The Immune Landscape of Molecular Bacterial Vaginosis and HPV Natural History

Mykhaylo Usyk

Albert Einstein College of Medicine

Nicolas Schlecht

Roswell Park Cancer Institute

Sarah Pickering

Mount Sinai Medical Center

LaShanda Williams

Albert Einstein College of Medicine

Christopher Sollecito

Albert Einstein College of Medicine

Ana Gradissimo

Albert Einstein College of Medicine

Mahboobeh Safian

Roche (United States)

Ligia A Pinto

Frederick National Laboratory for Cancer Research

Rolando Herrero

Agencia Costarricense de Investigaciones Biomédicas

Carolina Porras

Agencia Costarricense de Investigaciones Biomédicas

Mark Shiffman

National Cancer Institute

Shankar Viswanathan

Albert Einstein College of Medicine

Anne Nucci-Sack

Mount Sinai Medical Center

Angela Diaz

Mount Sinai Medical Center

Robert Burk (✉ robert.burk@einsteinmed.org)

Albert Einstein College of Medicine

Article

Keywords: Bacterial vaginosis (BV), sexually transmitted infections, adverse reproductive outcomes

Posted Date: June 22nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-613095/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Bacterial vaginosis (BV) is a highly prevalent condition that is associated with acquisition of sexually transmitted infections and adverse reproductive outcomes. It has been proposed that BV's role as a pathogenic condition is mediated via bacteria-induced local inflammation. However, the complex interplay between vaginal microbes and host immune factors has yet to be clearly elucidated. We demonstrate that 16S rRNA amplicon sequencing and a novel pipeline can be used to generate a molecular Nugent BV score (molBV) corresponding to the Nugent score 0 - 10. This algorithm is independent of the region of 16S rRNA amplified, the sequencing platform and source population. We further identify two local immune cytokine patterns associated with this molecular Nugent score (q -values <0.001). The main immune response is represented by an elevated IL-1 β /IP-10 ratio, whereas a second pattern consists of an increased TNF- α /MIP-1 β ratio. To evaluate the biological significance of molBV-BV and the local immune response, we show that clearance of high-risk HPV (HR-HPV) infections (HZ=1.86, 95% CI: 1.19-2.9) was associated with immune profiles, but not molBV-BV when both were considered in the model. In contrast, the TNF- α /MIP-1 β signature was associated with progression of incident infections to CIN2+ (OR = 2.81, 95% CI: 1.62-5.42), but not HR-HPV clearance. Thus, BV is a heterogeneous condition that activates different arms of the local immune response, which in turn are independent risk factors for HR-HPV clearance and progression.

Introduction

Bacterial vaginosis (BV) is defined as vaginal dysbiosis with inflammation and accompanying symptoms including vaginal discharge^{1,2}. According to Centers for Disease Control and Prevention (CDC) and NHANES studies, the prevalence of BV was 29.2% amongst reproductive aged women living in the United States^{3,4}. Globally this condition is estimated to have an economic burden of approximately \$5-billion per year^{5,6}.

In addition to its ubiquity, BV is a urogenital condition that has been associated with adverse reproductive health outcomes including infertility⁷, increased risk for pre-term birth⁸, and low birth weights⁹. Moreover, an active state of BV is associated with an elevated risk for transmission of a variety of sexually transmitted infections (STIs) ranging from bacterial pathogens such as *Chlamydia*¹⁰ and *Mycoplasma*¹¹, to viral agents including HIV^{12,13} and human papillomavirus (HPV)¹⁴. There is increasing interest in understanding the relationship between the cervicovaginal microbiome (CVM) and HPV natural history and progression to cancer¹⁵⁻¹⁸. In fact, differences in the CVM might explain why some high-risk HPV (HR-HPV) infections clear, while others persist and progress. Lastly, BV is also associated with non-reproductive health issues such as obesity¹⁹.

Clinical BV is primarily diagnosed using Amsel criteria²⁰, which requires the presence of three out of four signs or symptoms: (1) homogeneous, thin, white discharge that smoothly coats the vaginal walls; (2) clue cells in a wet mount; (3) pH of vaginal fluid >4.5 ; and (4) a fishy odor from the vaginal discharge before or after addition of 10% KOH (i.e., whiff test). Although commonly used, this approach has been widely criticized for a considerable rate of misdiagnosis²¹. An alternative to Amsel's criteria is the Nugent score that creates a composite score based on counts of key bacteria morphologically identified on a Gram stain (i.e., *Lactobacillus*, *Gardnerella* and curved gram-negative rods)²². Although this method is more sensitive than the

Amsel criteria²²⁻²⁴, it has been shown to suffer from interobserver variability²⁵ and its use has primarily been limited to research settings due to the amount of time, expertise and costs required to perform the test²⁶. The term, molecular BV has been introduced recently¹³ and there are various meanings depending on the system used for molecular detection and the correlation with clinical, bacteriologic and/or microscopic BV^{2,27}

Bacterial vaginosis has features of an inflammatory state that is associated with alterations of cervicovaginal cytokines at the mucosal level²⁸⁻³¹. A number of studies have reported the association of elevated IL-1 β and BV^{30,31}, whereas most immune markers associated with BV appear to differ across studies. It has been proposed that this variability may be due to small sample sizes, heterogeneity of study populations, and/or differing microbial taxa within the CVM. It is important to identify the source of this variability since the pathogenic effects of BV appear to be associated with local inflammation^{13,31-34}.

In this study, we demonstrate that 16S rRNA amplicon sequencing and an algorithm described herein can reproducibly categorize *molBV*-BV using a Nugent-like 0-10 score across a variety of populations including those from the US and Africa. Using this molecular approach to identify BV, we report the association of a set of cervicovaginal cytokines with *molBV*-BV. In particular, we demonstrate that elevated levels of *Lactobacillus iners* may in part explain the detection of a BV-like inflammatory signature amongst *molBV*-BV negative women. Although there appears to be a predominant host immune response to *molBV*-BV, the CVM can induce alternative forms of inflammation through the activity of specific microbial agents. We utilize these observations to explore risk factors for the rate of clearance and progression of oncogenic HPV. We provide evidence that an inflammatory cervical profile underlies the association of HR-HPV natural history with *molBV*-BV. Surprisingly, the alternative inflammatory pathway is associated with progression of HR-HPV infections to neoplastic lesions. This study provides evidence of multiple host inflammatory pathways associated with the cervicovaginal microbiome that influence the outcome of cervicovaginal HPV infection and possibly other pathologic outcomes of bacterial vaginosis.

Results

Developing a molecular bacterial vaginosis scoring system

Initially, 30 young women with and without symptoms of BV were recruited for evaluation of Nugent, Amsel and 16S amplicon sequencing. Three samples were inadequate for study leaving a training set of 57 participants (mean age = 21, range 15-25 years). Based on Amsel's criteria, 22 were classified as BV-positive; whereas, Nugent score evaluations categorized 26 with BV, 8 as intermediate for BV, and 23 as inconsistent with BV.

The 16S rRNA gene V4 region was amplified from cervicovaginal samples of all 57 participants, as it has been shown to robustly detect bacterial species from the cervicovaginal region³⁵. There was an average of 16,580 $\pm$ 487 16S reads/sample. Fungal sequencing of an ITS1 region amplicon using recently validated primers³⁶ resulted in an average of 16,290 $\pm$ 3,486 ITS reads per sample. Following taxonomic assignments and clustering by Euclidian distances, (**Fig. 1A**) samples formed two primary clusters that were either defined by a dominance of two major species from the genus *Lactobacillus* (n=27) or a state of polymicrobialism (n=30). There was a highly significant tendency of the BV-positive samples to sort to the polymicrobial clade and BV

negative samples to the *Lactobacillus* clade based on either the Amsel ($p < 0.001$) or Nugent BV diagnosis ($p < 0.001$). Hierarchical clustering using fungal communities revealed two primary clades: one dominated by *Candida albicans*, the other with a dominance of *Malassezia restricta* (**Supplemental Fig. 1**). The fungal community clustering showed no significant association with binary BV diagnosis, although some clustering was observed for both the Nugent ($p\text{-value}=0.18$) and Amsel BV-positive samples ($p\text{-value}=0.22$).

All alpha diversity measures (Chao1, Fisher, and Shannon, all $p < 0.001$) were highly associated with both Amsel and Nugent outcomes of BV (**Fig. 1B**). Beta diversity analysis using Jensen-Shannon Divergence (JSD) distances subsampled to 10,000 reads revealed that both the Amsel and Nugent criteria for BV were also significantly associated with the vaginal microbiome $R^2=0.25$, $p < 0.001$ (**Fig. 1C**) and $R^2=0.59$, $p < 0.001$ (**Fig. 1D**), respectively. To identify specific taxa associated with BV using ANCOM, we focused on the samples with concordant results for BV by Amsel and Nugent criteria (**Supplemental Table 1**); 52 differentially abundant genera ($FDR < 0.05$, **Fig. 1E**) were identified, with *Lactobacillus* being the dominant genus elevated in BV-negative women and a mixture of anaerobic gram negative bacteria such as *Gardnerella* elevated in BV-positive, as expected³⁷⁻³⁹. There were no significant associations of fungal alpha or beta measures, or specific fungal taxa identified with BV states (**Supplemental Fig. 2A-C**).

We sought to define a single molecular score from the 16S rRNA amplicon next generation sequencing data that would maximize generalizability of such a metric. Thus, we limited the markers to those taxa present in $>80\%$ of all samples at a relative abundance of $\geq 0.001\%$ after subsampling to 10,000 reads. We identified 11 genera meeting these criteria including *Lactobacillus*, *Prevotella*, *Gardnerella*, *Megasphaera*, *Parvimonas*, *Clostridium*, *Porphyromonas*, *Adlercreutzia*, *Dialister*, *Atopobium* and *Sneathia*. We then derived a *molBV* score using robust regression modeling as described in the Methods section and created an averaged score of microbial taxa ratios for each sample providing a score of 0 – 10 similar to a Nugent score output.

molBV prediction of Nugent BV using 3 independent clinical datasets

molBV was evaluated in publicly available datasets that included 16S rRNA amplicon NGS data and measures of bacterial vaginosis. One testing set contained 388 American women with available 16S data sequenced using the V1-V2 16S rRNA region (different from the V4 region used above) and clinical Nugent scores³⁸. In addition, we identified two African populations, one collected in Cape Town ($n=90$) and the other in Soweto ($n=78$) that sequenced the V4 region of 16S and had measures of BV⁴⁰. We ran the 16S amplicon NGS reads through the *molBV* pipeline to generate Nugent-like scores and observed a strong correlation between the clinical Nugent scores and the *molBV* scores in all three cohorts with ICC values between 0.71 – 0.81 (**Supplemental Fig. 3A**). We next assessed the *molBV* score as a discriminant tool for BV using a dichotomous outcome from the available data (BV-negative = 0-3 or BV-positive = 7-10)⁴¹. The *molBV* score showed high AUC values (0.88 – 0.98) in all three datasets and outperformed other measures of the microbiome such as alpha diversity measures Chao1 and Shannon and the relative abundance of *Lactobacillus* (**Supplemental Fig. 3B-D**). Thus, the *molBV* pipeline is a robust tool to convert 16S NGS data into BV categories independent of 16S amplicon region and population characteristics.

The inflammatory landscape of BV

Previous studies indicated that vaginal dysbiosis is associated with an innate immune response^{42,43}. To further investigate the host immune landscape and bacterial vaginosis, we utilized the *molBV* tool to recapitulate categories of vaginal dysbiosis where other measures of BV were unavailable. We utilized 431 samples from individual women participating in the placebo arm of the Costa Rica Vaccine Trial that had 32 cytokine proteins (i.e., cytokines, chemokines and soluble receptors) quantitated from cervical secretions collected with a sponge (see Methods)⁴⁴. Using three ordinal categories of BV derived from the *molBV* scores equivalent to Nugent BV negative (*molBV* 0-3, n = 179), intermediate (*molBV* 4-7, n = 70) and positive (*molBV* 7-10, n = 182), we identified 13 cytokines significantly associated with a trend across the three BV states (**Fig. 2A**, all markers $q < 0.001$). The strongest association of categorical *molBV* states was with IL-1 β (unit increase OR = 1.73, 95% CI: 1.56-1.92), whereas IP-10 was inversely associated with BV (OR = 0.76, 95% CI: 0.68-0.85). **Supplemental Table 2** shows the cytokines associated with *molBV*-BV in univariate analysis; 6 have been previously associated with BV and 7 additional cytokines are described in this report. Given that inflammation is a complex host response, we analyzed the interrelationships amongst the 32 cytokines to identify patterns of expression. A correlation network was constructed connecting those cytokines with a Pearson correlation > 0.6 (see **Fig. 3B**). Given that 18 of the markers show a strong correlation to at least one other marker, we sought to further improve the identification of differentially abundant inflammatory signals by using cytokine ratios that overcome some issues with compositional data and relative abundances⁴⁵. **Fig. 2C** shows a volcano plot indicating cytokine ratios (in red) with a $q < 3.77 \times 10^{-44}$ threshold. The ratios with the strongest effects were IL-1 β /IP-10 and IP-10/TNF- α , shown at the far right and left, respectively. To further examine the relationships of cytokine ratios, we created a matrix showing the pairwise correlations associated with *molBV*-BV (**Fig. 2D**). Six highly correlated ratios share the IL-1 β cytokine and 2 other highly correlated ratios share TNF- α (the 1st and 2nd highest ORs for a given cytokine, respectively, in the univariate analysis shown in **Fig. 2A**). Of the BV associated ratios, IL-1 β /IP-10 had the strongest overall effect based on absolute odds ratio. Interestingly, IL-1 β and IP-10 have been previously shown to be strongly associated with BV⁴⁶, making this ratio very attractive for further consideration.

Exploring the distribution of the IL-1 β /IP-10 ratio across the *molBV*-BV states revealed that despite the clear and consistent association (linear trend p -value = 3×10^{-48}) with BV (**Fig. 2E**), 24% (85/349) of women with *molBV*-BV did not have elevated IL-1 β /IP-10 levels. Whereas, surprisingly 28% (88/309) of BV negative women, had elevated IL-1 β /IP-10 levels (see left violin plot, **Fig. 2E**). We used ANCOM to explore potential causes of *molBV*-negative women having high IL-1 β /IP-10 levels. This analysis identified elevated levels of *L. iners* and *G. vaginalis*, whereas those with elevated levels of *L. crispatus* had low levels of IL-1 β /IP-10 (W-stat threshold > 10 , FDR < 0.05 , **Fig. 2F**).

To determine whether women with BV who did not show an elevated IL-1 β /IP-10 signature had an alternative form of inflammation, we compared this group to BV-negative women with similarly low levels of IL-1 β /IP-10 (i.e., below the cohort median, **Fig. 2G**). The analysis revealed that TNF- α /MIP-1 β was significantly associated with *molBV*-BV (OR = 1.24, 95% CI: 1.19-1.29) in the absence of elevated IL-1 β /IP-10. In addition, there were *molBV*-BV-negative women with elevated IL-1 β /IP-10 inflammatory signatures that, in fact, had underlying pathogenic bacteria (e.g., *L. iners*) not triggering a BV diagnosis with the *molBV* pipeline.

Molecular BV, cervical inflammation and the natural history of HPV

We previously reported that increased diversity of the cervicovaginal microbiome contributed to HPV natural history¹⁵. To evaluate the impact of bacterial vaginosis and directly test whether cervicovaginal inflammation could be mediating the effect of BV on the natural history of HPV, we utilized the previously reported¹⁵ prospective cohort sub-study from the CVT trial. We utilized 16S rRNA NGS data from cervicovaginal DNA¹⁵ to calculate the *molBV* scores across two study visits. Women who had sustained low levels of *molBV* (*molBV* = 0-3 for both visits) vs. those that had sustained high *molBV* scores (*molBV* = 7-10 for both visits) were more likely to clear HR-HPV over time (**Fig. 3A**, $p=0.02$). Similarly, sustained inflammation vs. not, as determined by IL-1 β /IP-10 was associated with HR-HPV clearance (**Fig. 3B**).

To determine whether *molBV*-BV or sustained inflammatory markers acted as independent risk factors for HR-HPV clearance, we used a cox-proportional hazard model after covariate adjustment. **Table 1** presents the effect of sustained elevated IL-1 β /IP-10 levels using cox-proportional hazard models adjusted for age, smoking status and HPV16. Model 1 considered the effect of both IL-1 β /IP-10 and dichotomous *molBV* states. In the adjusted analysis, *molBV* levels were not associated with HR-HPV clearance when IL-1 β /IP-10 was also in the model suggesting that it was an independent driver of HPV clearance (HZ = 1.90, 95% CI: 1.08-3.40). Given the strong correlation between *molBV* and IL-1 β /IP-10 levels (**Fig. 2E**), we made an additional parsimonious model (Model 2) to more accurately measure the effect of sustained elevated IL-1 β /IP-10 levels on HR-HPV clearance (HZ = 1.86, 95% CI: 1.19-2.90). This model did not change the hazard ratio of sustained low IL-1 β /IP-10 with HR-HPV clearance, but it did significantly reduce the p-value supporting the analyses that it is the true driver of HR-HPV clearance.

We next evaluated the association of BV and HR-HPV progression to CIN2+¹⁵. Briefly, the original study considered the binary outcome of persistent HR-HPV progressing to CIN2+ (diagnosed ~2 years after sample visit¹⁵) vs. HR-HPV infection clearance with CVM components serving as predictors. In the current analysis, we first tested whether molecular BV and/or the primary BV inflammatory signature (i.e. IL-1 β /IP-10) was associated with progression to CIN2+ using a generalized linear model (**Table 2, Model 1**). The results indicated that a continuous *molBV* score was prospectively associated with CIN2+ progression with the odds of progression increasing by 1.24 (95% CI: 1.02-1.55) per unit increase of *molBV*. However, the dominant BV IL-1 β /IP-10 signature was not significant when *molBV* was also in the model (**Table 2, Model 2**, OR = 1.15, 95% CI: 0.92-1.46). Surprisingly, the alternative inflammatory signature represented by TNF- α /MIP-1 β levels was predictive of progression when either *molBV* (**Table 2, Model 3**, OR = 2.71, 95% CI: 1.46-5.61) or *molBV* and IL-1 β /IP-10 were in the model (**Table 2, Model 4**, OR = 2.65, 95% CI: 1.39-5.68). Given the association of TNF- α /MIP-1 β with *molBV* (**Fig. 2C**) and the modest correlation with IL-1 β /IP-10 (**Fig. 2D**), an additional parsimonious model was constructed to more accurately measure the effect of TNF- α /MIP-1 β with progression to CIN2+ (**Table 2, Model 5**, OR = 2.81, 95% CI: 1.65-5.42).

Discussion

In this study, we develop a relatively simple means to characterize cervicovaginal samples for BV using 16S rRNA gene amplicon next-generation sequencing (NGS). It is well known that features of the CVM are strongly associated with BV^{27,37,38,47-50}, and our study takes this relationship further and provides a quantitative score 0 – 10, equivalent to the Nugent score. This method is particularly useful for the high-throughput analysis to

determine a “molecular” Nugent-like score (*molBV*-BV) in women that might not have been evaluated for BV. Development of this method used a stringent diagnosis of BV including subjects concurrent for BV both by Nugent and Amsel criteria. The method was validated in 3 cohorts suggesting the generalizability of this particular molecular approach to generate a Nugent-like score. Although it is known that there is substantial variation between the cervicovaginal microbiome between African women and women with European ancestry⁵¹, use of the *molBV* algorithm in two African populations revealed high diagnostic AUCs of 0.97 and 0.88 for Soweto and Cape Town sets, respectively. Moreover, the *molBV* diagnostic was also used to evaluate the local host inflammatory response in a population without available BV measures (i.e., Nugent or Amsel).

Identification of *molBV*-BV in the CVT cohort was strongly associated with 13 different cytokines, 6 previously reported to be associated with BV (Supplemental Table 2 and ³¹). The cytokine ratios were used in order to better address the interrelated data structure and observed that the IL-1 β /IP-10 ratio was strongly correlated with molecular BV states (linear trend p-value=3.67x10⁻⁴⁸). IL-1 β /IP-10 ratio was previously postulated to be a relevant signature for BV and the identification of women at higher risk of STI transmission⁴⁶.

Despite a strong correlation between IL-1 β /IP-10 and *molBV* states, there were still 24-28% of women that had an elevated ratio despite a *molBV* score of 0-3. Upon further analyses, elevated *L. iners* and *G. vaginalis* were identified in this enigmatic group. This finding is of interest since dominance of *Lactobacillus* species in the CVM is typically associated with vaginal health ^{38,48}. It is of interest that women in the discovery set within a *Lactobacillus* dominated clade and having clinical features of BV also had elevated levels of *L. iners* (**Fig. 1A**). Furthermore, in the CVT cohort *L. crispatus* was protective against IL-1 β /IP-10 inflammation consistent with the discovery panel (**Fig. 1A**) and previous reports including *L. crispatus* ³⁸. Our data indicate that *L. iners* may be able to induce a BV-like inflammatory state even when a state of polymicrobialism, typically associated with BV⁵², is absent. It is possible that the variable association of *L. iners* with an inflammatory state is due to strain level variation, since the biology of *L. iners* is enigmatic ⁵³⁻⁵⁵. Deeper sequencing of the CVM will be required to validate this hypothesis. Alternatively, community context of *L. iners* might influence its behavior and association with inflammation. It appears that specific bacteria could have context specific functionality depending on the observed molecular BV state at time of sampling. Although the *molBV* algorithm was set to clinical parameters of BV, the current work suggests that additional stratification to identify women at risk for sub-clinical inflammation is necessary. This is especially important given the presented data showing the significance of these inflammatory shifts in the context of HR-HPV infections and possible implications for other diseases in which BV acts as a risk factor.

Another interesting aspect of the cytokine analysis of this study is the heterogeneity of immune markers that were elevated with the IL-1 β /IP-10 across the strata of molecular BV states. It would be reasonable to predict that having an elevation of this ratio would yield similar distribution of other cytokines given the correlation levels observed in **Fig.s 2B** and **2D**. However, when comparing the molecular BV-positive to BV-negative women with levels of IL-1 β /IP-10 lower than the cohort median, we observed that the BV-positive women had elevated levels of TNF- α /MIP-1 β inflammatory cytokines. Furthermore the secondary inflammatory marker identified in **Fig. 2C** and **2D** was in fact the dominant signature elevated in molecular BV-positive women with low levels of IL-1 β /IP-10. These results indicate that BV seems to be heterogeneous as to the exact type of sub-clinical host inflammatory response. Given the identifiable variance of inflammatory levels within

individual strata of BV due to specific organisms such as *L. iners*, it is likely specific taxa, or bacterial networks, might be associated with this observation.

There is increased interest in uncovering the relationship between the cervicovaginal microbiome and HPV natural history since a number of studies show correlation between vaginal microbial diversity, BV and HPV clearance¹⁵⁻¹⁸. It is not currently known why certain high-risk infections clear while a small minority persist for years and eventually progress to pre-cancer⁵⁶. We utilized HR-HPV detection data within the Costa Rica vaccine trial cohort⁴⁴, to evaluate cervical cytokine profiles, 16S rRNA data and the newly developed molecular BV states to interrogate possible mechanisms of HR-HPV infection clearance. Crude Kaplan-Meier analysis revealed similar clearance patterns between sustained IL-1 β /IP-10 levels and *molBV* scores, with the cytokine measure showing slightly better sample separation although these measures are highly correlated. To better understand these relationships, we utilized Cox-proportional hazard modeling with adjustment for age, smoking status and HPV16 infection. The association with *molBV*-BV was eliminated once we adjusted for an elevated IL-1 β /IP-10 ratio.

Another relevant HPV outcome is progression of a persistent HR-HPV infection to pre-cancer (CIN2+) ⁵⁷. The CVM was previously reported to be predictive of this outcome^{15,34,58}. However, when we tested the continuous IL-1 β /IP-10 levels in a model, this cytokine signature did not add any additional information beyond *molBV* for HR-HPV progression. Interestingly, when we tested the minor TNF- α /MIP-1 β signature associated with *molBV* we found that it was associated with CIN2+ progression even after adjustment for *molBV* as well as *molBV* + IL-1 β /IP-10 levels (**Table 2**). The final analysis revealed that a single unit increase in the TNF- α /MIP-1 β ratio was associated with an odds of 2.81 (95% CI: 1.65-5.42) of developing CIN2+ within 2-years of sample collection. These results may explain why certain studies do not see an association with CVM diversity (a surrogate for molecular BV, **Fig. 1B**), but do see signals when analyzing specific taxa⁵⁹. These data suggest that BV is a highly heterogeneous condition and although consistently associated with certain microbial shifts and overall community structure (e.g., higher alpha diversity), it is also modified by the activity of specific taxa. Based on the observations reported in this study, these variations appear to have a substantial effect on the immune response, which in turn has an effect on HR-HPV clearance and progression to CIN2+.

The currently reported analyses have several weaknesses that should be taken into account to interpret the data. All of the analysis of the CVM utilized 16S rRNA amplicon sequencing. This method limits the taxonomic resolution of bacteria and other organisms constituting the microbiome. Deeper exploration of BV using techniques such as shotgun metagenomics may provide more thorough explanation as to why there is significant heterogeneity in the inflammatory response to BV amongst different women. Additionally the developed *molBV* score is dependent on a relatively small panel of bacteria. This decision was made in order to ensure robustness of the measure across cohorts. This appears to have been effective based on the associations seen within African populations, but may present a limitation in populations with different structures of the CVM. Most of the women in the reported study are adolescent or young adults and we do not know if the analyses extend to women of all ages and geographic locations. Moreover, it is possible that other inflammatory signatures might exist in the cervicovaginal region that were not measured in the current report.

Here we present a comprehensive molecular characterization of BV using 16S rRNA amplicon sequencing and a curated panel of cytokines. We demonstrate using multiple cohorts that 16S rRNA amplicon sequencing can

be reliably used to diagnose BV using the newly developed *molBV* score. We further demonstrate that this score was strongly correlated with a heterogeneous inflammatory landscape within the cervicovaginal region. Exploring these inflammatory markers further revealed a complex system of interactions between individual taxa, specific cytokines and molecular BV states. Furthermore, we demonstrated that there is potential clinical relevance of the findings through the use of HR-HPV outcomes. We specifically show that the different possible inflammatory states in BV are either associated with persistence of HR-HPV infection (i.e. IL-1 β /IP-10) or the progression of infection to precancer (i.e., TNF- α /MIP-1 β). In support of a role for TNF- α in progression of HR-HPV infections, a recent report indicated that TNF- α was the main discriminatory biomarker associated with invasive cervical cancer ⁶⁰.

Whether the adverse health outcomes from BV are all based on the host inflammatory response remains to be seen. Deeper exploration of these associations is warranted using more robust techniques such as machine learning in order to further understand why certain women experience inflammation in BV, while others do not and its impact on health. Lastly, implication of BV inducing a local inflammatory response might imply signaling systemic inflammation, which was not assessed in this study. The role BV-induced inflammation might have in immune conditions more prevalent in women remains an interesting hypothesis that could have profound diagnostic and therapeutic ramifications.

Methods

Bacterial Vaginosis training set

This study was conducted within an ongoing HPV study at Mount Sinai Adolescent Health Center (MSAHC) in New York City ^{61,62}. Cervicovaginal samples were collected from female patients, 15 to 25 years of age, with vaginal symptoms suggestive of BV (n=30) or no symptoms (n=30), both groups were recruited sequentially from the same clinic. Pregnant women were excluded. The parent study and BV sub-study were approved by the Institutional Review Board at The Icahn School of Medicine at Mount Sinai.

Diagnosis of bacterial vaginosis

Subjects were evaluated for Amsel criteria by the examining physician and the presence of 3 out of 4 Amsel criteria established a diagnosis of BV ⁶³. Vaginal swabs were collected for Nugent scores by carefully inserting a sterile swab into the vagina about two inches, gently rotating against the vaginal wall for 10 to 30 seconds, and then withdrawn without touching the skin to avoid contamination. De-identified swabs were placed in a plastic culturette tube and shipped overnight to an outside clinical laboratory for Nugent scoring following standardized criteria ²². A composite score from 0-10 was generated with a diagnosis of Nugent BV assigned as follows: 0-3 was considered negative, 4-6 was considered intermediate, and 7-10 was considered diagnostic of BV.

Microbiome sample collection and DNA extraction

Samples for microbiome analyses were collected using a Cytobrush® placed in PreservCyt transport medium (ThinPrep®; Hologic, Marlborough, MA). Samples were stored immediately at -20°C until transport to the research lab at the Albert Einstein College of Medicine. In the lab, the samples were transferred to a 15 ml tube

and gently centrifuged at 1500 RPM for 5 minutes. After removing the supernatant by decanting, the pellets were rinsed in 3 ml of TE (10 mM Tris, 1.0 mM EDTA). This solution was then vortexed and centrifuged at 1500 RPM for 5 minutes and the supernatant was removed by decanting. The remaining pellet and leftover solution (~150 µl) was used for DNA isolation via column processing with the QIAamp Mini spin column (Qiagen, Valencia, CA) following the manufacturer's protocol. The purified DNA was eluted in 150 µl of elution buffer (10mM Tris/0.5mM EDTA, pH 9).

PCR Amplification

PCR for bacterial communities was performed using forward (515F) GTGYCAGCMGCCGCGGTA and reverse (806R) GGACTACHVGGGTWTCTAAT primers which amplify the V4 hypervariable region of the prokaryotic 16S rRNA gene^{64,65}. All primers contained unique Golay barcodes to allow for dual indexing of each sample. PCRs were conducted in a 25 µl reaction with 2 µl input of template DNA, 16.75 µl of ddH₂O, 2.5 µl of Platinum 10X PCR buffer (Invitrogen, Waltham, MA), 0.75 µl of MgCl₂ (50 mM, Invitrogen), 0.5 µl of dNTP mix (10mM each, Roche, Basel, Switzerland), 0.25 µl AmpliTaq Gold, polymerase (5 U/µl, Applied Biosystems, Carlsbad, CA), 0.25 µl of Platinum Taq DNA Polymerase (10 U/µl, Invitrogen), and 1 µl (5 µM) of each primer (IDT, Coralville, IA). Thermocycling conditions included an initial denaturation of 95°C for 5 mins, followed by 15 cycles of 95°C for 1 m, 55°C for 1 m, 72°C for 1 m, followed by 15 cycles of 95°C for 1 m, 60°C for 1 m, 72°C for 1 m, and a final extension of 72°C for 10 min.

PCR for fungal communities was performed using forward (48F) ACACACCGCCCGTCGCTACT and reverse (217R) TTTCTGCTGCGTTCTTCATCG primers that amplify the ITS1 region of the prokaryotic ribosomal gene cluster^{36,66}. PCRs were conducted in a 25 µl reaction with 10 µl input of template DNA, 8.75 µl of ddH₂O, 2.5 µl of Platinum 10x PCR buffer (Invitrogen), 0.75 µl of MgCl₂ (50 mM, Invitrogen), 0.5 µl of dNTP mix (10 mM each, Roche), 0.25 µl AmpliTaq Gold polymerase (5 U/µl, Applied Biosystems), 0.25 µl of Platinum Taq DNA Polymerase (10 U/µl, Invitrogen), and 1 µl (5 µM) of each primer (IDT, Coralville, IA). Thermocycling conditions included an initial denaturation of 95°C for 5 mins, followed by 35 cycles of 95°C for 30s, 55°C for 30s, 72°C for 2 min, followed by a final extension of 72°C for 10 min. All PCRs were conducted in a GeneAmp PCR System 9700 (Applied Biosystems) and PCR products were verified by gel electrophoresis.

Next Generation Library Preparation and Sequencing

PCR products for each sample were pooled by PCR assay (16S and ITS1) in approximately equal concentrations and 100 µl of the pooled products were loaded into a 3% agarose gel and run at 80V for 3 hours to separate the DNA fragments. The DNA fragment for each assay was excised and purified with a QIAquick Gel Extraction Kit (Qiagen) and quantified using a Qubit High Sensitivity dsDNA assay (Invitrogen). NGS library preparation was conducted on the purified pooled PCR samples from each assay with a KAPA LTP Library Preparation Kit (KAPA Biosystems, Wilmington, MA) according to the manufacturer's protocol. The library amplicons were validated on a 2100 Bioanalyzer (Agilent Technologies, Santa Clara CA) and sequencing of libraries was carried out on an Illumina MiSeq with 2x300 bp paired-end reads kit at the Genomics Core of the Albert Einstein College of Medicine.

Bioinformatics

Illumina reads were initially right trimmed to remove bases that fell below PHREAD score 25 using PRINSEQ-lite⁶⁷. Reads were then demultiplexed using NovoBarcode based on unique dual Golay barcode combinations⁶⁸.

QIIME2⁶⁹ was used to identify amplicon sequence variants using DADA2⁷⁰ for both the 16SV4 rRNA and ITS1 amplicon data. For 16SV4 rRNA amplicon sequence variants (ASVs) the naïve Bayesian classifier⁷¹ was used to assign taxonomy using the lab's custom database that is comprised of GreenGenes 13.8⁷², HOMD⁷³ and vaginal reference sequences⁷⁴. For fungal taxonomic assignments, BLAST⁷⁵ was used with the UNITE database⁷⁶. Taxonomic assignment was combined with the ASV data using custom bash scripts into a biome file⁷⁷ and further processed with R⁷⁸.

Statistical Analysis

The *phyloseq*⁷⁹ package was used to import microbiome data into R⁷⁸ and to calculate the Chao1, Fisher and Shannon alpha diversity metrics as well as the Jensen-Shannon diversity index for beta diversity analysis. The *vegan*⁸⁰ package was used to run the PERMANOVA. The *pROC* package was used for the AUC analysis⁸¹. All data visualization was achieved using the *ggplot2* package⁸².

Significance of belonging to taxon-specific hierarchical clusters in the heatmap analysis was assessed using Fisher's exact test. Pairwise statistical significance in alpha diversity was determined using the Wilcoxon test. Significance in beta diversity was determined using PERMANOVA. ANCOM⁸³ was used for bacterial taxa (i.e., biomarker) discovery. A linear model was used to determine significance of trends in the cytokine analyses and extract ordinal ORs. Pearson coefficient was used for correlation analysis. The *q-value*⁸⁴ package was used to correct the calculated linear trend p-values for multiple testing. Standard error of the mean was used to represent variation in sequencing depth between samples. Cox proportional hazard model was used in order to adjust the data for age, smoking and HPV16 status in the survival analyses. Age was treated as a continuous variable, HPV16 as binary (0 or 1) based on PCR result (V1) and smoking status as ordinal (0 = never smoker, 1 = former smoker and 2 = current smoker) at V1.

Calculating molBV

ANCOM⁸³ was used to determine which bacterial genera were associated with BV through the use of microbial reference frames⁸⁵ (**Fig. 1E**). Only those samples that were positive, or negative, for BV by both Amsel and Nugent criteria were used in the analysis (n = 18 BV+, n = 22 BV-). Out of the identified biomarkers only those that were present in $\geq 80\%$ of samples with at least a 0.001% relative abundance after subsampling were retained for the calculation of *molBV*. These taxa included- *Lactobacillus*, *Prevotella*, *Gardnerella*, *Megasphaera*, *Parvimonas*, *Clostridium*, *Porphyromonas*, *Adlercreutzia*, *Dialister*, *Atopobium* and *Sneathia*.

To create the microbial reference frames, log ratios were created between *Lactobacillus* and markers elevated in the BV positive group. The log ratios were then analyzed using a robust regression with the Nugent scores serving as the outcome and each ratio as the predictor. The beta coefficients and intercepts for each of the ratios were extracted and are presented in **Supplemental Table 3**.

To calculate *molBV*, which is the imputed continuous Nugent score, the log ratios between *Lactobacillus* and BV specific markers were generated (e.g., *Lactobacillus:Prevotella*, *Lactobacillus:Gardnerella*, *Lactobacillus:Shuttleworthia*, etc.) and used, along with the data from **Supplemental Table 3**, to calculate *molBV*.

In the formula X_i represents the log ratio i , B_{0i} is the ratio's corresponding intercept and B_{1i} is the beta coefficient. For an estimate to be valid for a given reference taxa, there had to be both *Lactobacillus* and a BV marker detected with a minimum of 1 read each. The final molecular BV score is the average of the valid log ratio estimates that approximates the clinical Nugent score. Given the nature of regression prediction *molBV* is not bound by the 0-10 range of the Nugent score and may take on non-integer values. To make the two scales more comparable *molBV* is fit into the 0-10 range with scores below the range becoming 0 and those above becoming 10.

Similar to Nugent scoring, ranges of the continuous *molBV* score, are used to define BV status. A *molBV* score of 0-3 is considered negative for BV, 4-6 is considered intermediate, and a score of 7-10 is consistent with BV.

Confirmation Cohorts

Three cohorts were used to confirm the *molBV* classifier. Full details about sample collection and processing can be obtained from the cited studies. The United States (USA) confirmation cohort was composed of 388 women with collection from three separate locations (two in Baltimore and one in Atlanta) with the women having a median age of 31 years³⁸. The Cape Town cohort was composed of 90 women with a median age of 18 years^{40,42}. The Soweto cohort was composed of 78 women with a median age of 18 years^{40,42}.

Cervical immune cytokines, chemokines, and soluble receptors

Cervical sponge samples were collected from women participating in the HPV Costa Rica Vaccine Trial (CVT) using a Merocel sponge (Medtronic Xomed, Jacksonville, FL) as previously described⁸⁶. A customized panel including 32 cytokines, chemokines, and soluble receptors were quantitated using Luminex-based Milliplex Map Multiplex Assays (Millipore, Billerica, MA) as previously described⁸⁷.

HPV Natural History Exposure/Outcome Definitions

Residual DNA from cervical samples from the placebo arm of the Costa Rica vaccine trial⁴⁴ were used to test the prospective association of molecular BV and cervicovaginal inflammation with HPV natural history stages. The analyzed outcome was time to clearance of high-risk HPVs (i.e., HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58 and/or 59⁸⁸) detected at the sub-study visit V0. Women were selected from the CVT placebo arm where an incident high-risk HPV infection was detected. Clearance was recorded if a woman cleared all high-risk types detected at the incident visit within a 2-year observation window. The two core exposures were sustained high/low *molBV* and sustained high/low cervical inflammation as determined from cytokine marker ratios. Inflammation and *molBV* sustained/persistent status was determined by median stratification; specifically the sustained status categories refer to a per-protocol approach where the exposure status had to be similar across the two analyzed visits. For example the median *molBV* value across all study visits was 5.4 and if a women had a *molBV* value of 10 for visit 1 and 8 for visit 2, she would be placed in the sustained

molBV high category. Participants with discordant *molBV* values were not included in this analysis in order to measure the per-protocol effect of sustained inflammation in the context of HPV natural history.

Declarations

Acknowledgements

We would like to thank Jo-Ann Passmore, Smritee Dabee, Katie Viljoen, Heather Jaspan, Christina Balle and other members of the WISH cohort for providing the clinical Nugent scores for our Cape Town and Soweto confirmation sets.

Ethics approval and consent to participate

The study was approved by the Institutional Review Board at Icahn School of Medicine at Mount Sinai, Manhattan, New York, and written informed consent was obtained from all study participants and from guardians accompanying minors before enrollment.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no conflict of interest.

Authors' contributions

MU, NFS, and RDB conceptualized and designed the study. NSF, RDB, and SP were involved with patient recruitment and clinical study design. MS, MS and LP were involved in the CVT clinical sub-study design. CCS and AG processed and prepared samples for sequencing. MU, SV and NFS performed statistical analyses. MU and RB wrote the manuscript with the help of all co-authors. All authors read and edited the final manuscript.

Availability of supporting data

Sequence files and metadata for all samples used in this study have been uploaded to SRA (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA641099>).

CVT Cohort Declaration and Acknowledgements

Investigators in the International Agency for Research on Cancer/World Health Organization

Where authors are identified as personnel of the International Agency for Research on Cancer/ World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer / World Health Organization.

Investigators in the Costa Rica HPV Vaccine Trial (CVT) Group

Proyecto Epidemiológico Guanacaste, Fundación INCIENSA, San José, Costa Rica—Bernal Cortés (specimen and repository manager), Paula González (LTFU: co-principal investigator) (deceased), Rolando Herrero (CVT: co-principal investigator), Silvia E. Jiménez (trial coordinator), Carolina Porras (co-investigator), Ana Cecilia Rodríguez (co-investigator).

United States National Cancer Institute, Bethesda, MD, USA—Allan Hildesheim (co-principal investigator & NCI co-project officer), Aimée R. Kreimer (LTFU: co-principal investigator & NCI co-project officer), Douglas R. Lowy (HPV virologist), Mark Schiffman (CVT: medical monitor & NCI co-project officer), John T. Schiller (HPV virologist), Mark Sherman (CVT: QC pathologist), Sholom Wacholder (statistician)(deceased).

Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, Frederick, MD, USA (HPV Immunology Laboratory)—Ligia A. Pinto, Troy J. Kemp

Georgetown University, Washington, DC, USA—Mary K. Sidawy (CVT: histopathologist)

DDL Diagnostic Laboratory, Netherlands (HPV DNA Testing)—Wim Quint, Leen-Jan van Doorn, Linda Struijk.

University of California, San Francisco, CA, USA—Joel M. Palefsky (expert on anal HPV infection and disease diagnosis and management), Teresa M. Darragh (pathologist and clinical management)

University of Virginia, Charlottesville, VA, USA—Mark H. Stoler (QC pathologist)

Acknowledgments

We extend a special thanks to the women of Guanacaste and Puntarenas, Costa Rica, who gave of themselves in participating in this effort. In Costa Rica, we acknowledge the tremendous effort and dedication of the staff involved in this project; we would like to specifically acknowledge the meaningful contributions by Carlos Avila, Loretto Carvajal, Rebeca Ocampo, Cristian Montero, Diego Guillen, Jorge Morales and Mario Alfaro. In the United States, we extend our appreciation to the team from Information Management Services (IMS) responsible for the development and maintenance of the data system used in the trial and who serve as the data management center for this effort, especially Jean Cyr, Julie Buckland, John Schussler, and Brian Befano. We thank Dr. Diane Solomon (CVT: medical monitor & QC pathologist) for her invaluable contributions during the randomized blinded phase of the trial and the design of the LTFU and Nora Macklin (CVT) and Kate Torres (LTFU) for the expertise in coordinating the study. We thank the members of the Data and Safety Monitoring Board charged with protecting the safety and interest of participants during the randomized, blinded phase of our study (Steve Self, Chair, Adriana Benavides, Luis Diego Calzada, Ruth Karron, Ritu Nayar, and Nancy Roach) and members of the external Scientific HPV Working Group who have contributed to the success of our efforts over the years (Joanna Cain and Elizabeth Fontham, Co-Chairs, Diane Davey, Anne Gershon, Elizabeth Holly, Silvia Lara, Henriette Raventós, Wasima Rida, Richard Roden, Maria del Rocío Sáenz Madrigal, Gypsyamber D'Souza, and Margaret Stanley).

References

- 1 Ravel, J., Moreno, I. & Simón, C. Bacterial vaginosis and its association with infertility, endometritis, and pelvic inflammatory disease. *American Journal of Obstetrics and Gynecology* (2020).

- 2 Redelinghuys, M. J., Geldenhuys, J., Jung, H. & Kock, M. M. Bacterial vaginosis: current diagnostic avenues and future opportunities. *Frontiers in Cellular and Infection Microbiology* **10** (2020).
- 3 Koumans, E. H. *et al.* The prevalence of bacterial vaginosis in the United States, 2001–2004; associations with symptoms, sexual behaviors, and reproductive health. *Sexually transmitted diseases* **34**, 864-869 (2007).
- 4 Allsworth, J. E. & Peipert, J. F. Prevalence of bacterial vaginosis: 2001–2004 national health and nutrition examination survey data. *Obstetrics & Gynecology* **109**, 114-120 (2007).
- 5 Peebles, K., Velloza, J., Balkus, J. E., McClelland, R. S. & Barnabas, R. V. High Global Burden and Costs of Bacterial Vaginosis: A Systematic Review and Meta-Analysis. *Sex. Transm. Dis.* **46**, 304-311 (2019).
- 6 Bradshaw, C. S. & Sobel, J. D. Current treatment of bacterial vaginosis—limitations and need for innovation. *The Journal of infectious diseases* **214**, S14-S20 (2016).
- 7 Salah, R. M., Allam, A. M., Magdy, A. M. & Mohamed, A. S. Bacterial vaginosis and infertility: cause or association? *European journal of obstetrics & gynecology and reproductive biology* **167**, 59-63 (2013).
- 8 Brabant, G. Bacterial vaginosis and spontaneous preterm birth. *J. Gynecol. Obstet. Biol. Reprod. (Paris)*. **45**, 1247-1260 (2016).
- 9 Tellapragada, C. *et al.* Risk factors for preterm birth and low birth weight among pregnant Indian women: a hospital-based prospective study. *Journal of Preventive Medicine and Public Health* **49**, 165 (2016).
- 10 Bautista, C. T. *et al.* Bacterial vaginosis: a synthesis of the literature on etiology, prevalence, risk factors, and relationship with chlamydia and gonorrhea infections. *Military Medical Research* **3**, 4 (2016).
- 11 Rumyantseva, T., Khayrullina, G., Guschin, A. & Donders, G. Prevalence of Ureaplasma spp. and Mycoplasma hominis in healthy women and patients with flora alterations. *Diagn. Microbiol. Infect. Dis.* **93**, 227-231 (2019).
- 12 Eastment, M. C. & McClelland, R. S. Vaginal microbiota and susceptibility to HIV. *AIDS (London, England)* **32**, 687-698 (2018).
- 13 McKinnon, L. R. *et al.* The evolving facets of bacterial vaginosis: implications for HIV transmission. *AIDS research and human retroviruses* **35**, 219-228 (2019).
- 14 King, C. C. *et al.* Bacterial vaginosis and the natural history of human papillomavirus. *Infectious diseases in obstetrics and gynecology* **2011** (2011).
- 15 Usyk, M. *et al.* Cervicovaginal microbiome and natural history of HPV in a longitudinal study. *PLoS pathogens* **16**, e1008376 (2020).
- 16 Quan, L. *et al.* Simultaneous detection and comprehensive analysis of HPV and microbiome status of a cervical liquid-based cytology sample using Nanopore MinION sequencing. *Scientific reports* **9**, 1-13 (2019).

- 17 Champer, M. *et al.* The role of the vaginal microbiome in gynaecological cancer. *BJOG: An International Journal of Obstetrics & Gynaecology* **125**, 309-315 (2018).
- 18 Moscicki, A.-B., Shi, B., Huang, H., Barnard, E. & Li, H. Cervical-Vaginal Microbiome and Associated Cytokine Profiles in a Prospective Study of HPV 16 Acquisition, Persistence, and Clearance. *Frontiers in cellular and infection microbiology* **10** (2020).
- 19 Brookheart, R. T., Lewis, W. G., Peipert, J. F., Lewis, A. L. & Allsworth, J. E. Association between obesity and bacterial vaginosis as assessed by Nugent score. *Am. J. Obstet. Gynecol.* (2019).
- 20 Amsel, R. *et al.* Nonspecific vaginitis: diagnostic criteria and microbial and epidemiologic associations. *The American journal of medicine* **74**, 14-22 (1983).
- 21 Schwiertz, A., Taras, D., Rusch, K. & Rusch, V. Throwing the dice for the diagnosis of vaginal complaints? *Annals of clinical microbiology and antimicrobials* **5**, 4 (2006).
- 22 Nugent, R. P., Krohn, M. A. & Hillier, S. L. Reliability of diagnosing bacterial vaginosis is improved by a standardized method of gram stain interpretation. *Journal of clinical microbiology* **29**, 297-301 (1991).
- 23 Chaijareenont, K., Sirimai, K., Boriboonhirunsarn, D. & Kiriwat, O. Accuracy of Nugent's score and each Amsel's criteria in the diagnosis of bacterial vaginosis. *J. Med. Assoc. Thai.* **87**, 1270-1274 (2004).
- 24 Hilbert, D. W. *et al.* Development and validation of a highly accurate quantitative real-time PCR assay for diagnosis of bacterial vaginosis. *J. Clin. Microbiol.* **54**, 1017-1024 (2016).
- 25 Mohanty, S., Sood, S., Kapil, A. & Mittal, S. Interobserver variation in the interpretation of Nugent scoring method for diagnosis of bacterial vaginosis. *Indian J. Med. Res.* **131** (2010).
- 26 Coleman, J. S. & Gaydos, C. A. Molecular diagnosis of bacterial vaginosis: an update. *Journal of clinical microbiology* **56**, e00342-00318 (2018).
- 27 Coleman, J. S. & Gaydos, C. A. Molecular diagnosis of bacterial vaginosis: an update. *Journal of clinical microbiology* **56** (2018).
- 28 Mitchell, C. & Marrazzo, J. Bacterial vaginosis and the cervicovaginal immune response. *American Journal of Reproductive Immunology* **71**, 555-563 (2014).
- 29 Alcaide, M. L. *et al.* A bio-behavioral intervention to decrease intravaginal practices and bacterial vaginosis among HIV infected Zambian women, a randomized pilot study. *BMC infectious diseases* **17**, 1-10 (2017).
- 30 Turovskiy, Y., Sutyak Noll, K. & Chikindas, M. L. The aetiology of bacterial vaginosis. *Journal of applied microbiology* **110**, 1105-1128 (2011).
- 31 Dabee, S., Passmore, J.-A. S., Heffron, R. & Jaspan, H. B. The complex link between the female genital microbiota, genital infections and inflammation. *Infection and Immunity* (2021).

- 32 Denney, J. M. & Culhane, J. F. in *Seminars in Fetal and Neonatal Medicine*. 200-203 (Elsevier).
- 33 Romero, R., Chaiworapongsa, T., Kuivaniemi, H. & Tromp, G. Bacterial vaginosis, the inflammatory response and the risk of preterm birth: a role for genetic epidemiology in the prevention of preterm birth. *American Journal of Obstetrics & Gynecology* **190**, 1509-1519 (2004).
- 34 Gillet, E. *et al.* Bacterial vaginosis is associated with uterine cervical human papillomavirus infection: a meta-analysis. *BMC infectious diseases* **11**, 1-9 (2011).
- 35 Van Der Pol, W. J. *et al.* In silico and experimental evaluation of primer sets for species-level resolution of the vaginal microbiota using 16s ribosomal rna gene sequencing. *The Journal of infectious diseases* **219**, 305-314 (2018).
- 36 Usyk, M., Zolnik, C. P., Patel, H., Levi, M. H. & Burk, R. D. Novel ITS1 fungal primers for characterization of the mycobiome. *MSphere* **2**, e00488-00417 (2017).
- 37 Fredricks, D. N., Fiedler, T. L. & Marrazzo, J. M. Molecular identification of bacteria associated with bacterial vaginosis. *New England Journal of Medicine* **353**, 1899-1911 (2005).
- 38 Ravel, J. *et al.* Vaginal microbiome of reproductive-age women. *Proceedings of the National Academy of Sciences* **108**, 4680-4687 (2011).
- 39 Kacerovsky, M. *et al.* Cervical Gardnerella vaginalis in women with preterm prelabor rupture of membranes. *Plos one* **16**, e0245937 (2021).
- 40 Lennard, K. *et al.* Microbial composition predicts genital tract inflammation and persistent bacterial vaginosis in South African adolescent females. *Infection and immunity* **86** (2018).
- 41 Burton, J. P. & Reid, G. Evaluation of the bacterial vaginal flora of 20 postmenopausal women by direct (Nugent score) and molecular (polymerase chain reaction and denaturing gradient gel electrophoresis) techniques. *The Journal of infectious diseases* **186**, 1770-1780 (2002).
- 42 Dabee, S. *et al.* Defining characteristics of genital health in South African adolescent girls and young women at high risk for HIV infection. *PloS one* **14**, e0213975 (2019).
- 43 Dabee, S., Passmore, J.-A. S., Heffron, R. & Jaspan, H. B. The Complex Link between the Female Genital Microbiota, Genital Infections, and Inflammation. *Infection and Immunity* **89** (2021).
- 44 Herrero, R. *et al.* Rationale and design of a community-based double-blind randomized clinical trial of an HPV 16 and 18 vaccine in Guanacaste, Costa Rica. *Vaccine* **26**, 4795-4808 (2008).
- 45 Morton, J. T. *et al.* Establishing microbial composition measurement standards with reference frames. *Nature communications* **10**, 1-11 (2019).
- 46 Masson, L. *et al.* Inflammatory cytokine biomarkers to identify women with asymptomatic sexually transmitted infections and bacterial vaginosis who are at high risk of HIV infection. *Sexually transmitted*

infections **92**, 186-193 (2016).

- 47 Mitchell, C. M. *et al.* Vaginal microbiota and mucosal immune markers in women with vulvovaginal discomfort. *Sexually transmitted diseases* **47**, 269-274 (2020).
- 48 Srinivasan, S. *et al.* Bacterial communities in women with bacterial vaginosis: high resolution phylogenetic analyses reveal relationships of microbiota to clinical criteria. *PloS one* **7**, e37818 (2012).
- 49 Cartwright, C. P., Pherson, A. J., Harris, A. B., Clancey, M. S. & Nye, M. B. Multicenter study establishing the clinical validity of a nucleic-acid amplification–based assay for the diagnosis of bacterial vaginosis. *Diagnostic microbiology and infectious disease* **92**, 173-178 (2018).
- 50 France, M. T. *et al.* VALENCIA: a nearest centroid classification method for vaginal microbial communities based on composition. *Microbiome* **8**, 1-15 (2020).
- 51 Fettweis, J. M. *et al.* Differences in vaginal microbiome in African American women versus women of European ancestry. *Microbiology* **160**, 2272 (2014).
- 52 Oakley, B. B., Fiedler, T. L., Marrazzo, J. M. & Fredricks, D. N. Diversity of human vaginal bacterial communities and associations with clinically defined bacterial vaginosis. *Applied and environmental microbiology* **74**, 4898-4909 (2008).
- 53 van de Wijgert, J. H. The vaginal microbiome and sexually transmitted infections are interlinked: consequences for treatment and prevention. *PLoS medicine* **14**, e1002478 (2017).
- 54 Petrova, M. I., Reid, G., Vaneechoutte, M. & Lebeer, S. Lactobacillus iners: friend or foe? *Trends in microbiology* **25**, 182-191 (2017).
- 55 Vaneechoutte, M. Lactobacillus iners, the unusual suspect. *Research in microbiology* **168**, 826-836 (2017).
- 56 Schiffman, M. & Wentzensen, N. From human papillomavirus to cervical cancer. *Obstetrics & Gynecology* **116**, 177-185 (2010).
- 57 Schiffman, M. *et al.* Carcinogenic human papillomavirus infection. *Nature reviews Disease primers* **2**, 1-20 (2016).
- 58 Mitra, A. *et al.* Cervical intraepithelial neoplasia disease progression is associated with increased vaginal microbiome diversity. *Scientific reports* **5**, 1-11 (2015).
- 59 Piyathilake, C. J. *et al.* Cervical microbiota associated with higher grade cervical intraepithelial neoplasia in women infected with high-risk human papillomaviruses. *Cancer Prevention Research* **9**, 357-366 (2016).
- 60 Łaniewski, P. *et al.* Features of the cervicovaginal microenvironment drive cancer biomarker signatures in patients across cervical carcinogenesis. *Scientific reports* **9**, 1-14 (2019).

- 61 Schlecht, N. F. *et al.* Cervical, anal and oral HPV in an adolescent inner-city health clinic providing free vaccinations. *PloS one* **7**, e37419 (2012).
- 62 Schlecht, N. F. *et al.* Risk of delayed human papillomavirus vaccination in inner-city adolescent women. *The Journal of infectious diseases* **214**, 1952-1960 (2016).
- 63 Gutman, R. E., Peipert, J. F., Weitzen, S. & Blume, J. Evaluation of clinical methods for diagnosing bacterial vaginosis. *Obstetrics & Gynecology* **105**, 551-556 (2005).
- 64 Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME journal* **6**, 1621 (2012).
- 65 Wang, Y. & Qian, P.-Y. Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. *PloS one* **4**, e7401 (2009).
- 66 Rosenbaum, J. *et al.* evaluation of oral Cavity DNA extraction Methods on Bacterial and Fungal Microbiota. *Scientific reports* **9**, 1531 (2019).
- 67 Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**, 863-864, doi:10.1093/bioinformatics/btr026 (2011).
- 68 Hercus, C. Novocraft short read alignment package. *Website* <http://www.novocraft.com> (2009).
- 69 Bolyen, E. *et al.* QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science. Report No. 2167-9843, (PeerJ Preprints, 2018).
- 70 Callahan, B. J. *et al.* DADA2: high-resolution sample inference from Illumina amplicon data. *Nature methods* **13**, 581 (2016).
- 71 Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261-5267 (2007).
- 72 DeSantis, T. Z. *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72**, 5069-5072 (2006).
- 73 Chen, T. *et al.* The Human Oral Microbiome Database: a web accessible resource for investigating oral microbe taxonomic and genomic information. *Database* **2010** (2010).
- 74 Peterson, J. *et al.* The NIH human microbiome project. *Genome research* **19**, 2317-2323 (2009).
- 75 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *Journal of molecular biology* **215**, 403-410 (1990).
- 76 Kõljalg, U. *et al.* UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi. *New Phytologist* **166**, 1063-1068 (2005).

- 77 McDonald, D. *et al.* The Biological Observation Matrix (BIOM) format or: how I learned to stop worrying and love the ome-ome. *GigaScience* **1**, 7 (2012).
- 78 Team, R. C. R: A language and environment for statistical computing. (2013).
- 79 McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS one* **8**, e61217 (2013).
- 80 Oksanen, J. *et al.* Vegan: community ecology package. R package version 1.17-4. <http://cran.r-project.org>. Acesso em **23**, 2010 (2010).
- 81 Robin, X. *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC bioinformatics* **12**, 77 (2011).
- 82 Wickham, H. *ggplot2: elegant graphics for data analysis*. (Springer, 2016).
- 83 Mandal, S. *et al.* Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microbial ecology in health and disease* **26**, 27663 (2015).
- 84 Dabney, A., Storey, J. D. & Warnes, G. qvalue: Q-value estimation for false discovery rate control. *R package version 1* (2010).
- 85 Morton, J. T. *et al.* Establishing microbial composition measurement standards with reference frames. *Nature communications* **10**, 2719 (2019).
- 86 Kemp, T. J. *et al.* Evaluation of two types of sponges used to collect cervical secretions and assessment of antibody extraction protocols for recovery of neutralizing anti-human papillomavirus type 16 antibodies. *Clinical and Vaccine Immunology* **15**, 60-64 (2008).
- 87 Koshiol, J. *et al.* Evaluation of a multiplex panel of immune-related markers in cervical secretions: A methodologic study. *International journal of cancer* **134**, 411-425 (2014).
- 88 Amaro-Filho, S. M. *et al.* HPV73 a nonvaccine type causes cervical cancer. *International Journal of Cancer* **146**, 731-738 (2020).

Tables

Table 1. Risk factors for time to clearance of incident HR-HPV infections

Variable	Model 1: Sustained IL-1 β /IP-10 and molBV			Model 2: Parsimonious		
	HZ	CI	p.val	HZ	CI	p.val
IL-1 β /IP-10: Sustained-high (ref)	-	-	-	-	-	-
IL-1 β /IP-10: Sustained-low	1.90	1.08-3.40	0.026*	1.86	1.19-2.90	0.006**
Age	1.05	0.97-1.10	0.21	1.06	0.98-1.20	0.13
Smoking	0.75	0.53-1.10	0.11	0.75	0.53-1.10	0.12
HPV16 Status	0.98	0.61-1.60	0.93	0.98	0.62-1.60	0.94
molBV: Sustained-low (ref)	-	-	-	-	-	-
molBV: Became-high	0.98	0.52-1.80	0.94	-	-	-
molBV: Became-low	0.39	0.15-1.01	0.06	-	-	-
molBV: Sustained-high	1.08	0.58-2.00	0.81	-	-	-

This table shows the adjusted model for the outcome of incident HR-HPV infections. In Model 1, sustained IL-1 β /IP-10 levels are modeled with the inclusion of the molecular BV categories based on molBV at Visit 1 and Visit 2. For these analyses molBV was dichotomized as either high (above the median) or low (below the median). Model 2 shows the adjusted hazard ratio of IL-1 β /IP-10 levels without the molecular BV categories.

Table 2. Risk factors for progression to CIN2+

	Model 1: molBV			Model 2: IL-1 β /IP-10			Model 3: TNF- α /MIP-1 β			Model 4: IL-1 β /IP-10 and TNF- α /MIP-1 β			Model 5: Parsimonious		
	OR	CI	p.val	OR	CI	p.val	OR	CI	p.val	OR	CI	p.val	OR	CI	p.val
IL-1 β /IP-10	-	-	-	1.15	0.92-1.46	0.22	-	-	-	1.04	0.80-1.34	0.73	-	-	-
TNF- α /MIP-1 β	-	-	-	-	-	-	2.71	1.46-5.61	0.002**	2.65	1.39-5.58	0.004**	2.81	1.62-5.42	<0.001***
Age	1.24	1.02-1.55	0.039*	1.16	0.92-1.48	0.23	1.03	0.81-1.31	0.82	1.01	0.78-1.32	0.92	-	-	-
Smoking	0.92	0.92-0.77	0.33	0.92	0.77-1.10	0.40	0.86	0.71-1.03	0.10	0.86	0.71-1.04	0.12	0.86	0.71-1.02	0.08
HPV16 Status	1.47	0.66-3.23	0.33	1.51	0.67-3.36	0.31	1.37	0.59-3.14	0.54	1.39	0.59-3.21	0.53	1.35	0.59-3.05	0.56
molBV	1.26	0.48-3.29	0.63	1.27	0.48-3.35	0.67	1.62	0.58-4.69	0.44	1.60	0.57-4.65	0.45	1.63	0.58-4.70	0.44

Two inflammatory signatures are evaluated. Levels of IL-1 β /IP-10 and TNF- α /MIP-1 β were dichotomized into positive or negative based on above or below the median, respectively.

Figures

Figure 1

Microbial Community Features of Bacterial Vaginosis Panel A is a heatmap showing the abundance of the 15 most prevalent bacterial species as listed to the right of the heatmap. Each column represents a participant. Hierarchical clustering separates samples into two primary clades: one dominated by *Lactobacillus* and one with polymicrobialism. There is a significant tendency of the BV positive cases to be found in the polymicrobial clade and BV negative in the one dominated by *Lactobacillus* based on either the Amsel or Nugent diagnosis (see Materials and Methods) ($p < 0.001$ for both). Panel B shows the alpha diversity differences between BV diagnosed by either Amsel or Nugent criteria and the microbial communities based on the Chao1, Fisher and Shannon diversity indices (all $p < 0.001$). Panels C and D show beta diversity analyses

using PCoA and Jensen Shannon Diversity index for the Amsel BV, above ($R^2=0.25$, $p<0.001$) and Nugent BV diagnosis, below ($R^2=0.59$, $p<0.001$). Panel E shows the top 20 microbial markers (based on W-stat) for detecting BV using the “clean” BV status sample set (Amsel+/Nugent+ vs Amsel-/Nugent-). The y-axis represents the ANCOM W-stat, while the x-axis represents the mean relative abundance difference between BV+ and BV- cases for each bacterial taxon. The size of the circles represents relative abundance.

Figure 2

Cytokines associated with molBV Categorical States in the CVT dataset A. Forest plot showing the OR and associated confidence intervals computed using a linear model between the cytokine quartiles and the three categorical states of molBV (i.e., BV-negative, BV-intermediate, and BV-positive). Only cytokines with an adjusted q-value<0.001 are presented. ORs in panel A represent the odds of moving to the immediate next ordinal BV state. Panel B shows the correlation network between cytokines in the analyzed panel with a Pearson correlation >0.6. Panel C is a volcano plot showing the ORs on the x-axis when comparing molecular BV-negative vs. BV-positive and the $-\log(q\text{-value})$ on the y-axis. The ratios that had a q-value $<3.77 \times 10^{-44}$ (i.e. $-\log(q\text{-value}) > 100$) are highlighted and labeled in red. Panel D shows the pairwise Pearson correlation of the highly significant ratios presented in panel C (colored red). Ratios that had a OR <1.0 were inverted for

symmetry of data presentation. Panel E presents a box and violin plot for the log(IL-1 β /IP-10) ratio for BV-negative (colored in blue), BV-intermediate (colored in yellow) and BV-positive (colored in red). Panel F shows the bacterial species identified by ANCOM predicting IL-1 β /IP-10 inflammation (i.e., above the median) in women within the BV-negative group. Panel G shows the ORs for top 10 ratio combinations of the 32-cytokines (based on adjusted q-value) when comparing BV-negative to BV-positive women that had IL-1B/IP-10 levels below the cohort median.

Figure 3

Molecular BV, inflammation and HR-HPV Clearance Panel A shows the Kaplan-Meier curve for HR-HPV clearance colored by sustained BV status (i.e., molBV high) with the unadjusted p-value presented in the bottom left corner and number of women at risk at each time point presented in accompanying table directly below the plot. Panel B shows the Kaplan-Meier curve comparing women with either sustained or absent IL-1 β /IP-10 inflammation across the two analyzed visits.

Supplementary Files

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

- [naturecommunicationsscript.zip](#)
- [SupplementalMaterials.pdf](#)