Evaluation of the association between the concentrations of key vaginal bacteria and the increased risk of HIV acquisition in African women from five cohorts: a nested case-control study

R Scott McClelland, Jairam R Lingappa*, Sujatha Srinivasan, John Kinuthia, Grace C John-Stewart, Walter Jaoko, Barbra A Richardson, Krista Yuhas, Tina L Fiedler, Kishorchandra N Mandaliya, Matthew M Munch, Nelly R Mugo, Craig R Cohen, Jared M Baeten, Connie Celum, Julie Overbaugh, David N Fredricks*

Summary

Background Disruptions of vaginal microbiota might increase women's susceptibility to HIV infection. Advances in molecular microbiology have enabled detailed examination of associations between vaginal bacteria and HIV acquisition. Therefore, this study aimed to evaluate the association between the concentrations of specific vaginal bacteria and increased risk of HIV acquisition in African women.

Methods We did a nested case-control study of participants from eastern and southern Africa. Data from five cohorts of African women (female sex workers, pregnant and post-partum women, and women in serodiscordant relationships) were used to form a nested case-control analysis between women who acquired HIV infection versus those who remained seronegative. Deep sequence analysis of broad-range 16S rRNA gene PCR products was applied to a subset of 55 cases and 55 controls. From these data, 20 taxa were selected for bacterium-specific real-time PCR assays, which were examined in the full cohort as a four-category exposure (undetectable, first tertile, second tertile, and third tertile of concentrations). Conditional logistic regression was used to generate odds ratios (ORs) and 95% CIs. Regression models were stratified by cohort, and adjusted ORs (aORs) were generated from a multivariable model controlling for confounding variables. The Shannon Diversity Index was used to measure bacterial diversity. The primary analyses were the associations between bacterial concentrations and risk of HIV acquisition.

Findings Between November, 2004, and August, 2014, we identified 87 women who acquired HIV infection (cases) and 262 controls who did not acquire HIV infection. Vaginal bacterial community diversity was higher in women who acquired HIV infection (median 1·3, IQR 0·4–2·3) than in seronegative controls (0·7, 0·1–1·5; p=0·03). Seven of the 20 taxa showed significant concentration-dependent associations with increased odds of HIV acquisition: *Parvimonas* species type 1 (first tertile aOR 1·67, 95% CI 0·61–4·57; second tertile 3·01, 1·13–7·99; third tertile 4·64, 1·73–12·46; p=0·005) and type 2 (first tertile 3·52, 1·63–7·61; second tertile 0·85, 0·36–2·02; third tertile 2·18, 1·01–4·72; p=0·004), *Gemella asaccharolytica* (first tertile 2·09, 1·01–4·36; second tertile 2·02, 0·98–4·17; third tertile 3·03, 1·46–6·30; p=0·010), *Mycoplasma hominis* (first tertile 1·46, 0·69–3·11; second tertile 1·40, 0·66–2·98; third tertile 2·76, 1·36–5·63; p=0·048), *Leptotrichia/Sneathia* (first tertile 2·04, 1·02–4·10; second tertile 1·45, 0·70–3·00; third tertile 2·59, 1·26–5·34; p=0·046), *Eggerthella* species type 1 (first tertile 1·79, 0·88–3·64; second tertile 2·62, 1·31–5·22; third tertile 1·53, 0·72–3·28; p=0·041), and vaginal *Megasphaera* species (first tertile 3·15, 1·45–6·81; second tertile 1·43, 0·65–3·14; third tertile 1·32, 0·57–3·05; p=0·038).

Interpretation Differences in the vaginal microbial diversity and concentrations of key bacteria were associated with greater risk of HIV acquisition in women. Defining vaginal bacterial taxa associated with HIV risk could point to mechanisms that influence HIV susceptibility and provide important targets for future prevention research.

Funding National Institute of Child Health and Human Development.

Introduction

Compared with other parts of the world, where men account for most new infections, 56% of new HIV infections in Africa in 2015 were in women.[1] Bacterial vaginosis, a condition characterised by the presence of complex anaerobic vaginal bacterial communities, might contribute to HIV transmission and the disproportionate burden of HIV infection in African women.[2] The specific bacteria underlying the association between bacterial vaginosis and HIV infection remain poorly understood.

Advances in molecular microbiology have enhanced the understanding of normal and dysbiotic human microbiota.[3] These approaches have facilitated identification of distinct vaginal bacterial community types, ranging from low diversity and lactobacillus-dominated bacterial communities to heterogeneous and highly diverse bacterial vaginosis-associated communities, characterised by an abundance of anaerobic species.[3–5] Bacterial species vary in terms of their associations with bacterial vaginosis,[6] particular symptoms,[7] and vaginal inflammation.[8] A recent study from South Africa showed

Lancet Infect Dis 2018

Published Online
January 25, 2018
http://dx.doi.org/10.1016/S1473-3099(18)30058-6

See Online/Comment
http://dx.doi.org/10.1016/S1473-3099(18)30061-6

*Contributed equally

Department of Medicine (Prof R S McClelland MD, Prof J R Lingappa MD, Prof J M Baeten MD, Prof C Celum MD, Prof D N Fredricks MD), **Department of Epidemiology** (Prof R S McClelland, Prof G C John-Stewart MD, Prof J M Baeten, Prof C Celum), **Department of Global Health** (Prof R S McClelland, Prof J R Lingappa, Prof G C John-Stewart, K Yuhas MSc, Prof K N Mandaliya MBChB, Prof J M Baeten, Prof C Celum), **Department of Pediatrics** (Prof J R Lingappa), **and Department of Biostatistics** (Prof B A Richardson PhD), **University of Washington, Seattle, WA, USA; Institute of Tropical and Infectious Diseases** (Prof R S McClelland) **and Department of Medical Microbiology** (Prof W Jaoko MBChB), **University of Nairobi, Nairobi, Kenya; Vaccine and Infectious Disease Division** (S Srinivasan PhD, T L Fiedler BS, M M Munch BS, Prof D N Fredricks), **Public Health Sciences Division** (Prof J Overbaugh PhD), **and Human Biology Division** (Prof B A Richardson, Prof J Overbaugh), **Fred Hutchinson Cancer Research Center, Seattle, WA, USA; Kenyatta National Hospital, Nairobi, Kenya** (J Kinuthia MBChB); **Kenya Medical Research Institute, Nairobi, Kenya** (N R Mugo MBChB);

and Department of Obstetrics, Gynecology & Reproductive Sciences, University of California at San Francisco, San Francisco, CA, USA
(Prof C R Cohen MD)

Correspondence to:
Prof R Scott McClelland, Department of Medicine, University of Washington, Seattle, WA 98104, USA
mcclell@uw.edu

Research in context

Evidence before this study

We did a PubMed search on Nov 22, 2017, using the search terms "((((vagina*) AND bacteria*) AND HIV) AND acquisition) NOT review", without any date or language restrictions. The search returned 80 articles, of which nine addressed the hypothesis that vaginal microbiota might influence women's risk of acquiring HIV infection. All five prospective cohort studies in this group found that bacterial vaginosis was associated with increased risk of HIV acquisition. A single recent cohort study from South Africa used broad range bacterial PCR with deep sequencing to characterise the vaginal microbiome. In this cohort, young women with high-diversity vaginal bacterial communities had increased risk of HIV acquisition compared with women with low-diversity *Lactobacillus crispatus*-dominated communities. Higher relative abundances of *Prevotella melaninogenica, Veillonella montpellierensis, Mycoplasma, Prevotella bivia,* and *Sneathia sanguinegens* were also associated with increased risk of HIV acquisition. Individual bacterial species were not measured using quantitative PCR methods that are more sensitive compared with deep sequencing approaches, and allow for assessment of absolute concentrations. Further research is needed to clarify the relationship between individual bacterial quantities and women's risk of HIV acquisition.

Added value of this study

This study was a large nested case-control study of women from diverse regions within Africa and represented three important risk groups: female sex workers, pregnant and post-partum women, and HIV-negative women in serodiscordant relationships. This study is the first to show significant associations between the concentrations of specific vaginal bacteria and women's risk of HIV acquisition. The use of broad range PCR with deep sequencing to identify bacterial taxa for further study followed by taxon-directed real-time PCR to test the hypotheses related to individual bacterial taxa illustrates the value of this sequential experimental approach in microbiome studies.

Implications of all the available evidence

Concentrations of key vaginal bacteria were strongly associated with women's risk of acquiring HIV. High concentrations of some bacteria were substantially more predictive of HIV risk compared with a microscopic diagnosis of bacterial vaginosis. Key bacteria could increase HIV susceptibility through multiple potential pathways, including inflammation, production of HIV inducing factors, and disruption of physical and chemical barriers to infection. Defining vaginal bacterial taxa associated with HIV risk could point to mechanisms influencing HIV susceptibility and provide important targets for future prevention research.

that young women with high-diversity vaginal bacterial communities had increased numbers of activated genital mucosal CD4-positive T cells and a four-times increased risk of HIV acquisition compared with women with low-diversity *Lactobacillus crispatus*-dominated communities.[9] Increased relative abundance of several bacterial taxa (*Prevotella melaninogenica, Veillonella montpellierensis, Mycoplasma, Prevotella bivia,* and *Sneathia sanguinegens*) were also associated with increased risk. One limitation of relative abundance data is that they do not provide absolute concentrations of bacteria, which can vary widely in women with the same relative abundance. Therefore in this study, we test the hypothesis that concentrations of specific vaginal bacteria are associated with increased risk of HIV acquisition in African women.

Methods

Study design and participants

We did a nested case-control study of participants from five cohorts in eastern and southern Africa. In the Mombasa Cohort,[10] we included women who were aged 16 years or older, HIV seronegative, and self-identified as exchanging sex for cash or in-kind payment. In the Mama Salama Study,[11] we included women who were aged 14 years or older, pregnant, and HIV seronegative either at enrolment or documented during routine antenatal care within the past 3 months. In the three cohorts of HIV-serodiscordant heterosexual couples (the Partners in Prevention Herpes Simplex Virus/HIV Transmission Study,[12] the Couples Observational Study,[13] and the Partners Pre-Exposure Prophylaxis Study),[14] we included women who were aged 18 years or older, HIV seronegative, and had an HIV-seropositive male partner aged 18 years or older. We have published detailed procedures for each cohort.[10–14] Each protocol has received approval from country-specific and investigator-affiliated ethical review boards, and participants provided written informed consent.

Procedures

We enrolled and collected demographic, medical, and sexual history data, as well as vaginal samples for microbiota analyses across all cohorts. We assessed and treated for sexually transmitted infections at baseline using standard regimens for the respective countries. We asked participants to return every 1–3 months for HIV testing at the respective research clinics. We defined women who acquired HIV as cases, and we compared them with women without HIV infection (matched controls) from the same cohort. Additionally, we evaluated plasma samples using nucleic acid amplification tests from pre-seroconversion visits in women who seroconverted for antibodies to HIV infection. We used these data, together with HIV serology results, to identify the first visit with evidence of HIV infection (serum anti-HIV antibody or plasma HIV RNA, or both). We selected pre-HIV-infection genital samples from visits at which participants were

both HIV antibody negative and HIV RNA negative. All participants received risk-reduction education and free condoms. Table 1 provides additional details about the procedures for each cohort.[10–14]

Microbiota analyses

We collected vaginal samples for microbiota analyses using vaginal swabs for DNA extraction and bacterial PCR. These samples were stored at –80°C and transported on dry ice to the Fred Hutchinson Cancer Research Center (Seattle, WA, USA) for analysis. The appendix (pp 1–3) provides details about the laboratory methods used. However, to summarise, a sequential approach was used in these experiments. First, broad-range 16S rRNA gene PCR with pyrosequencing was done for a subset of all cases and one randomly selected control per case. These data illustrated the overall distribution of bacterial taxa in cases versus controls. Second, relative abundance data were used to identify key bacteria to analyse with use of real-time (rt) PCR in all cases and controls.

Statistical analysis

This study targeted at least 80 cases and 240 controls. Assuming an α level of 0·05, this sample size provides more than 90% power to detect a 2·6-times difference or more in the odds of detecting a vaginal bacterial taxon in cases versus controls, assuming 20% prevalence or more of the organism in controls.

The first step of the analysis, using deep sequencing data from 55 cases and 55 matched controls, generated a large number of tests of association, so we considered this step to be hypothesis generating. On the basis of initial comparisons, we selected 20 bacteria for directed hypothesis testing using rtPCR in all cases and controls.

See Online for appendix

Pyrosequencing data were used to calculate two measures of bacterial community structure. The Chao1 Index provides an estimate of community richness, reflecting the number of different taxa.[16] The Shannon Diversity Index is a measure of diversity accounting for both the number of different taxa and the evenness of their distribution.[17] We calculated statistics of ecological diversity and richness separately for each sample using

	Study dates	Countries	Risk group	Study design	Visit interval	Sample timing and type	Case identification	Control selection
Mombasa Cohort[10]	May, 2010, to August, 2014	Kenya	Female sex workers	Prospective cohort study	Monthly	Enrolment and monthly vaginal dry polyester swab*	Screening ELISA with positive results confirmed by second ELISA†	Incidence density sampling‡
Mama Salama Study[11]	May, 2011, to August, 2014	Kenya	Pregnant and post-partum women	Prospective cohort study	1–3 months§	Enrolment and 1–3 monthly vaginal dry polyester swab*	Transcription mediated amplification¶	Incidence density sampling‡
Partners in Prevention HSV/HIV Transmission Study[12]	November, 2004, to October, 2008	Botswana, Kenya, South Africa, Tanzania, Uganda, and Zambia	HIV-serodiscordant couples	Phase 3 clinical trial‖	3 months	Enrolment cervical swab in media**	HIV rapid assay with positive results confirmed by ELISA and in batch by HIV western blot††	Frequency matched by cohort‡‡
Couples Observational Study[13]	August, 2007, to January, 2010	Uganda and South Africa	HIV-serodiscordant couples	Prospective cohort study	3 months	Enrolment cervical swab in media**	HIV rapid assay with positive results confirmed by ELISA and in batch by HIV western blot††	Frequency matched by cohort‡‡
Partners PrEP Study[14]§§	July, 2008, to December, 2012	Kenya and Uganda	HIV-serodiscordant couples	Phase 3 clinical trial¶¶	Monthly	Enrolment and annual vaginal dry polyester swab‖‖	HIV rapid assay with positive results confirmed by ELISA and in batch by HIV western blot††	Frequency matched by cohort‡‡

In all cohorts, the exposure variable for the primary analyses was the quantity of individual bacterial taxa detected with use of real-time PCR assays. For the Partners in Prevention HSV/HIV Study, Couples Observational Study, and Partners PrEP Study, where vaginal samples were collected less frequently, cases were restricted to women who had a swab collected within 12 study months before seroconversion. ELISA=enzyme-linked immunoassay. HSV=herpes simplex virus. PrEP=pre-exposure prophylaxis. NAAT=nucleic acid amplification test. *Dry, large-bulb polyester tipped push-off swab (SpinEze, Spring Park, MN, USA). Swabs were placed on ice (Mombasa Cohort) or in liquid nitrogen dry shippers (Mama Salama Study) immediately after collection and transported to a central laboratory for inventory and storage at –80°C. †HIV1,2 Antigen Antibody ELISA Kit (Pishtaz Teb Diagnostics, Tehran, Iran) was used as the initial screening test and all positive samples were confirmed with the Vironostika HIV Uni-Form II Ag/Ab ELISA kit (bioMerieux, Marcy l'Etoile, France). ‡For each case, we selected three controls from the same cohort who remained HIV negative, matched in calendar time to the cases, by randomly selecting from among all HIV-negative participants who contributed a follow-up visit with a vaginal swab for microbiota analyses within 28 days after or before the case's exposure swab date. In addition to cases in women who were HIV negative at enrolment, the Mombasa Cohort and the Mama Salama Study included women defined as having acute HIV infection (ie, positive NAAT but a negative plasma antibody or rapid HIV test at enrolment). The Mama Salama Study also included enrolment of seroconverters with a documented negative HIV rapid test less than 3 months before enrolment, and both a positive rapid test and positive NAAT for HIV at enrolment. Because all of these women were newly diagnosed, none were on antibiotic prophylaxis. §Pregnant women were enrolled at any stage of gestation. During pregnancy, they were asked to return at 20, 24, 32, and 36 weeks of gestation. Following delivery, they were asked to return at 2, 6, 10, and 14 weeks, as well as at 6 and 9 months post partum. ¶At each visit, testing for HIV was done with use of the first-generation Gen-Probe HIV viral load assay (Hologic/Gen-Probe, San Diego, CA, USA). ‖Herpes suppression with acyclovir did not significantly reduce HIV transmission risk compared with placebo, so women in both trial groups were eligible for the present analysis. **Gen-Probe cervical collection swab sample collected in Gen-Probe medium (Hologic/Gen-Probe, San Diego, CA, USA). Female genital microbiome studies provided similar results with cervical and vaginal swabs.[15] Swabs were placed on ice after collection and transported to a central laboratory for inventory and storage at –80°C. ††Dual rapid HIV antibody tests were done at the clinic and confirmed with HIV enzyme immunoassay. HIV serostatus at enrolment visits for all participants and at follow-up visits for HIV seroconverters was assessed with use of western blot (Genetics Systems HIV, Hercules, CA, USA) at the University of Washington (Seattle, WA, USA) in batch at the end of each study. ‡‡For each HIV-serodiscordant couples cohort, a pool of non-seroconverting controls was identified that was frequency-matched by biological sex and site to represent the distribution of the enrolled study cohort. Controls for this study were then selected at random from each cohort's set of controls to match a three-to-one ratio of controls to cases. §§One participant initially identified to be a case was not confirmed on HIV testing done at the conclusion of the trial. This putative case was removed from the analysis. Because this cohort of HIV-serodiscordant couples used frequency matching of controls, no controls were excluded after removing that case. However, two controls were excluded for not having a swab at the relevant visit. ¶¶Participants randomly allocated to PrEP had a reduction in HIV acquisition risk compared with those randomly allocated to placebo; therefore, only women in the placebo group were eligible for the present analysis. ‖‖COPAN flocked swabs (COPAN Diagnostics, Murrieta, CA, USA) were used and placed on ice after collection, and transported to a central laboratory for inventory and storage at –80°C.

Table 1: **Summary of procedures and selection of cases and controls in the five cohorts of African women**

For the **R microbiome package** see http://microbiome.github.io/microbiome/"

read numbers classified to their most specific taxonomic rank, using implementations of the R microbiome package (version 3.3.2). We compared index values in cases and controls using Wilcoxon rank-sum tests.

	Total (n=349)	Full cohort		Deep sequencing subset	
		Cases (n=87)	Controls (n=262)	Cases (n=55)	Controls (n=55)
Median age, years	28 (22–35)	26 (22–30)	29 (23–36)	26 (22–30)	29 (20–35)
Country					
Kenya	210 (60·2%)	53 (60·9%)	157 (59·9%)	35 (63·6%)	35 (63·6%)
Uganda	112 (32·1%)	27 (31·0%)	85 (32·4%)	20 (36·4%)	19 (34·6%)
South Africa	13 (3·7%)	3 (3·5%)	10 (3·8%)	0	0
Tanzania	6 (1·7%)	0	6 (2·3%)	0	1 (1·8%)
Botswana	5 (1·4%)	1 (1·1%)	4 (1·5%)	0	0
Zambia	3 (0·9%)	3 (3·4%)	0	0	0
Median number of years of education	8 (6–10)	8 (7–10)	8 (6–10)	8 (6–10)	7 (5–9)
Married	265 (75·9%)	66 (75·9%)	199 (76·0%)	44 (80·0%)	43 (78·2%)
Enrolment cohort					
Mombasa Cohort	40 (11·5%)	10 (11·5%)	30 (11·5%)	7 (12·7%)	7 (12·7%)
Mama Salama Study	112 (32·1%)	28 (32·2%)	84 (32·1%)	20 (36·4%)	20 (36·4%)
Partners in Prevention HSV/HIV Transmission Study	52 (14·9%)	13 (14·9%)	39 (14·9%)	0	0
Couples Observational Study	32 (9·2%)	8 (9·2%)	24 (9·2%)	0	0
Partners PrEP Study	113 (32·4%)	28 (32·2%)	85 (32·4%)	28 (50·9%)	28 (50·9%)
Pregnancy and contraception status					
Not pregnant, no modern contraception	173 (49·6%)	35 (40·2%)	138 (52·7%)	17 (30·9%)	31 (56·4%)
Not pregnant, oral contraceptive	18 (5·2%)	7 (8·0%)	11 (4·2%)	5 (9·1%)	2 (3·6%)
Not pregnant, DMPA	55 (15·8%)	18 (20·7%)	37 (14·1%)	13 (23·6%)	8 (14·5%)
Not pregnant, IUD	8 (2·3%)	2 (2·3%)	6 (2·3%)	2 (3·6%)	2 (3·6%)
Not pregnant, implant	18 (5·2%)	5 (5·7%)	13 (5·0%)	5 (9·1%)	2 (3·6%)
Pregnant	77 (22·1%)	20 (23·0%)	57 (21·8%)	13 (23·6%)	10 (18·2%)
Number of recent sexual partners*					
0	70 (20·1%)	14 (16·1%)	56 (21·4%)	10 (18·2%)	14 (25·5%)
1	268 (76·8%)	71 (81·6%)	197 (75·2%)	44 (80·0%)	38 (69·1%)
>1	11 (3·2%)	2 (2·3%)	9 (3·4%)	1 (1·8%)	3 (5·5%)
Median frequency of sex in the past month†	3 (1–5)	3 (1–6)	3 (1–5)	3 (1–7)	2 (0–5)
Any recent unprotected sex‡	121 (34·7%)	36 (41·4%)	85 (32·4%)	22 (40·0%)	18 (32·7%)
On examination§					
Abnormal vaginal discharge¶	41 (14·0%)	12 (16·2%)	29 (13·2%)	9 (19·1%)	8 (17·8%)
Genital ulceration‖	6 (2·0%)	2 (2·7%)	4 (1·8%)	0	0
Cervical mucopus‖	4 (1·4%)	3 (4·1%)	1 (0·5%)	3 (6·4%)	0
Vaginal Gram stain Nugent score**					
Normal (0–3)	161 (52·6%)	28 (36·8%)	133 (57·8%)	21 (38·9%)	36 (69·2%)
Intermediate (4–6)	45 (14·7%)	16 (21·1%)	29 (12·6%)	12 (22·2%)	9 (17·3%)
Bacterial vaginosis (7–10)	100 (32·7%)	32 (42·1%)	68 (29·6%)	21 (38·9%)	7 (13·5%)

(Table 2 continues on next page)

To identify potentially important species for quantitative analysis, we applied unadjusted logistic regression to the relative abundance data, with case status as the outcome and relative abundance percentage for each taxon separately as the exposure. To select the subset for rtPCR testing, we ranked the bacteria identified through pyrosequencing in descending order by score statistic using logistic models run on each taxon in rank order of score statistic until a p value of 0·2 was reached in the univariate logistic regressions. 16 bacteria were identified for further study using rtPCR on the basis of the magnitude of odds ratios (ORs) in logistic regression. The *Parvimonas* taxon was represented by rtPCR assays to detect species types 1 and 2, both of which were linked to HIV risk. The *Megasphaera* rtPCR was a combined assay detecting vaginal *Megasphaera* types 1 and 2, both of which have been linked to bacterial vaginosis. The *Prevotella* rtPCR was a genus-directed assay. This approach was chosen on the basis of pyrosequencing data, in which higher relative abundance of *Prevotella timonensis, P bivia,* and additional undifferentiated *Prevotella* taxa were all associated with similarly increased odds of HIV acquisition. We tested four additional bacteria using rtPCR despite lower score statistics. We included *Gardnerella vaginalis* because of its long-standing association with bacterial vaginosis and role in biofilm formation.[18] Additionally, we included *L crispatus* and *Lactobacillus jensenii* because of their well recognised association with vaginal health.[4,6] Finally, the relative abundance of *Atopobium vaginae* was significantly higher in cases versus controls in serodiscordant couples, so rtPCR for this species was done in all cohorts.

In the rtPCR analyses, each bacterial taxon was analysed as a four-category exposure, including undetectable (reference category), first tertile, second tertile, and third tertile of concentrations. The one exception was the *Prevotella* genus, which had few undetectable samples, so was modelled in four quartiles. Conditional logistic regression was used to generate ORs and 95% CIs testing the hypothesis that increasing quantities of the targeted bacteria were associated with increased or decreased risk for HIV acquisition. Regression models were stratified by cohort to address clustering of individuals within cohorts. The modelling approach assumed a baseline odds for each cohort cluster. This was treated as a nuisance parameter (ie, a parameter not of direct interest but must be accounted for) and conditioned out of the likelihood in the conditional logistic regression models.[19] For each of 20 bacterial taxa in this primary analysis, a single joint p value was used to assess the significance of the overall association between bacterial quantity and HIV acquisition.

We selected potential confounders of the association between vaginal bacterial concentrations and HIV acquisition a priori on the basis of biologically plausible confounding effects. These included age (continuous), pregnancy and contraceptive status (categorical), number of sexual partners in the past month (continuous),

frequency of sex in the past month (continuous), and recent self-reported unprotected sex (binary). We included all potential confounders in a multivariable model stratified by cohort. We calculated Spearman correlation values for pairwise comparisons of bacterial concentrations.

To illustrate the difference in HIV risk associated with individual bacterial taxa compared with microscopic criteria for bacterial vaginosis used in earlier studies, we repeated analyses using Nugent scores,[20] comparing normal microbiota (scores 0–3) to intermediate microbiota (scores 4–6) and bacterial vaginosis (scores 7–10).

To examine the effect of the two-step experimental approach using analysis of pyrosequencing data to generate hypotheses for further investigation using rtPCR assays, we evaluated the association between detection of individual taxa with use of rtPCR assays and HIV acquisition in a validation subset of 128 women not included in the relative abundance analysis. Because the validation sample was independent of that used in the pyrosequencing analysis, it was possible to apply a Benjamini–Hochman false discovery rate of 0·20 to the 20 bacterial taxa examined. Three adjustments were made to the analytical approach because of the smaller dataset. First, this analysis used detection rather than quantiles of each bacterial taxon. Second, confounders in the multivariable model were restricted to age, pregnancy or contraceptive status, and recent unprotected intercourse. Third, a firth logistic regression was used to generate 95% CIs and p values if data were sparse (expected cell count less than five in a cross tabulation of exposure and outcome).

We did sensitivity analyses, focusing on bacteria significantly associated with HIV acquisition in the primary analysis. First, analyses were repeated with the female sex worker cohort, pregnant or post-partum cohort, and HIV-serodiscordant couples cohorts separately to assess whether results were similar in each population. Second, since vaginal microbiota change over time, the primary analysis was repeated after excluding cases and their matched controls in which the case sample was collected more than 90 days before the first visit where HIV infection was identified. Third, the analysis was repeated after excluding cases sampled during acute or early HIV infection and their controls. Fourth, because inflammation caused by classical sexually transmitted infections might abrogate an effect of vaginal microbiota on HIV susceptibility mediated through an inflammatory mechanism, analyses were repeated in the subset of women without *Neisseria gonorrhoeae, Chlamydia trachomatis,* or *Trichomonas vaginalis* detected at the study visit. Fifth, because vaginal washing is a suspected risk factor for both HIV acquisition and bacterial vaginosis,[21,22] an analysis adjusting for vaginal washing status was done in the subset of women in which these data were captured. Finally, analyses were repeated after stratifying by sample collection method (vaginal *vs* cervical swab).

	Total (n=349)	Full cohort		Deep sequencing subset	
		Cases (n=87)	Controls (n=262)	Cases (n=55)	Controls (n=55)
(Continued from previous page)					
Laboratory-confirmed sexually transmitted infections at analysis visits					
Neisseria gonorrhoeae††	12 (4·6%)	2 (2·9%)	10 (5·2%)	2 (4·4%)	2 (4·9%)
Chlamydia trachomatis‡‡	9 (3·8%)	3 (5·0%)	6 (3·4%)	2 (5·0%)	0
Trichomonas vaginalis§§	27 (8·0%)	11 (12·9%)	16 (6·3%)	4 (7·3%)	1 (1·8%)
Syphilis seropositive¶¶	6 (3·1%)	2 (4·4%)	4 (2·6%)	0	1 (3·7%)
HSV-2 seropositive||||	165 (83·3%)	45 (90·0%)	120 (81·1%)	32 (94·1%)	27 (87·1%)
Vaginal yeast on wet mount***	29 (19·1%)	9 (23·7%)	20 (17·5%)	8 (29·6%)	5 (18·5%)
Any antibiotic use in the past 90 days†††	22 (11·5%)	7 (14·3%)	15 (10·5%)	4 (10·5%)	4 (10·5%)

Data are n (%) or median (IQR). HSV=herpes simplex virus. PrEP=pre-exposure prophylaxis. DMPA=depo medroxyprogesterone acetate. IUD=intrauterine contraceptive device. *Past week for Mombasa Cohort, otherwise past month for the other cohorts. †Imputed for Mombasa Cohort as the past weeks frequency multiplied by four to get frequency per month. ‡Past week for Mombasa Cohort, otherwise past month for the other cohorts. §55 women in the Mama Salama Study were not examined at the swab collection date (per study schedule). One woman in the Partners in Prevention HSV/HIV Study was not examined for genital ulceration and cervical mucopus. ¶n=294 for full cohort (n=74 cases and n=220 controls) and n=92 for deep sequencing subset (n=47 cases and n=45 controls). ||n=293 for full cohort (n=73 cases and n=220 controls) and n=92 for deep sequencing subset (n=47 cases and n=45 controls). **Not available for the Couples Observational Study and missing for 11 women from other cohorts. n=306 for full cohort (n=76 cases and n=230 controls) and n=106 for deep sequencing subset (n=54 cases and n=52 controls). ††n=262 for full cohort (n=68 cases and n=194 controls) and n=86 for deep sequencing subset (n=45 cases and n=41 controls). 74 women in the Mama Salama Study were not assessed for *Neisseria gonorrhoeae* at the swab collection date (per study schedule), and *N gonorrhoeae* swabs are missing for 13 women from other cohorts. ‡‡n=239 for full cohort (n=60 cases and n=179 controls) and n=77 for deep sequencing subset (n=40 cases and n=37 controls). 74 women in the Mama Salama Study and 23 women in the Mombasa Cohort were not assessed for *Chlamydia trachomatis* at the swab collection date (per study schedules), and *C trachomatis* swabs are missing for 13 women from other cohorts. §§n=339 for full cohort (n=85 cases and n=254 controls) and n=110 for deep sequencing subset (n=55 cases and n=55 controls). Samples were missing for ten women from the Partners in Prevention HSV/HIV Study, Couples Observational Study, and Partners PrEP Study. ¶¶n=196 for full cohort (n=45 cases and n=151 controls) and n=51 for deep sequencing subset (n=24 cases and n=27 controls). Syphilis status at the study visit was not available for the Mama Salama Study and was not available at the study visit for 41 women from the other cohorts. ||||n=198 for full cohort (n=50 cases and n=148 controls) and n=65 for deep sequencing subset (n=34 cases and n=31 controls). HSV-2 serostatus was not available for the Mama Salama Study and was missing for 39 women from the other cohorts. ***n=152 for full cohort (n=38 cases and n=114 controls) and n=54 for deep sequencing subset (n=27 cases and n=27 controls). Yeast data were not available for the Partners in Prevention HSV/HIV Study, Couples Observational Study, and Partners PrEP study. †††n=192 for full cohort (n=49 cases and n=143 controls) and n=76 for deep sequencing subset (n=38 cases and n=38 controls). Antibiotic use data were not available for the Partners in Prevention HSV/HIV Study or the Couples Observational Study. Data on antibiotic use in the past 90 days were not available for the study visit for 73 women from the other cohorts. In the Partners PrEP Study, only antibiotics provided for sexually transmitted infections were captured.

Table 2: **Baseline characteristics**

We did the analyses using IBM SPSS Statistics (version 23), Stata (version 13), and R (version 3.3.2) using the ggplot2 and RColorBrewer packages.

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

Between November, 2004, and August, 2014, we identified 87 women who acquired HIV infection (cases) and 262 controls who did not acquire HIV infection. For

	Median relative abundance (%)		SD	Odds ratio (95% CI) per 1-SD change	p value
	Cases (n=55)	Controls (n=55)			
Lactobacillus iners	8·36 (0–99·85)	76·46 (0–99·94)	42·02	0·54 (0·36–0·80)	0·002
Gemella asaccharolytica	0·01 (0–5·70)	0 (0–0·82)	0·94	12·01 (2·26–63·78)	0·004
Eggerthella species type 1	0 (0–2·47)	0 (0–1·70)	0·51	2·06 (1·23–3·45)	0·006
Leptotrichia amnionii	0·01 (0–26·40)	0 (0–13·71)	5·40	2·67 (1·26–5·65)	0·01
Dialister species type 2	0 (0–7·61)	0 (0–5·14)	1·70	1·85 (1·11–3·08)	0·02
Parvimonas micra	0 (0–9·98)	0 (0–2·31)	1·12	3·26 (1·17–9·07)	0·02
Dialister micraerophilus	0·08 (0–1·60)	0 (0–1·43)	0·33	1·62 (1·03–2·53)	0·04
*Dialister**	0 (0–0·29)	0 (0–0·10)	0·04	2·17 (1·04–4·53)	0·04
*Mycoplasma**	0 (0–1·53)	0 (0–0·02)	0·17	$1·09 \times 10^{8}$ ($0·67–1·79 \times 10^{16}$)†	0·06
Lactobacillus reuteri/ *Lactobacillus vaginalis*	0 (0–2·76)	0 (0–15·15)	1·90	0·33 (0·10–1·07)	0·07
Prevotella timonensis	0·10 (0–16·68)	0 (0–19·11)	3·96	1·52 (0·97–2·37)	0·07
BVAB2	0 (0–8·66)	0 (0–5·10)	1·31	1·65 (0·96–2·82)	0·07
Mycoplasma hominis	0·03 (0–3·59)	0 (0–1·24)	0·61	2·36 (0·92–6·08)	0·08
Corynebacterium amycolatum	0 (0–0·06)	0 (0–0·19)	0·03	0·51 (0·24–1·08)	0·08
*Aerococcus**	0 (0–2·66)	0 (0–0·85)	0·31	2·01 (0·89–4·58)	0·09
Porphyromonas	0 (0–4·46)	0 (0–0·80)	0·49	2·24 (0·86–5·83)	0·10
Megasphaera species type 1	0 (0–17·76)	0 (0–4·39)	2·45	2·57 (0·82–8·07)	0·11
*Prevotella**	0·04 (0–15·12)	0 (0–3·32)	1·71	1·97 (0·86–4·52)	0·11
*Coriobacteriaceae**	0 (0–0·01)	0 (0–0·08)	0·01	0·35 (0·09–1·33)	0·12
Prevotella bivia	0 (0–68·15)	0 (0–32·23)	9·54	1·75 (0·86–3·59)	0·13
*Megasphaera**	0 (0–0·81)	0 (0–0·17)	0·10	2·05 (0·80–5·24)	0·13

Data are median (range), unless otherwise stated. Only bacterial taxa that were associated with HIV acquisition with p≤0·15 were included. BVAB2=bacterial vaginosis associated bacterium 2. *Some taxa were identified only to the genus level, whereas others were identified to the species level. In instances for which this level of identification was used, the genus level group excluded the individual species that were identified separately. †OR and CI are highly sensitive to two outlier *Mycoplasma* values; when these values are removed, the OR was 30·23 (95% CI 0·93–984·53; p=0·06).

***Table 3*: Univariate logistic regression comparing relative abundance of vaginal bacteria identified by pyrosequencing in women who acquired HIV infection with women who remained uninfected**

72 (82·8%) of 87 cases, vaginal samples were collected with a median of 141 days (IQR 84–250) before HIV detection. For the remaining 15 (17·2%) cases, samples were collected during acute or early HIV infection. Table 2 presents the baseline characteristics.

The appendix (p 22) shows the overall vaginal bacterial community diversity in 55 cases versus 55 controls evaluated by pyrosequencing. The Shannon Diversity Index was significantly higher in cases (median 1·3, IQR 0·4–2·3) than in controls (0·7, 0·1–1·5; p=0·03). Community richness, using the Chao1 Index, was also higher in cases (median 38·5, IQR 14·0–59·0) than in controls (25·0, 10·0–54·2), although this difference was not significant (p=0·17). Histograms illustrating Shannon and Chao1 distributions are shown in the appendix (p 23).

Relative abundance of individual bacterial taxa was compared between cases and controls. Table 3 summarises the taxa that showed a statistical trend towards an association with HIV acquisition (p<0·15). Higher relative abundance of *Dialister* genus (OR 2·17, 95% CI 1·04–4·53), *Dialister* species type 2 (1·85, 1·11–3·08), *Dialister micraerophilus* (1·62, 1·03–2·53), *Gemella asaccharolytica* (12·01, 2·26–63·78), *Eggerthella* species type 1 (2·06, 1·23–3·45), *Parvimonas micra* (3·26, 1·17–9·07), and *Leptotrichia amnionii* (2·67, 1·26–5·65) was associated with significantly higher odds of HIV acquisition. By contrast, higher relative abundance of *Lactobacillus iners* (OR 0·54, 95% CI 0·36–0·80) was associated with significantly lower odds of HIV acquisition.

In the univariate analyses of the 20 taxa selected for rtPCR testing, *Parvimonas* species type 1, *G asaccharolytica*, *Mycoplasma hominis*, *Leptotrichia*/*Sneathia*, *Porphyromonas* species type 1, *Parvimonas* species type 2, *G vaginalis*, *Eggerthella* species type 1, and *Megasphaera* showed significant associations with HIV acquisition (table 4). Results were similar after adjustment for potential confounders, although *G vaginalis* and *Porphyromonas* species type 1 were no longer significantly associated with HIV acquisition. Four taxa, *Parvimonas* species type 1 (first tertile adjusted OR [aOR] 1·67, 95% CI 0·61–4·57; second tertile 3·01, 1·13–7·99; third tertile 4·64, 1·73–12·46; p=0·005), *G asaccharolytica* (first tertile 2·09, 1·01–4·36; second tertile 2·02, 0·98–4·17; third tertile 3·03, 1·46–6·30; p=0·010), *M hominis* (first tertile 1·46, 0·69–3·11; second tertile 1·40, 0·66–2·98; third tertile 2·76, 1·36–5·63; p=0·048), and *Leptotrichia*/*Sneathia* (first tertile 2·04, 1·02–4·10; second tertile 1·45, 0·70–3·00; third tertile 2·59, 1·26–5·34; p=0·046), showed associations that were strongest at the highest concentrations. Figure 1 shows the aORs for the highest quantile of each taxon, and illustrates how some species typically associated with vaginal dysbiosis and bacterial vaginosis were strongly associated with HIV acquisition, whereas other bacterial species were not. This analysis identified individual bacterium-specific associations despite strong correlations between many bacteria included in the rtPCR analyses (figure 2).

In the validation subset, after adjustment for potential confounding factors and a 20% false discovery rate, detection by rtPCR was associated with higher odds of HIV acquisition for five of the seven bacteria that showed concentration-related associations in the primary analysis (appendix pp 10, 11): *M hominis* (aOR 2·71, 95% CI 1·13–6·49; p=0·026), *Eggerthella* species type 1 (2·50, 1·07–5·85; p=0·035), *Leptotrichia*/*Sneathia* (2·47, 0·98–6·22; p=0·056), *G asaccharolytica* (2·45, 1·04–5·78; p=0·040), and *Parvimonas* species type 2 (2·43, 1·03–5·70; p=0·042).

The Nugent scoring system was applied to 76 cases and 230 controls with vaginal Gram stains available. Compared with women with normal microbiota, those with intermediate microbiota (aOR 2·50, 95% CI 1·15–5·40) and bacterial vaginosis (2·10, 1·14–3·88) had increased risk for HIV acquisition (joint test p=0·018).

Several sensitivity analyses were applied to the seven bacterial taxa associated with HIV acquisition in our primary multivariable analysis. Associations were similar in female sex workers, pregnant and post-partum women, and serodiscordant couples (appendix p 12), despite the differences in demographics, risk factors, and

	Concentration (log_{10} genome copies per swab)	Total (n=349)	Cases (n=87)	Controls (n=262)	Odds ratio (95% CI)	p value	Adjusted odds ratio* (95% CI)	p value
Aerococcus christensenii	..	..	..	..	..	0·257	..	0·350
Undetectable	..	170 (48·7%)	35 (40·2%)	135 (51·5%)	1·00	..	1·00	..
Tertile 1	2·2–5·2	60 (17·2%)	16 (18·4%)	44 (16·8%)	1·36 (0·67–2·73)	..	1·34 (0·64–2·80)	..
Tertile 2	5·3–6·9	60 (17·2%)	17 (19·5%)	43 (16·4%)	1·58 (0·79–3·14)	..	1·42 (0·70–2·91)	..
Tertile 3	6·9–8·8	59 (16·9%)	19 (21·8%)	40 (15·3%)	1·91 (0·96–3·80)	..	1·90 (0·92–3·89)	..
Atopobium vaginae	..	..	..	..	..	0·529	..	0·777
Undetectable	..	134 (38·4%)	28 (32·2%)	106 (40·5%)	1·00	..	1·00	..
Tertile 1	2·0–6·1	71 (20·3%)	18 (20·7%)	53 (20·2%)	1·28 (0·65–2·51)	..	1·24 (0·61–2·49)	..
Tertile 2	6·1–7·6	72 (20·6%)	21 (24·1%)	51 (19·5%)	1·56 (0·81–3·00)	..	1·37 (0·69–2·70)	..
Tertile 3	7·6–9·1	72 (20·6%)	20 (23·0%)	52 (19·8%)	1·48 (0·75–2·92)	..	1·35 (0·66–2·74)	..
BVAB2	..	..	..	..	..	0·063	..	0·127
Undetectable	..	226 (64·8%)	46 (52·9%)	180 (68·7%)	1·00	..	1·00	..
Tertile 1	2·0–5·1	41 (11·7%)	13 (14·9%)	28 (10·7%)	1·79 (0·86–3·70)	..	1·78 (0·83–3·81)	..
Tertile 2	5·2–6·6	41 (11·7%)	15 (17·2%)	26 (9·9%)	2·26 (1·11–4·62)	..	2·10 (1·00–4·44)	..
Tertile 3	6·6–8·1	41 (11·7%)	13 (14·9%)	28 (10·7%)	1·85 (0·87–3·91)	..	1·72 (0·78–3·78)	..
Dialister micraerophilus	..	..	..	..	..	0·086	..	0·305
Undetectable	..	76 (21·8%)	13 (14·9%)	63 (24·0%)	1·00	..	1·00	..
Tertile 1	2·3–5·0	92 (26·4%)	20 (23·0%)	72 (27·5%)	1·41 (0·64–3·08)	..	1·32 (0·58–2·98)	..
Tertile 2	5·1–6·6	90 (25·8%)	25 (28·7%)	65 (24·8%)	1·98 (0·91–4·32)	..	1·55 (0·69–3·50)	..
Tertile 3	6·6–8·5	91 (26·1%)	29 (33·3%)	62 (23·7%)	2·56 (1·16–5·66)	..	2·14 (0·94–4·92)	..
Dialister species type 2	..	..	..	..	..	0·061	..	0·165
Undetectable	..	162 (46·4%)	35 (40·2%)	127 (48·5%)	1·00	..	1·00	..
Tertile 1	2·0–5·2	62 (17·8%)	11 (12·6%)	51 (19·5%)	0·78 (0·37–1·66)	..	0·67 (0·30–1·48)	..
Tertile 2	5·3–7·3	63 (18·1%)	19 (21·8%)	44 (16·8%)	1·52 (0·78–2·96)	..	1·32 (0·66–2·64)	..
Tertile 3	7·3–8·8	62 (17·8%)	22 (25·3%)	40 (15·3%)	2·09 (1·07–4·07)	..	1·72 (0·85–3·50)	..
Eggerthella species type 1	..	..	..	..	..	0·020	..	0·041
Undetectable	..	186 (53·3%)	35 (40·2%)	151 (57·6%)	1·00	..	1·00	..
Tertile 1	2·3–6·0	55 (15·8%)	16 (18·4%)	39 (14·9%)	1·77 (0·89–3·50)	..	1·79 (0·88–3·64)	..
Tertile 2	6·0–7·5	54 (15·5%)	21 (24·1%)	33 (12·6%)	2·77 (1·43–5·36)	..	2·62 (1·31–5·22)	..
Tertile 3	7·5–9·1	54 (15·5%)	15 (17·2%)	39 (14·9%)	1·72 (0·84–3·55)	..	1·53 (0·72–3·28)	..
Gemella asaccharolytica	..	..	..	..	..	0·003	..	0·010
Undetectable	..	196 (56·2%)	35 (40·2%)	161 (61·5%)	1·00	..	1·00	..
Tertile 1	2·0–5·2	51 (14·6%)	15 (17·2%)	36 (13·7%)	1·91 (0·95–3·87)	..	2·09 (1·01–4·36)	..
Tertile 2	5·3–6·7	51 (14·6%)	17 (19·5%)	34 (13·0%)	2·34 (1·17–4·69)	..	2·02 (0·98–4·17)	..
Tertile 3	6·7–8·8	51 (14·6%)	20 (23·0%)	31 (11·8%)	3·22 (1·59–6·49)	..	3·03 (1·46–6·30)	..
Gardnerella vaginalis	..	..	..	..	..	0·024	..	0·077
Undetectable	..	38 (10·9%)	5 (5·7%)	33 (12·6%)	1·00	..	1·00	..
Tertile 1	2·3–6·6	104 (29·8%)	21 (24·1%)	83 (31·7%)	1·68 (0·59–4·80)	..	1·97 (0·67–5·80)	..
Tertile 2	6·6–8·7	104 (29·8%)	36 (41·4%)	68 (26·0%)	3·52 (1·26–9·80)	..	3·24 (1·13–9·27)	..
Tertile 3	8·8–10·3	103 (29·5%)	25 (28·7%)	78 (29·8%)	2·18 (0·75–6·35)	..	1·90 (0·63–5·74)	..
Lactobacillus crispatus	..	..	..	..	..	0·625	..	0·469
Undetectable	..	268 (76·8%)	69 (79·3%)	199 (76·0%)	1·00	..	1·00	..
Tertile 1	2·1–5·3	27 (7·7%)	4 (4·6%)	23 (8·8%)	0·51 (0·17–1·51)	..	0·44 (0·14–1·37)	..
Tertile 2	5·4–7·5	27 (7·7%)	7 (8·0%)	20 (7·6%)	1·01 (0·41–2·49)	..	1·07 (0·42–2·75)	..
Tertile 3	7·6–9·1	27 (7·7%)	7 (8·0%)	20 (7·6%)	1·01 (0·41–2·51)	..	1·13 (0·44–2·91)	..
Lactobacillus iners	..	..	..	..	..	0·896	..	0·946
Undetectable	..	61 (17·5%)	17 (19·5%)	44 (16·8%)	1·00		1·00	..
Tertile 1	1·9–7·3	97 (27·8%)	25 (28·7%)	72 (27·5%)	0·92 (0·44–1·92)	..	1·10 (0·51–2·41)	..
Tertile 2	7·3–8·7	96 (27·5%)	22 (25·3%)	74 (28·2%)	0·76 (0·36–1·60)	..	0·88 (0·40–1·93)	..
Tertile 3	8·7–9·9	95 (27·2%)	23 (26·4%)	72 (27·5%)	0·81 (0·38–1·73)	..	0·94 (0·42–2·09)	..

(Table 4 continues on next page)

	Concentration (log_{10} genome copies per swab)	Total (n=349)	Cases (n=87)	Controls (n=262)	Odds ratio (95% CI)	p value	Adjusted odds ratio* (95% CI)	p value
(Continued from previous page)								
Lactobacillus jensenii	..	..	..	..	..	0·564	..	0·589
Undetectable	..	301 (86·2%)	78 (89·7%)	223 (85·1%)	1·00	..	1·00	..
Tertile 1	2·0–4·4	16 (4·6%)	4 (4·6%)	12 (4·6%)	0·96 (0·30–3·07)	..	1·21 (0·36–4·05)	..
Tertile 2	4·5–6·2	16 (4·6%)	3 (3·4%)	13 (5·0%)	0·66 (0·18–2·36)	..	0·61 (0·16–2·28)	..
Tertile 3	6·3–8·2	16 (4·6%)	2 (2·3%)	14 (5·3%)	0·41 (0·09–1·83)	..	0·44 (0·09–2·06)	..
Leptotrichia/Sneathia	..	..	..	..	..	0·025	..	0·046
Undetectable	..	143 (41·0%)	25 (28·7%)	118 (45·0%)	1·00	..	1·00	..
Tertile 1	1·7–4·7	69 (19·8%)	20 (23·0%)	49 (18·7%)	1·98 (1·01–3·91)	..	2·04 (1·02–4·10)	..
Tertile 2	4·7–7·5	68 (19·5%)	18 (20·7%)	50 (19·1%)	1·73 (0·86–3·46)	..	1·45 (0·70–3·00)	..
Tertile 3	7·5–9·6	69 (19·8%)	24 (27·6%)	45 (17·2%)	2·78 (1·39–5·57)	..	2·59 (1·26–5·34)	..
Megasphaera	..	..	..	..	..	0·024	..	0·038
Undetectable	..	233 (66·8%)	48 (55·2%)	185 (70·6%)	1·00	..	1·00	..
Tertile 1	2·1–6·3	39 (11·2%)	17 (19·5%)	22 (8·4%)	3·00 (1·47–6·14)	..	3·15 (1·45–6·81)	..
Tertile 2	6·3–7·7	39 (11·2%)	11 (12·6%)	28 (10·7%)	1·50 (0·70–3·22)	..	1·43 (0·65–3·14)	..
Tertile 3	7·7–9·0	38 (10·9%)	11 (12·6%)	27 (10·3%)	1·57 (0·71–3·50)	..	1·32 (0·57–3·05)	..
Mycoplasma hominis	..	..	..	..	..	0·018	..	0·048
Undetectable	..	206 (59·0%)	41 (47·1%)	165 (63·0%)	1·00	..	1·00	..
Tertile 1	2·0–5·1	48 (13·8%)	13 (14·9%)	35 (13·4%)	1·48 (0·72–3·03)	..	1·46 (0·69–3·11)	..
Tertile 2	5·1–6·5	47 (13·5%)	13 (14·9%)	34 (13·0%)	1·58 (0·76–3·28)	..	1·40 (0·66–2·98)	..
Tertile 3	6·5–8·4	48 (13·8%)	20 (23·0%)	28 (10·7%)	3·02 (1·52–6·01)	..	2·76 (1·36–5·63)	..
Parvimonas species type 1	..	..	..	..	..	0·002	..	0·005
Undetectable	..	287 (82·2%)	60 (69·0%)	227 (86·6%)	1·00	..	1·00	..
Tertile 1	2·4–3·1	21 (6·0%)	7 (8·0%)	14 (5·3%)	1·99 (0·77–5·18)	..	1·67 (0·61–4·57)	..
Tertile 2	3·1–4·2	21 (6·0%)	9 (10·3%)	12 (4·6%)	3·02 (1·20–7·60)	..	3·01 (1·13–7·99)	..
Tertile 3	4·2–7·0	20 (5·7%)	11 (12·6%)	9 (3·4%)	4·93 (1·92–12·63)	..	4·64 (1·73–12·46)	..
Parvimonas species type 2	..	..	..	..	..	0·001	..	0·004
Undetectable	..	226 (64·8%)	45 (51·7%)	181 (69·1%)	1·00	..	1·00	..
Tertile 1	2·4–5·4	41 (11·7%)	19 (21·8%)	22 (8·4%)	3·82 (1·82–7·98)	..	3·52 (1·63–7·61)	..
Tertile 2	5·4–7·0	41 (11·7%)	8 (9·2%)	33 (12·6%)	0·95 (0·41–2·20)	..	0·85 (0·36–2·02)	..
Tertile 3	7·0–8·2	41 (11·7%)	15 (17·2%)	26 (9·9%)	2·30 (1·10–4·80)	..	2·18 (1·01–4·72)	..
Porphyromonas asaccharolytica/ Porphyromonas uenonis	..	..	..	..	..	0·673	..	0·740
Undetectable	..	76 (21·8%)	16 (18·4%)	60 (22·9%)	1·00	..	1·00	..
Tertile 1	2·0–3·4	92 (26·4%)	26 (30·0%)	66 (25·2%)	1·55 (0·73–3·28)	..	1·40 (0·65–3·05)	..
Tertile 2	3·5–5·2	90 (25·8%)	21 (24·1%)	69 (26·3%)	1·21 (0·55–2·67)	..	1·00 (0·44–2·26)	..
Tertile 3	5·2–8·6	91 (26·1%)	24 (27·6%)	67 (25·6%)	1·44 (0·64–3·23)	..	1·22 (0·52–2·86)	..
Porphyromonas species type 1	..	..	..	..	..	0·029	..	0·068
Undetectable	..	204 (58·5%)	43 (49·4%)	161 (61·5%)	1·00	..	1·00	..
Tertile 1	2·0–3·3	49 (14·0%)	11 (12·6%)	38 (14·5%)	1·18 (0·54–2·55)	..	1·14 (0·52–2·52)	..
Tertile 2	3·4–4·9	48 (13·8%)	13 (14·9%)	35 (13·4%)	1·48 (0·71–3·07)	..	1·22 (0·57–2·62)	..
Tertile 3	4·9–8·0	48 (13·8%)	20 (23·0%)	28 (10·7%)	2·93 (1·45–5·94)	..	2·74 (1·30–5·76)	..
Porphyromonas bennonis	..	..	..	..	..	0·907	..	0·951
Undetectable	..	245 (70·2%)	62 (71·3%)	183 (69·8%)	1·00	..	1·00	..
Tertile 1	1·9–2·6	35 (10·0%)	7 (8·0%)	28 (10·7%)	0·73 (0·30–1·79)	..	0·79 (0·31–1·97)	..
Tertile 2	2·6–3·6	34 (9·7%)	9 (10·3%)	25 (9·5%)	1·06 (0·46–2·43)	..	0·99 (0·42–2·33)	..
Tertile 3	3·6–6·4	35 (10·0%)	9 (10·3%)	26 (9·9%)	1·01 (0·42–2·44)	..	0·85 (0·34–2·13)	..

(Table 4 continues on next page)

	Concentration (log_{10} genome copies per swab)	Total (n=349)	Cases (n=87)	Controls (n=262)	Odds ratio (95% CI)	p value	Adjusted odds ratio* (95% CI)	p value
(Continued from previous page)								
Prevotella	..	..	..	..	..	0·343	..	0·543
Quartile 1	0–4·3	88 (25·2%)	18 (20·7%)	70 (26·7%)	1·00	..	1·00	..
Quartile 2	4·4–5·9	87 (24·9%)	20 (23·0%)	67 (25·6%)	1·26 (0·59–2·69)	..	1·30 (0·60–2·83)	..
Quartile 3	5·9–7·9	87 (24·9%)	22 (25·3%)	65 (24·8%)	1·40 (0·67–2·91)	..	1·26 (0·59–2·68)	..
Quartile 4	8·0–9·8	87 (24·9%)	27 (31·0%)	60 (22·9%)	1·98 (0·92–4·26)	..	1·80 (0·80–4·04)	..

Data are n (%), unless otherwise stated. BVAB2=bacterial vaginosis associated bacterium 2. *Adjustment was made in a multivariable model stratified by confounding variables.

Table 4: **Odds ratios and adjusted odds ratios showing the association between vaginal bacterial concentration by real-time PCR and HIV acquisition in women who became HIV infected versus women who remained HIV uninfected**

incidence observed in these cohorts (table 1). Additionally, analyses limited to cases where samples were collected less than 90 days before identification of HIV infection, excluding cases with samples from acute HIV infection, excluding cases with sexually transmitted infections, incorporating adjustment for vaginal washing, or stratifying by swab collection method had similar point estimates compared with the primary analyses; however, despite the similar point estimates, some associations were no longer significant, possibly because of the smaller sample sizes in these subset analyses (appendix pp 13–21).

Discussion

Using two sophisticated bacterial PCR approaches and data from five cohorts spanning six sub-Saharan African countries, this study was the first to show significant associations between the quantity of specific vaginal bacteria and women's risk of HIV acquisition. Concentrations of *Parvimonas* species types 1 and 2, *G asaccharolytica*, *M hominis*, *Leptotrichia/Sneathia*, *Eggerthella* species type 1, and *Megasphaera* were significantly associated with increased HIV risk. There were strong correlations between concentrations of many of the 20 bacteria evaluated with rtPCR, suggesting they might frequently be found together, establishing high-risk bacterial communities. These findings were consistent across three distinct risk groups: female sex workers, pregnant and post-partum women, and women in serodiscordant relationships.

Vaginal microbiota could influence women's risk of HIV acquisition at multiple levels.[23] First, genital inflammation, mediated by the presence of particular bacterial taxa or communities, is likely to influence HIV susceptibility.[9,24] A recent study identified six vaginal bacterial genera independently associated with pro-inflammatory cytokines.[5] Two of these genera, *Sneathia* and *Gemella*, correspond to vaginal bacteria showing concentration-dependent associations with HIV risk in the present analysis. Second, vaginal dysbiosis has been associated with HIV-inducing factors in vaginal fluid.[25] Third, many bacteria associated with bacterial vaginosis produce sialidases and mucinases that disrupt the protective cervicovaginal mucus layer.[7]

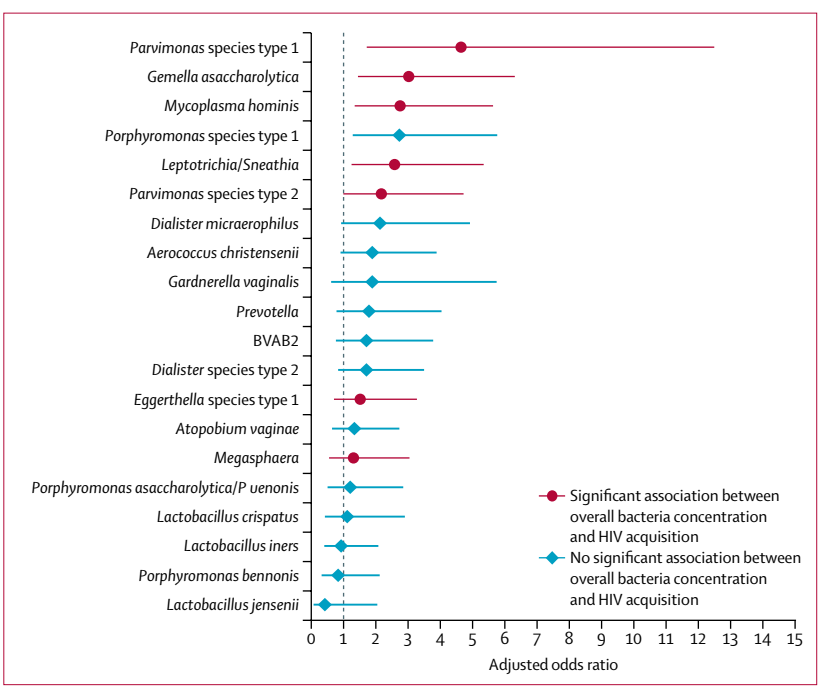

Figure 1: **Association between the highest bacterial quantile of concentration and HIV acquisition for each of the 20 bacterial taxa**

Error bars are 95% CIs. Quantiles were defined as undetectable, first tertile, second tertile, and third tertile; however, *Prevotella* was modelled as four quartiles as it had few samples with bacterial DNA concentrations that were not detected. BVAB2=bacterial vaginosis associated bacterium 2.

In one earlier study, the presence of cultivable *Lactobacillus* species was associated with a decreased risk of acquiring HIV.[26] Additionally, a recent study using molecular characterisation of vaginal microbiota found that women with vaginal bacterial communities deficient in non-*iners* species of *Lactobacillus* were at increased risk for HIV infection.[9] The present analysis showed an association between lower relative abundance of *L iners* and HIV acquisition. However, the primary analysis using rtPCR assays in the full dataset showed no significant associations between concentrations of

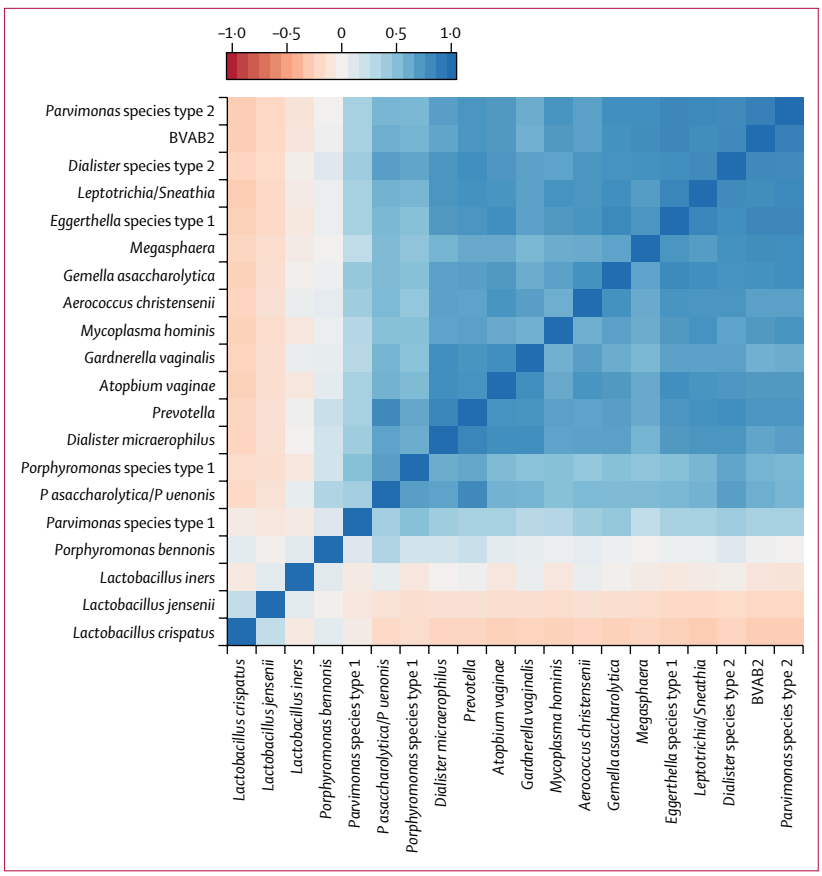

Figure 2: **Heat map illustrating Spearman's correlation for quantity of the 20 bacterial taxa**
Cohort data were of 349 women from eastern and southern Africa. BVAB2=bacterial vaginosis associated bacterium 2. *P asaccharolytica=Porphyromonas asaccharolytica.*

Lactobacillus species (*L iners, L crispatus,* and *L jensenii*) and HIV acquisition. Given the negative correlations between concentrations of lactobacilli and bacterial vaginosis-associated species, low relative abundance of *Lactobacillus* species might simply reflect the presence of high concentrations of bacterial vaginosis-associated bacteria that affect HIV susceptibility.

This study used a novel sequential experimental approach that used complementary methods to evaluate the vaginal microbiome. In the first step, relative abundance data were used to show the association between vaginal bacterial diversity and HIV acquisition, and to guide selection of a restricted set of bacteria for further investigation. In the second step, highly sensitive taxon-directed rtPCR assays were used to test the hypothesis that concentrations of 20 key bacteria would be associated with HIV risk. The two steps captured related but distinct exposures. Specifically, relative abundance is not the same as absolute quantity of a bacterial taxon. Additionally, the rtPCR assays are more sensitive but might be less specific for detection of individual bacteria than use of broad range PCR with pyrosequencing.

This study included a validation analysis in a subset of samples that were not included in the pyrosequencing step, facilitating independent testing of hypotheses generated from pyrosequencing data. Two important points should be considered in comparing the primary analysis with the validation analysis. First, because of the smaller sample size in the validation analysis than in the primary analysis, bacterial taxa were modelled in a binary fashion rather than as four quantiles. Second, the validation subset was not a randomly generated group of cases and controls, so distribution across the five cohorts diverged from that of the full dataset. Despite these caveats, the validation confirmed an association between bacterial taxa and HIV acquisition for five of the seven bacteria identified in the primary analysis.

An important strength of this study was the large and geographically diverse sample, with individuals representing three distinct risk groups. A further strength in the study design was the collection of vaginal microbiota samples before HIV acquisition in more than 80% of cases, and shortly after HIV acquisition in the remainder of cases. Furthermore, the analyses were robust in multiple sensitivity analyses testing assumptions in the experimental approach. This study also had limitations. First, as an observational study, these analyses do not provide definitive evidence that the associations detected are caused by bacteria increasing HIV susceptibility. Second, these analyses did not explore mechanisms through which individual bacterial taxa might increase HIV risk. Such mechanistic data will help to further evaluate the likelihood of a causal link between vaginal bacteria and HIV susceptibility, and will be the focus of future studies. Third, despite adjustment for potential confounding factors, residual confounding is possible because of measurement error or unmeasured confounding factors. Fourth, longer intervals between sample collection and HIV acquisition could attenuate the observed associations,[27] although this source of variability was minimised by avoiding sample collection during menses. Fifth, sampling methods and laboratory procedures varied across the five cohorts included in this analysis. Related to this point, laboratory data on sexually transmitted infections and vaginal yeast were not available at all analysis visits, so the primary analysis does not include adjustment for these conditions. Finally, despite inclusion of multiple risk groups in this study, all participants were from eastern and southern Africa. Although this represents the region affected the most by the HIV epidemic, the findings might not be generalisable to all geographical regions.

Higher diversity vaginal bacterial communities not dominated by lactobacilli are more common in African and Hispanic women than in women of Asian or European origin,[4,28] leading to the hypothesis that racial differences in vaginal microbiota might contribute to population-level differences in HIV transmission and prevalence.[29] Underlining this point, recent studies

suggest that vaginal dysbiosis accounts for 20–30% of the population attributable risk of HIV acquisition in African women.[30,31] Because bacterial vaginosis is an extremely heterogeneous condition,[3,4] defining individual vaginal bacteria that are associated with HIV risk in women could provide additional specific targets and inform future strategies for HIV prevention research.

Contributors

RSM, JRL, and DNF developed the study concept. RSM, JRL, GCJ-S, JMB, CC, JO, and DNF secured funding for the parent P01 grant and for the individual cohorts that contributed samples and data. RSM, JRL, JK, GCJ-S, WJ, KNM, NRM, CRC, JMB, and CC oversaw clinical field work. DNF oversaw the molecular microbiological work. SS and TLF developed the real-time (rt) PCR assays. TLF and MMM applied the rtPCR assays to the samples. TLF did the rtPCR assays. SS and MMM generated and curated the pyrosequencing data. RSM, JRL, BAR, and KY developed the statistical analysis plan. KY did the statistical analyses. RSM and JRL developed the initial draft of the manuscript. All authors contributed to editing of the manuscript and approved the final draft for submission.

Declaration of interests

RSM receives research funding, paid to the University of Washington (Seattle, WA, USA), from Hologic/Gen-Probe for a study of human papilloma virus screening. JRL is on the scientific advisory board for Prosetta Biosciences. All other authors declare no competing interests.

Acknowledgments

This study was supported by the National Institute of Child Health and Human Development of the National Institutes of Health (NIH; P01-HD64915). Support for the Mombasa Cohort was through the National Institute of Allergy and Infectious Diseases of the NIH (R37-AI38518). Support for the clinical trials in HIV-serodiscordant couples was through the Bill & Melinda Gates Foundation (OPP47674, 26469, and 41185). Infrastructure and logistical support was received from the University of Washington and the Fred Hutchinson Cancer Research Center's Center for AIDS Research (P30-AI27757). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. We acknowledge the contributions of our clinical, laboratory, and administrative staff in Kenya, Uganda, Tanzania, Zambia, Botswana, South Africa, and the USA. We also acknowledge the women in the cohorts that contributed to this analysis.

References

1 UNAIDS. Global AIDS update 2016. 2016. http://www.unaids.org/sites/default/files/media_asset/global-AIDS-update-2016_en.pdf (accessed Jan 18, 2018).

2 Low N, Chersich MF, Schmidlin K, et al. Intravaginal practices, bacterial vaginosis, and HIV infection in women: individual participant data meta-analysis. *PLoS Med* 2011; **8:** e1000416.

3 Fredricks DN, Fiedler TN, Marrazzo JM. Molecular identification of bacteria associated with bacterial vaginosis. *N Engl J Med* 2005; **353:** 1899–911.

4 Ravel J, Gajer P, Abdo Z, et al. Vaginal microbiome of reproductive-age women. *Proc Natl Acad Sci USA* 2011; **108** (suppl 1)**:** 4680–87.

5 Anahtar MN, Byrne EH, Doherty KE, et al. Cervicovaginal bacteria are a major modulator of host inflammatory responses in the female genital tract. *Immunity* 2015; **42:** 965–76.

6 Fredricks DN, Fiedler TL, Thomas KK, Oakley BB, Marrazzo JM. Targeted PCR for detection of vaginal bacteria associated with bacterial vaginosis. *J Clin Microbiol* 2007; **45:** 3270–76.

7 Srinivasan S, Hoffman NG, Morgan MT, et al. Bacterial communities in women with bacterial vaginosis: high resolution phylogenetic analyses reveal relationships of microbiota to clinical criteria. *PLoS One* 2012; **7:** e37818.

8 Srinivasan S, Morgan MT, Fiedler TL, et al. Metabolic signatures of bacterial vaginosis. *mBio* 2015; **6:** e00204–15.

9 Gosmann C, Anahtar MN, Handley SA, et al. Lactobacillus-deficient cervicovaginal bacterial communities are associated with increased HIV acquisition in young South African women. *Immunity* 2017; **46:** 1–9.

10 McClelland RS, Richardson BA, Cherutich P, et al. A 15-year study of the impact of community antiretroviral therapy coverage on HIV incidence in Kenyan female sex workers. *AIDS* 2015; **29:** 2279–86.

11 Kinuthia J, Drake AL, Matemo D, et al. HIV acquisition during pregnancy and postpartum is associated with genital infections and partnership characteristics. *AIDS* 2015; **29:** 2025–33.

12 Celum C, Wald A, Lingappa JR, et al. Acyclovir and transmission of HIV-1 from persons infected with HIV-1 and HSV-2. *N Engl J Med* 2010; **362:** 427–39.

13 Lingappa JR, Petrovski S, Kahle E, et al. Genomewide association study for determinants of HIV-1 acquisition and viral set point in HIV-1 serodiscordant couples with quantified virus exposure. *PLoS One* 2011; **6:** e28632.

14 Baeten JM, Donnell D, Ndase P, et al. Antiretroviral prophylaxis for HIV prevention in heterosexual men and women. *N Engl J Med* 2012; **367:** 399–410.

15 Gorgos LM, Sycuro LK, Srinivasan S, et al. Relationship of specific bacteria in the cervical and vaginal microbiotas with cervicitis. *Sex Transm Dis* 2015; **42:** 475–81.

16 Chao A. Nonparametric estimation of the number of classes in a population. *Scand J Stat* 1984; **11:** 265–70.

17 Shannon CE. A mathematical theory of communication. *Bell Syst Tech J* 1948; **27:** 379–423.

18 Swidsinski A, Mendling W, Loening-Baucke V, et al. An adherent *Gardnerella vaginalis* biofilm persists on the vaginal epithelium after standard therapy with oral metronidazole. *Am J Obstet Gynecol* 2008; **198:** 97.

19 Langholz B, Goldstein L. Conditional logistic analysis of case-control studies with complex sampling. *Biostatistics* 2001; **2:** 63–84.

20 Nugent RP, Krohn MA, Hillier SL. Reliability of diagnosing bacterial vaginosis is improved by a standardized method of Gram stain interpretation. *J Clin Microbiol* 1991; **29:** 297–301.

21 McClelland RS, Lavreys L, Hassan WM, Mandaliya K, Ndinya-Achola JO, Baeten JM. Vaginal washing and increased risk of HIV-1 acquisition among African women: a 10-year prospective study. *AIDS* 2006; **20:** 269–73.

22 McClelland RS, Richardson BA, Graham SM, et al. A prospective study of risk factors for bacterial vaginosis in HIV-1-seronegative African women. *Sex Transm Dis* 2008; **35:** 617–23.

23 Schellenberg JJ, Plummer FA. The microbiological context of HIV resistance: vaginal microbiota and mucosal inflammation at the viral point of entry. *Int J Inflam* 2012; **2012:** 131243.

24 Passmore JA, Jaspan HB, Masson L. Genital inflammation, immune activation and risk of sexual HIV acquisition. *Curr Opin HIV AIDS* 2016; **11:** 156–62.

25 Cohn JA, Hashemi FB, Camarca M, et al. HIV-inducing factor in cervicovaginal secretions is associated with bacterial vaginosis in HIV-1-infected women. *J Acquir Immune Defic Syndr* 2005; **39:** 340–46.

26 Martin HL, Richardson BA, Nyange PM, et al. Vaginal lactobacilli, microbial flora, and risk of human immunodeficiency virus type 1 and sexually transmitted disease acquisition. *J Infect Dis* 1999; **180:** 1863–68.

27 Srinivasan S, Liu C, Mitchell CM, et al. Temporal variability of human vaginal bacteria and relationship with bacterial vaginosis. *PLoS One* 2010; **5:** e10197.

28 Zhou X, Brown CJ, Abdo Z, et al. Differences in the composition of vaginal microbial communities found in healthy caucasian and black women. *ISME J* 2007; **1:** 121–33.

29 Buve A, Jespers V, Crucitti T, Fichorova RN. The vaginal microbiota and susceptibility to HIV. *AIDS* 2014; **28:** 2333–44.

30 van de Wijgert JH, Morrison CS, Brown J, et al. Disentangling contributions of reproductive tract infections to HIV acquisition in African women. *Sex Transm Dis* 2009; **36:** 357–64.

31 Masese L, Baeten JM, Richardson BA, et al. Changes in the contribution of genital tract infections to HIV acquisition among Kenyan high-risk women from 1993 to 2012. *AIDS* 2015; **29:** 1077–85.