

UCSF

UC San Francisco Previously Published Works

Title

Unique microbial diversity, community composition, and networks among Pacific Islander endocervical and vaginal microbiomes with and without Chlamydia trachomatis infection in Fiji.

Permalink

<https://escholarship.org/uc/item/2h57q74n>

Journal

mBio, 15(1)

Authors

Bommana, Sankhya

Hu, Yi-Juan

Kama, Mike

et al.

Publication Date

2024-01-16

DOI

10.1128/mbio.03063-23

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Unique microbial diversity, community composition, and networks among Pacific Islander endocervical and vaginal microbiomes with and without *Chlamydia trachomatis* infection in Fiji

Sankhya Bommana,¹ Yi-Juan Hu,² Mike Kama,³ Ruohong Wang,¹ Reshma Kodimerla,¹ Kenan Jijakli,⁴ Timothy D. Read,⁴ Deborah Dean^{1,5,6,7,8}

AUTHOR AFFILIATIONS See affiliation list on p. 16.

ABSTRACT Pacific Islanders (PI) are a vulnerable population with a high prevalence of *Chlamydia trachomatis* (*Ct*) sexually transmitted infections and remain underrepresented in research. Here, 258 vaginal and 92 paired endocervical samples from women of diverse ethnicities in Fiji were analyzed using metagenomic shotgun sequencing. Using VALENCIA, a sub-community state type (subCST) classifier, 93 (36.04%) vaginal and 38 (41.30%) endocervical microbiomes did not match reference subCSTs. We, therefore, designed four unique subCSTs, IV-D0, IV-D1, IV-D2, and IV-E—differentially dominated by *Gardnerella vaginalis*, *Prevotella* spp., and *Lactobacillus iners*—to improve microbiome classifications. *Ct*-infected paired endocervical and vaginal microbiomes were significantly more likely to have differential species relative abundance (81.58%) than uninfected pairs (35.84%; adjusted odds ratio, 7.93; 95% confidence interval, 2.93–21.93; $P < 0.0001$). Among paired microbiomes, the endocervix had a significantly higher alpha diversity and independently a higher alpha diversity for high-risk human papilloma virus (HPV) types compared to low-risk and no HPV. The *Ct*-infected endocervical network compared to all others had smaller clusters—indicating lower species abundance and decreased weight of interactions among bacterial groups within clusters—and no interactions with potentially beneficial *Lactobacillus* spp. These data suggest a distinct influence on this network from *Ct*-infected endocervical cells. Furthermore, the high relative abundance of *G. vaginalis* and *Prevotella* spp. may generate polymicrobial biofilms that predispose to and/or promote *Ct* and possibly HPV persistence and pathogenicity. While these data expand on the existing repertoire of endocervical and vaginal microbiomes and fill in knowledge gaps regarding PIs, prospective studies are needed to define causal associations to develop successful interventions.

IMPORTANCE *Chlamydia trachomatis* (*Ct*) is the most common sexually transmitted bacterium globally. Endocervical and vaginal microbiome interactions are rarely examined within the context of *Ct* or among vulnerable populations. We evaluated 258 vaginal and 92 paired endocervical samples from Fijian women using metagenomic shotgun sequencing. Over 37% of the microbiomes could not be classified into sub-community state types (subCSTs). We, therefore, developed subCSTs IV-D0, IV-D1, IV-D2, and IV-E—dominated primarily by *Gardnerella vaginalis*—to improve classification. Among paired microbiomes, the endocervix had a significantly higher alpha diversity and, independently, higher diversity for high-risk human papilloma virus (HPV) genotypes compared to low-risk and no HPV. *Ct*-infected endocervical networks had smaller clusters without interactions with potentially beneficial *Lactobacillus* spp. Overall, these data suggest that *G. vaginalis* may generate polymicrobial biofilms that predispose

Editor Sheryl Justice, The Ohio State University School of Medicine, Columbus, Ohio, USA

Address correspondence to Deborah Dean, deborah.dean@ucsf.edu.

Timothy D. Read and Deborah Dean contributed equally to this article. Their order was determined by seniority.

The authors declare no conflict of interest.

See the funding table on p. 16.

Received 14 November 2023

Accepted 15 November 2023

Published 20 December 2023

Copyright © 2023 Bommana et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

to and/or promote *Ct* and possibly HPV persistence and pathogenicity. Our findings expand on the existing repertoire of endocervical and vaginal microbiomes and fill in knowledge gaps regarding Pacific Islanders.

KEYWORDS Pacific Islanders, *Chlamydia trachomatis*, metagenomic shotgun sequencing, pathogenesis, microbiome networks

Chlamydia trachomatis (*Ct*) is the most common bacterial cause of sexually transmitted infections (STIs) worldwide. The World Health Organization estimates that over 130 million global cases occur each year (1, 2). The distribution of these infections is greatest in Africa and the Western Pacific Region (WPR) where there are about 61 million cases (1). The prevalence of *Ct* STIs ranges from 30% to 44% among teens and young adults in the Pacific Island Countries and Territories (PICT) of the WPR (3–6). Here, the combination of asymptomatic, untreated infection and syndromic management—treatment based on signs and symptoms as a proxy for infection—can lead to life-threatening ectopic pregnancy, infertility, and an increased risk of cervical cancer and HIV (7–10). Indeed, PICTs have the third highest prevalence of infertility in the world today (11).

The vaginal microbiome is influenced by resident microbes but also those associated with bacterial vaginosis (BV) and STIs, including *Ct*, *Neisseria gonorrhoeae* (*Ng*), *Trichomonas vaginalis* (*Tv*), *Mycoplasma genitalium* (*Mg*), HIV, and human papilloma virus (HPV) (12–14). Community state type I (CST-I), CST-II, CST-III, CST IV, and CST-V have been developed to classify vaginal microbiomes (15). CSTs I–III and V are dominated by *Lactobacillus crispatus*, *Lactobacillus gasseri*, *Lactobacillus iners*, or *Lactobacillus jensenni* that maintain vaginal homeostasis by secreting lactic acid, bacteriocins, and antimicrobial compounds, but not hydrogen peroxide, to inhibit anaerobic bacterial growth and protect against BV and STIs (16). CST-IV is deficient in *Lactobacillus* spp. and rich in anaerobic bacteria such as *Prevotella*, *Dialister*, *Atopobium*, *Gardnerella*, *Megasphaera*, *Peptoniphilus*, *Sneathia*, *Eggerthella*, *Aerococcus*, *Fingoldia*, and *Mobiluncus* genera. A diverse polymicrobial microbiome can increase the risk of *Ct* infection and other STIs (17, 18). Vaginal microbiome data from European, Asian, African, and Hispanic populations have distinct CSTs among these racial, ethnic, and geographic groups regardless of STI status (19–23). However, data from other ethnic and racial populations such as indigenous Australian, Native American, and Pacific Islander (PI) populations are lacking.

To provide insights into the epidemiology, structure, and function of the vaginal microbiome, the vaginal non-redundant gene catalog (VIRGO) (24) and sub-community state type (subCST) classifier VALENCIA (25) were developed. VIRGO was constructed from 264 vaginal metagenomes and 308 draft genomes of urogenital bacterial isolates from North American women and tested using 91 vaginal metagenomes from North American, African, and Chinese women. These data provided over 95% microbiome coverage applicable to these populations. VALENCIA classifies microbiomes into 13 subCSTs based on their similarity to a set of reference centroids built using a set of 13,160 16S rRNA sequencing-based taxonomic profiles from 1,975 North American, Black, White, Hispanic, and Asian women. It was validated using data sets of reproductive age eastern and southern African women and adolescent girls and ethnically and geographically diverse samples of postmenopausal women, although the number of women in each data set was not available.

We recently implemented VIRGO and VALENCIA on paired endocervical and vaginal metagenomes of PIs in Fiji as a pilot study, which involved age matching five *Ct*-infected women to those without *Ct* (26). VALENCIA could not accurately classify 4 of 10 vaginal and 4 of 10 endocervical microbiomes. Furthermore, because VIRGO does not contain viral, fungal, or protozoan pathogen sequences, such as HPV, *Candida*, and *T. vaginalis*, respectively, analyses for these pathogens were limited. We found that the microbiomes were largely dominated by species other than *Lactobacillus* regardless of STI status, and while endocervical and vaginal microbiomes were closely related, they were distinct

in taxonomic composition, subCSTs, and metabolic function. However, our study was relatively small, and only ethnic iTaukei PIs were studied.

There is an obvious need to expand metagenome studies of endocervical and vaginal microbiomes from diverse ethnic and global populations to better understand microbial diversity, compositions, and networks that drive pathogenesis. Building on these knowledge gaps, we expanded our data set to 258 vaginal and 92 paired endocervical samples using metagenomic shotgun sequencing (MSS) to characterize vaginal and endocervical microbial relative abundance, composition, and network interactions for the diversity of ethnicities among women residing in Fiji. We also explored differences between paired microbiomes with and without the presence of *Ct*, other STIs, *Candida*, and BV.

RESULTS

Characteristics of study subjects, infection prevalence, and quality statistics of endocervical and vaginal MSS

Women 18 to 40 years of age attending Fiji Ministry of Health and Medical Services (MoHMS) health centers were enrolled in the parent study as described (3). The ethnicities in Fiji include ~57% iTaukei and ~37% Indo-Fijians (27, 28). Other ethnic groups include Chinese, European, mixed race, other PIs—IKiribati, Banabans, Tuvaluans, Tongans, Samoans, and Wallisians—and expatriates of various nationalities (28).

From the parent study, 258 vaginal samples, of which 92 had paired endocervical samples, were available. Paired samples were collected at the same time point from the same woman. Data on participants' age, ethnicity, BV status, and infection status for *Ct*, *Ng*, *Tv*, HPV, and *Candida* were provided (Table S1). Age above 24 years and PI ethnicities were significantly associated with *Ct* infection (Table 1).

Results for BV using Amsel criteria (29) and for *Tv* and *Candida*—detected by wet prep on vaginal samples (provided by the parent study) and MetaPhlan v3.0 (this study) for both vagina and endocervix—are shown in Table S1. Using multivariate logistic regression, there was a significant association with *Ct* for *Ng* in the endocervix and vagina; for high-risk (hr) HPV, *Tv*, and BV in the vagina; and for *Mg* in the endocervix, although the numbers were small (Table 2).

The 258 vaginal and 92 endocervical metagenomes yielded a total of ~26 billion and 9.5 billion raw reads of which 22.8 billion (87.36%) and 8.76 billion (91.4%) were identified as human contamination, respectively (Table S2).

Based on VIRGO, *Mg* reads were detected in 9 (3.5%) of 258 vaginal and 7 (7.6%) of 92 endocervical metagenomes (see Materials and Methods; Table S3). HPVViewer (30)

TABLE 1 Bivariate analysis for age group and ethnicity associated with *C. trachomatis* infection for study participants^c

Variables	N ^a	Endocervix (n = 46)				N ^a	Vagina (n = 111)			
		Ct	No Ct	P-value	OR (95% CI)		Ct	No Ct	P-value ^d	OR (95% CI)
Age (years)	46	22	24			111	60	51		
18–24	15	4	11	–	ref	37	14	23	–	ref
25–40	31	18	13	0.0520	3.81 (1.05, 16.3)	74	46	28	0.0168	2.70 (1.21, 6.21)
Ethnicity ^b	46	22	24			111	60	51		
Indo-Fijian	11	1	10	–	ref	28	4	24	–	ref
iTaukei	26	16	10	0.0136	16.0 (2.50, 317)	64	45	19	1.17e–05	14.2 (4.74, 53.5)
Other PI	4	3	1	0.0292	30.0 (1.98, 1283)	10	7	3	0.0026	14.0 (2.76, 91.8)
Other	4	1	3	0.4402	3.33 (0.109, 104)	8	3	5	0.158462	3.60 (0.568, 22.2)
Missing	1	1	0	–	–	1	1	0	–	–

^aThe total number of women in the study was 111; 46 endocervical samples were obtained from these women at the same time point as the vaginal samples, which we consider paired samples; paired endocervical and vaginal samples were also available at a second time point for the same 46 women, for a total of 350 samples (92 endocervical and vaginal samples paired from one time point + 92 paired from a second time point + 166 unpaired vaginal samples from the remaining women who did not have paired endocervical and vaginal samples).

^bOther PI ethnicity: IKiribati, Banabans, Tuvaluans, Tongans, Samoans, and Wallisians; other ethnicity: Chinese, European, and mixed race.

^cCt, *C. trachomatis* infection; ref, reference for calculating P-value and OR, OR, odds ratio; CI, confidence interval; PI, Pacific Islander.

^dP-values that have been highlighted in bold are those that are statistically significant for that variable.

TABLE 2 Bivariate analysis and multivariate model for sexually transmitted infections, including HPV, *M. genitalium*, *N. gonorrhoeae*, and *T. vaginalis*, in addition to BV and their association with *Ct* infection for endocervical and vaginal samples^a

Variables	Endocervix (n = 92)						Vagina (n = 258)						
	N ^b	Ct	No Ct	P-value	OR (95% CI)	P-value	N ^b	Ct	No Ct	P-value	OR (95% CI)	P-value ^c	AOR (95% CI)
HPV	92	40	52				258	99	159				
No infection	38	13	25	-	ref	-	101	33	68	-	ref	-	ref
lrHPV	29	15	14	0.152	2.06 (0.771, 5.64)	-	116	45	71	0.349	1.31 (0.748, 2.29)	-	-
hrHPV	25	12	13	0.276	1.78 (0.633, 5.05)	-	41	21	20	0.0410	2.16 (1.03, 4.57)	0.0329	2.45 (1.08, 5.66)
Missing	0	0	0	-	-	-	0	0	0	-	-	-	-
<i>M. genitalium</i>	92	40	52				258	99	159				
Absent	85	34	51	-	ref	-	249	95	154	-	ref	-	-
Present	7	6	1	0.0463	9.00 (1.45, 174)	0.0572	9	4	5	0.7040	1.30 (0.314, 5.02)	-	-
Missing	0	0	0	-	-	-	0	0	0	-	-	-	-
<i>N. gonorrhoeae</i>	92	40	52				258	99	159				
Absent	53	15	38	-	ref	-	197	67	130	-	ref	-	ref
Present	39	25	14	0.0008	4.52 (1.90, 11.3)	0.0132	61	32	29	0.0104	2.14 (1.20, 3.85)	0.0410	1.94 (1.03, 3.68)
Missing	0	0	0	-	-	-	0	0	0	-	-	-	-
<i>T. vaginalis</i>	92	40	52				258	99	159				
Absent	86	37	49	-	ref	-	232	84	148	-	ref	-	ref
Present	6	3	3	0.740	1.32 (0.234, 7.51)	-	26	15	11	0.0368	2.40 (1.06, 5.60)	0.282	1.60 (0.685, 3.85)
Missing	0	0	0	-	-	-	0	0	0	-	-	-	-
Bacterial vaginosis													
Absent	-	-	-	-	-	-	258	99	159	-	-	-	-
Present	-	-	-	-	-	-	174	58	116	-	ref	-	ref
Missing	-	-	-	-	-	-	39	24	15	0.00150	3.20 (1.58, 6.68)	0.0153	2.64 (1.22, 5.93)

^aThe total number of women in the study was 111; 46 endocervical samples were obtained from these women at the same time point as 46 vaginal samples; paired endocervical and vaginal samples were available at a second time point for the same 46 women, which we consider paired samples, for a total of 350 samples (92 endocervical and vaginal samples paired from one time point + 92 endocervical and vaginal samples paired from a second time point + 166 unpaired vaginal samples from the remaining women who did not have paired endocervical and vaginal samples) for HPV, *M. genitalium*, *T. vaginalis*, and *N. gonorrhoeae* associations. The paired and unpaired vaginal samples were used in the BV analyses. Note that BV cannot be diagnosed in the endocervix.

^bCt, *C. trachomatis* infection; ref, reference for calculating P-value and OR; OR, odds ratio; CI, confidence interval; PI, Pacific Islander; AOR, adjusted OR; HPV, human papilloma virus; hrHPV, high-risk HPV; lrHPV, low-risk HPV.

^cP-values that have been highlighted in bold are those that are statistically significant for that variable.

(Table S3) identified HPV in 157 (60.9%) of 258 vaginal and 54 (58.7%) of 92 endocervical samples. hrHPV types were found in 92 (35.66%) and 38 (41.30%) samples, respectively; not all paired samples had the same HPV type(s). The most common low-risk (lr) HPV types were 57 ($n = 45$; 25 vaginal, 20 endocervical) followed by HPV 62 ($n = 42$; 32 vaginal, 10 endocervical) and 90 ($n = 40$; 26 vaginal, 14 endocervical). The most common hrHPV types were 52 ($n = 35$; 24 vaginal, 11 endocervical) and 39 ($n = 30$; 17 vaginal, 13 endocervical). hrHPV types 16 and 18—responsible for most HPV-related cancers—were found in 24 (9.30%) of 258 vaginal and 11 (11.96%) of 92 endocervical samples and 18 (6.98%) vaginal and 21 (22.83%) endocervical samples, respectively. For *Tv*, six samples were positive by MSS alone in the endocervix and 22 in the vagina—10 by wet prep, 4 by wet prep and MSS, and 8 by MSS alone (Table S1). Few samples were positive for *Candida* in the vagina, and none, in the endocervix (Table S3).

Expansion of existing CST classification system to include new subCSTs unique to endocervical and vaginal microbiomes of PIs

The microbiomes resolved into five CSTs (15) and 13 subCSTs based on VALENCIA, a nearest-centroid algorithm (25). The algorithm assigns a similarity score of 0.0 (no shared taxa) to 1.0 (all taxa shared at similar relative abundance) of the microbial profile of the sample to 1 of 13 reference centroids. In our previous (26) and current studies, several samples had low scores (<0.1–0.25) but also scores above 0.4, ranging from 0.019 to 0.808, where there was a discrepancy between microbiome species abundance and archetypal profile of the closest centroid. We, therefore, manually compared microbial relative abundance profiles of each microbiome with that of the assigned subCST, re-classifying samples regardless of their similarity score.

The relative abundance profiles for 93 (36.04%) of 258 vaginal and 38 (41.3%) of 92 endocervical microbiomes did not match their assigned subCST (Table S4). These profiles indicated that they could be grouped into four new subCST categories within CST IV: IV-D0, dominated by *G. vaginalis* ($n = 51$); IV-D1, high to moderate relative abundance of *G. vaginalis* with *Prevotella* spp. ($n = 18$); IV-D2, high to moderate relative abundance of *G. vaginalis* with *L. iners* ($n = 18$); and IV-E, moderate *Prevotella* spp. with other species ($n = 25$) (Fig. 1, orange highlights). The new subCSTs were constructed by averaging the relative abundance of microbiota for the microbiomes ($n = 131$) that fit new subCST designation (Table S5).

Microbiomes classified as subCST IV-D0, IV-D1, and IV-D2 were originally assigned subCST IV-B by VALENCIA (Fig. 2). While subCST IV-B had a high to moderate relative abundance of *G. vaginalis* (33.75%) and *A. vaginae* (12.16%), the new subCSTs were different in composition. subCST IV-D0 had a higher relative abundance of *G. vaginalis* at 85.15% and lower abundance of *A. vaginae* at 2.46%. subCST IV-D1 had a higher relative abundance of both *G. vaginalis* (52.72%) and *Prevotella* spp. (20%), and subCST IV-D2 had a higher relative abundance of *G. vaginalis* (61.84%) and *L. iners* (23.2%) compared to CST IV-B (Fig. 2). Compared to more abundant bacteria and the diversity of species shown in the figures, individual species that were in relatively low abundance such as *Ct* are not visible.

Samples that were originally assigned as CST III-B, IV-A, IV-B, IV-C0, IV-C1, and IV-C4 by VALENCIA were re-classified as subCST IV-E (Fig. 3). CST IV-E had ~50% relative abundance of *Prevotella* spp. and 12.29% for *G. vaginalis*.

Of the 95 study participants whose microbiomes did not match the assigned subCSTs, iTaukei and other PI ethnicities (68/74 [91.89%]) were significantly more likely to not match compared to Indo-Fijian and other ethnicities (27/36 [75.00%]; adjusted odds ratio [AOR]: 3.78; 95% confidence interval [CI]: 1.23–11.64; $P = 0.0205$) (Table S4). Age group was not associated with a lack of matching to the assigned subCSTs.

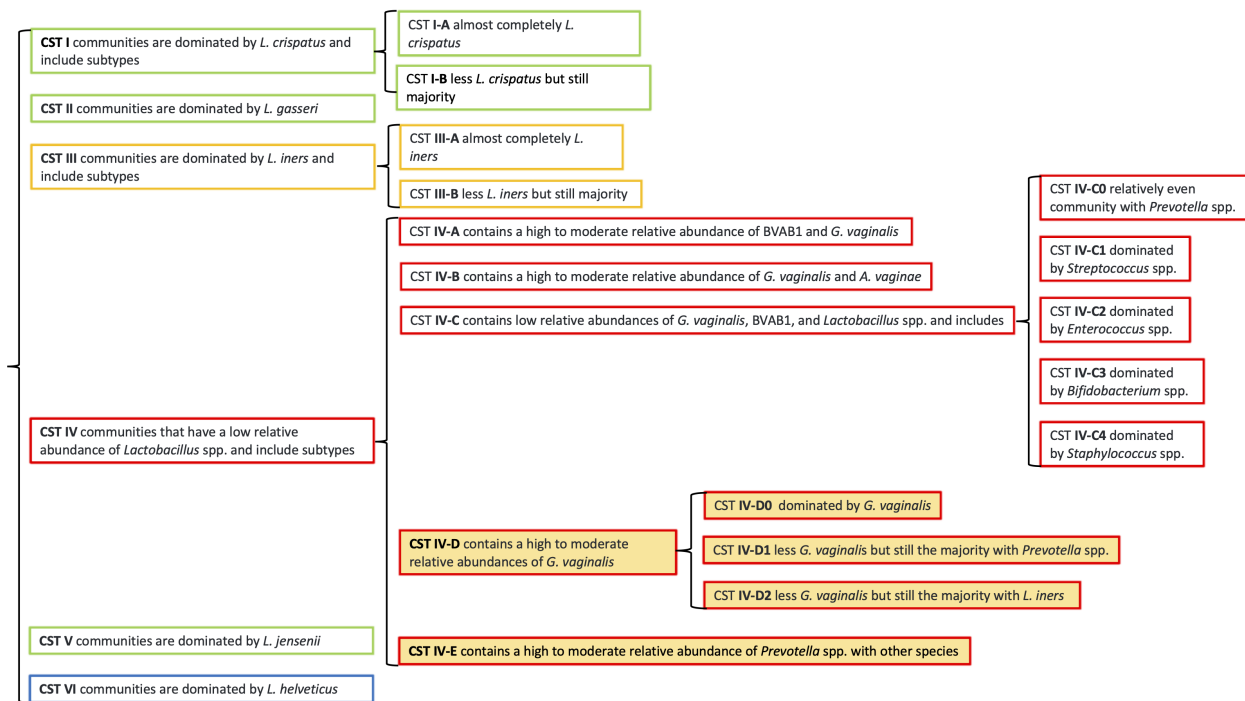


FIG 1 Classification of the endocervical and vaginal CSTs by VALENCIA with inclusion of four new subCSTs (IV-D0, IV-D1, IV-D2, and IV-E) identified among PI residing in Fiji. The new subCSTs (denoted in orange background) were constructed by averaging the relative abundance of all microbiota for the samples that fit the new subCST designation ($n = 112$; see Materials and Methods). A new CST VI suggested by Li et al. (31) is also included in the figure (blue outlined box).

Within-host paired endocervical and vaginal species composition differs for women with and without Ct infection

Figure 4 shows the relative abundance of taxa in paired vaginal and endocervical microbiomes based on Ct status. For Ct-uninfected endocervical and vaginal pairs, 19 (35.84%) of 53 pairs had differential species relative abundance between the anatomic sites, reflecting a difference in subCST (boxed in black, Fig. 4). The remaining 33 (63.46%) pairs had relatively similar species relative abundance. The presence of *L. crispatus*-dominated microbiomes in 5 (15.15%) of 33 similar pairs was unique to the uninfected group (boxed in green, Fig. 4). The majority of paired microbiomes in the uninfected group were dominated by *G. vaginalis* (23/52; 44.23%), *L. iners* (16/52; 30.76%), and *Prevotella* spp. (4/52; 7.69%).

Ct-infected paired microbiomes were statistically more likely to have differential species relative abundance (31/38 [81.57%]) between the two sites compared to Ct-uninfected pairs (19/53 [35.84%]; AOR: 7.93; 95% CI: 2.93–21.93; $P < 0.0001$) (boxed in black, Fig. 4). There was a higher relative abundance of *G. vaginalis* (18/38; 47.36%) compared to *L. iners* (7/38; 18.42%). Some profiles were dominated by *Prevotella* spp. (3/38; 7.89%) and *E. coli* (2/38; 5.26%). Interestingly, 6 (15.79%) of 38 pairs (subjects 56, 57, 70, 86, 87, and 90) had a moderate to high relative abundance of Ct in the endocervical but not vaginal microbiomes.

Higher microbial diversity and species evenness were found in endocervical compared to vaginal microbiomes

Alpha diversity was measured by Chao1 and Shannon index as shown in Fig. 5 using the linear decomposition model (LDM) (32). For vaginal microbiomes, iTaukei PIs had a significantly higher alpha diversity compared to Indo-Fijian and other ethnicities controlling for age and Ct (Chao1 $P = 0.0255$, Shannon $P = 0.0322$). Alpha diversity also varied significantly among subCSTs with higher diversity in subCST-IV (Chao1 $P = 2e-04$,

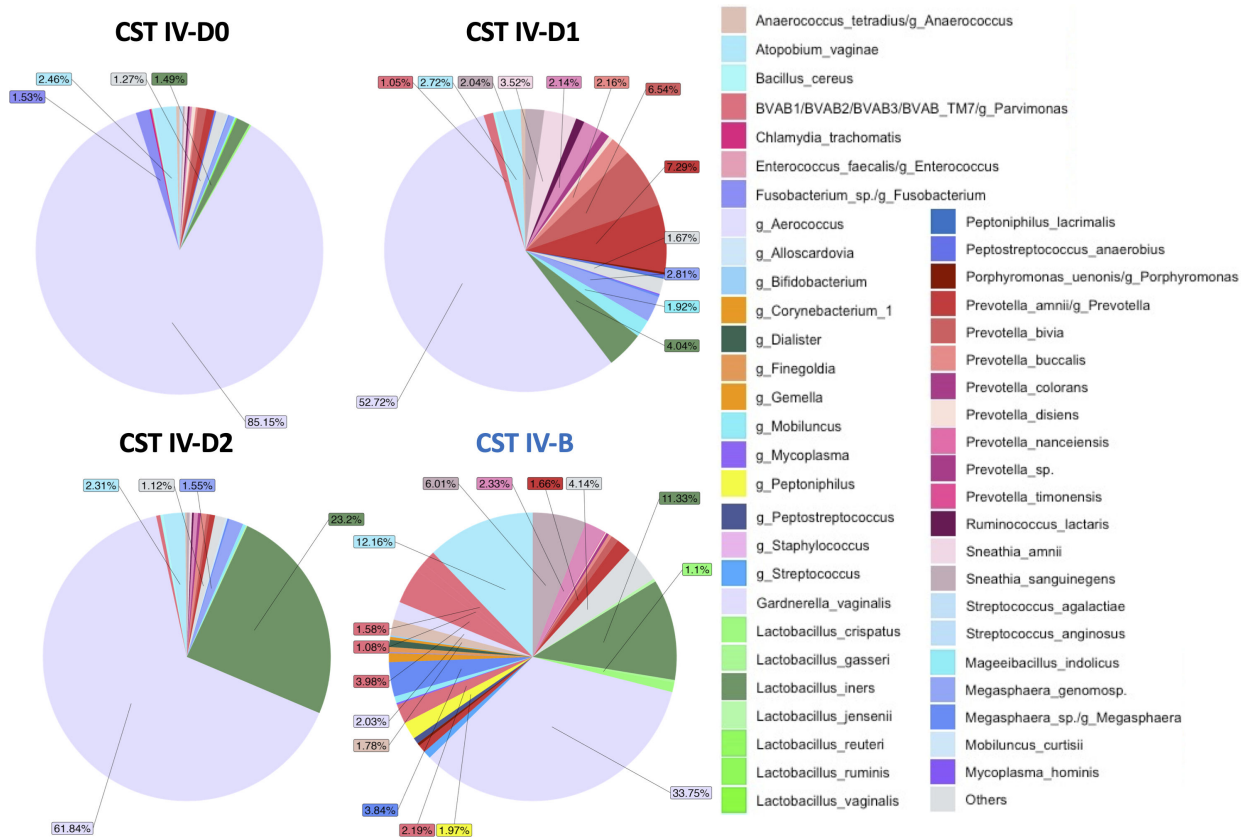


FIG 2 Pie chart representation of unique subCSTs IV-D0, IV-D1, and IV-D2 developed for endocervical and vaginal microbiomes from the Fiji metagenome database output from VIRGO. The microbiomes were originally assigned to subCST IV-B (right lower pie graph) using the VALENCIA classifier but were manually re-classified to the new subCSTs based on their relative abundance profiles that were not a good match with IV-B (see Materials and Methods). The new subCSTs were defined as IV-D0, dominated by *G. vaginalis*; IV-D1, less *G. vaginalis* but still the majority with *Prevotella* spp.; and IV-D2, less *G. vaginalis* but still the majority with *L. iners*. Relative abundance data were used to plot the pie charts of the new CSTs (Table S5).

Shannon $P = 2e-04$). Vaginal microbiomes with BV had higher diversity than those without BV controlling for *Ct* (Chao1 $P = 2e-04$, Shannon $P = 0.0014$), as did *Ct*-infected versus uninfected microbiomes (Chao1 $P = 0.0072$, Shannon $P = 0.0022$) controlling for ethnicity and age.

Endocervical microbiomes had a significantly higher alpha diversity for all subCST-IV classifications compared to other subCSTs (Chao $P = 0.0168$, Shannon $P = 2e-04$) (Fig. 5B). A higher alpha diversity was also found for hrHPV types compared to lrHPV types and no HPV infection (Chao $P = 0.007$, Shannon $P = 0.0226$).

Alpha diversity between paired vaginal and endocervical microbiomes differed significantly based on Chao1 ($P = 0.0024$) (Fig. 5C) that was influenced by the presence of *Ct* (Chao1: $P = 0.024$) (Fig. 5D) and less so by the absence of *Ct* (Chao1: $P = 0.043$) (Fig. 5E).

Divergence in microbial composition between samples was measured by beta diversity using Bray-Curtis and Jaccard metric. Principal component analysis (PCoA) showed significant segregation of vaginal microbiomes based on ethnic groups ($P = 0.039$ and $P = 0.0044$, respectively), subCST categories ($P = 2e-04$ and $P = 2e-04$), BV ($P = 8e-04$ and $P = 2e-04$), *Candida* ($P = 0.0068$ and $P = 0.04$), and *Ct* status ($P = 0.012$ and $P = 0.04$) (Fig. 6). In contrast, endocervical microbiomes were significantly segregated based on subCST-IV ($P = 2e-04$ and $P = 2e-04$) and *Ct* status (Jaccard only, $P = 0.024$) (Fig. 6).

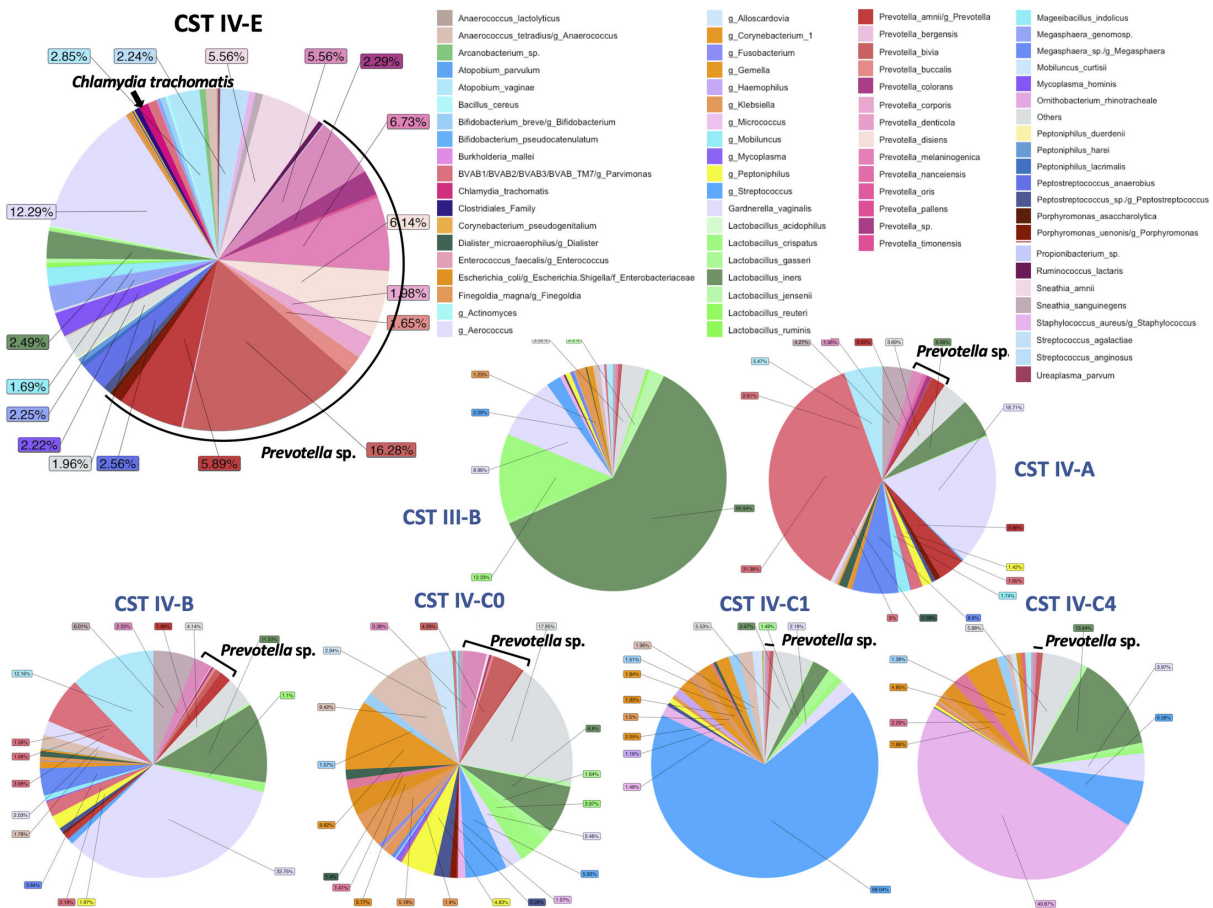


FIG 3 Pie chart representation of the unique subCST IV-E (upper left) developed for endocervical and vaginal microbiomes from the Fiji metagenome database output from VIRGO. The microbiomes were originally assigned to subCSTs III-B, IV-A, IV-B, IV-C0, IV-C1, and IV-C4 using the VALENCIA classifier (denoted by smaller pie charts and subCSTs in blue font) but were manually re-classified based on their relative abundance profile that was not a good match to the other subCSTs (see Materials and Methods). CST IV-E contains a high to moderate relative abundance of *Prevotella* spp. with other species. Relative abundance data were used to plot the subCST IV-E pie chart (Table S5).

Ethnicity, BV, Ct, and anatomic site-specific associated species based on differential abundance and presence or absence of bacterial species

Based on LDM differential abundance analysis (32) of vaginal microbiomes, iTaukei and other PI ethnicities had a significantly higher abundance of pathogenic bacteria *S. amni*, *Leptotrichia goodfellowii*, and *Parvimonas* sp. compared to Indo-Fijian and other ethnic groups; *D. microaerophilus*, *Prevotella* sp., *Prevotella salivae*, and *Bifidobacterium animalis* were associated with iTaukei ethnicity only (Fig. S1). The high abundance of vaginal *L. crispatus* and *Actinomyces europaeus* was strongly associated with the “other” ethnic group. *Ruminococcus lactaris*, *M. indolicus*, *S. sanguinegens*, *Actinobaculum massiliae*, *Megashaera* sp., and *P. histicola* were enriched in endocervical microbiomes and significantly associated with iTaukei and/or other PI ethnicities, while *L. aciophilus*, *B. dentium*, and BVAB1 were associated with the other ethnicity (Fig. S2).

The 34 species significantly differentially abundant in BV-positive compared to BV-negative vaginal microbiomes included *Dialister*, *Leptotrichia*, *Mageeibacillus*, *Megashaera*, *Mobiluncus*, *Mycoplasma*, *Parvimonas*, *Prevotella*, *Porphyromonas*, and *Sneathia* (Fig. S3). Based on presence or absence, 33 species were significantly higher in BV-positive microbiomes, overlapping with differential abundance data except for *L. crispatus* and *L. jensenii*, which were significantly higher in BV-negative microbiomes (Fig. S3).

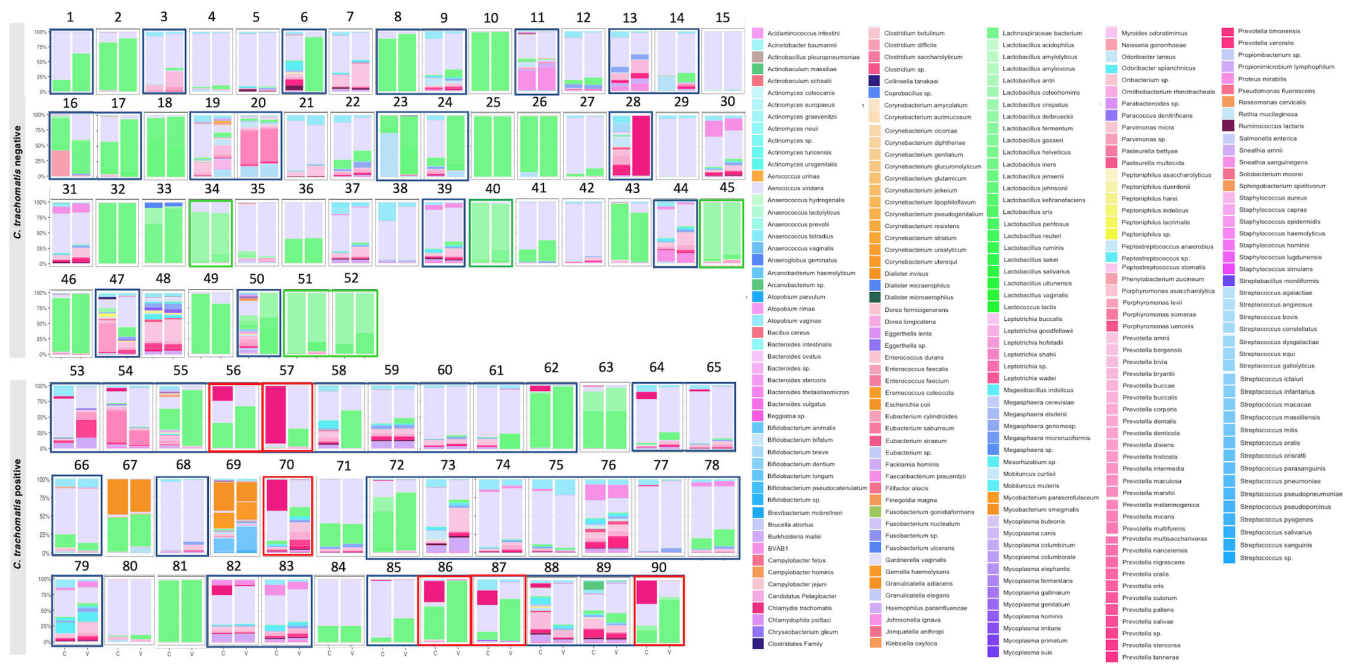


FIG 4 Relative abundance bar plots for *Ct*-infected and for uninfected paired endocervical and vaginal microbiomes. Only 90 pairs are shown due to lack of relative abundance data for two endocervical samples. Profiles in black boxes indicate paired endocervical and vaginal microbiomes with different microbial profiles. These microbiomes were dissimilar in species relative abundance for 19 (36.53%) of 52 *C. trachomatis* uninfected pairs and 31 (81.6%) of 38 *C. trachomatis* infected pairs. Green boxes show *L. crispatus*-dominant paired endocervical and vaginal microbiomes with similar profiles, which were only found in *C. trachomatis*-uninfected pairs (five [9.6%] of 52 pairs). Red boxes show moderate to high relative abundance for *C. trachomatis* endocervical microbiomes with dissimilar profiles, which were only found in *C. trachomatis*-infected pairs (six [15.79%] of 38 pairs). C, endocervix; V, vagina.

Differentially abundant species significantly associated with *Ct*-infected vaginal microbiomes included *Prevotella* sp., *Tannerella forsythensis*, *L. goodfellowii*, *Lactobacillus pentosus*, *S. amnii*, and *Megasphaera* sp. (Fig. S4). For presence-absence data, *L. crispatus* was significantly higher in *Ct*-uninfected microbiomes, while *Sneathia sanguinegens*, *S. amnii*, *Tannerella forsythensis*, *Prevotella* sp., *Lactobacillus pentosus*, *Megasphaera micronuciformis*, *Mageeibacillus indolicus*, *Prevotella pallens*, *Prevotella amnii*, and *Bacteroides* sp. were significantly higher in *Ct*-infected microbiomes. No species were differentially abundant in *Ct*-infected endocervical microbiomes (data not shown.)

For *Ct*-infected paired microbiomes, 56 species were significantly differentially abundant in endocervical compared to vaginal microbiomes, and 12 species were differentially abundant in vaginal compared to endocervical biomes (Fig. S5). Based on species presence or absence, 71 species were significantly higher in endocervical versus vaginal biomes, and 19, in vaginal compared to endocervical biomes (Fig. S6).

For *Ct*-uninfected paired microbiomes, 26 species were significantly differentially abundant in vaginal compared to endocervical microbiomes, and six were differentially abundant in the endocervix (Fig. S7). For presence or absence data, 28 species were significantly higher in vaginal compared to endocervical microbiomes, and four, in the endocervix compared to the vagina (Fig. S8).

When considering anatomic site-specific differences regardless of *Ct* status, based on differential abundance, 67 species were significantly abundant in endocervical compared to paired vaginal microbiomes, and 18 species were significantly abundant in vaginal compared to endocervical microbiomes (Fig. S9). Based on presence or absence, 63 species were significantly higher in endocervical compared to paired vaginal microbiomes, and 12 species, in the vagina compared to the endocervix (Fig. S10).

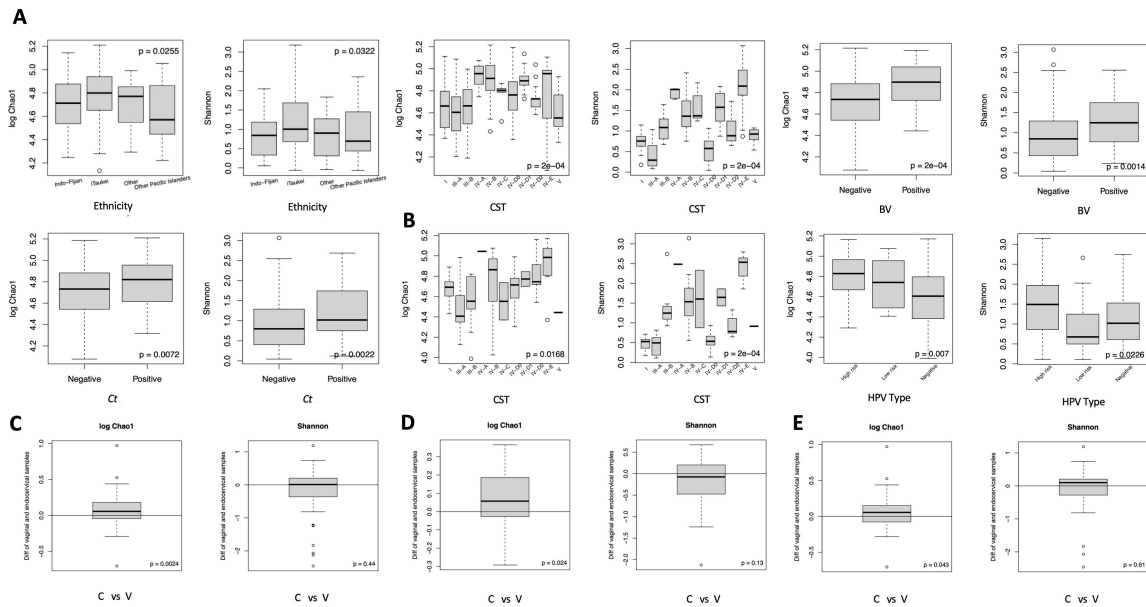


FIG 5 Alpha diversity metrics for the vaginal and endocervical microbiomes with categorical variables and estimated *P*-values based on the linear decomposition model (LDM; see Materials and Methods). (A) Shown in the four sets of bar plots for the vaginal microbiomes are the significantly higher alpha diversity (i.e., alpha diversity calculated by log Chao1 and Shannon) that was found when comparing groups within the categorical variables of ethnicity (i.e., Indo-Fijian, iTaukei, other PIs, other), CSTs, including subCSTs, BV status (i.e., present compared to not present), and *Ct* status (i.e., *Ct* infection compared to no infection). Groups within categorical variables that were not found to have a significantly higher alpha diversity are not shown. (B) Shown in the two sets of bar plots for the endocervical microbiomes are the significantly higher alpha diversity found when comparing groups within the categorical variables of CSTs (i.e., comparisons across all CSTs) and HPV types (i.e., comparing high-risk HPV, low-risk HPV, and no HPV; see Tables S1 and S3 for HPV risk types). HPV, human papilloma virus; hrHPV, high-risk HPV. (C) Shown in the set of bar plots for paired endocervical and vaginal microbiomes are the significantly higher alpha diversity found overall regardless of *Ct* infection (i.e., endocervical vs vaginal) and (D) independently for *C. trachomatis* infection (i.e., endocervical vs vaginal) and (E) for no *C. trachomatis* infection (i.e., endocervical vs vaginal). Log Chao1 and Shannon indices were both used to measure alpha diversity (see Materials and Methods). C, endocervix; V, vagina.

Paired endocervical and vaginal microbiomes show stable networks with anatomic site variation depending on *Ct* status

To understand bacterial community networks in vaginal and endocervical microbiomes, we employed Pearson’s correlation coefficient with a significance cutoff of *P* < 0.05 between different bacterial taxa in the respective microbiomes. Microbiota community clusters (i.e., nodes) within the networks were identified; modality was calculated to measure the degree of stability within densely connected communities in the network, which was greater than 0.8 for all networks, indicating high stability (33).

Figure 7 shows numbered groups that represent a single species or related species; species within groups and clusters are listed in Table S6. Nodes are comprised of two or more groups, and their size denotes abundance of those species in the community, imbuing a greater weight of interactions for a node within the network (Fig. 7). The edges represent mutualistic relationships within species of the community (Fig. 7). Nine clusters were identified among *Ct*-infected and *Ct*-uninfected paired endocervical and vaginal microbiomes (Fig. 7). Cluster 1 consists of gut bacteria from groups 27 and 26. Cluster 1 areas were 35.60, 39.35, 36.94, and 37.04 for *Ct*-infected endocervical and vaginal and *Ct*-uninfected endocervical and vaginal microbiomes, respectively. Cluster 2 consists of group 1 gut bacteria and group 2 BV-associated bacteria. Cluster 3 comprises groups 1, 2, and 5; cluster 4, groups 1, 2, and 15; cluster 5, groups 4, 6, and 20; cluster 6, groups 2, 3, and 27; cluster 7, groups 19 and 20; cluster 8, groups 6 and 10; and cluster 9, groups 8 and 10.

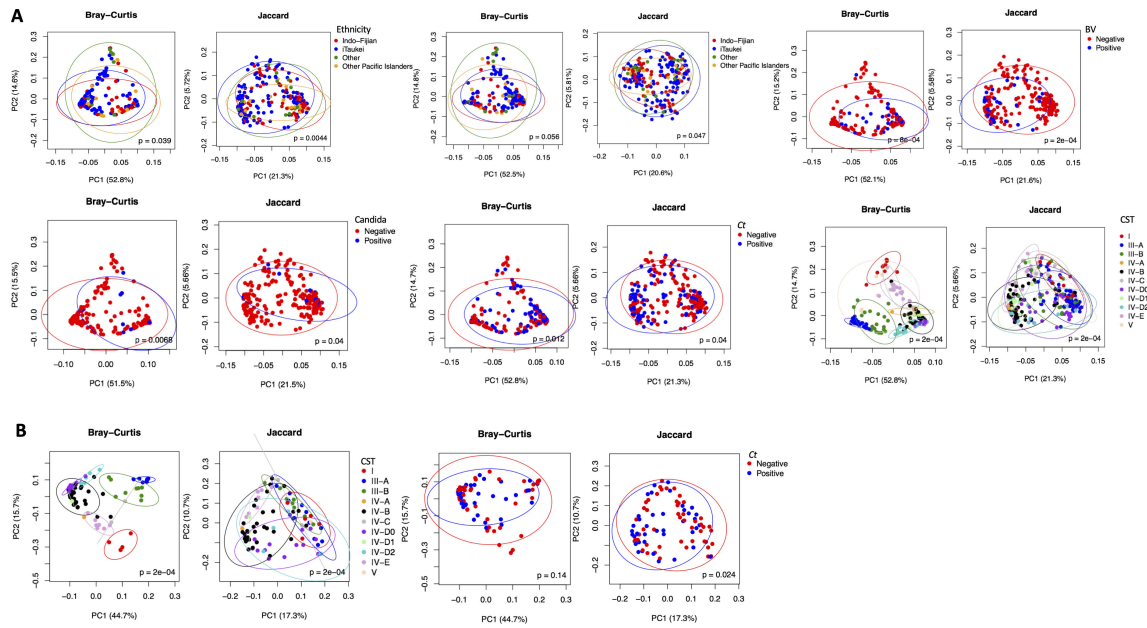


FIG 6 Beta diversity metrics for the vaginal and endocervical microbiomes with categorical variables and *P*-values estimated based on permutational multivariate analysis of variance (PERMANOVA). (A) Significant associations with the vaginal microbiomes are shown for ethnicity with and without controlling for *Ct* infection and age, CSTs, including subCSTs, BV, *Candida*, and *Ct* status. (B) Significant associations are shown for CSTs, including subCSTs, and *C. trachomatis* status with the endocervical microbiomes. Bray-Curtis and Jaccard indices were used to measure beta diversity (see Materials and Methods).

DISCUSSION

Major advances in our ability to analyze the microbial composition of vaginal microbiomes have revolutionized our understanding of the role they play in both health and disease. However, data on vulnerable populations from diverse ethnic and global populations are largely missing. Moreover, the relationship between endocervical and vaginal microbiomes has barely been examined. The former is of particular importance because the endocervix—not the vagina—is a primary site of infection for sexually transmitted *Ct*, *Ng*, and HPV. Here, we expanded on our pilot study (26) by evaluating 258 vaginal and 92 paired endocervical microbiomes for women of all ethnicities residing in Fiji. We characterized the microbial relative abundance, community composition, and networks of these microbiomes and examined the unique relationship between paired endocervical and vaginal microbiomes.

We previously found that women of iTaukei ethnicity were significantly more likely to have *Ct* than Indo-Fijians (3), consistent with the present study. Furthermore, there was a significantly higher microbial diversity for iTaukei ethnicity compared to Indo-Fijian, other PIs, and other ethnicities. Many BV-associated species such as *B. animalis*, *D. microaerophilus*, *Prevotella* sp., *P. salivae*, *S. amnii*, and *Parvimonas* sp. were also significantly associated with this ethnicity. Moreover, iTaukei and other PIs had a significantly diverse community composition of anaerobes compared to a predominance of *Lactobacillus* spp. such as *L. crispatus* for Indo-Fijian and other ethnicities. Such ethnicity- or race-based microbial clustering in microbiomes have also been reported for ethnic/racial subgroups of African descent and European/Caucasian background (21, 22).

Vaginal microbiomes with BV were significantly associated with *Ct* infection and also had higher microbial diversity with differences in community composition compared to those without BV, as has been reported (34, 35). Not surprisingly, BV-associated bacteria including *Megasphaera* spp., *Sneathia* spp., *G. vaginalis*, and *Prevotella* spp. were present at a significantly higher relative abundance compared to BV-negative microbiomes. *A. intestenii*, *T. forsythensis*, and *E. saburreum* were newly associated with BV. *A. intestenii* is an anaerobic intestinal commensal that has previously been found in the vaginal flora

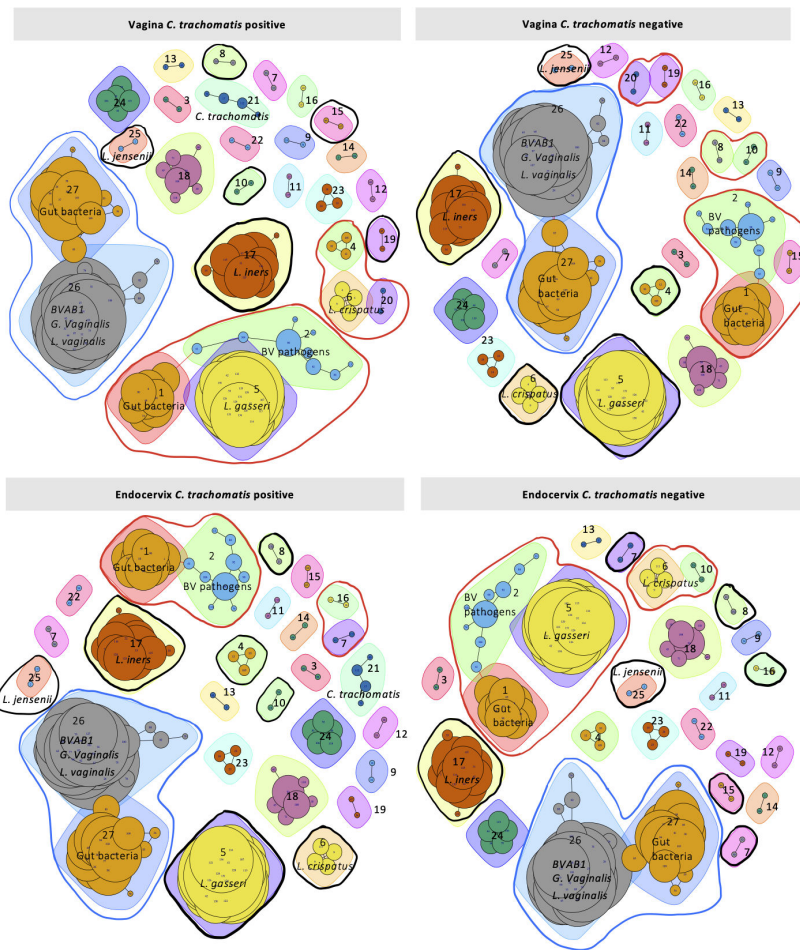


FIG 7 Network analysis based on Pearson’s correlation of the 92 *Ct*-infected or *Ct*-uninfected paired endocervical and vaginal microbiomes. Graph partitioning was performed to identify microbial community clusters within each network, and “modularity” was calculated to measure the degree of stability within the densely connected compartments or communities (see Materials and Methods). Nodes represent two or more groups—single or related species shown as different colored shapes—where the larger the size of the node, the greater the weight of the interactions of the groups within the network. Red lines show nodes with differential groups between paired endocervical and vaginal *C. trachomatis*-infected or *C. trachomatis*-uninfected networks and between endocervical and vaginal anatomic site networks. Black lines show differential independent groups between pairs and also between anatomic sites in the networks. Blue lines show similar nodes for all four *C. trachomatis*-infected and *C. trachomatis*-uninfected endocervical and vaginal networks.

(36). The other species are present in oral biomes and linked to periodontal disease and/or respiratory infections (37–39). Their association with BV is not understood, although oral to genital autoinoculation or transfer during sex is a possibility.

Xpert CT/NG and wet prep tests on vaginal samples were able to detect *Ct*, *Ng*, *Tv*, and *Candida*, while MSS detected each along with *Mg*. In our previous study, *Mg* was found in 100% of samples using One Codex. In contrast, only 3.5% and 7.6% of vaginal and endocervical microbiomes had *Mg* using VIRGO (24). This discrepancy is not surprising, since complete *Mg* genomes were used to build the VIRGO database, which is therefore highly specific.

HPV types were identified using HPVviewer (30), providing much needed data on the distribution of genotypes among PIs in the WPR. The only published study of HPV genotypes in PICTs identified hrHPV genotypes 16, 18, 33, 39, 45, and 52 among cervical cancer specimens along with IrHPV genotypes 54 and 67 that are infrequently found in

these malignancies (40, 41); no data were available for women without cervical cancer. hrHPV 16, 18, 39, and 52 were also found in our Fijian population. Of the common hrHPV types present in the population, HPV 57 has been associated with oral papillomas (42). A possible oral to genital auto-inoculation or sexual transmission could explain the unexpected presence of this genotype in genital microbiomes.

Although no significant association of endocervical hrHPV genotypes with ethnicity, age, or *Ct* infection was found, not surprisingly, these high-risk types were associated with higher microbial diversity compared to low-risk types or no infection in the endocervical microbiome. This finding is consistent with two recent studies in China that showed a significantly increased alpha diversity in the HPV-infected cervix compared to those without HPV (43, 44).

Many endocervical (41.3%) and vaginal (36.04%) microbial relative abundance profiles did not match their assigned subCST based on VALENCIA. By manually comparing the species relative abundance of each microbiome with their assigned subCSTs, we developed four new subCSTs—IV-D0, IV-D1, IV-D2, and IV-E—that provide a more specific classification for Fijian microbiomes. Microbiomes classified as IV-B required reclassification as IV-D0, IV-D1, and IV-D2, and those assigned as CST III-B, IV-A, IV-B, IV-C0, IV-C1, or IV-C4 were re-classified as CST IV-E. The need to expand CST classifications is supported by a 2022 study of ethnic Han Chinese where a predominance of *L. helveticus* was discovered based on 16S rRNA sequencing of vaginal microbiomes not reflected in the VALENCIA reference centroids (31). They proposed a new CST-VI to represent communities that are dominated by this species. Additional CSTs will likely be required to further define endocervical and vaginal microbial landscapes, as more diverse and globally represented populations are studied.

The new subCSTs IV-D0, IV-D1, IV-D2, and IV-E had a predominance of *G. vaginalis*, *L. iners*, and *Prevotella* spp. along with other BV-associated pathogens. Many of these microbiomes may have a predisposition to *Ct* and/or HPV infections in this population. While there was no association between new subCSTs that did not match assigned subCSTs and *Ct* infection, there was a significant association with iTaukei and Other PI ethnicities.

Paired microbiomes had unique attributes depending on *Ct* status. Among *Ct*-uninfected pairs, while a little over a third had differential species relative abundance between anatomic sites, 15.15% of the pairs with similar species were dominated by *L. crispatus* with a slightly greater relative abundance in the endocervix. This is in contrast to a study where uninfected Chinese women were found to have a higher abundance of this species in the vagina compared to the endocervix (45). *L. crispatus* produces D-lactic acid and bacteriocins that may trap or inhibit pathogen attachment to host cells, protecting against STIs and also BV to maintain microbiome eubiosis (46–48). Similar to studies of South African adolescents and Italian women, we found that *L. crispatus*-dominated microbiomes were less likely to be associated with *Ct* (13, 49).

Among *Ct*-infected pairs, over 80% had differential species relative abundance with 72 species showing a significant difference between the two anatomic sites. Not surprisingly, these pairs were significantly more likely to have differential species relative abundance between sites compared to *Ct*-uninfected pairs. The *Ct*-infected endocervix also had a higher microbial diversity compared to the paired vaginal microbiome. These data are consistent with our previous study (26) and a study of adolescents where paired endocervical and vaginal microbiomes were also examined (13). A recent investigation of HPV infections found similar results (50). Additionally, the high dominance of *G. vaginalis* with other BV-associated pathogens such as *Prevotella*, *Megasphaera*, and *Atopobium* among *Ct*-infected microbiomes is supported by investigations in South Africa and Italy, although 16S rRNA-based sequencing was used instead of metagenomics (13, 49). The relatively lower dominance of *L. iners* (18.4%) compared to *Ct*-uninfected pairs (30.7%) could be explained by the fact that *L. iners* lacks the essential amino acid synthesis repertoire, forcing it to rely on exogenous host amino acids, which would make it more sensitive to community composition and perhaps less successful in competing with

other microbes (51). However, *L. iners* has been shown in other studies to be associated with *Ct* infection and is more prevalent among women of African descent than other racial or ethnic groups (52–54).

Six (15.79%) of the *Ct*-infected endocervical microbiomes were dominated by a moderate to high abundance of *Ct*, but not in the paired vaginal microbiomes. Although the columnar epithelial cells of the endocervix are the target for *Ct* infection, this level of abundance has not previously been reported and was a unique occurrence in PIs. Indeed, few studies that evaluated the endocervical microbiome did not find this level of *Ct* abundance among, for example, Malaysian, Chinese, Indian, Italian, and South African women (13, 49, 55). One concern is that a high burden of *Ct* may contribute to upper genital tract infections and the high rate of infertility in Fiji (11, 56). A larger prospective study is required to assess the frequency of our findings in PI ethnicities and the significance in terms of microbial composition and disease pathogenesis, including fertility.

Nine clusters or nodes were identified in the paired endocervical and vaginal microbiome networks with shifts in size and composition depending on *Ct* status and anatomic site. Cluster 1, comprising a combination of gut bacteria and BV-associated species, was predominant in all networks. This finding reflects the very nature of PI microbiomes in that they tend to be highly diverse and rich in pathogenic bacteria regardless of *Ct* status. However, in the *Ct*-infected endocervical network, this cluster was appreciably smaller compared to all others, and no clusters included *Lactobacillus* spp. (i.e., groups 5, 6, 17, or 25); two of the three clusters in the *Ct*-infected vaginal microbiome contained *L. crispatus* and *L. gasseri*. Since *Ct* infects the endocervical epithelium and is found either inside sloughed endocervical cells or loose in the vagina, the influence of *Ct* on each anatomic site might be distinct. Nodes containing *Lactobacillus* spp. may interact with BV-associated species and gut bacteria to dampen the effects of these pathogenic species or vice versa, and *Ct*-infected endocervical cells may actually work to dissociate or prevent *Lactobacillus* spp. interactions. This would be consistent with the smaller nodes that lack *Lactobacillus* spp. in the *Ct*-infected endocervical networks. In contrast, *Ct*-uninfected endocervical networks had much larger nodes where *L. gasseri* (group 5) was associated with gut (group 1) and BV-associated (group 2) bacteria. Our hypothesis may also explain the similarity between *Ct*-uninfected endocervical and *Ct*-infected vaginal networks where *Ct* does not infect the vaginal squamous epithelial cells. These network interactions are, therefore, likely driven by the microbial milieu of gut and BV-associated bacteria that are influenced by the actual *Ct* infection of cells in the endocervix and not by the simple presence of *Ct*, free or otherwise, as in the vagina. In addition, the high abundance of *G. vaginalis* may form biofilms but also interact with other BV-associated species to generate polymicrobial biofilms that increase the risk and persistence of *Ct* and other STIs such as HPV (57, 58). Additional studies will be needed to tease out these pathogenic interactions.

MATERIALS AND METHODS

Sample collection and patient characteristics

Women attending Fijian MoHMS health centers in the Central Division, Viti Levu, Fiji, were enrolled after informed consent as part of the parent study (3). Human subject's approval was obtained from the institutional review boards of the Fijian MoHMS and the University of California San Francisco. Women of various ethnicities residing in Fiji aged 18 to 40 years were enrolled unless they had HIV, untreated syphilis, a diagnosis of cancer, or received antibiotics within the prior month. The parent study provided 258 de-identified vaginal and 92 paired endocervical samples collected from the same women at the same time point. All swabs were collected by trained clinicians. Vaginal samples were tested using the Xpert CT/NG assay (Cepheid); endocervical samples were screened for *Ct* using an in-house qPCR as described (26). Samples were additionally

tested for *Tv* and *Candida* by wet prep. Amsel criteria were used to diagnose BV required three or more Amsel criteria (29).

Sample processing, metagenomic shotgun sequencing, and taxonomic analysis

Genomic (g)DNA was extracted from swabs using a lysozyme cocktail prior to gDNA purification using the QIAmp DNA Mini Kit (Qiagen, Germantown, MD) as described (26). DNA concentrations were measured by the Qubit dsDNA assay kit (Invitrogen, Carlsbad, CA). MSS libraries were constructed using Illumina Nextera XT kits and sequenced using 150 nucleotide paired-end reads on an Illumina Novaseq PE150 platform.

Sequence data were processed using BBDuk v38.96 for Illumina adapter and quality trimming (59) to an average quality score of >10 and excised if scores were <10. Human read contaminants were removed using BBSplit v38.96 (59) with version 38 of the human genome (RefSeq: GCF_000001405.26). Ambiguous reads or reads not clearly human in origin were preserved for downstream processing. Data were then profiled using Metaphlan v3.0.14 for species identification (60, 61).

VIRGO (24) was run on post-processed microbial reads as described (26). The number of confirmed reads for each STI and *Candida* were divided by the total number of microbial reads for that sample and then multiplied by 1 million to arrive at a value for reads per million for that pathogen (62). HPVViewer (30) was used to identify and quantify HPV genotypes in the metagenomes.

CST assignments

The abundance data generated from VIRGO (24) was used for classification of vaginal and endocervical microbiomes into subCSTs using VALENCIA (25) based on the highest similarity score to 13 reference centroids. From our previous study (26) and a much larger sample size in the present study, there was a high discrepancy between microbiome species composition and subCST scores. We therefore manually compared VALENCIA assigned subCSTs with species composition of the 258 vaginal and 92 endocervical microbiomes to determine whether any new subCSTs could be developed.

Statistical analysis and visualization

Microbiome analysis was performed using R version 4.0.3 including stats (4.0.3), dplyr (1.0.8) and broom (1.0.1) packages (63), Vegan (v.1.4–5) (64), Adespatial (v.0.3–14) (65), ggplot2 (v.3.3.5) (66), CGPfunctions (v.0.6.3) (67), FSA (v.0.9.3) (68), ggpubr (v.0.4.0) (69), and LDM (32). Bivariate and multivariate analyses of association between covariates and *Ct* were performed controlling for confounding variables. Results were expressed as odds ratios adjusted with 95% CI. Differences in log Chao1 and also Shannon indices between paired endocervical and vaginal microbiomes were estimated using odds ratio with 95% CI and tested by paired two-sample *t* test, χ^2 test of association, or Fisher's exact test when cells were small. Differences between *Ct*-infected and *Ct*-uninfected microbiomes were tested by unpaired two-sample *t* test. The relative abundance of the top 25 species within each patient's vaginal and endocervical microbiome was visualized using the ggplot2 (v.3.3.5) library (66). PCoA was used to visualize clustering based on Bray-Curtis distances and Jaccard with statistical hypothesis testing performed by permutational multivariate analysis of variance (PERMANOVA) (see below). Statistical comparisons between *Ct*-infected and *Ct*-uninfected groups was performed using Kruskal-Wallis test and LDM.

Correlations between different bacterial taxa in the microbiomes based on *Ct* status were determined by Pearson's correlation coefficients using R packages igraph (v. 1.3.0) (33), Hmisc (v. 4.6-0) (70), and Matrix (v. 1.4-1) (71). Graph partitioning of networks was performed to identify community clusters within each network, and "modularity" was calculated to measure the degree of stability within the densely connected communities (33).

Differential relative abundance association testing was performed using LDM, which is based on linear models of regressing individual taxon data on the covariate to test, controlling for confounding variables (32, 72). Taxon data are taken to be relative abundances, arcsin-root-transformed relative abundances, and presence-absence status, and the results from analyzing all data scales were combined to produce overall tests of differential abundance at both species and community levels. LDM bases the inference on a permutation procedure, which is important to preserve data structure and thus correlations in our data. The permutation procedure was performed by stratifying groups of subjects who each had the same number of visits. For each group of subjects, covariate values from a subject were shuffled as a set with the covariate values from another subject. Additionally, covariate values within a subject were also shuffled. The species-level analysis detected associated species at the nominal False discovery rate (FDR) level of 10%. The community-level analysis aggregated information from the species-level analysis. Both the global *P*-value and individual *P*-values in boxplots were produced by LDM. LDM for presence-absence analysis (72) is based on rarefaction, and LDM aggregates information from the all rarefied taxa count table. *P*-values for testing associations of each beta diversity metric with a variable was based on the PERMANOVA method (32, 73) using the same permutation procedure described above. *P*-values were adjusted for multiple testing by Sandve's algorithm integrated in the LDM package (72).

ACKNOWLEDGMENTS

We thank the parent study for providing the de-identified data for this study and Fijian colleagues: Dr. Rachel Devi, Dr. Kinisimere Nadredre, Dr. Mere Kurulo, and Dr. Darshika Balak.

This research was supported by Public Health Service grant from the National Institutes of Health R01 AI151075 (to D.D. and T.D.R.).

AUTHOR AFFILIATIONS

¹Department of Pediatrics, University of California San Francisco, Oakland, California, USA

²Department of Biostatistics and Bioinformatics, Emory University, Atlanta, Georgia, USA

³Ministry of Health and Medical Services, Suva, Fiji

⁴Department of Medicine, Emory University School of Medicine, Atlanta, Georgia, USA

⁵Department of Medicine, University of California San Francisco, San Francisco, California, USA

⁶Department of Bioengineering, Joint Graduate Program, University of California San Francisco and University of California Berkeley, San Francisco, California, USA

⁷Bixby Center for Global Reproductive Health, University of California San Francisco, San Francisco, California, USA

⁸University of California San Francisco, Benioff Center for Microbiome Medicine, San Francisco, California, USA

AUTHOR ORCIDs

Sankhya Bommana  <http://orcid.org/0000-0002-4469-7925>

Timothy D. Read  <http://orcid.org/0000-0001-8966-9680>

Deborah Dean  <http://orcid.org/0000-0002-4490-1746>

FUNDING

Funder	Grant(s)	Author(s)
HHS NIH National Institute of Allergy and Infectious Diseases (NIAID)	R01 AI151075	Timothy D. Read Deborah Dean

AUTHOR CONTRIBUTIONS

Sankhya Bommana, Data curation, Formal analysis, Methodology, Validation, Visualization, Writing – original draft, Writing – review and editing | Yi-Juan Hu, Formal analysis, Methodology, Software, Validation, Writing – review and editing | Mike Kama, Investigation, Supervision, Writing – review and editing | Ruohong Wang, Formal analysis, Visualization, Writing – review and editing | Reshma Kodimerla, Formal analysis, Visualization, Writing – review and editing | Kenan Jijakli, Data curation, Formal analysis, Validation | Timothy D. Read, Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Resources, Supervision, Writing – review and editing | Deborah Dean, Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review and editing

DIRECT CONTRIBUTION

This article is a direct contribution from Deborah Dean, a Fellow of the American Academy of Microbiology, who arranged for and secured reviews by Georg Häcker, Institute of Microbiology and Hygiene, University Medical Center Freiburg, and Huizhou Fan, Rutgers Robert Wood Johnson Medical School.

DATA AVAILABILITY

MSS reads are available as FASTQ files with associated metadata at the NCBI-SRA under BioProject ID [PRJNA982400](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA982400). Table S7 provides accession numbers for all samples in NCBI. All scripts on data analysis are available at <https://github.com/ddeanlab/Microbial-PI-Ct-MSS>.

ADDITIONAL FILES

The following material is available [online](#).

Supplemental Material

Supplemental Figures (mBio03063-23-s0001.pdf). Figures S1 to S10.

Table S1 (mBio03063-23-s0002.xlsx). Participant characteristics.

Table S2 (mBio03063-23-s0003.xlsx). Metagenomic shotgun sequencing results and quality control statistics.

Table S3 (mBio03063-23-s0004.xlsx). MSS data for *N. gonorrhoeae* and *M. genitalium* confirmed by VIRGO, *T. vaginalis* and *C. albicans* confirmed by MetaPhlan v3.0, and HPV by confirmed HPVviewer.

Table S4 (mBio03063-23-s0005.xlsx). New classification of subCSTs for Pacific Islanders and metrics for age, ethnicity and *C. trachomatis* infection status.

Table S5 (mBio03063-23-s0006.xlsx). Average of the relative abundance data for each species of the newly assigned sub community state types (subCSTs).

Table S6 (mBio03063-23-s0007.xlsx). List of all species within the groups and clusters of vaginal and endocervical microbiome networks (see Fig. 7).

Table S7 (mBio03063-23-s0008.xlsx). Accession numbers for all metagenomic samples submitted to NCBI.

REFERENCES

- World Health Organization. 2016. Global health sector strategy on sexually transmitted infections. World Health Organization.
- Programmes STI. 2021. Global progress report on HIV, viral hepatitis and sexually transmitted infections. World Health Organization.
- Svigals V, Blair A, Muller S, Sahu Khan A, Faktaufon D, Kama M, Tamani T, Esfandiari L, O'Brien M, Dean D. 2020. Hyperendemic *Chlamydia trachomatis* sexually transmitted infections among females represent a high burden of asymptomatic disease and health disparity among Pacific Islanders in Fiji. PLOS Negl Trop Dis 14:e0008022. <https://doi.org/10.1371/journal.pntd.0008022>
- Cliffe SJ, Tabrizi S, Sullivan EA. 2008. Pacific Islands second generation HIV surveillance group. *Chlamydia* in the Pacific region, the silent epidemic. Sex Transm Dis 35:801–806.
- van Gemert C, Stooze M, Kwarteng T, Bulu S, Bergeri I, Wanyeki I, Badman S, Malverus J, Vella A, Tarivonda L, Johnston LG. 2014. Chlamydia prevalence and associated Behaviours among female sex

- workers in Vanuatu: Results from an integrated bio-behavioural survey. *AIDS Behav* 18:2040–2049. <https://doi.org/10.1007/s10461-014-0791-2>
6. Valley LM, Toliman P, Ryan C, Rai G, Wapling J, Tomado C, Huliafi S, Munnnull G, Rarau P, Phuanukoannon S, Wand H, Siba P, Mola GD, Kaldor JM, Valley AJ. 2016. Prevalence and risk factors of *Chlamydia trachomatis*, *Neisseria gonorrhoeae*, *Trichomonas vaginalis* and other sexually transmissible infections among women attending antenatal clinics in three provinces in Papua New Guinea: a cross-sectional survey. *Sex Heal* 13:420–427. <https://doi.org/10.1071/SH15227>
 7. Greub G. 2014. Chlamydial infection: a clinical and public health perspective. *Emerg Infect Dis* 20:1266. <https://doi.org/10.3201/eid2007.140490>
 8. Mania-Pramanik J, Kerkar S, Sonawane S, Mehta P, Salvi V. 2012. Current *Chlamydia trachomatis* infection, a major cause of infertility. *J Reprod Infertil* 13:204–210.
 9. Peterman TA, Newman DR, Maddox L, Schmitt K, Shiver S. 2015. Risk for HIV following a diagnosis of syphilis, gonorrhoea or chlamydia: 328,456 women in Florida, 2000–2011. *Int J STD AIDS* 26:113–119. <https://doi.org/10.1177/0956462414531243>
 10. Koskela P, Anttila T, Bjørge T, Brunsvig A, Dillner J, Hakama M, Hakulinen T, Jellum E, Lehtinen M, Lenner P, Luostarinen T, Pukkala E, Saikku P, Thoresen S, Youngman L, Paavonen J. 2000. Chlamydia Trachomatis infection as a risk factor for invasive Cervical cancer. *Int J Cancer* 85:35–39. [https://doi.org/10.1002/\(sici\)1097-0215\(20000101\)](https://doi.org/10.1002/(sici)1097-0215(20000101)85:35-39)
 11. Sorchik R, Than J, Carter K, Linhart C, Haberkorn G. 2019. Fertility trends in Pacific Island countries and territories
 12. van der Veer C, Bruisten SM, van der Helm JJ, de Vries HJC, van Houdt R. 2017. The cervicovaginal microbiota in women notified for *Chlamydia trachomatis* infection: a case-control study at the sexually transmitted infection outpatient clinic in Amsterdam, the Netherlands. *Clin Infect Dis* 64:24–31. <https://doi.org/10.1093/cid/ciw586>
 13. Balle C, Lennard K, Dabee S, Barnabas SL, Jaumdally SZ, Gasper MA, Maseko V, Mbulawa ZZA, Williamson A-L, Bekker L-G, Lewis DA, Passmore J-AS, Jaspan HB. 2018. Endocervical and vaginal microbiota in South African adolescents with asymptomatic *Chlamydia trachomatis* infection. *Sci Rep* 8:11109. <https://doi.org/10.1038/s41598-018-29320-x>
 14. Brotman RM. 2011. Vaginal microbiome and sexually transmitted infections: an epidemiologic perspective. *J Clin Invest* 121:4610–4617. <https://doi.org/10.1172/JCI57172>
 15. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SSK, McCulle SL, Karlebach S, Gorle R, Russell J, Tacket CO, Brotman RM, Davis CC, Ault K, Peralta L, Forney LJ. 2011. Vaginal microbiome of reproductive-age women. *Proc Natl Acad Sci U S A* 108 Suppl 1:4680–4687. <https://doi.org/10.1073/pnas.1002611107>
 16. Gong Z, Luna Y, Yu P, Fan H, Yang X. 2014. Lactobacilli inactivate *Chlamydia trachomatis* through lactic acid but not H2O2. *PLoS ONE* 9:e107758. <https://doi.org/10.1371/journal.pone.0107758>
 17. Di Paola M, Sani C, Clemente AM, Iossa A, Perissi E, Castronovo G, Tanturli M, Rivero D, Cozzolino F, Cavaliere D, Carozzi F, De Filippo C, Torcia MG. 2017. Characterization of Cervico-vaginal Microbiota in women developing persistent high-risk human Papillomavirus infection. *Sci Rep* 7:10200. <https://doi.org/10.1038/s41598-017-09842-6>
 18. Tamarelle J, Thiébaud ACM, de Barbeyrac B, Bébéar C, Ravel J, Delarocque-Astagneau E. 2019. The vaginal microbiota and its association with human papillomavirus, *Chlamydia trachomatis*, *Neisseria gonorrhoeae* and *Mycoplasma genitalium* infections: a systematic review and meta-analysis. *Clin Microbiol Infect* 25:35–47. <https://doi.org/10.1016/j.cmi.2018.04.019>
 19. Zhou X, Brown CJ, Abdo Z, Davis CC, Hansmann MA, Joyce P, Foster JA, Forney LJ. 2007. Differences in the composition of vaginal microbial communities found in healthy caucasian and black women. *ISME J* 1:121–133. <https://doi.org/10.1038/ismej.2007.12>
 20. Gupta VK, Paul S, Dutta C. 2017. Geography, ethnicity or subsistence-specific variations in human microbiome composition and diversity. *Front Microbiol* 8:1162. <https://doi.org/10.3389/fmicb.2017.01162>
 21. Borgdorff H, van der Veer C, van Houdt R, Alberts CJ, de Vries HJ, Bruisten SM, Snijder MB, Prins M, Geerlings SE, Schim van der Loeff MF, van de Wijgert JHHM. 2017. The association between ethnicity and vaginal microbiota composition in Amsterdam, the Netherlands. *PLoS One* 12:e0181135. <https://doi.org/10.1371/journal.pone.0181135>
 22. Fettweis JM, Brooks JP, Serrano MG, Sheth NU, Girerd PH, Edwards DJ, Strauss JF, Jefferson KK, Buck GA. 2014. Differences in vaginal microbiome in African American women versus women of European ancestry. *Microbiology (Reading)* 160:2272–2282. <https://doi.org/10.1099/mic.0.081034-0>
 23. MacIntyre DA, Chandiramani M, Lee YS, Kindinger L, Smith A, Angelopoulos N, Lehne B, Arulkumaran S, Brown R, Teoh TG, Holmes E, Nicholson JK, Marchesi JR, Bennett PR. 2015. The vaginal microbiome during pregnancy and the postpartum period in a European population. *Sci Rep* 5:8988. <https://doi.org/10.1038/srep08988>
 24. Ma B, France MT, Crabtree J, Holm JB, Humphrys MS, Brotman RM, Ravel J. 2020. A comprehensive non-redundant gene catalog reveals extensive within-community intraspecies diversity in the human vagina. *Nat Commun* 11:940. <https://doi.org/10.1038/s41467-020-14677-3>
 25. France MT, Ma B, Gajer P, Brown S, Humphrys MS, Holm JB, Waetjen LE, Brotman RM, Ravel J. 2020. VALENCIA: a nearest centroid classification method for vaginal microbial communities based on composition. *Microbiome* 8:166. <https://doi.org/10.1186/s40168-020-00934-6>
 26. Bommana S, Richards G, Kama M, Kodimerla R, Jijakli K, Read TD, Dean D. 2022. Metagenomic shotgun sequencing of endocervical, vaginal, and rectal samples among Fijian women with and without *Chlamydia trachomatis* reveals disparate microbial populations and function across anatomic sites: a pilot study. *Microbiol Spectr* 10:e0010522. <https://doi.org/10.1128/spectrum.00105-22>
 27. McKinney IM. 2006. Statistical news. encyclopedia of statistical sciences. John Wiley & Sons, Inc, Hoboken, NJ, USA.
 28. Naidu V, Matadradra A, Sahib M, Osborne J. 2013. Fiji: the challenges and opportunities of diversity. The University of the South Pacific.
 29. Amsel R, Totten PA, Spiegel CA, Chen KCS, Eschenbach D, Holmes KK. 1983. Nonspecific vaginitis. Diagnostic criteria and microbial and epidemiologic associations. *American J Med* 74:14–22. [https://doi.org/10.1016/0002-9343\(83\)91112-9](https://doi.org/10.1016/0002-9343(83)91112-9)
 30. Hao Y, Yang L, Galvao Neto A, Amin MR, Kelly D, Brown SM, Branski RC, Pei Z. 2018. HPVviewer: sensitive and specific genotyping of human papillomavirus in metagenomic DNA. *Bioinformatics* 34:1986–1995. <https://doi.org/10.1093/bioinformatics/bty037>
 31. Li H, Jiang J, Nie C, Xiao B, Li Q, Yu J. 2022. Community structure and ecological network's changes of vaginal microbiome in women right after delivery. *Front Pediatr* 10:750860. <https://doi.org/10.3389/fped.2022.750860>
 32. Hu Y-J, Satten GA. 2020. Testing hypotheses about the microbiome using the linear decomposition model (LDM). *Bioinformatics* 36:4106–4115. <https://doi.org/10.1093/bioinformatics/btaa260>
 33. CsardiG, NepuszT. 2006. The igraph software package for complex network research. *InterJournal, complex systems* 1695:1–9.
 34. Srinivasan S, Hoffman NG, Morgan MT, Matsen FA, Fiedler TL, Hall RW, Ross FJ, McCoy CO, Bumgarner R, Marrazzo JM, Fredricks DN. 2012. Bacterial communities in women with bacterial vaginosis: high resolution phylogenetic analyses reveal relationships of microbiota to clinical criteria. *PLoS One* 7:e37818. <https://doi.org/10.1371/journal.pone.0037818>
 35. Onderdonk AB, Delaney ML, Fichorova RN. 2016. The human microbiome during bacterial vaginosis. *Clin Microbiol Rev* 29:223–238. <https://doi.org/10.1128/CMR.00075-15>
 36. D'Auria G, Galán J-C, Rodríguez-Alcayna M, Moya A, Baquero F, Latorre A. 2011. Complete genome sequence of *Acidaminococcus intestini* RYC-MR95, a Gram-negative bacterium from the phylum firmicutes. *J Bacteriol* 193:7008–7009. <https://doi.org/10.1128/JB.06301-11>
 37. Eribe ERK, Olsen I. 2017. *Leptotrichia* species in human infections II. *J Oral Microbiol* 9:1368848.
 38. Sharma A. 2010. Virulence mechanisms of *Tannerella forsythia*. *Periodontology* 2000 54:106–116. <https://doi.org/10.1111/j.1600-0757.2009.00332.x>
 39. Hamlet SM, Taiyeb-Ali TB, Cullinan MP, Westerman B, Palmer JE, Seymour GJ. 2007. *Tannerella forsythensis* prtH genotype and association with Periodontal status. *J Periodontol* 78:344–350. <https://doi.org/10.1902/jop.2007.060161>
 40. Schisler TM, Bhavsar AK, Whitcomb BP, Freeman JH, Washington MA, Blythe JW, Nahid MA, Dietrich CS. 2018. Human papillomavirus genotypes in Pacific Islander cervical cancer patients. *Gynecol Oncol Rep* 24:83–86. <https://doi.org/10.1016/j.gore.2018.04.007>

41. de Sanjose S, Quint WG, Alemany L, Geraets DT, Klaustermeier JE, Lloveras B, Tous S, Felix A, Bravo LE, Shin H-R, et al. 2010. Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study. *Lancet Oncol* 11:1048–1056. [https://doi.org/10.1016/S1470-2045\(10\)70230-8](https://doi.org/10.1016/S1470-2045(10)70230-8)
42. Cubie HA. 2013. Diseases associated with human papillomavirus infection. *Virology* 445:21–34. <https://doi.org/10.1016/j.virol.2013.06.007>
43. Zhang Y, Xu X, Yu L, Shi X, Min M, Xiong L, Pan J, Zhang Y, Liu P, Wu G, Gao G. 2022. Vaginal microbiota changes caused by HPV infection in Chinese women. *Front Cell Infect Microbiol* 12:814668. <https://doi.org/10.3389/fcimb.2022.814668>
44. Lin W, Zhang Q, Chen Y, Dong B, Xue H, Lei H, Lu Y, Wei X, Sun P. 2022. Changes of the vaginal microbiota in HPV infection and cervical intraepithelial neoplasia: a cross-sectional analysis. *Sci Rep* 12:2812. <https://doi.org/10.1038/s41598-022-06731-5>
45. Chen C, Song X, Wei W, Zhong H, Dai J, Lan Z, Li F, Yu X, Feng Q, Wang Z, et al. 2017. The microbiota continuum along the female reproductive tract and its relation to uterine-related diseases. *Nat Commun* 8:875. <https://doi.org/10.1038/s41467-017-00901-0>
46. Breshears LM, Edwards VL, Ravel J, Peterson ML. 2015. *Lactobacillus crispatus* inhibits growth of *Gardnerella vaginalis* and *Neisseria gonorrhoeae* on a porcine vaginal mucosa model. *BMC Microbiol* 15:276. <https://doi.org/10.1186/s12866-015-0608-0>
47. Fuochi V, Cardile V, Petronio G, Furneri PM. 2019. Biological properties and production of bacteriocins-like-inhibitory substances by *Lactobacillus* sp. strains from human vagina. *J Appl Microbiol* 126:1541–1550. <https://doi.org/10.1111/jam.14164>
48. Nunn KL, Wang Y-Y, Harit D, Humphrys MS, Ma B, Cone R, Ravel J, Lai SK. 2015. Enhanced trapping of HIV-1 by human cervicovaginal mucus is associated with *Lactobacillus crispatus*-dominant microbiota. *mBio* 6:e01084-15. <https://doi.org/10.1128/mBio.01084-15>
49. Ceccarani C, Foschi C, Parolin C, D'Antuono A, Gaspari V, Consolandi C, Laghi L, Camboni T, Vitali B, Severgnini M, Marangoni A. 2019. Diversity of vaginal microbiome and metabolome during genital infections. *Sci Rep* 9:14095. <https://doi.org/10.1038/s41598-019-50410-x>
50. Zhang Z, Li T, Zhang D, Zong X, Bai H, Bi H, Liu Z. 2021. Distinction between vaginal and Cervical Microbiota in high-risk human papilloma virus-infected women in China. *BMC Microbiol* 21:90. <https://doi.org/10.1186/s12866-021-02152-y>
51. France MT, Mendes-Soares H, Forney LJ. 2016. Genomic comparisons of *Lactobacillus crispatus* and *Lactobacillus iners* reveal potential ecological drivers of community composition in the vagina. *Appl Environ Microbiol* 82:7063–7073. <https://doi.org/10.1128/AEM.02385-16>
52. van Houdt R, Ma B, Bruisten SM, Speksnijder AGCL, Ravel J, de Vries HJC. 2018. *Lactobacillus iners*-dominated vaginal microbiota is associated with increased susceptibility to *Chlamydia trachomatis* infection in Dutch women: a case-control study. *Sex Transm Infect* 94:117–123. <https://doi.org/10.1136/sextrans-2017-053133>
53. Lennard K, Dabee S, Barnabas SL, Havyarimana E, Blakney A, Jaumdally SZ, Botha G, Mkhize NN, Bekker L-G, Lewis DA, Gray G, Mulder N, Passmore J-AS, Jaspan HB. 2018. Microbial composition predicts genital tract inflammation and persistent bacterial vaginosis in South African adolescent females. *Infect Immun* 86:e00410-17. <https://doi.org/10.1128/IAI.00410-17>
54. Molenaar MC, Singer M, Ouburg S. 2018. The two-sided role of the vaginal microbiome in *Chlamydia trachomatis* and *Mycoplasma genitalium* pathogenesis. *J Reprod Immunol* 130:11–17. <https://doi.org/10.1016/j.jri.2018.08.006>
55. Cheong HC, Yap PSX, Chong CW, Cheok YY, Lee CYQ, Tan GMY, Sulaiman S, Hassan J, Sabet NS, Looi CY, Gupta R, Arulanandam B, AbuBakar S, Teh CSJ, Chang LY, Wong WF. 2019. Diversity of endocervical microbiota associated with genital *Chlamydia trachomatis* infection and infertility among women visiting obstetrics and gynecology clinics in Malaysia. *PLoS One* 14:e0224658. <https://doi.org/10.1371/journal.pone.0224658>
56. Mascarenhas MN, Flaxman SR, Boerma T, Vanderpoel S, Stevens GA, Low N. 2012. National, regional, and global trends in infertility prevalence since 1990: a systematic analysis of 277 health surveys. *PLoS Med* 9:e1001356. <https://doi.org/10.1371/journal.pmed.1001356>
57. Filardo S, Di Pietro M, Tranquilli G, Sessa R. 2019. Biofilm in genital ecosystem: a potential risk factor for *Chlamydia trachomatis* infection. *Can J Infect Dis Med Microbiol* 2019:1672109. <https://doi.org/10.1155/2019/1672109>
58. Pekmezovic M, Mogavero S, Naglik JR, Hube B. 2019. Host-pathogen interactions during female genital tract infections. *Trends Microbiol* 27:982–996. <https://doi.org/10.1016/j.tim.2019.07.006>
59. BBMap. SourceForge. 2022.
60. Beghini F, McIver LJ, Blanco-Míguez A, Dubois L, Asnicar F, Maharjan S, Mailyan A, Manghi P, Scholz M, Thomas AM, Valles-Colomer M, Weingart G, Zhang Y, Zolfo M, Huttenhower C, Franzosa EA, Segata N. 2021. Integrating Taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *Elife* 10:e65088. <https://doi.org/10.7554/eLife.65088>
61. CQLS Software Support. 2022. Metaphlan. Available from: <https://software.cqls.oregonstate.edu/updates/metaphlan-3.0.14/>
62. Babiker A, Bradley HL, Stittleburg VD, Ingersoll JM, Key A, Kraft CS, Waggoner JJ, Piantadosi A. 2020. Metagenomic sequencing to detect respiratory viruses in persons under investigation for COVID-19. *J Clin Microbiol* 59:e02142-20. <https://doi.org/10.1128/JCM.02142-20>
63. R: a language and environment for statistical computing. 2023. <https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing>.
64. Community Ecology Package (R package vegan version 2.6-4). 2023. Comprehensive Archive Network (CRAN); 2022. Available from: <https://CRAN.R-project.org/package=vegan>
65. Adespatial: Multivariate Multiscale Spatial Analysis. 2023. Comprehensive R Archive Network (CRAN). Available from: <https://cran.r-project.org/web/packages/adespatial/index.html>
66. Wickham H, Grolemund G. 2023. R for data science: import, tidy, transform, visualize, and model data. Available from: <https://r4ds.had.co.nz>
67. 2020. PlotXTabs: plot a cross tabulation of two variables using dplyr in CGPfunctions: powell miscellaneous functions for teaching and learning statistics. Available from: <https://rdr.io/cran/CGPfunctions/man/PlotXTabs.html>
68. 2023. Comprehensive R Archive Network (CRAN). Available from: <https://CRAN.R-project.org/package=FSA>
69. Github. Kassambara A. *Ggpubr*: “Ggplot2” based publication ready plots. Available from: <https://github.com/kassambara/ggpubr>
70. Harrell FE J, Harrell M. 2023. Comprehensive R Archive Network (CRAN). Available from: <https://cran.r-project.org/web/packages/Hmisc/index.html>
71. Matrix.org. 2023. Matrix V1.4 release. Available from: <https://matrix.org/blog/2022/09/29/matrix-v-1-4-release>
72. Hu Y-J, Lane A, Satten GA. 2021. A rarefaction-based extension of the LDM for testing presence-absence associations in the microbiome. *Bioinformatics* 37:1652–1657. <https://doi.org/10.1093/bioinformatics/btab012>
73. McArdle BH, Anderson MJ. 2001. Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology* 82:290–297. [https://doi.org/10.1890/0012-9658\(2001\)082\[0290:FMTCJ\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2001)082[0290:FMTCJ]2.0.CO;2)