Different and diverse anaerobic microbiota were seen in women living with HIV with unsuppressed HIV viral load and in women with recurrent bacterial vaginosis: a cohort study

Chelsea Elwood [a,b], Arianne Y. K. Albert [b], Elisabeth McClymont [a,b], Emily Wagner [b], Daljeet Mahal [b], Keshini Devakandan [a,b], Bonnie L. Quiqley [c,1], Zahra Pakzad [b], Mark H. Yudin [d], Janet E. Hill [c], Deborah Money [a,b], and the VOGUE Research Group

[a] Department of Obstetrics and Gynecology, University of British Columbia, Vancouver, BC, Canada

[b] Women's Health Research Institute, Vancouver, BC, Canada

[c] Department of Veterinary Microbiology, University of Saskatchewan, Saskatoon, SK, Canada

[d] Department of Obstetrics and Gynecology, St. Michael's Hospital, University of Toronto, Toronto, ON, Canada

[1] Current address: GeneCology Research Centre, University of the Sunshine Coast, Queensland, Australia

Corresponding author: D Money, 317 – 2194 Health Sciences Mall, University of British Columbia, Vancouver, BC, Canada, V6T 1Z3. Email: deborah.money@ubc.ca Telephone: 1-604-827-0327.

Abstract

Objective: To compare the vaginal microbiota of women living with HIV (WLWH) to the vaginal microbiota of women with recurrent bacterial vaginosis (BV) and healthy women without HIV to determine if there are differences in the vaginal microbiome, what factors influence these differences, and to characterize HIV clinical parameters including viral load and CD4 count in relation to the vaginal microbiome.

Design: Observational cohort study.

Setting: Canada.

Population: Women aged 18-49 years who were premenopausal and not pregnant were recruited into three cohorts: healthy women, WLWH, and women with recurrent BV.

Methods: Demographic and clinical data were collected via interviews and medical chart reviews. Vaginal swabs were collected for Gram stain assessment and microbiome profiling utilizing the *cpn*60 barcode sequence.

Main Outcome Measures: To compare overall community composition differences, we used compositional data analysis methods, hierarchical clustering, and Kruskal-Wallis tests where appropriate.

Results: Clinical markers such as odour and abnormal discharge, but not irritation, were associated with higher microbial diversity. WLWH with unsuppressed HIV viral loads were more likely than other groups to have non-*Gardnerella* dominated microbiomes. HIV was associated with higher vaginal microbial diversity and this was related to HIV viral load, with unsuppressed women demonstrating significantly higher relative abundance of *Megasphaera genomosp.* 1, Atopobium *vaginae*, and *Clostridiales sp.* (all $p<0.05$) compared to all other groups.

Conclusions: In WLWH, unsuppressed HIV viral loads were associated with a distinct dysbiotic profile consisting of very low levels of *Lactobacillus* and high levels of anaerobes.

Funding: Canadian Institutes of Health Research and Genome British Columbia.

Keywords: Vaginal microbiome, HIV, *cpn*60, bacterial vaginosis, women's health

Tweetable Abstract

Vaginal microbiomes in WLWH with viral load >50 have distinct dysbiotic profiles with high levels of anaerobes.

Introduction

While the healthy vaginal microbiome has been increasingly studied, the microbiome in the context of recurrent bacterial vaginosis (BV) and other conditions including infection with human immunodeficiency virus (HIV) has not been explored to the same extent.[1] Fully understanding the dimensions of dysbiosis in different populations remains elusive. In addition, women living with HIV (WLWH) have higher rates of BV than women without HIV, as diagnosed by Nugent's score and affecting as many as 30% of WLWH.[2, 3] The vaginal microbiome has been identified as having significant clinical implications in the effectiveness of pre-exposure prophylaxis for the prevention of HIV transmission. In particular, women with a perturbed microbiome (i.e., BV) and the presence of *Gardnerella vaginalis* experienced decreased effectiveness of vaginal tenofovir for the prevention of HIV transmission, thought to be due to direct degradation of tenofovir by non-*Lactobacillus* vaginal bacteria.[4] In addition to the role the vaginal microbiome plays in the acquisition of HIV,[5, 6] HIV itself has been found to play a role in increased lower genital tract infections.[1]

The majority of studies exploring the vaginal microbiome have used 16S rRNA gene PCR and sequencing, and few of these studies have included direct comparisons to a population of women without HIV.[7, 8] Our previous research has increased the understanding of the vaginal microbiome in healthy, reproductive aged women in North America and has resolved previously unrecognized "community state types" (CSTs) using profiling based on chaperonin-60 (*cpn*60) barcode sequences.[9] Utilization of *cpn*60 sequences allows for better species and strain level differentiation of microorganisms.[10] Given the importance of the vaginal microbiome in reproductive health and its role in

HIV transmission and other health outcomes in WLWH, determination of the composition of the microbiome in dysbiosis is important. In the present study we sought to compare the vaginal microbiota of women with recurrent BV and WLWH to healthy women to determine if there are differences between these women, what potential factors influence these differences, and in WLWH to characterize HIV clinical parameters including viral load and CD4 count in relation to the vaginal microbiome.

Methods

Participants enrolled in the HIV-negative cohort were volunteer women recruited from the greater Vancouver area of BC, Canada, without symptoms of recurrent BV as previously described.[9] Women in the recurrent BV cohort were recruited, by convenience samples, from two referral Reproductive Infectious Diseases clinics (Vancouver and Toronto) if they had experienced symptoms in the previous two weeks, had an intermediate to high Nugent's score by clinical swab at the visit, and also had at least four self-identified episodes of vulvovaginitis in the past 12 months. WLWH were recruited from a comprehensive women's HIV clinic where women access both HIV and gynecologic care (the Oak Tree Clinic, BC Women's Hospital). They had to be between the ages of 18-49 years, premenopausal, and not pregnant. Patients were not involved in the development of this study.

Demographic and clinical data were collected via interview and by reviewing medical charts. Vaginal swab samples were collected from the lateral vaginal wall and posterior fornix during a clinically indicated speculum examination and were sent for Gram stain assessment and microbiome profiling.[11] Swab collection, sample storage, and

119 Nugent's scoring were conducted as previously described.[9] Total DNA extraction from
120 vaginal swabs, generation of *cpn*60 PCR amplicon libraries and sequencing were carried
121 out as previously described.[9] The *cpn*60 barcode sequence has been demonstrated to
122 provide species or even strain level resolution of bacteria,[10, 12] including those comprising
123 the vaginal microbiota.[11, 13-16]

124 Amplification primer sequences were removed using cutadapt, followed by
125 quality trimming with trimmomatic (quality cut-off 30, minimum length 150). Quality
126 trimmed reads were loaded into QIIME2[17] for sequence variant calling and read
127 frequency calculation with DADA2[18] and a truncation length of 250. For taxonomic
128 identification, variant sequences were compared to the cpnDB_nr reference database
129 (version 20190305, downloaded from www.cpndb.ca) using watered-BLAST.[16] Sequence
130 variants having the same best database reference were grouped together into nearest
131 neighbor "species" by summing their total read counts within samples. These nearest
132 neighbour taxonomic labels are used in the subsequent analyses.

133 Detection of Mollicutes (Mycoplasma and/or Ureaplasma) was accomplished
134 using a family-specific, semi-nested PCR targeting the 16S rRNA gene.[19] *Ureaplasma*
135 spp. were detected with a PCR assay targeting the multiple-banded antigen, in which *U.*
136 *parvum* and *U. urealyticum* can be differentiated based on PCR product size.[20]

137 A number of statistical tests were employed to provide a comprehensive
138 understanding of the diversity and composition of the vaginal microbiome in four groups:
139 healthy women, women with recurrent BV, WLWH who have suppressed HIV viral
140 loads (VL<50 copies/mL), WLWH who have unsuppressed HIV viral loads (VL>50
141 copies/mL). Additional variables considered were: use of any antibiotics in the past three

142 months, age (years), body mass index (BMI, kg/m^2), current tobacco use (yes/no),
143 abnormal odour in the past two weeks (yes/no), abnormal irritation in the past two weeks
144 (yes/no), abnormal discharge in the past two weeks (yes/no), and number of sexual
145 partners in the last two months (0, 1, 2, or more).

146 We compared Shannon's diversity using Kruskal-Wallis tests. To compare overall
147 community composition differences, we used compositional data analysis methods,[21, 22]
148 then visualized communities with PCA and hierarchical clustering. Community state
149 types (CST)[9, 23, 24] were identified from these analyses using cluster fit statistics.[25] We
150 compared the relative abundance (center-log ratio transformed) of a subset of taxa,
151 among the clinical variables described above using linear regressions assessing all
152 variables together in multivariable models. Although there was some deviation from the
153 assumption of normality of residuals, the deviation was not extreme in any case, and the
154 overall sample size is large justifying this approach. Confidence intervals for the
155 estimated regression coefficients were calculated using a BCa bootstrap technique with
156 resampling of cases and 10,000 bootstrap iterations. Post-hoc pairwise tests with p-value
157 correction were performed to determine which groups were different. All statistical
158 analyses were carried out using R v3.3.1.[26]

159

160 **Results**

161 The cohort of 306 reproductive aged women without a history of BV or HIV was
162 used for comparison in this study.[9] A total of 39 women with recurrent BV and 54
163 WLWH and were compared to the cohort of healthy women. Of the women with
164 recurrent BV, 12 samples (31%) came from the clinic in Toronto, Ontario while all other

samples in the study were from women in the Greater Vancouver, British Columbia region. There were a total of 399 samples with adequate sequence reads and associated clinical data. WLWH had a greater average age (38.1 ± 5.8 and 34.0 ± 6.8 for those with VL <50 and VL>50, respectively, vs. 30.2 ± 7.6 in healthy women) and had a greater average BMI (26.1 ± 6.0 and 29.1 ± 7.2 for those with VL<50 and VL>50, respectively, vs. 23.9 ± 5.3 in healthy women) (Table 1). Healthy women and women with recurrent BV were more likely to be White (64.4% and 61.5%, respectively) while there was a greater ethnic mix among the WLWH with 35% of those with VL<50 and VL>50 being non-White ethnicities. Women with recurrent BV were most likely and WLWH were least likely to have completed some post-secondary education. There was no difference among the groups in the number of sexual partners in the previous two months. In the cohort of WLWH, 94.4% of participants were on combination antiretroviral therapy with 63% having an undetectable HIV viral load and 72.2% having a CD4 count ≥350 cells/mm^3 (Table S1).

There was a difference in microbial diversity among the groups (p=0.001) (Table S1). There was no significant difference between the BMI categories (p=0.05) but there was a trend for obese women to have slightly higher diversity (mean Shannon index = 1.08 (±0.85) compared to 0.77 (±0.65) for normal weight). There was also no significant difference in diversity between those with irritation in the previous two weeks, any antibiotic use in the past three months, or number of sexual partners in the previous two months (all p > 0.05). There was significantly higher microbial diversity among those with abnormal odour in the previous two weeks (mean Shannon index = 1.2 (±0.72)) compared to no abnormal odour (mean Shannon index = 0.78 (±0.69); p = 0.001), and

among those with abnormal discharge in the previous two weeks (mean Shannon index = 1.13 (±0.73)) compared to those with no abnormal discharge (mean Shannon index = 0.76 (±0.68); p = 0.0001).

Overall community composition (beta-diversity) was visualized using PCA. The first principal component (PC1) describes the differences in relative abundance of *Lactobacillus crispatus* vs. the relative abundance of *Megasphaera* genomosp 1, *Prevotella timonensis, Atopobium vaginae,* an unidentified *Clostridiales* sp., *Dialister micraerophilus*, and others (Figure 1; Table S2). PC2 describes differences in the relative abundance of *Lactobacillus iners* and *L. jensenii* relative to all other taxa. Figure 1A demonstrates how these PC scores relate to the participant groups, and suggests a separation of the groups along these axes with WLWH, particularly those with VL>50, having higher PC1 scores on average than the other groups (i.e., lower relative abundance of *L. crispatus* with higher relative abundances of anaerobic taxa as described above).

Clustering statistics gave reasonable support for four clusters of microbiome profiles (Table S3). Visual examination of the heatmap and dendrograms supported four broad classes, some with subclusters, for a total of ten identifiably different groupings of samples. The four broad classes divide: 1) those with higher relative abundance of mostly *L. crispatus* (corresponding to previously described CST I[9, 23, 24]) and with low amounts of other taxa, 2) those with higher relative abundance of *Lactobacillus* spp. (Mixed lactobacilli) with some other varied taxa present, 3) those with high relative abundance of *Gardnerella* spp., but with lower abundances of some other dysbiotic taxa, most notably *Clostridiales*, *Porphyromonas uenonis*, and some *Prevotella* (CST IVC), and 4) those with a more uniform distribution of traditionally BV-associated taxa (*Megasphaera*

genomosp 1, *Prevotella timonensis, Atopobium vaginae, Dialister micraerophilus, Porphyromonas uenonis,* an unidentified *Clostridiales* sp.*,* several *Prevotella* spp., and several *Gardnerella* spp. (CST IVD).

The subclusters divide the Mixed lactobacilli group into five CST (IB, IC, II, III, and IVA) dominated by combinations of *L. iners, crispatus, jensenii*, *gasseri*, *coleohominis*, and *L. reuteri,* and smaller relative amounts of unusually seen taxa including *Pseudomonas* spp., *Paraburkholderia* spp., and *Bifidobacterium breve* (Figure 1C). The subclusters also divide the *Gardnerella* dominated IVC group into three: IVC.1 with samples that have higher relative abundance of *Gardnerella vaginalis,* IVC.2 with *Gardnerella vaginalis* and *Gardnerella swidsinskii*, and IVC.3 with *Gardnerella vaginalis, Gardnerella leopoldii,* and *Gardnerella piotii*. There was a significant difference in the proportion of samples from each group of participants in each CST ($p<0.0001$; Figure 2).

Eighty-one percent of participants in the healthy cohort had sequence profiles that were clustered within one of the *Lactobacillus* dominated CST, compared to 48.7% of the participants with recurrent BV, 67.7% of the WLWH with VL<50, and 35% of the WLWH with VL>50. When we restricted the data to samples that had a diagnosis of BV via Nugent's scoring, and combined the two WLWH groups to compare their community compositions, there was a trend for differences in proportion of samples in each CST among the groups ($p = 0.06$) with a very small proportion of samples from women with recurrent BV in CST IVD (11% compared to 50% in WLWH, and 25% in the healthy cohort).

We compared the clr-relative abundance of the eight non-*Lactobacillus* taxa that loaded onto the first Principal Component with a score of ≥0.2 (Figure 3; Table S4). WLWH with VL>50 had significantly higher relative abundance of *Megasphaera genomosp* 1, *Atopobium vaginae*, and *Clostridiales* sp. compared to women in all other groups, controlling for other variables in the models. There was significantly higher relative abundance of *Dialister micraerophilus* in both HIV groups compared to the healthy cohort, and significantly higher relative abundance of *Prevotella amnii* in the WLWH VL>50 group compared to the healthy group and the recurrent BV group. Finally, there was higher relative abundance of *Gardnerella swidsinskii* in the WLWH VL<50 compared to the healthy group, but no difference among any other groups.

Among the other variables in the models, there was significantly higher relative abundance of *Prevotella timonesis* and *Dialister micraerophilus* in samples from current tobacco users compared to non-users (Table S4). Abnormal discharge was associated with a higher relative abundance of *Dialister micraerophilus*, *Prevotella amnii*, and *Gardnerella swidsinskii* compared to those with normal discharge, while abnormal odour was associated with a higher relative abundance of *Atopobium vaginae* compared to those with normal odour. Finally, having had two or more sexual partners in the last two months was associated with higher relative abundance of *Atopobium vaginae* compared to those with no sexual partners.

We also assessed the presence of all Mollicutes in each cohort (Table 1). Significantly more WLWH with VL<50 were positive for Mollicutes than healthy women. The women with recurrent BV and the WLWH with VL>50 also had higher prevalence of *Ureaplasma urealyticum* than the healthy cohort.

Discussion

Main Findings

This is the first study to use the *cpn*60 barcode sequence to determine the composition of the vaginal microbiome in women with recurrent BV and WLWH. Of importance, we have shown that VL>50 was associated with less than 20% of WLWH having a CST dominated by *Lactobacillus,* and 40% having microbiome profiles of CST IVD a highly diverse and dysbiotic state. Women whose VL>50 had higher relative abundances of *Megasphaera genomosp* 1, *Atopobium vaginae*, *Prevotella amnii*, and *Clostridiales* sp. compared to those with VL<50 and those without HIV including women with recurrent BV, suggesting that there are specific anaerobes that are more prevalent in women with unsuppressed viral loads. HIV was associated with a higher relative abundance of *Dialister micraerophilus* compared to the healthy cohort. Additionally, WLWH with VL<50 had a higher relative abundance of *Gardnerella swidsinskii* compared to the healthy group.

Prevotella timonesis and *Dialister micraerophilus* presence correlated with tobacco use, independent of HIV status, and *Dialister micraerophilus* was also seen to correlate with HIV status. The mechanism whereby tobacco use could influence the vaginal microbiome remains unclear but it is a known modifiable risk factor for vaginal dysbiosis.[28] Obligate anaerobes are necessarily more difficult to study in laboratory and their diagnostic cultivation is not routine; it is therefore perhaps not surprising that there is much less knowledge about the role of obligate anaerobes in the vaginal microbiome. New species are continuing to be identified[29] and the roles of anaerobes and their

279 fermentation products in vaginal microbial ecology are being explored, but more work is
280 needed.

281 The vaginal microbiome, and specifically *Gardnerella vaginalis*, has been
282 implicated as a negative influence on the effectiveness of vaginal tenofovir for pre-
283 exposure prophylaxis and the prevention of HIV transmission.[4, 30] Vaginal tenofovir is
284 metabolized by the vaginal microbiome and *Gardnerella vaginalis* was able to alter *in*
285 *vitro* drug levels as well as *in vivo* vaginal and plasma drug levels.[4, 30] In addition, *L.*
286 *crispatus* can alter permeability of cervical mucus to HIV virions, and presence of altered
287 vaginal microbiota and lack of *L. cripatus* has been proposed to increase the risk of HIV
288 and other STI acquisition.[5] *Gardnerella* species differ in sialidase activity, which can
289 alter barrier function of the vaginal and cervical mucosa, with only *Gardnerella piotii*
290 exhibiting sialidase activity.[15] In this study, *Gardnerella piotii* was not important for
291 defining overall community differences, and was not significantly different in average
292 relative abundance among the groups (data not shown). Many other BV-associated
293 bacteria produce hydrolytic enzymes including sialidase and mucinase[31] and it has been
294 suggested that some of the "virulence factors" of *Gardnerella* may have more of a role in
295 initiating dysbiosis than they do in directly contributing to BV symptoms by establishing
296 an environment that supports the growth of other anaerobes and the formation of multi-
297 species biofilms.[32] In future studies, the recent reclassification of *Gardneralla vaginalis*[33]
298 and its potential phenotypic differences may prove important for understanding clinical
299 differences associated with the same Nugent's score.

300 Our data also re-affirm what is known in clinical practice regardless of HIV
301 status, which is that odour and abnormal discharge are associated with diversity,[34] as

302 opposed to irritation which did not correlate with diversity. Antibiotic use within the last
303 three months (largely in the recurrent BV cohort) did not show any significant association
304 with diversity; however, women with BV have significant treatment failure rates up to
305 50%.[35] In this population, antibiotic use was primarily clindamycin or metronidazole,
306 which is used to treat BV but is not resulting in changes in diversity in this study. Results
307 of the principal components analysis suggest that the large variation between the samples
308 occur due to variance in the relative abundances of a handful of taxa. These are
309 associated with differences in BV-associated taxa, recurrent BV status, HIV viral load,
310 and presence of symptoms. If further studies confirm that these variations are due to a
311 small number of taxa and are potentially correlated with clinical variables such as viral
312 load, targeted PCR, which is more readily available than microbiome sequencing, may be
313 of use for clinical diagnostics and more targeted antimicrobial therapy.
314
315 Strengths and Limitations
316 This study is the first of its kind to use a higher resolution *cpn*60 marker to
317 characterize the vaginal microbiome in women with recurrent BV and WLWH with more
318 taxonomic detail than is available using 16S rRNA. It is one of the larger studies in North
319 America to date. A limitation of our study is that it evaluated the vaginal microbiome of
320 patients at a single time point and does not represent the dynamic nature of the
321 microbiome.[36] The small relative sample of WLWH with VL>50 makes it difficult to
322 determine if small effect associations were missed for this group in particular. The
323 observational nature of the study and the potential for missed confounders mean that the
324 conclusions are not definitive but instead point to potential areas for future investigation.

This analysis does however provide important insights into the variability in vaginal dysbiosis and specifically differences between WLWH and women without HIV, but whether these differences correlate with adverse clinical outcomes remains to be seen. In addition, WLWH in this cohort were generally well engaged in care in North America, thus, our findings may not be generalizable to all WLWH globally.

Interpretation

Women with dysbiosis with and without HIV can be well characterized using this highly discerning platform permitting greater capabilities to study changes in the microbiome over time, under different conditions of health and disease, and to monitor treatment effects. Clinical symptoms such as odour and discharge appear to correlate with dysbiosis, thus reaffirming the value of these symptoms in the clinic; however, they are not sufficient for identifying the specific taxa that may be causing symptoms. In our cohort of WLWH, unsuppressed HIV viral load appears to influence the vaginal microbiome with more women having dysbiosis when VL>50 copies/mL, and with differing microbiota composition dependent on viral load suppression. This has not been seen previously, thus clinicians should have a higher index of suspicion for dysbiosis in women with unsuppressed viral loads and differing treatment options may need to be developed in the future depending on viral load suppression. We also noted that smoking, a potentially modifiable behaviour, was associated with specific anaerobes associated with BV. Future studies utilizing the higher resolution *cpn*60 barcode are required to correlate these findings with adverse outcomes given that WLWH are at higher risk of adverse reproductive outcomes such as preterm birth and cervical cancer.

Conclusion

This study demonstrates the high utility of the *cpn*60 barcode in differentiating subgroups of women with dysbiosis including women with recurrent BV and WLWH. WLWH who have unsuppressed HIV viral loads are more likely to demonstrate vaginal dysbiosis than WLWH with suppressed HIV viral loads and women without HIV. Additionally, the composition of dysbiotic vaginal microbiomes in unsuppressed WLWH is different from the composition of dysbiotic vaginal microbiomes in suppressed WLWH, with the vaginal microbiomes of dysbiotic unsuppressed WLWH more likely to contain very low levels of *Lactobacillus* and very high levels of anaerobic bacteria such as *Megasphaera*. These results point to important differences for interpretation of diagnostics for dysbiosis as well as treatment approaches for these varying microbes.

Acknowledgements

The authors wish to thank the women who took the time to participate in this study, as well as the clinicians and research staff at BC Women's Hospital and Health Centre for their important contributions, and Champika Fernando for technical support in the laboratory. The VOGUE Research Group, in alphabetical order, is: Alan Bocking (University of Toronto), Tim Dumonceaux (Agriculture and Agri-Food Canada), Gregory Gloor (University of Western Ontario), Sean Hemmingsen (National Research Council Canada), Janet E. Hill (University of Saskatchewan), Matthew Links (University of Saskatchewan), Deborah Money (University of British Columbia), Kieran O'Doherty (University of Guelph), Gregor Reid (University of Western Ontario), Julianne van

Schalkwyk (University of British Columbia), Patrick Tang (University of British Columbia), Mark H. Yudin (University of Toronto).

Disclosure of Interests

Dr. Chelsea Elwood has served as an advisory board member for Gilead, as a reproductive infectious disease specialist.

Dr. Deborah Money has received past funding for unrelated work from Merck, GSK, Sanofi, and Novartis.

All authors declare funding for this study from the Canadian Institutes of Health Research and Genome British Columbia.

Contribution to Authorship

DMoney conceived and initiated the study and was Principal Investigator. DMoney, JH, AA and CE contributed to study design. DMahal, KD, EW, and MY conducted study visits. JH and BC carried out microbiome profiling. AA and EW managed the database. AA conducted the analyses, produced the tables and figures, and wrote sections of the manuscript. CE, AA, DMahal, DMoney, and JH interpreted the results. CE, EM and DMahal wrote the first draft of the manuscript. CE, EM, DMahal, BC, AA, ZP, JH, and DMoney contributed to several rounds of manuscript drafting and approved the final manuscript.

Details of Ethics Approval

Ethical approval for these cohorts was obtained from the University of British Columbia Children's & Women's Research Ethics Board and the St. Michael's Hospital Research Ethics Board. Certificate numbers for the healthy cohort, the cohort of women living with HIV, the cohort of women with recurrent bacterial vaginosis in Vancouver and the cohort of women with recurrent bacterial vaginosis in Toronto were H10-02535 (approved 9 Dec 2010), H11-00119 (approved 11 Apr 2011), H11-01912 (approved 1 Jun 2012), and 15-161 (approved 27 Oct 2015), respectively.

Funding

This study was funded by the Canadian Institutes of Health Research (F10-02113) and Genome British Columbia (Vaginal Microbiota Project Team [VOGUE]).

References

1. Cu-Uvin S, Hogan JW, Warren D, Klein RS, Peipert J, Schuman P, et al. Prevalence of lower genital tract infections among human immunodeficiency virus (HIV)-seropositive and high-risk HIV-seronegative women. HIV Epidemiology Research Study Group. Clin Infect Dis. 1999 Nov;29(5):1145-50.
2. Atashili J, Poole C, Ndumbe PM, Adimora AA, Smith JS. Bacterial vaginosis and HIV acquisition: a meta-analysis of published studies. AIDS. 2008 Jul 31;22(12):1493-501.
3. Mirmonsef P, Krass L, Landay A, Spear GT. The role of bacterial vaginosis and trichomonas in HIV transmission across the female genital tract. Curr HIV Res. 2012 Apr;10(3):202-10.
4. Klatt NR, Cheu R, Birse K, Zevin AS, Perner M, Noel-Romas L, et al. Vaginal bacteria modify HIV tenofovir microbicide efficacy in African women. Science. 2017 Jun 02;356(6341):938-45.
5. 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.
6. Gosmann C, Anahtar MN, Handley SA, Farcasanu M, Abu-Ali G, Bowman BA, et al. Lactobacillus-Deficient Cervicovaginal Bacterial Communities Are Associated with Increased HIV Acquisition in Young South African Women. Immunity. 2017 Jan 17;46(1):29-37.

7. Hummelen R, Fernandes AD, Macklaim JM, Dickson RJ, Changalucha J, Gloor GB, et al. Deep sequencing of the vaginal microbiota of women with HIV. PLoS One. 2010;5(8):e12078.
8. Benning L, Golub ET, Anastos K, French AL, Cohen M, Gilbert D, et al. Comparison of lower genital tract microbiota in HIV-infected and uninfected women from Rwanda and the US. PLoS One. 2014;9(5):e96844.
9. Albert AY, Chaban B, Wagner EC, Schellenberg JJ, Links MG, van Schalkwyk J, et al. A Study of the Vaginal Microbiome in Healthy Canadian Women Utilizing cpn60-Based Molecular Profiling Reveals Distinct Gardnerella Subgroup Community State Types. PLoS One. 2015;10(8):e0135620.
10. Links MG, Dumonceaux TJ, Hemmingsen SM, Hill JE. The chaperonin-60 universal target is a barcode for bacteria that enables de novo assembly of metagenomic sequence data. PloS One. 2012;7(11):e49755.
11. Chaban B, Links MG, Jayaprakash TP, Wagner EC, Bourque DK, Lohn Z, et al. Characterization of the vaginal microbiota of healthy Canadian women through the menstrual cycle. Microbiome. 2014;2(1):23.
12. Vancuren SJ, Hill JE. Update on cpnDB: a reference database of chaperonin sequences. Database: the journal of biological databases and curation. 2019;2019.
13. Albert AYK, Chaban B, Wagner EC, Schellenberg JJ, Links MG, van Schalkwyk J, et al. A Study of the Vaginal Microbiome in Healthy Canadian Women Utilizing cpn60-Based Molecular Profiling Reveals Distinct Gardnerella Subgroup Community State Types. PloS One. 2015;10(8):e0135620.
14. Paramel Jayaprakash T, Schellenberg JJ, Hill JE. Resolution and characterization of distinct cpn60-based subgroups of Gardnerella vaginalis in the vaginal microbiota. PloS One. 2012;7(8):e43009-e.
15. Schellenberg JJ, Paramel Jayaprakash T, Withana Gamage N, Patterson MH, Vaneechoutte M, Hill JE. Gardnerella vaginalis Subgroups Defined by cpn60 Sequencing and Sialidase Activity in Isolates from Canada, Belgium and Kenya. PLoS One. 2016;11(1):e0146510.
16. Schellenberg J, Links MG, Hill JE, Dumonceaux TJ, Peters GA, Tyler S, et al. Pyrosequencing of the Chaperonin-60 Universal Target as a Tool for Determining Microbial Community Composition. Applied and Environmental Microbiology. 2009;75(9):2889-98.
17. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nature Methods. 2010;7(5):335.
18. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods. 2016;13(7):581-3.
19. Yoshida T, Maeda S-I, Deguchi T, Ishiko H. Phylogeny-Based Rapid Identification of Mycoplasmas and Ureaplasmas from Urethritis Patients. Journal of Clinical Microbiology. 2002;40(1):105-10.
20. Watson HL, Blalock DK, Cassell GH. Variable antigens of Ureaplasma urealyticum containing both serovar-specific and serovar-cross-reactive epitopes. Infection and Immunity. 1990;58(11):3679-88.

21. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome Datasets Are Compositional: And This Is Not Optional. Frontiers in Microbiology. 2017;8.
22. Mandal S, Van Treuren W, White RA, Eggesbø M, Knight R, Peddada SD. Analysis of composition of microbiomes: a novel method for studying microbial composition. Microbial Ecology in Health and Disease. 2015;26:27663-7.
23. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SSK, McCulle SL, et al. Vaginal microbiome of reproductive-age women. Proceedings of the National Academy of Sciences of the United States of America. 2011;108 Suppl 1(Suppl 1):4680-7.
24. Gajer P, Brotman RM, Bai G, Sakamoto J, Schütte UME, Zhong X, et al. Temporal dynamics of the human vaginal microbiota. Science Translational Medicine. 2012;4(132):132ra52-ra52.
25. Hennig C. fpc: Flexible Procedures for Clustering; 2018.
26. Team RC. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing
; 2018.
27. Shokoohi M, Bauer GR, Kaida A, Lacombe-Duncan A, Kazemi M, Gagnier B, et al. Substance use patterns among women living with HIV compared with the general female population of Canada. Drug and Alcohol Dependence. 2018;191:70-7.
28. Nelson TM, Borgogna JC, Michalek RD, Roberts DW, Rath JM, Glover ED, et al. Cigarette smoking is associated with an altered vaginal tract metabolomic profile. Scientific Reports. 2018;8(1):852-13.
29. Diop K, Diop A, Michelle C, Richez M, Rathored J, Bretelle F, et al. Description of three new Peptoniphilus species cultured in the vaginal fluid of a woman diagnosed with bacterial vaginosis: Peptoniphilus pacaensis sp. nov., Peptoniphilus raoultii sp. nov., and Peptoniphilus vaginalis sp. nov. MicrobiologyOpen. 2019;8(3):e00661-n/a.
30. Sharon L. Hillier LAM, Katherine Bunge, Michele Austin, Bernard J. Moncla, Charlene S. Dezzutti, Jill Schwartz, Mark A. Marzinke, Craig Hendrix, Lisa C. Rohan. Impact of Vaginal Microbiota on Genital Tissue and Plasma Concentrations of Tenofovir. Conference on Retroviruses and Opportunistic Infections Seattle, WA; 2017.
31. Wiggins R, Hicks SJ, Soothill PW, Millar MR, Corfield AP. Mucinases and sialidases: their role in the pathogenesis of sexually transmitted infections in the female genital tract. Sexually Transmitted Infections. 2001;77(6):402-8.
32. Hardy L, Cerca N, Jespers V, Vaneechoutte M, Crucitti T. Bacterial biofilms in the vagina. Research in Microbiology. 2017;168(9-10):865-74.
33. Vaneechoutte M, Guschin A, Van Simaey L, Gansemans Y, Van Nieuwerburgh F, Cools P. Emended description of Gardnerella vaginalis and description of Gardnerella leopoldii sp. nov., Gardnerella piotii sp. nov. and Gardnerella swidsinskii sp. nov., with delineation of 13 genomic species within the genus Gardnerella. International journal of systematic and evolutionary microbiology. 2019;69(3):679-87.
34. Livengood CH. Bacterial vaginosis: an overview for 2009. Reviews in obstetrics & gynecology. 2009;2(1):28.
35. Bradshaw CS, Morton AN, Hocking J, Garland SM, Morris MB, Moss LM, et al. High recurrence rates of bacterial vaginosis over the course of 12 months after oral metronidazole therapy and factors associated with recurrence. J Infect Dis. 2006 Jun 1;193(11):1478-86.

36. Mehta SD, Donovan B, Weber KM, Cohen M, Ravel J, Gajer P, et al. The vaginal microbiota over an 8- to 10-year period in a cohort of HIV-infected and HIV-uninfected women. PLoS One. 2015;10(2):e0116894.
37. Fernandes AD, Macklaim JM, Linn TG, Reid G, Gloor GB. ANOVA-like differential expression (ALDEx) analysis for mixed population RNA-Seq. PloS one. 2013;8(7):e67019.
38. Fernandes AD, Reid JN, Macklaim JM, McMurrough TA, Edgell DR, Gloor GB. Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. Microbiome. 2014;2(1):15-.
39. Quinn TP, Richardson MF, Lovell D, Crowley TM. propr: An R-package for Identifying Proportionally Abundant Features Using Compositional Data Analysis. Scientific reports. 2017;7(1):16252-9.
40. Quinn TP, Erb I, Richardson MF, Crowley TM. Understanding sequencing data as compositions: an outlook and review. Bioinformatics. 2018;34(16):2870-8.
41. Jari Oksanen FGB, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs and Helene Wagner vegan: Community Ecology Package; 2019.

Table/Figure Caption List

Table 1: Summary of demographic and clinical variables by cohort and overall.

Figure 1: A and B) Principal components results. Raw read counts were summed by nearest neighbour taxon label from cpnDB with at least 55% identity. Center-log ratio transformation was performed using ALDEx::*aldex.clr*[37, 38]and then combined using propr::*aldex2propr.*[39, 40] 48 nearest neighbour taxa with non-zero read counts in at least 5% of the samples were retained for further analysis. PCA was performed on the Euclidean distance matrix. The first three principal components explained 40% of the variance in relative abundance: (24.2%, 9.2%, and 6.6%). PERMANOVA (vegan::*adonis*)[41] R^2= 0.05, p=0.001 for distance among groups for overall community composition. A) Points are samples and the distance between them is proportional to their compositional distance. Grey = healthy cohort, green = recurrent BV, orange = WLWH

VL<50, and magenta = WLWH VL>50. The circles encompass 1SD of the samples in each group. B) Principal component loadings for the nearest neighbour taxa. Distance and direction from the origin are proportional to the SD of the taxon, distance between points is inversely proportional to their compositional association. Points close together tend to vary together in relative abundance across samples. E.g. *Megasphaera genomesp*1, and *Prevotella timonensis* tend to both be at high relative abundance, or both at low relative abundance across samples. Viewing A and B together suggests that samples from WLWH with VL>50 have on average higher relative abundance of the taxa to the right in panel B.
C) Heatmap of clr-transformed relative abundance of the 48 nearest neighbour taxa with Ward hierarchical clustering on the Euclidean distance matrix. Shown are 4 clusters with 10 subclusters.

Figure 2: Percent of samples from each group in each of the four broad community state types (CST). CST 1 contains samples dominated by *Lactobacillus crispatus*, CST IVC samples have high relative abundance of *Gardnerella* spp, and CST IVD samples with a more uniform distribution of traditionally BV-associated taxa (*Megasphaera* genomosp 1, *Prevotella timonensis, Atopobium vaginae, Dialister micraerophilus, Porphyromonas uenonis,* an unidentified *Clostridiales* sp.*,* several *Prevotella* spp., and several *Gardnerella* spp.

Figure 3: Boxplots of clr-transformed relative abundance of eight taxa with PC1 loadings ≥0.2. The horizontal lines indicate the medians, the boxes extend to the interquartile

ranges, the whiskers extend to 1.5 times the IQR, and the points indicate outliers. Letters above boxes show significant pairwise comparisons by post-hoc tests with p-value correction. Groups sharing the same letter are not significantly different ($p>0.05$).

Table S1: Summary of additional demographic and clinical variables by cohort and overall.

Table S2: Principal component score loadings for the 48 taxa included in the analysis.

Table S3: Cluster statistics.

Table S4: Comparisons of clr-transformed relative abundance of selected taxa by clinical variables.

Table 1. Summary of demographic and clinical variables by cohort and overall.

	Total No. 399	Healthy No. 306	Recurrent BV No. 39	WLWH <50 VL No. 34	WLWH >50 VL No. 20	P-value
Age (years)						
Mean (SD)	31.2 (±7.8)	30.2 (±7.6)	32.1 (±7.3)	38.1 (±5.8)	34.0 (±6.8)	< 0.0001
BMI						
Mean (SD)	24.3 (±5.5)	23.9 (±5.3)	22.8 (±3.9)	26.1 (±6.0)	29.1 (±7.2)	< 0.0001
Missing	4 (1.0%)	3 (1.0%)	1 (2.6%)	0 (0%)	0 (0%)	
Any antibiotic use last 3 months						
yes	52 (13.0%)	18 (5.9%)	20 (51.3%)	7 (20.6%)	7 (35.0%)	< 0.0001
no	347 (87.0%)	288 (94.1%)	19 (48.7%)	27 (79.4%)	13 (65.0%)	
Number of sexual partners in the last 2 months						
0	98 (24.6%)	75 (24.5%)	7 (17.9%)	12 (35.3%)	4 (20.0%)	0.23
1	282 (70.7%)	220 (71.9%)	26 (66.7%)	20 (58.8%)	16 (80.0%)	
2 or more	15 (3.8%)	10 (3.3%)	4 (10.3%)	1 (2.9%)	0 (0.0%)	
Missing	4 (1.0%)	1 (0.3%)	2 (5.1%)	1 (2.9%)	0 (0.0%)	
Discharge past 2 weeks						
no	331 (83.0%)	279 (91.2%)	7 (17.9%)	30 (88.2%)	15 (75.0%)	< 0.0001
yes	62 (15.5%)	22 (7.2%)	31 (79.5%)	4 (11.8%)	5 (25.0%)	
Missing	6 (1.5%)	5 (1.6%)	1 (2.6%)	0 (0.0%)	0 (0.0%)	
Irritation past 2 weeks						
no	330 (82.7%)	272 (88.9%)	9 (23.1%)	32 (94.1%)	17 (85.0%)	< 0.0001
yes	63 (15.8%)	29 (9.5%)	29 (74.4%)	2 (5.9%)	3 (15.0%)	
Missing	6 (1.5%)	5 (1.6%)	1 (2.6%)	0 (0.0%)	0 (0.0%)	
Odour past 2 weeks						
no	360 (90.2%)	296 (96.7%)	18 (46.2%)	30 (88.2%)	16 (80.0%)	< 0.0001
yes	33 (8.3%)	5 (1.6%)	20 (51.3%)	4 (11.8%)	4 (20.0%)	
Missing	6 (1.5%)	5 (1.6%)	1 (2.6%)	0 (0.0%)	0 (0.0%)	

Table 1. Summary of demographic and clinical variables by cohort and overall.

	Total No. 399	Healthy No. 306	Recurrent BV No. 39	WLWH <50 VL No. 34	WLWH >50 VL No. 20	P-value
Any symptoms past 2 weeks						
no	303 (75.9%)	260 (85.0%)	1 (2.6%)	28 (82.4%)	14 (70.0%)	< 0.0001
yes	91 (22.8%)	41 (13.4%)	38 (97.4%)	6 (17.6%)	6 (30.0%)	
Missing	5 (1.3%)	5 (1.6%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	
Nugent category						
Not consistent with BV	289 (72.4%)	247 (80.7%)	10 (25.6%)	25 (73.5%)	7 (35.0%)	< 0.0001
Intermediate	42 (10.5%)	24 (7.8%)	13 (33.3%)	2 (5.9%)	3 (15.0%)	
Consistent with BV	64 (16.0%)	32 (10.5%)	15 (38.5%)	7 (20.6%)	10 (50.0%)	
Missing	4 (1.0%)	3 (1.0%)	1 (2.6%)	0 (0.0%)	0 (0.0%)	
Mollicutes						
Negative	109 (27.3%)	92 (30.1%)	11 (28.2%)	3 (8.8%)	3 (15.0%)	0.028
Positive	290 (72.7%)	214 (69.9%)	28 (71.8%)	31 (91.2%)	17 (85.0%)	
Ureaplasma						
Negative	190 (47.6%)	159 (52.0%)	11 (28.2%)	16 (47.1%)	4 (20.0%)	0.0006
Parvum	167 (41.9%)	125 (40.8%)	18 (46.2%)	13 (38.2%)	11 (55.0%)	
Parvum & Urealyticum	2 (0.5%)	1 (0.3%)	1 (2.6%)	0 (0.0%)	0 (0.0%)	
Urealyticum	40 (10.0%)	21 (6.9%)	9 (23.1%)	5 (14.7%)	5 (25.0%)	
Shannon's Diversity Index (H)						
Mean (SD)	0.82 (0.70)	0.74 (0.67)	1.11 (0.75)	0.97 (0.76)	1.18 (0.77)	0.001

P-values are from Wilcoxon rank sum tests for continuous variables and Fisher's exact tests for categorical variables

583