



Microbial Composition Predicts Genital Tract Inflammation and Persistent Bacterial Vaginosis in South African Adolescent Females

Katie Lennard,^{a,b} Smritee Dabee,^{a,c} Shaun L. Barnabas,^{a,c} Enock Havyarimana,^{a,c,e} Anna Blakney,^d Shameem Z. Jaumdally,^{a,c} Gerrit Botha,^{a,b} Nonhlanhla N. Mkhize,ⁱ Linda-Gail Bekker,^{a,e} David A. Lewis,^{f,g,h,i} Glenda Gray,^{j,k} Nicola Mulder,^{a,b} Jo-Ann S. Passmore,^{a,c,l} Heather B. Jaspan^{a,c,m,n,o}

^aInstitute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa

^bDepartment of Integrative Biomedical Sciences, University of Cape Town, Cape Town, South Africa

^cDepartment of Pathology, University of Cape Town, Cape Town, South Africa

^dDepartment of Bioengineering, University of Washington, Seattle, Washington, USA

^eDesmond Tutu HIV Centre, University of Cape Town, Cape Town, South Africa

^fWestern Sydney Sexual Health Centre, Parramatta, Australia

^gMarie Bashir Institute for Infectious Diseases and Biosecurity, University of Sydney, Sydney, Australia

^hSydney Medical School—Westmead, University of Sydney, Sydney, Australia

ⁱNational Institute for Communicable Diseases, Sandringham, Johannesburg, South Africa

^jPerinatal HIV Research Unit, University of the Witwatersrand, Johannesburg, South Africa

^kSouth African Medical Research Council, Cape Town, South Africa

^lNational Health Laboratory Service, Johannesburg, South Africa

^mSeattle Children's Research Institute, Seattle, Washington, USA

ⁿDepartment of Pediatrics, University of Washington, Seattle, Washington, USA

^oDepartment of Global Health, University of Washington, Seattle, Washington, USA

ABSTRACT Young African females are at an increased risk of HIV acquisition, and genital inflammation or the vaginal microbiome may contribute to this risk. We studied these factors in 168 HIV-negative South African adolescent females aged 16 to 22 years. Unsupervised clustering of 16S rRNA gene sequences revealed three clusters (subtypes), one of which was strongly associated with genital inflammation. In a multivariate model, the microbiome compositional subtype and hormonal contraception were significantly associated with genital inflammation. We identified 40 taxa significantly associated with inflammation, including those reported previously (*Prevotella*, *Sneathia*, *Aerococcus*, *Fusobacterium*, and *Gemella*) as well as several novel taxa (including increased frequencies of bacterial vaginosis-associated bacterium 1 [BVAB1], BVAB2, BVAB3, *Prevotella amnii*, *Prevotella pallens*, *Parvimonas micra*, *Megasphaera*, *Gardnerella vaginalis*, and *Atopobium vaginae* and decreased frequencies of *Lactobacillus reuteri*, *Lactobacillus crispatus*, *Lactobacillus jensenii*, and *Lactobacillus iners*). Women with inflammation-associated microbiomes had significantly higher body mass indices and lower levels of endogenous estradiol and luteinizing hormone. Community functional profiling revealed three distinct vaginal microbiome subtypes, one of which was characterized by extreme genital inflammation and persistent bacterial vaginosis (BV); this subtype could be predicted with high specificity and sensitivity based on the Nugent score (≥ 9) or BVAB1 abundance. We propose that women with this BVAB1-dominated subtype may have chronic genital inflammation due to persistent BV, which may place them at a particularly high risk for HIV infection.

KEYWORDS 16S RNA, HIV susceptibility, HIV target cells, female genital tract microbiome, inflammation, vaginal microbiome

Received 8 June 2017 Returned for modification 21 August 2017 Accepted 10 October 2017

Accepted manuscript posted online 16 October 2017

Citation Lennard K, Dabee S, Barnabas SL, Havyarimana E, Blakney A, Jaumdally SZ, Botha G, Mkhize NN, Bekker L-G, Lewis DA, Gray G, Mulder N, Passmore J-AS, Jaspan HB. 2018. Microbial composition predicts genital tract inflammation and persistent bacterial vaginosis in South African adolescent females. *Infect Immun* 86:e00410-17. <https://doi.org/10.1128/IAI.00410-17>.

Editor Vincent B. Young, University of Michigan-Ann Arbor

Copyright © 2018 American Society for Microbiology. All Rights Reserved.

Address correspondence to Heather B. Jaspan, hbjaspan@gmail.com.

Adolescent females and young women, particularly in sub-Saharan Africa, are at an extreme risk for adverse sexual and reproductive outcomes. Apart from sociobehavioral factors, genital inflammation is a major factor associated with these outcomes (1–3), and women with genital inflammation have a 3-fold-higher risk of later human immunodeficiency virus (HIV) acquisition. Several factors have been proposed to cause genital inflammation, yet the attributable risk and interdependence of these factors remain poorly understood. Hormonal contraception (HC), particularly the long-acting injectable progestin-only contraceptive depot medroxyprogesterone acetate (DMPA), may be associated with increased levels of proinflammatory cytokines (4), and meta-analyses have demonstrated a moderate increase (adjusted hazards ratio [aHR], 1.43; 95% confidence interval [CI], 1.23 to 1.67) in the HIV risk associated with DMPA use (5, 6). Sexually transmitted infections (STIs), including chlamydia and gonorrhea (7–9), are also associated with increased levels of female genital tract inflammation (10). However, these factors still do not fully explain the high incidence of HIV seen in South African adolescent females and young women (3). Bacterial vaginosis (BV), which is characterized by increased vaginal microbial diversity and a decrease in the proportion of lactobacilli, is associated with an increased risk of STIs (11–13), including HIV (14–16). BV rates vary by ethnicity and geographical location, with increased rates in women of African descent (17–19) and extremely high rates in South Africa (34 to 58%) relative to those in other countries (20).

Possible microbiota-mediated mechanisms of increased HIV risk include increased levels of specific inflammatory cytokines (7, 21–23), increased HIV target cell frequency or activation, and decreased levels of antimicrobial factors or epithelial barrier function in the lower female genital tract (21). While some studies have found increased interleukin-1 β (IL-1 β) and decreased secretory leukocyte protease inhibitor (SLPI) levels in BV-positive (BV⁺) women (24), the relationship with other cytokines such as IL-8, which are also relevant to HIV acquisition, remains unclear (24). Anahtar et al. described a high-diversity, *Prevotella*-mixed (a vaginal microbiota subtype referred to as a cervicotype 4 by Anahtar et al. that lacked a consistent dominant species, but communities all included *Prevotella*) female genital tract (FGT) (16S rRNA gene sequencing-based) microbiome subtype that was more strongly associated with genital inflammation than were other common STIs or classic BV status by Nugent scoring or Amsel criteria (23).

Here, we examine the vaginal microbiota of 16- to 22-year-old HIV-negative but high-risk South African women and relate it to genital inflammation and HIV target cell frequency and activation. We aimed to identify microbial factors that best predict genital inflammation and/or increased HIV target cell activation in adolescent females in an unbiased manner. We considered both microbiome composition and function to identify subtypes that are most predictive of genital inflammation while incorporating their importance relative to those of other potentially inflammatory variables, including BV status, STIs, and HC usage.

RESULTS

A total of 298 black, HIV-negative, 16- to 22-year-old South African women were recruited as part of the Women's Initiative in Sexual Health (WISH) study (61). Of the 185 women for whom 16S rRNA gene amplicon sequencing was performed, 168 passed sequencing quality control ($\geq 5,000$ reads/sample) and were used for downstream analyses: 78 were from Johannesburg (JHB) and 90 were from Cape Town (CPT). Of these women, 74 (44%) were BV positive, 22 (13%) were BV intermediate, and the remainder (43%) were BV negative, as determined by Nugent scoring (see Table S5 in the supplemental material), which is similar to data from previous reports on older South African women (20). The prevalences of BV and STI differed significantly by geographic location, with rates of BV being higher in CPT (46 women [51%]) than in JHB (28 women [36%]) ($P = 0.04$) and with overall rates of STIs also being higher in CPT than in JHB (54 women [60%] versus 19 women [24%]; $P = 4.8e-6$). In particular, the prevalence of *Chlamydia trachomatis* was 3-fold higher in CPT than in JHB (43% [$n = 39$] and 17% [$n = 13$], respectively; $P = 0.0002$) (Table S5). The other STIs tested for in

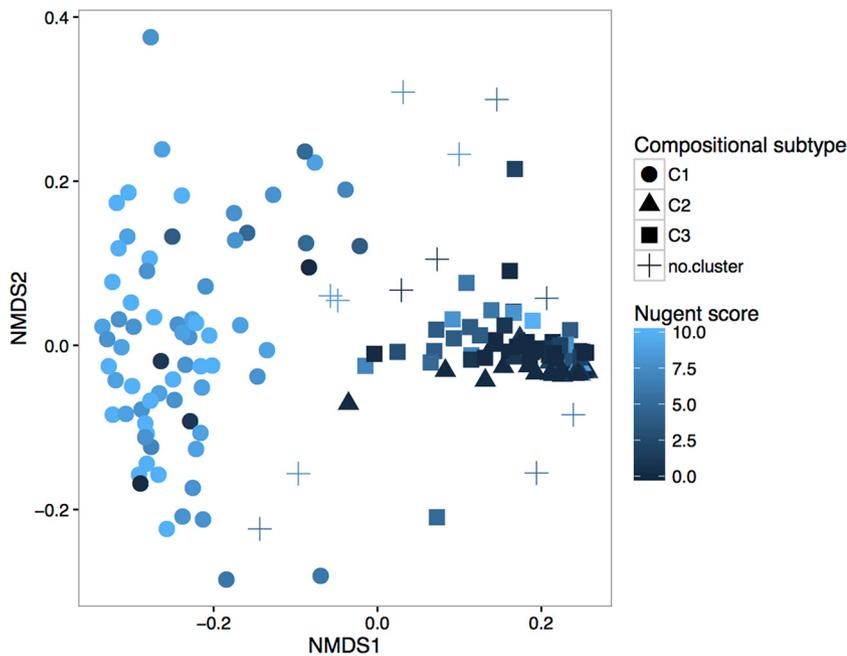


FIG 1 Nonmetric multidimensional scaling (NMDS) of microbial compositional subtypes established by using Fuzzy clustering with weighted UniFrac distances. Samples that did not meet the $\geq 60\%$ minimum probability of belonging to any of these clusters were excluded from downstream analyses (referred to as “no.cluster” in the key).

this study (*Neisseria gonorrhoeae*, *Trichomonas vaginalis*, *Mycoplasma genitalium*, herpes simplex virus 1 [HSV-1], HSV-2, *Haemophilus ducreyi*, *Treponema pallidum*, and lymphogranuloma venereum [LGV]) were present at much lower rates; after *Chlamydia trachomatis*, the next most prevalent causative organism was *N. gonorrhoeae* (14% in CPT versus 5% in JHB); *T. vaginalis*, *M. genitalium*, and HSV-2 were present in 1 to 7% of women, while HSV-1, *H. ducreyi*, *T. pallidum*, and LGV were not detected in any of the women (Table S5).

In addition, CPT women more commonly used the injectable HC DMPA or norethisterone enanthate (Net-En) (78 women [87%] in CPT versus 25 women [32%] in JHB; $P = 4e-13$) or any HC ($P < 2.2e-16$) and had high body mass indexes (BMIs) (CPT median, 25.4; JHB median, 22.5; $P = 0.02$) and low vaginal pH (CPT median, pH 4.7; JHB mean, pH 5; $P = 0.03$) compared to their JHB counterparts. CPT women almost exclusively reported Xhosa ancestry (99%), whereas the JHB site included women of Zulu (35%) and Xhosa (27%) ancestries and roughly equal proportions of Ndebele, Pedi, Sotho, Swati, Tsonga, Tswana, and Venda women ($P < 2.2e-16$).

Host immune markers as a proxy for genital inflammation. We defined genital inflammation based on the unsupervised clustering of all 47 markers (including cytokines, chemokines, and growth factors) for the 283 women for whom complete cytokine data were available. The use of all 47 cytokines, versus only the ones considered inflammatory, did not significantly alter the clusters obtained. We opted to use all 47 cytokines since there was clearly a global pattern in the cytokine data (see Fig. S1a in the supplemental material). Two unique clusters emerged, as judged by the average silhouette width for partitioning around medoids (PAM) clustering, with 62% ($n = 173$) and 38% ($n = 107$) of women being classified as having either high inflammation (inflammation-H) or low inflammation (inflammation-L) (Fig. S1a).

Genital microbiota type is strongly associated with genital inflammation. To characterize the vaginal microbiota, we employed an unsupervised approach to the 16S rRNA gene sequence data using Fuzzy clustering with weighted UniFrac distances, which identified three distinct clusters (Fig. 1) that we named cluster 1 (C1), C2, and C3. C1 was dominated by participants with high Nugent scores (BV positive). C2 and C3

TABLE 1 Odds ratios for predicting high versus low inflammation using various microbiome subtyping methods showing the distribution of these factors by inflammation category

Factor	OR (95% CI)	Chi-square P value	Distribution (%)	
			Inflammation-H	Inflammation-L
Compositional subtype (C1 vs C2/C3)	18.07 (7.30, 44.77)	<0.001	69 C1	11 C1
BV (positive/intermediate vs negative)	14.17 (6.63, 30.27)	<0.001	81 BV ⁺ /intermediate	23 BV ⁺ /intermediate
Subtype determined by Anahtar et al. (<i>Prevotella</i> mixed vs all others)	11.18 (4.42, 28.32)	<0.001	52 <i>Prevotella</i> type	9 <i>Prevotella</i> type
Functional subtype (F2 + F3 vs F1)	11.17 (5.25, 23.77)	<0.001	77 F2/F3	23 F2/F3
Functional subtype (F3 vs F2 + F1)	Inf (NaN, Inf) ^a	<0.001	39 F3	0 F3
BV (positive vs intermediate/negative)	9.46 (4.38, 20.41)	<0.001	64 BV ⁺	16 BV ⁺

^aNaN, not a number; Inf, infinite, since 0% inflammation-L cases were of the F3 type.

were made up of predominantly *Lactobacillus crispatus* and *Lactobacillus iners*, respectively. C1 was highly predictive of genital inflammation, with an odds ratio (OR) of 18.07 (95% CI, 7.30, 44.77) (Table 1), compared with C2/C3 (which were grouped together given their shared *Lactobacillus*-dominant compositions). The OR for BV was 14.17 or 9.46, depending on whether BV-intermediate cases were grouped with BV-positive or -negative cases, respectively. We compared our clustering to that reported previously by Anahtar et al. (23), who described four clusters based on dominant taxa (*L. iners*, non-*L. iners* lactobacilli, *Gardnerella*, and *Prevotella* mixed). There was overlap between our clusters and those of Anahtar et al., with their *Gardnerella*- and *Prevotella*-mixed groups together corresponding to our subtype C1, their *Lactobacillus* (not *L. iners*) subtype corresponding to our subtype C2, and their *L. iners*-dominant subtype corresponding to our subtype C3. However, there was no difference in genital inflammation between the *Gardnerella* (a vaginal microbiota subtype referred to as a cervicotype 3 by Anahtar et al. where *Gardnerella* was the dominant taxon) and *Prevotella*-mixed groups in our cohort, and only 8% of women (versus 28% as determined by Anahtar et al.) were *Gardnerella* dominant (Fig. 2). Furthermore, our C1 was more strongly predictive of inflammation than was their inflammatory *Prevotella*-mixed cervicotype (Table 1).

In a multivariate logistic regression including all factors *a priori* considered to be potentially associated with inflammation (geographical location, hormonal contraceptive usage, STI [any], microbiota subtype, ethnicity, and BMI), both vaginal microbiota composition (adjusted OR [aOR], 23; 95% CI, 7 to 83) and any HC use (aOR, 14; 95% CI, 3 to 73) were significantly associated with vaginal inflammation. There were no significant differences in the proportions of inflammation-H cases between different types of HC; however, small sample sizes for some of the groups (e.g., $n = 10$ for oral contraceptives) may have precluded the detection of significant differences. Of the HCs considered, women on DMPA and an etonogestrel implant had the highest rates of genital inflammation, at 81% ($n = 21/26$) and 86% (6/7), respectively, with a rate of only 59% (41/70) among Net-EN users. DMPA was the only HC associated with significantly higher rates of genital inflammation than male condom usage only (adjusted P value of 0.006). The presence of any STI, BMI, geographic location, and ethnicity did not significantly influence genital inflammation (Table 2). Given the high burden of *Chlamydia* in this cohort and its reported association with inflammation, a second multivariate analysis was performed with *Chlamydia* as a categorical variable instead of any STI. *Chlamydia* by itself was also not significantly associated with inflammation in this cohort ($P = 0.9$), while microbiota subtype and HC use were.

Bacterial taxa associated with genital inflammation. We used MetagenomeSeq and random forest analysis to identify specific bacterial taxa associated with inflammation, adjusting for the use of HC. *Gardnerella vaginalis*, *Sneathia sanguinegens*, *Prevotella amnii*, *Prevotella pallens*, *Megasphaera*, bacterial vaginosis-associated bacterium 1 (BVAB1), BVAB2, BVAB3, *Gemella*, *Sneathia*, *Clostridium*, *Atopobium vaginae*, *Dialister microaerophilus*, and *Parvimonas micra*, among others, were found in significantly higher relative abundances in inflammation-H samples. Conversely, several lactobacillus species, including *Lactobacillus crispatus*, *L. jensenii*, *L. reuteri*_cluster, and

TABLE 2 Multivariate logistic regression analysis of predictors of genital inflammation

Predictor ^b	Coefficient	P value	Adjusted odds ratio ^a	95% CI
Location (JHB)	0.7	0.4		
Hormonal contraception (yes)	2.6	9.00e−04	14	3–72
STI (any)	0.4	0.4		
Microbiota subtype (C1 vs C2/C3)	3.1	3.00e−07	23	8–84
Ethnicity (Xhosa vs other)	0.3	0.7		
BMI	−0.02	0.7		

^aAdjusted for all other variables in the table.

^bSTI, sexually transmitted infection; JHB, Johannesburg.

L. iners, as well as *Fingoldia magna*, *Morganella morganii*, and *Varibaculum cambriense* were found in significantly lower abundances in inflammation-H cases (Fig. 3A; see also Table S6 in the supplemental material). While most of these taxa differed significantly in terms of relative abundance only, *Gemella* and Rs-045 were also significantly more commonly present in inflammation-H samples, while *L. reuteri*_cluster and *Varibaculum* were significantly more commonly detected in inflammation-L samples (Table S6). Random forest analysis confirmed these results, scoring *Prevotella*, *Dialister*, *P. micra*, *S. sanguinegens*, *L. crispatus*, and *L. johnsonii_gasseri* as being the most highly influential taxa in predicting inflammation (Fig. 3B). The cross-validated prediction performance of models with sequentially smaller numbers of predictors (ranked by variable importance) was determined by using a nested cross-validation procedure. The use of only the 10 most influential taxa (Fig. 3B) resulted in a cross-validated error rate of 25%. The use of these 10 taxa only to predict inflammation in the test set resulted in an area under the curve (AUC) of 0.8, a positive predictive value (PPV) of 0.67, and a negative predictive value (NPV) of 0.78.

Community functional profiling reveals alternative microbiota types. The Human Microbiome Project revealed that microbial community function remains relatively stable between individuals and body sites, despite variation in microbiome compositions among individuals, with a stable high-abundance housekeeping “core” and more variable niche-specific pathways/functions (25). We therefore hypothesized that among women of a given compositional microbiota subtype, a minority of genital microbial community functions or pathways may substantially impact host mucosal immunology, despite relatively similar microbial compositions. Although we and others have shown that compositional subtypes are good predictors of inflammation (20), we investigated whether community functional subtypes might reveal novel associations with host immunology/inflammation. Bacterial gene content was predicted by using phylogenetic investigation of communities by reconstruction of unobserved states (PICRUSt). Three distinct functional groups (functional group 1 [F1], F2, and F3) were identified (Fig. 4A). F1, F2, and F3 differed significantly by Nugent score and BV status (with BV⁺ rates of 11%, 63%, and 94%, respectively) (Table 3). F3 had significantly higher Nugent scores than did F2 or F1 ($P < 0.05$), even when considering only BV-positive (Nugent score of ≥ 7) samples between F2 and F3, which had median Nugent scores of 8 and 10, respectively. In fact, having a Nugent score of ≥ 9 could predict F3 versus F2 with sensitivity and specificity of 80 and 88%, respectively, and redundancy analysis (RDA) revealed that 30% of the functional variation observed between samples could be explained by Nugent score or BV, with none of the other variables contributing significantly.

PICRUSt analysis revealed that C1 consisted of two functionally distinct subtypes (F2 and F3) that were not discernible from the compositional data alone (Fig. 4B). This is in contrast to compositional data that revealed two distinct, predominantly BV-negative

FIG 2 Legend (Continued)

16S rRNA gene-based microbiome profiles; the functional subtype was determined from inferred functional data (described in Materials and Methods).

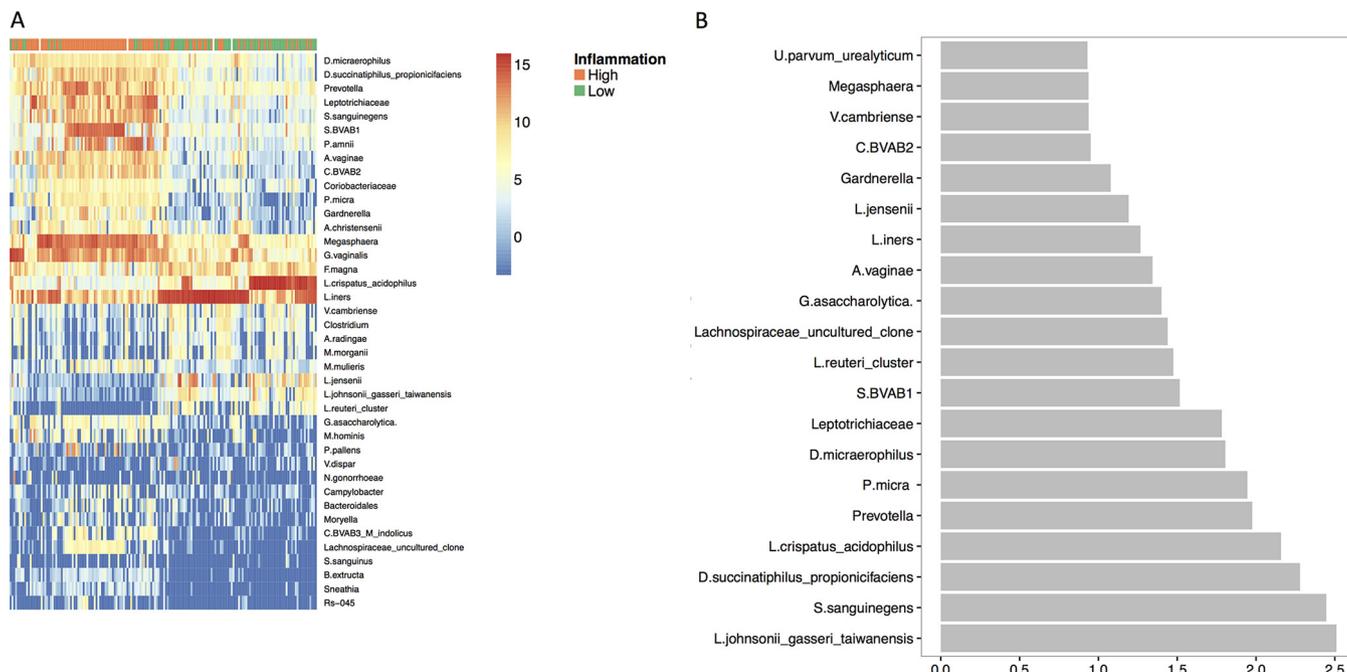


FIG 3 Taxa significantly altered between inflammation-high and inflammation-low cases (adjusted for any versus no HC use). (A) Taxa that were significantly differentially abundant and/or frequent by inflammation category using MetagenomeSeq (FDR of ≤ 0.05 ; coefficient of ≥ 1.25 ; taxa present in $\geq 20\%$ of samples in at least one of the two groups being compared). Shown is the unsupervised clustering of samples (columns) by Bray-Curtis distances; the heat map scale is on \log_2 -transformed standardized counts. M.hominis, *Mycoplasma hominis*; U.parvum, *Ureaplasma parvum*; D.succinatiphilus, *Dialister succinatiphilus*; A.christensenii, *Aerococcus christensenii*; A.radingae, *Actinomyces radingae*; M.mulieris, *Mobiluncus mulieris*; G.asaccharolytica, *Gemella asaccharolytica*; V.dispar, *Veillonella dispar*. (B) The top 20 most influential taxa by random forest analysis. The x axis indicates the mean decrease in the Gini index (lengths of bars represent the predictive ability of each taxon).

subtypes (C2 and C3) and one predominantly BV-positive subtype (C1). Functional subtype F3 consisted entirely of inflammation-H cases (100%), while F2 and F1 had 69% and 29% inflammation-H cases, respectively (Fig. 4C). F3 consisted exclusively of C1 samples, F2 included both C1 (85%) and C2 (15%) samples, and F1 consisted of C2 (44%) and C3 (54%) samples. Together, F2 and F3 were stronger predictors of inflammation than F3 alone (OR, 11.17 [95% CI, 5.25 to 23.77]; $P < 0.001$; PPV = 82%; NPV = 71%) but were not as strong as C1. Of the C1 samples, those in F3 were more commonly inflammation-H than those in F2 ($P = 0.004$) (Table 3), suggesting that the presence of lower-abundance taxa, with distinct community functions, might be associated with inflammation. Based on these findings, we hypothesized that F3 represented a unique subset of inflammation-H, BV⁺ samples with biologically relevant attributes.

High relative abundance of BVAB1 is related to inflammation and BV persistence. Of the 42 BV-positive cases with available Nugent scoring at follow-up visits (every 2 to 3 months depending on the type of HC used), F3 women were significantly more likely to exhibit persistent BV than were F2 women ($P = 0.02$). Differential

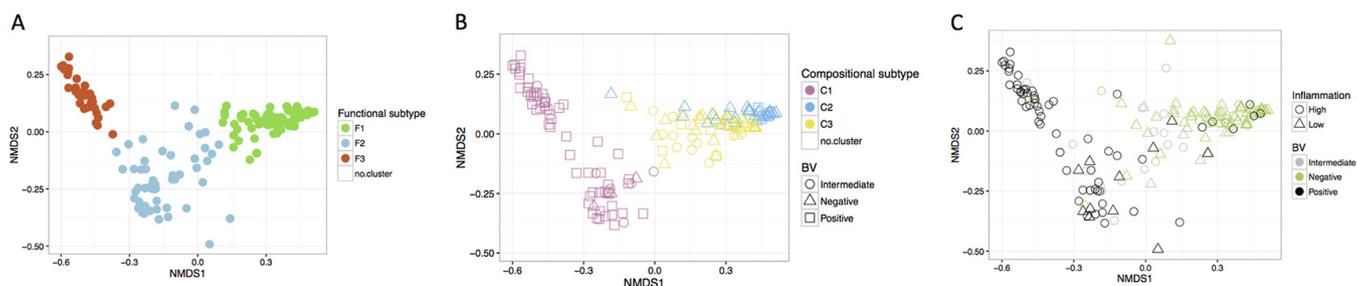


FIG 4 Community functional subtypes. Shown is nonmetric multidimensional scaling of samples colored by functional subtype (A); samples colored by compositional subtype, with BV subtypes displayed as shapes (B); and samples colored by BV status, with inflammation-H/L displayed as shapes (C).

TABLE 3 Participant summary by community functional subtype (F1, F2, and F3)

Parameter	Value for group			P value comparing all groups ^f	P value for F3 vs F2 ^f
	F1 (n = 74)	F2 (n = 51)	F3 (n = 35)		
No. (%) of participants at location				0.02	0.8
CPT	31 (42)	32 (63)	23 (66)		
JHB	43 (58)	19 (37)	12 (34)		
Median age (yr) ^a	18	19	19	0.24	
Median pH ^b	4.7	5	5	4.00e−06	0.8
No. (%) of participants with HPV status				0.05	0.7
High risk	21 (28)	24 (57)	19 (54)		
Low risk	19 (26)	13 (26)	6 (17)		
Negative	34 (46)	14 (27)	10 (29)		
No. (%) of participants with BV status				<2.2e−16	0.003
BV positive	8 (11)	32 (63)	33 (94)		
BV intermediate	6 (8)	11 (21)	1 (3)		
BV negative	60 (81)	8 (16)	1 (3)		
No. (%) of participants of compositional subtype ^c				<2.2e−16	3.00e−04
C1	0 (0)	35 (85)	35 (100)		
C2	32 (44)	0 (0)	0 (0)		
C3	41 (56)	6 (15)	0 (0)		
Median Nugent score	0	8	10	<2.2e−16	1.00e−04
No. (%) of participants with STI					
Any	17 (23)	29 (57)	19 (54)	1.00e−04	0.8
<i>C. trachomatis</i>	13 (18)	20 (39)	16 (46)	0.003	0.7
No. (%) of participants of inflammatory category ^d				4.00e−12	1.00e−04
High	21 (29)	34 (69)	35 (100)		
Low	51 (71)	15 (31)	0 (0)		
Median BMI ^e	22.5	23.3	25.8	0.0001	1
Median α diversity value (Shannon)	1	2.3	2.3	<2.2e−16	0.8
No. (%) of participants using HC	48 (65)	38 (76)	28 (80)	0.2	0.8

^aEight women had missing information.^bSix women had missing information.^cEleven women were not classified as belonging to C1/C2/C3 (i.e., clustering criteria were not met).^dFour women had missing information.^eSeven women had missing information.^fSignificant adjusted *P* values are in boldface type.

abundance testing of the compositional data using MetagenomeSeq revealed that the functional differences between F3 and F2 could be ascribed to significant changes in the abundances of several taxa, including increases in the abundances of BVAB1, BVAB2, BVAB3, *Lachnospiraceae_uncultured_clone*, *Gardnerella*, *Prevotella amnii*, *Megasphaera*, and *Prevotella pallens* and significant decreases in the abundances of *L. reuteri_cluster*, *Veillonella montpellierensis*, *Streptococcus anginosus*, *Prevotella bivia*, and *Fusobacterium* (Fig. 5B; see also Table S7 in the supplemental material). These results were supported by random forest analysis (Fig. 5A), which further demonstrated that F2 could be distinguished from F3 almost perfectly (with a 10-fold cross-validated error rate of 2.3% in the training set and a validation prediction error rate of 2.3% in the test set), using only BVAB1 as a predictor. These PICRUSt predictions depend on the genome similarity between the taxa detected here and their nearest sequenced reference genome (from which the functional data are inferred). The nearest sequenced reference genome for BVAB1 has only 89% identity across the 16S rRNA gene V4 region with *Shuttleworthia* (26), and it is therefore not possible to draw definitive conclusions based on these results, which require further validation by whole-genome sequencing.

Endogenous hormones, hormonal contraceptive use, and BMI are associated with vaginal microbiota subtypes. We next determined which physiological or

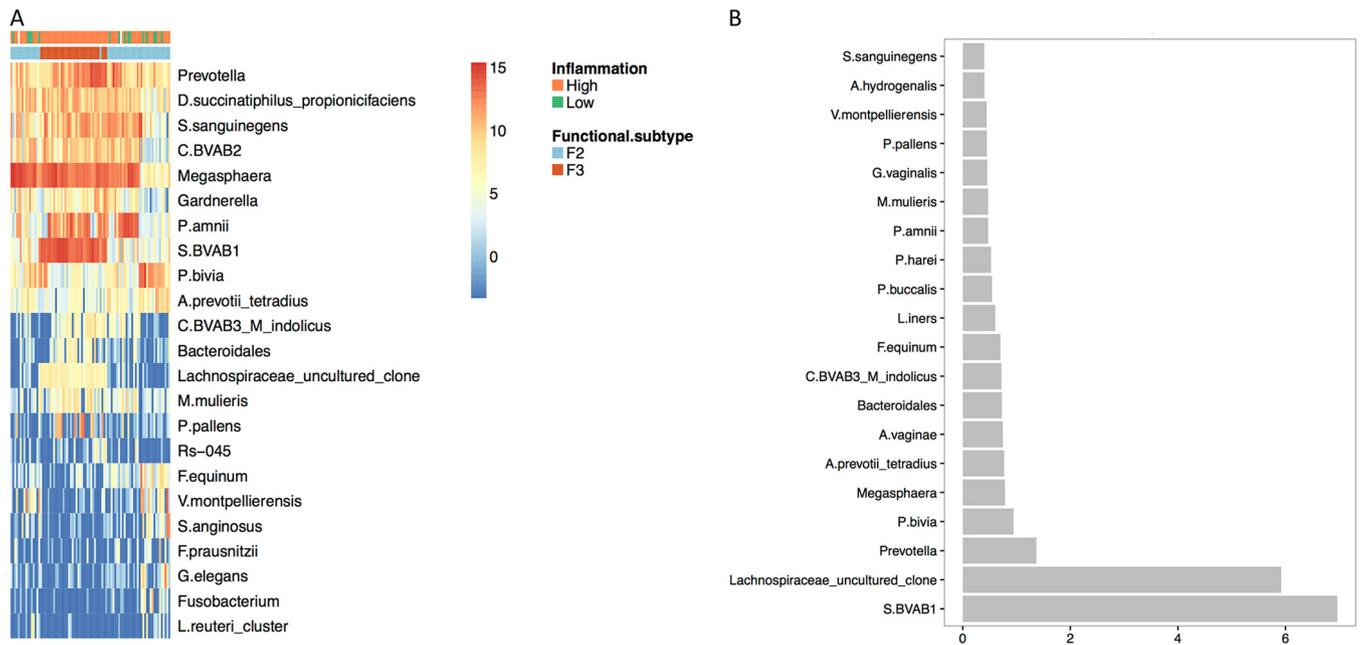


FIG 5 Taxa significantly different in F3 versus F2. (A) Top 20 most important taxa by random forest analysis. The x axis shows the mean decrease in the Gini index (lengths of bars represent the relative predictive ability of each taxon). *A.prevotii*, *Anaerococcus prevotii*; *M.indolicus*, *Mageeibacillus indolicus*; *F.equinum*, *Fusobacterium equinum*; *F.prausnitzii*, *Faecalibacterium prausnitzii*; *G.elegans*, *Granulicatella elegans*; *P.harei*, *Peptinophilus harei*. (B) Taxa that were significantly differentially abundant and/or frequent in F3 versus F2 (FDR of ≤ 0.05 ; coefficient of ≥ 1.25 ; taxa present in $\geq 20\%$ of samples in at least one of the two groups being compared). Shown is hierarchical clustering (Bray-Curtis distance); the heat map scale shows \log_2 -transformed standardized counts. *P.buccalis*, *Prevotella buccalis*.

demographic factors were most strongly associated with genital microbiota compositional subtypes. In a multivariate model with C1 versus C2/C3 as the outcome, inflammation-H (OR, 13 [95% CI, 4 to 49]; $P = 3e-5$) and high vaginal pH (OR, 4.7 [95% CI, 1.6 to 16]; $P = 0.008$) were associated with C1, as expected. However, higher BMI (OR, 1.2 [95% CI, 1 to 1.3]; $P = 0.01$), no use of HCs (OR, 6 [95% CI, 1.2 to 37]; $P = 0.04$), living in CPT (OR, 6 [95% CI, 1.4 to 30]; $P = 0.02$), and the presence of one or more STIs (OR, 3.1 [95% CI, 1.1 to 9.1]; $P = 0.03$) were additionally associated with C1, while age was not.

The relationship among endogenous hormones (estradiol, luteinizing hormone [LH], and progesterone), microbiota subtypes, and genital inflammation was assessed only for women not utilizing any form of HC: of these 47 women, 17 (36%) were BV positive (Nugent score of ≥ 7), while a further 3 women (6%) were BV intermediate (Nugent score of 4 to 6). The three BV-intermediate cases were included with the BV-negative cases for downstream analyses, due to the small size of this group. The estradiol level was significantly lower in C1 (median, 0.19 nmol/liter) than in C2 (0.58 nmol/liter; $P = 0.001$) and C3 (0.3 nmol/liter; $P = 0.07$) women. Furthermore, inflammation-H cases had significantly lower estradiol levels than did inflammation-L cases (0.18 versus 0.34 nmol/liter; $P = 0.02$), similarly to F3 (0.14 nmol/liter) and F2 (0.2 nmol/liter) relative to F1 (0.49 nmol/liter; $P = 0.01$ and $P = 0.02$, respectively). The LH levels were significantly lower in C1 (4.9 IU/liter) than in C2 (8.7 IU/liter; $P = 0.02$) and C3 (9.3 IU/liter, $P = 0.02$) and were also significantly lower in F3 (4.9 IU/liter) and F2 (5.8 IU/liter) than in F1 (9.8 IU/liter; $P = 0.01$ and $P = 0.04$, respectively) but did not differ by inflammation status. Progesterone levels were significantly higher in C2 than in C3 (median, 12 nmol/liter versus 1.25 nmol/liter; $P = 0.02$) but did not differ by microbiome functional subtype or inflammation status.

Specific host immunological factors are associated with microbiota subtypes.

Since genital inflammation and CD4⁺ target cell availability are important host factors determining susceptibility to HIV, we next examined the relationship between these immunological factors and microbiota subtypes. We found that most of the immuno-

Downloaded from <http://iai.asm.org/> on March 6, 2018 by INSERM

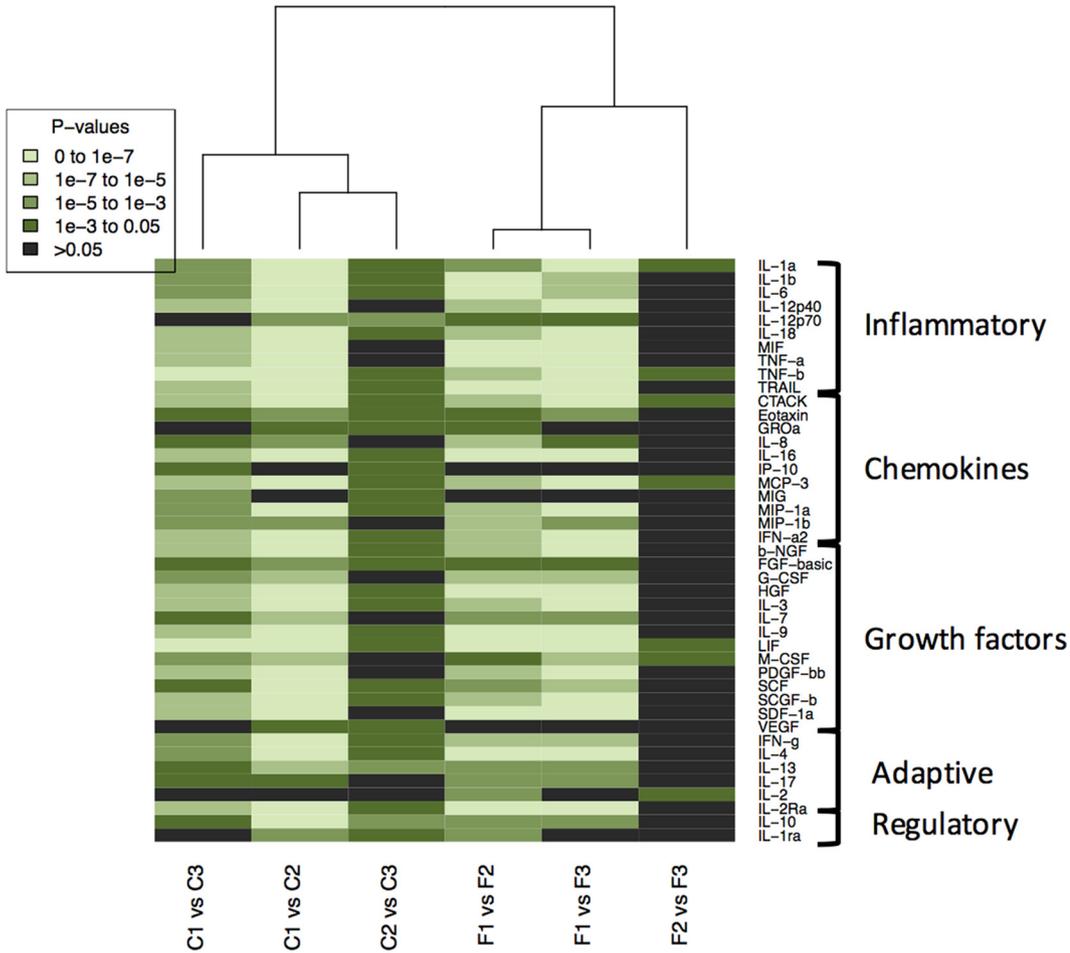


FIG 6 Heat map summary of cytokines by microbiome compositional (C1, C2, and C3) and functional (F1, F2, and F3) subtypes highlighting cytokines significantly different between F2 and F3. Results that were significant by a Kruskal-Wallis test were evaluated by Dunn’s *post hoc* test.

logical factors measured (44/47) were significantly different by microbial compositional subtype (false discovery rate [FDR] of ≤ 0.05) (Fig. 6). Strikingly, of these 44 factor, the majority (27) had levels that were significantly higher in C1 than in C2 and C3. GRO- α , IL-1R α , IL-12p70, and vascular endothelial growth factor (VEGF) levels were significantly higher in C1 than in C2 but not C3, and MIG and IP-10 levels were significantly higher in C3 than in both C2 and C1 (Fig. 6).

Similarly, the levels of 43 cytokines varied significantly by functional subtype (FDR, ≤ 0.05) (Fig. 6). Not surprisingly, the majority (40/43) showed higher levels in F3 and F2 than in F1. Importantly, however, the levels of certain proinflammatory cytokines/chemokines, including IL-1 α , CTACK, and tumor necrosis factor beta (TNF- β), were increased in F3 (which was associated with BV persistence) relative to both F1 and F2, as were the growth factors LIF and macrophage colony-stimulating factor (M-CSF) and the chemokine monocyte chemoattractant protein 3 (MCP-3). The IL-2 level was decreased in F3 relative to F2 (Fig. 6; see Fig. S2 in the supplemental material).

The mechanism by which elevated levels of genital cytokines increase HIV risk is unknown but is presumed to be due to disruptions in mucosal barrier integrity or to their ability to attract and activate HIV target cells. For this reason, we compared the frequencies of CD4⁺ T cells expressing CCR5, the HIV coreceptor, and HLA-DR and CD38, activation markers, on cervical T cells immediately *ex vivo* and compared these values between compositional and functional subtypes. Surprisingly, there were no significant differences in the frequencies of cervical CD4⁺ T cells expressing CCR5,

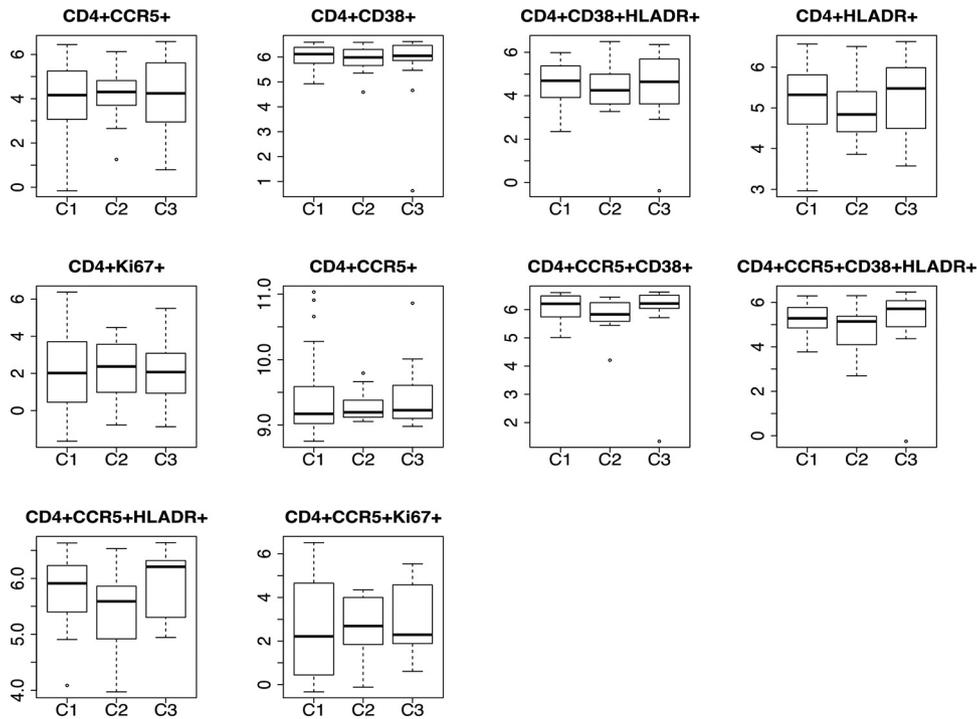


FIG 7 Box plot summaries of (\log_2) target cell activation frequencies by compositional subtype (C1, C2, and C3).

CD38, or HLA-DR, alone or in combination, or Ki67 between compositional subtypes (Fig. 7). These CD4 T cell frequencies also did not vary significantly by functional subtype (Fig. S3), suggesting that the mechanisms by which microbiota increase HIV susceptibility may be related to epithelial barrier integrity rather than HIV target cell availability.

DISCUSSION

In young South African women, the genital microbiota is an important predictor of genital inflammation. Here, we describe specific microbiota subtypes and individual taxa that are associated with inflammation. Furthermore, using community functional inference, we identify BVAB1 as a likely contributor to BV persistence.

Of all the potential factors considered to be associated with increased concentrations of inflammatory cytokines in the lower genital tract in young women, the microbiota subtype was the strongest predictor of inflammation, with HC use being the only other independently predictive factor in a multivariate analysis. DMPA was the HC most significantly associated with inflammation after adjustment for microbiota subtype and other potential confounders, despite the relatively small number of women in this group compared to the number of women who use Net-En. Net-En is used extensively in South Africa but differs from DMPA in its relatedness to progesterone, molecular mechanisms of action, and serum protein binding capabilities (28). Most importantly, in a recent systematic review, DMPA was associated with around a 1.5-fold increased risk of HIV acquisition, while Net-En was not (29). Here, HC use was independently associated with both inflammation and microbiota compositional subtype, suggesting that all three of these factors contribute to HIV risk in these young women. Although STIs were previously associated with genital inflammation (7–10) and were more prevalent among inflammation-H women in our study, this association was not significant in multivariate models in this cohort. Rather, microbiota was a stronger driver of inflammation. The reasons for this are unclear but may be related to the age of our study population (adolescents) or the possibility that previous studies examining the relationship between STIs and genital inflammation did not take into account vaginal microbiota. Since vaginal microbiota may modify the risk of chlamydial infec-

tion (30), this may have been a confounding factor in the apparent relationship between *C. trachomatis* infection and genital inflammation in previous studies. We identified 40 taxa significantly associated with high or low genital inflammation in these black South African adolescents, only 5 of which (*Prevotella*, *Gemella*, *Sneathia*, *Aerococcus*, and *Mobiluncus*) were previously reported as such (23). The abundances of *S. sanguinegens*, *P. micra*, BVAB1/2/3, *A. vaginae*, *P. amnii*, *Megasphaera*, and *Gardnerella* were increased 2- to 4-fold in women with high inflammation, while the abundances of all *Lactobacillus* species (including *L. iners*) were decreased 1.5- to 4-fold, with the greatest decrease being seen for *L. crispatus*. These findings support the use of *L. reuteri* (and other lactobacilli) as probiotics to promote vaginal health. *P. amnii* may be particularly relevant in the African context, as African American women are more likely to be colonized by *P. amnii* than are European women (31).

We did not find any association between vaginal microbiota and the frequency of activated or proliferating HIV target (CD4⁺ CCR5⁺) T cells in the cervix. This implies that the vaginal microbiota mediates HIV risk via mechanisms unrelated to the recruitment of HIV target cells. One such mechanism may be mucosal barrier disruption, which would allow virions easier access to target cells in the submucosa and is consistent with the functional prediction of the most inflammatory vaginal subtype, F3. A related possibility is that lactobacillus-dominated microbiota may trap HIV virions more readily (32). This finding is in contrast to previously reported findings for adult women in Kwazulu-Natal, where the *Prevotella*-mixed cervicotype was associated with 17-fold-higher numbers of CD4⁺ CCR5⁺ CD38⁺ HLA-DR⁺ cells in cervical cytobrushes (22). Here, we used frequencies of CD4⁺ cells expressing CCR5 alone or in combination with one or both of these activation markers and found no such association. Possible reasons for this discrepancy are that cytobrush cell numbers are potentially increased with barrier integrity disruptions but that the proportions of cells that are HIV target cells are not. Another potential explanation could be that HIV susceptibility mechanisms could differ between adolescents and adults.

The levels of the endogenous hormones estradiol and LH were both decreased in women with C1 microbiota. Estradiol levels were negatively associated with inflammation, while progesterone was not associated with inflammation. In postmenopausal women, low-dose estrogen treatment significantly increases the abundance of lactobacilli and decreases the abundance of *Gardnerella* (33). Estradiol promotes lactobacillus growth by stimulating epithelial cell production of glycogen. Importantly, estradiol confers protection against simian immunodeficiency virus (SIV) in animal models (34, 35), potentially via decreasing inflammation, CD4⁺ T cell-dependent antiviral effects (36), and lactobacillus-dependent virus inhibition, particularly in women with lactobacillus-deficient microbiota (i.e., C1 subtype).

Using PICRUSt, we report two functionally distinct, predominantly BV-positive subtypes, F2 and F3, that can be distinguished with high specificity and sensitivity based on Nugent score (≥ 9) or BVAB1 abundance. PICRUSt infers gene content based on the nearest sequenced taxon. Our nearest-sequenced-taxon index (NSTI) scores indicate that the majority of inferences could be made at the species/genus level (see Materials and Methods); however, many of the genes that discriminate F3 from F2 were (predicted to be) contributed by BVAB1, for which predictions were based on the genome of *Shuttleworthia*. Nonetheless, the distinct phenotype of F3 suggests that functional annotation may have important clinical implications, which is supported by the significant increase in the IL-1 α level in F3 women. We previously demonstrated a 1.65-fold increased odds of HIV acquisition per log increase of the IL-1 α concentration in the genital tract in an adult South African cohort (3).

Despite the limitations of PICRUSt as a predictor of gene functionality from 16S amplicon sequencing data, we found similarities between our results and those of previously reported metabolomics studies. Srinivasan et al. performed metabolomics analysis on vaginal swabs from a predominantly North American cohort in relation to absolute abundances of key vaginal bacteria and found two clusters of metabolically different BV-associated bacteria, one of which was characterized by BVAB1, BVAB3, and

P. amnii (37). In our cohort of young African women, this BVAB1-associated subtype showed extremely high vaginal inflammation. These women may have chronic genital inflammation due to persistent BV (possibly mediated by adherent bacterial biofilms), which may place them at a particularly high risk of HIV infection. Biofilm formation is a prominent virulence mechanism that enhances bacterial attachment to epithelial surfaces, allowing bacteria to reach high concentrations (38), and prevents antimicrobial agents from reaching organisms, resulting in persistent and recurrent BV (38). Evidence for a higher occurrence of biofilms among F3 women include increased levels of *G. vaginalis*, a major constituent of vaginal biofilms (39). Increased functions related to flagellar and motility proteins can also promote biofilm formation by facilitating the initial cell-to-surface interactions (27) and biofilm architecture (40). The highly specific increase in the abundance of BVAB1 (which we and others have reported in association with BV persistence [41]) may point toward a hitherto underappreciated role for BVAB1 in vaginal biofilms, which are themselves associated with refractory or persistent BV.

In conclusion, we show that compositional and functional microbiota subtyping produces complementary results and subtypes that are more strongly associated with genital inflammation than BV status. In addition to genital microbiota, HC usage was associated with high genital inflammation. We identify several novel taxa associated with genital inflammation and demonstrate that women with Nugent scores of ≥ 9 and high levels of BVAB1 are more likely to have persistent BV and high levels of genital inflammation, which may place them at an increased risk of HIV infection. Most importantly, we found no differences in the frequencies of activated, CCR5⁺ cervical CD4⁺ T cells between microbiota subtypes, which raises the possibility that the microbiota could instead increase HIV susceptibility by compromising epithelial barrier integrity.

MATERIALS AND METHODS

Participant selection and sample collection. A total of 298 black, 16- to 22-year-old, HIV-negative South African women were recruited as part of the Women's Initiative in Sexual Health (WISH) study (61), from low-income, high-population-density communities in Cape Town (CPT) and Johannesburg (JHB). Approval was obtained for the study from the Research Ethics Committees of the Universities of Cape Town and Witwatersrand. All participants ≥ 18 years of age provided informed consent, while assent and parental consent were obtained for participants < 18 years of age. Young women were enrolled if they were HIV negative, in general good health, and not pregnant or menstruating at the time of sampling and if they had not had unprotected sex or douched in the last 48 h. Additional exclusion criteria were the use of antibiotics in the previous 2 weeks. Study visits were scheduled 2 weeks after injection for participants on injectable progestin contraceptives or otherwise during the luteal phase of their menstrual cycles (between days 14 and 28) if they were not using any HCs or if they were using oral HCs. Before specimen collection, the following procedures were performed: an HIV pretest and risk reduction counseling, a rapid HIV test (Aleré Determine HIV-1/2 Ag/Ab Combo; Aleré, Waltham, MA), a pregnancy test (U-test pregnancy strip; Humor Diagnostica, Pretoria, South Africa), and a general physical examination. The following genital samples were collected, in order: cervicovaginal fluid, using a disposable menstrual cup (Softcup) placed over the cervix for an hour, two vulvovaginal swabs (after cup removal) for STI testing, Nugent scoring and microbiota analysis, and the use of a cervical cytobrush for the isolation of cervical T cells for flow cytometry. CPT women were monitored longitudinally for a total of three visits, every 2 months if they were using Net-EN, combined oral contraceptives, or barrier contraception only or every 3 months if they were using DMPA.

STI and BV testing. Vulvovaginal swabs were collected for STI testing by multiplex PCR (*Chlamydia trachomatis*, *Neisseria gonorrhoeae*, *Trichomonas vaginalis*, *Mycoplasma genitalium*, HSV-1 and -2, *Haemophilus ducreyi*, *Treponema pallidum*, and lymphogranuloma venereum), as previously described (42). Chlamydial infections were differentiated into LGV-associated serovars or non-LGV-associated serovars by the use of a LGV-specific PCR (no LGV strains were detected). Blood was obtained for rapid HIV and HSV-2 serological testing. Endocervical swabs were collected for human papillomavirus (HPV) detection and genotyping by using a Roche linear array (62). The following HPV types were considered high-risk HPV types: 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and 68 (43). For analyses referring to STIs, women considered positive had to have at least one of the STIs tested for in this study (excluding HPV). Lateral wall/posterior fornix swabs were collected for Nugent scoring to classify samples as being BV negative (Nugent score of 0 to 3), BV intermediate (Nugent score of 4 to 6), or BV positive (Nugent score of 7 to 10), and vaginal pH was measured by using color-fixed indicator strips (Macherey-Nagel, Düren, Germany). Women were classified as having persistent BV if they had Nugent scores of ≥ 7 at visits 1 and 2, for women for whom we had Nugent scores for only visits 1 and 2, or at visits 1, 2, and 3, for women for whom we had Nugent scores for all three visits (see Table S1 in the supplemental material).

Host mucosal immune and endogenous hormone profiling. The concentrations of 47 cytokines, chemokines, and growth factors in cervicovaginal fluid were profiled by using a Luminex assay (Bio-Rad

Laboratories, Hercules, CA, USA) as previously described (7). These factors included IL-10 and IL-1 α (considered regulatory cytokines); gamma interferon (IFN- γ), IL-4, IL-13, IL-17, IL-2, IL-2R α , and IL-5 (considered adaptive cytokines); β -NGF, bFGF, G-CSF, HGF, IL-3, IL-7, IL-9, LIF, M-CSF, PDGF-BB, SCF, SCGF- β , SDF-1 α , GM-CSF, and VEGF (considered growth factors); CCL27/CTACK, eotaxin, CXCL1/GRO- α , IL-8, IL-16, CXCL10/IP-10, MCP-1, MCP-3, CXCL9/MIG, MIP-1 α , MIP-1 β , RANTES, and IFN- α 2 (considered chemokines); and IL-1 α , IL-1 β , IL-6, IL-12p40, IL-12p70, IL-18, MIF, TNF- α , TNF- β , and TRAIL (considered inflammatory cytokines).

Participants were categorized as having high or low genital tract inflammation (inflammation-H or inflammation-L, respectively) by unsupervised clustering of these 47 proteins applied to the 283 women for whom complete cytokine data were available. Unsupervised clustering was performed by PAM, available in the R package "cluster" (44), with k being equal to 2 (i.e., 2 clusters, where $k = 2$ produced the maximum average silhouette width) (see Fig. S1 in the supplemental material). Importantly, the application of the same method of unsupervised clustering to only the cytokines and chemokines considered strictly inflammatory gave clustering results (97% overlap in clustering) very similar to those using all markers; we decided to use all 47 markers based on the strong distinction seen across all markers (Fig. S1) between the two groups referred to here as inflammation-H and -L.

Cells were harvested from cytobrushes *ex vivo*, and cellular markers were assessed by flow cytometry as previously described (45). Briefly, a cervical cytobrush was collected, transported to the laboratory at 4°C, and processed within 4 h of collection. An eight-color polychromatic flow cytometry panel was used to measure the levels of expression of the HIV coreceptor CCR5 (CCR5-allophycocyanin [APC]; BD Biosciences, Franklin Lakes, NJ, USA), cellular activation (CD38-phycoerythrin [PE]-Cy7 and HLA-DR-PE; BD Biosciences, Franklin Lakes, NJ, USA), and proliferation (Ki67-fluorescein isothiocyanate [FITC]; BD Biosciences, Franklin Lakes, NJ, USA) in CD4⁺ and CD8⁺ T cells isolated from cervical mononuclear cells (CMCs) (CD3-APC-H7 [BD Biosciences], CD4-peridinin chlorophyll protein [PerCP]-Cy5.5 [BD Pharmingen, San Diego, CA, USA], and CD8-Qdot605 [Invitrogen, Carlsbad, CA, USA]). A dump channel was included in the panel to exclude all dead cells (using the live/dead marker ViViD-Pacific Blue [PacBlue]; Invitrogen, Carlsbad, CA, USA), B cells (CD14-PacBlue; BD Pharmingen, San Diego, CA, USA), and monocytes (CD19-PacBlue; Invitrogen, Carlsbad, CA, USA).

For the quantitation of endogenous hormone levels, blood was collected in a serum-separating tube and tested at the National Health Laboratory Services (NHLS) using a competitive electrochemiluminescence immunoassay (ECLIA) (Cobas; Roche Diagnostics, Switzerland) to measure estradiol (E2), progesterone, and LH levels.

Amplification and sequencing of the V4 region of the 16S rRNA gene. All available visit 1 samples with sufficient DNA concentrations were processed for 16S rRNA gene amplification by PCR with subsequent amplicon sequencing ($n = 185$). Swabs were thawed; treated with a cocktail of mutanolysin (25 kU/ml; Sigma-Aldrich), lysozyme (450 kU/ml; Sigma-Aldrich), and lysostaphin (4 kU; Sigma-Aldrich); and then mechanically disrupted with a bead beater (Thermo Savant FastPrep 120 cell disrupter system) 3 times for 30 s at speed setting m/s 5.5 (equivalent to 5,500 rpm). DNA was extracted by using the MoBio PowerSoil DNA extraction kit (MoBio, Carlsbad, CA). While there are advantages and disadvantages to selecting the V1-V3 region versus the V4 region for 16S rRNA gene amplification (the V4 region allows an improved resolution of certain BV-associated bacteria such as *G. vaginalis*, while the V1-V3 region allows a better resolution of *Lactobacillus* spp.), we chose to amplify the V4 region for an improved resolution of BV-associated bacteria (which was a major interest in this study) and so that our results would be more directly comparable with those of relevant studies reported previously (22, 23). The V4 region of the 16S rRNA gene was amplified by using the following universal primers that were modified to encode the Illumina MiSeq sequencing primer sequence at the 5' end (46): 515F (TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG NNN NNG TGC CAG CMG CCG CGG TAA) and 806R (GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GNN NNN GGA CTA CHV GGG TWT CTA AT), where NNNNN indicates five randomly incorporated nucleotides for increased complexity. The 5' end (underlined) is the Illumina Nextera adapter, and the sequences following the N's are complementary to the V4 rRNA gene region.

Quality was checked by using a Qubit fluorometer (Invitrogen). Pooled duplicate samples were purified with AMPure XP beads (Beckman Coulter, Brea, CA, USA) and quantified by using the PicoGreen double-stranded DNA (dsDNA) assay (Invitrogen, Carlsbad, CA, USA). Dual indices and Illumina sequencing adapters were attached by using the Nextera XT DNA Prep kit (Illumina). Samples were again purified by using AMPure XP beads, quantified by using a Qubit fluorometer (Invitrogen), and pooled for sequencing. Purified libraries consisting of ~96 pooled samples were paired-end sequenced on an Illumina MiSeq platform (300-bp paired-end reads with V3 chemistry).

Following demultiplexing, raw reads were preprocessed as follows: forward and reverse reads were merged by using usearch7 (47), allowing a maximum of three mismatches; merged reads were quality filtered by using usearch7 (reads with E scores of >0.1 were discarded); primer sequences were removed by using a custom python script; and merged, filtered reads were truncated at 250 bp. Next, sequences were dereplicated while recording the level of replication for each sequence by using usearch7. Dereplicated sequences were sorted by abundance (highest to lowest) and clustered *de novo* into operational taxonomic units (OTUs) at 97% similarity using usearch7. Chimeric sequences were detected (against the Gold database) by using UCHIME (48) and removed. Individual sequences were assigned to the specific identifiers using a 97% similarity threshold. Taxonomic assignment was performed with QIIME 1.8.0 (49) using the RDP classifier (using the default confidence level of 0.5) against the Greengenes 13.8 reference taxonomy for 97% identity. To increase species-level resolution, we constructed a custom taxonomic database appropriate for V4 region 16S rRNA gene amplicon sequencing based on the custom vaginal 16S rRNA gene reference database created by Fettweis et al. (50). The database created

by Fettweis et al. was described in 2012 and was originally designed for the taxonomic annotation of V1-V3 16S rRNA gene amplicons; entries of <1,300 bp (which might exclude the V4 region) were therefore first replaced with the corresponding full 16S sequence, from the same strain where possible. Similarly, full-genome entries were present for certain species, which were replaced with the corresponding 16S sequence to keep the database as small as possible. A minority of entries had subsequently been updated with new accession numbers in the NCBI database and were updated accordingly in our custom vaginal 16S reference database. A detailed list of the updates made to the database of Fettweis et al. can be found in Table S2 in the supplemental material.

We used the `usearch_global` command implemented in VSEARCH (51) to search our representative sequences of *de novo*-picked OTUs against the updated custom vaginal 16S reference database, accepting all hits with $\geq 97\%$ identity (Table S3). Several taxonomic annotations made by using this custom vaginal 16S reference database conflicted with the corresponding Greengenes annotations; we therefore performed BLAST searches against the NCBI nucleotide database (excluding uncultured organisms) for all OTUs with hits against the custom vaginal 16S reference database and manually curated the results. OTUs that mapped to more than one species (with the same identity score) were annotated as follows: if an OTU mapped to two or three species, the OTU would be named genus species A_species B or genus species A_species B_species C, respectively, and if an OTU mapped to more than three species but one species was clearly associated with vaginal microbiota (based on prior knowledge), the OTU was named genus species_cluster, where "species" was selected based on the majority of hits; e.g., *L. reuteri*_cluster indicates the case where the majority of hits were for *L. reuteri* but there were several other species with equal identity scores present. A full list of the taxonomic annotations for the 149 OTUs (out of 376) that had hits in the vaginal 16S reference database (50) can be found in Table S4. Importantly, OTUs annotated as *Shuttleworthia* using the Greengenes database were incorrect; these assignments were all made based on Greengenes OTU 2277000, which is actually BVAB1, as evidenced from alignments against the custom vaginal 16S reference database and the NCBI nucleotide database. This incorrect annotation was also noted previously by Noecker et al. (52). BVAB1/2/3 is not currently annotated in commonly used 16S rRNA gene amplicon databases, including Greengenes and SILVA. Several studies previously described *Shuttleworthia* in the vaginal microbiota, which in many cases may instead be more appropriately annotated BVAB1 or other uncultured bacteria related to *Lachnospiraceae*. In our cohort, OTU_4 had 100% identity with BVAB1 (GenBank accession number [GQ046543.1](https://www.ncbi.nlm.nih.gov/nuccore/GQ046543.1)), while the rest of the OTUs originally annotated as *Shuttleworthia* had 95.3 to 96.8% identity to BVAB1: these were annotated *Lachnospiraceae_uncultured_clone*.

Samples with $\geq 5,000$ reads were selected for downstream analyses. The OTU table was standardized (i.e., transformed to relative abundance and multiplied by the median sample read depth) and filtered so that each OTU had at least 10 counts in at least 2% of samples or a relative abundance of at least 0.001%.

Microbiota functional prediction. Microbiota functional prediction was performed on 16S rRNA gene sequence data by using PICRUSt (53). PICRUSt requires closed reference OTUs to perform matching against an existing database with Greengenes identifiers. Our *de novo* OTU table was therefore filtered to exclude OTUs that did not map to Greengenes 13.8 identifications (with $\geq 97\%$ identity), and *de novo* identifications for the remaining matching OTUs were replaced with their corresponding Greengenes OTU identifications. The standardized, filtered OTU table was submitted to the command line version of PICRUSt; Greengenes 13.5 KEGG Orthology (KO) terms were used for PICRUSt predictions. First, the OTU table was normalized to account for 16S rRNA gene copy numbers (using the PICRUSt script `normalize_by_copy_number.py`). Next, metagenomes were predicted by using the PICRUSt script `predict_metagenomes.py`; individual KEGG genes were collapsed to higher hierarchies (e.g., pathways) by using the PICRUSt `categorize_by_function.py` script. Finally, contributions by taxa were calculated by using the PICRUSt script `metagenome_contributions.py`. NSTI scores were calculated as a measure of the quality/accuracy of predictions made. NSTI scores provide a measure of how closely microorganisms in a given sample are related to sequenced genomes. The median NSTI score was 0.05, which indicates that the majority of bacteria within our samples could be predicted by using a relative from the same (95%) species/genus.

Statistical analyses. All downstream statistical analyses were performed in R, using the packages phyloseq (54) for beta diversity analyses, MetagenomeSeq (55) for differential abundance testing, vegan (56) for ordinations and redundancy analysis, and NMF (57) for annotated heat maps. Microbiota compositional and functional subtypes were established by fuzzy clustering using the R package cluster (44), with k being equal to 3, using 16S rRNA gene sequence count data and the PICRUSt-inferred community metagenome as the inputs, respectively. Clustering parameters were as follows: the membership exponent was 1.25, the dissimilarity measure was the Bray-Curtis measure, and members with a <60% probability of belonging to any of the three clusters were excluded from downstream analyses ($n = 13$). Comparison with the subtypes described previously by Anahar et al. was made as described previously (23); briefly, samples were classified as being *L. iners*, *Lactobacillus* (excluding *L. iners*), *Gardnerella*, or *Prevotella* dominant based on which of these taxa were dominant in each sample.

Differences in microbial compositions between groups of interest were assessed by using the MetagenomeSeq MRFulltable function with a custom filter to determine significance: merged taxa were deemed significantly different if they exhibited a fold change (beta coefficient) of ≥ 1.25 and had an adjusted P value of ≤ 0.05 and if at least one of the two groups being compared had $\geq 20\%$ of samples with the given OTU/taxon, or, if the result of Fisher's exact test was significant (after multiple-testing correction [MTC] by the Benjamini-Hochberg method [58]), OTUs were first merged at the lowest available taxonomic level (i.e., for OTUs with *Lactobacillus* as the lowest available taxonomic annotation,

counts were summed, while OTUs with additional species-level annotation, e.g., *L. iners*, were summed at the species level instead). For identifying taxa significantly different between inflammation-H versus -L, HC use (any versus none/male condoms) was included as a covariate in the model.

Differences in predicted functions were assessed by using STAMP (26), which employs Welch's *t* test for two-group comparisons (entries with adjusted *P* values of ≤ 0.05 , ratios of proportions of ≥ 1.5 , and differences in mean proportions of ≥ 0.1 were deemed significant). Random forest analyses were conducted on merged OTUs to determine which taxa best predicted (i) inflammation-H versus -L and (ii) functional subtype F3 versus F2, using the R packages randomForest (59) and ROCR (60) for receiver operating characteristic (ROC) analysis. For random forest analyses, the data were randomly divided into training and test sets, comprising one-third and two-thirds of the data, respectively. Differences in individual cytokines between groups were assessed by using Kruskal-Wallis tests with Dunn's *post hoc* test.

Data availability. Raw sequence data for 16S rRNA gene amplicon sequences are available at <http://www.ebi.ac.uk/> under BioProject accession number PRJEB15497. R analysis scripts are available upon request.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/IAI.00410-17>.

SUPPLEMENTAL FILE 1, PDF file, 1.0 MB.

ACKNOWLEDGMENTS

This study was supported by grants from the European and Developing Countries Clinical Trials Partnership (EDCTP) Strategic Primer (SP-2011-41304-038) and the South African Department of Science and Technology (DST/CON 0260/2012). K.L. is supported by the Harry Crossley Foundation and the Suid-Afrikaanse Akademie vir Wetenskap en Kuns. H.B.J. is supported in part by grant K08HD069201. S.L.B. was supported by the HIV Vaccine Trials Network SHAPe Program, the Fogarty Foundation, and the South African Medical Research Council (MRC). S.D. was supported by the National Research Foundation of South Africa. The Desmond Tutu HIV Foundation (DTHF) also recognizes support from Viiv Health Care in their YouthShield program. The Perinatal HIV Research Unit (PHRU) was supported through funding from the South African Medical Research Council. A.B. is supported in part by the National Science Foundation GRFP and GROW with USAID.

We thank the WISH Study Teams, particularly Pinky Ngobo, Sr.; Nozipho Hadebe, Sr.; Janine Nixon; and all the young women who kindly participated in the study. We thank Lynn Morris and Raveshni Durgiah from the National Institute for Communicable Diseases for their help with sample processing. Computations were performed by using facilities provided by the University of Cape Town ICTS High Performance Computing team (<http://hpc.uct.ac.za/>).

H.B.J., J.-A.S.P., and K.L. conceived of and designed the experiments. G.G., S.L.B., and L.-G.B. recruited the cohort. S.D., S.L.B., A.B., E.H., and S.Z.J. performed wet-lab experiments. D.A.L. performed STI and BV testing. K.L., S.D., J.-A.S.P., and G.B. analyzed the data. K.L., J.-A.S.P., H.B.J., S.L.B., S.D., D.A.L., L.-G.B., A.B., E.H., N.M., N.N.M., S.Z.J., G.G., and G.B. wrote the manuscript.

REFERENCES

- Vannuccini S, Clifton VL, Fraser IS, Taylor HS, Critchley H, Giudice LC, Petraglia F. 2016. Infertility and reproductive disorders: impact of hormonal and inflammatory mechanisms on pregnancy outcome. *Hum Reprod Update* 22:104–115. <https://doi.org/10.1093/humupd/dmv044>.
- Romero R, Dey SK, Fisher SJ. 2014. Preterm labor: one syndrome, many causes. *Science* 345:760–765. <https://doi.org/10.1126/science.1251816>.
- Masson L, Passmore JS, Liebenberg LJ, Werner L, Baxter C, Arnold KB, Williamson C, Little F, Mansoor LE, Naranbhai V, Lauffenburger DA, Ronacher K, Walz G, Garrett NJ, Williams BL, Couto-Rodriguez M, Hornig M, Lipkin WI, Grobler A, Abdool Karim Q, Abdool Karim SS. 2015. Genital inflammation and the risk of HIV acquisition in women. *Clin Infect Dis* 61:260–269. <https://doi.org/10.1093/cid/civ298>.
- Deese J, Masson L, Miller W, Cohen M, Morrison C, Wang M, Ahmed K, Agot K, Crucitti T, Abdellati S, Van Damme L. 2015. Injectable progestin-only contraception is associated with increased levels of pro-inflammatory cytokines in the female genital tract. *Am J Reprod Immunol* 74:357–367. <https://doi.org/10.1111/aji.12415>.
- Morrison CS, Chen P-L, Kwok C, Baeten JM, Brown J, Crook AM, Van Damme L, Delany-Moretlwe S, Francis SC, Friedland BA, Hayes RJ, Hefron R, Kapiga S, Karim QA, Karpoff S, Kaul R, McClelland RS, McCormack S, McGrath N, Myer L, Rees H, van der Straten A, Watson-Jones D, van de Wijgert JHHM, Stalter R, Low N. 2015. Hormonal contraception and the risk of HIV acquisition: an individual participant data meta-analysis. *PLoS Med* 12:e1001778. <https://doi.org/10.1371/journal.pmed.1001778>.
- Ralph LJ, McCoy SI, Shiu K, Padian NS. 2015. Hormonal contraceptive use and women's risk of HIV acquisition: a meta-analysis of observational studies. *Lancet Infect Dis* 15:181–189. [https://doi.org/10.1016/S1473-3099\(14\)71052-7](https://doi.org/10.1016/S1473-3099(14)71052-7).

7. Masson L, Misana K, Little F, Werner L, Mkhize NN, Ronacher K, Gamiel-dien H, Williamson C, McKinnon LR, Walz G, Abdool Karim Q, Abdool Karim SS, Passmore J-AS. 2014. Defining genital tract cytokine signatures of sexually transmitted infections and bacterial vaginosis in women at high risk of HIV infection: a cross-sectional study. *Sex Transm Infect* 90:580–587. <https://doi.org/10.1136/sextrans-2014-051601>.
8. Reddy BS, Rastogi S, Das B, Salhan S, Verma S, Mittal A. 2004. Cytokine expression pattern in the genital tract of Chlamydia trachomatis positive infertile women—implication for T-cell responses. *Clin Exp Immunol* 137:552–558. <https://doi.org/10.1111/j.1365-2249.2004.02564.x>.
9. Fichorova RN, Desai PJ, Gibson FC, III, Genco CA. 2001. Distinct proinflammatory host responses to Neisseria gonorrhoeae infection in immortalized human cervical and vaginal epithelial cells. *Infect Immun* 69:5840–5848. <https://doi.org/10.1128/IAI.69.9.5840-5848.2001>.
10. Misana K, Naicker N, Werner L, Roberts L, Van Loggerenberg F, Baxter C, Passmore JAS, Grobler AC, Sturm AW, Williamson C, Ronacher K, Walz G, Abdool Karim SS. 2012. Symptomatic vaginal discharge is a poor predictor of sexually transmitted infections and genital tract inflammation in high-risk women in South Africa. *J Infect Dis* 206:6–14. <https://doi.org/10.1093/infdis/jis298>.
11. Wiesenfeld HC, Hillier SL, Krohn MA, Landers DV, Sweet RL. 2003. Bacterial vaginosis is a strong predictor of Neisseria gonorrhoeae and Chlamydia trachomatis infection. *Clin Infect Dis* 36:663–668. <https://doi.org/10.1086/367658>.
12. Gallo MF, Macaluso M, Warner L, Fleenor ME, Hook EW, Brill I, Weaver MA. 2012. Bacterial vaginosis, gonorrhoea, and chlamydial infection among women attending a sexually transmitted disease clinic: a longitudinal analysis of possible causal links. *Ann Epidemiol* 22:213–220. <https://doi.org/10.1016/j.annepidem.2011.11.005>.
13. Balkus JE, Richardson BA, Rabe LK, Taha TE, Mgodini N, Kasaro MP, Ramjee G, Hoffman IF, Abdool Karim SS. 2014. Bacterial vaginosis and the risk of Trichomonas vaginalis acquisition among HIV-1-negative women. *Sex Transm Dis* 41:123–128. <https://doi.org/10.1097/OLQ.0000000000000075>.
14. Taha TE, Hoover DR, Dallabetta GA, Kumwenda NI, Mtiamvalye LA, Yang LP, Liomba GN, Broadhead RL, Chipangwi JD, Miotti PG. 1998. Bacterial vaginosis and disturbances of vaginal flora: association with increased acquisition of HIV. *AIDS* 12:1699–1706. <https://doi.org/10.1097/00002030-199813000-00019>.
15. Myer L, Denny L, Telerant R, de Souza M, Wright TC, Kuhn L. 2005. Bacterial vaginosis and susceptibility to HIV infection in South African women: a nested case-control study. *J Infect Dis* 192:1372–1380. <https://doi.org/10.1086/462427>.
16. Atashili J, Poole C, Ndumbe PM, Adimora AA, Smith JS. 2008. Bacterial vaginosis and HIV acquisition: a meta-analysis of published studies. *AIDS* 22:1493–1501. <https://doi.org/10.1097/QAD.0b013e3283021a37>.
17. Srinivasan S, Hoffman NG, Morgan MT, Matsen FA, Fiedler TL, Hall RW, Ross FJ, McCoy CO, Bumgarner R, Marrazzo JM, Fredricks DN. 2012. Bacterial communities in women with bacterial vaginosis: high resolution phylogenetic analyses reveal relationships of microbiota to clinical criteria. *PLoS One* 7:e37818. <https://doi.org/10.1371/journal.pone.0037818>.
18. Buvé A, Jaspers V, Crucitti T, Fichorova RN. 2014. The vaginal microbiota and susceptibility to HIV. *AIDS* 28:2333–2344. <https://doi.org/10.1097/QAD.0000000000000432>.
19. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SSK, Mcculle SL, Ault K, Peralta L, Forney LJ. 2010. Vaginal microbiome of reproductive-age women. *Proc Natl Acad Sci U S A* 108:4680–4687. <https://doi.org/10.1073/pnas.1002611107>.
20. Kenyon C, Colebunders R, Crucitti T. 2013. The global epidemiology of bacterial vaginosis: a systematic review. *Am J Obstet Gynecol* 209:505–523. <https://doi.org/10.1016/j.ajog.2013.05.006>.
21. Thurman AR, Kimble T, Herold B, Mesquita PM, Fichorova RN, Dawood HY, Fashemi T, Chandra N, Rabe L, Cunningham TD, Anderson S, Schwartz J, Doncel G. 2015. Bacterial vaginosis and subclinical markers of genital tract inflammation and mucosal immunity. *AIDS Res Hum Retroviruses* 31:1139–1152. <https://doi.org/10.1089/aid.2015.0006>.
22. Gosmann C, Anahtar MN, Handley SA, Walker BD, Virgin HW, Kwon DS, Gosmann C, Anahtar MN, Handley SA, Farcasanu M, Abu-Ali G, Bowman BA. 2017. Lactobacillus-deficient cervicovaginal bacterial communities are associated with increased HIV acquisition in young South African women. *Immunity* 46:29–37. <https://doi.org/10.1016/j.immuni.2016.12.013>.
23. Anahtar MN, Byrne EH, Doherty KE, Bowman BA, Yamamoto HS, Soumillon M, Padavattan N, Ismail N, Moodley A, Sabatini ME, Ghebremichael MS, Nusbaum C, Huttenhower C, Virgin HW, Ndung'u T, Dong KL, Walker BD, Fichorova RN, Kwon DS. 2015. Cervicovaginal bacteria are a major modulator of host inflammatory responses in the female genital tract. *Immunity* 42:965–976. <https://doi.org/10.1016/j.immuni.2015.04.019>.
24. Mitchell C, Marrazzo J. 2014. Bacterial vaginosis and the cervicovaginal immune response. *Am J Reprod Immunol* 71:555–563. <https://doi.org/10.1111/aji.12264>.
25. Human Microbiome Project Consortium. 2012. Structure, function and diversity of the healthy human microbiome. *Nature* 486:207–214. <https://doi.org/10.1038/nature11234>.
26. Parks DH, Tyson GW, Hugenholtz P, Beiko RG. 2014. STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics* 30:3123–3124. <https://doi.org/10.1093/bioinformatics/btu494>.
27. Toole GAO, Kolter R. 1998. Flagellar and twitching motility are necessary for Pseudomonas aeruginosa biofilm development. *Mol Microbiol* 30:295–304. <https://doi.org/10.1046/j.1365-2958.1998.01062.x>.
28. Hapgood JP, Koubovec D, Louw A, Africander D. 2004. Not all progestins are the same: implications for usage. *Trends Pharmacol Sci* 25:554–557. <https://doi.org/10.1016/j.tips.2004.09.005>.
29. Polis CB, Curtis KM, Hannaford PC, Phillips SJ, Chipato T, Kiarie JN, Westrich DJ, Steyn PS. 2016. An updated systematic review of epidemiological evidence on hormonal contraceptive methods and HIV acquisition in women. *AIDS* 30:2665–2683. <https://doi.org/10.1097/QAD.0000000000001228>.
30. Ziklo N, Huston WM, Hocking JS, Timms P. 2016. Chlamydia trachomatis genital tract infections: when host immune response and the microbiome collide. *Trends Microbiol* 24:750–765. <https://doi.org/10.1016/j.tim.2016.05.007>.
31. Fettweis JM, Brooks JP, Serrano MG, Sheth NU, Girerd PH, Edwards DJ, Strauss JF, Jefferson KK, Buck GA. 2014. Differences in vaginal microbiome in African American women versus women of European ancestry. *Microbiology* 160:2272–2282. <https://doi.org/10.1099/mic.0.081034-0>.
32. Nunn KL, Wang Y, Harit D, Humphrys MS, Ma B, Cone R, Ravel J. 2015. Enhanced trapping of HIV-1 by human cervicovaginal mucus is associated with Lactobacillus crispatus-dominant microbiota. *mBio* 6:e01084-15. <https://doi.org/10.1128/mBio.01084-15>.
33. Shen J, Song N, Williams CJ, Brown CJ, Yan Z, Xu C, Forney LJ. 2016. Effects of low dose estrogen therapy on the vaginal microbiomes of women with atrophic vaginitis. *Sci Rep* 6:24380. <https://doi.org/10.1038/srep24380>.
34. Smith SM, Baskin GB, Marx PA. 2000. Estrogen protects against vaginal transmission of simian immunodeficiency virus. *J Infect Dis* 182:708–715. <https://doi.org/10.1086/315776>.
35. Smith SM, Mefford M, Sodora D, Klase Z, Singh M, Alexander N, Hess D, Marx PA. 2004. Topical estrogen protects against SIV vaginal transmission without evidence of systemic effect. *AIDS* 18:1637–1643. <https://doi.org/10.1097/01.aids.0000131393.76221.cc>.
36. Anipindi VC, Bagri P, Roth K, Dizzell SE, Nguyen PV. 2016. Estradiol enhances CD4+ T-cell anti-viral immunity by priming vaginal DCs to induce Th17 responses via an IL-1-dependent pathway. *PLoS Pathog* 12:e1005589. <https://doi.org/10.1371/journal.ppat.1005589>.
37. Srinivasan S, Morgan MT, Fiedler TL, Djukovic D, Hoffman NG, Raftery D, Marrazzo JM, Fredricks N. 2015. Metabolic signatures of bacterial vaginosis. *mBio* 6:e00204-15. <https://doi.org/10.1128/mBio.00204-15>.
38. Swidsinski A, Mendling W, Loening-Baucke V, Swidsinski S, Dörffel Y, Scholze J, Lochs H, Verstraelen H. 2008. An adherent Gardnerella vaginalis biofilm persists on the vaginal epithelium after standard therapy with oral metronidazole. *Am J Obstet Gynecol* 198:97.e1–97.e6. <https://doi.org/10.1016/j.ajog.2007.06.039>.
39. Swidsinski A, Mendling W, Loening-Baucke V, Ladhoff A, Swidsinski S, Hale LP, Lochs H. 2005. Adherent biofilms in bacterial vaginosis. *Obstet Gynecol* 106:1013–1023. <https://doi.org/10.1097/01.AOG.0000183594.45524.d2>.
40. Wood TK, Gonzalez Barrios AF, Herzberg M, Lee J. 2006. Motility influences biofilm architecture in Escherichia coli. *Appl Microbiol Biotechnol* 72:361–367. <https://doi.org/10.1007/s00253-005-0263-8>.
41. Marrazzo JM, Thomas KK, Fiedler TL, Ringwood K, Fredricks D. 2008. Relationship of specific vaginal bacteria and bacterial vaginosis treatment failure in women who have sex with women. *Ann Intern Med* 149:20–28. <https://doi.org/10.7326/0003-4819-149-1-20080710-00006>.
42. Lewis DA, Müller E, Steele L, Sternberg M, Radebe F, Lyall M, Ballard RC, Paz-Bailey G. 2012. Prevalence and associations of genital ulcer and urethral pathogens in men presenting with genital ulcer syndrome to

- primary health care clinics in South Africa. *Sex Transm Dis* 39:880–885. <https://doi.org/10.1097/OLQ.0b013e318269cf90>.
43. Jacobs MV, Snijders PJ, van den Brule AJ, Helmerhorst TJ, Meijer J, Walboomers JM. 1997. A general primer GP5⁺/GP6⁺-mediated PCR-enzyme immunoassay method for rapid detection of 14 high-risk and 6 low-risk human papillomavirus genotypes in cervical scrapings. *J Clin Microbiol* 35:791–795.
 44. Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K. 2016. cluster: cluster analysis basics and extensions. R package version 2.0.4. R Foundation for Statistical Computing, Vienna, Austria.
 45. Jaspan HB, Liebenberg L, Hanekom W, Burgers W, Coetzee D, Williamson AL, Little F, Myer L, Coombs RW, Sodora D, Passmore JA. 2011. Immune activation in the female genital tract during HIV infection predicts mucosal CD4 depletion and HIV shedding. *J Infect Dis* 204:1550–1556. <https://doi.org/10.1093/infdis/jir591>.
 46. Pearce MM, Hilt EE, Rosenfeld AB. 2014. The female urinary microbiome: a comparison of women with and without urgency urinary incontinence. *mBio* 5:e01283-14. <https://doi.org/10.1128/mBio.01283-14>.
 47. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>.
 48. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. 2011. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27:2194–2200. <https://doi.org/10.1093/bioinformatics/btr381>.
 49. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7:335–336. <https://doi.org/10.1038/nmeth.f.303>.
 50. Fettweis JM, Serrano MG, Sheth NU, Mayer CM, Glascock AL, Brooks JP, Jefferson KK, Buck GA. 2012. Species-level classification of the vaginal microbiome. *BMC Genomics* 13(Suppl 8):S17.
 51. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. 2016. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4:e2584. <https://doi.org/10.7717/peerj.2584>.
 52. Noecker C, Eng A, Srinivasan S, Theriot CM, Young VB, Jansson JK, Fredricks DN, Borenstein E. 2016. Microbiome taxonomic and metabolomic profiles elucidates mechanistic links between ecological and metabolic variation. *mSystems* 1:e00013-15. <https://doi.org/10.1128/mSystems.00013-15>.
 53. Langille MGI, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepille DE, Vega Thurber RL, Knight R, Beiko RG, Huttenhower C. 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol* 31: 814–821. <https://doi.org/10.1038/nbt.2676>.
 54. McMurdie PJ, Holmes S. 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8:e61217. <https://doi.org/10.1371/journal.pone.0061217>.
 55. Paulson JN, Stine OC, Bravo HC, Pop M. 2013. Differential abundance analysis for microbial marker-gene surveys. *Nat Methods* 10:1200–1202. <https://doi.org/10.1038/nmeth.2658>.
 56. Oksanen AJ, Blanchet FG, Kindt R, Legendre P, Minchin PR, Hara RBO, Simpson GL, Solymos P, Stevens MHH, Wagner H. 2016. Package “vegan.” R Foundation for Statistical Computing, Vienna, Austria.
 57. Gaujoux R. 2014. Generating heatmaps for nonnegative matrix factorization. R Foundation for Statistical Computing, Vienna, Austria.
 58. Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc* 57: 289–300.
 59. Liaw A, Wiener M. 2002. Classification and regression by randomForest. *R News* 2(3):18–22.
 60. Sing T, Sander O, Beerenwinkel N, Lengauer T. 2005. ROCr: visualizing classifier performance in R. *Bioinformatics* 21:3940–3941. <https://doi.org/10.1093/bioinformatics/bti623>.
 61. Barnabas S, Dabee S, Passmore J, Jaspan HB, Lewis D, Jaumdally SZ, Gamielidien H, Masson L, Nuller E, Maseko V, MKhize N, Mbulawa Z, Williamson AL, Gray CM, Hope TJ, Chiodi F, Dietrich J, Gray G, Bekker LG for the Women’s Initiative in Sexual Health (WISH) Study Team. *Int J STD AIDS*, in press.
 62. Mbulawa ZZ, Coetzee D, Williamson AL. 2015. Human papillomavirus prevalence in South African women and men according to age and human immunodeficiency virus status. *BMC Infect Dis* 15:459. <https://doi.org/10.1186/s12879-015-1181-8>.