Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Host Genetic Determinants of the Vaginal Microbiome in Kenyan Women

Supriya D Mehta (Supriyad@uic.edu)
University of Illinois at Chicago https://orcid.org/0000-0002-7926-2489
Drew R Nannini
Northwestern University
Fredrick Otieno
Nyanza Reproductive Health Society
Stefan J. Green
University of Illinois at Chicago
Walter Agingu
Nyanza Reproductive Health Society
Alan Landay
Rush University
Yinan Zheng
Northwestern University
Lifang Hou
Northwestern University

Research

Keywords: Vaginal microbiome, Vaginal microbiota, Bacterial vaginosis, L. crispatus, L. iners, G. vaginalis, Shannon diversity index, community state type, genome wide association study, pathway analysis, toll-like receptors, innate immune response, Kenya

Posted Date: March 13th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-17153/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Abstract

Background Women with vaginal microbial community state types (CST) with high diversity and a paucity of Lactobacillus crispatus have increased risk of HIV acquisition. Identifying host genetic factors associated with the vaginal microbial composition may aid in elucidating the biological mechanisms regulating these microbial traits and inter-individual variations in associated diseases.

Methods We conducted genome-wide associations studies (GWASs) on vaginal microbiome traits on 171 Kenyan women. Study participants were genotyped using the Infinium Global Screening Array and 16S rRNA gene amplicon sequencing was performed to characterize the vaginal microbiome. Linear and logistic regression were performed, adjusting for age and principal components of genetic ancestry, to evaluate the association between L. crispatus, L. iners, G. vaginalis, Shannon diversity index, and CST with host genetic single nucleotide polymorphisms (SNPs). Pathway enrichment analyses were performed to identify biological processes putatively associated with the vaginal microbiome traits.

Results At baseline, the median age of study participants was 22 years (IQR: 22 – 25) with 22% having Bacterial vaginosis (Nugent score 7-10). L. crispatus and L. iners were present in 24% and 83% of samples with a mean relative abundance of 31% and 45%, respectively. The most significant SNPs associated were: rs73330467 located between LOC101927488-GRAMD2B (P =4.79x10 -6) for L. crispatus ; rs527430 in the FOXD2-TRABD2B (P =6.98x10 -7) region for L. iners ; rs1229660 in the SNX10-LOC441204 (P =4.65x10 -6) region for G. vaginalis; rs972741 in the ZKSCAN2-HS3ST4 (P =8.52x10 -7) region for Shannon diversity index; and rs2302902 in ELK3 (P =3.09x10 -6) for CST. During pathway enrichment analysis, Toll-like receptors, cytokine production, and other components of innate immune response were associated with L. crispatus, L. iners, and CST. Multiple genomic loci were replicated, including IL-8 (Shannon, CST), TIRAP (L. iners , Shannon), TLR2 (Shannon, CST), MBL2 (L. iners , G. vaginalis , CST), and MYD88 (L. iners, Shannon).

Conclusions We identified numerous genetic loci on several pathways related to host immunity and infection that were associated with vaginal microbiome traits, providing insight into potential host genetic influences on vaginal microbiome composition. This new information should guide larger longitudinal studies, with genetic and functional comparison across microbiome sites within individuals, and across populations.

Introduction

Bacterial vaginosis (BV) is a condition of clinical and public health significance. BV is prevalent in 20–50% of women in sub-Saharan Africa [1]. In a meta-analysis, Atashili et al. estimated the relative risk of HIV acquisition to be 1.6 times higher for women with BV, equating to a population attributable fraction of 15% due to the high population prevalence [2]. BV is also associated with an increased likelihood of sexually transmitted infections (STI) and adverse pregnancy outcomes [3, 4]. BV represents a polymicrobial shift in the vaginal microbiome, often from a Lactobacillus dominant community to one that is diverse, with multiple species of anaerobic bacteria [5]. In particular, L. crispatus enrichment has been shown to be protective against BV, HIV, and STIs [6].

BV is considered a sexually enhanced infection, and individuals with new or multiple sex partners, those who engage in unprotected vaginal sex, or those whose sex partner is uncircumcised are at an increased risk of BV [3, 7–9]. In addition to sexual risk factors, non-sexual risk factors that are associated with BV include intravaginal and vaginal hygiene practices [10], cigarette smoking [11], and race/ethnicity [12]. In numerous studies, women of African descent have increased risk of BV and non-optimal vaginal microbiome composition [13]. These associations with race persist even when controlling for socio-demographics and sexual practices [14, 15], leading to the hypothesis that genetic factors may influence the vaginal microbiome composition and subsequently BV [16].

To date, a limited number of studies have examined the host genetic contribution to the vaginal flora and BV. Primarily these studies have focused on host differences in candidate genes responsible for inflammatory mucosal immune response in other infectious conditions [16, 17]. As reviewed by Turovskiy et al., studies have targeted single nucleotide polymorphisms (SNPs) in: (1) Toll-like receptor (TLR) genes because of their role in recognizing potential threats and initiating inflammatory responses and activating other immune cells, and (2) pro-inflammatory and inflammatory cytokines (e.g., IL-1β, TNF-α) [15]. Most of these studies, however, evaluated SNPs in relation to BV or to specific taxa associated with BV (e.g., G. vaginalis, A. vaginae spp.) and none of the studies measured the vaginal microflora or host genetic contribution comprehensively, such as with cultivation-independent molecular microbial characterization or a genome wide association study (GWAS), respectively.

To address this gap, we conducted GWAS on vaginal microbiome traits among native Kenyan women. We hypothesized that by conducting GWAS, we would identify novel genomic loci associated with vaginal microbiome traits, furthering our understanding of the underlying genetic factors. Identifying genetic variants may aid in elucidating the biological mechanisms associated with these complex vaginal microbiome traits and associated diseases. Complementary to GWAS, we conducted replication analyses of SNPs previously reported to be associated with vaginal microflora or BV.

Results

Study Sample

Among 171 women included in this analysis, the prevalence of BV at baseline was 22% (**Table 1**). Over half (53%) of women were HSV-2 seropositive and 10% were HIV positive, which is in keeping with the prevalence of HIV among women in this age range in the general population in this region of western Kenya [18]. *L. crispatus* was present in 25% of samples (28% mean relative abundance), *L. iners* was present in 83% of samples (mean relative abundance 45%), and *G. vaginalis* was present in 75% of samples (25% mean relative abundance).

Genome-wide association results

The genomic control (GC) inflation factor, l, was 1.01, 1.00, 1.00, 1.00, and 1.00 for *L. crispatus*, *L. iners*, *G. vaginalis*, the Shannon diversity index, and CST, respectively, after adjusting for age and the first three principal components. Upon visual inspection of the Q-Q plots (**Supplementary Figure 1**), the observed *P* values do not deviate from the null, except at the extreme tail. Together with the genomic inflation factors, the Q-Q plots indicate proper control of population stratification.

Figure 1A-E displays the Manhattan plots of the genome-wide *P* values for each vaginal microbiome trait. **Table 2** summarizes the top SNPs ($P < 1 \times 10^{-5}$) for each vaginal microbiome trait and minor allele frequencies (MAF) derived from our data. Overall, no SNP reached genome-wide significance (0.05/336,151 = 1.49E-07). The SNP with the lowest adjusted p-value associated with presence of *L. crispatus* was rs73330467 ($P = 4.79 \times 10^{-6}$, **Figure 1A**), located in an intergenic region between *LOC101927488* and *GRAMD2B* on chromosome 5. The minor allele G (MAF = 0.07) was associated with an 11.6-fold increased odds of detecting *L. crispatus*. The most significant SNP associated with relative abundance of *L. iners* was rs527430 ($P = 6.98 \times 10^{-7}$, **Figure 1B**) on chromosome 1, located in an intergenic region between *FOXD2* and *TRABD2B*. The minor allele A (MAF = 0.06) was associated with an increase in relative abundance of *L. iners* (β = 1.05). The most significant SNP associated with *G. vaginalis* was rs1229660 ($P = 4.65 \times 10^{-6}$, **Figure 1C**) located between *SNX10-LOC441204* on chromosome 7. The minor allele C (MAF = 0.10) was associated with a decrease

in *G. vaginalis* relative abundance (β = -0.99). The most significant SNP associated with Shannon diversity index was rs972741 ($P = 8.52 \times 10^{-7}$, **Figures 1D**) located between *ZKSCAN2* and *HS3ST4* on chromosome 16. The minor allele A (MAF = 0.19) was associated with an increase in the Shannon diversity index (β= 0.66). The most significant SNP associated with CST was rs2302902 ($P = 3.09 \times 10^{-6}$, **Figure 1E**) located in *ELK3* on chromosome 12. The minor allele T (MAF = 0.20) was associated with an increased likelihood of membership in lower diversity CST-I and CST-II as compared to CST-IV (β= 0.41).

Results from Imputed SNPs

We performed genotype imputation in order to expand genomic coverage and interrogate additional SNPs not directly genotyped. After removing SNPs with low imputation quality and low frequency (Rsq < 0.30 and MAF < 1%), several imputed SNPs (but no additional genomic regions) were identified. **Figure 2** presents the regional SNP association plots for each of the vaginal microbiome traits. Within the *ZKSCAN2-HS3ST4* intergenic region on chromosome 16, several imputed SNPs were more significantly associated with the Shannon diversity index than those directly genotyped (**Figure 2E**). The most significant SNP in this region was rs115869045 ($P = 4.77 \times 10^{-7}$, Rsq = 0.70).

Conditional Analyses

We conducted conditional analyses for the genomic loci in which the most significant SNP was identified for each vaginal microbiome trait to determine whether additional SNPs contribute to the observed associations. **Supplementary Figure 2** displays the regional SNP associations conditioning on the most significant directly genotyped SNP by including the SNP as a covariate in the regression model. As shown in each plot, all of the immediate SNP associations reduced towards the null, suggesting the SNPs identified are the leading SNPs for the observed associations.

Analysis of Previously Reported Loci

We investigated the associations of previously reported SNPs in relation to BV or vaginal microflora to determine whether these associations are relevant in a Kenyan population. **Figure 3** and **Supplemental Table 1** summarize these findings. Overall, 49 autosomal SNPs were identified from previous studies, of which 45 SNPs exhibited high imputation quality (Rsq ³ 0.30). When analyzed, 7 SNPs for *L. iners*, 8 SNPs for *L. crispatus*, 6 SNPs for *G. vaginalis*, 10 SNPs for the Shannon Diversity Index, and 8 SNPs for CST had a $P < 0.05$ in our sample. These findings, however, were not significant after applying a Bonferroni correction (0.05 / 49 = 0.00102).

To replicate genomic loci associated with the microbiome traits, we extracted the most significant imputed SNP within ± 100 kb of the previously reported SNPs for each microbiome trait in our study. This approach resulted in 6 genomic regions for *L. iners*, 1 genomic region for *G. vaginalis*, 7 genomic regions for Shannon Diversity Index, and 4 genomic regions for CST that were significant after Bonferroni correction (**Figure 3, Supplemental Table 1**). No regions were significant after multiple testing correction for *L. crispatus*. The replicated genomic regions include *IL1RN* (CST), *IL5* (Shannon), *IL5RA* (*L. iners,* Shannon), *IL-8* (Shannon, CST), *TIRAP* (*L. iners*, Shannon), *TLR2* (Shannon, CST), *TLR3* (*L. iners, Shannon)*, *TLR4* (*L. iners*), *MBL2* (*L. iners*, *G. vaginalis*, CST), and *MYD88* (*L. iners,* Shannon).

Pathway and Pleiotropy Analyses

Table 3 presents the top five KEGG and Reactome pathways from WebGestalt for each vaginal microbiome trait. Four pathways were significantly associated with *L. crispatus* after Benjamini-Hochberg multiple testing adjustment, including G-protein-coupled receptor (GPCR) ligand binding and neutrophil degranulation. Five pathways were associated with *L. iners*: 4 related to TLR cascades, and one to the TLR upstream component (MyD88:Mal TIRAP; MyD88-adaptor-like TIR-domain containing adaptor protein). Two biological pathways related to GPCR signaling were associated with *G. vaginalis*, as well as 'neutrophil degranulation', though it did not meet Bonferroni significance. Similarly, two pathways were associated with CST, including the most significant pathway 'Class B/2 (secretin family receptors)', and one pathway was associated with the Shannon diversity index, 'Class B/2 (secretin family receptors)'.

Table 4 summarizes the top five DisGeNET, GLAD4U, OMIM, and Human Phenotype Ontology terms from WebGestalt for each vaginal microbiome trait. Five phenotypes were significantly associated with *L. crispatus* after Benjamini-Hochberg multiple testing adjustment, including 'Bacterial infections' ($P = 0.011$) as the most significant phenotype. Three phenotypes were associated with *L. iners*, with 'Autosomal dominant inheritance' the most significant ($P = 0.018$). No phenotypes were significantly associated with *G. vaginalis*. One phenotype was associated with the Shannon diversity index, 'Abnormality of the female genitalia' ($P = 0.037$). Five phenotypes were associated with CST, including 'Abnormality of the integument' ($P = 0.00023$).

Discussion

To our knowledge, this is the first GWAS of vaginal microbiome traits. In native Kenyan women, we identified novel genomic loci and biological pathways with biological relevance to host immunity, cell signaling, and infection, supporting the role of host genetics in inter-individual variability in vaginal microbiome traits.

In GWAS, the most significant SNP associated with CST (rs2302902) is located in an intron of ELK3, a member of the ETS transcription factor family, which has not been previously associated with any vaginal microbiome traits or BV. Proteins in the ELK3 subfamily are recruited by serum response factor and participate in transcription regulation, and has been associated in various cancers [19–21], including HPV-positive tumors of oropharyngeal cancer [22]and HPV16 in cervical tumors [23]. In meta-analysis, molecular and clinical measure of BV is associated with two-fold increase in risk of high-grade cervical epithelial neoplasia (CIN) or cancer [24]. Non-optimal vaginal microbiome leads to mucosal disruption and a pro-inflammatory environment that can facilitate HPV acquisition [25], but there may also be underlying shared or synergistic mechanism with ELK3. Other SNPs identified in GWAS occurred primarily in intergenic regions that haven't previously been associated with BV or vaginal microbiome traits, and should be assessed for replication in future studies.

BV was absent from CST-I (L. crispatus dominant) and infrequent (5.1%) in CST-III. Although CST-IV accounted for 89% of BV cases in this sample, 57% of women in CST-IV did not have BV. This is in keeping with many other studies: women with diverse CSTs are much more likely to have BV, but as many as 50% of women with these CSTs do not have BV [26]. Host genetics may partially explain this variability. For example, women with similar vaginal bacterial colonization (such as those within CST-IV) may have different response to bacterial colonization depending on genetic traits. For women who mount an inflammatory response, this immune response may potentiate BV rather being solely a response to BV. In our study, restricting analyses to HIV uninfected women did not change our findings. This is not entirely unexpected, as associations between HIV and subsequent impact on the vaginal microbiome are variable, with some finding no effect [27, 28]. Conversely, there is substantial data that demonstrate the temporal association between non-optimal vaginal microbiome composition, mucosal inflammation, and subsequent HIV risk [29], and our results suggest that host genetic

differences may underpin this. Well-powered longitudinal studies that examine how host genetics is associated with varying CST, mucosal immune markers, and BV trajectories could be informative to understanding this variability in pathology within non-optimal CSTs. Additionally studies are needed to explore the potential long range regulation of these SNPs on genes, as well as evaluate for expression or methylation quantitative trait loci associations.

Complex traits are polygenic, and our analysis of previously reported loci identified multiple SNPs across numerous genetic loci with Bonferroni corrected significance. Several SNPs in the MBL2 region were significantly associated with L. iners, G. vaginalis, and CST. Kalia et al. studied these previously reported MBL2 SNPs in association with BV in women in India due to associations between mannose binding lectin (MBL) insufficiency and increased susceptibility to infectious diseases [30]. Among Brazilian women, a specific MBL2 polymorphism was associated with increased odds of recurrent BV compared to controls [31], though a study among Italian women that screened for three specific MBL2 SNPs found no differences for women with recurrent BV as compared to controls [32]. The relevance of this gene may be population specific or subject to interaction with other factors such as the prevalence of co-infections (e.g., HSV-2, HIV), sexual exposures (e.g., partner circumcision status, multiple sex partners, condom use), or non-sexual factors (e.g., cigarette smoking, intravaginal practices). Additionally, we identified SNPs in IL-1RN (CST), IL-5 (Shannon), and IL5RA (L. iners, Shannon). IL-1RN (the gene for interleukin 1 receptor agonist, IL-1RA) inhibits the inflammatory IL-1 cytokines, IL-1α, and IL-1β. In a study of 62 Italian women, compared to women with Nugent score 0–3, women with BV (Nugent score 7–10) had increased IL-1RA, and increased inflammatory IL-5 was associated with decreased Lactobacillus [33]. In candidate gene analysis, Si et al. targeted genetic variants of IL-5 finding association with increased abundances of various Prevotella [34]. Our findings further these studies by analyzing genetic factors and vaginal microbiome traits in a novel population and by demonstrating trans-ethnic associations for these previously identified loci.

In pathway analysis, GPCR pathways were identified with L. crispatus, G. vaginalis, and CST, and Class B/2 (secretin family receptors) within the GPCR family were identified with Shannon diversity index and CST. GPCRs play a significant role in intracellular signaling in response to a broad range of stimuli (e.g., hormones, neurotransmitters, proteins, etc.) and across a wide range of functions (e.g., growth, nutrition requirements, response to disease, etc.) [35]. This likely includes vaginal microbiome modulation, and variation in GPCR signaling may related to variations in innate immune response. LL-37 is elevated in women with BV and in an ex vivo endocervical model, after application of a GPCR inhibitor, LL-37 mediated induction of IL-8 production was inhibited [36]; this is relevant given that IL-8 is often elevated in women with BV [37]. In analyses of previously reported SNPs, we found SNPs in the IL-8 gene region were associated with Shannon diversity index and CST. In an experimental study, Mares et al. demonstrate that application of CVL from women with BV induced higher levels of IL-8 and NFkappaB in human monocytes than CVL from women without BV [38]. Among women in the CAPRISA 004 trial, those with vaginal CST-IV had increased mucosal inflammation (including elevated IL-8) and increased risk of HIV seroconversion [29]. The effects of variation in GPCR signaling on innate immune response may extend to other outcomes. Among Dutch women, GWAS and pathway analysis found that genes encoding GPCR signaling were enriched for 71 Chlamydia trachomatis seropositive women compared to 169 control women [39]. GPCR signaling encompasses a broad range of actions, but in conjunction with our pathway results implicating IL-8 and TLRs, and other studies showing potential roles for GPCR in cervicovaginal health, GPCR signaling merits further study as a potential pathway affecting acquisition or maintenance of vaginal L. crispatus, CST membership, and subsequent BV status.

Our pathway analysis additionally linked neutrophil degranulation with L. crispatus and G. vaginalis. Neutrophils are effector cells in innate immunity and degranulation can have serious adverse consequences for the host due to release of damaging molecules. We are unaware of other studies linking neutrophil degranulation to the vaginal microbiome, BV, or other vaginal infections. However, myeloperoxidase (MPO) is a measure of neutrophils and is elevated in women with sexually transmitted infection (STI) [40]. Whether neutrophil degranulation is an important process in vaginal

microbiome composition may be evaluated in subsequent studies, such as by targeting genes encoding molecules associated with this pathway (e.g., β-arrestins, Rac2) [41].

Lastly, pathway analysis identified 'MyD88:MAL (TIRAP) cascade' with L. iners, and we identified SNPs within the MyD88 gene in relation to L. iners and Shannon diversity index. MyD88:Mal (TIRAP) is an adaptor to MyD88, the first downstream component of TLR4 and TLR2, and part of the larger overall process for inflammatory cytokine effects [42]. In pathway analysis, several TLR signaling associations were associated with L. iners relative abundance, and in analysis of previously reported SNPs, TLR2, and TLR3 had associations with Shannon diversity index, CST, and L. iners. As summarized in Supplementary Table 1, several investigators have found polymorphisms in TLR genes related to altered risk of BV and enrichment of BV-related bacteria. TLRs, especially TLR2 alone or in combination with TLR-1/TLR-6 and TLR-4, have been recognized for their association with BV and BV-related bacteria, as these TLRs precipitate release of pro-inflammatory cytokines and recruitment of inflammatory cells [43–47]. As summarized by Taylor et al., TLR gene variants can lead to inadequate or excess inflammatory immune response, which can affect disease susceptibility or progression [44]. Additionally studies are needed to evaluate the potential biological role TLRs have on vaginal microbiome composition. Given that the microbiome is a complex trait and complex traits are pleiotropic, this may provide understanding of how TLR variants contribute to risk of BV, response to BV treatment, or other outcomes of non-optimal vaginal microbiome composition, such as adverse pregnancy outcomes. Taken together, our findings demonstrate numerous biologically relevant pathways and phenotypes associated with vaginal microbiome traits.

Limitations

We found substantial functional overlap in several biologically relevant innate immunity pathways. This may in part be driven by the compositionality of individual taxa. To mitigate this, we limited examination to a few traits of known importance in conjunction with comprehensive community characterization. We also confirmed numerous SNPs in close proximity to genomic loci associated with innate immunity reported in previous studies of host genetic traits associated with BV and vaginal taxa. Although we identified novel candidate regions associated with vaginal microbiome traits, additional studies are needed to replicate our findings. Evaluation of functional impact with multi-omic studies (such as gene expression, epigenetics, and fine mapping) would allow mechanistic inference and aid in elucidating regulatory elements, which could lead to translational advances in optimizing the vaginal microbiome and preventing BV and associated consequences.

Conclusions

Given the multiple associations of important microbiome traits with various immune signaling pathways, the composition of the vaginal microbiome (and subsequently, BV risk) likely is influenced by host genetics through immune phenotypes. Replication studies and functional studies are needed to confirm and further elucidate understanding of how vaginal bacteria interact with the genetic make-up of the host to effect non-optimal vaginal microbiome and related outcomes. Given the co-occurrence of risk for BV, non-optimal vaginal microbiome, and HIV and STI among African and African American women, well-powered cross-national studies are needed to answer whether common variants are replicated across populations, and to identify important gene by environment interactions that inform how much risk for non-optimal vaginal microbiome may be modifiable, such as through behavioral modification or prevention or treatment of co-infections (e.g., HSV-2, HIV).

Methods

This study was approved by the Ethical Review Committee of Maseno University (Kisumu, Kenya; MSU/DRPC/MUERC/00054/13; January 13, 2014), and the Institutional Review Board of the University of Illinois at

Chicago (USA; 2013-0511; February 12, 2014).

Study Design and Participants

Subjects in this analysis were enrolled in *Afya Jozi, Afya Jamii* (Kiswahili for "Healthy Pair, Healthy Community"), a prospective cohort study of heterosexual couples in Kisumu, Kenya. Recruitment and eligibility criteria have been previously published [48]. Briefly, we included men aged 18-35 years and their female partners aged 16 years and older, who independently confirmed a sexual relationship of at least 6 months duration. After the baseline visit, couples were scheduled for follow-up at 1 month, 6 months, and 12 months. Couples were enrolled from April 1, 2014, through June 22, 2016 and 12-month follow-up was completed June 21, 2017.

Specimen Collection

At each scheduled study visit, participants underwent a standardized medical history and physical examination, followed by a personal interview to obtain socio-demographic and behavioral information. At baseline and each follow-up visit, clinicians collected vaginal swabs for assessment of BV. Gram stained slides were evaluated according to Nugent's criteria, and a score of 7-10 was defined as BV following clinical recommendations [49]. Vaginal swab collection was followed by collection of cervicovaginal lavage (CVL) in 10 ml sterile water. Aliquots of 2mL were stored at -80º until shipment to Chicago for processing.

Vaginal Microbiome Characterization

We used the vaginal microbiome as measured from the baseline visit for this analysis, because antimicrobial treatment was provided for BV and other cervicovaginal infections which may confound subsequent vaginal microbiome observations. DNA extraction, library preparation and sequencing were performed at the University of Illinois at Chicago Sequencing Core (UICSQC). DNA was extracted from CVL specimens using a PowerFecal DNA kit (Qiagen). A two-stage PCR protocol was used to amplify the V3-V4 variable region of bacterial 16S rRNA genes [50]. Illumina MiSeq was used to sequence the amplicons after pooling, implementing V3 chemistry (600 cycles). Quality control and taxonomic annotation were conducted by University of Maryland Institute for Genomic Science (UMD IGS) following previously published protocol [51]. Community state types (CSTs) were determined by hierarchical clustering of relative abundances of taxa, using VALENCIA algorithm implemented by UMD IGS, which uses a distance-based metric to classify each sample to a CST based on the similarity of that sample to the centroid of CSTs identified in a reference set [52]. Raw sequence data files are available in the Sequence Read Archive (National Center for Biotechnology Information; BioProject identifier PRJNA516684).

Genotyping and Quality Control

Genomic DNA was obtained from 200 Kenyan women using oral swabs and was extracted using QIAamp DNA blood mini kit (Qiagen). DNA extraction and isolation was performed at the Center for Population Epigenetics at Northwestern University. Study participants were genotyped using the Infinium Global Screening Array (~640,000 markers; Illumina, Inc., San Diego). The software Illumina GenomeStudio (v2.0.4, Illumina, Inc., San Diego) was used to call single nucleotide polymorphisms (SNPs) at the Genomic Facility, University of Chicago. Study participants with a genotyping

call rate less than 95% were excluded from analysis as were study participants with missing vaginal microbiome traits. Due to cryptic relatedness identified during quality control, we restricted all analyses to the maximum number of unrelated individuals. Implementation of these exclusion criteria resulted in 171 study participants for analysis. SNPs with a call rate less than 95%, a minor allele frequency less than 1%, or a Hardy-Weinberg equilibrium *P* value less than 10^{-6} were omitted. After these exclusion criteria, 336,151 remained for further analysis.

Genotype Imputation

We performed genotype imputation to identify additional SNPs associated with the vaginal microbiome traits that were not directly genotyped using the Luhya in Webuye, Kenya reference panel from the 1000 Genomes Project [53]. The program Shapeit2 [54] was used to phase genotypes and imputation was performed using Minimac3 [55]. Imputed genotypes were coded as allelic dosages (estimated counts ranging from 0 to 2). Low quality imputed SNPs, i.e. Rsq < 0.30 and MAF < 1%, were excluded from further analyses.

Statistical Analysis

We evaluated the association between single nucleotide polymorphisms (SNPs) with five vaginal microbiome traits: (1) presence of *L. crispatus*, relative abundances of (2) *L. iners* and (3) *G. vaginalis*, (4) Shannon diversity index, and (5) CST. We chose *L. crispatus* for its known importance in vaginal microbiome health, *L. iners* due to its predominance among *Lactobacillus* species in our sample and other African samples, and *G. vaginalis* due to its inclusion as a morphotype in Nugent scoring and determination of BV. For alpha diversity measures, we selected Shannon diversity index, as all alpha diversity measures were highly correlated (Spearman correlations, 0.77 – 0.98). Prior to analyses, the relative abundance of *L. iners* and Shannon diversity index were inverse log-transformed to approximate a more normal distribution (**Supplemental Figure 2**). As transformation did not lead to normal approximation, the relative abundance of *G. vaginalis* was analyzed by quartile. Given the infrequency of *L. crispatus* (present in 24% of samples) we analyzed it as presence vs. absence. Of 171 subjects in the analysis, 168 (98%) baseline vaginal microbial communities were classified as CST-I (n=14; 0% with BV), CST-III (n=77; 5.2% with BV), or CST-IV (n=77; 43% with BV). There were three outliers (n=2 with CST-II, n=1 with CST-V), and these observations were excluded from CST analyses. BV was not analyzed as a separate outcome because it was collinear with CST. We analyzed CST as a linear outcome given the ordinal nature of BV risk, and to meet parameters for statistical software.

Principal components of genetic ancestry were generated using the program EIGENSOFT [56] and the first three principal components were retained and included as covariates. Quantile-quantile (Q-Q) plots were generated for each trait to visualize test statistics and the genomic control inflation factors were calculated. We used PLINK (v1.90) [57] to test the associations between *L. iners*, *G. vaginalis*, Shannon diversity index, and CST with SNPs using linear regression, and association between *L. crispatus* and SNPs using logistic regression, adjusting for age and principal components. We assumed an additive genetic effects model for all analyses. To account for genotype uncertainty, the allelic dosages from imputed SNPs were analyzed using Mach2QTL for linear regression and Mach2DAT for logistic regression (see Web Resources). SNPs with a *P* value less than 1.49×10^{-7} (0.05/336,151) were declared genome-wide significant. The software R (3.5.0) [58] and LocusZoom [59] (hg19/1000 Genomes AFR) were used for graphing.

Pathway and Pleiotropy Analysis

We performed pathway and pleiotropy analysis using SKAT-O [60] and WebGestalt [61] to examine the combined genetic effects on the underlying biological mechanisms influencing the vaginal microbiome traits. We mapped all directly genotyped SNPs to autosomal genes based on the genomic positions of the GRCh37/hg19 assembly and included ±20 kb gene boundary to capture proximal regulatory and functional elements that contribute to gene regulation. Gene-set associations were performed using SKAT-O, adjusting for age and the first three principal components. Over-representation analyses of pathways (Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome) and diseases / phenotypes (DisGeNET, GLAD4U, Online Mendelian Inheritance in Man (OMIM), and Human Phenotype Ontology) was analyzed using WebGestalt. Pathways, diseases, and phenotypes with a false discovery rate (FDR) P value < 0.05 after Benjamini-Hochberg multiple testing adjustment were declared significant. The top five pathways and diseases / phenotypes from WebGestalt are reported.

Web Resources

Mach2DAT, http://csg.sph.umich.edu/abecasis/MACH/download/

Mach2QTL, http://csg.sph.umich.edu/abecasis/MACH/download/

Supplemental Figure Information

Supplemental Figure 1A-1E. Q-Q plots for five vaginal microbiome traits measured in 171 Kenyan women.

Legend: Inspection of Q-Q plots for each microbiome trait evaluated indicate proper control of population stratification.

Supplemental Figure 2A-E. Regional conditional association plots of the top genomic loci associated with vaginal microbiome traits.

Legend: P values (−log10) of the GWAS (solid circle) on the y-axis are plotted against the genomic positions of each SNP on the x-axis for each microbiome trait. Genes in the region are shown below. The linkage disequilibrium (LD) values (r^2) between the lead SNP and the other SNPs are indicated in different colors.

Declarations

- **Ethics approval and consent to participate**: This study was approved by the Ethical Review Committee of Maseno University (Kisumu, Kenya; MSU/DRPC/MUERC/00054/13; January 13, 2014), and the Institutional Review Board of the University of Illinois at Chicago (USA; 2013-0511; February 12, 2014). Written informed consent was obtained from all participants in their language of choice (English, DhoLuo, or Kiswahili).
- **Consent for publication**: Not applicable.
- **Availability of data and material**: The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.
- **Competing interests**: The authors declare that they have no competing interests
- **Funding**: This study was supported by grant number R01-AI110369 (PI: Mehta) from the National Institutes of Health, National Institute of Allergy and Infectious Diseases, Division of Microbiology, and by the National Cancer Institute of the National Institutes of Health under award number U54CA221205 (MPI: Hou, Murphy). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

- **Authors' contributions**:
  - **SDM**: Obtained funding; study conceptualization and design; drafted manuscript.
  - **DN**: Design and implementation of statistical analysis approach; visualization; critical review and revision of manuscript
  - **FO**: Study oversight and management to ensure integrity to protocols; critical review and revision of manuscript
  - **SJG**: Development and oversight of protocols for amplicon sequencing; microbiologic analyses and interpretation; critical review and revision of manuscript
  - **WA**: Development, implementation, and oversight of laboratory protocols in Kenya; acquisition of data; microbiologic analyses and interpretation; critical review and revision of manuscript.
  - **AL**: Critical review and revision of manuscript.
  - **YZ**: Design and interpretation of statistical analysis approach; critical review and revision of manuscript
  - **LH**: Selection of GWAS methods; design of analysis; interpretation of results; critical review and revision of manuscript.
- **Acknowledgements:** Not applicable
- **Authors' information (optional)**: Not applicable

References

1. Torrone EA, Morrison CS, Chen PL, Kwok C, Francis SC, Hayes RJ, Looker KJ, McCormack S, McGrath N, van de Wijgert J, et al: **Prevalence of sexually transmitted infections and bacterial vaginosis among women in sub-Saharan Africa: An individual participant data meta-analysis of 18 HIV prevention studies.** *PLoS Med* 2018, **15:**e1002511.
2. Atashili J, Poole C, Ndumbe PM, Adimora AA, Smith JS: **Bacterial vaginosis and HIV acquisition: a meta-analysis of published studies.** *AIDS* 2008, **22:**1493-1501.
3. Lewis FM, Bernstein KT, Aral SO: **Vaginal Microbiome and Its Relationship to Behavior, Sexual Health, and Sexually Transmitted Diseases.** *Obstet Gynecol* 2017, **129:**643-654.
4. van de Wijgert J, Jespers V: **The global health impact of vaginal dysbiosis.** *Res Microbiol* 2017, **168:**859-864.
5. Fredricks DN: **Molecular methods to describe the spectrum and dynamics of the vaginal microbiota.** *Anaerobe* 2011, **17:**191-195.
6. van de Wijgert JH, Borgdorff H, Verhelst R, Crucitti T, Francis S, Verstraelen H, Jespers V: **The vaginal microbiota: what have we learned after a decade of molecular characterization?** *PLoS One* 2014, **9:**e105998.
7. Fethers KA, Fairley CK, Hocking JS, Gurrin LC, Bradshaw CS: **Sexual risk factors and bacterial vaginosis: a systematic review and meta-analysis.** *Clin Infect Dis* 2008, **47:**1426-1435.
8. Gray RH, Kigozi G, Serwadda D, Makumbi F, Nalugoda F, Watya S, Moulton L, Chen MZ, Sewankambo NK, Kiwanuka N, et al: **The effects of male circumcision on female partners' genital tract symptoms and vaginal infections in a randomized trial in Rakai, Uganda.** *Am J Obstet Gynecol* 2009, **200:**42 e41-47.
9. Verstraelen H, Verhelst R, Vaneechoutte M, Temmerman M: **The epidemiology of bacterial vaginosis in relation to sexual behaviour.** *BMC Infect Dis* 2010, **10:**81.
10. Low N, Chersich MF, Schmidlin K, Egger M, Francis SC, van de Wijgert JH, Hayes RJ, Baeten JM, Brown J, Delany-Moretlwe S, et al: **Intravaginal practices, bacterial vaginosis, and HIV infection in women: individual participant data meta-analysis.** *PLoS Med* 2011, **8:**e1000416.
11. Nelson TM, Borgogna JC, Michalek RD, Roberts DW, Rath JM, Glover ED, Ravel J, Shardell MD, Yeoman CJ, Brotman RM: **Cigarette smoking is associated with an altered vaginal tract metabolomic profile.** *Sci Rep* 2018, **8:**852.

12. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SS, McCulle SL, Karlebach S, Gorle R, Russell J, Tacket CO, et al: **Vaginal microbiome of reproductive-age women.** *Proc Natl Acad Sci U S A* 2011, **108 Suppl 1:**4680-4687.
13. Peebles K, Velloza J, Balkus JE, McClelland RS, Barnabas RV: **High Global Burden and Costs of Bacterial Vaginosis: A Systematic Review and Meta-Analysis.** *Sex Transm Dis* 2019, **46:**304-311.
14. Alcendor DJ: **Evaluation of Health Disparity in Bacterial Vaginosis and the Implications for HIV-1 Acquisition in African American Women.** *Am J Reprod Immunol* 2016, **76:**99-107.
15. Fettweis JM, Brooks JP, Serrano MG, Sheth NU, Girerd PH, Edwards DJ, Strauss JF, The Vaginal Microbiome C, Jefferson KK, Buck GA: **Differences in vaginal microbiome in African American women versus women of European ancestry.** *Microbiology* 2014, **160:**2272-2282.
16. Martin DH, Marrazzo JM: **The Vaginal Microbiome: Current Understanding and Future Directions.** *J Infect Dis* 2016, **214 Suppl 1:**S36-41.
17. Turovskiy Y, Sutyak Noll K, Chikindas ML: **The aetiology of bacterial vaginosis.** *J Appl Microbiol* 2011, **110:**1105-1128.
18. **Kenya Ministry of Health** [http://nacc.or.ke/wp-content/uploads/2016/12/Kenya-HIV-County-Profiles-2016.pdf]
19. Yang H, Schramek D, Adam RC, Keyes BE, Wang P, Zheng D, Fuchs E: **ETS family transcriptional regulators drive chromatin dynamics and malignancy in squamous cell carcinomas.** *Elife* 2015, **4:**e10870.
20. Heo SH, Lee JY, Yang KM, Park KS: **ELK3 Expression Correlates With Cell Migration, Invasion, and Membrane Type 1-Matrix Metalloproteinase Expression in MDA-MB-231 Breast Cancer Cells.** *Gene Expr* 2015, **16:**197-203.
21. Lee JH, Hur W, Hong SW, Kim JH, Kim SM, Lee EB, Yoon SK: **ELK3 promotes the migration and invasion of liver cancer stem cells by targeting HIF-1alpha.** *Oncol Rep* 2017, **37:**813-822.
22. Suarez E, Gonzalez L, Perez-Mitchell C, Ortiz AP, Ramirez-Sola M, Acosta J, Bernabe-Dones RD, Gonzalez-Aquino C, Montes-Rodriguez I, Cadilla CL: **Pathway Analysis using Gene-expression Profiles of HPV-positive and HPV-negative Oropharyngeal Cancer Patients in a Hispanic Population: Methodological Procedures.** *P R Health Sci J* 2016, **35:**3-8.
23. Thorland EC, Myers SL, Gostout BS, Smith DI: **Common fragile sites are preferential targets for HPV16 integrations in cervical tumors.** *Oncogene* 2003, **22:**1225-1237.
24. Brusselaers N, Shrestha S, van de Wijgert J, Verstraelen H: **Vaginal dysbiosis and the risk of human papillomavirus and cervical cancer: systematic review and meta-analysis.** *Am J Obstet Gynecol* 2019, **221:**9-18 e18.
25. Mitra A, MacIntyre DA, Marchesi JR, Lee YS, Bennett PR, Kyrgiou M: **The vaginal microbiota, human papillomavirus infection and cervical intraepithelial neoplasia: what do we know and where are we going next?** *Microbiome* 2016, **4:**58.
26. Ravel J, Brotman RM: **Translating the vaginal microbiome: gaps and challenges.** *Genome Med* 2016, **8:**35.
27. Reimers LL, Mehta SD, Massad LS, Burk RD, Xie X, Ravel J, Cohen MH, Palefsky JM, Weber KM, Xue X, et al: **The Cervicovaginal Microbiota and Its Associations With Human Papillomavirus Detection in HIV-Infected and HIV-Uninfected Women.** *J Infect Dis* 2016, **214:**1361-1369.
28. Watts DH, Springer G, Minkoff H, Hillier SL, Jacobson L, Moxley M, Justman J, Cejtin H, O'Connell C, Greenblatt RM: **The occurrence of vaginal infections among HIV-infected and high-risk HIV-uninfected women: longitudinal findings of the women's interagency HIV study.** *J Acquir Immune Defic Syndr* 2006, **43:**161-168.
29. Eastment MC, McClelland RS: **Vaginal microbiota and susceptibility to HIV.** *AIDS* 2018, **32:**687-698.
30. Kalia N, Singh J, Sharma S, Kaur M: **Impact of SNPs interplay across the locus of MBL2, between MBL and Dectin-1 gene, on women's risk of developing recurrent vulvovaginal infections.** *Cell Biosci* 2019, **9:**35.
31. Giraldo PC, Babula O, Goncalves AK, Linhares IM, Amaral RL, Ledger WJ, Witkin SS: **Mannose-binding lectin gene polymorphism, vulvovaginal candidiasis, and bacterial vaginosis.** *Obstet Gynecol* 2007, **109:**1123-1128.

32. Milanese M, Segat L, De Seta F, Pirulli D, Fabris A, Morgutti M, Crovella S: **MBL2 genetic screening in patients with recurrent vaginal infections.** *Am J Reprod Immunol* 2008, **59:**146-151.
33. Campisciano G, Zanotta N, Licastro D, De Seta F, Comar M: **In vivo microbiome and associated immune markers: New insights into the pathogenesis of vaginal dysbiosis.** *Sci Rep* 2018, **8:**2307.
34. Si J, You HJ, Yu J, Sung J, Ko G: **Prevotella as a Hub for Vaginal Microbiota under the Influence of Host Genetics and Their Association with Obesity.** *Cell Host Microbe* 2017, **21:**97-105.
35. Hilger D, Masureel M, Kobilka BK: **Structure and dynamics of GPCR signaling complexes.** *Nat Struct Mol Biol* 2018, **25:**4-12.
36. Frew L, Makieva S, McKinlay AT, McHugh BJ, Doust A, Norman JE, Davidson DJ, Stock SJ: **Human cathelicidin production by the cervix.** *PLoS One* 2014, **9:**e103434.
37. St John E, Mares D, Spear GT: **Bacterial vaginosis and host immunity.** *Curr HIV/AIDS Rep* 2007, **4:**22-28.
38. Mares D, Simoes JA, Novak RM, Spear GT: **TLR2-mediated cell stimulation in bacterial vaginosis.** *J Reprod Immunol* 2008, **77:**91-99.
39. Roberts CH, Ouburg S, Preston MD, de Vries HJC, Holland MJ, Morre SA: **Pathway-Wide Genetic Risks in Chlamydial Infections Overlap between Tissue Tropisms: A Genome-Wide Association Scan.** *Mediators Inflamm* 2018, **2018:**3434101.
40. Spear GT, Kendrick SR, Chen HY, Thomas TT, Bahk M, Balderas R, Ghosh S, Weinberg A, Landay AL: **Multiplex immunoassay of lower genital tract mucosal fluid from women attending an urban STD clinic shows broadly increased IL1ss and lactoferrin.** *PLoS One* 2011, **6:**e19560.
41. Lacy P: **Mechanisms of degranulation in neutrophils.** *Allergy Asthma Clin Immunol* 2006, **2:**98-108.
42. Gangloff M, Gay NJ: **MD-2: the Toll 'gatekeeper' in endotoxin signalling.** *Trends Biochem Sci* 2004, **29:**294-300.
43. Royse KE, Kempf MC, McGwin G, Jr., Wilson CM, Tang J, Shrestha S: **Toll-like receptor gene variants associated with bacterial vaginosis among HIV-1 infected adolescents.** *J Reprod Immunol* 2012, **96:**84-89.
44. Taylor BD, Totten PA, Astete SG, Ferris MJ, Martin DH, Ness RB, Haggerty CL: **Toll-like receptor variants and cervical Atopobium vaginae infection in women with pelvic inflammatory disease.** *Am J Reprod Immunol* 2018, **79**.
45. Taylor BD, Darville T, Ferrell RE, Ness RB, Kelsey SF, Haggerty CL: **Cross-sectional analysis of Toll-like receptor variants and bacterial vaginosis in African-American women with pelvic inflammatory disease.** *Sex Transm Infect* 2014, **90:**563-566.
46. Mackelprang RD, Scoville CW, Cohen CR, Ondondo RO, Bigham AW, Celum C, Campbell MS, Essex M, Wald A, Kiarie J, et al: **Toll-like receptor gene variants and bacterial vaginosis among HIV-1 infected and uninfected African women.** *Genes Immun* 2015, **16:**362-365.
47. Ryckman KK, Williams SM, Krohn MA, Simhan HN: **Genetic association of Toll-like receptor 4 with cervical cytokine concentrations during pregnancy.** *Genes Immun* 2009, **10:**636-640.
48. Mehta SD, Nordgren RK, Agingu W, Otieno F, Odongo W, Odhiambo F, Bailey RC: **Sexual Quality of Life and Association With HIV and Sexually Transmitted Infections Among a Cohort of Heterosexual Couples in Kenya.** *J Sex Med* 2018, **15:**1446-1455.
49. Nugent RP, Krohn MA, Hillier SL: **Reliability of diagnosing bacterial vaginosis is improved by a standardized method of gram stain interpretation.** *J Clin Microbiol* 1991, **29:**297-301.
50. Naqib A, Poggi S, Wang W, Hyde M, Kunstman K, Green SJ: **Making and Sequencing Heavily Multiplexed, High-Throughput 16S Ribosomal RNA Gene Amplicon Libraries Using a Flexible, Two-Stage PCR Protocol.** *Methods Mol Biol* 2018, **1783:**149-169.
51. Holm JB, Humphrys MS, Robinson CK, Settles ML, Ott S, Fu L, Yang H, Gajer P, He X, McComb E, et al: **Ultrahigh-Throughput Multiplexing and Sequencing of >500-Base-Pair Amplicon Regions on the Illumina HiSeq 2500 Platform.**

mSystems 2019, **4**.

52. Brown SE, Schwartz JA, Robinson CK, O'Hanlon DE, Bradford LL, He X, Mark KS, Bruno VM, Ravel J, Brotman RM: **The Vaginal Microbiota and Behavioral Factors Associated With Genital Candida albicans Detection in Reproductive-Age Women.** *Sex Transm Dis* 2019, **46:**753-758.
53. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR: **A global reference for human genetic variation.** *Nature* 2015, **526:**68-74.
54. Delaneau O, Marchini J, Genomes Project C, Genomes Project C: **Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel.** *Nat Commun* 2014, **5:**3934.
55. Das S, Forer L, Schonherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, Chew EY, Levy S, McGue M, et al: **Next-generation genotype imputation service and methods.** *Nat Genet* 2016, **48:**1284-1287.
56. Patterson N, Price AL, Reich D: **Population structure and eigenanalysis.** *PLoS Genet* 2006, **2:**e190.
57. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ: **Second-generation PLINK: rising to the challenge of larger and richer datasets.** *Gigascience* 2015, **4:**7.
58. R Development Core Team: **In R Foundation for Statistical Computing,.** Vienna, Austria; 2009.
59. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ: **LocusZoom: regional visualization of genome-wide association scan results.** *Bioinformatics* 2010, **26:**2336-2337.
60. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, Team NGESP-ELP, Christiani DC, Wurfel MM, Lin X: **Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies.** *Am J Hum Genet* 2012, **91:**224-237.
61. Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B: **WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs.** *Nucleic Acids Res* 2019, **47:**W199-W205.

Tables

Table 1. Baseline participant characteristics and vaginal microbiome composition overall and stratified by community state type*.

	Total, N=171 n (%)	CST I, N=14 n (%)	CST-III, N=77 n (%)	CST-IV, N=77 n (%)
Median age in years (IQR)	22 (20-25)	24.5 (19-28)	23 (20-25)	22 (20-25)
HIV status				
Negative	152 (89.9)	14 (100)	70 (92.1)	65 (85.5)
Positive	17 (10.1)		6 (7.9)	11 (14.5)
Missing	2		1	1
HSV-2 status				
Negative	81 (47.4)	12 (85.7)	39 (50.6)	28 (36.4)
Positive	90 (52.6)	2 (14.3)	38 (49.4)	49 (63.6)
Circumcised male sex partner	91 (53.2)	6 (42.9)	43 (55.8)	40 (52.0)
BV (Nugent score 7-10) at baseline	37 (21.6)	0 (0)	4 (5.2)	33 (42.9)
Proportion of visits with BV over follow-up				
0%	89 (52.1)	14 (100)	49 (63.6)	25 (32.5)
25%	22 (12.9)		10 (13.0)	11 (14.3)
>25%-50%	33 (20.2)		11 (14.3)	21 (27.3)
>50%-75%	19 (11.1)		7 (9.1)	12 (15.5)
100%	8 (4.7)			8 (10.4)
Presence of *L. crispatus* at baseline	42 (24.6)	14 (100)	22 (28.6)	6 (7.8)
Mean RA *L. crispatus* if present (SD)	28.2 (35.3)	72.5 (21.6)	7.71 (13.0)	0.21 (0.34)
Presence of *L. iners* at baseline	142 (83.0)	5 (35.7)	77 (100)	58 (75.3)
Mean RA *L. iners* if present (SD)	44.8 (42.4)	1.84 (3.24)	80.3 (23.2)	2.99 (5.11)
Presence of *G. vaginalis* at baseline	129 (75.4)	7 (50.0)	46 (59.7)	73 (94.8)
Mean RA *G. vaginalis* if present (SD)	23.1 (24.8)	20.9 (15.0)	6.81 (10.8)	33.7 (26.7)
Median Shannon diversity index (IQR)	0.94 (0.30 – 1.82)	0.80 (0.33 – 0.98)	0.34 (0.16 – 0.83)	1.84 (1.20 – 2.17)

IQR = Interquartile range; BV = Bacterial vaginosis; RA = Relative abundance; SD = Standard Deviation

* CST-IV combined sub-types A (n=9), B (n=62), and C (n=6). Three observations (n=2 with CST-II, n=1 with CST-V) are not included in this table.

Table 2. Top single nucleotide polymorphisms associated with five vaginal microbiome traits.

	SNP	Chr	Position	Gene	A1/A2	MAF	b	*P*
L. crispatus								
	rs73330467	5	125694420	*LOC101927488-GRAMD2B*	G/T	0.07	11.58*	4.79E-06
L. iners								
	rs527430	1	47918821	*FOXD2-TRABD2B*	A/G	0.06	1.05	6.98E-07
	rs77007265	2	31384829	*GALNT14-CAPN14*	C/A	0.08	-0.93	2.07E-06
	rs17010778	2	31384974	*GALNT14-CAPN14*	C/T	0.06	-1.00	3.60E-06
	rs12221275	10	122972398	*WDR11-FGFR2*	A/C	0.01	2.22	6.95E-06
G. vaginalis								
	rs1229660	7	26437429	*SNX10-LOC441204*	C/T	0.10	-0.99	4.65E-06
	rs10414170	19	57246309	*ZNF835-ZIM2-AS1*	C/A	0.30	-0.62	6.56E-06
Shannon Diversity Index								
	rs7632135	3	154455745	*GPR149-MME*	G/A	0.24	-0.59	4.37E-06
	rs3097137	5	73330562	*ARHGEF28-LINC01335*	A/G	0.20	0.56	4.25E-06
	rs112627544	7	1929410	***MAD1L1***	T/G	0.42	-0.47	9.04E-06
	rs6970796	7	1947895	***MAD1L1***	T/C	0.36	-0.50	2.25E-06
	rs56952063	14	95107145	***SERPINA13P***	C/T	0.08	-0.90	9.65E-06
	rs972741	16	25468083	*ZKSCAN2-HS3ST4*	A/G	0.19	0.66	8.52E-07
Community State Type								
	rs419816	5	52571758	*LOC257396-FST*	C/T	0.13	0.45	9.99E-06
	rs1929353	9	3759975	*RFX3-AS1-GLIS3*	G/A	0.24	0.35	9.51E-06
	rs2302902	12	96617304	***ELK3***	T/C	0.20	0.41	3.09E-06

SNP, Single nucleotide polymorphism; Chr, Chromosome; A1/A2, allele 1 / allele 2; MAF, Minor allele frequency.

SNPs with a $P < 1 \acute{} 10^{-5}$ for each trait are included in the table. Gene name is in boldface if the SNP is located within the gene.

* Reported as odds ratio

Table 3. Results of pathway analysis: Top KEGG and Reactome pathways associated with vaginal microbiome traits.

Trait	Pathway	Observed/Total Genes	*P*	Adjusted *P**
L. crispatus				
	Metabolism of amino acids and derivatives	29/370	7.27E-06	1.10E-02
	GPCR ligand binding	32/457	2.46E-05	1.37E-02
	PCP/CE pathway	11/92	1.50E-04	3.97E-02
	Beta-catenin independent WNT signaling	14/145	2.04E-04	4.72E-02
	Neutrophil degranulation	30/479	3.12E-04	5.31E-02
L. iners				
	Toll-like receptor cascades	17/155	8.41E-06	9.76E-03
	MyD88:MAL(TIRAP) cascade initiated on plasma membrane	13/95	8.82E-06	9.76E-03
	Toll like receptor TLR6:TLR2 cascade	13/95	8.82E-06	9.76E-03
	Toll like receptor TLR1:TLR2 cascade	13/98	1.25E-05	9.76E-03
	Toll like receptor 2 (TLR2) cascade	13/98	1.25E-05	9.76E-03
G. vaginalis				
	Signaling by GPCR	66/1162	5.35E-07	1.46E-03
	GPCR downstream signaling	63/1100	7.45E-07	1.46E-03
	Neutrophil degranulation	29/479	3.11E-04	2.44E-01
	G alpha (s) signaling events	31/536	4.29E-04	2.80E-01
	Formation of the cornified envelope	12/129	5.23E-04	2.93E-01
Shannon Diversity Index				
	Class B/2 (secretin family receptors)	12/95	1.89E-05	3.71E-02
	RNA polymerase II transcription	62/1292	6.53E-05	5.12E-02
	Gene expression (transcription)	66/1430	1.19E-04	6.52E-02
	Neuronal system	24/368	1.99E-04	6.52E-02
	Cleavage of growing transcript in the termination region	8/67	6.80E-04	9.52E-02
Community State Type				
	Human T-cell leukemia virus 1 infection	20/255	8.77E-05	9.30E-03
	Class B/2 (secretin family receptors)	11/95	1.23E-04	1.06E-02
	GPCR ligand binding	29/457	1.25E-04	1.06E-02
	Post-translational modification: synthesis of GPI-anchored proteins	10/92	4.17E-04	1.89E-02
	Transmission across chemical synapses	17/227	4.95E-04	2.12E-02

*P-values are Benjamini-Hochberg adjusted

Table 4. Results of pleiotropy analysis: Top phenotypes associated with vaginal microbiome traits.

Trait	Phenotype	Observed/Total Genes	*P*	Adjusted *P**
L. crispatus				
	Bacterial infections	23/260	9.25E-06	1.10E-02
	Meckel syndrome (MKS)	17/164	1.78E-05	1.37E-02
	Hypertension, pulmonary	15/137	2.88E-05	1.37E-02
	Varicose veins	11/79	3.64E-05	1.45E-02
	Gram-positive bacterial infections	17/181	6.29E-05	2.15E-02
L. iners				
	Autosomal dominant inheritance	72/1403	3.17E-05	1.78E-02
	Growth abnormality	88/1822	3.89E-05	1.91E-02
	Abnormality of the cardiovascular system	86/1796	6.49E-05	2.83E-02
	Aplasia/hypoplasia involving the vertebral column	8/55	3.08E-04	8.05E-02
	Abnormality of hair density	20/267	3.35E-04	8.05E-02
G. vaginalis				
	Pancreatic neoplasm	11/88	6.51E-05	8.52E-02
	Carcinoma, pancreatic ductal	13/139	2.95E-04	2.44E-01
	Pancreatic neoplasms	19/281	9.08E-04	3.56E-01
	Neoplasm of the endocrine system	9/89	1.42E-03	4.11E-01
	Thyroid neoplasms	13/165	1.48E-03	4.11E-01
Shannon Diversity Index				
	Abnormality of the female genitalia	33/505	1.31E-05	3.71E-02
	Abnormality of female internal genitalia	29/444	4.38E-05	5.12E-02
	Abnormal internal genitalia	30/473	5.57E-05	5.12E-02
	Abnormality of the uterus	15/176	1.92E-04	6.52E-02
	Abnormality of the ovary	16/196	1.96E-04	6.52E-02
Community State Type				
	Serotonin syndrome	23/134	1.18E-11	4.39E-08
	Abnormality of the integument	99/1920	5.16E-08	1.01E-04
	Autosomal dominant inheritance	74/1403	1.33E-06	1.50E-03
	Prader-Willi syndrome	20/195	1.70E-06	1.50E-03
	Abnormality of the skin	80/1572	1.92E-06	1.50E-03

**P*-values are Benjamini-Hochberg adjusted

Supplemental Table 1. Excel File. Details of Candidate Gene Analysis Findings

Figures

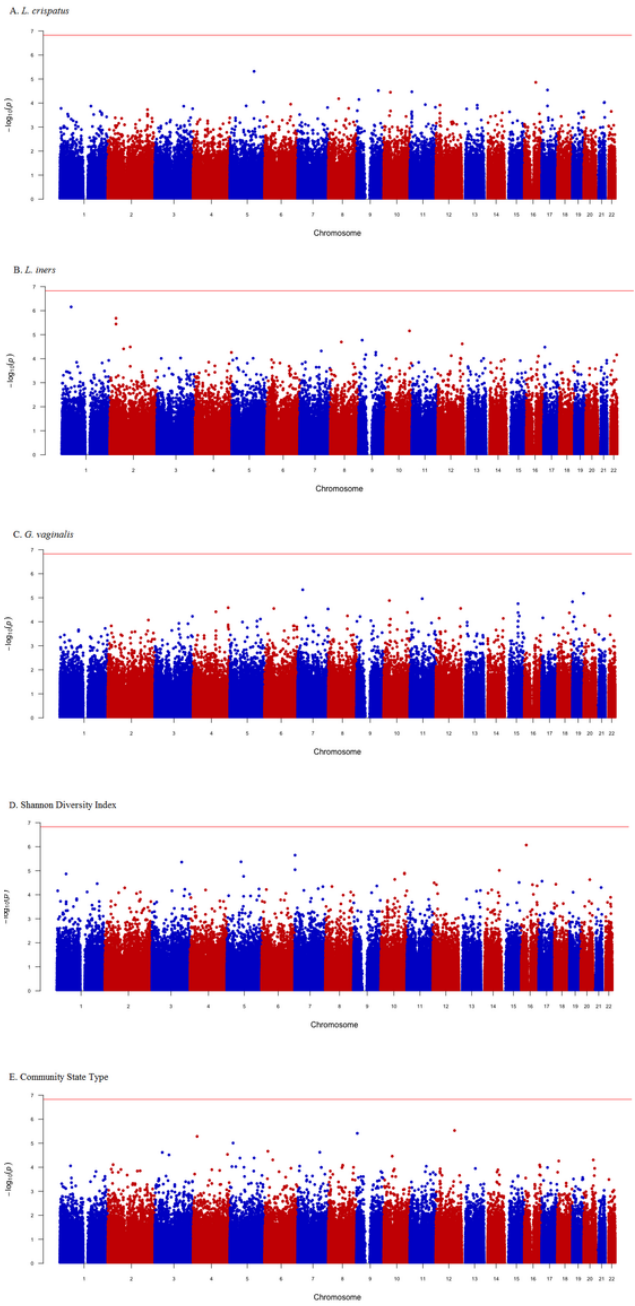

Figure 1

Manhattan plots for single nucleotide polymorphisms associated with vaginal microbiome traits. Legend: The x-axis corresponds to the genomic position, and the y-axis shows the −log10 of the P-value. The horizontal dashed line corresponds to the genome-wide significance threshold.

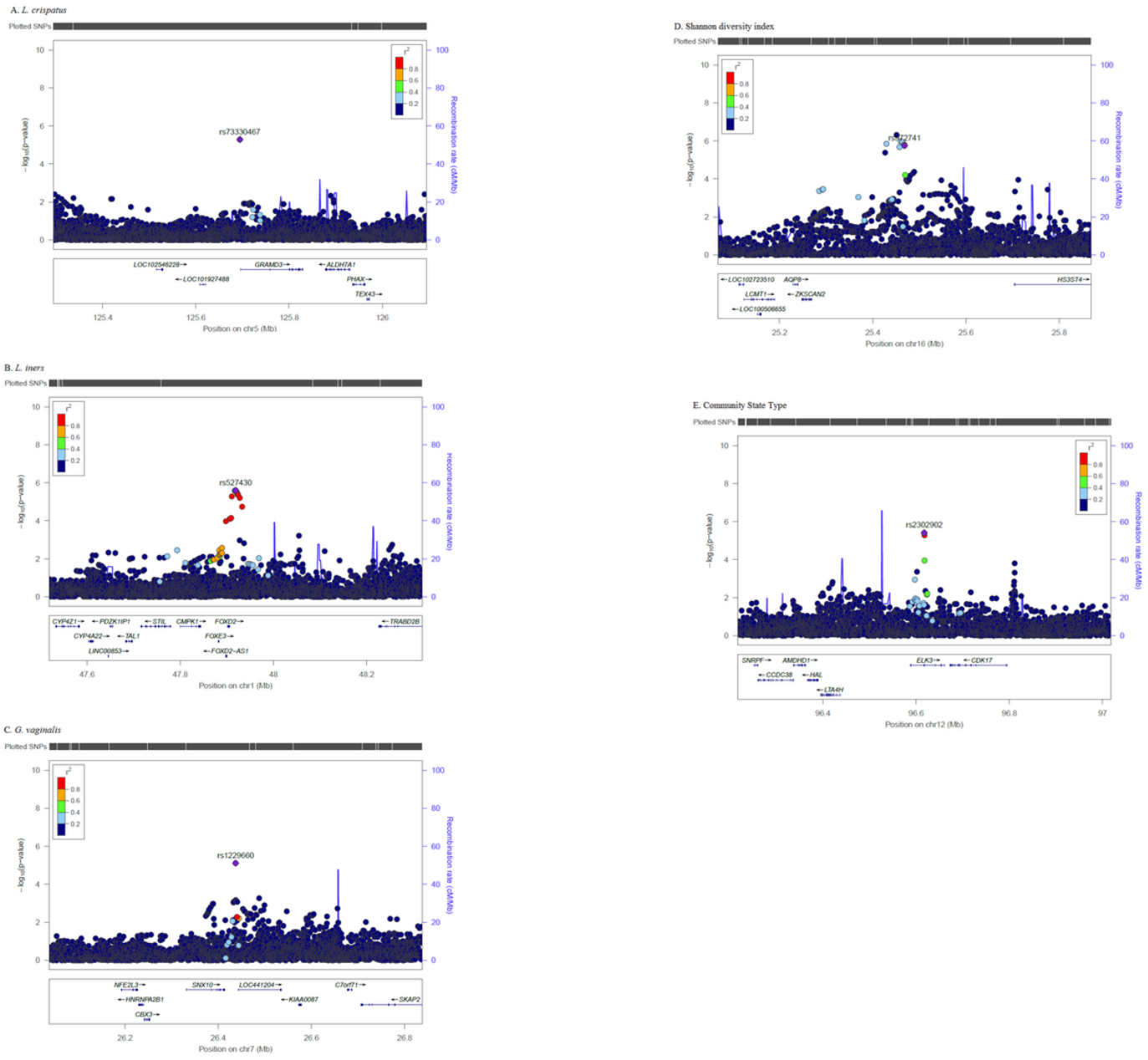

Figure 2

Regional association plots of the top genomic loci associated with vaginal microbiome traits. Legend: P values (−log10) of the GWAS (solid circle) on the y-axis are plotted against the genomic positions of each SNP on the x-axis for each microbiome trait. Genes in the region are shown below. The linkage disequilibrium (LD) values (r2) between the lead SNP and the other SNPs are indicated in different colors.

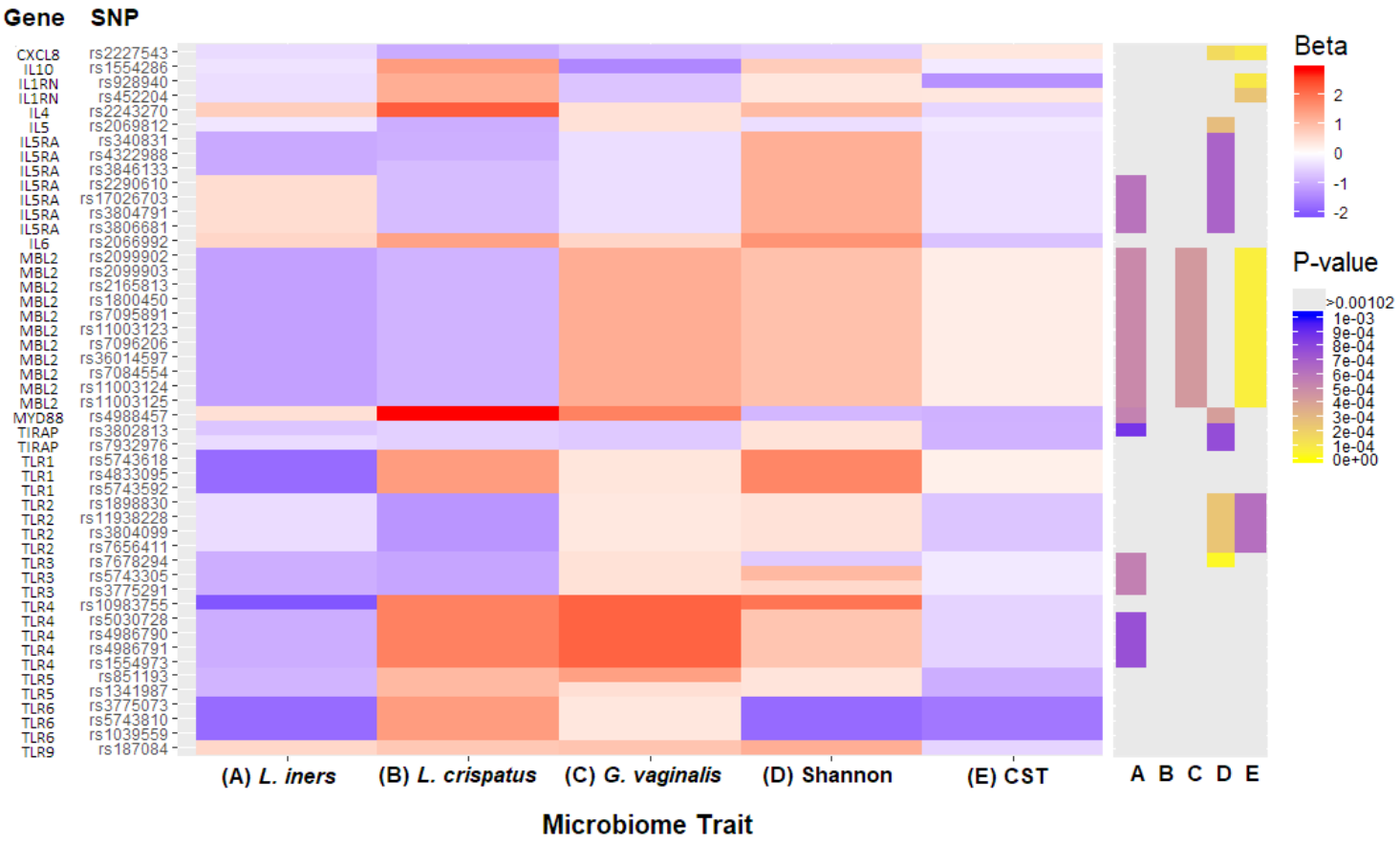

Figure 3

Heatmap summarizing beta coefficients for the most significant SNPs within 100kbp of previoulsy reported SNPs. Legend: The heatmap on the left represents the direction and magnitude of the coefficients (beta) for the SNPs within 100kb associated with five vaginal microbiome traits: (A) relative abundance L. iners, (B) presence of L. crispatus, (C) relative abundance G. vaginalis, (D) Shannon diversity index, (E) Community State Type (CST), with CST-IV as the reference. Negative coefficients are shaded in blue and positive coefficients are shaded in red, with deeper intensity representing the magnitude of the coefficient. The heatmap on the right represents the P-value of the corresponding coefficients for SNPs and vaginal microbiome traits (A-E). All associations shown have p-value<0.05. The Bonferroni cutoff for significance is <0.00102, and these p-values are shaded yellow to purple. P-values >0.00102-0.05 are shaded grey.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementalFigure1BLinersqq.png
- SupplementalFigure1ECSTqq.png
- SupplementalFigure2ALcrispcondzoom.png
- SupplementalFigure2CGvagcondzoom.png
- SupplementalFigure2DShannoncondzoom.png
- SupplementalFigure2ECSTcondzoom.png
- SupplementalFigure2BLinerscondzoom.png

- SupplTableCandidateDetailsFinalComplete.xlsx
- GWASN171finalanalyticdataset.csv
- SupplementalFigure1CGvagqq.png
- SupplementalFigure1DShannonqq.png
- SupplementalFigure1ALcrispqq.png