Comparative Analysis of Vaginal Bacterial Diversity in Northern-Chinese Women Associated With or Without Bacterial Vaginosis

Yajuan Liu1, Wenbo Hao1, Chaohui Duan2, Jianping Tan1, Weixun Xu1, Shimeng Zhang1, Xiaqing Liao1, Ming Li1*, Shuhong Luo1*

1Institute of Antibody Engineering, School of Biotechnology, Southern Medical University, 1383 N. Guangzhou Avenue, Guangzhou People’s Republic of China
2Laboratory of Clinical Immunology, the Sun Yat-Sen Memorial hospital, Sun Yat-Sen University, Guangzhou People’s Republic of China

Abstract

Background: Bacterial Vaginosis (BV) is the common vaginal infection in women and it has been linked to enhanced risks for pre-term birth, pelvic inflammatory disease and sexually transmitted diseases. BV is caused by a disorder of vaginal microbiota which changes from the normal lactobacillus dominated community to a more diverse community of non-lactobacillus bacteria. Several previous reports analyzed the overall vaginal microbial communities of volunteers from limited sampling area and they suggested a possible link between the vaginal microbial contents and the ethnicity of women. Here, we analyzed the diversity of vaginal microbiota in 10 subjects associated with BV (BV+) and 10 subjects without BV (BV-) from the metropolitan area of Herbing in Northern China using full-length 16S rDNA.

Results: The vaginal bacterial communities detected in subjects with BV were much more taxon rich and diverse than those without BV. At a 97% sequence similarity cutoff, the number of operational taxonomic units (OTUs) per 10 subjects with BV was nearly three times greater than 10 subjects without BV by 29.4 ± 9.3 versus 11.7 ± 7.8 (Mean ± SD). Our data confirmed that there is a shift in the abundance of bacterial species present in the vaginal environment when BV and non-BV groups were compared. Each sequence read was assigned to a genus or a species when possible. Principal component analysis was performed at genus levels. Most BV+ samples clustered together while there were two clusters among BV- samples. Several bacteria have been found to be associated with BV, including Gardnerella, Atopobium, Peptoniphilus, Leptotrichia/Sneathia, Prevotella, Parvimonas and Dialister. Based on result of classification, four possible novel phyotype microorganisms were found.

Conclusions: The data presented here on the composition and richness of the vaginal microbial ecosystem in BV and health state will provide the depth insight in the etiology of BV.

Keywords: Bacterial vaginosis; 16S rDNA amplification; Clone library; Vaginal flora; Microbiology

Abbreviations: BV: Bacterial vaginosis; OTUs: Operational taxonomic units; PCA: Principal component analysis

Background

Bacterial vaginosis (BV) is the common vaginal infection in women of reproductive age, and is the most common etiology of vaginal symptoms which prompt women to seek medical care [1]. Numerous health problems including preterm labor resulting in low birth weight [1], pelvic inflammatory disease [2], and acquisition of HIV [3] are closely related to BV.

BV is caused by an imbalance of naturally occurring bacterial flora in healthy vagina with the common normal bacterial flora of Lactobacillus crispatus and Lactobacillus iners. Lactobacilli include some hydrogen peroxide-producing species that help to prevent other vaginal microorganisms from excessive growth. Previous reports showed that BV patients have dromatic changes in vaginal from the normal lactobacillus from excessive growth. Previous reports showed that BV patients have dromatic changes in vaginal from the normal lactobacillus dominated community to the diversity of non-lactobacillus bacteria [4]. Alone the absence of Lactobacillus does not define an unhealthy state. Complementarily, the presence of solely, or a combination of, Atopobium, Gardnerella, Peptostreptococcus, Prevotella, Pseudomonas, and/or Streptococcus (often nosuxia bacteria when in/on humans) does not define an unhealthy state [5]. In the past, the composition and diversity of human vagina microbiota were detected by cultural methods which had many limitations for the really circumstance. The advent of PCR based techniques and pyrosequencing made it possible to further examine this complex microbial niche. The small ribosomal subunit or 16S rDNA gene was often used as the most common target for molecular identification of bacteria. Bacterial 16S rDNA gene is present in all bacteria and has conservative regions that can be targeted with broad range PCR primers [6]. Areas of amplified sequences can be used to characterize and identify the bacteria origin, in order to determine their phylogenetic relationships.

In this study, we used the full-length 16S rDNA to determine the diversity of vaginal microbiota in 10 BV+ and 10 BV- women from Haerbin in Northern China and to get the basic data for the bacterial imbalance when BV occured. Our results revealed that analysis of vaginal microbiota and the clustering of microbial profile could help in the identification of bacterial vaginosis and assessment of women’s vaginal health.

*Corresponding author: Dr. Shuhong Luo, Institute of Antibody Engineering, School of Biotechnology, Southern Medical University, Guangzhou Avenue, Guangzhou, Guangdong Province, China; Tel: 8620-61648325 ; Fax : 8620-61648554; E-mail: shluo815@yahoo.com

Received March 17, 2012; Accepted May 22, 2012; Published May 30, 2012


Copyright: © 2012 Liu Y, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
Methods

Human subjects

This study was approved by both the ethics committee of Southern Medical University and by the Heilongjiang Provincial People’s Hospital. Twenty women were recruited for this study. All the women enrolled provided informed consent and none of them declined participation. The study was supervised by experts from Southern Medical University and Heilongjiang Provincial People’s Hospital. These subjects were at various (recorded) times of their menstrual cycle with 19-55 years of age, including 10 BV positive (BV+ group) women (BV+ group, aged 36.8 ± 15.02.) and 10 BV negative (BV-) women (BV-group, aged 29.1 ± 4.43), who came to the Department of Obstetrics and Gynecology, Heilongjiang Provincial People's Hospital, for routine gynecology examination from August 2009 to July 2011.

BV status was assessed based on Amsel criteria for all subjects [7]. The participants who met three or more of the following criteria were clinically diagnosed as BV+: homogenous vaginal discharge, >20% clue cells on wet mount microscopy, elevated pH (≥ 4.5) of vaginal fluid, and release of a fishy amine odor on addition of 10% potassium hydroxide solution to vaginal fluid (“whiff” test). Diagnosis was confirmed using BV Quick-Test sialidase test kit (Zhuhai Livzon Pharmaceutical Group Inc, China, Cat. No.2401325). Participants without these symptoms were defined as BV negative group (BV-).

Participants were excluded by any of the following exclusion criteria : <18 years of age, pregnancy, taking contraceptive steroids, complaints of urogenital symptoms or noticeable infection on physical examination of the urogenital tract, diabetes mellitus, the use of antibiotics or vaginal antimicrobials in the previous month, presence of an intrauterine device, vaginal intercourse within the latest 3 days, antibiotic or vaginal antimicrobials in the previous month, presence of an intrauterine device, vaginal intercourse within the latest 3 days, any active infection due to Chlamydia, yeast, HPV, HSV-2, or HIV-1 infection. The clinical data of each participant were summarized in Table 1.

Sample collection and preparation

The sample swabs were collected as described by Kin et al. [8]. Two swabs were taken from each woman. One was used for sialidase test, vaginal smear, and saline microscopy, the other was used for bacterial genomic DNA extraction. This vaginal swab was placed in sterile cryovial with 1 ml normal saline, packaged in ice packs and transferred to the laboratory immediately in ice-box, and stored at -80 °C until use.

Extraction of genomic DNA

The bacterial cells retrieved on swabs were vigorously agitated to dislodge cells. The cells were pelleted by centrifugation at 10,000 g for 10 min. Then bacterial genomic DNA was extracted using QIAamp DNeasy Blood and Tissue Kit (QIAGEN, Hilden, Germany; Cat. No. 69504) followed the manufacturer’s instructions. Briefly, the bacterial pellet was suspended in 200 µl of normal saline, and then added 20 µl of proteinase K solution and 200 µl of lysis buffer (Buffer AL). Vertex the mixtures for homogenization and incubated at 56 °C for 30 min. Then, 200 µl of 100% ethanol was added and mixed thoroughly, the mixture was then pipeted into the QIAamp spin column and centrifuged at 8,000 g for 1 min. The QIAamp spin column was placed in a new 2 ml collection micro-tube, and the containing filtrate was discarded. The spin columns were washed with 500 µl buffer AW1 and AW2 provided in the kit. Finally, the genomic DNA was eluted

<table>
<thead>
<tr>
<th>Sample No</th>
<th>Age</th>
<th>BV Test</th>
<th>Clue cell</th>
<th>rod-shaped bacteria</th>
<th>WBC Count</th>
<th>Clinical Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>BV+ Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>124</td>
<td>30</td>
<td>+</td>
<td>+</td>
<td>++</td>
<td>W++</td>
<td>BV</td>
</tr>
<tr>
<td>127</td>
<td>27</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>W++</td>
<td>BV</td>
</tr>
<tr>
<td>130</td>
<td>55</td>
<td>+</td>
<td>+</td>
<td>++</td>
<td>W+++</td>
<td>BV</td>
</tr>
<tr>
<td>131</td>
<td>19</td>
<td>+</td>
<td>+</td>
<td>++</td>
<td>W++</td>
<td>BV</td>
</tr>
<tr>
<td>133</td>
<td>30</td>
<td>+</td>
<td>+</td>
<td>++</td>
<td>W+++</td>
<td>BV</td>
</tr>
<tr>
<td>227</td>
<td>57</td>
<td>+</td>
<td>+</td>
<td>++</td>
<td>W+++</td>
<td>BV</td>
</tr>
<tr>
<td>228</td>
<td>58</td>
<td>+</td>
<td>+</td>
<td>++</td>
<td>W++</td>
<td>BV</td>
</tr>
<tr>
<td>232</td>
<td>29</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>W+++</td>
<td>BV</td>
</tr>
<tr>
<td>245</td>
<td>21</td>
<td>+</td>
<td>+</td>
<td>++</td>
<td>W++++</td>
<td>BV</td>
</tr>
<tr>
<td>251</td>
<td>42</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>W++</td>
<td>BV</td>
</tr>
<tr>
<td>BV- Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>113</td>
<td>29</td>
<td>-</td>
<td>-</td>
<td>Miscellaneous+</td>
<td>W+</td>
<td>PE</td>
</tr>
<tr>
<td>125</td>
<td>21</td>
<td>-</td>
<td>-</td>
<td>Miscellaneous+</td>
<td>W+</td>
<td>PE</td>
</tr>
<tr>
<td>136</td>
<td>28</td>
<td>-</td>
<td>-</td>
<td>Miscellaneous+</td>
<td>W+</td>
<td>PE</td>
</tr>
<tr>
<td>138</td>
<td>30</td>
<td>-</td>
<td>-</td>
<td>Miscellaneous+</td>
<td>W+</td>
<td>PE</td>
</tr>
<tr>
<td>146</td>
<td>32</td>
<td>-</td>
<td>-</td>
<td>Miscellaneous+</td>
<td>W+</td>
<td>PE</td>
</tr>
<tr>
<td>223</td>
<td>22</td>
<td>-</td>
<td>-</td>
<td>Miscellaneous+</td>
<td>W+</td>
<td>PE</td>
</tr>
<tr>
<td>229</td>
<td>32</td>
<td>-</td>
<td>-</td>
<td>Miscellaneous+</td>
<td>W+</td>
<td>PE</td>
</tr>
<tr>
<td>235</td>
<td>31</td>
<td>-</td>
<td>-</td>
<td>Miscellaneous+</td>
<td>W+</td>
<td>PE</td>
</tr>
<tr>
<td>237</td>
<td>35</td>
<td>-</td>
<td>-</td>
<td>Miscellaneous+</td>
<td>W+</td>
<td>PE</td>
</tr>
<tr>
<td>249</td>
<td>31</td>
<td>-</td>
<td>-</td>
<td>W+</td>
<td>PE</td>
<td></td>
</tr>
</tbody>
</table>

*BV test is based on a sialidase test kit. W values are the white blood cell counts per viewing field under microscope. W+ indicates a value greater than 10 and W++, W+++, W++++ greater than 20, 30, and 40 respectively.

PCR amplification

Highly conserved sequences of the Escherichia coli 16S rDNA was used to design the primers for PCR amplification. The forward primer was 8f: 5′-AGAGTTTGAT CMTGGCTCAG-3′, M = A + C. In the European Ribosomal RNA Database, 8f is called BSF8/20 (www.psb.ugent.be/rRNA). The reverse primer was 1492r: 5′-TACG

<table>
<thead>
<tr>
<th>Age</th>
<th>BV Test</th>
<th>Clue cell</th>
<th>rod-shaped bacteria</th>
<th>WBC Count</th>
<th>Clinical Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>249</td>
<td>31</td>
<td>-</td>
<td>W+</td>
<td>PE</td>
<td></td>
</tr>
</tbody>
</table>

16S rDNA gene cloning library construction and sequencing

16S rDNA PCR products were cloned into the pMD19T (Takara, Dalian) vector and transferred to E. coli TOP 10 competent cells (TianGen) according to the manufacturer’s recommendations. The TOP 10 cells were plated onto Luria–Bertani agar plates containing 100 µl of elution buffer (Buffer AE). The concentration of DNA was determined by NanoDrop ND-1000 spectrophotometer (Thermo Electron Corporation, USA). The integrity and size of the extracted DNA were estimated and checked by 1% agarose gel electrophoresis containing 0.5 mg/ml ethidium bromide. All the DNA samples were stored at -20 °C until use.

Table 1: Patient information of volunteer subjects.

with 100 µl of elution buffer (Buffer AE). The concentration of DNA was determined by NanoDrop ND-1000 spectrophotometer (Thermo Electron Corporation, USA). The integrity and size of the extracted DNA were estimated and checked by 1% agarose gel electrophoresis containing 0.5 mg/ml ethidium bromide. All the DNA samples were stored at -20 °C until use.

PCR amplification

Highly conserved sequences of the Escherichia coli 16S rDNA was used to design the primers for PCR amplification. The forward primer was 8f: 5′-AGAGTTTGAT CMTGGCTCAG-3′, M = A + C. In the European Ribosomal RNA Database, 8f is called BSF8/20 (www.psb.ugent.be/rRNA). The reverse primer was 1492r: 5′-TACG

<table>
<thead>
<tr>
<th>Age</th>
<th>BV Test</th>
<th>Clue cell</th>
<th>rod-shaped bacteria</th>
<th>WBC Count</th>
<th>Clinical Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>249</td>
<td>31</td>
<td>-</td>
<td>W+</td>
<td>PE</td>
<td></td>
</tr>
</tbody>
</table>

16S rDNA gene cloning library construction and sequencing

16S rDNA PCR products were cloned into the pMD19T (Takara, Dalian) vector and transferred to E. coli TOP 10 competent cells (TianGen) according to the manufacturer’s recommendations. The TOP 10 cells were plated onto Luria–Bertani agar plates containing 100 µl of elution buffer (Buffer AE). The concentration of DNA was determined by NanoDrop ND-1000 spectrophotometer (Thermo Electron Corporation, USA). The integrity and size of the extracted DNA were estimated and checked by 1% agarose gel electrophoresis containing 0.5 mg/ml ethidium bromide. All the DNA samples were stored at -20 °C until use.

PCR amplification

Highly conserved sequences of the Escherichia coli 16S rDNA was used to design the primers for PCR amplification. The forward primer was 8f: 5′-AGAGTTTGAT CMTGGCTCAG-3′, M = A + C. In the European Ribosomal RNA Database, 8f is called BSF8/20 (www.psb.ugent.be/rRNA). The reverse primer was 1492r: 5′-TACG

<table>
<thead>
<tr>
<th>Age</th>
<th>BV Test</th>
<th>Clue cell</th>
<th>rod-shaped bacteria</th>
<th>WBC Count</th>
<th>Clinical Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>249</td>
<td>31</td>
<td>-</td>
<td>W+</td>
<td>PE</td>
<td></td>
</tr>
</tbody>
</table>

16S rDNA gene cloning library construction and sequencing

16S rDNA PCR products were cloned into the pMD19T (Takara, Dalian) vector and transferred to E. coli TOP 10 competent cells (TianGen) according to the manufacturer's recommendations. The TOP 10 cells were plated onto Luria–Bertani agar plates containing 100 µl of elution buffer (Buffer AE). The concentration of DNA was determined by NanoDrop ND-1000 spectrophotometer (Thermo Electron Corporation, USA). The integrity and size of the extracted DNA were estimated and checked by 1% agarose gel electrophoresis containing 0.5 mg/ml ethidium bromide. All the DNA samples were stored at -20 °C until use.

PCR amplification

Highly conserved sequences of the Escherichia coli 16S rDNA was used to design the primers for PCR amplification. The forward primer was 8f: 5′-AGAGTTTGAT CMTGGCTCAG-3′, M = A + C. In the European Ribosomal RNA Database, 8f is called BSF8/20 (www.psb.ugent.be/rRNA). The reverse primer was 1492r: 5′-TACG

<table>
<thead>
<tr>
<th>Age</th>
<th>BV Test</th>
<th>Clue cell</th>
<th>rod-shaped bacteria</th>
<th>WBC Count</th>
<th>Clinical Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>249</td>
<td>31</td>
<td>-</td>
<td>W+</td>
<td>PE</td>
<td></td>
</tr>
</tbody>
</table>
100 µg/ml ampicillin and 20 µg/ml X-gal and incubated overnight at 37°C. Approximately 200 white, well-isolated colonies were randomly selected from each of library. Single colonies were picked from the agar plates and transferred with sterile tips to the wells of a 96 well plate filled with LB Broth supplemented with 200 µg ampicillin/ml and incubated at 37°C for 3 h. For each sample, at least one 96-well plate containing a library with >90% of the colonies carrying cloned inserts was sequenced by using Applied Biosystems BigDye-terminator chemistry and the Applied Biosystems 3730 DNA Analyzer (Carlsbad, CA) with universal primer M13 reverse –48.

**Sequence processing and data analysis**

The clone sequences were edited and processed by soft package. Raw sequence data were base-calling by PHRED and assigns a quality score as Q20 to each base edited and processed to accurate reads for alignment [12,13]. In all cases, vector and primer sequences were removed and quality trimmed from the clone sequences by using the LUCY program [14,15]. 37 clones of sequences are shorter than 200 bps or empty vector without insert sequence, thus were cut off. The related bacterial 16S rDNA sequences in the RDP II database (http://rdp.cme.msu.edu/) with high Seqmatch scores (>0.9), were selected and used to assign by RDP naïve Bayesian rRNA Classifier for further analysis [15,16]. Analysis of the 8 clone sequences, which did not match, revealed that nearly all of them were probably not unique or unknown phylotypes. Based on the Seqmatch results, the clone sequences could be assigned to the type strains in the RDP II.

**Statistical analysis**

Bray-Curtis distance analysis was used to determine the differences in microbial population profiles. The Bray-Curtis dissimilarity index [17,18], which is equivalent to a doubly weighted form of the Sørensen-Dice dissimilarity index, was calculated according to the equation when \( \sum_{ixi} = 1 \) and \( \sum_{jxj} = 1 \), where \( d_{ij} \) is the dissimilarity or distance between communities \( j \) and \( k \), and \( x_{j} \) and \( x_{k} \) are the fractional or percentage populations of component \( i \) found in communities \( j \) and \( k \). This distance was used to calculate the distance matrix for cluster analysis using the unweighted-pair group method using average linkages in the PHYLIP package. Phylotype diversity of the combined vaginal libraries within each subject was estimated by Simpson's reciprocal and Shannon-Wiener's diversity indices, using the EstimateS software (http://viceroy.eeb.uconn.edu/estimates). The OTU, rarefaction, Chao1 and ACE estimation were analyzed using the mothur [19] (v.1.6.0, http://www.mothur.org/wiki/Main-Page). The statistical significance of differences in microbial community composition, and Shannon and Simpson index between sample categories was determined by SSPP with chi-test and t-test. For bootstrap values, three methods (Fitch, parsimony, and maximum likelihood) with 100 re-samplings were performed using FITCH, DNAPARS, DNAMLK, SEQBOOT, and CONSENSE in PHYLIP.

**Results**

**Sequence analysis**

Clones were sequenced with a single read using M13 rev–48 primer and showed that nearly complete 16S rDNA PCR product was cloned into pMD19T vector. Total of 3245 sequences passed quality control and were used for data analysis; 1893 sequences for BV+ group and 1346 sequences for BV- group. Sequences were classified based on self-similarities rather than matches to an external database and were grouped into operational taxonomic units (OTUs) with cutoffs of each sequence similarity using the MOTHUR package (http://www.mothur.org/wiki/Main-Page) [19]. As described previously [20], rarefaction curves were generated for unique, 1%, 2%, 3%, sequence dissimilarities. For phylogenetic analyses, the core data set of representative sequences with 97% 16S rDNA sequence similarity to define the phylogenetic species. As shown in Figure 1, the rarefaction curve in BV- group almost reached the saturation level. However, there was much more richness in bacterial diversity in BV+ individuals than in BV- individuals at the 97% similarity level. The number of OTUs continued to increase at the 97% or 98% similarity level in BV+ group (Figure 1), which indicated that additional sampling was needed to determine the true microbial diversity in BV+ vaginal community.

The taxonomic assignments of vaginal bacterial community members and the associated diversity estimation for each subject are shown in Table 2. The sequences identified in these groups covered the majority of the clone libraries presented in our detected samples for the Good’s coverage were more than 98.0% for all sequences in the two groups according to the estimated species diversity indices. Furthermore, as estimated by Chao1 and ACE diversity indices, the number of OTUs detected was close to the total number of OTUs for the two groups of vaginal communities analyzed at the 3% dissimilarity level. Based on the number of sequences or OTUs and relative proportions, Chao1 and ACE richness estimators were measured for the number of microbial species in the samples. This also proved that our clone libraries were well-covered during sequencing. Taxon definitions of 99%, 98%, and 97% OTUs were assigned using the Mothur package. The numbers of OTU taxa per subjects with BV of Mean ± SD were 39.8 ± 11.9, 31.2 ± 11.1, 29.4 ± 9.3 versus 16.2 ± 11.6, 11.3 ± 8.5, 11.7 ± 7.8 correspond to the subjects without BV at 99%, 98%, 97% similarity respectively. The OUT taxa detected per subject were dramatically higher (P ≤ 0.001) in subjects with BV than in subjects without BV for all phylotype definitions (Figure 2). Similarly,

![Figure 1: Rarefaction curves for BV+ and BV- groups used to estimate richness at a 97% or 98% similarity level.](image-url)
diversity was higher (P <0.005) in subjects with BV than that without BV. The Shannon diversity index in subjects with BV was 2.79 ± 0.5, and in subjects without BV was 1.56 ± 0.9 based on the 97% similarity taxonomic definition.

The Shannon and Simpson indices [21], for both clone libraries were far from the critical values, Shannon Index (3.401 in BV- group Vs. 4.469 in BV+ group) and Simpson Index (0.068199 in BV- group Vs.0.02544 in BV+ group). The diversity of both groups was abundant; however the diversity was higher in subjects with BV, which the Shannon diversity index was 1.3 times greater than subjects without BV. The vaginal community in BV+ group (Evenness = 0.592) was more even than in BV- group (Evenness=0.472) (Table 2) indicating that bacterial community in BV+ group had greater species diversity.

Differential bacterial abundance found in BV+ and BV- Groups

A significant difference (P <0.001) was found between the subjects with BV and those without BV according to the mean number of taxa in each sample of Mean ± SD with the 11.8 ± 4.44 in BV+ group and 3.9 ± 3.35 in BV- group.

The overall abundance was presented as pie charts in Figure 3. At phylum level, Firmicutes was the most abundant member in BV- group and Firmicutes and Actinobacteria was more abundant in BV+ group (Figure 3A and B). Fusobacteria and Tenericutes were only found in samples from subjects with BV. There were differences in the relative proportions of major groups of bacteria for the taxonomic distribution of phytophylotypes. Taxonomic affiliations and relative proportions of sequences were summarized at the genus (Figure 4A) and phylum (Figure 4B) levels.

Proportions of Actinobacteria, and Bacteroidetes were 140 times and 40 times higher respectively in subjects with BV than those without BV. However Firmicutes in subjects with BV was only 0.62 times of that without BV. The dominant microflora is Lactobacillus iners in subjects without BV, which accounted for 83.73% of all bacterial communities (Figure 4A). In contrast, subjects with BV did not possess a single dominant taxon but rather harbored a diverse array of vaginal bacteria, many present at low relative abundance, including Gardnerella, Atopobium, Prevotella, Gemella, Aerococcus, L.iners, Anaerococcus, Parvimonas, Peptoniphilus, Megasphaera, Dialiste, and Sneathia. From the above, we can see that BV+ group has a higher degree of microbial diversity.

The abundance of bacterial phylotypes from each subject was summarized in Table 3. Several species showed preferential occurrence in subjects of BV+ group with t-test p value < 0.1 and this preference was analyzed by chi-test according to presence of species in the subject (Table 4). Most of these species were more prevalent in BV+ group.
Gardnerella vaginalis was one of the bacteria first found to be associated with BV [22]. In our research, Gardnerella vaginalis was found in 4 subjects in BV+ group, with the occurrence of 10.3%, 17.37%, 0.87% and 15.18% in the subject 124, 131, 228 and 245 respectively. In BV- group, Gardnerella vaginalis was found in only one subject 146 with the occurrence of 0.65%. In this study, the t-test p value for abundance was only < 0.1 and chi-test p value for presence was < 0.5. Although Gardnerella vaginalis was not more prevalent in BV+ group, its presence was more common in BV+ group (Figure 4).

There was a significant difference between BV+ and BV- group in almost genera obtained from the vagina (Table 4). Not surprisingly, the Atopobium, Prevotella, Sneathia (P<0.01 for chi-test, and P<0.05 for t-test) had significantly strong association with BV. Among these predominant genera, Gemella, Anaerococcus, Parvimonas, Peptoniphilus, Dialiste, Megaspheara had weak associations with BV (P<0.05 for only one test). Although they had been detected by previously studies, they were not identified as BV-associated bacteria [5,23-26].

Lactobacillus bacteria were more abundant in BV- group, consisting of 97.6% ± 9.6% and significantly higher than that in BV+ group, which was 12.2% ± 26.6% (P<0.001) (Figure 4). Among the Lactobacillus species, L. iners, L. jensenii, L. crispatus, L. vaginalis and L. gasseri were found with overall occurrence of 83.73%, 3.25%, 10.21%, 0.07% and 0.07% in BV+ group. Only L. iners and L. coleohominis were found with 13.63% and 0.05% in BV+ group. L. iners was more prevalent in BV-group with t-test p value of 0.001 and chi-test p value of 0.001 (Table 3). Chinese women are more likely to be colonized with L. iners than other Lactobacillus species in the vaginal ecosystem. This is consistent with Ravel, that among 96 Asian women, 41 had a vaginal microbiota dominated by L. iners [26]. The diversity of vaginal microbiota in healthy women from the Haerbing, Northern China, compared with the other area of China investigated [27,28]. L. vaginalis was present in subject 125 with 0.41% prevalence, and L. gasseri in subject 229 with 0.94%.

Based on results of classification, four possible novel phylotype microorganisms were found. Two uncultured bacterium classified into the Ruminococcaceae family were apparent to be associate with BV, that uncultured bacterium AY958888 showed significant BV-associate (P<0.05 for chi-test and P<0.01 for t-test), another uncultured Acetivibrio bacterium (P<0.1 for chi-test) which was indifined as BVAB3-Strain 2 by Fredricks [23].

Two additional unidentified bacteria were also found in this study, although not associated with BV. One classified into Coriobacteriaceae bacterium was presented in subject 130, 133 and 232 with prevalence of 1.65%, 0.55% and 2.5%. The other classified into Clostridiales bacterium was presented in subject 131 with 0.52%.

**Principal component analysis**

Principal Component Analysis (PCA) on the genera composition of microbes in each subject that represents an environment was performed using Canoco (Version 4.51) [29]. The clustering of the samples and the vectors contributing to the spread of the data points...
were showed by the biplot in Figure 5. *Lactobacillus*, *Anaerococcus*, *Prevotella* and *Sneathia* contributed to the formation of a cluster, respectively. *Parvimonas*, *Atopobium*, *Gemella*, *Gardnerella*, and *Dialister* altogether contributed to the separation of a cluster. When the clusters from the PCA results were compared with the sialidase based BV test, nine out of ten subjects in BV+ group fall into the same cluster. The only exception was subject 245 with the dominant consist of *L. iners*. Although subject 245 had been treated for BV, the sialidase test showed that the prevalence of *L. iners* was significantly higher in BV+ group than in BV- group. 

Table 4: Genera of bacteria with high overall prevalence or with differential occurrence in BV+ vs BV- group.
In the previous reports, the dominating vaginal microbiota was Lactobacilli [27,28,30-32]. The species Lactobacillus iners, Lactobacillus crispatus, Lactobacillus jensenii, and Lactobacillus gasseri were found in the Asian Japanese, white and black women [30-33] while only the Lactobacillus iners, Lactobacillus crispatus were found in Eastern Chinese (Shanghai and Zhejiang) women [27,28]. In our study, the diversity of vaginal microbiota in healthy women from the Haerbing, Northern China was investigated. L. vaginalis was found in this investigation. Among the Lactobacillus species, L. iners, L. crispatus, L. jensenii, L. gasseri and L. coleohominis were found with overall occurrence of 42.91%, 4.26%, 1.35%, 0.03%, 0.03% respectively. L. iners and L. crispatus were more prevalent with t-test p value of 0.000, 0.237 and chi-test p value of 0.000, 0.023 (Table 4, Figure 4). Our results showed that the vaginal bacterial communities of healthy women are not always dominated by Lactobacillus species. The difference may be caused by the host genetic factors, cultural, behavioral and environmental differences.

In our research, we observed total of 213 taxa across all subjects with BV when taxon defined as 97% OTU definition, but richness estimates of the true number of taxa present ranged from 217 to 236 (95% confidence interval), with an average estimate of 222 taxa (Table 2). Maybe the more samples could really reveal even greater richness and diversity in subjects with BV. While to subjects without BV, our samples were enough because we got 92 taxa with a 97% OTU definition with the total ranged from 93 to 107 (95% confidence interval) (Table 2).

The variability in community membership and structure has important implications for understanding the etiology of BV and for developing diagnostic tools. Our data revealed that the vaginal bacterial communities found in subjects with and without BV are distinctly different. Notably, the structures of these vaginal bacterial communities have high inter subject variability within each clinical group. This is consistent with the reported inter subject variability in the vagina [23] and other areas of the human body [33, 34]. The possible reasons especially why the taxonomic composition of BV-associated bacteria is so different for each subject are: 1) differences in host immune response; 2) differences in geographic location; 3) expression of legends for bacterial attachment to epithelial cells; 4) the chemical and physical environments of the host; 5) intra- and inter specific microbial competition.

We identified several non-lactobacillus bacteria associated with BV, including Atopobium, Prevotella, Gemella, Anaerococcus, Parvimonas and Peptoniphilus, and Dialister, Sneathia. At phylum level, BV-group consists of mostly Firmicutes while BV+ group also has higher abundance in Bacteroidetes and Actinobacteria. We found that species diversity was more richness in BV+ group.

The preferential occurrence of several species in subjects of BV+ group was tested by t-test and analyzed strictly according to presence of absence of the species in the subject by chi-test. The results showed that most of these species were more prevalent in BV+ group. Gardnerella vaginalis was first found to be associated with BV. Gardnerella vaginalis was not more prevalent in BV+ group, but its presence was more common in BV+ group. Notably, Among the Lactobacillus species, Only L. iners and L. coleohominis were found with overall occurrence of 12.16% and 0.03% for BV + group. At genus level, the subjects in BV- group has fewer bacterial genera (3.9 ± 3.35) than BV+ groups (11.8 ± 4.44). Several genera were found predominantly including Atopobium, Porphyromonas, Prevotella, Parvimonas, Peptoniphilus, Dialister, Aerococcus, Acetivibrio BVAB3, Leptotrichia and Sneathia.
in BV+ group, this had been previously identified as BV-associated bacteria [23-26].

In previous reports about the microbial ecosystem in human gastrointestinal tract showed that the host immune system and the cell surface receptors may influence human–microbial symbioses and dictate adhesion of microorganisms [35,36]. McFall-Ngai [37] put forward the point of view that adaptive immunity plays a role in recognizing and managing the complex community composition of beneficial microorganisms in vertebrates. Based on the above, we proposed that similar selectivity may occur in vagina and the local vaginal immune system may play an important role in determining the composition of vaginal microbial communities.

Conclusions

The data presented here about the composition and richness of the vaginal microbial ecosystem in BV and health state will provide insight into the etiology of BV and the basis for the prevention and control of BV of microbial ecosystem in women.

Acknowledgements

This work was supported by the start grant from Guangdong province and Southern Medical University (to Shuhong Luo). We thank Yilong Luo, Yanghui Zheng and Yang Li for the sample collection. We thank Dr. Mengfei Ho and Dr. Brenda A. Wilson for critical reading of the manuscript.

Authors’ contributions

SL and ML participated in the design of the study, the analysis of the data. YL, WH, CH, WX, SZ, JT and XL participated in the sample collection, DNA extraction, PCR library construction, sequencing and interpretation of the data. SL, ML and YL wrote the manuscript. All authors read and approved the final manuscript.

References


