

Vaginal microbiota varies by geographical location in South African women

Authors:

Katie Lennard¹, Smritee Dabee², Shaun L Barnabas², Enock Havyarimana², Anna Blakney³, Shameem Z Jaumdally, Gerrit Botha¹, Nonhlanhla N Mkhize², Linda-Gail Bekker⁴, David A Lewis^{5,6,7}, Glenda Gray^{8,9}, Nicola Mulder¹, Jo-Ann S Passmore^{2,10}, Heather B Jaspan^{2,11*}

Affiliations:

¹Institute of Infectious Disease and Molecular Medicine and Computational Biology Division, Department of Integrative Biomedical Sciences, University of Cape Town, South Africa

²Institute of Infectious Disease and Molecular Medicine and Department of Pathology, University of Cape Town, South Africa

³Department of Bioengineering, University of Washington, United States

⁴Institute of Infectious Disease and Molecular Medicine, Desmond Tutu HIV Centre, University of Cape Town, South Africa;

⁵Western Sydney Sexual Health Centre, Australia

⁶Marie Bashir Institute for Infectious Diseases and Biosecurity and Sydney Medical School-Westmead, University of Sydney, Australia

⁷National Institute for Communicable Diseases, South Africa

⁸Perinatal HIV Research Unit, University of the Witwatersrand, South Africa

⁹South African Medical Research Council, Cape Town, South Africa

¹⁰National Health Laboratory Service, South Africa

¹¹Seattle Children's Research Institute, Department of Pediatrics and Global Health, University of Washington, USA

Corresponding author:

Dr Heather Jaspan
hbjaspan@gmail.com

Dates:

Received: 17/09/2018

Accepted: 12/04/2019

Published:

Women of African descent are more likely to have bacterial vaginosis than women of other ethnicities. Both diversity and likely specific taxa in these microbial communities are important to sexual and reproductive health, such as HIV risk. However, whether the specific taxa also vary by geographical location and/or ethnicity requires further investigation.

Here, we compare the vaginal microbiota of 16–22-year-old black, HIV-negative South African women from two geographically disparate but low-income high population density communities, one in Cape Town (CPT) and one in Johannesburg (JHB). Vaginal microbiota composition was assessed by 16S rRNA gene amplicon sequencing of lateral vaginal wall swabs.

Geographical location was significantly associated with vaginal microbiota composition by permutational analysis of variance (PERMANOVA) ($p=0.02$), as were body mass index BMI ($p=0.015$) and human papilloma virus (HPV) risk type ($p=0.005$), while the presence of one or more sexually transmitted infections (STIs) ($p=0.053$) and hormonal contraceptive (HC) usage ($p=0.4$) were not. Geographical location remained a significant determinant of microbiota composition independent of BMI, STI status and HPV-risk. Together, geographical location, BMI and HPV-risk explained 10% of the variance in microbiota composition with a large proportion of the variance remaining unexplained. Several taxa differed significantly between geographical location – some by frequency and others by relative abundance.

Our results therefore suggest that HIV prophylactic approaches targeting the vaginal microbiota should be geographically tailored.

Geografiese ligging beïnvloed vaginale mikrobiële profiele in Suid Afrikaanse vroue:

Vroue van Afrika-afkoms is meer vatbaar vir bakteriële vaginose (BV) in vergelyking met Europese vroue. Beide mikrobiële diversiteit (soos met BV) sowel as spesifieke bakteriële taksa speel 'n rol in seksuele en reproductiewe gesondheid insluitende MIV vatbaarheid.

Die moontlike rol van geografiese ligging en etnisiteit op die verhouding tussen mikrobiële samestelling en seksuele en reproductiewe gesondheid bly egter onbekend.

In hierdie studie vergelyk ons dus die vaginale mikrobiota van 16–22-jarige swart, HIV-negatiewe Suid Afrikaanse vroue van twee geografiese-uiteenlopende liggings, beide lae-inkomste, hoë bevolkingsdigtheidsgemeenskappe, een in Kaapstad, en een in Johannesburg. Vaginale mikrobiële profiele is bepaal met behulp van 16S rRNS volgordebepaling van laterale muur deppers.

Ons pas permutasie variansieanalise (PERMANOVA) toe en vind statisties betekenisvolle assosiasies tussen vaginale mikrobiële samestelling en geografiese ligging ($p=0.02$), asook met liggaamsmassa-indeks (LMI) ($p=0.015$) en menslike papilloomvirus (MPV) risikotipe ($p=0.005$), maar nie met die voorkoms van een of meer seksueel-oordraagbare infeksies (SOI's) ($p=0.053$) of met hormonale kontrasepsie verbruik nie. ($p=0.4$)

Geografiese ligging was 'n statisties betekenisvolle determinant van mikrobiële samestelling, ongeag verskille in LMI, SOI status en MPV-risiko tipes tussen Kaapstad en Johannesburg vroue. Geografiese ligging, LMI en MPV-risiko verduidelik gesamentlik 10% van die variansie in mikrobiële samestelling, met 'n groot persentasie van onbekende oorsprong. Verskeie taksa het statisties betekenisvol verskil in terme van frekwensie of relatiewe vlakke van voorkoms tussen die geografiese liggings.

Ons resultate stel voor dat MIV profilaktiese metodes wat die vaginale mikrobiota teiken die effek van geografiese ligging in ag moet neem.

How to cite this article:

Katie Lennard, Smritee Dabee, Shaun L Barnabas, Enock Havyarimana, Anna Blakney, Shameem Z Jaumdally, Gerrit Botha, Nonhlanhla N Mkhize, Linda-Gail Bekker, David A Lewis, Glenda Gray, Nicola Mulder, Jo-Ann S Passmore, Heather B Jaspan, Vaginal microbiota varies by geographical location in South African women, *Suid-Afrikaanse Tydskrif vir Natuurwetenskap en Tegnologie* 38(1) (2019)

'n Afrikaanse vertaling van die manuskrip is aanlyn beskikbaar by <http://www.satnt.ac.za/index.php/satnt/article/view/685>

Copyright:

© 2019. Authors.

Licensee: *Die Suid-Afrikaanse Akademie vir Wetenskap en Kuns*. This work is licensed under the Creative Commons Attribution License.

Introduction

Vaginal microbiota profiles vary by ethnicity (Srinivasan et al. 2012; Buvé et al. 2014; Ravel et al. 2010). Women of African descent less commonly have *Lactobacillus*-dominant vaginal microbiota compared with Caucasian women (Ravel et al. 2010; Anahtar et al. 2015b; Lennard et al. 2017). This finding appears to be generalisable to African American and Hispanic women from North America who frequently have decreased relative abundance of *Lactobacilli* (Anahtar et al. 2015a; Fettweis et al. 2014; Zhou et al. 2007). It is less clear to what extent geographical location affects vaginal microbiota composition among women of the same ethnicity. Bacterial vaginosis (BV) rates vary by ethnicity and geographical location (with potentially large variation in the proportion of BV among different African countries) (Kenyon, Colebunders, and Crucitti 2013). Yet, detailed description of vaginal microbiota composition by geographical location is currently lacking.

It has long been recognised that bacterial vaginosis (a vaginal dysbiosis) is associated with adverse sexual and reproductive health outcomes, including sexually transmitted infections (Wiesenfeld et al. 2003; Gallo et al. 2012; Balkus et al. 2014) and adverse birth outcomes (Leitich and Kiss 2007; Holst, Goffeng, and Andersch 1994; Nelson et al. 2015). Recently, with the advent of next generation sequencing, specific taxa have been implicated in these outcomes – such as preterm births (Freitas et al. 2018; Tabatabaei et al. 2018; Vinturache et al. 2016) and HIV risk (McClelland et al. 2018). In a study conducted on five separate cohorts from Kenya, Uganda, South Africa, Tanzania, Botswana and Zambia, McClelland et al. identified taxa that were associated with increased odds of HIV acquisition across all cohorts considered, some of which were significantly so (*Parvimonas* species type 1 and 2, *Gemella asaccharolytica*, *Mycoplasma hominis*) (McClelland et al. 2018). The question as to whether we can define a robust microbiota signature of HIV risk that is generalisable across geographical locations/ethnicities, or whether location-specific taxa should be identified for HIV risk assessments remains.

Here, we compare the vaginal microbiota of 16–22-year-old black HIV-negative South African women from two low-income high population density communities; one in Cape Town (CPT) and the other Johannesburg (JHB).

Materials and Methods

Participant selection and sample collection

The cohorts have been described previously in detail (Barnabas et al. 2017; Lennard et al. 2017). Briefly, 298 black, 16–22-year-old HIV-negative South African women were recruited as part of the Women's Initiative in Sexual Health (WISH) study (Barnabas et al. 2017) from low-income, high population density communities in Cape Town (CPT) and

Johannesburg (JHB). Approval was obtained for the study from the Research Ethics Committees of the Universities of Cape Town and Witwatersrand. All participants ≥ 18 years provided informed consent, while assent and parental consent were obtained for participants ≤ 18 years. Young women were enrolled if they were HIV-negative, in general good health, not pregnant or menstruating at the time of sampling, and if they had not had unprotected sex or douched in the last 48 hours. Additional exclusion criteria were use of antibiotics in the prior two weeks. Study visits were scheduled two weeks after injection for participants on injectable progestin contraceptives, or otherwise during the luteal phase of their menstrual cycles (between day 14–28) if they were not using any HCs or if they were using oral HCs. Before specimen collection, the following were performed: HIV pre-test and risk-reduction counselling, an HIV rapid test (Alere Determine™ HIV-1/2 Ag/Ab Combo, Alere, Waltham, MA), a pregnancy test (U-test Pregnancy strip, Humor Diagnostica, Pretoria, South Africa) and a general physical examination. Cervicovaginal fluid via disposable menstrual cup (Softcup®), one vulvovaginal swab for STI testing and one lateral vaginal wall swab for Nugent scoring and microbiome analysis were collected.

STI and BV testing

Vulvo-vaginal swabs were assayed for nucleic acid of the following STIs by multiplex PCR: *Chlamydia trachomatis*, *Neisseria gonorrhoeae*, *Trichomonas vaginalis*, *Mycoplasma genitalium*, HSV-1 and -2, *Haemophilus ducreyi*, *Treponema pallidum* and lymphogranuloma venereum as previously described (Lewis et al. 2012). Endo-cervical swabs were collected for HPV detection and genotyping by Roche Linear Array (Mbulawa et al. 2018). The following HPV types were considered high-risk HPV: 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68 (Jacobs et al. 1997). For analyses referring to STI (any), women considered positive had at least one of the STIs tested for in this study, excluding HPV. Lateral wall/posterior fornix swabs were collected for Nugent scoring to classify samples as BV negative (Nugent 0–3), intermediate (Nugent 4–6) or positive (Nugent 7–10); and vaginal pH was measured using colour-fixed indicator strips (Macherey-Nagel, Düren, Germany).

16S sequencing and analysis

Swabs were thawed, treated with a cocktail of mutanolysin (25kU/ml, Sigma Aldrich), lysozyme (450kU/ml, Sigma Aldrich), and lysostaphin (4kU, Sigma Aldrich), then mechanically disrupted with a bead-beater. DNA was extracted using the MoBio PowerSoil DNA extraction kit (MoBio, Carlsbad, CA). The V4 region of the 16S rRNA gene was amplified using universal primers that were modified to encode the Illumina MiSeq sequencing primer sequence at the 5' end (46): 515F (TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG NNN NNG TGC CAG CMG CCG CGG TAA) and 806R (GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GNN NNN GGA

CTA CHV GGG TWT CTA AT), where NNNNN indicates five randomly incorporated nucleotides for increased complexity (Pearce, Hilt, and Rosenfeld 2014). The 5' end is the Illumina Nextera adapter, and the sequences following the Ns are complementary to the V4 rRNA gene region.

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

Following demultiplexing, raw reads were preprocessed as follows: forward and reverse reads were merged using `usearch7` (Edgar 2010), allowing a maximum of three mismatches; merged reads were quality filtered using `usearch7` (reads with E scores larger than 0.1 were discarded); primer sequences were removed using a custom python script; and merged, filtered reads were truncated at 250bp. Next, sequences were de-replicated whilst recording the level of replication for each sequence using `usearch7`. De-replicated sequences were sorted by abundance (highest to lowest) and clustered de novo into operational taxonomic units (OTUs) at 97% similarity using `usearch7`. Chimeric sequences were detected (against the Gold database) using UCHIME (Edgar et al. 2011) and removed. Individual sequences were assigned to the specific identifiers using a 97% similarity threshold. Taxonomic assignment was performed in QIIME 1.8.0 (Caporaso et al. 2010) using the RDP classifier (using the default confidence level of 0.5) against the GreenGenes 13.8 reference taxonomy for 97% identity. To increase species-level resolution, we constructed a custom taxonomic database appropriate for V4 region 16S rRNA gene amplicon sequencing based on the custom vaginal 16S rRNA gene reference database created by Fettweis et al. (Fettweis et al. 2012). This database was updated for the V4 region and used to increase species-level resolution as previously described (Lennard et al. 2017). OTUs that mapped to more than one species (with the same identity score) were annotated as follows: if an OTU mapped to two or three species, the OTU would be named *Genus speciesA_speciesB* or *Genus speciesA_speciesB_speciesC*, respectively, and if an OTU mapped to more than three species but one species was clearly associated with vaginal microbiota (based on prior knowledge), the OTU was named *Genus species_cluster*, where "species" was selected based on the majority of hits; e.g., *L. reuteri_cluster* indicates the case where the majority of hits were for *L. reuteri* but there were several other species with equal identity scores present.

Samples with ≥ 5000 reads were selected for downstream analyses. The OTU table was standardised (i.e. transformed to relative abundance and multiplied by the median sample read depth), and filtered so that each OTU had to have at least 10 counts in at least 2% of samples or have a relative abundance of at least 0.001%.

Statistical analyses

All downstream statistical analyses were performed in R, using the packages `phyloseq` (McMurdie and Holmes 2013) for beta diversity analyses, `metagenomeSeq` (Paulson et al. 2013) for differential abundance testing, `vegan` (Oksanen et al. 2016) for ordinations and redundancy analysis, and `NMF` (Gaujoux 2014) for annotated heat maps.

Permutational multivariate analysis of variance (PERMANOVA) was performed using the `adonis` and `adonis2` functions from the R package `vegan` (Oksanen et al. 2016); for the `adonis` function the order of predictor variables matter, while the order of terms do not affect results in the `adonis2` function. Because we did not wish to make assumptions regarding the relative importance of predictor variables, `adonis2` was used to obtain p-values for individual variables, while `adonis()` was used to obtain adjusted R^2 values (which are not available when using `adonis2`). The assumption for PERMANOVA of homogeneity of variance between groups was assessed using the `betadisper()` function from the R package `vegan` (Oksanen et al. 2016). This assumption was met when using Bray-Curtis as distance metric, but not when using UniFrac distance or weighted UniFrac distance; hence we used Bray-Curtis distance. In the final model ethnicity (for which 24 participants had missing information) was excluded as ethnicity was not significant when performing PERMANOVA on the subset of participants for whom we did have ethnicity information.

Distance-based redundancy analysis (db-RDA) was performed on Bray-Curtis dissimilarity matrix using the `dbrda()` function from the R package `vegan` and the ordination was constrained on geographical location, STI (any), BMI and HPV risk (variables that were not significant by PERMANOVA were excluded from the final model used for visual presentation in Figure 1).

Differences in microbial composition between groups of interest were assessed using the R package `metagenomeSeq`'s `MRfulltable` function with a custom filter to determine significance: merged taxa were deemed significantly different if they exhibited a fold change (beta coefficient) of ≥ 1.5 , had an adjusted p-value of ≤ 0.01 and if at least one of the two groups being compared had $\geq 20\%$ of samples with the given taxon OR the Fisher's exact test result was significant (after multiple testing correction). OTUs were first merged at the lowest available taxonomic level

(i.e. for OTUs with *Lactobacillus* as the lowest available taxonomic annotation counts were summed, while OTUs with additional species-level annotation e.g. *L. iners* were summed at species-level instead). Composite barplots (Figure 2) were also created based on this merged table. The most abundant taxa were selected as follows: For each sample the most abundant taxa were determined (based on standardised, merged taxon counts), after ranking taxa for each sample by read counts (high to low) and selecting those taxa that cumulatively made up the first 50% of reads for that sample. This resulted in a list of 28 unique taxa across all samples, which was then limited to the subset that had been classified as 'abundant' in at least two samples, reducing the number of abundant taxa to 12 (Figure 2).

Random forests analyses were conducted on merged taxa to determine which taxa best predicting geographical location using the R packages *randomForest* (Liaw and Wiener 2002) and *ROCR* (Sing et al. 2005) for ROC analysis. The full dataset in question was used to train random forests models, i.e. the data were not divided into training and test sets.

Results

Microbiota profiling was performed by 16S rRNA gene amplicon sequencing for 102 women from CPT and 79 women from JHB (Table 1). The two groups were well

matched in terms of age (median 18 years for both locations). Hormonal contraceptive usage differed significantly with 100% of CPT women compared with 41% of JHB women on some form of hormonal contraceptive, likely due to differences in recruitment approaches between the two sites (CPT participants were recruited through a family planning clinic while JHB participants were recruited from a broader population). CPT women had higher BV prevalence (55 vs. 35%), STI prevalence (59 vs. 24%), BMI (25.4 vs. 22.5), had higher levels of genital inflammation and were of more homogenous ethnicity than JHB women (Table 1).

To identify factors influencing vaginal microbiota profiles, permutational multivariate analysis of variance (PERMANOVA) was performed. Factors considered included ethnicity, age, hormonal contraceptive usage (yes/no), the presence of any one or more STI excluding HPV (yes/no), HPV-risk (high/low/negative), geographical location and BMI. Ethnicity was not included in the final model since there were 24 participants for whom we did not have ethnicity information and ethnicity was not a significant factor when performing PERMANOVA on the subset for whom we did have ethnicity information. Age was also not included in the final model as there was no significant difference in age between JHB and CPT (Table 1). Factors significantly associated with vaginal microbiota composition were geographical location ($p=0.02$), BMI ($p=0.015$), and

TABLE 1: Participant summary by geographical location

Feature	Cape Town (N=102)	Johannesburg (N=79)	P value
Median age, years	18	18	0.6
BV prevalence, n (%)			0.008
<i>BV positive</i>	56 (55)	28 (35)	
<i>BV intermediate</i>	7 (7)	15 (19)	
<i>BV negative</i>	39 (38)	36 (46)	
Nugent score (median)	8	4	0.01
STI (any), n (%)	60 (59)	19 (24)	2.7e-6
<i>C. trachomatis</i>	45 (44)	13 (17)	1e-4
<i>N. gonorrhoeae</i>	14 (14)	4 (5)	0.08
<i>T. vaginalis</i>	6 (6)	3 (4)	0.7
<i>M. genitalium</i>	4 (4)	2 (3)	0.7
<i>HSV-2 (DNA)</i>	6 (6)	1 (1)	0.1
HPV risk, n (%)			0.3
<i>High</i>	42 (41)	33 (42)	
<i>Low</i>	29 (28)	15 (19)	
<i>Negative</i>	31 (30)	31 (39)	
Hormonal contraceptives [§] , n (%)			< 2.2e-16
<i>DMPA</i>	19 (19)	9 (12)	
<i>Implanon</i>	8 (8)	0 (0)	
<i>Nur isterate</i>	70 (69)	12 (15)	
<i>OCP</i>	4 (4)	6 (8)	
<i>Male condom</i>	0 (0)	36 (46)	
<i>Nuvaring</i>	1 (1)	0 (0)	
<i>Injectable (type not specified)</i>	0 (0)	5 (6)	
<i>None</i>	0 (0)	10 (13)	
Using hormonal contraceptives, n (%)			< 2.2e-16
<i>Any hormonal contraceptive</i>	102 (100)	32 (41)	
<i>Condoms/none</i>	0 (0)	46 (59)	
Ethnicity [#] , n (%)			< 2.2e-16
<i>Xhosa</i>	94 (99)	17 (28)	
<i>Other^²</i>	1 (1)	44 (72)	
BMI (median)	25.3	22.5	0.04

Y=Yes; N=No; [§]one woman had incomplete HC data; [#]25 women had incomplete ethnicity data

HPV-risk ($p=0.005$), while STI ($p=0.053$) and HC use ($p=0.4$) were not. Together these factors explained $\sim 10\%$ of the variation in microbiota composition with the remaining 90% of unknown origin. Given the large discrepancy in HC use between geographical locations (Table 1) PERMANOVA was redone on the subset of women who used HC, excluding those who used condoms only or no form of contraceptive. Again location ($p=0.02$), BMI ($p=0.03$) and HPV-risk ($p=0.05$) were significant while STI use was not ($p=0.7$).

Distance-based redundancy analysis (db-RDA) was performed on the Bray-Curtis dissimilarity matrix, which confirmed the PERMANOVA results (Figure 1); db-RDA is a constrained principal coordinates analysis, which allows the use of non-Euclidean dissimilarity indices such as Bray-Curtis, therefore more suited to 16S rRNA gene microbiome

data. Factors included in the db-RDA ordination included geographical location, STI other than HPV (yes/no), HPV-risk and BMI (i.e. factors that were significant by PERMANOVA, with STIs $p=0.053$). To further confirm the significance of geographical location on microbiota composition, factors that vary significantly by location (STI(yes/no), HPV-risk and BMI) were partialled out in the db-RDA model, yet location remained significant ($p=0.02$).

The most abundant taxa are summarised by geographical location in Figure 2.

To determine which taxa significantly differed between CPT and JHB, differential abundance analysis was performed using the R package metagenomeSeq. The analysis was performed on taxa merged at the lowest available

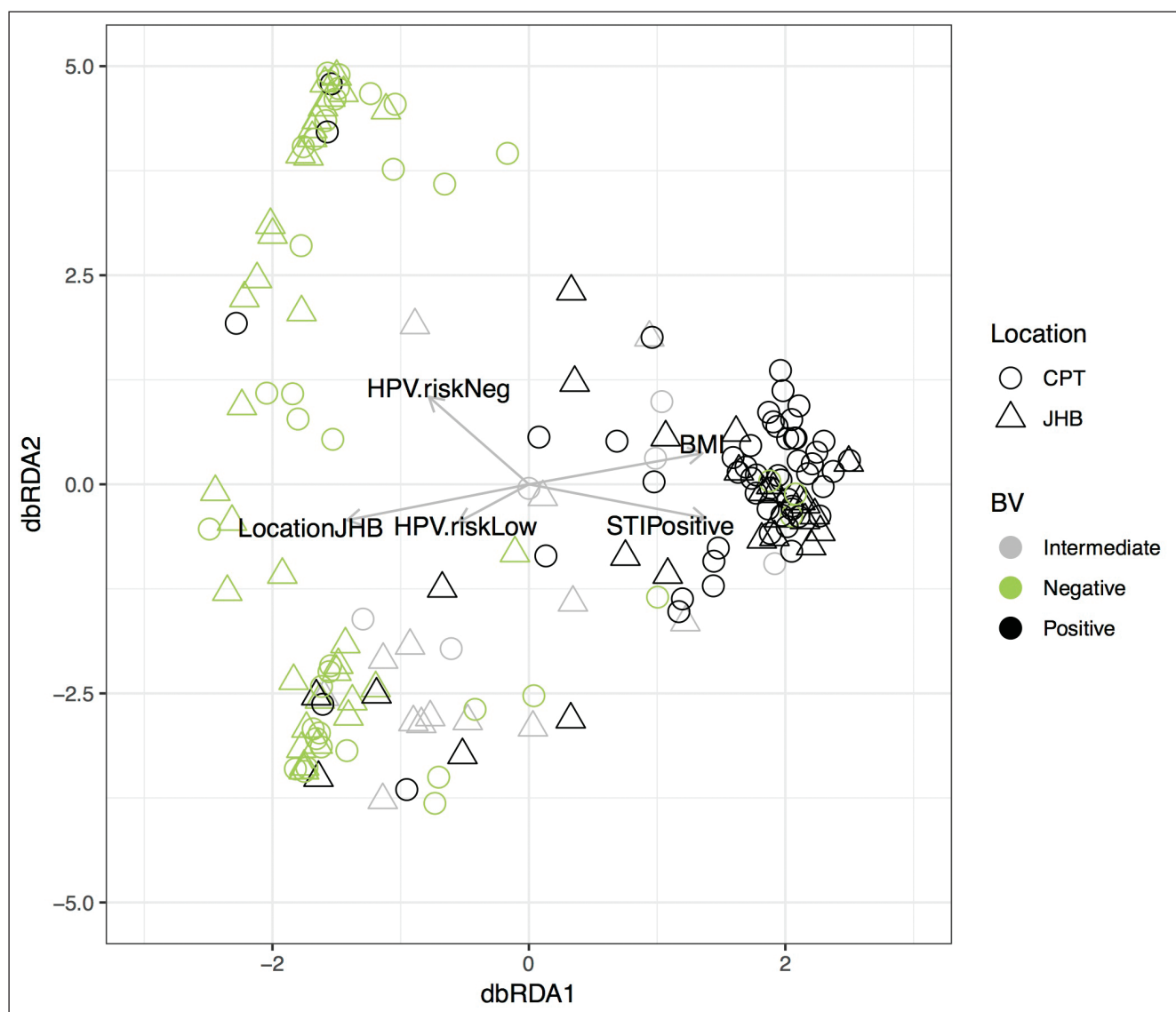


FIGURE 1: Distance-based redundancy analysis (dbRDA) on the Bray-Curtis dissimilarity matrix by geographical location, STI(yes/no), HPV-risk (high/medium/low) and BMI

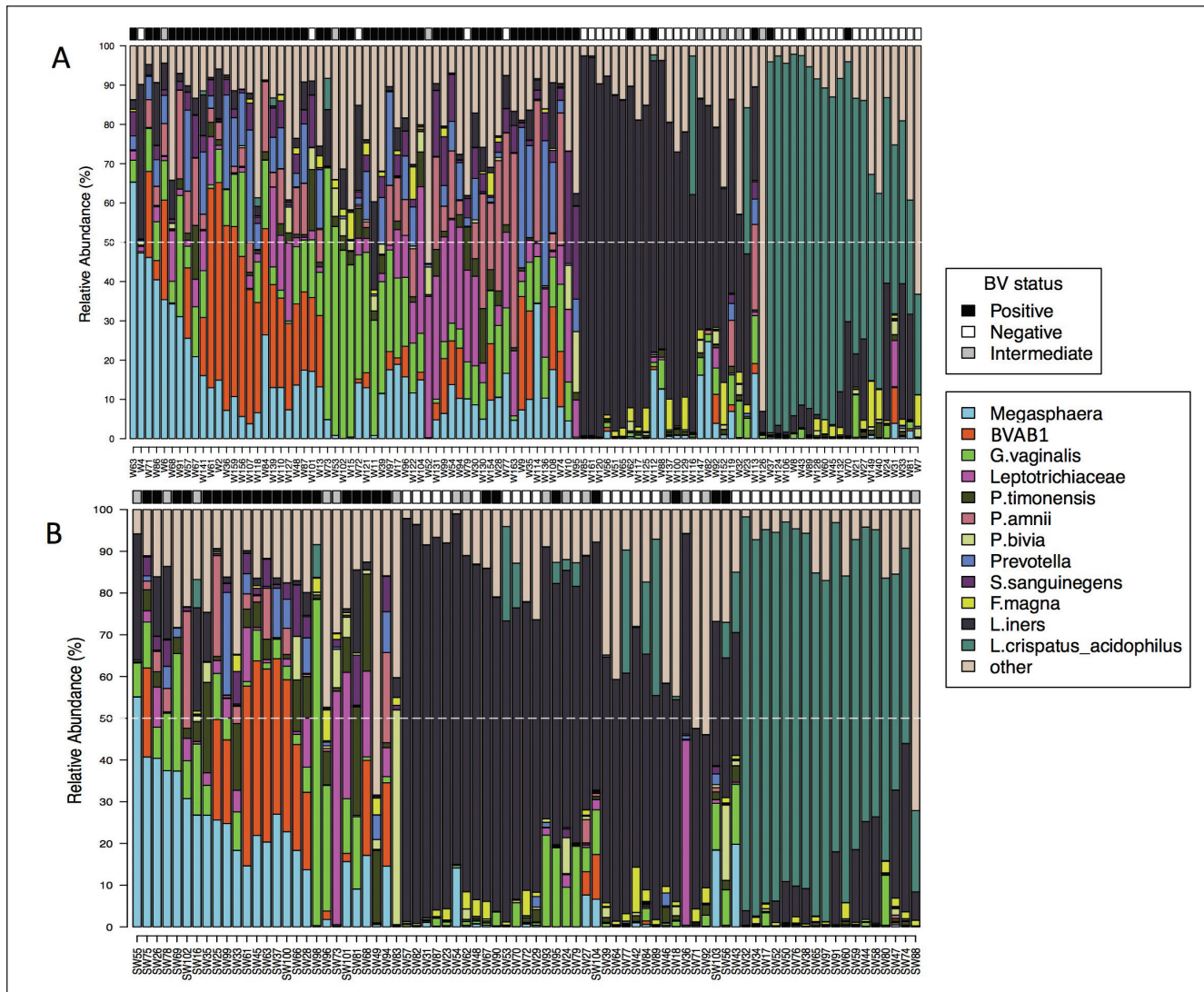


FIGURE 2: Composite barplot highlighting the most abundant taxa in A) CPT and B) JHB. Taxa included in the legend were selected from those taxa that made up the first 50% of each sample when ranked by abundance. Participants are ordered based on their most dominant taxon, matching the order of the figure legend (e.g. participants for whom *Megasphaera* was the most abundant taxon are listed first); other: the summed abundance of all taxa not included in the figure legend.

taxonomic annotation (see Methods for details). Eighteen taxa differed significantly in terms of frequency and/or relative abundance between JHB and CPT (Figure 3).

Taxa that were significantly higher in frequency (i.e. proportion of positive samples) in CPT compared to JHB included *Bifidobacterium*, *Prevotella pallens*, *Pseudomonas*, *Elizabethkingia meningoseptica*, *Brevundimonas*, *Mycoplasmataceae* and *Chryseobacterium* whereas *Lactobacillus coleohominis*, *Lactobacillus reuteri_cluster*, *Morganella morganii* and *Varibaculum cambriense* were more common in JHB women. Taxa that were present at similar frequencies between the CPT and JHB but varied in terms of relative abundance were *Leptotrichiaceae*, *Sneathia sanguinegens*, *P. amnii*, *Prevotella* and *BVAB3* (*Ageibacillus indolicus*), all of which had higher relative abundance in samples from CPT.

Random forest analysis identified *M. morganii* and *V. cambriense* as the highest ranked taxa to distinguish samples from JHB vs. CPT (training AUC=0.95, PPV=0.91, NPV=0.89).

Discussion

Vaginal microbiota profiles are known to vary by ethnicity and geographical location. Here we demonstrate differences in the relative abundance and frequency of colonisation of specific vaginal microbiota in African women from CPT and JHB, of similar ages and socioeconomic backgrounds. These differences could not fully be explained by factors that differed by geographical location, including hormonal contraceptive usage, ethnicity, BMI, HPV-risk or the occurrence of STIs. Together, geographical location, BMI and HPV-risk explained 10% of the variance in microbiota composition with a large proportion of the variance remaining unexplained.

McClelland et al. found in five different African cohorts, that the concentration of certain taxa were associated with later HIV seroconversion. In sensitivity analyses using frequency, however, certain of these taxa were clearly of more importance in specific cohorts. For example, the

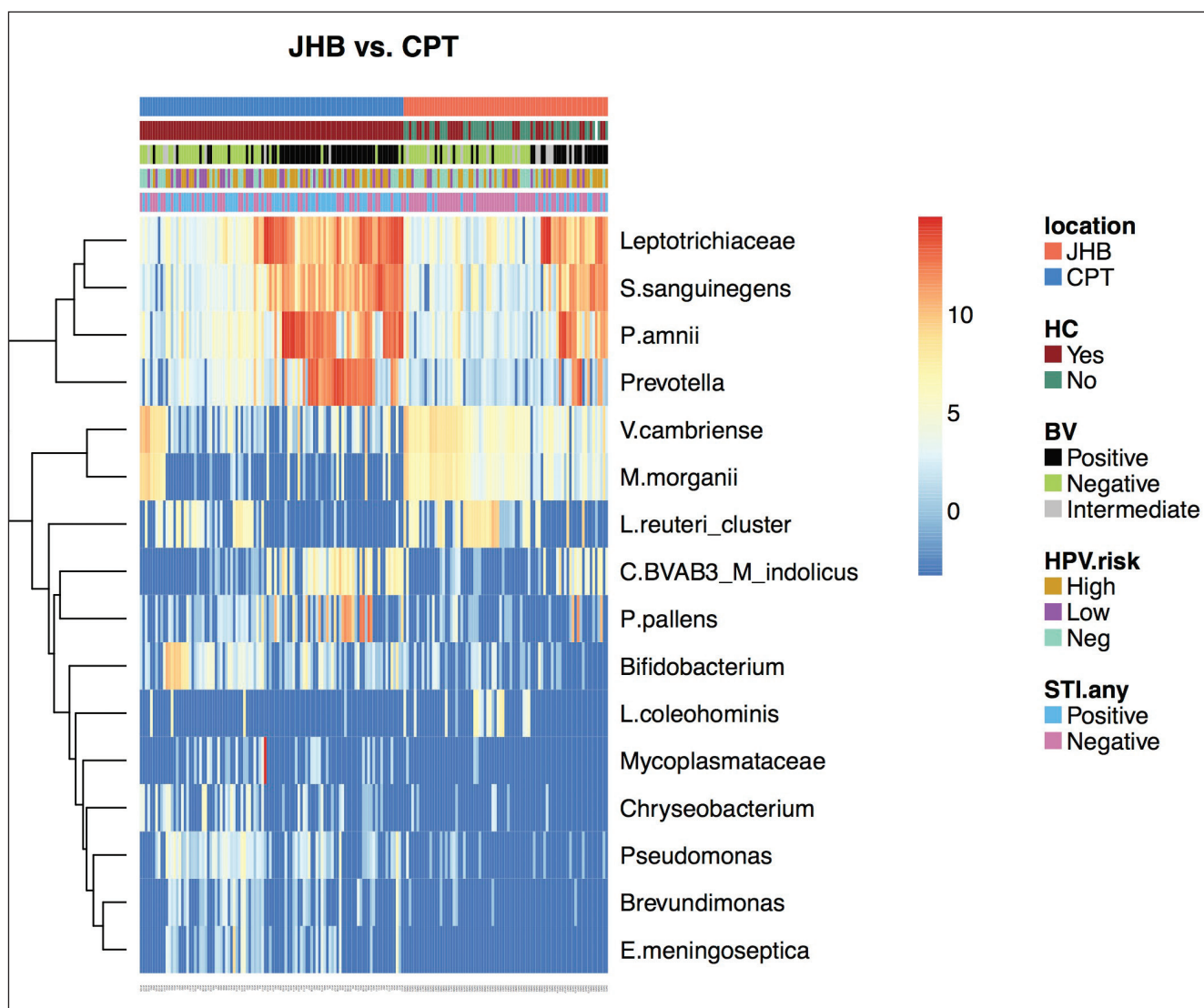


FIGURE 3: Taxa (merged at lowest available taxonomic annotation) that differed significantly in terms of relative abundance and/or frequency between CPT and JHB. Columns have been manually sorted by location, but hierarchical clustering was performed within each location subset. Results were filtered as follows: $FDR \leq 0.01$, beta coefficient ≥ 1.5 and each taxon had to be present in at least 20% of samples from either or both locations.

presence of detectable *Mycoplasma hominis* played a role in HIV risk in Kenyan female sex workers but not in serodiscordant couples from Uganda and South Africa, where *Gemella* and *Parvoimonas* were more important players in the latter cohort (McClelland et al. 2018). In a study of women from KwaZulu Natal, Williams et al found that the relative abundance of *Prevotella bivia* to be the taxon most predictive of later HIV seroconversion (Williams, AIDS Conference 2016). Finally, Gossman et al, also in a cohort from KZN but younger than the CAPRISA cohort, found that relative abundances of *P. melaninogenica* and *Veillonella montpellierensis* were the taxa most predictive of later HIV seroconversion (Gosmann et al. 2017).

Several studies of vaginal microbiota and preterm birth have found Lactobacillus-dominant vaginal microbiota to be protective, however, no taxa have consistently been associated with increased risk of this outcome (Dingens 2016, Romero 2014). Freitas et al. found the concentration of

Mollicutes to be a potential risk factor (Freitas et al. 2018). Although *Gardnerella* and *Ureaplasma* relative abundance were predictive of preterm birth in a predominantly Caucasian cohort from California, Callahan et al. were unable to replicate these findings in a predominantly African American cohort from Alabama (Callahan et al. 2017).

In summary, although there may be a subset of taxa consistently associated with adverse sexual and reproductive outcomes across a range of geographical locations, several clinically relevant taxa may be missed if geographical context is ignored. It remains unclear what might be driving these geographical differences in vaginal microbiota composition – environmental/community microbiota composition during early-life establishment of the microbiome likely plays an important role. Independent of the origin of these differences, our results strongly argue for geographically tailored microbiome-based diagnostics and therapeutics, even within the same country.

Acknowledgements

This study was supported by grants from the European and Developing Countries Clinical Trials Partnership (EDCTP) Strategic Primer grant [SP•2011•41304•038] and the South African Department of Science and Technology [DST/CON 0260/2012]. KL was supported by the National Research Foundation and the Suid-Afrikaanse Akademie vir Wetenskap en Kuns. HBJ was supported in part by K08HD069201. SLB was supported by the HIV Vaccine Trials Network SHAPe Program, the Fogarty Foundation and the South African Medical Research Council (MRC). SD was supported by the National Research Foundation of South Africa. The DTHF also recognises the support from ViiV health care in their YouthShield program. PHRU was supported through funding from the South African Medical Research Council. We thank the WISH Study Teams, particularly Ms Pinky Ngobo, Sr Nozipho Hadebe, Sr Janine Nixon, and all the young women who kindly participated in the study. We thank Prof Lynn Morris, David Lewis, Venessa Maseko and Raveshni Durgiah from the National Institute for Communicable Diseases for their help with sample processing. Computations were performed using facilities provided by the University of Cape Town's ICTS High Performance Computing team: <http://hpc.uct.ac.za>

References

- Anahar, M. N., Byrne, E.H., Doherty, K. E., Bowman, B. A., Yamamoto, H. S., Soumillon, M., Padavattan, N., et al. 2015a. Cervicovaginal Bacteria Are a Major Modulator of Host Inflammatory Responses in the Female Genital Tract. *Immunity* 42 (5). Elsevier Inc.:965-76. <https://doi.org/10.1016/j.immuni.2015.04.019>.
- Anahar, M.N., Byrne, E. H., Doherty, K. E., Bowman, B. A., Yamamoto, H. S., Soumillon, M., Padavattan, N., et al. 2015b. Cervicovaginal Bacteria are a Major Modulator of Host Inflammatory Responses in the Female Genital Tract. *Immunity* 42 (5). Elsevier Inc.:965-76. <https://doi.org/10.1016/j.immuni.2015.04.019>.
- Balkus, J. E., Richardson, B. A., Rabe, L. K., Taha, T. E., Mgodji, N., Kasaro, M. P., Ramjee, G., Hoffman, I. F., and Karim, S. S. A. 2014. Bacterial Vaginosis and the Risk of Trichomonas Vaginalis Acquisition among HIV-1-Negative Women. *Sexually Transmitted Diseases* 41 (2):123-28. <https://doi.org/10.1097/OLQ.0000000000000075>.
- Barnabas, S. L., Dabee, S., Passmore, J-A. S., Jaspán, H. B., Lewis, D. A., Jaumdally, S. Z., Gamielidien, H., et al. 2017. Converging Epidemics of Sexually Transmitted Infections and Bacterial Vaginosis in Southern African Female Adolescents at Risk of HIV. *International Journal of STD & AIDS*.
- Buvé, A., Jaspers, V., Crucitti, T. and Fichorova, R. N. 2014. The Vaginal Microbiota and Susceptibility to HIV. *Aids* 28 (16):2333-44. <https://doi.org/10.1097/QAD.0000000000000432>.
- Callahan, B. J., DiGiulio, D. B., Goltsman, D. S. A., Sun, C. L., Costello, E. K., Jeganathan, P., Biggio, J. R., et al. 2017. Replication and Refinement of a Vaginal Microbial Signature of Preterm Birth in Two Racially Distinct Cohorts of US Women. *Proceedings of the National Academy of Sciences* 114 (37):201705899. <https://doi.org/10.1073/pnas.1705899114>.
- Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., et al. 2010. QIIME Allows Analysis of High-Throughput Community Sequencing Data. *Nature Methods* 7 (5):335-36.
- Edgar, R. C. 2010. Search and Clustering Orders of Magnitude Faster than BLAST. *Bioinformatics* 26 (19):2460-61. <https://doi.org/10.1093/bioinformatics/btq461>.
- Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. and Knight, R. 2011. UCHIME Improves Sensitivity and Speed of Chimera Detection. *Bioinformatics* 27 (16):2194-2200. <https://doi.org/10.1093/bioinformatics/btr381>.
- Fettweis, J. M., Brooks, J. P., Serrano, M. G., Sheth, N. U., Girerd, P. H., Edwards, D. J., Strauss, J. F., Jefferson, K. K., and Buck, G. A. 2014. Differences in Vaginal Microbiome in African American Women versus Women of European Ancestry. *Microbiology (United Kingdom)* 160 (2014):2272-82. <https://doi.org/10.1099/mic.0.081034-0>.
- Fettweis, J. M., Serrano, M. G., Sheth, N. U., Mayer, C. M., Glascock, A. L., Brooks, J. P., Jefferson, K. K. and Buck, G. A.. 2012. Species-Level Classification of the Vaginal Microbiome. *BMC Genomics* 13 (Suppl 8):1-9.
- Freitas, A. C., Bocking, A., Hill, J. E. and Money, D. M. 2018. Increased Richness and Diversity of the Vaginal Microbiota and Spontaneous Preterm Birth. *Microbiome* 6 (1). Microbiome:1-15. <https://doi.org/10.1186/s40168-018-0502-8>.
- Gallo, M. F., Macaluso, M., Warner, L., Fleenor, M. E., Hook, E. W., Brill, I. and Weaver, M. A. 2012. Bacterial Vaginosis, Gonorrhoea, and Chlamydial Infection Among Women Attending a Sexually Transmitted Disease Clinic: A Longitudinal Analysis of Possible Causal Links. *Annals of Epidemiology* 22 (3):213-20. <https://doi.org/10.1016/j.annepidem.2011.11.005>.
- Gaujoux, R. 2014. Generating Heatmaps for Nonnegative Matrix Factorization.
- Gosmann, C., Anahar, M. N., Handley, S. A., Walker, B. D., Virgin, H. W., Kwon, D. S., Gosmann, C. et al. 2017. Lactobacillus -Deficient Cervicovaginal Bacterial Communities Are Associated with Increased HIV Acquisition in Young South African Women Report Lactobacillus -Deficient Cervicovaginal Bacterial Communities Are Associated with Increased HIV Acquisition In . *Immunity* 46. Elsevier Inc.:1-9. <https://doi.org/10.1016/j.immuni.2016.12.013>.
- Holst, E., Goffeng, A. R. and Andersch B. 1994. Bacterial Vaginosis and Vaginal Microorganisms in Idiopathic Premature Labor and Association with Pregnancy Outcome. *Journal of Clinical Microbiology* 32 (1):176-86.
- Jacobs, M. V., Sniijders, P. J., Helmerhorst, T. J., Meijer, J. and Walboomers, J. M. 1997. A General Primer GP5+ / GP6+ -Mediated PCR-Enzyme Immunoassay Method for Rapid Detection of 14 High-Risk and 6 Low-Risk Human Papillomavirus Genotypes in Cervical Scrapings. *Journal of Clinical Microbiology* 35 (3):791.
- Kenyon, C.s, Colebunders, R. and Crucitti, T. 2013. The Global Epidemiology of Bacterial Vaginosis: A Systematic Review. *American Journal of Obstetrics and Gynecology* 209 (6). Elsevier Inc:505-23. <https://doi.org/10.1016/j.ajog.2013.05.006>.
- Leitch, H. and Kiss, H. 2007. Asymptomatic Bacterial Vaginosis and Intermediate Flora as Risk Factors for Adverse Pregnancy Outcome. *Best Practice and Research: Clinical Obstetrics and Gynaecology* 21 (3):375-90. <https://doi.org/10.1016/j.bpobgyn.2006.12.005>.
- Lennard, K., Dabee, S., Barnabas, S. L., Havyarimana, E., Blakney A., Jaumdally, S. Z., Botha, G., et al. 2017. Microbial Composition Predicts Genital Tract Inflammation and Persistent Bacterial Vaginosis in Adolescent South African Women. *Infection and Immunity*, no. October:IAI.00410-17. <https://doi.org/10.1128/IAI.00410-17>.
- Lewis, D. A, Müller, E., Steele, L., Sternberg, M., Radebe, F., Lyall, M., Ballard, R.C. and Paz-Bailey, G. 2012. Prevalence and Associations of Genital Ulcer and Urethral Pathogens in Men Presenting with Genital Ulcer Syndrome to Primary Health Care Clinics in South Africa. *Sexually Transmitted Diseases* 39 (11):880-85. <https://doi.org/10.1097/OLQ.0b013e318269cf90>.
- Liaw, A. and Wiener, M. 2002. Classification and Regression by RandomForest. *R News* 2 (3):18-22. <https://doi.org/10.1177/154405910408300516>.
- Mbulawa, Z. Z. A., Van Schalkwyk, C., Nai-chung Hu, Meiring, T. L., Barnabas, S., Dabee, S., Jaspán, H., et al. 2018. High Human Papillomavirus (HPV) Prevalence in South African Adolescents and Young Women Encourages Expanded HPV Vaccination Campaigns. *PLoS One* 13 (1):1-15.
- McClelland, R. S., Lingappa, J. R., Srinivasan, S., Kinuthia, J, John-Stewart, G. C., Jaoko, W., Richardson, B. A., et al. 2018. 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. *The Lancet Infectious Diseases* 3099 (18). Elsevier Ltd:1-11. [https://doi.org/10.1016/S1473-3099\(18\)30058-6](https://doi.org/10.1016/S1473-3099(18)30058-6).
- McMurdie, P. J. and Holmes, S. 2013. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* 8 (4):e61217. <https://doi.org/10.1371/journal.pone.0061217>.
- Nelson, D. B., Hanlon, A. L., Wu, G., Liu, C. and David Fredricks, D. N. 2015. First Trimester Levels of BV-Associated Bacteria and Risk of Miscarriage Among Women Early in Pregnancy. *Maternal and Child Health Journal* 19 (12). Springer US:2682-87. <https://doi.org/10.1007/s10995-015-1790-2>.
- Oksanen, A. J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., Hara, R. B. O., Simpson, G. L., Solymos, P., Stevens, M. H. H., and Wagner, H. 2016. Package 'Vegan'.
- Paulson, J. N., Stine, O. C., Bravo, H. C. and Pop, M. 2013. Differential Abundance Analysis for Microbial Marker-Gene Surveys. *Nature Methods* 10 (12):1200-1202. <https://doi.org/10.1038/nmeth.2658>.
- Pearce, M. M., Hilt, E. E. and Rosenfeld, A. B. 2014. The Female Urinary Microbiome: A Comparison of Women with and without Urgency Urinary Incontinence. *5 (4):1-12*. <https://doi.org/10.1128/mBio.01283-14>. Editor.
- Ravel, J., Gajer, P., Abdo, Z., Schneider, G. M., Koenig, S. S. K., Mcculle, S. L., Ault, K., Peraalta, L. and Forney, L. J. 2010. Vaginal Microbiome of Reproductive-Age Women. *PNAS* 108:4680-87. <https://doi.org/10.1073/pnas.1002611107/-/DCSupplemental>. www.pnas.org/cgi/doi/10.1073/pnas.1002611107.
- Sing, T., Sander, O., Beerenwinkel, N. and Lengauer, T. 2005. ROCr: Visualizing Classifier Performance in R. *Bioinformatics* 21 (20):3940-41. <https://doi.org/10.1093/bioinformatics/bti623>.

- Srinivasan, S., Hoffman, N. G., Morgan, M. T., Matsen, F. A., Fiedler, T. L., Hall, R. W., Ross, F. J., et al. 2012. Bacterial Communities in Women with Bacterial Vaginosis : High Resolution Phylogenetic Analyses Reveal Relationships of Microbiota to Clinical Criteria. *PLoS One* 7 (6). <https://doi.org/10.1371/journal.pone.0037818>.
- Tabatabaei, N., Eren, A. M., Barreiro, L. B., Yotova, V., Dumaine, A., Allard, C. and Fraser, W. D. 2018. Vaginal Microbiome in Early Pregnancy and Subsequent Risk of Spontaneous Preterm Birth: A Case-Control Study. *BJOG: An International Journal of Obstetrics & Gynaecology*, 8-11. <https://doi.org/10.1111/1471-0528.15299>.
- Vinturache, A. E., Gyamfi-Bannerman, C., Hwang, J., Mysorekar, I. U. and Jacobsson, B. 2016. Maternal Microbiome – A Pathway to Preterm Birth. *Seminars in Fetal and Neonatal Medicine* 21 (2):94-99. <https://doi.org/10.1016/j.siny.2016.02.004>.
- Wiesenfeld, H. C., Hillier, S. L., Krohn, M. A., Landers, D. V. and Sweet, R. L. 2003. Bacterial Vaginosis Is a Strong Predictor of Neisseria Gonorrhoeae and Chlamydia Trachomatis Infection. *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America* 36 (5):663-68. <https://doi.org/10.1086/367658>.
- Zhou, X., Brown, C. J., Abdo, Z., Davis, C. C., Hansmann, M. A., Joyce, P., Foster, J. A. and Forney, L. J. 2007. Differences in the Composition of Vaginal Microbial Communities Found in Healthy Caucasian and Black Women. *ISME Journal* 1 (2):121-33. <https://doi.org/10.1038/ismej.2007.12>.