**RESULTS**

**Table 1. Subject Demographics and cervicovaginal lavages microbial composition**

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Nugent Score</th>
<th>Incident STI</th>
<th>Vaginal pH</th>
<th>Vaginal race</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>343442</td>
<td>6</td>
<td>No</td>
<td>6.5</td>
<td>BVI</td>
<td>37.8</td>
</tr>
<tr>
<td>342826</td>
<td>10</td>
<td>No</td>
<td>5.3</td>
<td>Asym</td>
<td>34.5</td>
</tr>
<tr>
<td>342587</td>
<td>8</td>
<td>No</td>
<td>5.5</td>
<td>Asym</td>
<td>27.5</td>
</tr>
<tr>
<td>342495</td>
<td>9</td>
<td>No</td>
<td>5.5</td>
<td>Asym</td>
<td>30</td>
</tr>
<tr>
<td>342491</td>
<td>7</td>
<td>No</td>
<td>5.5</td>
<td>White</td>
<td>37.8</td>
</tr>
</tbody>
</table>

**Figure 3. Circular representation of a closed “Ca. Lachnocurva vaginae” genome (left).** Produced by PacBio Sequel II Sequencing. The cMAG of “Ca. Lachnocurva vaginae” is 1,649,642 bp in size with 31.8% GC content, encodes 1,578 genes, and means with long read coverage of 124X. Whole-genome alignments of cMAG and MAGs (right). Regions of genomic fluidity exist around genomic islands and phase.

**Figure 4. Genomic features of BVAB1.** Genes coding for the following functions were conserved across all genomes: Methyl-accepting chemotaxis (MCRT) and subsequent flagella assembly, Sec-SRP secretion systems, choline import and metabolism, bacteriocin production and export, mannose, fructose, and L-ascorbate influx and metabolism resulting in pyruvate and D-lactate, biosynthesis of only 4 amino acids.

**Figure 5. Accessory gene-encoded proteins and flagellar and extrachromosomal element functions.**

**CONCLUSIONS**

We present here a circularized MAG of “Ca. Lachnocurva vaginae” and six MAGs of the candidate species, previously known as BVAB1, an important member of the human vaginal microbiota associated with bacterial vaginosis and other adverse outcomes. Short-read metagenomic assemblies do not perform well and lead to sub-optimal assemblies with missing regions, whereas long-read metagenomic assemblies are promising and can generate circularized metagenome-assembled genomes, as shown in this study. Our inability to culture this bacterium has limited our understanding of its ecological role in the vaginal microbiome. We have shown that the MAGs of “Ca. Lachnocurva vaginae” have genomic potential for chemotaxis, and is likely capable of resisting several antibiotics via drug efflux systems. Our analysis indicates this candidate species may contribute to the fibrous odor characteristic of bacterial vaginosis through the production of TMA from choline. This crucial genomic data could be used in metabolic modeling experiments to define a culture medium suitable for the cultivation of “Ca. Lachnocurva vaginae”, a critical step to further understand its role in the vaginal ecosystem.

**ACKNOWLEDGMENTS**

We are grateful to the NIH F32 AI136400 NSF awarded to J. Holm and R15 AI115439 awarded to R. Brotman for funding this work. We are also grateful to Elizabeth O’Hanlon for providing the Gram stain images.

**REFERENCES**

Bacterial Vaginosis Associated bacterium 1 (BVAB1) is an as-yet uncultured bacterial species found in the human vagina that belongs to the family Lachnospiraceae within the order Clostridiales. As its name suggests, this bacterium is often associated with bacterial vaginosis (BV), a common vaginal disorder that has been shown to increase a woman’s risk for HIV, Chlamydia trachomatis, and Neisseria gonorrhoeae infections as well as preterm birth. Further, BVAB1 is associated with the persistence of BV following metronidazole treatment, increased vaginal inflammation, and adverse obstetrics outcomes. There is no available complete genome sequence of BVAB1, which has made it difficult to mechanistically understand its role in disease.

We present here a circularized metagenome-assembled genome (cMAG) of BVAB1 as well as a comparative analysis including an additional six metagenome-assembled genomes (MAGs) of this species.

These sequences were derived from cervicovaginal samples of seven separate women. The cMAG is 1.649 Mb in size and encodes 1,578 genes.

We propose to rename BVAB1 to “Candidatus Lachnocurva vaginae” based on phylogenetic analyses, and provide genomic evidence that this candidate species may metabolize D-lactate, produce trimethylamine (one of the chemicals responsible for BV-associated odor), and be motile. The cMAG and the six MAGs are valuable resources that will further contribute to our understanding of the heterogeneous etiology of bacterial vaginosis.

**METHODS**

The cMAG source sample was a self-collected Copan Eswab re-suspended in 1 mL Amies transport medium, frozen at -20°C for no more than a week, and then transferred to -80°C until analyzed. DNA was extracted using the MasterPure DNA purification kit (Lucigen) with two phenol/chloroform cleanups prior to DNA precipitation and quantified on a TapeStation 2200 instrument run with a Genomic DNA tape (Agilent). The sequencing library was prepared from 50 ng of DNA in 200 µl with SMRTbell Template Prep Kit K10 and was size-selected on a BluePippin (Sage Science) with a cutoff of 5 kb. The library was barcoded and sequenced as part of a multiplexed run with four other unrelated samples.

Sequencing was performed on a Pacific Sequel II instrument with an 8K single read 450 flowcell and a single end run of 450 cycles. The raw data was demultiplexed with (version 1.9.0) using default parameters except for minimum barcode score set at 26 and a minimum read length of 50 after clipping of the barcode was enforced. Both tools are part of the SMRTLink 5.0.1 software package with updated CCLS version 5.0.1. Human reads were detected using phage of v4.f (Pacific Biosciences) and the human genome build 38 (GRCh38p12). Remaining reads were corrected and assembled via Canu v1.7 and the “--pacbio” protocol (Koren et al., 2017). The largest resulting contig was 0.6 Mb in length, much larger than the second longest contig (450 bp). Four copies of the 15S rRNA gene were detected on this contig and were identical to the existing BVAB1 15S rRNA gene reference (AY724739 (Fredrick et al., 2008)). The contig had 6.5 kb of overlapping ends and was determined to be circular by Canu.

The other MAGs were from cervicovaginal lavage samples from six participants collected as part of the NIH Longitudinal Study for Vaginal Flora (Kalosanof et al., 2004). Gram-stain smears were prepared for Nugent scoring as per (Nugent et al., 1995). DNA was extracted from 200 µL of lavage fluid using the MagAttract Microbial DNA Kit (QIAGEN Inc.). Metagenomic libraries for the six samples from the LSV study were prepared using the KAPA HyperPlus Kit (Kapa Biosystems) with KAPA Single-Index Adapter Kit Set B. A fixed volume (35 µL) of genomic DNA was used as input. Libraries were sequenced (8 libraries per lane) on the Illumina HiSeq 4000 instrument using the 150 bp paired-end protocol, quality filtered with Trimmomatic, human reads removed, and assembled using SPAdes and the cMAG in the trusted-contigs option.

Genes were annotated with the IGS Prokaryotic Annotation pipeline. Pfage and genomic islands were detected using PHASTER (Zhou et al., 2011) and IslandViewer4 (Bertelli et al., 2017), respectively.