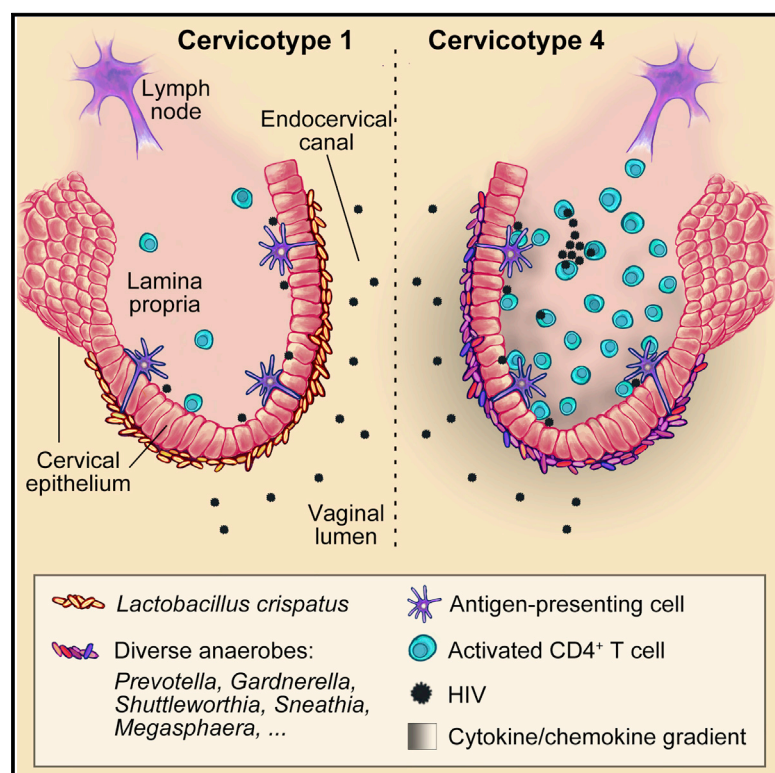# *Lactobacillus*-Deficient Cervicovaginal Bacterial Communities Are Associated with Increased HIV Acquisition in Young South African Women

Graphical Abstract

Authors

Christina Gosmann, Melis N. Anahtar, Scott A. Handley, ..., Bruce D. Walker, Herbert W. Virgin, Douglas S. Kwon

Correspondence

dkwon@mgh.harvard.edu

In Brief

The potential impact of cervicovaginal bacteria on HIV susceptibility is not well-defined. Gosmann et al. (2017) identify anaerobic cervicovaginal bacterial communities and specific taxa highly prevalent in young healthy South African women that increase their HIV risk. These findings might be leveraged to reduce HIV acquisition in women.

Highlights

- Women with high-diversity genital bacterial communities acquire HIV at 4× higher rates

- Activated mucosal CD4[+] T cell numbers are elevated in women with high-risk bacteria

- Specific genital bacterial taxa are linked with reduced or elevated HIV acquisition

- I.vag administered *Prevotella* increases activated genital CD4[+] T cell numbers in mice

CellPress

Lactobacillus-Deficient Cervicovaginal Bacterial Communities Are Associated with Increased HIV Acquisition in Young South African Women

Christina Gosmann,[1,2] Melis N. Anahtar,[1,2] Scott A. Handley,[3] Mara Farcasanu,[1] Galeb Abu-Ali,[4] Brittany A. Bowman,[1] Nikita Padavattan,[5] Chandni Desai,[3] Lindsay Droit,[3] Amber Moodley,[6] Mary Dong,[1,6] Yuezhou Chen,[7] Nasreen Ismail,[5] Thumbi Ndung'u,[1,5] Musie S. Ghebremichael,[1,2] Duane R. Wesemann,[7] Caroline Mitchell,[8] Krista L. Dong,[1,6] Curtis Huttenhower,[4] Bruce D. Walker,[1,9,10,11] Herbert W. Virgin,[3] and Douglas S. Kwon[1,2,10,12,*]

[1]Ragon Institute of MGH, MIT, and Harvard, Massachusetts General Hospital, Cambridge, MA 02139, USA
[2]Harvard Medical School, Boston, MA 02115, USA
[3]Department of Pathology and Immunology, Washington University School of Medicine, Saint Louis, MO 63110, USA
[4]Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115, USA
[5]HIV Pathogenesis Programme, Nelson R. Mandela School of Medicine, University of KwaZulu-Natal, Durban, KwaZulu-Natal, 4001, South Africa
[6]Females Rising through Education, Support, and Health, Durban, KwaZulu-Natal, 4066, South Africa
[7]Department of Medicine, Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA
[8]Vincent Obstetrics & Gynecology, Massachusetts General Hospital, Boston, MA 02115, USA
[9]Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA
[10]Division of Infectious Diseases, Massachusetts General Hospital, Boston, MA 02115, USA
[11]Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[12]Lead Contact
*Correspondence: dkwon@mgh.harvard.edu
http://dx.doi.org/10.1016/j.immuni.2016.12.013

SUMMARY

Elevated inflammation in the female genital tract is associated with increased HIV risk. Cervicovaginal bacteria modulate genital inflammation; however, their role in HIV susceptibility has not been elucidated. In a prospective cohort of young, healthy South African women, we found that individuals with diverse genital bacterial communities dominated by anaerobes other than *Gardnerella* were at over 4-fold higher risk of acquiring HIV and had increased numbers of activated mucosal CD4+ T cells compared to those with *Lactobacillus crispatus*-dominant communities. We identified specific bacterial taxa linked with reduced (*L. crispatus*) or elevated (*Prevotella*, *Sneathia*, and other anaerobes) inflammation and HIV infection and found that high-risk bacteria increased numbers of activated genital CD4+ T cells in a murine model. Our results suggest that highly prevalent genital bacteria increase HIV risk by inducing mucosal HIV target cells. These findings might be leveraged to reduce HIV acquisition in women living in sub-Saharan Africa.

INTRODUCTION

24 million people in sub-Saharan Africa are infected with HIV, with the majority of new transmission events occurring following heterosexual sex. Young African women have an up to 8-fold increased HIV prevalence compared to young men (UNAIDS, 2014), emphasizing the need for a better understanding of the factors in the female genital tract (FGT) that influence HIV acquisition.

Elevated inflammation is associated with an increased risk of HIV infection (Masson et al., 2015), and cervicovaginal bacteria have been shown to impact baseline inflammation in the FGT (Anahtar et al., 2015), suggesting a potential role for FGT bacteria in modulating acquisition risk. A small number of prospective studies indicate that women with cervicovaginal bacteria deficient in lactobacilli, a bacterial genus considered beneficial for vaginal health (Spurbeck and Arvidson, 2011), are at higher risk of acquiring HIV (Low et al., 2011; Martin et al., 1999; Myer et al., 2005). However, the depth of these studies has been limited by the use of Gram-stained vaginal smears to assess the genital bacterial composition and by investigation of sex-worker and clinical cohorts not representative of the average healthy female population.

The recent development of high-throughput sequencing technologies to characterize the human microbiome has significantly improved our understanding of the role of the microbiome in diseases such as obesity, inflammatory bowel disease, colorectal cancer (Marchesi et al., 2016), and HIV-induced AIDS (Handley et al., 2016; Monaco et al., 2016). While these studies have focused on the enteric microbiome, less is understood regarding the role of the cervicovaginal microbiome in disease susceptibility, particularly in the developing world. Additionally, the majority of microbiome studies have focused on bacteria while viruses have remained incompletely characterized, particularly in the FGT.

Here we performed bacterial 16S ribosomal (r)RNA gene and viral shotgun sequencing of cervicovaginal microbiota and assessed their role in HIV acquisition in participants of the FRESH
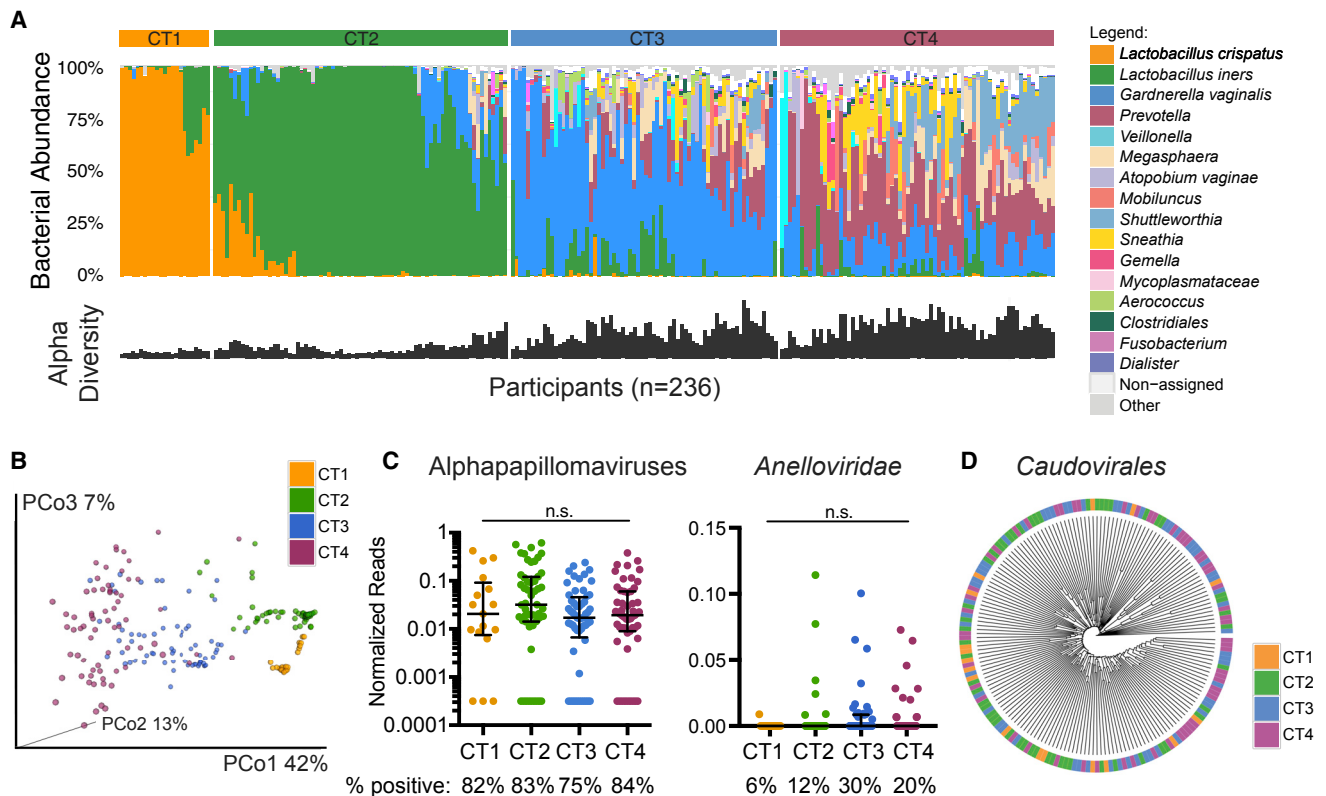
Figure 1. Bacterial, but Not Viral, Cervicovaginal Community Structures Are Distinct in HIV-Uninfected Women

(A) Stacked bar plot indicating the bacterial composition of the cervicovaginal microbiome of 236 HIV-uninfected women. Samples were grouped according to 4 cervicotypes (CTs), and the alpha diversity is shown below for each sample.

(B) Principal coordinates analysis (PCoA) plot of the 236 individuals, distinguished by CT.

(C) Normalized reads of alphapapillomaviruses and *Anelloviridae* detected in CVL fluid, grouped by CT (n = 180 individuals). Groups were compared using Kruskal-Wallis test. The percentage of samples in which the indicated virus was detected is shown below.

(D) Average linkage clustering of the Bray-Curtis dissimilarity of relative abundances of *Caudovirales* sequences detected in CVL fluid (n = 180 individuals). Lines in the plots indicate the median and the interquartile range (IQR) of the dataset.

(Females Rising through Education, Support and Health) study, a prospective cohort study of healthy HIV-uninfected black South African women aged 18–23 years, monitored with high frequency testing for incident HIV infection. We demonstrated a differential association of four distinct bacterial community structures with genital inflammation, cervical HIV target cell numbers, and HIV acquisition. We additionally identified specific bacterial taxa that were closely associated with genital inflammation and HIV acquisition and showed that these same bacteria induced inflammation in in vitro and in vivo models, thus establishing their mechanistic potential for directly increasing HIV acquisition.

RESULTS

South African Women Have Distinct Bacterial but Not Viral Cervicovaginal Communities

We followed 236 HIV-uninfected women prospectively with a median follow-up time of 336 days (IQR: 178.5 to 347 days) and a total follow-up time of 198.2 person-years. Despite intensive HIV prevention counseling, 31 women acquired HIV while 205 remained HIV-uninfected with an overall HIV incidence in the FRESH cohort of 8.43 per 100 person-years.

Sequencing of the V4 region of the bacterial 16S rRNA gene in these 236 individuals demonstrated the presence of distinct bacterial community types that, based on composition and diversity, clustered into 4 basic groups, which we referred to as "cervicotypes" (CTs). Similar to our previous findings in a smaller cohort (Anahtar et al., 2015), 10% of women had low diversity communities dominated by *Lactobacillus crispatus* (CT1, n = 23), and 32% were dominated by *Lactobacillus iners* (CT2, n = 74) (Figure 1A). This contrasts with white women in developed countries, 90% of whom have *Lactobacillus*-dominant cervicovaginal communities (Ravel et al., 2011). The cervicovaginal bacterial microbiome of the remaining 58% of FRESH participants consisted of high-diversity communities with low *Lactobacillus* abundance, with either *Gardnerella vaginalis* dominance (CT3, n = 68) or a bacterial genus other than *Lactobacillus* or *Gardnerella* as the dominant taxon (CT4, n = 70). The most abundant taxa in CT4 were *Prevotella*, *Gardnerella*, *Shuttleworthia*, *Sneathia*, and *Megasphaera* (Figure 1A). Principal coordinates analysis confirmed distinct clustering of these CT groups (Figure 1B).

Given the paucity of information regarding the community of viruses (the "virome") in the FGT, we used metagenomic sequencing to characterize DNA and RNA viral populations in
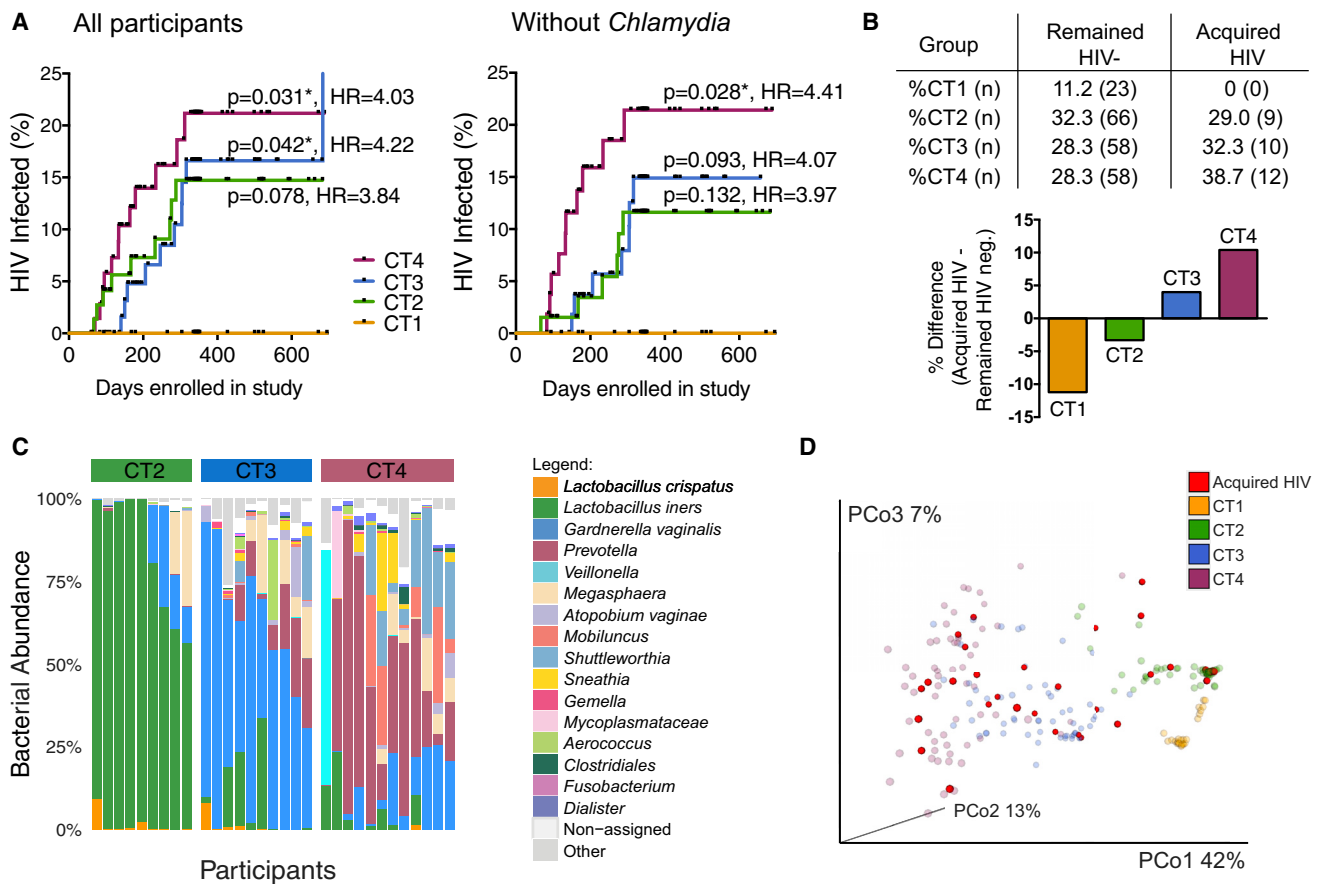
Figure 2. Cervicovaginal Bacteria Impact HIV Acquisition

(A) Kaplan-Meier curve showing HIV infections in each CT group over time, including or excluding individuals with chlamydia infection. Acquisition curves for CT2 (n = 74/65 individuals), CT3 (n = 68/57), and CT4 (n = 68/54) were compared with the acquisition curve for CT1 (n = 23/23). Log-rank (Mantel-Cox) test-based p values and Mantel-Haenszel hazard ratios (HR) are displayed.

(B) Distribution of CT groups in participants who remained HIV-uninfected (n = 205) versus those who acquired HIV (n = 31), shown as % of total, and the difference in distribution between these two groups.

(C) Stacked bar plot indicating the bacterial composition of the cervicovaginal microbiome of 31 women who subsequently acquired HIV. Samples were grouped by CT.

(D) Bacterial PCoA plot indicating study participants who acquired HIV (in red, n = 31) and CT groups for those who remained HIV-uninfected (faded coloring, n = 205).

See also Figure S1 and Table S1.

180 study participants. We identified alphapapillomaviruses and *Anelloviridae* as the predominant eukaryotic viral taxa present in the FGT in this cohort (Figure 1C). Numerous species of *Caudovirales*, the tailed bacteriophages, were also observed (Figure 1D). In contrast to our findings for bacteria, we identified no distinct viral community structures within our cohort, nor did we detect a difference in viral composition among different bacterial communities (Figures 1C and 1D). In summary, we observed four distinct cervicovaginal bacterial community types, the most prevalent ones being characterized by high diversity and low *Lactobacillus* abundance and no distinct viral community structures in the study participants.

HIV Acquisition Is Increased in Women with High-Diversity, Low *Lactobacillus* Abundance FGT Bacterial Communities

We next examined associations between cervicovaginal bacteria and HIV acquisition. We observed a greater than 4-fold increase

in HIV acquisition in women with high-diversity, low *Lactobacillus* abundance communities compared to women with *L. crispatus* dominance (CT4: HR = 4.03, 95% CI: 1.14 to 14.27, p = 0.031, and CT3: HR = 4.22, 95% CI: 1.06 to 16.88, p = 0.042) (Figure 2A). None of the women who acquired HIV had an *L. crispatus* dominant community. Women with CT2 communities were also underrepresented among participants who subsequently acquired HIV relative to the uninfected group, while those with CT3 and CT4 communities were more prevalent among women who went on to become infected (Figure 2B). The genital bacterial communities of women who subsequently acquired HIV were comparable to those of their respective CT group who remained HIV negative (Figure 2C, Figure S1A), with no acquisition-specific clustering within a CT (Figure 2D), suggesting a similar HIV infection risk for women within the same CT group. In contrast to cervicovaginal bacteria, viral taxa identified in this study were not different between women who became infected and those who remained uninfected (Figure S1B).
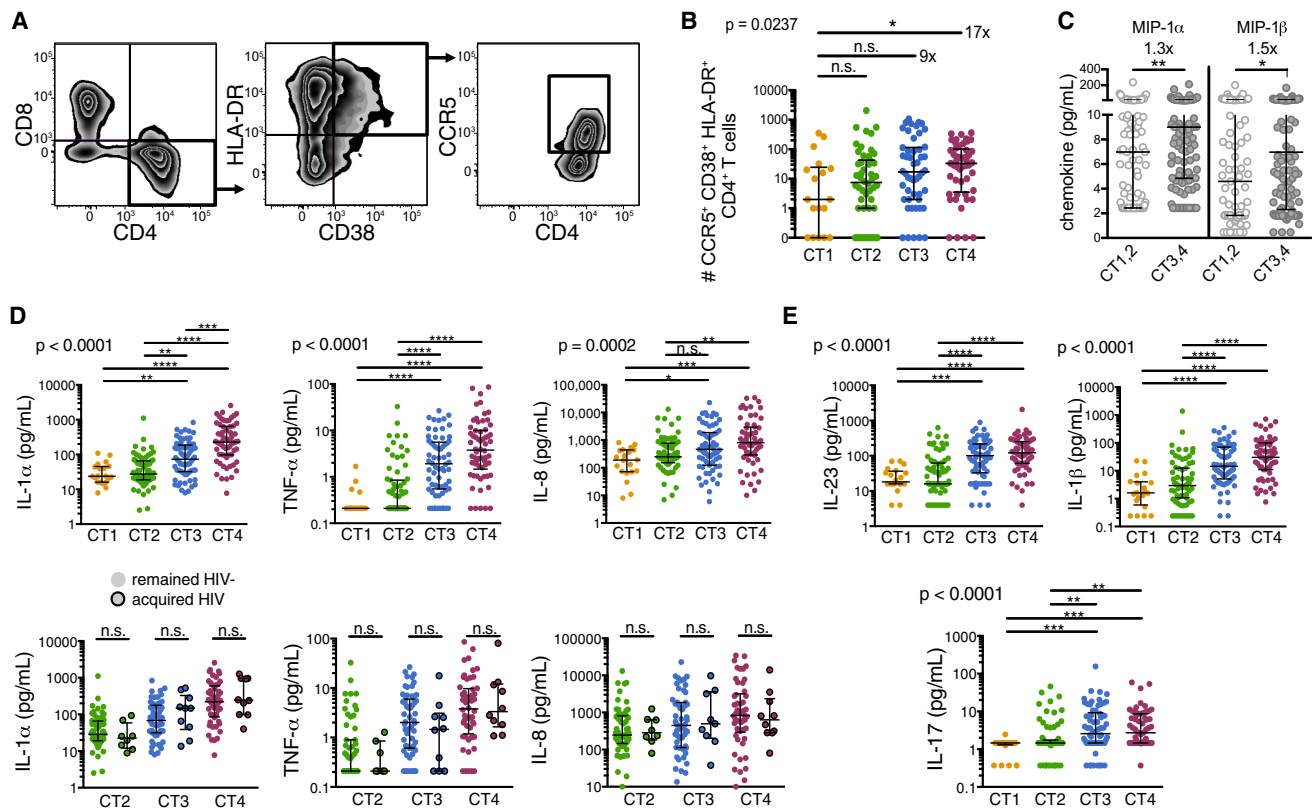
Figure 3. High-Risk Genital Bacterial Communities Are Associated with Increased HIV Target Cell Numbers and Pro-Inflammatory Cytokines

(A) Representative flow cytometry plots to identify HIV target cells isolated from cervical cytobrushes. Live T cells (CD45$^+$CD3$^+$) were gated on cells expressing CD4, HLA-DR, CD38, and CCR5.

(B) Flow cytometry analysis of HIV target cell numbers in cytobrushes from 169 individuals, grouped by CT.

(C–E) Chemokine and cytokine concentrations in CVL fluid from 219 individuals, grouped by CT. The bottom figure panel in (D) further distinguishes between samples of women who became infected with HIV and those who remained HIV-uninfected.

Lines indicate the median and the interquartile range (IQR) of each dataset. Groups were compared using Kruskal-Wallis test with Dunn's post hoc analyses. See also Figures S2 and S3.

We next assessed factors associated with different genital bacterial communities in our cohort, with particular attention to those that might increase HIV susceptibility (Byrne et al., 2016; McKinnon and Karim, 2016; Ward and Rönn, 2010) (Table S1). We observed no differences between bacterial communities and demographic factors, condom use, frequency or type of sexual acts, or number of sexual partners. The use of injectable progestin contraceptives (IPCs), a known HIV risk factor in the FRESH cohort, was also similar among women with different CTs, which is consistent with our previous finding that cervicovaginal bacterial communities did not vary with IPC use (Byrne et al., 2016). Sexually transmitted infections (STIs) were equally distributed among all groups, except for an increase in *Chlamydia trachomatis* infection from CT1 through CT4 (Table S1). After removing all *Chlamydia*-positive samples from the analysis, HIV acquisition in women with CT4 but not CT3 communities was still significantly increased compared to women with *L. crispatus*-dominant communities (HR: 4.41, 95% CI: 1.17 to 16.61, p = 0.028) (Figure 2A). Similar results were obtained after adjusting for chlamydia infection. Collectively, these results demonstrate that women with high-diversity anaerobic communities have a more than 4-fold increase in HIV acquisition compared to women with an *L. crispatus*-dominant cervicovaginal microbiome.

Women with Anaerobic FGT Bacteria Have Increased Genital HIV Target Cell Numbers and Th17-Associated Cytokines

In most women, HIV infection is established following viral replication in CD4$^+$ T cells present in the genital mucosa (Haase, 2011). Activated CD4$^+$ T cells expressing the HIV co-receptor CCR5 ("HIV target cells") are not only more prone to infection but also support higher degrees of viral replication than resting cells (Meditz et al., 2011; Stevenson et al., 1990), thus increasing HIV susceptibility (Haase, 2011; Koning et al., 2005). We examined the relationship between the bacterial microbiome and HIV target cell numbers in the genital tract and observed a 17-fold increase in HIV target cells in the FGT of women with CT4 versus *L. crispatus*-dominant communities (Figures 3A and 3B; Figure S2). Consistent with this finding, we demonstrated elevated levels of the chemokines MIP-1α and MIP-1β, which attract CCR5-expressing cells, in women with high-diversity bacterial communities (Figure 3C). We also observed higher concentrations of cytokines and chemokines associated with

increased HIV acquisition (Masson et al., 2015), such as interleukin-1α (IL-1α), tumor necrosis factor-α (TNF-α), IL-8 (Figure 3D; upper row), IL-12p70, interferon-γ (IFN-γ), and IL-10 (Figure S3A), which confirmed our previous findings in a smaller cohort (Anahtar et al., 2015). Women who went on to become infected with HIV had similar pro-inflammatory cytokine levels as those who remained HIV-uninfected within the same CT group. This suggests comparable HIV susceptibility among women with similar genital communities (Figure 3D; lower row).

We further examined cytokines associated with specific T helper (Th) subsets as recent work suggests that Th17 cells are the first targets of SIV and HIV infection in the FGT (Rodriguez-Garcia et al., 2014; Stieh et al., 2016). While we detected no or little difference in Th2-associated cytokines IL-4, IL-5, and IL-13 between CT groups (Figure S3B), we measured a notable increase in IL-17 and IL-17-inducing cytokines IL-23 and IL-1β in the cervicovaginal lavage (CVL) fluid of women with CT3 and CT4 structures (Figure 3E; Figure S3E). This finding suggested that increased numbers of Th17 cells, the predominant producers of IL-17 in the genital mucosa (Gosmann et al., 2014; Stieh et al., 2016), were induced in the presence of CT3 and CT4 bacteria. Taken together, these results reveal an association of high-diversity genital bacterial communities with increased production of local pro-inflammatory cytokines and recruitment and activation of mucosal CD4+ T cells.

Specific Bacterial Taxa Are Differentially Associated with Genital Inflammation and HIV Acquisition

We next investigated whether specific bacterial taxa were associated with genital inflammation and HIV infection. We found that *L. crispatus* and, to a lesser extent, *L. iners* were associated with reduced inflammation as indicated by an inverse correlation with the first cytokine principal component (PC1), which explained 51% of the variation in cytokines (Figure 4A, Figures S3C and S3D). In contrast, anaerobic taxa were associated with increased inflammation, the five taxa most strongly correlated with cytokine PC1 being *Megasphaera*, *Clostridium*, *Prevotella*, *Atopobium vaginae*, and *Sneathia* (Figure 4A). Consistent with this finding, cervical epithelial cells produced higher concentrations of IL-6 and IL-8 when co-cultured with *Gardnerella vaginalis*, *Prevotella bivia*, and *Prevotella amnii* compared to *L. crispatus* (Figure 4B). To more closely examine bacterial taxa associated with HIV acquisition, we identified resolved sequence variants (RSV) from 16S rRNA gene amplicons. Non-*iners* lactobacilli were significantly more abundant in women who remained HIV-uninfected (Figure 4C). In contrast, *Prevotella melaninogenica*, *Veillonella montpellierensis*, *Mycoplasma*, *Prevotella bivia*, and *Sneathia sanguinegens*, most of which were associated with increased genital inflammation, were all positively associated with HIV acquisition (Figure 4C). In summary, we identified specific bacterial taxa associated with reduced (e.g., *L. crispatus*) or increased (e.g., *Prevotella* and *Sneathia*) inflammation and HIV acquisition.

Vaginal Inoculation of Germ-Free Mice with Anaerobic Bacteria Increases Numbers of Activated CD4+ T cells in the FGT

To more definitively assess a mechanistic role for these bacteria in inducing increased numbers of FGT HIV target cells, we inoc-ulated germ-free mice intravaginally with *L. crispatus* or *P. bivia* (Figures 4D and 4E). We detected higher numbers of activated (CD44+) CD4+ T cells in the genital mucosa, but not in the blood, of mice inoculated with *P. bivia* compared to mice that received *L. crispatus* (Figure 4E). Mice treated with *P. bivia* further had increased numbers of mucosal CCR5+ CD4+ T cells (Figure S4A) and showed a trend toward elevated genital TNF-α concentrations (Figure S4B) compared to mice inoculated with *L. crispatus*. Together, these results indicate that specific bacterial taxa may increase HIV infection risk by inducing elevated numbers of HIV target cells in the FGT.

DISCUSSION

Here we presented a comprehensive characterization of the genital bacterial and viral microbiome in healthy young women in sub-Saharan Africa, and identified a significant association of distinct cervicovaginal bacterial communities and specific bacterial taxa with HIV acquisition. The sequencing-based identification of cervicovaginal bacteria and viruses provides a substantial advancement over previous prospective studies that used Gram-stained vaginal smears (Low et al., 2011; Martin et al., 1999; Myer et al., 2005).

Our results indicate that high-diversity, low *Lactobacillus* abundance bacterial communities in the FGT negatively impact vaginal health by increasing the risk of HIV acquisition. In addition, prior studies have reported associations between *Lactobacillus*-deficient genital microbiota and other poor reproductive outcomes such as preterm delivery, late miscarriage (Hay et al., 1994; Lamont et al., 2011), and cervicitis (Gorgos et al., 2015). The increased rate of HIV acquisition in women with high-diversity, low *Lactobacillus* abundance bacterial communities observed in our study could be explained by the sensing of specific bacterial antigens or components such as lipopolysaccharide (LPS); we previously demonstrated that CT4 communities were enriched in LPS relative to CT1, and that LPS signaling pathways were significantly upregulated in cervical antigen-presenting cells of women with high-diversity communities (Anahtar et al., 2015). This likely results in activation and recruitment of HIV target cells to the female genital mucosa. In support of this hypothesis, we show a 17-fold increase in HIV target cell numbers in the FGT mucosa of women with CT4 versus CT1 communities. This finding is further corroborated by the presence of increased numbers of activated CD4+ T cells in the genital mucosa of germ-free mice intravaginally inoculated with *P. bivia* compared to mice inoculated with *L. crispatus*, providing direct evidence of the impact of these bacteria on the recruitment of activated CD4+ T cells to the FGT. Increased concentrations of Th17-inducing cytokines IL-23 and IL-1β, and of IL-17 in the presence of CT3 and CT4 bacteria further indicated the presence of elevated numbers of Th17 cells, which are particularly susceptible to HIV infection and have been described as the first cells to be infected by lentivirus in the FGT (Christensen-Quick et al., 2016; Stieh et al., 2016). While the development of the Th17 subset in human epithelial tissues is not fully understood, in murine models, Th17 cells are specifically induced by bacteria in the gut (Ivanov et al., 2009). Thus, the bacterial composition of the FGT microbiome might influence HIV susceptibility not only by increasing local HIV target cell numbers, but also by
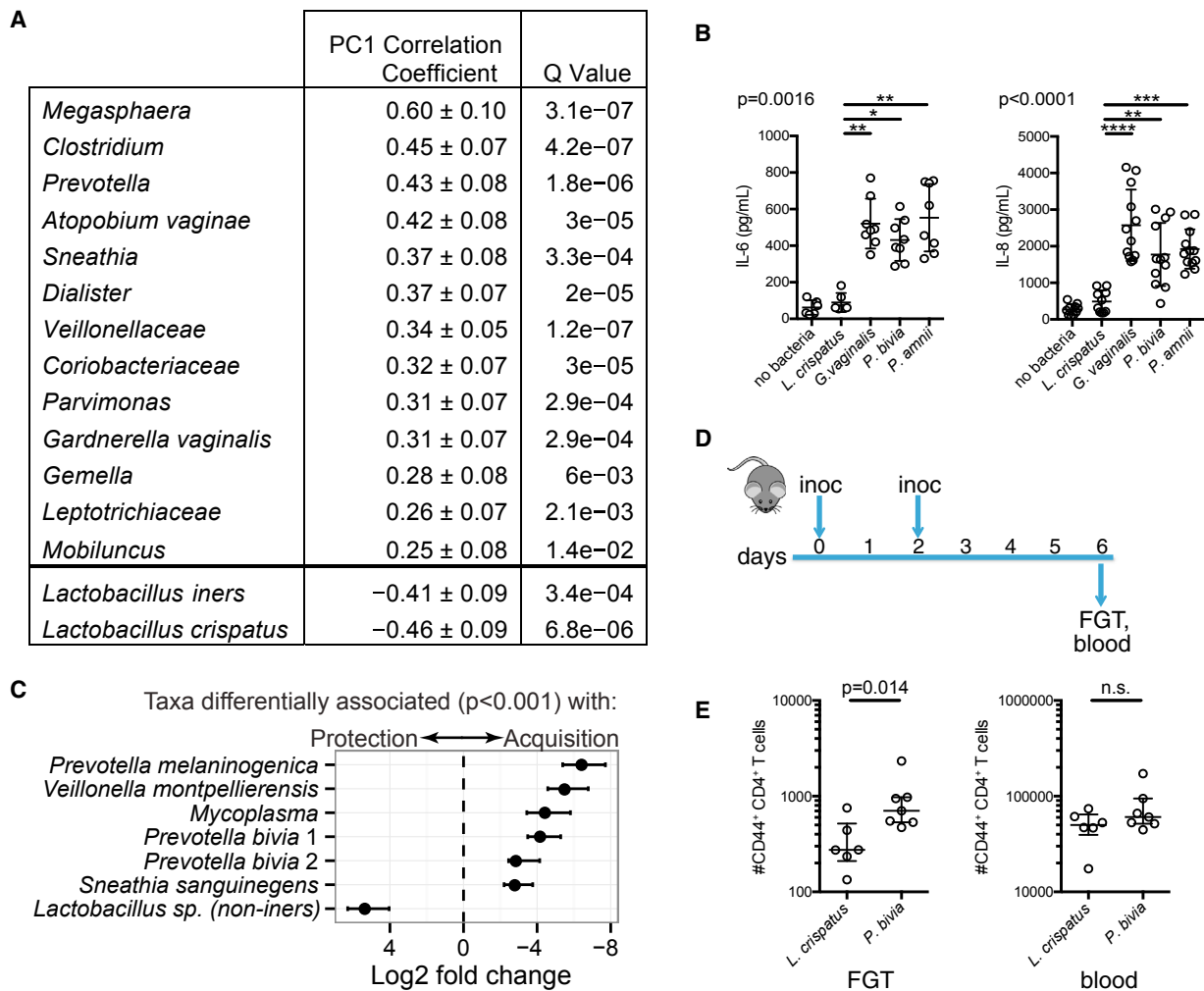
A

	PC1 Correlation Coefficient	Q Value
Megasphaera	0.60 ± 0.10	3.1e−07
Clostridium	0.45 ± 0.07	4.2e−07
Prevotella	0.43 ± 0.08	1.8e−06
Atopobium vaginae	0.42 ± 0.08	3e−05
Sneathia	0.37 ± 0.08	3.3e−04
Dialister	0.37 ± 0.07	2e−05
Veillonellaceae	0.34 ± 0.05	1.2e−07
Coriobacteriaceae	0.32 ± 0.07	3e−05
Parvimonas	0.31 ± 0.07	2.9e−04
Gardnerella vaginalis	0.31 ± 0.07	2.9e−04
Gemella	0.28 ± 0.08	6e−03
Leptotrichiaceae	0.26 ± 0.07	2.1e−03
Mobiluncus	0.25 ± 0.08	1.4e−02
Lactobacillus iners	−0.41 ± 0.09	3.4e−04
Lactobacillus crispatus	−0.46 ± 0.09	6.8e−06

B

C

Taxa differentially associated (p<0.001) with:

D

E

Figure 4. Specific Bacterial Taxa Are Associated with Genital Inflammation and HIV Infection

(A) Bacterial taxa in 219 individuals positively or negatively correlated with cytokine principal component 1 (PC1), an indicator of the presence of genital inflammation. Q values reflect corrected p values after multiple testing.

(B) Cytokine and chemokine concentrations measured in the supernatant of cervical epithelial cells co-cultured with the indicated bacteria for 24 hr (pooled results from 2 or 3 independent experiments; Kruskal-Wallis test).

(C) Bacterial taxa significantly differentially abundant in women who became infected with HIV (n = 31) and those who remained uninfected (n = 205), including two strains of *P. bivia*. Error bars indicate the standard error estimate for the log2 fold change in abundance.

(D) Germ-free mice were intravaginally inoculated with *L. crispatus* or *P. bivia* on day 0 and day 2 and sacrificed on day 6.

(E) Numbers of activated (CD44$^+$) CD4$^+$ T cells detected in the female genital tract (FGT) and in 100 μl blood of the gnotobiotic mice on day 6 of the experiment (pooled results from 3 independent experiments; Mann-Whitney test).

Lines in the plots indicate the median and the interquartile range (IQR) of the dataset.

See also Figure S4.

influencing T helper cell differentiation within the genital mucosa. Whether FGT bacteria contribute to shaping mucosal T helper cell subsets will need to be addressed in future studies.

Reduced HIV acquisition in women with *L. crispatus*-dominant communities might be based on their ability to impair the growth of inflammatory bacteria and pathogens through the production of lactic acid, hydrogen peroxide, and bactericidal compounds (Spurbeck and Arvidson, 2011), thus diminishing inflammation and HIV target cell numbers. Consistent with this hypothesis, *L. crispatus* did not co-exist with the more diverse anaerobic communities associated with HIV acquisition. In contrast, *L. iners* was frequently found in CT3 and CT4 communities,

and 29% of the women who became infected with HIV had *L. iners*-dominant communities while none had *L. crispatus* dominance. This observation might further be explained by reduced HIV virion diffusion rates in the mucus from women with *L. crispatus* dominated vaginal microbiota compared to women with *L. iners* dominance (Nunn et al., 2015). These findings identify *L. crispatus* as a *Lactobacillus* species associated with decreased HIV infection and highlight the importance of distinguishing between different lactobacilli in the context of HIV susceptibility. Our findings further indicate that individuals within a particular CT group are at similar risk of acquiring HIV. This suggests that more than half the population of young black

South African women is at increased HIV infection risk based on the composition of their cervicovaginal microbiome.

The absence of distinct viral signatures in women with different bacterial communities suggests a primary role for bacteria in modulating HIV acquisition risk. However, it is possible that a greater depth of sequencing, or expansion of viral databases will reveal associations that were not observed here. The stability in bacteriophage community structure across CT groups further suggests that the observed bacterial communities were not being modified by bacteriophage predation, a mechanism that could explain the formation of distinct bacterial communities in different individuals.

Clinical studies assessing the efficacy of regimens such as antibiotics and probiotic *Lactobacillus* vaginal suppositories in modifying the genital microbiome of women with bacterial vaginosis showed significant recurrence rates (Bradshaw et al., 2012; Woodman, 2016). Our findings underscore the importance of developing more effective regimens that achieve a sustained alteration of the genital microbiome since this could potentially reduce HIV acquisition. In this context, strategies targeting specific bacterial taxa associated with reduced or increased HIV acquisition might be useful, although the host mechanisms that maintain specific microbial communities in the genital tract need to be better understood. Our results advance our understanding of the cervicovaginal microbiome as an HIV risk factor and demonstrate the importance of considering the microbiome in the development of new treatments and preventive strategies to reduce HIV acquisition in young women living in sub-Saharan Africa.

EXPERIMENTAL PROCEDURES

Study Cohort
Study participants were enrolled in the Females Rising through Education, Support, and Health (FRESH) study, a prospective observational study of 18- to 23-year-old HIV uninfected women conducted in Umlazi, South Africa. At the time of data analysis, pre-infection mucosal samples from 31 women who became HIV-infected, and samples from 205 women who remained uninfected were available. The study protocol was approved by the Biomedical Research Ethics Committee of the University of KwaZulu-Natal and the Massachusetts General Hospital Institutional Review Board (2012P001812/MGH). Informed consent was obtained following explanation of the nature and possible consequences of the study.

Clinical Procedures
Twice per week, participants had a finger prick blood draw for HIV RNA viral load testing and every 3 months had a peripheral blood draw and a pelvic exam and completed an HIV risk questionnaire administered by a counselor. The pelvic exam involved the collection of ectocervical and midvaginal swabs (Catch-All, Epicenter), a cervicovaginal lavage (CVL), and an endocervical cytobrush.

Bacterial 16S rRNA V4 Gene Sequencing
Nucleic acid extracted from cervical swabs was amplified and sequenced as previously described (Anahtar et al., 2016), and analysis was performed using MacQIIME.

Definition of Cervicotypes
Samples with the relative majority of sequences assigned to the genus *Lactobacillus* (but not *Lactobacillus iners*) were defined as CT1. *Lactobacillus iners* was the predominant species in CT2, *Gardnerella vaginalis* in CT3, and CT4 had a dominant bacterial taxon other than *Lactobacillus*, *L. iners*, or *G. vaginalis*.

Virome Sequencing
Sequencing of viral DNA and RNA extracted from CVL fluid was performed as detailed in the Supplemental Experimental Procedures.

Sexually Transmitted Infection Testing
STI testing was performed by Global Labs, South Africa.

Measurement of Cytokines
Cytokine concentrations in CVL fluid were determined using a human cytokine/chemokine multiplexed bead assay (Millipore) and human and mouse ELISA (eBioscience, R&D Systems).

Flow Cytometry
Human and mouse cells were stained with a viability dye (Invitrogen) and labeled with monoclonal antibodies as specified in the Supplemental Experimental Procedures. Samples were acquired on a FACS Aria III or LSR II (BD). Data were analyzed using FlowJo Version 9.8.5 (FlowJo Enterprise).

Detection of Differentially Abundant Bacterial Taxa
Resolved sequence variants (RSV) from 16S amplicons were identified using the dada2 procedure (Callahan et al., 2016). Differentially abundant taxa were determined using the DESeq2 package (Love et al., 2014).

Co-Culture of Epithelial Cells and Bacteria
Human endocervical epithelial cells (End1/E6E7, ATCC) were co-cultured anaerobically with *Lactobacillus crispatus* (provided by Dr. David N. Fredricks), *Prevotella bivia* (ATCC 29303), *Prevotella amnii* (CCUG 53648T), or *Gardnerella vaginalis* (ATCC 14018) for 24 hr.

Mouse Experiments
All animal procedures were approved by the Animal Resources at Children's Hospital (ARCH) review committee (protocol # 16-08-3219R). CVL fluid was collected from 6- to 9-week-old female germ-free Swiss-Webster mice housed according to an out-of-the-isolator protocol (Faith et al., 2014), followed by application of 2×10^8 CFU/mL to the vaginal cavity. On the day of harvest, blood, CVL fluid, and, after CO_2 asphyxiation, the female reproductive tract (vagina, cervix) were collected, processed, and used for downstream analyses.

Statistics
For comparison of continuous data between two and more groups, Mann-Whitney test and Kruskal-Wallis test with Dunn's post hoc analyses, respectively, were used. The time to HIV acquisition was summarized using Kaplan-Meier curves, and significance was assessed by log-rank test. Fisher's exact test was performed for comparison of categorical data between two or more groups. Principal component analysis (PCA) was used to obtain summary measures for the multi-variable cytokine set and performed using FactoMineR package in R. Multivariate analysis of cytokine and bacterial data was performed using MaAsLin (https://huttenhower.sph.harvard.edu/maaslin) (Morgan et al., 2012) and the Benjamini-Hochberg method to adjust for multiple testing. P values are two-sided and indicated as follows: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$; n.s.: $p \geq 0.05$. The analyses were performed in Prism 6 (GraphPad) and R Studio.

AUTHOR CONTRIBUTIONS

C.G., M.N.A., and D.S.K. conceived the study and designed experiments. C.G., M.F., B.A.B., and M.N.A. performed the nucleic acid extraction from

REFERENCES

Anahtar, M.N., Byrne, E.H., Doherty, K.E., Bowman, B.A., Yamamoto, H.S., Soumillon, M., Padavattan, N., Ismail, N., Moodley, A., Sabatini, M.E., et al. (2015). Cervicovaginal bacteria are a major modulator of host inflammatory responses in the female genital tract. Immunity 42, 965–976.

Anahtar, M.N., Bowman, B.A., and Kwon, D.S. (2016). Efficient Nucleic Acid Extraction and 16S rRNA Gene Sequencing for Bacterial Community Characterization. J. Vis. Exp. 110, 53939.

Bradshaw, C.S., Pirotta, M., De Guingand, D., Hocking, J.S., Morton, A.N., Garland, S.M., Fehler, G., Morrow, A., Walker, S., Vodstrcil, L.A., and Fairley, C.K. (2012). Efficacy of oral metronidazole with vaginal clindamycin or vaginal probiotic for bacterial vaginosis: randomised placebo-controlled double-blind trial. PLoS ONE 7, e34540.

Byrne, E.H., Anahtar, M.N., Cohen, K.E., Moodley, A., Padavattan, N., Ismail, N., Bowman, B.A., Olson, G.S., Mabhula, A., Leslie, A., et al. (2016). Association between injectable progestin-only contraceptives and HIV acquisition and HIV target cell frequency in the female genital tract in South African women: a prospective cohort study. Lancet Infect. Dis. 16, 441–448.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J., and Holmes, S.P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nat. Methods 13, 581–583.

Christensen-Quick, A., Lafferty, M., Sun, L., Marchionni, L., DeVico, A., and Garzino-Demo, A. (2016). Human Th17 cells lack HIV-inhibitory RNases and are highly permissive to productive HIV infection. J. Virol. 90, 7833–7847.

Faith, J.J., Ahern, P.P., Ridaura, V.K., Cheng, J., and Gordon, J.I. (2014). Identifying gut microbe-host phenotype relationships using combinatorial communities in gnotobiotic mice. Sci. Transl. Med. 6, 220ra11.

Gorgos, L.M., Sycuro, L.K., Srinivasan, S., Fiedler, T.L., Morgan, M.T., Balkus, J.E., McClelland, S.R., Fredricks, D.N., and Marrazzo, J.M. (2015). Relationship of Specific Bacteria in the Cervical and Vaginal Microbiotas With Cervicitis. Sex. Transm. Dis. 42, 475–481.

Gosmann, C., Mattarollo, S.R., Bridge, J.A., Frazer, I.H., and Blumenthal, A. (2014). IL-17 suppresses immune effector functions in human papillomavirus-associated epithelial hyperplasia. J. Immunol. 193, 2248–2257.

Haase, A.T. (2011). Early events in sexual transmission of HIV and SIV and opportunities for interventions. Annu. Rev. Med. 62, 127–139.

Handley, S.A., Desai, C., Zhao, G., Droit, L., Monaco, C.L., Schroeder, A.C., Nkolola, J.P., Norman, M.E., Miller, A.D., Wang, D., et al. (2016). SIV Infection-Mediated Changes in Gastrointestinal Bacterial Microbiome and Virome Are Associated with Immunodeficiency and Prevented by Vaccination. Cell Host Microbe 19, 323–335.

Hay, P.E., Lamont, R.F., Taylor-Robinson, D., Morgan, D.J., Ison, C., and Pearson, J. (1994). Abnormal bacterial colonisation of the genital tract and subsequent preterm delivery and late miscarriage. BMJ 308, 295–298.

Ivanov, I.I., Atarashi, K., Manel, N., Brodie, E.L., Shima, T., Karaoz, U., Wei, D., Goldfarb, K.C., Santee, C.A., Lynch, S.V., et al. (2009). Induction of intestinal Th17 cells by segmented filamentous bacteria. Cell 139, 485–498.

Koning, F.A., Otto, S.A., Hazenberg, M.D., Dekker, L., Prins, M., Miedema, F., and Schuitemaker, H. (2005). Low-level CD4+ T cell activation is associated with low susceptibility to HIV-1 infection. J. Immunol. 175, 6117–6122.

Lamont, R.F., Nhan-Chang, C.L., Sobel, J.D., Workowski, K., Conde-Agudelo, A., and Romero, R. (2011). Treatment of abnormal vaginal flora in early pregnancy with clindamycin for the prevention of spontaneous preterm birth: a systematic review and metaanalysis. Am. J. Obstet. Gynecol. 205, 177–190.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550.

Low, N., Chersich, M.F., Schmidlin, K., Egger, M., Francis, S.C., van de Wijgert, J.H., Hayes, R.J., Baeten, J.M., Brown, J., Delany-Moretlwe, S., et al. (2011). Intravaginal practices, bacterial vaginosis, and HIV infection in women: individual participant data meta-analysis. PLoS Med. 8, e1000416.

Marchesi, J.R., Adams, D.H., Fava, F., Hermes, G.D., Hirschfield, G.M., Hold, G., Quraishi, M.N., Kinross, J., Smidt, H., Tuohy, K.M., et al. (2016). The gut microbiota and host health: a new clinical frontier. Gut 65, 330–339.

Martin, H.L., Richardson, B.A., Nyange, P.M., Lavreys, L., Hillier, S.L., Chohan, B., Mandaliya, K., Ndinya-Achola, J.O., Bwayo, J., and Kreiss, J. (1999). Vaginal lactobacilli, microbial flora, and risk of human immunodeficiency virus type 1 and sexually transmitted disease acquisition. J. Infect. Dis. 180, 1863–1868.

Masson, L., Passmore, J.A., Liebenberg, L.J., Werner, L., Baxter, C., Arnold, K.B., Williamson, C., Little, F., Mansoor, L.E., Naranbhai, V., et al. (2015). Genital inflammation and the risk of HIV acquisition in women. Clin. Infect. Dis. 61, 260–269.

McKinnon, L.R., and Karim, Q.A. (2016). Factors Driving the HIV Epidemic in Southern Africa. Curr. HIV/AIDS Rep. 13, 158–169.

Meditz, A.L., Haas, M.K., Folkvord, J.M., Melander, K., Young, R., McCarter, M., Mawhinney, S., Campbell, T.B., Lie, Y., Coakley, E., et al. (2011). HLA-DR+ CD38+ CD4+ T lymphocytes have elevated CCR5 expression and produce the majority of R5-tropic HIV-1 RNA in vivo. J. Virol. 85, 10189–10200.

Monaco, C.L., Gootenberg, D.B., Zhao, G., Handley, S.A., Ghebremichael, M.S., Lim, E.S., Lankowski, A., Baldridge, M.T., Wilen, C.B., Flagg, M., et al. (2016). Altered Virome and Bacterial Microbiome in Human Immunodeficiency Virus-Associated Acquired Immunodeficiency Syndrome. Cell Host Microbe 19, 311–322.

Morgan, X.C., Tickle, T.L., Sokol, H., Gevers, D., Devaney, K.L., Ward, D.V., Reyes, J.A., Shah, S.A., LeLeiko, N., Snapper, S.B., et al. (2012). Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol. 13, R79.

Myer, L., Kuhn, L., Stein, Z.A., Wright, T.C., Jr., and Denny, L. (2005). Intravaginal practices, bacterial vaginosis, and women's susceptibility to HIV infection: epidemiological evidence and biological mechanisms. Lancet Infect. Dis. 5, 786–794.

Nunn, K.L., Wang, Y.Y., Harit, D., Humphrys, M.S., Ma, B., Cone, R., Ravel, J., and Lai, S.K. (2015). Enhanced Trapping of HIV-1 by Human Cervicovaginal Mucus Is Associated with Lactobacillus crispatus-Dominant Microbiota. MBio 6, e01084–e15.

Ravel, J., Gajer, P., Abdo, Z., Schneider, G.M., Koenig, S.S., McCulle, S.L., Karlebach, S., Gorle, R., Russell, J., Tacket, C.O., et al. (2011). Vaginal microbiome of reproductive-age women. Proc. Natl. Acad. Sci. USA 108 (Suppl 1), 4680–4687.

Rodriguez-Garcia, M., Barr, F.D., Crist, S.G., Fahey, J.V., and Wira, C.R. (2014). Phenotype and susceptibility to HIV infection of CD4+ Th17 cells in the human female reproductive tract. Mucosal Immunol. 7, 1375–1385.

Spurbeck, R.R., and Arvidson, C.G. (2011). Lactobacilli at the front line of defense against vaginally acquired infections. Future Microbiol. 6, 567–582.

Stevenson, M., Stanwick, T.L., Dempsey, M.P., and Lamonica, C.A. (1990). HIV-1 replication is controlled at the level of T cell activation and proviral integration. EMBO J. 9, 1551–1560.

Stieh, D.J., Matias, E., Xu, H., Fought, A.J., Blanchard, J.L., Marx, P.A., Veazey, R.S., and Hope, T.J. (2016). Th17 Cells Are Preferentially Infected Very Early after Vaginal Transmission of SIV in Macaques. Cell Host Microbe 19, 529–540.

UNAIDS, U.J.P.o.H.A. (2014). The gap report 2014: Adolescent girls and young women.

Ward, H., and Rönn, M. (2010). Contribution of sexually transmitted infections to the sexual transmission of HIV. Curr. Opin. HIV AIDS 5, 305–310.

Woodman, Z. (2016). Can one size fit all? Approach to bacterial vaginosis in sub-Saharan Africa. Ann. Clin. Microbiol. Antimicrob. 15, 16.

Supplemental Information

Lactobacillus-Deficient Cervicovaginal Bacterial

Communities Are Associated with Increased HIV

Acquisition in Young South African Women

Christina Gosmann, Melis N. Anahtar, Scott A. Handley, Mara Farcasanu, Galeb Abu-Ali, Brittany A. Bowman, Nikita Padavattan, Chandni Desai, Lindsay Droit, Amber Moodley, Mary Dong, Yuezhou Chen, Nasreen Ismail, Thumbi Ndung'u, Musie S. Ghebremichael, Duane R. Wesemann, Caroline Mitchell, Krista L. Dong, Curtis Huttenhower, Bruce D. Walker, Herbert W. Virgin, and Douglas S. Kwon

Figure S1

A
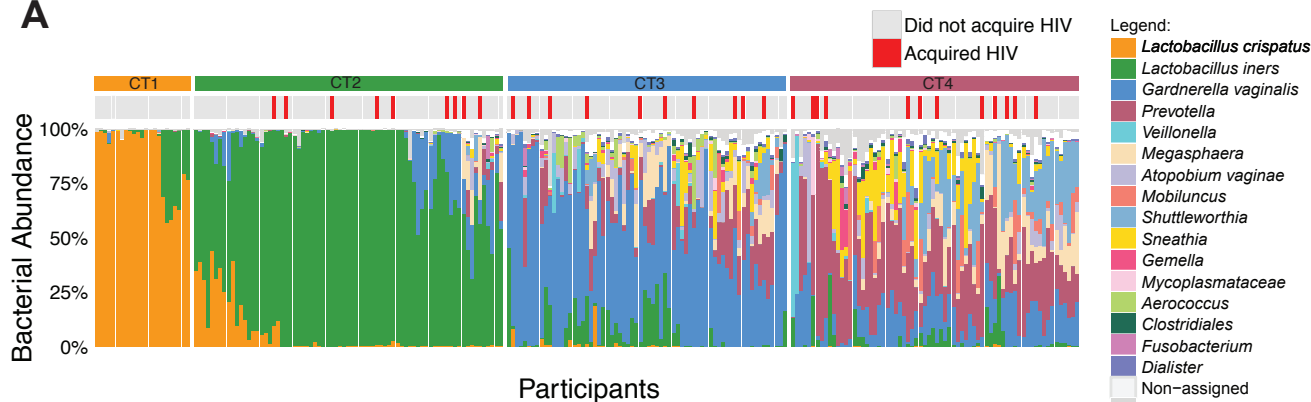
B
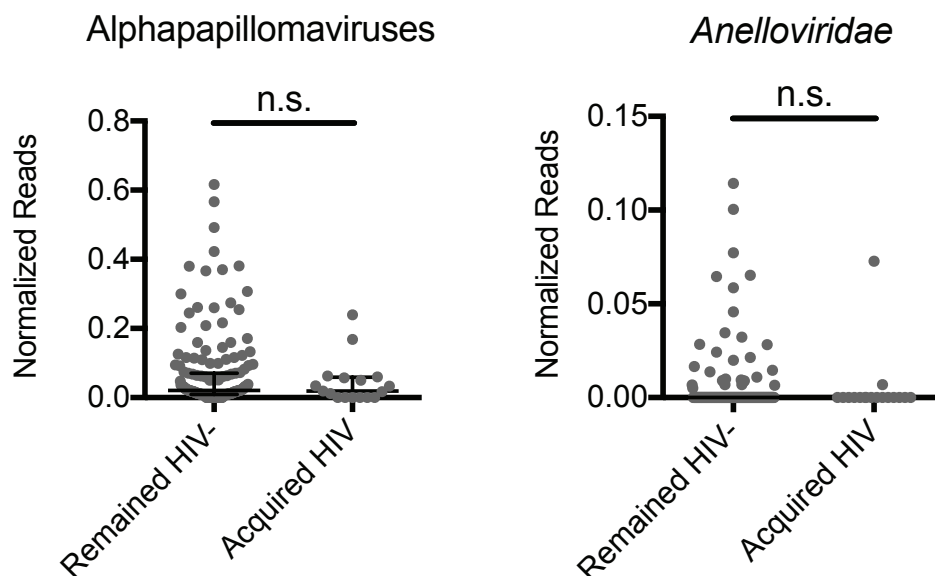
Figure S1. FGT bacteria and viruses in participants who subsequently acquired HIV. Related to Figure 2. (A) Stacked bar plot showing the bacterial composition of the cervicovaginal microbiome of 236 HIV-uninfected women. Participants who subsequently acquired HIV (n=31) are indicated. **(B)** Alphapapillomaviruses and *Anelloviridae* are not associated with HIV acquisition. Normalized reads of alphapapillomaviruses and *Anelloviridae*, grouped by women who remained HIV-uninfected (n=163) and those who became infected with HIV (n=17). Lines indicate the median and the interquartile range (IQR). Groups were compared using Mann-Whitney test.

Figure S2
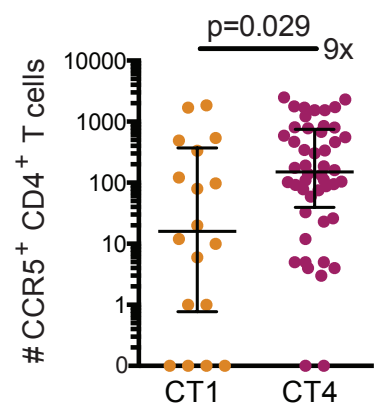
Figure S2. Increased mucosal CCR5⁺ CD4⁺ T cell numbers in participants with CT4 bacterial communities compared to individuals with CT1 communities. Related to Figure 3.

Flow cytometry analysis of cells isolated from cytobrushes from 66 individuals (n (CT1) = 18; n (CT4) = 48). Lines in the plots indicate the median and the interquartile range (IQR) of the dataset. Groups were compared by Mann-Whitney test.

Figure S3

A

B

C

D

	PC1	PC2
IFN-γ	0.27	0.01
IL-10	0.29	0.07
IL-12p70	0.30	0.01
IL-13	0.28	0.18
IL-17	0.26	0.04
IL-1α	0.14	-0.47
IL-1β	0.20	-0.30
IL-21	0.24	0.12
IL-23	0.29	-0.05
IL-4	0.25	0.23
IL-5	0.26	0.20
IP-10	-0.18	0.36
I-TAC	0.03	0.47
MIG	-0.24	0.31
MIP-1α	0.25	0.17
MIP-1β	0.24	0.19
TNF-α	0.26	-0.18

E

Figure S3. Additional cytokines measured in CVL fluid, and PCA loadings. Related to Figure 3.

(A, B) Chemokine and cytokine concentrations in CVL samples from 219 individuals, grouped by CT.
(C) Cytokine principal component (PC) 1 and PC2 values for 219 subjects, grouped by CT. Groups in
(**A**)-(**C**) were compared using Kruskal-Wallis test with Dunn's post-hoc analyses. **(D)** Cytokine loadings on
PC1 (51.3% of variation) and PC2 (14.6% of variation). **(E)** IL-17 concentrations in the CVL fluid of
participants who remained HIV negative (n=28) and those who acquired HIV (n=191) (Mann-Whitney
test). Lines indicate the median and the interquartile range (IQR) of each dataset.

Figure S4

A

B

Figure S4. Genital mucosal CCR5+ CD4+ T cells and TNF-α concentrations in germ-free mice intravaginally inoculated with *L. crispatus* and *P. bivia*. Related to Figure 4.

(A) Flow cytometry analysis of cells isolated from the FGT. Pooled results from 4 independent experiments.

(B) TNF-α concentrations measured in CVL fluid. Pooled results from 3 independent experiments.

Lines indicate the median and the interquartile range (IQR) of each dataset. Groups were compared using Mann-Whitney test.

Table S1

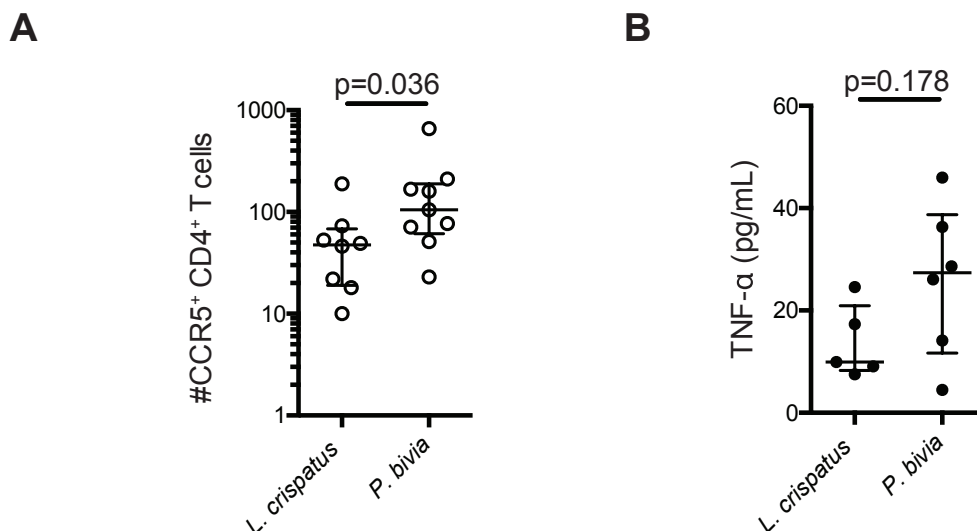|	All n=236	CT1 n=23	CT2 n=75	CT3 n=68	CT4 n=70	
Age of participant at enrollment (Median, IQR)	21.4 [20, 22.4]	20.7 [19.3, 22.1]	21.5 [20.1, 22.4]	21.5 [20.4, 22.7]	21.4 [20.2, 22.6]	p = 0.2352[e]
Age difference between participant and current sexual partner (Median, IQR)[f]	2.9 [1, 5.1]	2.5 [1.2, 3.7]	2.9 [0.6, 5]	3 [1.8, 5.9]	3 [1, 5.1]	p = 0.6630[e]
STIs:	% (n/t)	% (n/t)	% (n/t)	% (n/t)	% (n/t)	
Neisseria gonorrhoeae	4% (10/236)	9% (2/23)	3% (2/75)	4% (3/68)	4% (3/70)	p = 0.5889[a]
Chlamydia trachomatis	16% (37/236)	0% (0/23)	13% (10/75)	16% (11/68)	23% (16/70)	p = 0.0446[a]
Mycoplasma genitalium	6% (15/236)	4% (1/23)	8% (6/75)	4% (3/68)	7% (5/70)	p = 0.8562[a]
Trichomonas vaginalis	8% (18/236)	4% (1/23)	7% (5/75)	7% (5/68)	10% (7/70)	p = 0.8735[a]
Herpes simplex virus 2	2% (5/217)	0% (0/23)	0% (0/69)	3% (2/63)	5% (3/62)	p = 0.2541[a]
Herpes simplex virus 1	0% (0/195)	0% (0/19)	0% (0/64)	0% (0/59)	0% (0/53)	p = 1[a]
Experiencing STI symptoms at time of exam: [c]						
No	92% (217/236)	100% (23/23)	91% (68/75)	97% (66/68)	86% (60/70)	p = 0.1530[b]
Yes	6% (14/236)	0% (0/23)	5% (4/75)	3% (2/68)	11% (8/70)	
Not available	2% (5/236)	0% (0/23)	4% (3/75)	0% (0/68)	3% (2/70)	
Contraceptives:						
No family planning	52% (121/232)	61% (14/23)	48% (35/73)	45% (30/67)	61% (42/69)	p = 0.4540[b]
DMPA/Nuristerate/Implanon	44% (102/232)	39% (9/23)	48% (35/73)	52% (35/67)	33% (23/69)	
Oral contraception	3% (6/232)	0% (0/23)	3% (2/73)	3% (2/67)	3% (2/69)	
Other[d]	1% (3/232)	0% (0/23)	1% (1/73)	0% (0/67)	3% (2/69)	
Condoms in 30 days:						
Always	21% (48/236)	26% (6/23)	13% (10/75)	18% (12/68)	30% (21/70)	p = 0.0508[b]
Sometimes	34% (80/236)	35% (8/23)	41% (31/75)	34% (23/68)	26% (18/70)	
Never	20% (47/236)	13% (3/23)	27% (20/75)	25% (17/68)	11% (8/70)	
Not available	25% (58/236)	26% (6/23)	19% (14/75)	24% (16/68)	33% (23/70)	
Drying agent usage:						
Always	1% (2/236)	0% (0/23)	1% (1/75)	1% (1/68)	0% (0/70)	p = 0.4684[b]
Sometimes	10% (24/236)	13% (3/23)	9% (7/75)	15% (10/68)	6% (4/70)	
Never	74% (174/236)	65% (15/23)	79% (59/75)	66% (45/68)	79% (55/70)	
Not available	15% (36/236)	22% (5/23)	11% (8/75)	18% (12/68)	16% (11/70)	
# of sex episodes in past 30 days:						
One or more	74% (173/236)	70% (16/23)	80% (60/75)	76% (52/68)	67% (47/70)	p = 0.2649[b]
None	24% (56/236)	26% (6/23)	17% (13/75)	24% (16/68)	31% (22/70)	
Not available	2% (4/236)	4% (1/23)	3% (2/75)	0% (0/68)	1% (1/70)	
# of anal sex episodes in past 30 days:						
One or more	3% (6/236)	0% (0/23)	4% (3/75)	1% (1/68)	3% (2/70)	p = 0.6078[b]
None	72% (170/236)	74% (17/23)	76% (57/75)	75% (51/68)	64% (45/70)	
Not available	25% (60/236)	26% (6/23)	20% (15/75)	24% (16/68)	33% (23/75)	
# of regular sex partners in last 30 days:						
One	70% (165/236)	65% (15/23)	77% (58/75)	71% (48/68)	63% (44/70)	p = 0.4258[b]
More than one	0% (0/236)	0% (0/23)	0% (0/75)	0% (0/68)	0% (0/70)	
None	4% (9/236)	4% (1/23)	1% (1/75)	6% (4/68)	4% (3/70)	
Not available	26% (62/236)	30% (7/23)	21% (16/75)	24% (16/68)	33% (23/70)	
# of casual sex partners in last 30 days:						
One	4% (10/236)	0% (0/23)	4% (3/75)	4% (3/68)	6% (4/70)	p = 0.6433[b]
More than one	0% (0/236)	0% (0/23)	0% (0/75)	0% (0/68)	0% (0/70)	
None	70% (164/236)	70% (16/23)	76% (56/75)	72% (49/68)	63% (43/70)	
Not available	27% (62/236)	30% (7/23)	22% (16/75)	24% (16/68)	34% (23/70)	

a. Fisher's exact test; comparing the number of participants with each STI to those without that STI.
b. Fisher's exact test, comparing the number of participants in all the groups lists under the subheading.
c. Symptoms defined as itching, pain, burning, sores, foul smelling discharge, blood, and pain during sex.
d. Includes women reporting use of Loop/IUD or that underwent tubal ligation
e. Kruskal-Wallis test, shown with median and interquartile range.
f. Participant's age is subtracted from the age of the current partner

Table S1. Active sexually transmitted infections (STIs), contraceptive, condom and drying agent usage, and sexual behavior according to cervicotype (CT) for 236 HIV-uninfected individuals. Related to Figure 2.

Supplemental Experimental Procedures

Study cohort

The individuals analyzed in this study were enrolled in the Females Rising through Education, Support, and Health (FRESH) study, a prospective observational study conducted near Durban, South Africa, a region with high HIV prevalence. Participants were followed between Nov 26, 2012 and Sep 15, 2016 for a median of 336 days (IQR: 178.5 to 347 days) and a total of 198.2 person-years. The study protocol was approved by the Biomedical Research Ethics Committee of the University of KwaZulu-Natal and the Massachusetts General Hospital Institutional Review Board (2012P001812/MGH). Informed consent was obtained following the explanation of the nature and possible consequences of the study. Participants received intensive HIV prevention counselling, and condoms (both female and male) were provided at the study site.

Eligibility criteria

To be eligible for the study, participants had to be female, 18-23 years old, sexually active, HIV-negative, and able to understand the provided information and consent forms. They further had to be willing to adhere to study requirements, to have HIV tests performed twice per week and to have samples stored. Criteria for exclusion were pregnancy, anemia or enrolment in another study.

Clinical procedures

Twice per week, participants attended classes focused on personal empowerment, job skills training, and HIV prevention, and underwent finger prick blood draw for HIV RNA viral load testing by PCR. Every 3 months, participants had a peripheral blood draw and a pelvic exam (not performed during menstruation) that involved the collection of

ectocervical and midvaginal swabs (Catch-All, Epicentre), a cervicovaginal lavage (CVL), and an endocervical cytobrush as described previously (Anahtar et al., 2015). All sample collection was performed by a single nurse throughout the entire study. The participants further completed a detailed HIV risk questionnaire that was administered by a counselor and addressed the participant's sexual behavior, STI history, antibiotic usage, diet, and family planning.

Outcomes

Time to HIV-1 infection was defined as the time from enrolment to the first confirmed positive HIV RNA PCR result, or to the last day of viral load assessment for individuals that remained HIV-negative. Assessment of HIV status was made at study entry and twice per week during follow-up.

Selection of samples for analysis

At the time of data analysis, pre-infection mucosal samples from 31 women who became HIV-infected, and samples from 205 women who remained uninfected were available. The estimated median time to infection after the most recent pelvic exam was 35 days (IQR: 25 to 68.5 days), assuming that infection occurred on average 7 days prior to detection of HIV in the blood.

16S rRNA gene sequencing was performed at a single time point for all 236 participants, using the most recent pre-infection ectocervical swab for women who acquired HIV, and the first collected ectocervical swab for women who remained HIV-negative. We chose the latter based on the observed high stability of the vaginal microbiota in FRESH participants over time (Anahtar et al., 2015), and based on the availability of a larger number of cytobrush samples that had flow cytometry performed at this time point. All

further analyses (virome sequencing, Luminex, flow cytometry (HIV target cell analysis), STI and behavioral data) were conducted for the same time point as the 16S rRNA analysis, provided the availability of a matched sample. If no matched sample was available, the individual was excluded from the analysis. We performed Luminex analysis of 219 available paired CVL samples, including 28 pre-infection samples, virome analysis of 180 paired samples, including 17 pre-infection samples, and target cell frequency analysis of 169 paired samples, including 12 pre-infection samples.

Nucleic acid extraction from cervical swabs, PCR amplification and sequencing of the V4 region of the bacterial 16S rRNA gene

This method was described in detail previously (Anahtar et al., 2016; Anahtar et al., 2015).

Analysis of 16S sequencing results was performed using MacQIIME (Caporaso et al., 2010). Fastq files were quality filtered (excluding sequences with a Phred quality score < 30) and de-multiplexed using the split_libraries.py command. The resulting fasta files were combined, and operational taxonomic units (OTUs) were assigned using open-reference picking (97% identity, Greengenes v.13.8) with default parameters except for 0.1% subsampling. Following OTU picking, a median sequence count of 19,426 was obtained, with a minimum count of 5,175 per sample. For display of the bacterial community structure, taxa were summarized at the lowest taxonomic rank resolved by OTU picking, that is at the species level or higher. The core_diversity_analyses.py workflow was used to calculate Faith's Phylogenetic Diversity (diversity within a sample), and the Principal Coordinates Analysis (PCoA) plots in EMPeror (Vazquez-Baeza et al., 2013), using the weighted UniFrac method (Lozupone and Knight, 2005) with a rarefaction cutoff of 5,000 reads per sample.

Definition of cervicotypes

Cervicotypes (CTs) were defined as previously described (Anahtar et al., 2015). Briefly, samples with the relative majority of sequences assigned by QIIME open-reference OTU picking to the genus *Lactobacillus* (but not *Lactobacillus iners*) were defined as CT1. Shotgun sequencing and oligotyping analyses demonstrated that *Lactobacillus* species other than *L. iners* were mostly composed of *Lactobacillus crispatus* (Anahtar et al., 2015). *Lactobacillus iners* was the predominant species in samples assigned to CT2, and *Gardnerella vaginalis* in samples assigned to CT3. CT4 samples have a dominant bacterial taxon other than *Lactobacillus*, *L. iners* or *G. vaginalis*.

Virome Sequencing

For each sample, 300 μl of PBS was added to 200 μl of frozen CVL fluid. The CVL fluid was passed through a 0.45-μm-pore-size membrane, and DNase and Lysozyme treated to enrich for virus-like particles. Total RNA plus DNA was extracted using the Qiagen DNeasy Blood and Tissue kit (Qiagen) according to the manufacturer's recommendation. Purified nucleic acid was reverse transcribed and PCR amplified using barcoded primers consisting of a base-balanced 16-nucleotide-specific sequence upstream of a random 15-mer as previously described (Finkbeiner et al., 2009) and used for NEBNext library construction (New England BioLabs). Libraries were multiplexed (12 samples per flow cell) on an Illumina MiSeq instrument (Washington University Center for Genome Sciences) and sequenced using the paired-end 2 × 250 protocol.

Detection of viral sequences was done using VirusSeeker, a custom bioinformatics pipeline designed to detect sequences sharing nucleotide and protein level sequence

similarity to known viruses (Handley et al., 2016; Monaco et al., 2016). In brief, potential unique viral reads were queried against the NCBI nt/nr databases, and only reads matching exclusively to viral sequences were kept for further analysis. All sequences aligning to viruses were further classified into viral families based on the NCBI taxonomic identity of the best hit. Absolute read counts were normalized by dividing individual taxon sequence counts by the total assigned sequence count in a sample. Richness and diversity were calculated using the diversity function of the *vegan* package (Oksanen, 2016).

Sexually transmitted infection (STI) testing

HIV-1 infection was assessed with a rapid screening HIV RNA PCR assay. Infection with *Chlamydia trachomatis, Neisseria gonorrhoea, Trichomonas vaginalis*, HSV-1 and HSV-2 was tested using a posterior fornix swab (all performed by Global Labs, Durban, South Africa, as previously described (Byrne et al., 2016)). Positive results were followed by a second, confirmatory assay. All women with a positive STI test were referred for further care.

Measurement of cytokines

CVL fluid samples were thawed on ice and centrifuged at 800 x g (10 min, 4°C). The supernatant was transferred to a fresh Eppendorf tube, gently mixed by pipetting and assayed immediately using a customized high-sensitivity MILLIPLEX assay (EMD Millipore) to measure concentrations of IL-1α, IL-1β, TNF-α, IFN-γ, IL-4, IL-5, IL-6, IL-10, IL-12p70, IL-13, IL-17, IL-21, IL-23, MIP-1α, MIP-1β, MIP-3α, MIG, IP-10 and I-TAC. Samples were acquired on a Bio-Plex 3D Suspension Array System (Bio-Rad), and data

were analyzed using Bio-Plex Manager (Bio-Rad). IL-8 concentrations in human CVL fluid and tissue culture supernatant, IL-6 in tissue culture supernatant and TNF-α in murine CVL fluid were measured by ELISA (eBioscience, R&D Systems). Cytokine measurements below the limit of detection were plotted as half the minimum detectable concentration for that cytokine.

Flow cytometry

Human and mouse cells were stained with a LIVE/DEAD viability dye (Invitrogen) and fluorescently labeled monoclonal antibodies specific for the following human or mouse surface markers: hCD45 (HI30), hCD3 (UCHT1), hCD4 (SK3), hCCR5 (2D7), hHLA-DR (G46-6), hCD38 (HIT2), hCD25 (2A3), hCD8 (SK1), hCD11c (B-ly6), hCD14 (M5E2) and hCD19 (HIB19), mCD3 (145-2C11) (BD), mCD45 (30-F11), mCD4 (RM4-5), mCD8a (53-6.7), and mCD44 (IM7) (BioLegend). Samples were acquired on a FACS Aria III or on an LSR II (BD). The whole human cytobrush sample and mouse cells were acquired fresh within 6 hours of collection. Mouse cells were quantified using CountBright Absolute Counting Beads (Invitrogen). Data were analyzed using FlowJo Version 9.8.5 (FlowJo Enterprise). To assess the number of activated CD4$^+$ T cells in a sample, doublets and dead cells were excluded, and the remaining events were gated for CD45, CD3 and CD4$^+$ cells and CCR5, HLA-DR and CD38 (human) or CD44 (mouse). FMO controls were used to define gates for activation markers.

Detection of Differentially Abundant Bacterial Taxa

Resolved sequence variants (RSV) were identified from 16S amplicons using the dada2 procedure (Callahan et al., 2016). Prior to RSV resolution, sequences were quality trimmed to a 220 bases. RSV were inferred from pooled sequences and taxonomy was

assigned using a naïve Bayes classifier against GreenGenes August 2013 release database (gg_13_8_99) (McDonald et al., 2012). Differentially abundant taxa were identified using the DESeq2 package (Love et al., 2014).

Bacteria

Lactobacillus crispatus (provided by Dr. David N. Fredricks) was subcultured on Columbia Blood Agar and grown in Lactobacilli MRS broth (Hardy Diagnostics) anaerobically prior to experiments. *Prevotella bivia* (ATCC 29303), *Prevotella amnii* (CCUG 53648T) and *Gardnerella vaginalis* (ATCC 14018) were cultured on Brucella Blood Agar with Hemin and Vitamin K (Hardy Diagnostics) anaerobically. Inocula were prepared in ultra-pure PBS (Invitrogen) on the day of the experiment, using a spectrophotometer (GE Healthcare) to adjust the optical density as desired, and plated out on their respective culture medium to assess the bacterial concentration.

Epithelial co-culture assays

Human endocervical epithelial cells (End1/E6E7, ATCC) were grown in keratinocyte serum-free medium (KSF) (Invitrogen) supplemented with bovine pituitary extract (0.05 mg/mL), epidermal growth factor (0.1 ng/mL), and calcium chloride (0.4 nM). For experiments, cells were plated in 96-well plates at 2.5×10^4 per well, and allowed to adhere for 24 hours prior to addition of bacteria. 100 ul of a 1×10^7 - 10^8 cfu/mL bacterial suspension in tissue culture medium was added per well and incubated anaerobically at 37 °C. After 24 h of co-culture, the supernatant was aspirated, and an aliquot from each condition plated to confirm bacterial viability. Each treatment had 4 replicates, and each experiment was repeated at least twice.

Mice

All animal procedures were approved by the Animal Resources at Children's Hospital (ARCH) review committee (protocol # 16-08-3219R). Germ-free Swiss-Webster mice were bred and housed in germ-free isolators at Children's Hospital, Boston, MA. For intravaginal inoculation experiments, 6-9 week old female mice were transferred to autoclaved cages as previously described ("out-of-the-isolator protocol") (Faith et al., 2014).

Intravaginal inoculation of mice with bacteria

Mice were anaesthetized by i.p. injection of ketamine (60 mg/kg; Santa Cruz Animal Health) and xylazine (10 mg/kg; Hospira). CVL fluid was collected by gently pipetting 20 μl of ultra-pure PBS three times up and down the vaginal cavity. This was repeated twice. Then, 20 μl of inoculum containing 2x10^8 CFU/mL were released into the vaginal cavity using a 20 μl pipette tip.

Processing of mouse tissues

Blood was collected from the facial vein, and red blood cells were removed using an RBC lysis buffer (BioLegend). Mice were sacrificed via CO_2 asphyxiation, and the mouse female reproductive tract was excised. The uterine horns were removed, and the cervix and vagina were stored in RPMI supplemented with 10% FCS at 4 °C. In the lab, the tissue was cut into tiny pieces by the use of scalpels and subjected to enzymatic digestion with 0.5 mg/mL collagenase II (Sigma) and 3 μg/mL DNase I (Qiagen) for 30 min at 37 °C. The tissue fragments were subsequently pipetted vigorously, ground on a 40 um cell strainer using the sterile stamp of a 5 mL syringe, and frequently rinsed with

ice-cold FACS buffer. The resulting cell suspensions were centrifuged and used for downstream procedures.

Statistics

Descriptive measures, such as median, interquartile range (IQR), frequency and percent, were used to summarize the data. For comparison of continuous data between two groups, Mann-Whitney test was performed. Kruskal-Wallis test with Dunn's post hoc analyses was used for comparing continuous data between more than two groups. The time to HIV acquisition was summarized using Kaplan-Meier curves, and significance was assessed by Log-rank test. Fisher's exact test was performed for comparison of categorical data between two or more groups. Principal component analysis (PCA) was used to obtain summary measures for the multi-variable cytokine set and performed using FactoMineR package in R. Multivariate analysis using MaAsLin (https://bitbucket.org/biobakery/maaslin) (Morgan et al., 2012) was performed to identify bacteria correlated with cytokine principal component 1. More specifically, linear regression of logit-transformed bacterial abundances was performed on principal component 1 scores. Bacteria not present in at least 10% of the samples with at least 0.01% abundance were excluded from the analysis. Adjustments for multiple testing were made using the Benjamini-Hochberg method.

All p values shown are two-sided. P values indicated by asterisks should be interpreted as follows: *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001; n.s.: p≥0.05. Unless indicated otherwise, the analyses were performed in Prism 6 (GraphPad) and R Studio.

Supplemental References

Anahtar, M.N., Bowman, B.A., and Kwon, D.S. (2016). Efficient Nucleic Acid Extraction and 16S rRNA Gene Sequencing for Bacterial Community Characterization. Journal of visualized experiments : JoVE.

Anahtar, M.N., Byrne, E.H., Doherty, K.E., Bowman, B.A., Yamamoto, H.S., Soumillon, M., Padavattan, N., Ismail, N., Moodley, A., Sabatini, M.E.*, et al.* (2015). Cervicovaginal bacteria are a major modulator of host inflammatory responses in the female genital tract. Immunity *42*, 965-976.

Byrne, E.H., Anahtar, M.N., Cohen, K.E., Moodley, A., Padavattan, N., Ismail, N., Bowman, B.A., Olson, G.S., Mabhula, A., Leslie, A.*, et al.* (2016). Association between injectable progestin-only contraceptives and HIV acquisition and HIV target cell frequency in the female genital tract in South African women: a prospective cohort study. The Lancet Infectious Diseases *16*, 441-448.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J., and Holmes, S.P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nature methods *13*, 581-583.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I.*, et al.* (2010). QIIME allows analysis of high-throughput community sequencing data. Nature methods *7*, 335-336.

Faith, J.J., Ahern, P.P., Ridaura, V.K., Cheng, J., and Gordon, J.I. (2014). Identifying gut microbe-host phenotype relationships using combinatorial communities in gnotobiotic mice. Sci Transl Med *6*.

Finkbeiner, S.R., Holtz, L.R., Jiang, Y., Rajendran, P., Franz, C.J., Zhao, G., Kang, G., and Wang, D. (2009). Human stool contains a previously unrecognized diversity of novel astroviruses. Virology journal *6*, 161.

Handley, S.A., Desai, C., Zhao, G., Droit, L., Monaco, C.L., Schroeder, A.C., Nkolola, J.P., Norman, M.E., Miller, A.D., Wang, D.*, et al.* (2016). SIV Infection-Mediated Changes in Gastrointestinal Bacterial Microbiome and Virome Are Associated with Immunodeficiency and Prevented by Vaccination. Cell Host Microbe *19*, 323-335.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology *15*, 1-21.

Lozupone, C., and Knight, R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. Applied and environmental microbiology *71*, 8228-8235.

McDonald, D., Price, M.N., Goodrich, J., Nawrocki, E.P., DeSantis, T.Z., Probst, A., Andersen, G.L., Knight, R., and Hugenholtz, P. (2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J *6*, 610-618.

Monaco, C.L., Gootenberg, D.B., Zhao, G., Handley, S.A., Ghebremichael, M.S., Lim, E.S., Lankowski, A., Baldridge, M.T., Wilen, C.B., Flagg, M.*, et al.* (2016). Altered Virome and Bacterial Microbiome in Human Immunodeficiency Virus-Associated Acquired Immunodeficiency Syndrome. Cell Host Microbe *19*, 311-322.

Morgan, X.C., Tickle, T.L., Sokol, H., Gevers, D., Devaney, K.L., Ward, D.V., Reyes, J.A., Shah, S.A., LeLeiko, N., Snapper, S.B.*, et al.* (2012). Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol *13*, R79.

Oksanen, J.B., F. G.; Friendly, M.; Kindt, R.; Legendre, P.; McGlinn, D.; Minchin, P. R.; O'Hara, R. B.; Simpson, G. L.; Solymos, P.; Stevens, M. H. H.; Szoecs, E. and Wagner, H. (2016). vegan: Community Ecology Package. R package version 2.4-0. .

Vazquez-Baeza, Y., Pirrung, M., Gonzalez, A., and Knight, R. (2013). EMPeror: a tool for visualizing high-throughput microbial community data. GigaScience *2*, 16.