microbiota of 39 women at up to five timepoints as assessed by shotgun metagenomics (n=194, **A**). Relative contribution of each species to the metatranscriptome of the corresponding samples (n=180, **B**). Both estimates were derived using VIRGO mapping results and were corrected for gene length. Samples were sorted in the same order in the two datasets. Gaps indicate the fourteen samples for which no corresponding metatranscriptomic data were available. Communities were assigned to CSTs based on their taxonomic composition using VALENCIA [9] according to the following scheme: CST I–*L. crispatus* dominated, CST II–*L. gasseri* dominated, CST III–*L. iners* dominated, and CST V–*L. jensenii* dominated) and CST IV. Separate plots that display the longitudinal data series for each individual subject can be found in Supplementary File 1

both the metagenome and metatranscriptome (Fig. 1). The composition of a metagenome represents the relative abundances of taxa, whereas the composition of the metatranscriptome represents the relative activities of taxa (Supplementary Table 1).

We first determined whether the relative abundance of a species was consistent with its relative transcriptional activity. The ratio of a species relative activity divided by its relative abundance was referred to as its "relative expression" and represents the extent to which the species is over- or under-represented in the metatranscriptome, as compared to the metagenome. We calculated these relative expression values for the twenty-five most prevalent species study-wide which included the four common vaginal *Lactobacillus* species, *G. vaginalis, Atopobium vaginae, Ca.* Lachnocurva vaginae, and *Prevotella* spp., as well as several other facultative or obligate anaerobic bacteria (Fig. 2A). For each species and across samples, a wide range of relative expression values were observed ranging from over- to under-expressive. *G. vaginalis, A. vaginae, Finegoldia magna, Mobiluncus mulieris,* and *Streptococcus anginosus* were more likely to be under-expressive (median relative expression <0, Fig. 2A), while the other twenty species were more likely to be over-expressive (median relative expression >0), especially *Sneathia amnii* (median relative expression = 3.3, Fig. 2A) and *S. sanguinegens* (median relative expression = 2.2, Fig. 2A).

We next asked whether a species' observed relative expression was correlated with its' relative abundance. Because both datasets are compositional, our evaluation of relative expression was bounded at the highest and lowest relative abundances. If we estimate a species' relative abundance to be 90%, the maximum value for its relative expression is ~1.1. Conversely, if a species comprises a small fraction of the community, the minimum value for its relative expression is constrained by the depth at which the

Fig. 2 The contribution of a species to the metatranscriptome does not always match its relative abundance. Relative gene expression in the 25 most abundant species in vaginal communities was calculated by dividing a species contribution to the metatranscriptome by its relative abundance in the metagenome (**A**). Relationship between a species' relative expression and its relative abundance in a community (**B**–**G**, Supplementary File 2: Supplementary Fig. 1)
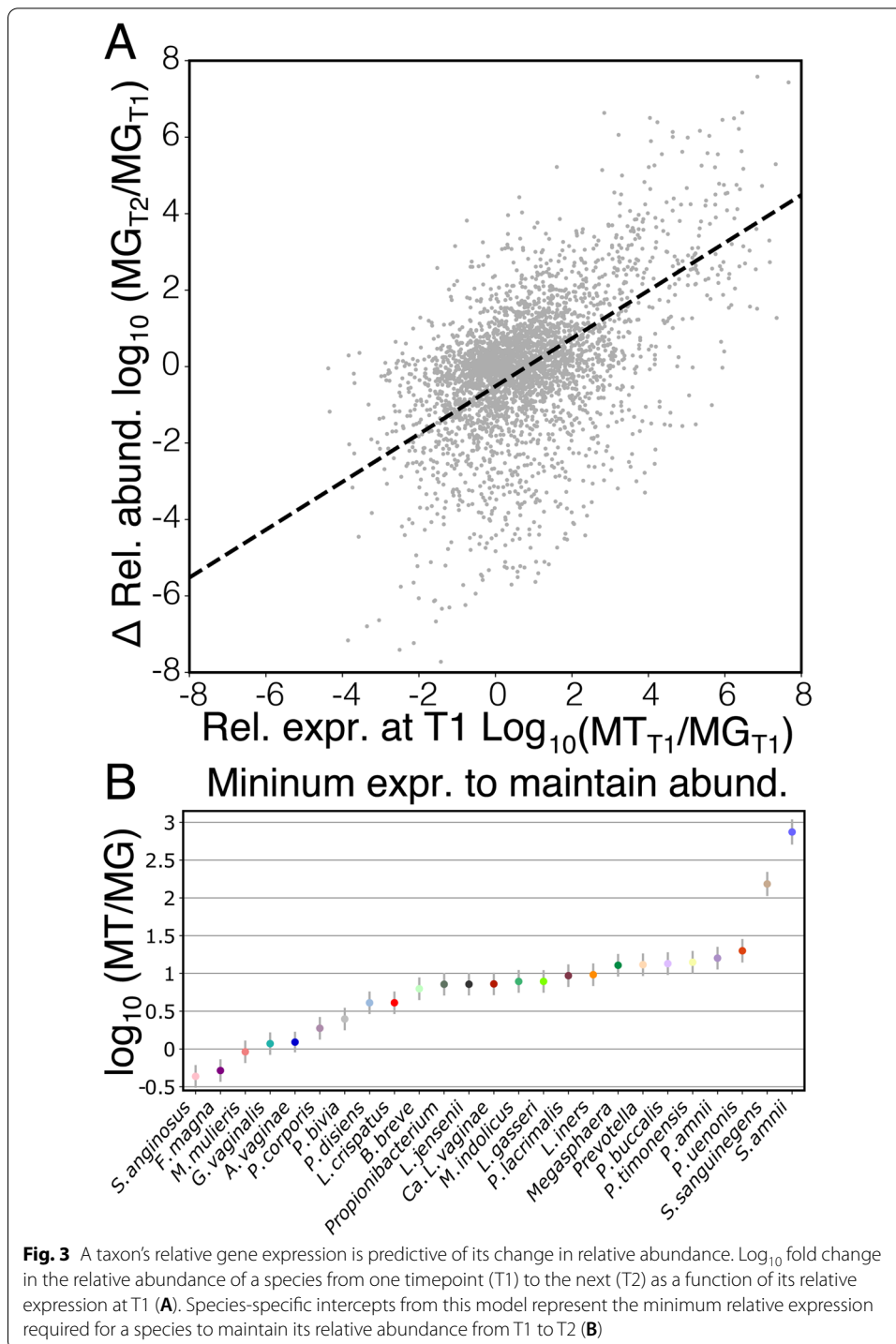
metatranscriptome was sequenced. We found a strong relationship between the abundance of a species and its' relative expression (Fig. 2B–G; Supplementary File 2: Supplementary Fig. 1; linear mixed model; $F_{1,4261}$=7167, p<0.001). When rare, vaginal bacteria were found to contribute more to the metatranscriptome than their relative abundance would suggest. Whereas species found at higher relative abundances tended to under contribute to the metatranscriptome. The relative abundance at which this switch from over- to under-expressive occurs varied by species (Fig. 2B–G; Supplementary File 2: Supplementary Fig. 1; linear mixed model; $F_{24,4261}$=198, p<0.001). Species such as *L. crispatus*, *L. iners*, and *G. vaginalis* (Fig. 2B–D), which tended to be found at higher relative abundances, maintained over-expression when common (x-intercept >10^{-2} relative abundance). Conversely, species like *A. vaginae* and *Ca.* L. vaginae (Fig. 2E, F), which tended to be found at moderate or low relative abundances, transitioned to under-expression at lower relative abundances (x-intercept <10^{-2} relative abundance). *S. amnii* (Fig. 2G) and *S. sanguinegens* (Supplementary File 2: Supplementary Fig. 1) are both routinely estimated to constitute a small proportion of the community, yet they were found to vastly over-contribute to the metatranscriptome.

A taxon's contribution to the metatranscriptome is predictive of future changes in its abundance
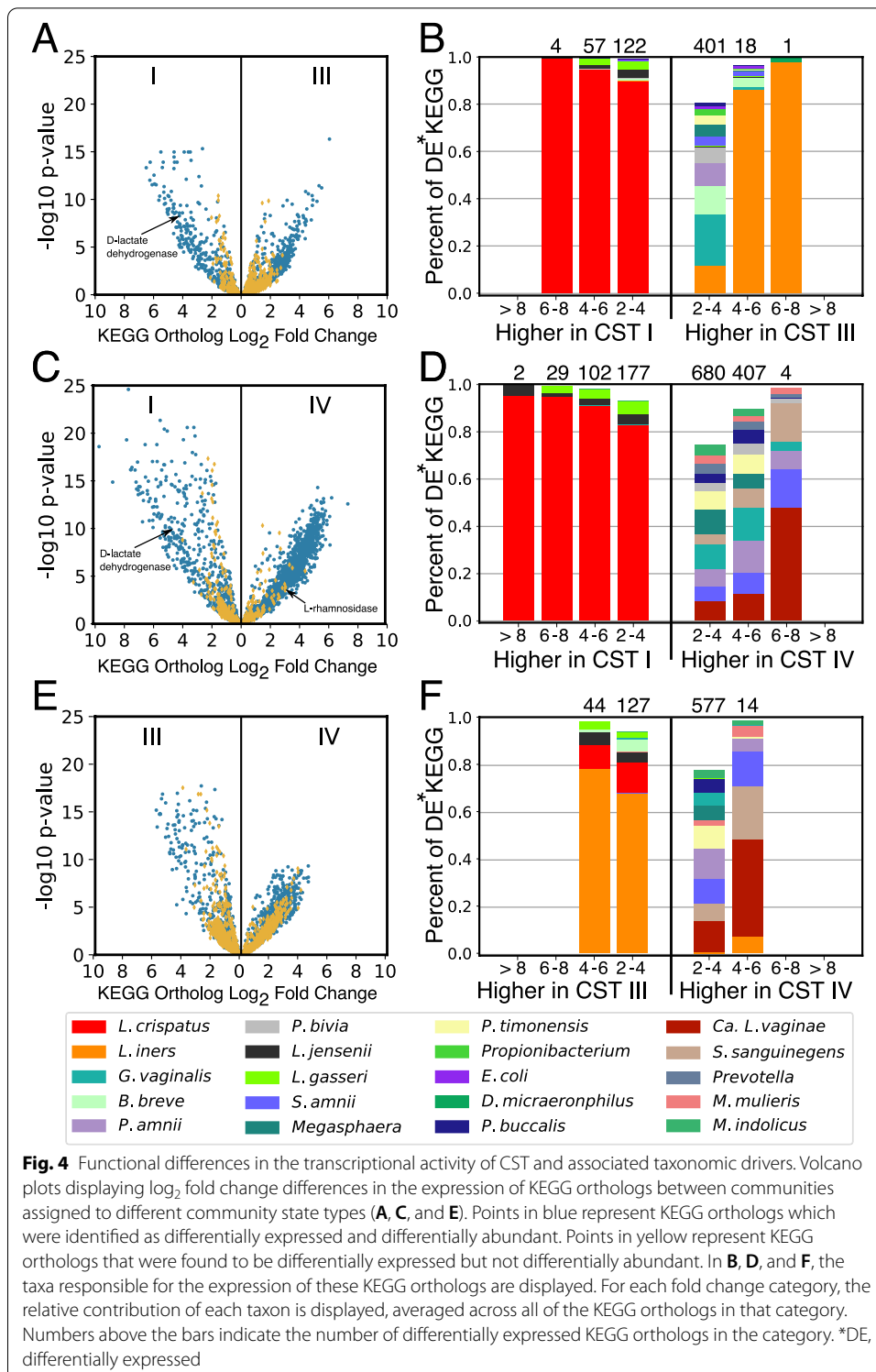
The relative transcriptional activity of a species or strain may have some bearing on its future abundance in the microbial community. One might expect that species which demonstrate a low relative expression may experience a decrease in relative abundance, while those that demonstrate a high relative expression might increase. Our longitudinal dataset allowed us to test this hypothesis. We assessed the change in the $\log_{10}$ transformed relative abundance of species between pairs of consecutive timepoints, spaced 2 weeks apart. As hypothesized, we found a strong relationship between a species' change in relative abundance and its level of expression at the earlier timepoint (Fig. 3A; linear mixed model; $F_{1,1847}$=649, p<0.001). Species which demonstrated low relative expression were found to decrease in relative abundance while those that demonstrated high relative expression were found to increase. We also found support for species-specific intercepts in this relationship (Fig. 3B; linear mixed model; $F_{24,1847}$=11.1, p<0.001), which represent the relative expression necessary for the species to maintain its relative abundance over the 2-week interval. Most taxa needed to be slightly over-expressive to maintain their relative abundance. Notable exceptions were *S. anginosus* and *F. magna* which were found to be able to maintain their relative abundance while being under-expressive, and *S. amnii* and *S. sanguinegens*, which needed to be highly expressive to maintain their relative abundance.

Functional differences in the transcriptional activity of communities assigned to different CSTs

By definition, CSTs have substantial differences in composition, and consequently, it is expected that there will also be substantial differences in their functional capacity. To characterize these differences, we performed a differential expression analysis at the level of KEGG orthologs, which represent molecular functions [42]. We focused our analysis on comparisons of CST I, CST III, and CST IV, as there were

Fig. 3 A taxon's relative gene expression is predictive of its change in relative abundance. Log_{10} fold change in the relative abundance of a species from one timepoint (T1) to the next (T2) as a function of its relative expression at T1 (**A**). Species-specific intercepts from this model represent the minimum relative expression required for a species to maintain its relative abundance from T1 to T2 (**B**)

few examples of CST II and V in this dataset. Comparisons between these CSTs revealed a large number of differentially expressed KEGG orthologs (Fig. 4A–C, full table in Supplementary Table 2). These included expected differences like a 4.6- and 4.8-fold higher expression of D-lactate dehydrogenase (K03778) in CST I compared to CSTs III and IV, which catalyzes the production of D-lactate from pyruvate. The

Fig. 4 Functional differences in the transcriptional activity of CST and associated taxonomic drivers. Volcano plots displaying log$_2$ fold change differences in the expression of KEGG orthologs between communities assigned to different community state types (**A**, **C**, and **E**). Points in blue represent KEGG orthologs which were identified as differentially expressed and differentially abundant. Points in yellow represent KEGG orthologs that were found to be differentially expressed but not differentially abundant. In **B**, **D**, and **F**, the taxa responsible for the expression of these KEGG orthologs are displayed. For each fold change category, the relative contribution of each taxon is displayed, averaged across all of the KEGG orthologs in that category. Numbers above the bars indicate the number of differentially expressed KEGG orthologs in the category. *DE, differentially expressed

number of differentially expressed KEGG orthologs and the magnitude of those differences was highest between CST I and IV (Fig. 4B). The majority of the identified differentially expressed KEGG functions were also differentially abundant in the metagenomic data, signifying an underlying disparity in the functional capacity of

these communities. However, there were some KEGG orthologs which were equally abundant but differentially expressed (denoted by yellow diamonds in Fig. 4A–C). For example, the *ptsAB* phosphate transport system (K02036 and K02038), which had an equal abundance in CSTs I, III, and IV, exhibited an ~4-fold higher expression in CST I versus CST III and IV. Another example is alpha-L-rhamnosidase (K05989), which had equal abundance in CSTs I and IV, but 3-fold higher expression in CST IV.

We next characterized the taxa which contributed to each of the differentially expressed KEGG orthologs in order to identify the drivers of functional variation in the activity of CSTs (Fig. 4D–F). For each KEGG ortholog, we calculated the average contribution of each taxon to its expression in each CST. As expected, we found that KEGG functions that are enriched in CST I, compared to either CST III or CST IV, originate mostly from *L. crispatus* (Fig. 4D, E). This was not surprising given that communities assigned to CST I characteristically contain a large proportion of this species. In contrast, functions that are enriched in CST III compared to CST I did not primarily originate from *L. iners*, the species which defines CST III. Instead, the expression of these functions primarily originated from species like *G. vaginalis*, *Bifidobacterium breve*, and several *Prevotella* spp. (2- to 4-fold difference, Fig. 4D). Only KEGG orthologs which demonstrated higher fold changes in this comparison could be primarily attributed to *L. iners*. Expression of KEGG orthologs which were found to be enriched in CST IV compared either to CST I or CST III, were attributed a diverse array of species. *G. vaginalis* and *A. vaginae*, two of the taxa characteristic of communities assigned to CST IV, were not responsible for a majority of these differentially expressed KEGG orthologs; *Ca.* L. vaginae, *Sneathia* spp. and *Prevotella* spp. instead express these functions.

Differentially expressed KEGG orthologs were also grouped into KEGG modules. This higher-level of functional classification describes KEGG orthologs which act in concert to perform a function. Hierarchical clustering was used to group the KEGG modules based on their pattern of associations with CSTs I, III, and IV. This distinguished seven groups of KEGG modules (Fig. 5). One group included KEGG modules that demonstrated higher expression in CST I than in CST III or IV. This group includedβ-glucoside, mannitol, phosphonate, phosphate, and putrescine transport systems, as well as the biosynthesis of cysteine, lysine, carotene, and coenzyme F_{420}. Another group of KEGG modules was associated with higher expression in CSTs I and III and included: Zn/Mn, glutamine, glucitol transporters, as well as the arbutin-like PTS system. There was no group of KEGG modules which demonstrated exclusively higher expression in CST III, likely reflecting *L. iners* limited functional capacity resulting from its reduced genome size [43, 44]. Instead, there were two groupings of modules which demonstrated higher expression in CST III and IV. The association with CST III of these modules was mostly driven by organisms other than *L. iners*. Included within these two groups were: glutamate transport, methionine degradation, iron and zinc transport, and histidine biosynthesis. Finally, there was a number of modules associated only with CST IV. This group included KEGG modules responsible for branched-chain amino acid transport, maltose/glucose transport, cytolysin transport, chemotaxis, galactosamine transport, LPS transport, and sulfur degradation.
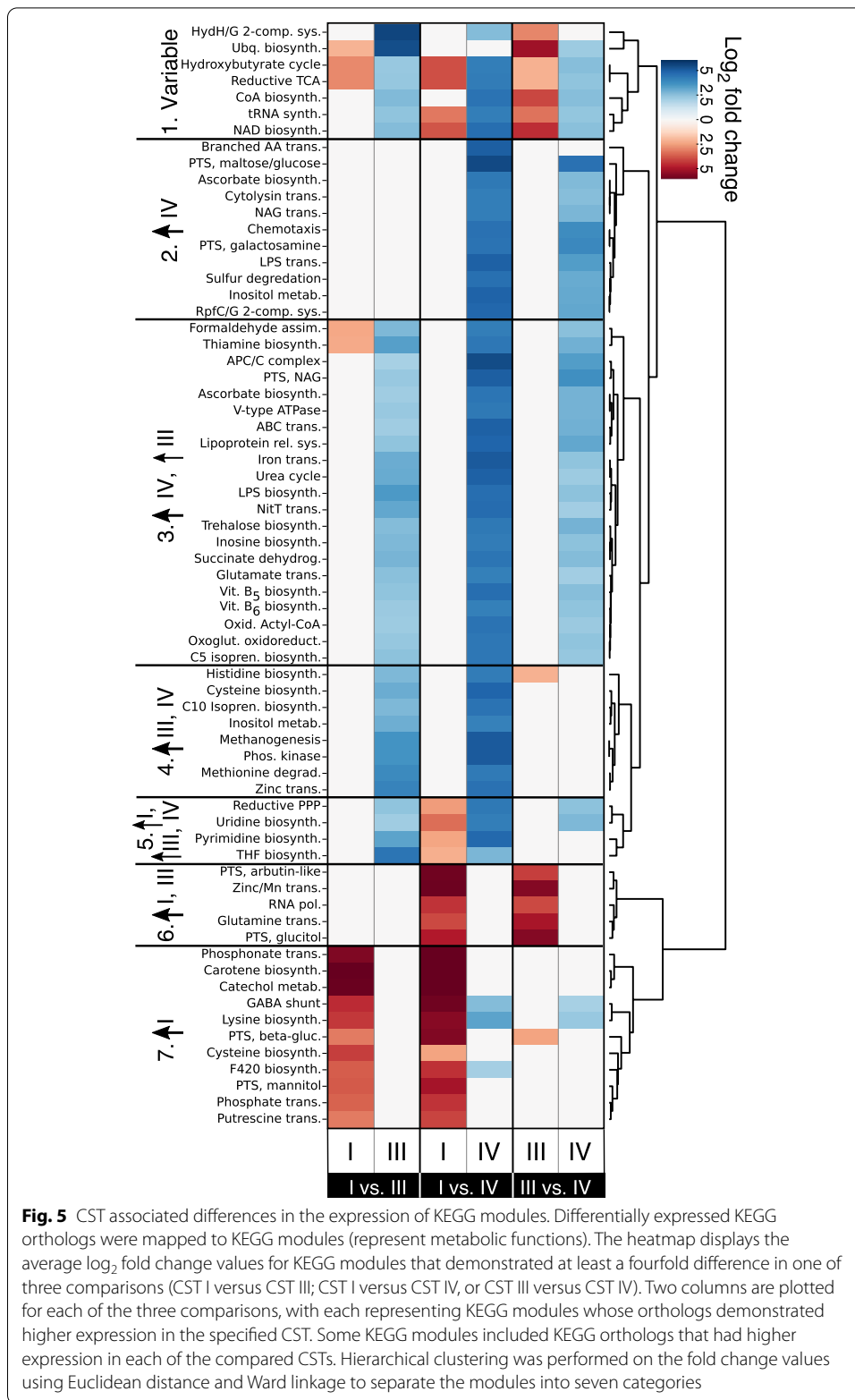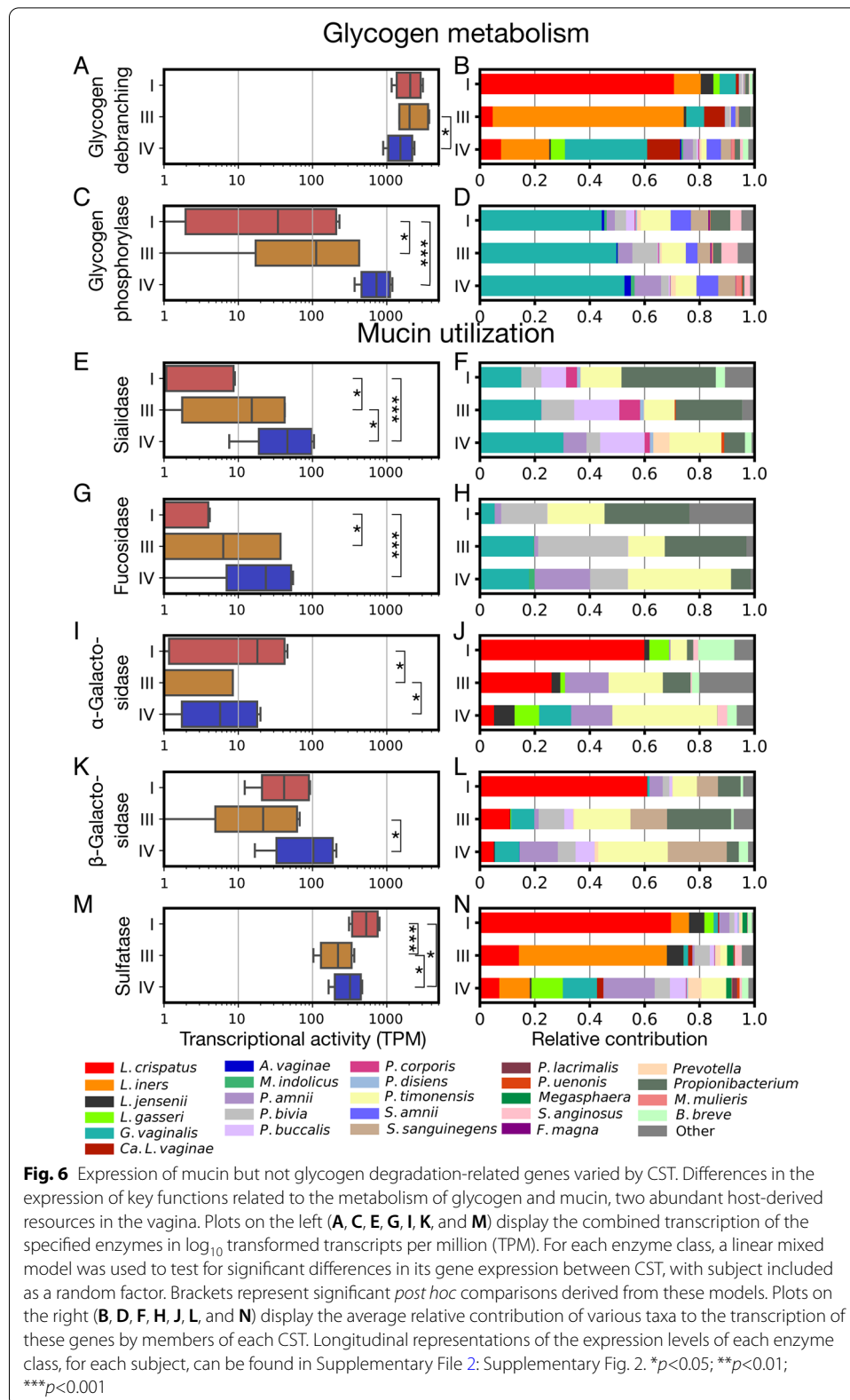
France *et al. Genome Biology*   (2022) 23:66

Page 9 of 26

Fig. 5 CST associated differences in the expression of KEGG modules. Differentially expressed KEGG orthologs were mapped to KEGG modules (represent metabolic functions). The heatmap displays the average log$_2$ fold change values for KEGG modules that demonstrated at least a fourfold difference in one of three comparisons (CST I versus CST III; CST I versus CST IV, or CST III versus CST IV). Two columns are plotted for each of the three comparisons, with each representing KEGG modules whose orthologs demonstrated higher expression in the specified CST. Some KEGG modules included KEGG orthologs that had higher expression in each of the compared CSTs. Hierarchical clustering was performed on the fold change values using Euclidean distance and Ward linkage to separate the modules into seven categories

Expression of glycogen and mucin utilization genes by the vaginal microbiota

Host-produced glycogen and mucin are both believed to be important sources of carbon and energy for bacteria in the vagina [45, 46]. Recent work has identified the genes responsible for the metabolism of these products by the microbiota including glycogen debranching enzymes in *Lactobacillus* spp. [47, 48] and sialidases in *G. vaginalis* [49]. We examined these and other enzymes involved in the degradation of glycogen and mucin to determine if their expression varied across communities assigned to different CSTs. VIRGO genes responsible for these activities were identified using the Carbohydrate Active enZymyes (CAZy) annotation schema [50]. Enzymes placed in the glycosyl-hydrolase (GH) family 13 and glycosyl-transferase (GT) family 35 were identified as likely having glycogen debranching and glycogen phosphorylase activities, respectively. The debranching enzymes act on α-glycoside linkages and can liberate trehalose, glucose, maltose, and maltotriose from glycogen. Expression of these enzymes was similar across the CSTs occurring at around 1250 transcripts per million (TPM), although it was slightly higher in CST III communities (linear mixed model; Fig. 6A). Glycogen debranching enzyme expression was attributed to many bacteria common to the vagina including *Lactobacillus* spp., *G. vaginalis*, *Ca*. L. vaginae, *Prevotella* spp., and *Sneathia* spp. and was consistent with the species composition of CSTs. For example, *L. crispatus* was responsible for most of the glycogen debranching enzyme expression in CST I (Fig. 6B). The expression of glycogen phosphorylase enzymes, which act to cleave individual glucose monomers from glycogen, did vary by CST with their expression being highest in CST IV, and lowest in CST I (linear mixed model, Fig. 6C). This was likely due to the absence of these genes in the genomes of *Lactobacillus* spp. which are dominant in CST I and III. The taxa responsible for the expression of glycogen phosphorylase enzymes was similar across CSTs (Fig. 6D).

Mucin produced by cervical cells is extensively modified with various carbohydrate residues that, when liberated, can be metabolized by the microbiota [51]. The removal of terminal residues, typically sialic acid or fucose, from these mucin-linked oligosaccharides is often the first step in their degradation [52]. The enzymes responsible for the removal of sialic acid and fucose are either sialidases (GH family 33) and fucosidases (GH family 29), respectively. The expression of both enzyme families was highest in CST IV communities (Fig. 6E, F) and lowest in CST I communities, indicating that mucin metabolism is likely lower when *Lactobacillus* spp. are dominant in the vaginal microbiota. The taxa responsible for the expression of sialidases in all CSTs included *G. vaginalis*, a number of *Prevotella* spp., along with a minor contribution from *B. breve* (Fig. 6F). In comparison, the expression of fucosidase originated primarily from *G. vaginalis*, *P. amnii*, *P. bivia*, *P. timonensis*, and *Propionibacterium* (Fig. 6H).

Galactose is another common component of mucin that can be released via the action of either α-galactosidases (GH family 36) or β-galactosidases (GH family 2). Expression of α-galactosidase was higher in CST I and IV communities than in CST III, communities, whereas expression of β-galactosidase was only higher in CST IV as compared to CST III communities (Fig. 6I, K). *L. iners* did not express enzymes in either class, explaining why CST III communities had the lowest expression of both enzyme families. Non-*Lactobacillus* spp. that expressed these enzymes included: *G. vaginalis*, *S. anginosus* (only α-galactosidase), *S. sanguinegens* (only β-galactosidase), and various *Prevotella*

Fig. 6 Expression of mucin but not glycogen degradation-related genes varied by CST. Differences in the expression of key functions related to the metabolism of glycogen and mucin, two abundant host-derived resources in the vagina. Plots on the left (**A**, **C**, **E**, **G**, **I**, **K**, and **M**) display the combined transcription of the specified enzymes in $\log_{10}$ transformed transcripts per million (TPM). For each enzyme class, a linear mixed model was used to test for significant differences in its gene expression between CST, with subject included as a random factor. Brackets represent significant *post hoc* comparisons derived from these models. Plots on the right (**B**, **D**, **F**, **H**, **J**, **L**, and **N**) display the average relative contribution of various taxa to the transcription of these genes by members of each CST. Longitudinal representations of the expression levels of each enzyme class, for each subject, can be found in Supplementary File 2: Supplementary Fig. 2. *$p<0.05$; **$p<0.01$; ***$p<0.001$

spp. Notably, *S. sanguinegens* was responsible for a large proportion of β-galactosidase expression. In sum, these data are consistent with mucin degradation being highest in CST IV communities, with limited degradation in CST I and III because these CSTs lack sialidase and fucosidase expression.

The final class of mucin utilization genes we examined were sulfatases, which act to cleave sulfate groups from mucin, which enables further removal of carbohydrate residues. These enzymes accounted for up to 0.2% of transcriptional activity in communities with their expression being highest in CST I and lowest in CST III (Fig. 6M). A wide range of taxa were responsible for their expression including *L. crispatus*, *L. iners*, and *G. vaginalis* (Fig. 6N).

Integration of metagenomic and metatranscriptomic data identify potential species interactions

In the above analyses, we focused on characterizations of the microbiota's total transcriptional output. Given the high depth of our sequence datasets and the tendency of the vaginal microbiota to exhibit low community evenness, we were also able to examine the transcriptional activity of individual species. The transcriptomes of *L. crispatus* (n=48), *L. iners* (n=74), and *G. vaginalis* (n=58) were found to have adequate coverage (>500,000 reads for the target species) in at least 30 metatranscriptomes. We asked whether the gene expression of these species was influenced by the composition of the community in which they resided using an integrative analysis of the metagenomic data, to establish community composition, and the metatranscriptomic data, to define the gene expression of the focal species. Sparse canonical correlation analysis [53] was used to identify axes of covariance between these two datasets, representing combinations of genes from the focal species whose expression was correlated with the relative abundances of combinations of species in the community. To overcome the variation in gene sequence and gene content across strains of the same species, we assessed expression at the level of vaginal orthologous protein groups (VOGs) cataloged in VIRGO [34]. In the following paragraphs, we describe the results of these analyses and point out aspects of the identified correlations that we find relevant to the ecology of the vaginal ecosystem. A complete list of the taxa and VOGs included in each correlation, as well as their weights, can be found in Supplementary Table 3.

For *L. crispatus*, we identified a set of 41 VOGs whose levels of gene expression were correlated with the relative abundance of *A. vaginae* and *G. vaginalis* (ρ=0.80, Supplementary File 2: Supplementary Fig. 3, panel A), indicating the abundance of these two species likely influences the transcriptional activity of *L. crispatus*. A principal component analysis (PCA) of the 41 VOGs and two taxa demonstrated that the combined relative abundances of *A. vaginae* and *G. vaginalis* varied along the first axis (Fig. 7A). Nineteen of the *L. crispatus* VOGs demonstrated a direct relationship, higher gene expression associated with increasing relative abundances of *A. vaginae* and *G. vaginalis*, and twenty-two VOGs demonstrated an inverse relationship, higher gene expression associated with decreasing relative abundances of *A. vaginae* and *G. vaginalis*. Included among the VOGs with a direct relationship were: two carbohydrate regulatory genes, *glpR* and *cggR*, iron and magnesium transport genes, the *ulaG* vitamin C utilization gene, the alternative sigma factor *sigL*, and the membrane-associated *asp23* alkaline shock

France *et al. Genome Biology*   (2022) 23:66

Page 13 of 26

response gene. VOGs demonstrating an inverse relationship included: the carboxylesterase *nlhH*, the lactose operon repressor *purR*, amino acid and phosphate transporters, the *dnaK* chaperone, and the glutamine synthetase gene *glnA*.

Our analysis of *L. iners* transcriptional activity revealed a set of 111 VOGs which were correlated with the relative abundances of ten taxa (ρ=0.84, Supplementary File 2: Supplementary Fig. 3, panel B), two with negative coefficients (*L. jensenii* and *F. magna*) and eight BV-associated taxa with positive coefficients (*P. amnii*, *Megasphaera*, *A. vaginae*, *P. timonensis*, *Prevotella spp.*, *G. vaginalis*, *P. buccalis*, and *Mageebacillus indolicus*). This result indicates that the gene expression of these VOGs is correlated with the ratio of positive to negative coefficient taxa and can be seen in a PCA of the combined dataset of correlated VOGs and taxa (Fig. 7B). Included among the 58 *L. iners* VOGs associated with a higher relative abundance of *L. jensenii* and *F. magna* was the carbamoyl-phosphate synthase gene (small and large chain components), galactose import and mutarotase genes, the cell shape controlling *rodZ* gene, the cell wall biosynthesis gene *murE*/*murF*, a potassium importer, the NH(3)-dependent NAD(+) synthetase gene, and several components of the DNA damage repair system (*mutL*, *mutS2*, *recJ*, and *recR*). On the other hand, the 53 *L. iners* VOGs demonstrating higher expression in the presence of the eight BV-associated organisms included: mannose/fructose/N-acetyl-galactosamine transport genes, the cholesterol-dependent cytolysin inerolysin, a gene encoding a putative mucin-binding protein, glucosamine kinase *gspK*, and the galactose-6-phosphate isomerase *lacAB* genes.

Lastly, we examined the transcriptional activity of *G. vaginalis* using the same approach. Our analysis indicated that a set of 185 *G. vaginalis* VOGs could be correlated with the relative abundances of nine bacterial taxa (ρ=0.80, Supplementary File 2: Supplemental Fig. 3, panel C). Of those nine taxa, five were identified as having positive coefficients (*L. crispatus*, *L. jensenii*, *L. gasseri*, *S. epidermidis*, and *Corynebacterium pseudogenitalium*) while four were identified with negative coefficients (*P. amnii*, *P. timonensis*, *M. indolicus*, and *Megasphaera*). A PCA of a combined dataset containing the correlated genes and taxa demonstrated that the samples varied in the ratio of the combined relative abundances of the positive and negative taxa along the first axis (Fig. 7C). Similar to what we observed with *L. iners,* among the VOGs associated with a higher relative abundance of BV-associated bacteria was the *G. vaginalis* cholesterol-dependent cytolysin vaginolysin. Other VOGs among the 128 associated with a higher relative abundance of these four BV-associated microbes included: a vitamin K epoxide reductase gene, several metal transporters genes, an endopeptidase gene, a serine protease gene, a maltose transport protein gene, and two putative gene encoding glycogen debranching enzymes with alpha-amylase domains. On the other side of the association,

(See figure on next page.)

Fig. 7 Community composition modulates species' gene expression. Principal component biplots of the VOGs and taxa were identified as correlated by sparse canonical correlation analysis (sCCA). The analysis was conducted on the transcriptional activity of the three most prevalent species: *L. crispatus* (**A**), *L. iners* (**B**), and *G. vaginalis* (**C**). The large circular points in each plot represent the samples. In **A**, these points are colored according to the combined $\log_{10}$ relative abundance the two negatively contributing taxa (*A. vaginae* and *G. vaginalis*). In **B** and **C**, these points are colored according to the ratio of the combined relative abundances of taxa that contribute either positively or negatively to the correlation. Factor loadings of the genes (positive, blue; negative, magenta) and taxa (positive, green; negative, gold) are included in the plot. For legibility, the labels for the taxa loadings, but not the VOG loadings are displayed

France *et al. Genome Biology*   (2022) 23:66

Page 14 of 26

Fig. 7 (See legend on previous page.)
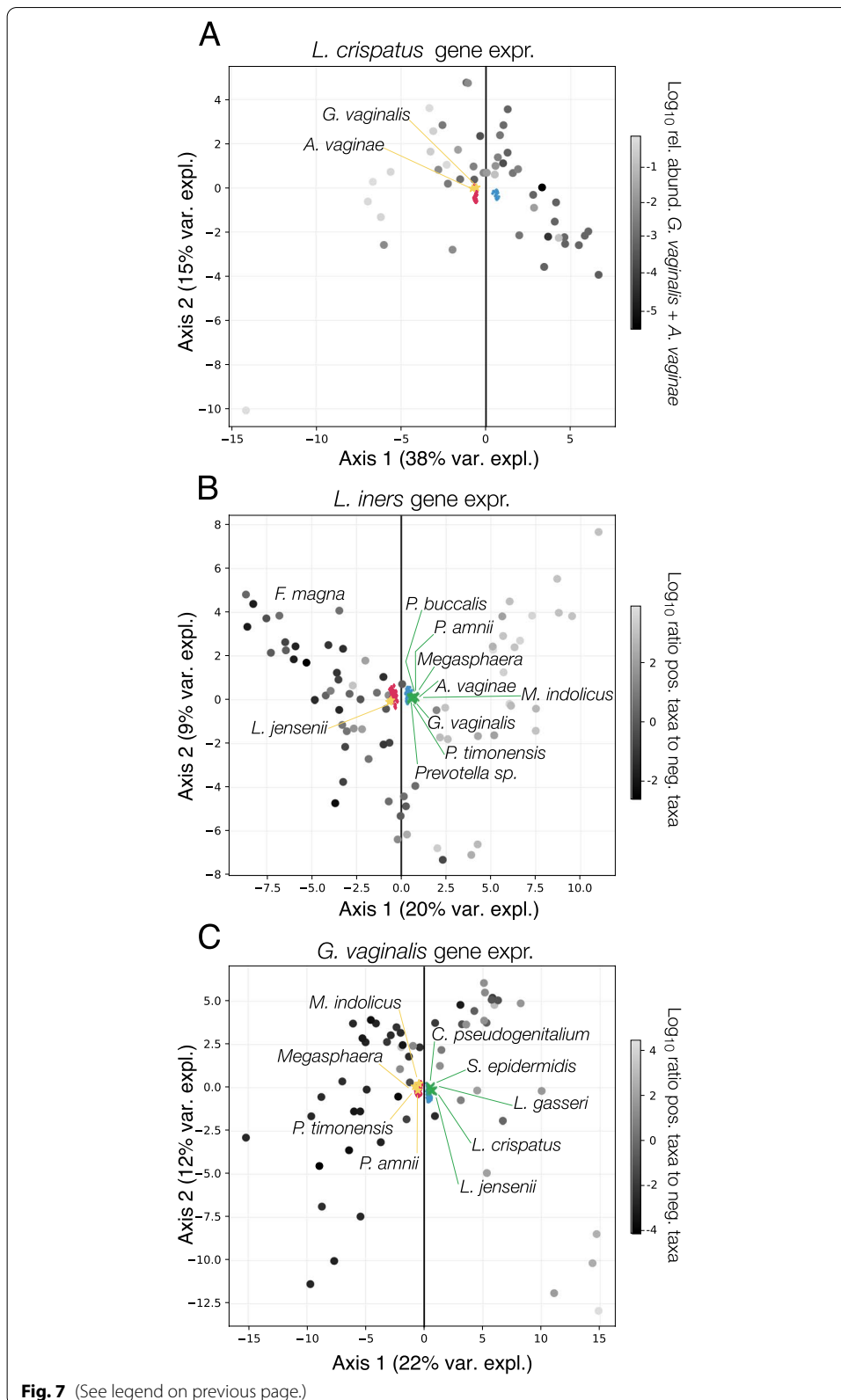
we found the expression of 57 VOGs to be correlated with higher relative abundances of *L. crispatus*, *L. jensenii*, *L. gasseri*, *S. epidermidis*, and *Corynebacterium pseudogenitalium*. Included among these were *hsp70*, a *groEL* like chaperonin, two genes bearing similarity to the staphylococcal gene *ebh*, two maltose transport genes, two putative glycogen debranching enzymes, and the alkyl-hydroperoxide reductase gene.

Discussion

Our integrative analysis of vaginal metatranscriptomes and metagenomes has shown that the relative abundance of a species does not necessarily reflect its transcriptional output. Taxa observed at lower relative abundances are often over-represented in the metatranscriptome while taxa observed at higher relative abundances are often under-represented. We have no reason to suspect these differences were due to bioinformatics procedures because the metagenomic and metatranscriptomics datasets were processed in the same manner. It could be that variation in the techniques used to extract the DNA and RNA contributed to the observed differences. However, we think it unlikely in this case because differences between the composition of the metagenome and metatranscriptome were found to be predictive of community dynamics. We instead posit that this discordance results from differential transcriptional activity among the species in the community. This interpretation has important implications for how we view these communities. If species which we estimate to be rare in the community are highly transcriptionally active, they might have a disproportionate influence on host tissues. For example, *S. amnii* and *S. sanguinigens* were almost always over-represented in the metatranscriptome, and often by several orders of magnitude. These two species are considered by many to be vaginal pathogens and have been shown to be associated with vaginal inflammation [12]. Some studies have further linked the abundance of these species to an increased risk of spontaneous preterm birth [36, 54, 55]. Associations between these species and adverse health outcomes may be strengthened by examining the transcriptional activity of the community, as the DNA-based compositional determinants are likely to underestimate the importance of *Sneathia* spp. when present in lower relative abundance.

We found that the transcriptional activity of a species could be used to predict changes in the relative abundance of taxa. Species which were under-represented in the metatranscriptome often decreased in relative abundance by our next timepoint and species which over-represented tended to increase in relative abundance. This result was somewhat surprising given that our timepoints were spaced approximately 2 weeks apart and that these communities can experience rapid change [31, 32]. RNA is less chemically stable than DNA and may therefore more accurately reflect the community at the time of sampling [56]. Compositional estimates based on DNA are also likely to experience a lag time while DNA from dead cells is degraded or washed out of the vagina [57]. Our results therefore suggest that shifts in community composition are effectuated earlier than our DNA-based estimations observe them. This has profound implications for our understanding of the factors that drive changes in the vaginal microbiota. For example, symptomatic bacterial vaginosis may be the manifestation of microbial processes put into motion days prior to its onset. Attempts to discern root causes of BV or to develop

novel BV diagnostics would benefit from a closer look at the microbiota's transcriptional activity before BV manifests.

We next examined functional differences in the transcriptional activity of communities assigned to different CSTs. As defined, communities assigned to different CSTs have fairly distinct taxonomic compositions, so we expected differences in their functional activities. However, functions that differ in expression between the CSTs represent metabolic pathways which could be leveraged to manipulate the composition of the vaginal microbiota, favoring *Lactobacillus*. For example, CST I and III communities express a transporter annotated as being capable of transporting glutamine, glutamate, and aspartate, while CST IV communities express a transporter annotated as only being capable of transporting glutamate. Glutamine is a peptidoglycan precursor and serves as the amine donor for both the conversion of fructose-6-phosphate into glucosamine-6-phosphate and the conversion of aspartate to asparagine in the oligopeptide crosslinks [58]. Based on this knowledge, we suggest that glutamine supplementation may favor *Lactobacillus* in the vaginal environment over the BV-associated organisms common to CST IV. This is supported by our observation that *L. crispatus* exhibits an increased expression of glutamine synthetase when co-resident with *G. vaginalis* and *A. vaginae*. We also find the expression of a metal transporter, predicted to transport zinc and manganese, to be higher in CST I and III than in CST IV. Previous studies have demonstrated that manganese is used by *Lactobacillus* as a defense against oxygen toxigenicity [59], and zinc supplementation has been shown to improve the growth of a probiotic *L. plantarum* strain [60].

Communities dominated by *L. crispatus* were also found to exhibit higher expression of phosphate and phosphonate transporters, consistent with a previous genome comparison that demonstrated *L. crispatus* to have more phosphate/phosphonate transporters than *L. iners* [44] and a previous observation of increased abundance of phosphate transport genes in the vaginal microbiota when vaginal pH is less than 4 [2]. Phosphate can buffer pH and intracellular accumulations of polyphosphate have been shown to be used by other *Lactobacillus* spp. as a mechanism of acid tolerance [61]. Although *L. crispatus* has not been shown to make polyphosphate, simply sequestering phosphate may allow *L. crispatus* to raise intracellular and lower extracellular pH. Similarly, we found *L. crispatus* dominant communities to exhibit higher expression of a putrescine/spermidine transporter, two biogenic amines capable of buffering pH. This is consistent with the results from metabolomic studies which have found *L. crispatus* dominant communities to have lower levels of putrescine and higher levels of spermidine [27, 62]. Manipulating the intracellular and extracellular concentrations of these pH buffering compounds could be a competitive strategy utilized by *L. crispatus* which allows it to inhibit the growth of competitors and facilitate its own metabolism.

As part of our functional analysis, we took a focused look at the expression of enzymes predicted to be involved in the degradation of glycogen and mucin, two host-produced potential sources of carbon and energy in the vaginal environment. Glycogen is primarily produced by vaginal epithelial cells [45] whereas mucin is primarily produced by cervical tissues and then secreted into the vagina as a constituent of mucus [46]. The ability of the vaginal microbiota to degrade glycogen has been debated for some time [63], but recent studies have demonstrated that many vaginal bacteria likely do have the ability to

degrade host-produced glycogen [47, 48]. We demonstrated that most common members of the vaginal microbiota express predicted glycogen degradation enzymes in vivo, although in vitro studies are needed to verify their activity. Expression of these enzymes did not appear to vary widely among communities assigned to different CSTs, indicating that glycogen degradation may occur at approximately the same rate in these women and that glycogen is an universal resource in the vaginal microenvironment. Glycogen supplementation alone, which some have suggested to treat BV [64], therefore, seems unlikely to favor the lactobacilli, and might actually favor the growth of BV-associated bacteria. This challenge might be overcome by complementing glycogen supplementation with a pH lowering agent in order to give the more acid-tolerant *Lactobacillus* an advantage in the metabolism of this common good.

Vaginal mucus has non-specific antibacterial properties and molecules such as lactoferrin, lysozyme, and immunoglobulins are often bound to it [65]. Cervicovaginal mucus plays critical functions in fertility [66] and protection against sexually transmitted infections [67], including HIV [68–70]. Mucin, the glycoprotein that comprises mucus, appears to be differentially degraded by the vaginal microbiota. Vaginal *Lactobacillus* spp. do not express sialidase or fucosidase enzymes meaning they are likely incapable of removing terminal sialic acid or fucose residues from mucin glycosylation chains. Taxa prevalent in CST IV, including *G. vaginalis* and *Prevotella* spp., do express these enzymes. *Gardnerella* is often featured in discussions of vaginal mucin degradation [49], but our results suggest that *Prevotella* spp., despite typically being at a lower relative abundance in CST IV communities in this cohort, are also likely responsible for a disproportionate amount of this activity. We conclude that the degree and pattern of mucin degradation likely vary with community composition. There have been some direct assessments of variation in vaginal mucin glycosylation chains both through time and between individuals [71]. However, these results are difficult to interpret given we know little about variation in mucin composition prior to any microbial tampering. Microbiota-driven differences in mucin composition could thus have important reproductive health implications considering the key roles mucins play in fertility and susceptibility to sexually transmitted infections, [68, 72].

In our final analysis, we examined the transcriptional activities of *L. crispatus*, *L. iners*, and *G. vaginalis*. We used an integrative analysis of metagenomic and metatranscriptomic data to ask whether there were sets of genes belonging to each of these species whose expression was correlated with the relative abundances of taxa in the community. The results of these analyses provided functional insight in the ecology of these species in the vaginal environment and hint at possible interactions between members of the microbiota. For *L. crispatus* we find a set of genes whose expression was correlated with the relative abundances of *A. vaginae* and *G. vaginalis*, two species that compete with *L. crispatus* in the vaginal environment. Included in the set of directly correlated *L. crispatus* genes is *asp23*, which encodes an alkaline shock protein. Higher expression of this gene may facilitate the growth of *L. crispatus* under higher pH conditions common to communities with a high relative abundance of *A. vaginae* and *G. vaginalis* [9]. Other correlated genes (e.g., carbohydrate metabolism regulators, amino acid and phosphate transporters, glutamine synthase) may signify metabolic shifts in the activities of *L. crispatus* when they co-reside with these bacteria. Additional in vitro studies on the

France *et al. Genome Biology*   (2022) 23:66

Page 18 of 26

shift in transcriptional activity of *L. crispatus* when co-resident with *A. vaginae* and *G. vaginalis* could inform strategies to modulate the microbiota towards an *L. crispatus* dominant state.

The results of our analysis of the *L. iners* and *G. vaginalis* transcriptomes demonstrated a number of commonalities which may reflect similarities in their ecology. *L. iners* and *G. vaginalis* both exhibited differences in their gene expression when co-resident with lactobacilli versus organisms associated with CST IV. *L. iners* is often found to co-occur with bacteria associated with CST IV communities, including *G. vaginalis* [73, 74]. For *L. iners,* the set of correlated lactobacilli only included *L. jensenii*, a species which often co-exists with *L. iners* [9]. For *G. vaginalis* the correlated lactobacilli included: *L. gasseri*, *L. jensenii*, and *L. crispatus.* Higher relative abundances of these species were associated with increased expression of alkyl-hydroperoxide reductase by G. *vaginalis*, a protein capable of reducing hydrogen peroxide. *L. crispatus*, *L. gasseri*, and *L. jensenii* produce H_2O_2 in vitro, and our transcriptomic data suggests *L. crispatus* expresses enzymes capable of generating H_2O_2 in vivo. Hydrogen peroxide is toxic to many of the anaerobes common to the vaginal tract, leading many to believe H_2O_2 to play a major role in the protection provided by the *Lactobacillus*, although this notion has been challenged [75, 76]. Arguments against its relevance include that it is not clear if the generation of H_2O_2 is possible at physiological oxygen concentrations in the vagina and that H_2O_2 is unlikely to diffuse far in the vagina far before it is quenched. Our observation is consistent with *Gardnerella* mounting a greater defense against peroxides when co-resident with these *Lactobacillus*, potentially minimizing the impact of *Lactobacillus*-produced H_2O_2 and providing another argument against H_2O_2 relevance as an antimicrobial. It is also possible that *Gardnerella* is instead responding to elevated environmental O_2 concentrations which might also favor the lactobacilli. This is an interesting finding, as it suggests that modulating the redox potential of the vaginal microenvironment could be used to favor *Lactobacillus* spp [77].

The second similarity between our analyses of the *L. iners* and *G. vaginalis* transcriptomes using sparse canonical correlation analysis involved their cholesterol-dependent cytolysins (inerolysin and vaginolysin [78–80]). Comparative genomic studies of these two species have indicated that both likely acquired their cytolysins via horizontal gene transfer since they are not present in other *Lactobacillus* species or other members of *Bifidobacteriaceae*. In fact, phylogenetic analyses have revealed that inerolysin is most similar to vaginolysin [44], although recent phenotypic comparisons have demonstrated some differences between them [81]. In both of our analyses, higher expression of the cytolysin was correlated with a greater relative abundance of several BV-associated anaerobes common to CST IV, although only *M. indolicus* was included in both species' correlations. This result indicates *L. iners* and *G. vaginalis* exhibit higher expression of their cytolysins when in the presence of these organisms, and lower expression when co-resident with other *Lactobacillus* species (*L. jensenii* for *L. iners*; *L. jensenii*, *L. gasseri*, and *L. crispatus* for *G. vaginalis*). Cholesterol dependent cytolysins act by forming pores in host cell membranes and have been shown to contribute to the pathogenicity of many bacteria [82], although the pathogenicity of the *L. iners* cytolysin is unclear given that the bacterium has not been definitively shown to be associated with adverse health conditions [73].

France *et al. Genome Biology*   (2022) 23:66

Page 19 of 26

Nevertheless, the pathogenic potential of both *L. iners* and *G. vaginalis* appears to be modulated by the overall composition of the community in which they reside. Communities which are dominated by *L. iners* may not be as destructive to host tissues as those that contain the *L. iners* co-resident with other BV-associated species.

Conclusion

A common limitation of many microbiome studies is that they rely on DNA-based assessments of community composition. These estimates, be they derived from 16S rRNA gene amplicon sequencing or shotguns metagenomics, represent genomic and functional *potential*. Not all genes on a genome are transcribed at the same time or at the same rate. Sequencing of the metatranscriptome describes the transcriptional activity of microbial communities in vivo and gives better insight into the function of these communities. However, metatranscriptomics does not provide a complete picture of the biochemical activity of microbial communities. Differences in mRNA stability and translation efficiency can introduce variation in the number of proteins produced per transcript and proteins can vary in their stability. If the transcript encodes an enzyme, its activity will also be governed by the concentration of substrates, products, and cofactors, as well as environmental conditions. Yet metatranscriptomics does bridge a large part of the gap between functional potential and realized biochemical activity. Epidemiological associations between the vaginal microbiome and health outcomes should be strengthened by examining the vaginal metatranscriptome. Our demonstration of the predictive power of the vaginal metatranscriptome could also be leveraged for the early diagnosis of vaginal disorders, like symptomatic BV. Further, results from our analysis of metatranscriptomic data could guide the development of innovative strategies to modulate the composition and activity of the vaginal microbiota to restore and maintain an optimal protective environment.

Methods

Cohort and sample description

The cohort described in this study included 39 North American reproductive-age women, between the ages of 19 and 45. Women included in the study self-identified as Black or African American ($n=24$), White or Caucasian ($n=10$), Hispanic or Latino ($n=4$), and Asian ($n=1$). They collected vaginal swabs daily, for 10 weeks, according to the protocol described in Ravel, Brotman [41]. For each subject, we selected up to five timepoints in this timeseries, spaced approximately 2 weeks apart for metagenomic ($n=194$) and metatranscriptomic ($n=180$) sequencing. Paired metagenomic and metatranscriptomic could not be generated for 14 of the samples resulting in six women with four samples, three women with three samples, and one woman with two samples. At the time of collection, vaginal swab specimens used in DNA extractions had been resuspended in 1 ml of Amies transport medium while RNA extractions had been stored in 2 ml 0.5x RNALater (Thermo Fisher; Waltham, MA) in Amies transport medium—both were preserved at − 80 °C. Participants also provided behavior and lifestyle information.

Shotgun metagenomics

Shotgun metagenomic data was generated for 194 samples (5 samples per subject). DNA was extracted from 200 μL of vaginal swab specimen resuspended in 1ml of Amies transport medium that had been preserved at − 80 °C. DNA extractions were performed using the MagAttract PowerMicrobiome DNA/RNA Kit (Qiagen) and bead-beating on a TissueLyser II according to the manufacturer's instructions and automated onto a Hamilton STAR robotic platform. Shotgun metagenomic sequence libraries were constructed from the DNA extracts using Illumina Nextera XT Flex kits according to manufacturer recommendations and then sequenced on an Illumina HiSeq 4000 platform (150 bp paired-end mode) at the Genomic Resource Center at the University of Maryland School of Medicine. Eight uniquely barcoded samples were included in each HiSeq 4000 lane yielding an average of 35 million read pairs for each sample.

Shotgun metatranscriptomics

Shotgun metatranscriptomics was generated for 180 samples (maximum of 5 samples per subject, minimum of 2) using a procedure described previously [34]. We were unable to obtain quality metatranscriptomic data for fifteen of the samples for which we obtained metagenomics data (information in Supplementary Table 3). For each sample, we extracted bulk RNA extracted from 1.5 ml of vaginal swab elute stored in 0.5x RNAlater solution (QIAGEN) that had been stored at − 80 °C. Prior to the extraction, 500 μl of ice-cold RNase-free phosphate-buffered saline (PBS) was added to the 1.5ml of stored swab specimen. To remove the RNAlater, the mixture was centrifuged at $8000 \times g$ for 10 min and then the resulting pellet was resuspended in 500 μl ice-cold RNase-free PBS with 10 μl of β-mercaptoethanol. RNA were extracted from the resulting suspension using the TRIzol (Sigma) procedure following the manufacturers recommendations. RNA was resuspended in 50 μl of DEPC-treated DNAase/RNAase-free water. Residual DNA was purged from 30 μl of RNA extract by treating once with Turbo DNase (Ambion, Cat. No. AM1907) for 30 min according to the manufacturer's protocol resulting in 30 μl of RNA suspension. DNA removal was confirmed via a PCR assay using 16S rRNA primer 27 F (5′-AGAGTTTGATCCTGGCTCAG-3′) and 534 R (5′-CATTACCGC GGCTGCTGG-3′). The quality of extracted RNA was verified using the Agilent 2100 Expert Bioanalyzer Nano chip. Ribosomal RNA removal was performed using 9 μl of the DNA-free RNA using the RiboZero Plus kit (Illumina) according to the manufacturer's protocol). Sequencing libraries were then prepared using SMARTer Stranded Total RNA-Seq - Pico Input Mammalian Kit (Takara Bio, Kyoto, Japan) according to manufacturer's protocols. The libraries were then sequenced on the NovaSeq platform (150bp, paired-end mode) with 40 libraries included per S2 chip. This provided approximately 85 million read pairs per library. Metatranscriptomic libraries were sequenced deeper than metagenomic libraries to account for the added sequence loss from the remaining rRNA.

Bioinformatic methods

To enable the downstream integration of metagenomic and metatranscriptomic data, the sequence data from both procedures were processed in the same way. First, reads originating from the human host were removed using BMtagger in combination with

the GRCh38 human reference genome [83]. Ribosomal RNA sequence reads were then removed using sortmeRNA [84]. Host-removed metatranscriptomic sequence datasets contained, on average, 31.7% residual rRNA sequences. Illumina adapters were excised and reads were trimmed using a 4 bp sliding window with an average quality score threshold of Q15 using Trimmomatic v0.3653 [85]. Reads trimmed to a length of less than 75bp were also removed. After these procedures, the median number of remaining reads per metagenomic dataset was 2.2×10^7 (6.6×10^5–9.9×10^7, Supplementary Table 4) and 1.4×10^7 per metatranscriptomic dataset (3.0×10^5–2.0×10^8, Supplementary Table 4). Reads were then mapped to the VIRGO non-redundant gene catalog for the vaginal microbiota using Bowtie [86] using VIRGO default settings [34]. Because each gene in the VIRGO database is annotated with a rich set of functional and taxonomic information, the results of these mappings give insight to the function and composition of the vaginal microbiome. To enable our analysis of carbohydrate catabolic enzymes, we further annotated the VIRGO database with the CAZy schema using dbCAN2 with default settings [87]. Annotations where two of the three methods used by dbCAN2 (Hmmer, DIAMOND, Hotpep) were accepted, as recommended. These new annotations have been made available via the VIRGO website (virgo.igs.umaryland.edu).

Analysis of the abundance and activity of taxa

We used the VIRGO mapping results to define the composition and structure of our metagenomic and metatranscriptomic data (Supplementary Table 1). The relative abundance of each taxon was determined by first correcting by dividing the number of mapped reads per gene by its length. These estimates of coverage were combined for each species and divided by sample total coverage in order to compute the relative abundances/activities. CSTs were assigned based on the taxonomic composition of the metagenome using Valencia [9]. For the 180 samples that had both metagenomic and metatranscriptomic data, we computed the relative expression of each taxon by dividing its relative contribution to the metatranscriptome by its relative abundance in the metagenome. We performed two statistical analyses of these taxonomic data as presented in Figs. 2 and 3. In the first, we modeled the $\log_{10}$ relative expression values in response to the $\log_{10}$ relative abundance of the taxa at that timepoint (linear mixed model; $F_{1,4261}=7167$, $p<0.001$), the taxa as a categorical predictor ($F_{24,4261}=198$, $p<0.001$), and their interaction (linear mixed model; $F_{24,4261}=11$, $p<0.001$) using a linear mixed model. The subject was also included as a categorial random factor to account for our longitudinal data. A second linear mixed model was used to determine whether there was predictive value in our estimates of relative expression. We modeled the $\log_{10}$ relative abundance of taxa as a function of the $\log_{10}$ relative expression at the previous timepoint ($F_{1,1847}=649$, $p<0.001$), with taxa ($F_{24,1847}=11.1$, $p<0.001$) and their combined interaction term ($F_{24,1847}=3.3$, $p<0.001$) included as well. Tukey adjusted post hoc comparisons of the taxa intercepts were conducted for both models. To demonstrate that these results were not affected by samples that yielded fewer reads we also conducted the analysis using datasets which contained only samples whose metagenomic and metatranscriptomic datasets contained at least 1 million microbial reads ($n=176$) and at least 5 million microbial reads ($n=135$). Results originating from the analysis of these filtered datasets did not differ from those obtained using the full dataset.

Differential expression analysis

Differential expression analyses were conducted between samples assigned to different community state types to identify functional variation in the vaginal microbiome. We conducted these analyses using the KEGG ortholog annotations [42] provided by the VIRGO database because these communities differ dramatically in taxonomic composition. For each KEGG ortholog, we summed the number of reads mapping to genes with the corresponding annotation. We then performed library size normalization using the edgeR package [88] and then modeled the differential expression of each KEGG ortholog using the dream function from the variancePartition package in R [89]. This package allowed us to fit mixed models which incorporated subject as random factor. Once a differentially expressed KEGG ortholog was identified, we used the VIRGO annotation schema to determine the taxa responsible for its expression. We calculated the relative contribution of each taxon to each differentially expressed KEGG ortholog, and then averaged these values for categories based on which CST they exhibited higher expression in, and the degree of their over-expression. The differential expression analysis was conducted on the metagenomic dataset to determine which KEGG orthologs also showed differential abundance in the microbial community. An additional differential expression analysis was conducted on just the enzymes we predicted using CAZy annotations to be involved in the degradation of glycogen or mucin. For each enzyme, we calculated its normalized expression in transcripts per million, which corrects for differences in gene length and library size. We performed individual linear mixed model analysis with CST as the predictor and subject included as a random factor to test for change in expression. Pairwise Tukey-adjusted post hoc comparisons were made between CST I, III, and IV.

Investigation of species interactions using sparse canonical correlation analysis

We used an integrative analysis of our metagenomic and metatranscriptomics datasets using sparse canonical correlation analysis or sCCA [53] to determine whether the composition of the microbiota influenced the transcriptional activities of *L. crispatus*, *L. iners*, and *G. vaginalis*. This method identifies axes of covariance between two datasets collected for the same set of samples. We used the individual transcriptomes of these three species as one dataset, and the composition of the microbiota, assessed via the metagenomics data, as the second dataset. First, we identified samples with at least 500,000 reads for the focal species in order to help mitigate differences in library size. Additionally, to help correct for strain differences in the populations of these species, we assessed the transcriptomes at the Vaginal Ortholog Gene (VOG) level [34] representing orthologous genes of the same species. We then removed VOGs that were expressed in less than half of the samples for the three species-specific datasets and converted the read counts per VOG into transcripts per million values. For the compositional dataset derived from the metagenomics data, we took the gene length-corrected relative abundance assessment for these samples. We first removed taxa that demonstrated a study-wide average relative abundance of less than 10^{-4} and transformed the relative abundance scores by taking the centered log ratio. To determine whether the pairs of datasets were correlated, we first calculated the RV coefficient of correlation [90]

between them using the ade4 R package [91]. sCCA was then performed for each of the three species using the PMA package in R [92]. Prior to the sCCA analysis, the pairs of datasets were centered and scaled. The permute function was used to select the optimal L1 penalties for both datasets independent. For *L. crispatus*, an L1 penalty of 0.1 was applied to the transcriptomics dataset and a 0.2 penalty was applied to the compositional dataset. These values were 0.2 and 0.4, respectively for *L. iners* and 0.2 and 0.35 for *G. vaginalis*. Final principal components analysis of the correlation VOGs and taxa were conducted using the ade4 package [91].

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-022-02635-9.

Additional file 1. Plots that displaying the longitudinal taxonomic composition of the metagenome and metatranscriptome for each subject.

Additional file 2: Supplementary Figure 1. Linear relationship of the $\log_{10}$ relative expression of each taxon in relation to its $\log_{10}$ relative abundance. Vertical gray lines represent the X intercept of the fit, representing the relative abundance at which each taxon transitions from over- to under-expressive. **Supplementary Figure 2.** Scatter plot of correlated canonical variables identified in the analysis of *L. crispatus* (a) *L. iners* (b) and *Gardnerella* (c) gene expression. **Supplemental Figure 3.** Scatter plot of correlated canonical variables identified in the analysis of *L. crispatus* (a) *L. iners* (b) and *Gardnerella* (c) gene expression.

Additional file 3: Supplementary Table 1. Tables containing the taxonomic composition of the metagenomic and metatranscriptomic data, as defined by VIRGO. Taxa relative abundances have been corrected for gene length.

Additional file 4: Supplementary Table 2. Tables containing results of differential expression analysis at the level of KEGG orthologs.

Additional file 5: Supplementary Table 3. Tables containing non-zero factor loadings for taxa and VOGs identified in sCCA analysis of *L. crispatus*, *L. iners*, *Gardnerella* gene expression.

Additional file 6: Supplementary Table 4. Post-processing sequence read counts and percent of those reads which mapped to the VIRGO database for the metagenomic and metatranscriptomic datasets.

Additional file 7. Review history.

Peer review information
Andrew Cosgrove was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Review history
The review history is available as Additional file 7.

Authors' contributions
Conceptualization, M. T. F. and J.R.; methodology, M. T. F., B. M., and P. G.; formal analysis, M. T. F.; investigation, L. F., L. R., H. Y., M. H., and S. N.; writing – original draft, M. T. F.; writing – review and editing, M. T. F., J. R., and L. J. F.; visualization, M. T. F.; supervision, J. R., M. H., and L. R.; funding acquisition, J. R. and L. J. F. The authors read and approved the final manuscript.

Availability of data and materials
Metagenomic and metatranscriptomic datasets generated as part of this study have been deposited in the NCBI Short Read Archive under the Bioproject PRJNA797778 [93]. Python and R code used in the described analyses and the generation of figures are accessible under a GNU General Public License (v2.0) from either Github (https://github.com/ravel-lab/MTMG_longitudinal) or Zenodo (https://zenodo.org/record/5932449) [94]. Additional information and requests for additional data and code should be directed to the Lead Contact, Jacques Ravel (jravel@som.umaryland.edu)

Declarations

Ethics approval and consent to participate
All human subject research studies were performed under approved protocols by the Institutional Review Board of the University of Maryland Baltimore under numbers HP-00040935 and HP-00041351. Samples were collected after

France *et al. Genome Biology*   (2022) 23:66

Page 24 of 26

obtaining written informed consent from all the participants and all experimental methods comply with the Helsinki Declaration.

Competing interests

Author details

[1]Institute for Genome Sciences and Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, MD, USA. [2]Department of Biological Sciences and Institute for Bioinformatics and Evolutionary Studies, University of Idaho, Moscow, ID, USA.

References

1. Cho I, Blaser MJ. The human microbiome: at the interface of health and disease. Nat Rev Genet. 2012;13(4):260–70.
2. Human Microbiome Project C. Structure, function and diversity of the healthy human microbiome. Nature. 2012;486(7402):207–14.
3. Pflughoeft KJ, Versalovic J. Human microbiome in health and disease. Annual Review of Pathology: Mechanisms of Disease: Annual Reviews; 2012. p. 99–122.
4. Miller EA, Beasley DE, Dunn RR, Archie EA. Lactobacilli dominance and vaginal pH: why is the human vaginal microbiome unique? Front Microbiol. 2016;7:1936.
5. Ma B, Forney LJ, Ravel J. Vaginal microbiome: rethinking health and disease. Annual review of microbiology; 2012. p. 371–89.
6. Yildirim S, Yeoman CJ, Janga SC, Thomas SM, Ho MF, Leigh SR, et al. Primate vaginal microbiomes exhibit species specificity without universal lactobacillus dominance. ISME J. 2014;44:2431.
7. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SSK, Mcculle SL, et al. Vaginal microbiome of reproductive-age women. In: Proceedings of the National Academy of Sciences; 2011. p. 4680–7.
8. Redondo-lopez V, Cook RL, Sobel JD. Emerging role of lactobacilli in the control and maintenance of the vaginal bacterial microflora. Rev Infect Dis. 1990;12:856–72.
9. France MT, Ma B, Gajer P, Brown S, Humphrys MS, Holm JB, et al. VALENCIA: a nearest centroid classification method for vaginal microbial communities based on composition. Microbiome. 2020;8(1):166.
10. Antonio MA, Hawes SE, Hillier SL. The identification of vaginal lactobacillus species and the demographic and microbiologic characteristics of women colonized by these species. J Infect Dis. 1999;180(6):1950–6.
11. Zhou X, Brown CJ, Abdo Z, Davis CC, Hansmann MA, Joyce P, et al. Differences in the composition of vaginal microbial communities found in healthy Caucasian and black women. ISME J. 2007;1(2):121–33.
12. Anahtar MN, Byrne EH, Doherty KE, Bowman BA, Yamamoto HS, Soumillon M, et al. Cervicovaginal bacteria are a major modulator of host inflammatory responses in the female genital tract. Immunity. 2015;42(5):965–76.
13. Tamarelle J, Barbeyrac BD, Hen IL, Thiébaut A, Bébéar C, Ravel J, et al. Vaginal microbiota composition and association with prevalent chlamydia trachomatis infection : a cross- sectional study of young women attending a STI clinic in France. Sex Transm Infect. 2018;94(8):616–8.
14. Gosmann C, Anahtar MN, Handley SA, Huttenhower C, Farcasanu M, Abu-Ali G, et al. Lactobacillus-deficient cervicovaginal bacterial communities are associated with increased HIV acquisition in young south African women. Immunity. 2017;46(1):29–37.
15. Brotman RM, Bradford LL, Conrad M, Gajer P, Ault K, Peralta L, et al. Association between Trichomonas vaginalis and vaginal bacterial community composition among reproductive-age women. Sex Transm Dis. 2012;39(10):807–12.
16. Elovitz MA, Gajer P, Riis V, Brown AG, Humphrys MS, Holm JB, et al. Cervicovaginal microbiota and local immune response modulate the risk of spontaneous preterm delivery. Nat Commun. 2019;10(1):1305.
17. Brown SE, Schwartz J, Robinson C, O'Hanlon ED, Bradford LL, Xin H, et al. The vaginal microbiota and behavioral factors associated with genital Candida albicans detection in reproductive-age women. Sex Transm Dis. 2019;46(11):753–8.
18. Richard DX, Brown G, Julian DXX, Lee S, Ann DXX, Denise DX, et al. Establishment of vaginal microbiota composition in early pregnancy and its association with subsequent preterm prelabor rupture of the fetal membranes. Transl Res. 2019; in press.
19. McClelland RS, Lingappa JR, Srinivasan S, Kinuthia J, John-Stewart GC, Jaoko W, et al. Evaluation of the association between the concentrations of key vaginal bacteria and the increased risk of HIV acquisition in African women from five cohorts: a nested case-control study. Lancet Infect Dis. 2018;18(5):554–64.
20. McKinnon LR, Achilles SL, Bradshaw CS, Burgener A, Crucitti T, Fredricks DN, et al. The evolving facets of bacterial Vaginosis: implications for HIV transmission. AIDS Res Hum Retrovir. 2019;35(3):219–28.
21. Boskey ER, Cone RA, Whaley KJ, Moench TR. Origins of vaginal acidity: high D/L lactate ratio is consistent with bacteria being the primary source. Human Reprod. 2001;16(9):1809–13.
22. Hanlon DEO, Moench TR, Cone RA. Vaginal pH and Microbicidal lactic acid when lactobacilli dominate the microbiota. PLoS One. 2013;8(11):e80074.
23. Smith SB, Ravel J. The vaginal microbiota, host defence and reproductive physiology. J Physiol. 2017;595(2):451–63.
24. Witkin S. The vaginal microbiome, vaginal anti-microbial defence mechanisms and the clinical challenge of reducing infection-related preterm birth. BJOG Int J Obstet Gynaecol. 2015;122(2):213–8.
25. Tachedjian G, Aldunate M, Bradshaw CS, Cone RA. The role of lactic acid production by probiotic lactobacillus species in vaginal health. Res Microbiol. 2017;168(9-10):782–92.

26. Brotman RM, Klebanoff MA, Nansel TR, Yu KF, Andrews WW, Zhang J, et al. Bacterial vaginosis assessed by gram stain and diminished colonization resistance to incident gonococcal, chlamydial, and trichomonal genital infection. J Infect Dis. 2010;202(12):1907–15.
27. Nelson TM, Borgogna JL, Brotman RM, Ravel J, Walk ST, Yeoman CJ. Vaginal biogenic amines: biomarkers of bacterial vaginosis or precursors to vaginal dysbiosis? Front Physiol. 2015;6:253.
28. Lefevre JC, Jean M, Averous S, Viraben R, Blanc C, Bauriaud R, et al. Etiology of bacterial vaginosis (non-specific vaginitis). J Gynecol Obstet Biol Reprod (Paris). 1985;14(6):703–8.
29. Allsworth JE, Peipert JF. Prevalence of bacterial vaginosis: 2001-2004 National Health and nutrition examination survey. Obstet Gynecol. 2007;109(1):114–20.
30. Callahan BJ, DiGiulio DB, Aliaga Goltsman DS, Sun CL, Costello EK, Jeganathan P, et al. Replication and refinement of a vaginal microbial signature of preterm birth in two racially distinct cohorts of US women; 2017.
31. Gajer P, Brotman RM, Bai G, Sakamoto J, Schutte UME, Zhong X, et al. Temporal dynamics of the human vaginal micro-biota. Sci Transl Med. 2012;4(132):132ra52.
32. Srinivasan S, Liu C, Mitchell CM, Fiedler TL, Thomas KK, Agnew KJ, et al. Temporal variability of human vaginal bacteria and relationship with bacterial vaginosis. PLoS One. 2010;5(4):e10197.
33. Feehily C, Crosby D, Walsh CJ, Lawton EM, Higgins S, McAuliffe FM, et al. Shotgun sequencing of the vaginal microbiome reveals both a species and functional potential signature of preterm birth. NPJ Biofilms Microb. 2020;6(1):50.
34. Ma B, France MT, Crabtree J, Holm JB, Humphrys MS, Brotman RM, et al. A comprehensive non-redundant gene catalog reveals extensive within-community intraspecies diversity in the human vagina. Nat Commun. 2020;11(1):940.
35. Ferretti P, Pasolli E, Tett A, Asnicar F, Gorfer V, Fedi S, et al. Mother-to-infant microbial transmission from different body sites shapes the developing infant gut microbiome. Cell Host Microbe. 2018;24(1):133–45 e5.
36. Fettweis JM, Serrano MG, Brooks JP, Edwards DJ, Girerd PH, Parikh HI, et al. The vaginal microbiome and preterm birth. Nat Med. 2019:1012–21.
37. Serrano MG, Parikh HI, Brooks JP, Edwards DJ, Arodz TJ, Edupuganti L, et al. Racioethnic diversity in the dynamics of the vaginal microbiome during pregnancy. Nat Med. 2019;25(6):1001–11.
38. Goltsman DSA, Sun CL, Proctor DM, DiGiulio DB, Robaczewska A, Thomas BC, et al. Metagenomic analysis with strain-level resolution reveals fine-scale variation in the human pregnancy microbiome. Genome Res. 2018;28(10):1467–80.
39. Macklaim JM, Fernandes AD, Di Bella JM, Hammond J-A, Reid G, Gloor GB. Comparative meta-RNA-seq of the vaginal microbiota and differential expression by lactobacillus iners in health and dysbiosis. Microbiome. 2013;1:12.
40. Deng Z-L, Gottschick C, Bhuju S, Masur C, Abels C, Wagner-Dobler I. Metatranscriptome analysis of the vaginal microbi-ota reveals potential mechanisms for recurrence and protection against metronidazole in bacterial vaginosis. mSphere. 2018;3(3):e0262-18.
41. Ravel J, Brotman RM, Gajer P, Ma B, Nandy M, Fadrosh DW, et al. Daily temporal dynamics of vaginal microbiota before , during and after episodes of bacterial vaginosis; 2013. p. 1–6.
42. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and Metagenome sequences. J Mol Biol. 2016;428(4):726–31.
43. Macklaim JM, Gloor GB, Anukam KC, Cribby S, Reid G. At the crossroads of vaginal health and disease, the genome sequence of lactobacillus iners AB-1. In: Proceedings of the National Academy of Sciences of the United States of America; 2011. p. 4688–95.
44. France MT, Mendes-Soares H, Forney LJ. Genomic comparisons of lactobacillus crispatus and lactobacillus iners reveal potential ecological drivers of community composition in the vagina. Appl Environ Microbiol. 2016;82(24):7063–73.
45. Ayre WB. The glycogen-estrogen relationship in the vaginal tract. J Clin Endocrinol Metab. 1951;11:103–10.
46. Elstein M. Cervical mucus: its physiological role and clinical significance. Adv Exp Med Biol. 1982;144:301–18.
47. van der Veer C, Hertzberger RY, Bruisten SM, Tytgat HLP, Swanenburg J, de Kat Angelino-Bart A, et al. Comparative genomics of human lactobacillus crispatus isolates reveals genes for glycosylation and glycogen degradation: implica-tions for in vivo dominance of the vaginal microbiota. Microbiome. 2019;7(1):49.
48. Nunn KL, Clair GC, Adkins JN, Engbrecht K, Fillmore T, Forney LJ. Amylases human vagina. mSphere. 2020;5(6):e00943–20.
49. Robinson LS, Schwebke J, Lewis WG, Lewis AL. Identification and characterization of NanH2 and NanH3, enzymes responsible for sialidase activity in the vaginal bacterium Gardnerella vaginalis. J Biol Chem. 2019;294(14):5230–45.
50. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014;42(Database issue):D490–5.
51. Carlstedt I, Lindgren H, Sheehan JK, Ulmsten U, Wingerup L. Isolation and characterization of human cervical-mucus glycoproteins. Biochem J. 1983;211:13–22.
52. Tailford LE, Crost EH, Kavanaugh D, Juge N. Mucin glycan foraging in the human gut microbiome. Front Genet. 2015;6:81.
53. Hardoon DR, Shawe-Taylor J. Sparse canonical correlation analysis. Mach Learn. 2010;83(3):331–53.
54. Nelson DB, Hanlon A, Nachamkin I, Haggerty C, Mastrogiannis DS, Liu C, et al. Early pregnancy changes in bacterial vaginosis-associated bacteria and preterm delivery. Paediatr Perinat Epidemiol. 2014;28(2):88–96.
55. Brown RG, Marchesi JR, Lee YS, Smith A, Lehne B, Kindinger LM, et al. Vaginal dysbiosis increases risk of preterm fetal membrane rupture, neonatal sepsis and is exacerbated by erythromycin. BMC Med. 2018;16(1):9.
56. Baudrimont A, Voegeli S, Viloria EC, Stritt F, Lenon M, Wada T, et al. Multiplexed gene control reveals rapid mRNA turno-ver. Sci Adv. 2017;3:e1700006.
57. Baloglu E, Senyigit ZA, Karavana SY, Bernkop-Schnurch A. Strategies to prolong the intravaginal residence time of drug delivery systems. J Pharm Pharmaceut Sci. 2009;12(3):312–36.
58. Chapot-Chartier MP, Kulakauskas S. Cell wall structure and function in lactic acid bacteria. Microb Cell Factories. 2014;13(Suppl 1):S9.
59. Archibald FS, Fridovich I. Manganese and defenses against oxygen toxicity in lactobacillus plantarum. J Bacteriol. 1981;145(1):442–51.
60. Mudroňová D, Gancarčíková S, Nemcová R. Influence of zinc Sulphate on the probiotic properties of lactobacillus plan-tarum CCM 7102. Folia Veterinaria. 2019;63(2):45–54.
61. Alcantara C, Blasco A, Zuniga M, Monedero V. Accumulation of polyphosphate in lactobacillus spp. and its involvement in stress resistance. Appl Environ Microbiol. 2014;80(5):1650–9.

France *et al. Genome Biology*   (2022) 23:66

Page 26 of 26

62. Srinivasan S, Morgan MT, Fiedler TL, Djukovic D, Hoffman NG, Raftery D, et al. Metabolic signatures of bacterial Vaginosis. mBio. 2015;6(2):e00204-15.
63. Nunn KL, Forney LJ. Unraveling the dynamics of the human vaginal microbiome. Yale J Biol Med. 2016;89:331–7.
64. Mirmonsef P, Hotton AL, Gilbert D, Burgad D, Landay A, Weber KM, et al. Free glycogen in vaginal fluids is associated with lactobacillus colonization and low vaginal pH. PLoS One. 2014;9(7):e102467.
65. Carlstedt I, Sheehan JK. Structure and macromolecular properties of cervical mucus glycoproteins. Symp Soc Exp Biol. 1989;43:289–316.
66. Nakano FY, Leão RBF, Esteves SC. Insights into the role of cervical musus and vaginal pH in unexplained infertility. MedicalExpress. 2015;2(2):M150207.
67. Wiggins R, Hicks SJ, Soothill PW, Millar MR, Corfield AP, Crowfoot D, et al. Mucinases and sialidases : their role in the pathogenesis of sexually transmitted infections in the female genital tract; 2001. p. 402–8.
68. Nunn KL, Wang YY, Harit D, Humphrys MS, Ma B, Cone R, et al. Enhanced trapping of HIV-1 by human Cervicovaginal mucus is associated with lactobacillus crispatus-dominant microbiota. mBio. 2015;6(5):e01084–15.
69. Hoang T, Toler E, DeLong K, Mafunda NA, Bloom SM, Zierden HC, et al. The cervicovaginal mucus barrier to HIV-1 is diminished in bacterial vaginosis. PLoS Pathog. 2020;16(1):e1008236.
70. Schroeder HA, Nunn KL, Schaefer A, Henry CE, Lam F, Pauly MH, et al. Herpes simplex virus-binding IgG traps HSV in human cervicovaginal mucus across the menstrual cycle and diverse vaginal microbial composition. Mucosal Immunol. 2018;11(5):1477–86.
71. Moncla BJ, Chappell CA, Debo BM, Meyn LA. The effects of hormones and vaginal microflora on the Glycome of the female genital tract: cervical-vaginal fluid. PLoS One. 2016;11(7):e0158687.
72. Mall AS, Habte H, Mthembu Y, Peacocke J, de Beer C. Mucus and Mucins: do they have a role in the inhibition of the human immunodeficiency virus? Virol J. 2017;14(1):192.
73. Petrova MI, Reid G, Vaneechoutte M, Lebeer S. Lactobacillus iners: friend or foe? Trends Microbiol. 2017;25(3):182–91.
74. Zozaya-hinchliffe M, Lillis R, Martin DH, Ferris MJ. Quantitative PCR assessments of bacterial species in women with and without bacterial vaginosis. J Clin Microbiol. 2010;48(5):1812–9.
75. O'Hanlon DE, Moench TR, Cone RA. In vaginal fluid, bacteria associated with bacterial vaginosis can be suppressed with lactic acid but not hydrogen peroxide. BMC Infect Dis. 2011;11(200):200.
76. Tachedjian G, O'Hanlon DE, Ravel J. The implausible "in vivo" role of hydrogen peroxide as an antimicrobial factor produced by vaginal microbiota. Microbiome. 2018;6(1):29.
77. Holmes KK, Chen KCS, Lipinski CM, Eschenbach DA. Vaginal redox potential in bacterial vaginosis (nonspecific vaginitis). J Infect Dis. 1985;152:379–82.
78. Rampersaud R, Planet PJ, Randis TM, Kulkarni R, Aguilar JL, Lehrer RI, et al. Inerolysin, a cholesterol-dependent cytolysin produced by lactobacillus iners. J Bacteriol. 2011;193(5):1034–41.
79. Cauci S, Monte R, Ropele M, Missero C, Not T, Quadrifoglio F, et al. Pore-forming and haemolytic properties of the Gardnerella vaginalis cytolysin. Mol Microbiol. 1993;9(6):1143–55.
80. Gelber SE, Aguilar JL, Lewis KL, Ratner AJ. Functional and phylogenetic characterization of Vaginolysin, the human-specific cytolysin from Gardnerella vaginalis. J Bacteriol. 2008;190(11):3896–903.
81. Ragaliauskas T, Pleckaityte M, Jankunec M, Labanauskas L, Baranauskiene L, Valincius G. Inerolysin and vaginolysin, the cytolysins implicated in vaginal dysbiosis, differently impair molecular integrity of phospholipid membranes. Sci Rep. 2019;9(1):10606.
82. Dowd KJ, Farrand AJ, Tweten RK. The cholesterol-dependent cytolysin signature motif: a critical element in the allosteric pathway that couples membrane binding to pore assembly. PLoS Pathog. 2012;8(7):e1002787.
83. Rotmistrovsky K, Agarwala R. BMTagger: best match tagger for removing human reads from metagenomic datasetsNCBI/NLM, NIH; 2011.
84. Kopylova E, Noé L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics. 2012;28(24):3211–7.
85. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence dataBioinformatics (Oxford, England): Oxford University Press; 2014. p. 2114–20.
86. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25.
87. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 2018;46(W1):W95–W101.
88. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139–40.
89. Hoffman GE, Schadt EE. variancePartition: interpreting drivers of variation in complex gene expression studies. BMC Bioinformatics. 2016;17(1):483.
90. Moonseong H, Ruben GK. A permutation test of association between configurations by means of the rv coefficient. Commun Stat- Simul Comput. 2010;27(3):843–56.
91. Dray S, Dufour A-B. The ade4 package: implementing the duality diagram for ecologists. J Stat Softw. 2007;22(4):1–20.
92. Witten DM, Tibshirani R, Hastie T. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. Biostatistics. 2009;10(3):515–34.
93. France M, Fu L, Rutt L, Yang H, Humphrys M, Narina S, et al. Insight into the ecology of vaginal bacteria through integrative analyses of metagenomic and metatranscriptomic data. SRA. 2022;PRJNA797778. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA797778.
94. France M, Fu L, Rutt L, Yang H, Humphrys M, Narina S, et al. Insight into the ecology of vaginal bacteria through integrative analyses of metagenomic and metatranscriptomic data. Zenodo. 2022. https://zenodo.org/record/5932449#.YhT1rfXML0o.

Publisher's Note