Disparity in the vaginal microbial community composition
of healthy Caucasian and Black women

Xia Zhou*, Celeste J. Brown*, Catherine C. Davis†, Zaid Abdo†, Melanie A. Hansmann†,
Paul Joyce†, James A. Foster†, and Larry J. Forney∗

*Department of Biological Sciences, †Departments of Mathematics and at the University of
Idaho, Moscow, ID 83844-3051; ‡FemCare Product Safety and Regulatory Affairs, Procter &
Gamble Company, Cincinnati. OH 45224.

Corresponding author:
Dr. Larry J. Forney
Department of Biological Sciences
Life Sciences South, Rm. 252
University of Idaho
Moscow, Idaho 83844
Tel. 1-208-885-6280
Fax. 1-208-885-7905
E-mail: lforney@uidaho.edu

Classification: Biological science: Ecology-Medical science
Word and character count:
Data deposition footnote: The sequences reported in this paper have been deposited in
GenBank and were assigned accession numbers AY995236-AY995274.
Abstract

The composition of vaginal bacterial communities in 144 Caucasian and Black women in North America were compared based on profiles of terminal restriction fragments of 16S rRNA genes, and by phylogenetic analysis of 16S rRNA gene sequences from the numerically dominant microbial populations. Among all the women sampled there were 95 kinds of vaginal communities (‘supergroups’) that occurred in the general populace at a frequency of at least 0.1 (p=0.99). There were striking differences within and between women in these racial groups. First, only three of the seven supergroups found in both Caucasian and Black women; four of the seven supergroups found in Caucasian women were not found in Black women, while two supergroups found in Black women were absent from Caucasian women. Second, the incidence of vaginal communities in which lactobacilli were not dominant was higher in Black women (33%) as compared to Caucasian women (7%). Third, communities dominated by roughly equal numbers of more than one species of Lactobacillus (supergroups III, IV, VI, and VII) were not found in Black women, but were common in Caucasian women. Finally, phylotypes related to Clostridiaceae constituted more than 5% of the clones in 75% of Black women, but were this common in only 10% of Caucasian women. We postulate that known disparities in the susceptibility of women in these racial groups to bacterial vaginosis and sexually transmitted diseases may be partly accounted for by differences in the composition of normal vaginal flora.
**Introduction**

Bacterial vaginosis (BV) more than doubles the risk of HIV infection and is the most common cause of vaginitis, vaginal discharge, and malodor in women. BV is often resistant to effective treatment (1-3) and it adversely affects women’s health in multiple ways. During pregnancy it is associated with preterm delivery of low birth weight infants, spontaneous abortion, premature rupture of membranes, amniotic fluid infections, postpartum endometritis, and endometritis following cesarean section (4-8). In nonpregnant women BV has been associated with infertility, endometritis, and pelvic inflammatory disease (9-11). Epidemiologic studies have also established that BV or abnormal vaginal microbial communities are significantly associated with an increased risk of acquiring sexually transmitted diseases and HIV (12, 13). Disparities in the incidence of BV among racial groups have been well documented (14, 15), and Black women are 2-3 fold more likely than Caucasian women to acquire BV (14, 15). The reasons for this are unknown but can not be explained by differences in socio-demographics, sexual activity, health behavior, or hygiene alone (15, 16). An accurate understanding of the composition and ecology of the vaginal ecosystems in normal healthy women may provide clues as to why this discrepancy exists.

Much of what is known about the microbial composition of the vaginal ecosystem is derived from descriptive studies that for the most part have relied on the characterization of readily cultivated bacterial populations (17, 18). Previously investigators have paid special heed to the species of *Lactobacillus* present in the vagina. These organisms are thought to play a critical role in suppressing the growth of pathogens in the human vagina by maintaining a low pH environment and through the production of hydrogen peroxide and other bactericidal compounds (19). While BV is thought to result from replacement of the normal hydrogen peroxide-producing *Lactobacillus* sp. in the vagina with high numbers of *Gardnerella vaginalis*, *Mycoplasma hominis*, and *Mobiluncus* sp. (20). This in turn leads to
the development of an oxygen-depleted environment that facilitates the growth of strict
anaerobes including Gram-negative species of *Prevotella*, *Porphyromonas*, *Bacteroides*, and *Peptostreptococcus*, as well as to higher cytokine levels in the cervix and vagina (21). The
cause(s) that trigger the depopulation of lactobacilli, changes in microbial community
structure, and the overgrowth of these other organisms are not fully understood.

Recently, culture-independent methods based on the analysis of 16S rRNA gene
sequences have been used to overcome the limitations of culture-based methods to identify
the microbes in the human vagina (22-29). Hyman *et al.* (2005) did an intensive study of
normal vaginal microbiota that entailed the analysis of samples from 20 premenopausal
women and found that in 12/20 women the communities were dominated by various
phylogenetically distinct lactobacilli; some of which were closely related to previously
characterized species of *Lactobacillus*, while others were related to uncultured *Lactobacillus*
lineages or novel. However, quite interestingly, clones related to lactobacilli were not
recovered from 8/20 women, indicating that they are either rare members or entirely
lacking from the communities in these women. This finding corroborates other studies that,
as Hyman *et al.* point out, have collectively shown that species of *Lactobacillus* were not
dominant members of the vaginal communities in 31% of women. Instead, they are
dominated by phylotypes related to *Atopobium*, *Pseudomonas*, *Bifidobacterium*,
*Streptococcus*, *Megasphaera*, *Gardnerella*, *Prevotella*, and various phylotypes within the
order *Clostridiales* (22-25). This discovery contradicts the widely held view that “the
“normal” vagina is dominated by hydrogen peroxide-producing *Lactobacillus* species,
particularly *L. crispatu*s and *L. jensenii* (30).

In this study, we sought to define the most common types of vaginal communities in
Caucasian and Black women in North America and determine their rank-abundance in these
racial groups. The data showed there are significant differences between Caucasian and
Black women, and we postulate that they may account for the differential risk to BV of women in these racial groups.
Results

Classification of vaginal microbial communities. Differences between vaginal microbial communities in Caucasian and Black women from North America were determined by analyzing profiles of terminal restriction fragment length polymorphisms (T-RFLP) of 16S rRNA genes. The data were subjected to cluster analysis to identify similar communities, and the number of clusters (kinds of communities) was independently assessed using three different statistical algorithms. Among the 75 Caucasian women sampled there were 16 kinds of bacterial communities, including eight major groups with two or more women, and eight kinds that were represented by a single individual (Fig. 1a). Among the 69 Black women sampled, there were 12 kinds of vaginal communities, including eight major kinds and 4 represented by a single individual (Fig. 1b).

To identify the numerically abundant bacterial populations in each kind of community, 16S rRNA gene libraries were constructed from samples that represented 85% of the diversity in each cluster, and from all clusters that consisted of a single individual (‘singleton’). The composition of each sample was determined by phylogenetic analysis of partial 16S rRNA gene sequences (Tables 1a and b). In total, 57 clone libraries were analyzed, and approximately 6000 clones (~90 from each library) were sequenced. Assuming the bacterial numbers in the original samples were on the order of $10^8$ cells per ml of vaginal secretion, then comparatively rare populations present at less than $\sim 10^6$ per ml would not have been sampled and are not reflected in data from the analysis of clone libraries. From these data it was apparent that the composition of communities from certain singletons closely resembled those of clusters that already included two or more women. Hence, these similar communities were combined into ‘supergroups’. After condensing the community types into supergroups, we found there were seven supergroups among Caucasian women and five supergroups among Black women. These supergroups represent all community types that occur in the populations sampled at a frequency of 0.1 ($P=0.99$).
The supergroups of vaginal bacterial communities in Caucasian and Black women could be readily distinguished from one another based on their compositions and the relative abundances of various phylotypes.

**Vaginal communities of Caucasian women.** Among Caucasian women, 6 of the 7 supergroups were dominated by organisms that were phylogenetically related to *L. iners, L. crispatus, L. jensenii, and L. gasseri* (Table 1a). These accounted for 89% of the women sampled. Clones related to two heterofermentative species of *Lactobacillus*, namely *L. vaginalis* and *L. coleohominis*, were rare members of a few community types. Overall, *L. iners* was the most common species of *Lactobacillus* in Caucasian women; it was recovered in 83% of the women sampled (62/75) women, and was the most abundant species in 48% (36/75) of women. Combinations of two or three species of lactobacilli, whose abundances were roughly equal, dominated vaginal communities in supergroups III, IV, VI, and VII, and these accounted for 32% of the women sampled. In contrast, communities of supergroup V had low numbers of lactobacilli, and exhibited greater species evenness through the inclusion of high numbers of *Atopobium, Lachnospiraceae, Anaerococcus, Dialister, Megasphaera, and Micromonas*. In addition, about 20-30% of clones from communities in supergroup V included an assortment of novel phylotypes within the phylum *Firmicutes* (Table 1a, Fig. S3b). Only three of 75 Caucasian women sampled (S12, S13 and S16) had vaginal communities that did not belong to one of the seven supergroups. Their frequency among all Caucasian women cannot be discerned from the available data.

**Vaginal communities of Black women.** The kinds of microbial communities found in the vaginas of Black women (Table 1b) differed significantly from those of Caucasian women (Table 1a). Only three of the seven supergroups (I, II and V) found in Caucasian women were also found in Black women. The proportion of Black women with vaginal communities
in supergroup V was roughly four times that of Caucasian women. Two supergroups, VIII and IX, were found in Black women that were not present in Caucasian women. These were dominated by *L. gasseri* (supergroup VIII) and a novel lineage within the order *Clostridiales* that had less than 90% 16S rRNA gene sequence similarity to known organisms. The incidence of vaginal communities dominated by various species of *Lactobacillus* was less in Black women (67%) than in Caucasian women (93%). At the same time the incidence of communities with a high proportion of phylotypes related to strictly anaerobic bacteria was considerably higher in Black women (30%, supergroups V and IX) than in Caucasian women (7%, supergroup V). Among the microbial communities sampled from Black women, 4 did not belong to any supergroup. Community S11 was dominated by *L. jensenii*, which accounted for 80.9% of the clones analyzed, and there were also high numbers of clones related to *Streptococcus* sp., *Catenella* sp., *Prevotella* sp. and *Finegoldia* sp. Singeltons S9, S10 and S12 differed from other communities; all had comparatively high numbers of clones related to *Gardnerella vaginalis*, *Escherichia coli*, *Shigella* sp., and *Mycoplasma* sp.. These three women may suffer from vaginal infections that were overlooked, or may have transiently high numbers of organisms of fecal origin.

**Rank abundance of community types in racial groups.** The rank abundance of community types differed among Black and Caucasian women (Fig. 2). Three findings were readily apparent. First, only three of the seven supergroups (I, II and V) found in Caucasian women were also found in Black women and, as with Caucasian women, the most common community type, supergroup I, was dominated by *L. iners*. Second, communities dominated by roughly equal numbers of more than one species of *Lactobacillus* (supergroups III, IV, VI, and VII) were not found in Black women. Third, communities akin to supergroup V, which include those dominated by *Atopobium* sp., *Firmicutes*, and other strict anaerobes, were four times more common in Black women than in Caucasian women. Thus, a
significant fraction of Black women have vaginal communities that differ in bacterial composition from those of Caucasian women.

**Diversity of numerically abundant of *Lactobacillus*.** The phylogenetic relationships of the *Lactobacillus* strains were determined by comparing the 16S rRNA gene sequences from this study to those of reference strains previously sequenced. Most lactobacilli found in the vaginal communities were phylogenetically related to *L. iners, L. crispatus, L. jensenii,* and *L. gasseri* (Fig. S3a), and were likely to be homofermentative. In contrast, *L. vaginalis* and *L. coleohominis* were phylogenetically distinct and related to heterofermentative species (31). The sequence heterogeneity among clones of *L. crispatus* and *L. jensenii* was greater than that of *L. iners* and *L. gasseri*, suggesting there are evolutionarily divergent subpopulations of *L. crispatus* and *L. jensenii* in vaginal communities. It should be noted that some clone sequences matched *L. crispatus* NCTC 4 (AJ242969), a reference strain that is distantly related to three other *L. crispatus* type strains. Therefore, these may actually represent a second distinct phylotype. In contrast to *L. crispatus* and *L. jensenii*, the clones of *L. iners* were highly related to one another and to a single reference strain, *Lactobacillus* sp. LSPY 17362. The occurrence of a virtually clonal lineage of *L. iners* in different women suggests there might be strong selection for specific phenotypic characteristics that are found in few strains.
Discussion

In this study we found that rather discrete differences exist in the composition of vaginal microbial communities of Caucasian and Black women in North America, and it was possible to classify them into one of nine supergroups. Some supergroups (I, II, and V) were common to both racial groups, while others (III, IV, VI, and VII) were only found in Caucasian women, and still others (VIII and IX) were only found in Black women. We postulate that these differences in the structure and composition of microbial communities may underlie well-known differences in the susceptibility of women in these racial groups to BV and various vaginal infections (14, 15, 32). There is little known about the development of vaginal communities in adolescent women, and whether the composition of a vaginal community remains constant throughout a woman's reproductive years. As a result, it is not known why a woman develops a particular kind of community rather than another, or if vaginal communities transition between supergroups at various times. However, we found it interesting that only four species of *Lactobacillus* were numerically important in the communities of women in both racial groups. Given the known diversity of lactobacilli in the world, and the limited number found in human vaginas, it would appear that there are host factors that select for specific organisms, that these species have unusual characteristics that allow them to successfully colonize the vagina, or both.

The resilience of an ecosystem is defined as its ability to resist or absorb perturbations and return to a stable equilibrium state (33). It is largely determined by the ability of indigenous species to adapt to stresses and disturbances, as well as the existence and nature of ecological interactions among species. Ecological theory and empirical data indicate that not all communities are equally resilient. Given this, we hypothesize that differences in the resilience of various vaginal microbial communities may account for the differential susceptibility of races and specific human populations to BV and other urogenital infectious diseases. If the resilience of a vaginal community is low then transitory changes...
to the structure of these communities may occur more readily in response to disturbances of various kinds, including menses, sexual intercourse, douching, and contraceptive practices. These changes in community structure may include shifts in population densities or species loss, and concomitant changes in community function may occur. These disturbed communities may be more susceptible to invasion by species that are not indigenous to the human vagina including transient species of fecal origin and opportunistic pathogens. Moreover, we speculate that the disturbed state may itself constitute the clinical syndrome of BV.

There is no information available on the relative resilience of different vaginal microbial communities. However, since the prevalence of BV in African American women is higher than in Caucasian women, it would make sense to focus attention on those community types that are more common or exclusively occur in African American women, namely, supergroups V, VIII, and IX. The dominant members of these supergroups were *Atopobium* and various genera in the order *Clostridiales*. A second hypothesis is that vaginal communities that are more common or only found in Caucasian women may be more resilient, and therefore more resistant to enduring ecological upsets or invasion by nonindigenous organisms. Again, it makes sense to focus attention on those community types that are more common or exclusively occur in Caucasian women. These community types, supergroups III, IV, VI, and VII, were characterized by the presence of roughly equal numbers of more than one species of *Lactobacillus*.

It is generally accepted that maintenance of a low pH in the vagina is important for precluding pathogenic organisms, and this function is most often attributed to lactic acid produced by species of *Lactobacillus*, which are commonly found in the vaginas of healthy women. Consistent with this we found that lactobacilli constituted ≥10% of vaginal communities in 71/72 Caucasian women and 52/64 African-American women. (Singletons were excluded from these calculations.) *Atopobium vaginae*, which also metabolizes glucose
via homolactic fermentation (34), was an important constituent of communities in
supergroup V. Thus, if one includes members of this genus in the tally, the vaginal
communities of all Caucasian women, and all but two African-Americans had high numbers
of lactic acid bacteria. The vaginal communities of these two African-American women were
dominated by a novel phylotype that is only distantly related to known genera so there is no
basis for speculation about its physiology. Nonetheless, it is clear that the ecological
function of vaginal bacterial communities – the formation of lactic acid and maintenance of a
low pH environment – was highly conserved among women of both racial groups despite the
variations seen in community composition. While this has not previously been demonstrated
for the human vagina, the conservation of function by communities that vary in composition
is a common theme in microbial ecology that has been observed in numerous other
habitats.

The presence of high numbers of lactic acid bacteria is often equated with “healthy”
and low numbers or their absence as being “abnormal”. The strong linkage between high
numbers of lactic acid bacteria and a normal microbial community is consistent with
Walker’s “drivers and passengers” hypothesis of community structure (35). This hypothesis
posit that ecosystem function is largely determined by individual “driver” species or guilds
(functional groups) of such species and that other species in the community are
“passengers” that have minor ecological impact. Driver species strongly influence the
species composition and structure of the biological communities in which they and
passenger species exist. Under this driver-passenger model, lactic acid bacteria would be
considered drivers because they strongly influence the ecosystem by maintaining a low pH
through lactic acid production, whereas the other species present would be less influential
and constitute passengers with little influence on the ecology of the system. The
consideration of such models may prove useful to studies of vaginal microbiology since they
provide a framework for understanding the ecology of the vagina, and the possible causes and prevention of ecological upsets such as BV.

Other recent studies have documented the occurrence of *Atopobium* in vaginal communities of reproductive age and post-menopausal women. For example, Ferris et al (29) reported that *A. vaginae* was present in a significant proportion (55%) of women diagnosed as having BV, but in only 2 of 24 women that had normal vaginal communities. Likewise, Burton et al (36) screened 35 postmenopausal women for the occurrence of BV and found *A. vaginae* in 44% of women with (asymptomatic) BV, while the organism was absent from subjects deemed healthy. Fredricks and Verhelst *et al.* (23, 28) discovered that *A. vaginae* was frequently detected in subjects with BV. Based on these findings, investigators have implicated *A. vaginae* in BV. In this study we found *Atopobium* in 77% of Black women and in 39% of Caucasian women that were healthy and showed no clinical symptoms of BV and did not report any abnormalities. Based on our findings, one can conclude that *Atopobium* is a commonly encountered constituent of vaginal communities, but its association with BV is unclear. In all of the studies cited above (except ours) the diagnosis of BV was based on the Nugent criteria in which a vaginal smear is Gram stained and examined for the occurrence and number of lactobacilli that adhere to vaginal epithelial cells. Low numbers or the absence of lactobacilli is the criterion used for the diagnosis of BV. Although *Atopobium* is a lactic acid bacterium, its cellular morphology is distinctly different from that of *Lactobacillus* (34). Thus, it seems entirely plausible to us that the women whose communities are dominated by *Atopobium* and lack appreciable numbers of lactobacilli may be misdiagnosed as having BV. This may partly explain the high incidence of so-called asymptomatic BV that has been reported by many investigators (2, 37). This postulate should be rigorously tested in future studies.

The data obtained in this study showed that a fair proportion of healthy Caucasian and African-American women host several fastidious, strictly anaerobic microorganisms that
belong to the order *Clostridiales*. These include appreciable numbers *Lachnospiraceae*, *Megasphaera*, *Dialister* and *Anaerococcus*, as well as many novel bacteria. These populations probably may not have been recovered from samples by using the cultivation methods commonly employed in previous studies of vaginal flora, or routinely used in clinical microbiology laboratories. Given that they appear to constitute more than 5% of communities in 10% of Caucasian and 75% of African-American women, efforts are needed to isolate and characterize these organisms to obtain clues to their function in vaginal communities. Importantly, the common occurrence of species classified in the *Clostridiales* may have important implications for the clinical diagnosis of BV. Members of this order are notorious for the production of malodorous compounds that include organic acids, amines, and thiols through the metabolism of glucose, other carbohydrates, and amino acids (38). Complaints of vaginal odor in the absence of infection are not uncommon. For example, Lander and Hillier *et al.* (39) studied 598 women with genital complaints and found that 23% women reported vaginal odor even though there were no outward signs of disease. However, the normal occurrence of *Clostridiales* in many vaginal communities suggests that some degree of odor may be normal for some women, which could lead to false positives in the diagnosis of BV. Indeed, the presence of amines in vaginal secretions is one of four criteria used for the diagnosis of BV based on the recommendations on Amsel *et al.* (40) that are commonly used in clinical settings. This also suggests that new diagnostic tests for BV that are based on amine production and odor formation should be used with caution (41, 42).

The incomplete understanding of normal vaginal flora has led to confusion and uncertainty regarding the nature and cause of BV (30, 43). The U.S. Centers for Disease Control and Prevention has characterized BV as a clinical syndrome that results from the replacement the normal hydrogen peroxide-producing *Lactobacillus* sp. in the vagina with high concentrations of anaerobic bacteria. Similarly vague characterizations are found
throughout the scientific literature (44, 45), and the underlying cause(s) of BV are unknown. The results of this study show that *Lactobacillus* sp. are not the only genus of lactic acid bacteria found in the vagina. In this regard, we concur with *Hyman* et al. (24) who concluded that the absence of *Lactobacillus* does not by itself define an unhealthy state. Thus, at the very least, the definitions used are incomplete and this may partly account for the enigmatic occurrence of asymptomatic BV in as much as 50% of women that are diagnosed with BV based on the criteria recommended by *Amsel* et al. (46) or *Nugent* et al. (47). Indeed, it occurs to us that given the occurrence of lactic acid bacteria in the vaginas of women that are not species of *Lactobacillus*, it could well be that some patients with recurring BV may simply be reflecting their normal vaginal bacterial flora which are mischaracterized as abnormal.

The data presented raise the specter that the diagnosis of BV based on commonly used criteria may yield an unacceptably high percentage of false positives. These misdiagnoses could partly explain clinical phenomena such women who repeatedly become symptomatic or fail to respond to after antibiotic treatment. In addition, the unnecessary treatment of patients with antibiotics is not only costly, but it could well lead to disturbances of intestinal or vaginal flora and cause problems that otherwise would not have occurred. To avoid this, health care practitioners may need to rethink what constitutes “normal” flora and the criteria used to diagnose BV in light of what is being learned about the differences in vaginal communities within and between racial groups.
Material and Methods

Subjects and clinical study design. As part of a study on the prevalence of *Staphylococcus aureus* carriage (48), 3012 healthy menstruating women between the ages of 13-40 were enrolled from five sites in North America. The subjects who were recruited by Hill Top Research Inc. in Cincinnati, OH, East Brunswick, NJ, St. Petersburg, FL, Scottsdale, AZ and Winnipeg, Manitoba, Canada. The racial profile of the women corresponded with that of the 1990 USA Census: 80% White, 12% Black, 5% Hispanic, and 3% Asian. A subset of 150 vaginal samples were included in our study; 75 were Caucasian and 69 were Black. The samples were chosen so that a total of 15 samples were drawn from each of the five geographic sites, and within each site there were an equal number of women from each of three age groups: 13-18, 19-35, and 36-40 years old. Subjects were eligible for enrollment if they had regular menstrual cycles (21 - 35 days); used tampons at least occasionally; were able to read, write and understand English; did not bathe or shower within 2 h of their scheduled visit; refrained from douching, vaginal medications, suppositories, feminine sprays, genital wipes, or contraceptive spermicides for 48 h prior to their scheduled visit; and were willing to comply with all other protocol requirements. Subjects were not eligible if they were participating in another clinical study; were pregnant, actively trying to get pregnant or suspected they were pregnant; had a gynecological abnormality as judged by the study medical personnel; had an infection of the genitals within the past 6 weeks; had been medically diagnosed as having diabetes, kidney failure, hepatitis, AIDS (HIV positive) or toxic shock syndrome; or were currently taking (within the last 30 d) immunosuppressive drugs, chemotherapy, systemic antimicrobial or antifungal drugs, or antimicrobials to treat a vaginal infection. Subjects completed a demographic questionnaire and classified themselves into one of four distinct racial groups: White, Black, Hispanic or Asian. The study protocol and informed consent document were reviewed and approved by Hill Top’s
Institutional Review Board. Documented informed consent was obtained from all subjects prior to participation in this study.

From each woman a sample was taken near the mid-vagina using a sterile swab. A saline-lubricated speculum was used to minimize contamination of the sample by the flora of the labia during entry and withdrawal of the swab. The swab was placed in a sterile cryovial and stored at -70°C until analysis. Upon collection of the vaginal sample, the attending health care practitioner noted any signs of possible genital infections (e.g., discharge, cervicitis, or foul odor). None of the subjects in this study had signs of vaginal infection.

**T-RFLP analysis of 16S rRNA genes.** Total microbial community DNA was isolated from vaginal swabs as previously described (25). Similarities and differences among the numerically dominant populations in the vaginal communities were determined by comparing profiles of terminal restriction fragment polymorphisms (T-RFLP) of 16S rRNA genes using a modified version of the method described by Zhou *et al.* (25). A detailed description of the methodology used is provided in the Supporting Information.

**Cluster analysis of T-RFLP data.** Cluster analysis of T-RFLP data was done to identify communities that had similar numerically abundant populations. Abdo *et al.* (49) previously described the algorithms used. First, true peaks were identified once a threshold (baseline) had been defined. Second, hierarchical clustering was done to identify those fragments with lengths close enough to justly group them in the same length category. Third, the Euclidean distances between T-RFLP profiles were calculated and these were hierarchically clustered based on average linkage (UPGMA) and a dendrogram was constructed. Finally, three clustering criteria were employed to identify a statistically meaningful number of groups in
the data: the Cubical Clustering Criteria, the pseudo $F$; and a statistic that can be
transformed to pseudo $T2$.

A ‘coverage sampling approach’ was employed to identify the fewest samples that
accounted for 85% of the phylotype diversity found within a cluster. In doing so, we assume
that in all profiles terminal restriction fragments with a unique size represented a single
phylotype. This reduced the total number of samples that needed to be analyzed while at
the same time assuring that each cluster was adequately sampled (49). The samples
specified were used to construct clone libraries of amplified 16S rRNA genes for subsequent
sequence analysis.

**Phylogenetic analysis of 16S rRNA gene sequences.** Primers FD1f and RD1r (50) were
used to generate nearly full length 16S rRNA amplicons of 16S rRNA genes in vaginal
samples. These mixtures of amplicons were used for library construction as previously
described (25). Approximately 100 clones from each library were randomly chosen and
sequenced. The details of DNA sequencing and sequence analysis are provided in the
Supporting Information.
Acknowledgements: We would like to thank Lin Yu, Audra Johnson, Maria G. Schneider and Jacob D. Pierson for their technical assistance; Stephen J. Bent and Ursel Schütte for their valuable advice on analyses of T-RFLP and sequence data; and William Ledger (Weill College of Medicine, Cornell University) for constructive suggestions regarding the manuscript. The authors gratefully acknowledge the Hilltop Research Facility (Cincinnati, OH) for conducting the clinical study.

Funding/Support: The project described was supported by grants P20RR16448 and P20RR016454 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH) and its contents are solely the responsibility of the authors and do not necessarily represent the official views of NCRR or NIH. Financial support from the Procter & Gamble Company, Cincinnati, OH was used to design and conduct the study.
References


Figure Legends

Fig. 1. Clustering of vaginal microbial communities in Caucasian and Black women based on T-RFLP analysis of 16S rRNA genes. Clusters with ≥2 women designated with a “C” followed by a number were from Caucasian women, while those designated with a “B” followed by a number were from Black women. An asterisk designates the samples used to construct clone libraries. “Clusters” that consisted of a single sample (singletons) were designated with an “S” followed by a number.

Fig. 2. The rank abundances of vaginal community supergroups found in Caucasian and Black women. Blue bars indicate supergroups found in Caucasian women and red bars indicate supergroups found in Black women.