Comparative Metagenome-Assembled Genome Analysis of "*Candidatus* Lachnocurva vaginae", Formerly Known as Bacterial Vaginosis-Associated Bacterium–1 (BVAB1)

Johanna B. Holm [1,2], Michael T. France [1,2], Bing Ma [1,2], Elias McComb [1], Courtney K. Robinson [1], Aditya Mehta [1], Luke J. Tallon [1], Rebecca M. Brotman [1,3] and Jacques Ravel [1,2]*

[1] Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD, United States, [2] Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, MD, United States, [3] Department of Epidemiology and Public Health, University of Maryland School of Medicine, Baltimore, MD, United States

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. BVAB1 has been further 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 was obtained from a metagenome sequenced with long-read technology on a PacBio Sequel II instrument while the others were derived from metagenomes sequenced on the Illumina HiSeq platform. 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 and metabolomic 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.

Keywords: women's health, gynecology, microbial genomics, odor, vaginal microbiome

INTRODUCTION

Bacterial vaginosis (BV) is a common vaginal infection affecting approximately 30% of US reproductive-aged women, with both African- and Mexican-Americans disproportionately afflicted (Allsworth and Peipert, 2007; Koumans et al., 2007). Though treatable with antibiotics, BV has a high rate of recurrence (Bradshaw et al., 2006; Schwebke and Desmond, 2007). The microbiological diagnosis of BV is established using Nugent scoring of Gram-stained vaginal smears and is defined by a low abundance of *Lactobacillus* spp. morphotypes and a wide array of strict and facultative Gram-negative anaerobes (Nugent et al., 1991). The clinical diagnosis of BV is made when 3 of the 4 Amsel's criteria are met: vaginal pH > 4.5, homogenous vaginal discharge, a positive whiff test, and the presence of clue cells upon wet mount examination (Amsel et al., 1983). Aside from the burdensome symptoms of vaginal discharge and fishy odor, BV is also associated with increased risk to adverse health outcomes including preterm birth (Leitich et al., 2003), increased risk of sexually transmitted infections acquisition and transmission, including HIV (Taha et al., 1998; Cherpes et al., 2003; Ness et al., 2005; Atashili et al., 2008; Cohen et al., 2012) and pelvic inflammatory disease (Ness et al., 2004).

A critical step along the path to understanding the ecology and pathogenic potential of a bacterial species is the characterization of its genome. Yet many of the bacteria associated with BV have thus far been uncultivable, further complicating genome sequencing efforts. BV-associated bacterium 1 (BVAB1) is one such organism for which there is limited genetic information available. BVAB1 was first identified by Fredricks et al. (2005) using 16S rRNA gene amplicons Sanger sequencing of samples associated with BV, and has eluded cultivation efforts since. Based on the sequence of its 16S rRNA gene, BVAB1 belongs to the family *Lachnospiraceae* (Muzny et al., 2014) and has often been misidentified as belonging to the genus *Shuttleworthia* (Lamont et al., 2011; Petrova et al., 2013). Interestingly, Gram-negative curved rods designated *Mobiluncus* morphotypes on Gram stain in Nugent scoring have been shown to likely be BVAB1 (Srinivasan et al., 2013). Further, vaginal communities in which BVAB1 16S rRNA gene sequence is detected have been associated with vaginal inflammation and persistent BV in African women (Lennard et al., 2018). BVAB1 remains uncultured and aside from detection of this species via partial 16S rRNA gene amplicon sequencing, little is known about its metabolism, pathogenic potential, or ecology in the vaginal environment, especially during BV. Further understanding of the genetic and physiological properties of BVAB1 will help to dissect the complex etiology of BV. Previously, a 94 contig BVAB1 metagenome-assembled genome from short-read sequencing was produced (Fettweis et al., 2019).

In this study, we characterize the first circularized metagenome assembled genome (cMAG) of BVAB1 constructed from a metagenome sequenced using the PacBio Sequel II long read platform. In addition, we compare this cMAG to an additional six metagenome-assembled genomes (MAGs). All genomes originate from different women with symptomatic or asymptomatic BV. Based on phylogenetic analysis of full-length 16S rRNA gene sequences obtained from the genomic assemblies, we propose to rename this bacterium "*Candidatus* Lachnocurva vaginae".

MATERIALS AND METHODS

Sample Collection

Vaginal samples used in this study were identified as containing a high relative abundance (> 60%) of BVAB1 using 16S rRNA gene amplicon sequencing of the V3–V4 regions as previously reported (Holm et al., 2019). Cervicovaginal lavages from six participants that were collected as part of the NIH Longitudinal Study for Vaginal Flora (LSVF) (Klebanoff et al., 2004) by washing the vaginal walls with 3 mL sterile, deionized water, aspiration from the vaginal vault via pipette and stored at −80°C, were included in this study. Gram stain smears were prepared for Nugent scoring as previously described (Nugent et al., 1991). DNA was extracted from 200 µL of lavage fluid using the MagAttract Microbial DNA Kit (QIAGEN Inc., Germantown MD) automated on a Hamilton Star robotic platform according the manufacturer recommendations. DNA was eluted in a final volume of 110 µL nuclease-free water.

An additional swab sample collected as part of the UMB-HMP study was used in this study (Ravel et al., 2013). The swab was self-collected by a participant using a Copan Eswab re-suspended into 1 mL Amies transport medium (ESwab, Copan Diagnostics Inc.), frozen at −20°C for no more than a week, and then transferred to −80°C until analyzed. High molecular weight DNA was extracted from this sample using the MasterPure DNA purification kit (Lucigen) with two phenol/chloroform cleanups prior to DNA precipitation. DNA extraction was quantified on a TapeStation 2200 instrument run with a Genomic DNA tape (Agilent).

Metagenomic Library Construction and Sequencing on the PacBio Sequel II Platform

The extracted DNA from the swab collected as part of the UMB-HMP study (Ravel et al., 2013) was found to be of sufficient concentration (5.49 ng/µL in 200 µL) for long-read sequencing using the Pacific Biosciences Sequel II platform (Pacific Biosciences). The sequencing library was prepared with SMRTBell Template Prep Kit 1.0 and was size-selected on a BluePippen (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 PacBio Sequel II instrument with an 8M cell loaded at 60 pM at the Genomic Resource Center of the University Maryland School of Medicine.

Long Read Quality Filtering, Host-Read Removal, and cMAG Construction

Raw reads were demultiplexed with *lima* (version 1.9.0) using default parameters except for minimum barcode score set at 26 and a minimum read length of 50 bp after clipping of the barcode was enforced. Both tools are part of the SMRTLink 6.0.1 software package with updated CCS version 3.4.1. Human reads were detected using pbalign v0.4.1 (Pacific Biosciences) and the

human genome build 38 (GRCh38.p12). Remaining reads were corrected and assembled via Canu v1.7 and the "-pacbio-raw" protocol (Koren et al., 2017). The largest resulting contig was ca. 1.6 Mb in length, much larger than the second longest contig (ca. 600 kb). Four copies of the 16S rRNA gene were detected on this contig and were identical to the existing BVAB1 16S rRNA gene reference AY724739 Fredricks et al., 2005 (NEJM). The contig had 6.5 kb of overlapping ends and was determined to be circular by Canu. This contig was then manually circularized using Geneious version 2019.2.1 (Galens et al., 2011), searched for the *dnaA* gene, and rotated so that *dnaA* was the first gene (Kearse et al., 2012). The circularized metagenome-assembled genome (cMAG) was annotated using the IGS Prokaryotic Annotation Pipeline (Galens et al., 2011). Translated gene sequences were assessed against KEGG using the BlastKOALA algorithm for insights into "*Candidatus* Lachnocurva vaginae" metabolism (Kanehisa et al., 2016). Bacteriophages were detected using PHASTER (Zhou et al., 2011; Arndt et al., 2016). Completeness and contamination of the cMAG was analyzed using CheckM version 1.0.18 (Parks et al., 2015) and the taxonomy_wf flag specifying Order *Clostridiales* and rerun specifying Family *Lachnospiraceae*. Metabolic reconstruction was examined with the cMAG using the metabolic modeling function in KBase (Arkin et al., 2018). Complete media for gapfilling and a Gram-positive template were used.

Metagenomic Library Construction and Sequencing on the Illumina HiSeq 4000 Platform

Metagenomic libraries for the six samples from the LSVF study were prepared using the KAPA HyperPlus Kit (Kapa Biosystems) with KAPA Single-Indexed Adapter Kit Set B. A fixed volume (35 µL) of genomic DNA was used as input, and libraries were prepared following the manufacturer's protocol with modifications based on their amount of input DNA as in **Supplemental Data Sheet 1**. For samples with 0.5 or 0.2 ng input DNA, the fragmentation enzyme was diluted 1:2 or 1:5 with water. All samples were fragmented at 37°C for 5 min. Adapter concentrations varied according to the input DNA as listed in **Supplemental Data Sheet 1**, and the adapter ligation was carried out overnight at 4°C for all samples. The post-ligation cleanup was performed with 0.8x Ampure XP beads (Beckman Coulter, Indianapolis IN) and 20 µL of sample was used in library amplification. Amplification library cycles varied by input DNA as listed in **Supplemental Data Sheet 1**. Post-amplification cleanup was performed with 1x Ampure XP beads; libraries with remaining adapter dimer peaks were cleaned a second time. The final elution was in 25 µL of nuclease-free water. Libraries were run on a TapeStation instrument with a D1000 tape (Agilent) to assess quality and concentration. Libraries were sequenced (8 libraries/lane) on an Illumina HiSeq 4000 instrument using the 150 bp paired-end protocol.

Short Read Quality Filtering, Host-Read Removal, and Metagenomic-Assembled Genome Reconstruction

Metagenomic reads were quality filtered using Trimmomatic v0.36 (Bolger et al., 2014) to remove sequencing adapters allowing for 2 mismatches, a palindromic clip threshold of 30, and a simple clip threshold of 10. Bases with quality scores < 3 were removed from the beginning and end of reads combined with a 4 bp sliding window which trimmed a read if the average quality score within that window fell below 15. Reads < 75 bp in length were removed. Human reads were detected by mapping to the human genome build 38 (GRCh38.p12) with Bowtie 2 v2.3.4.1 and default settings (Langmead and Salzberg, 2012), and removed using samtools v1.9 (Li et al., 2009) and bedtools v2.27.1 (Quinlan and Hall, 2010; Quinlan, 2014) (see **Supplemental Data Sheet 2** for specific scripts). Metagenomic assemblies were produced using SPAdes genome assembler v3.13.0 (Bankevich et al., 2012) with the careful setting. Resulting

TABLE 1 | Participants demographics and cervicovaginal lavage microbial compositions for samples used in metagenomic reconstruction of the "*Ca.* Lachnocurva vaginae" circularized metagenome-assembled genome using long-read (*) and shotgun sequencing (*).

Sample ID	Age	Race	Microbiota Relative Abundances of Vaginal Samples (16S rRNA Gene –V3V4)	BV Status	Nugent Score	Vaginal pH	Clinician-Observed Discharge	Positive Whiff Test
UAB071*	22	Black		Asym	8	5	-	Yes
Y3207	37	Black		No	6	6.5	No	No
Y2266	34	Black		Asym	10	5.3	Mild visible—Gray/White	Yes
Y2337	36	Black		Asym	8	5.5	Mild visible—Gray/White	Yes
Y3255	24	Black		Asym	9	5.5	No	Yes
Y2624	27	Black		Asym	7	5.5	Mild visible—"Mucousy"	Yes
Y2694	34	White		Sym	7	5.5	Mild visible—Gray/White	Yes

■, *Ca. Lachnocurva vaginae*; ■, *Gardnerella vaginalis*; ■, *Megasphaera*; ■, *Sneathia sanguinegens*; ■, *Prevotella*; ■, *Atopobium vaginae*; ■, *Prevotella amnii*;
■, *Mageeibacillus indolicus*; ■, *Other*.
Other (gray, includes all taxa except the 8 most abundant taxa). Asym, asymptomatic BV; Sym, symptomatic BV.

contigs were aligned to the cMAG using NUCmer with the –mum setting and allowing for 5,000 bp breaklength, filtered to remove aligned contigs < 100 bp, and minimum contig coverage of 25%. MAGs were annotated with the Live Annotate & Predict tool from Geneious version 2019.2.1 (Kearse et al., 2012) using the "*Ca.* Lachnocurva vaginae" cMAG as annotation
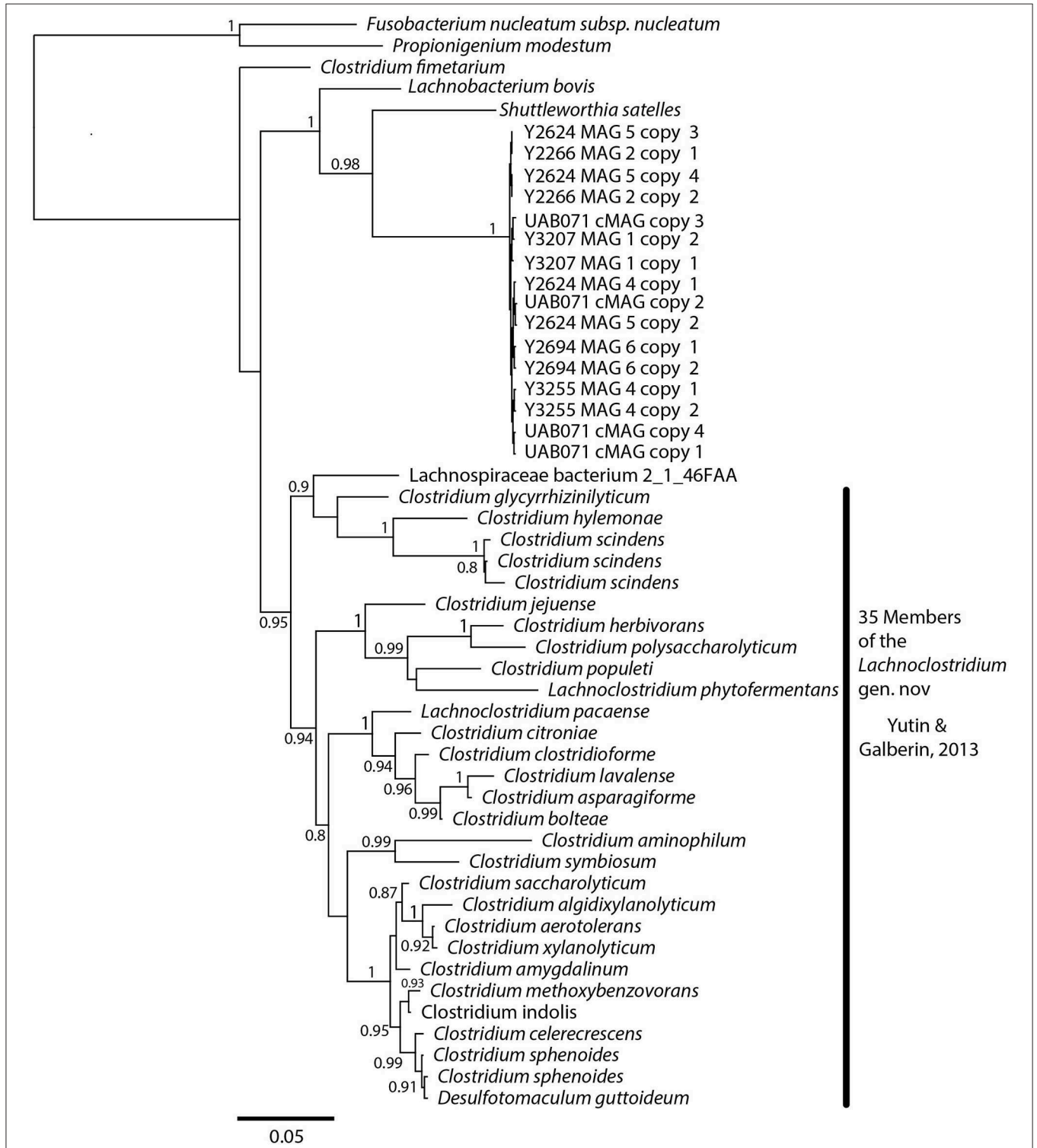

FIGURE 1 | "*Ca.* Lachnocurva vaginae" is related to *Shuttleworthia satelles*; the full-length 16S rRNA gene sequences are 89–90% identical.

source (sequence similarity of $\geq$ 95% required). Genes with no similarity to the cMAG were annotated with Prokka v1.13 (Seemann, 2014). Phages were detected using PHASTER (Zhou et al., 2011). A circle plot was constructed with the BLAST Ring Image Generator v0.95 (Alikhan et al., 2011), and the annotated cMAG.

Metabolomic Analysis

Metabolomics analysis was conducted as previously described (Nelson et al., 2018) by Metabolon Inc. using 200 µL of each LSVF lavage sample used in this study or 200 µL of a frozen dry swab eluted in 1 ml of PBS for the sample from the UMB-HMP study (Ravel et al., 2013). The abundances of 561 compounds were quantified using Ultrahigh Performance Liquid Chromatography-Tandem Mass Spectroscopy (UPLC-MS/MS). Quantities were corrected for instrument block variability and reported as normalized area-under-the-curve estimates. Figures were generated using ggplot2 v3.2.0 (Wickham, 2009).

Comparative Genome and Phylogenetic Analyses

Average nucleotide identities (ANI) were calculated using FastANI v1.1 with fragment lengths of 1,000 bp, which is about the mean length of coding sequences (Jain et al., 2018). Maximum-likelihood phylogenetic trees of full-length 16S rRNA gene sequences alignments were generated with MUSCLE v3.8.425 using default parameters (Edgar, 2004) and FastTree v2.1.11 using the Generalized Time Reversible model and default parameters (Price et al., 2009, 2010) in Geneious v11.0.3 (Kearse et al., 2012). The tree was rooted with *Fusobacterium nucleatum* (AJ133496) and *Propionigenium modestum* (X54275) (Domingo et al., 2009), and members of the genus *Lachnoclostridium* were included as neighbors (Yutin and Galperin, 2013) as well as 16S rRNA genes from *Shuttleworthia satelles* (NR_028827), and *Lachnospiraceae* bacterium 2_1_46FAA (NR_025127). In addition, the cMAG was submitted for taxonomic placement using the TrueBacID tool from EZBioCloud (Yoon et al., 2017). cMAG and MAG pseudomolecules were aligned and visualized using AliTV v1.0.6 (Ankenbrand et al., 2017). Genes were ordered by gene synteny of the cMAG. To examine homology to other reference sequences, the cMAG was aligned to the NCBI protein reference

database (downloaded 1/28/2019), as well as the genomes of *Shuttleworthia satelles* (NZ_ACIP00000000), *Lachnobacterium bovis* (GCF_900107245.1), and *Lachnospiraceae* bacterium 2_1_46FAA (ADLB02000001.1) using BLAST v2.8.1+. The top 3 hits were chosen by lowest e-value, highest percent identity, and longest alignment length, in that order. Genomic and pathogenicity islands were explored using IslandViewer4 (Bertelli et al., 2017).

RESULTS AND DISCUSSION

Participant Information

All six participants in the LSVF study were of reproductive age and had high Nugent scores (6–10) (**Table 1**). Four were diagnosed with asymptomatic Amsel-BV, one with symptomatic Amsel-BV, and one was negative for Amsel-BV (Mckinnon et al., 2019). The woman who participated in the UMB-HMP study (Ravel et al., 2013) had a Nugent score of 8 and was diagnosed with asymptomatic Nugent-BV. The vaginal microbiota of these 7 samples had >60% BVAB1 as defined by 16S rRNA gene V3-V4 amplicon sequencing (Ravel et al., 2013).

Phylogenetic Analysis

A phylogenetic analysis using the full-length 16S rRNA genes extracted from all MAGs revealed BVAB1 belongs in the *Clostridales* family *Lachnospiraceae*, and that *S. satelles* is the closest known relative (**Figure 1**), though nucleotide identity is only 89.2%. Similar results were obtained from EZBioCloud's TrueBacID (**Supplemental Data Sheet 3**). We therefore propose a new candidate species for BVAB1: "*Candidatus* Lachnocurva vaginae", which represents the phylogenetic placement, the curved morphology of the cells, and the source of this cMAG. Considering the entire cMAG, the ANI between "*Ca.* Lachnocurva vaginae" and the *S. satelles* draft genome (NZ_ACIP00000000) was 81.85%. Relative to each other, "*Ca.* Lachnocurva vaginae" 16S rRNA genes were >99.7% identical and the cMAG and MAG average nucleotide identities were also high (98.6–99.2%, **Table 2**).

Overall Genomic Features

The cMAG of "*Ca.* Lachnocurva vaginae" is 1,649,642 bp in size with 31.8% GC content, encodes 1,578 genes and was

TABLE 2 | Average Nucleotide Identity (ANI) between the "*Ca.* Lachnocurva vaginae" circularized metagenome-assembled genome, each metagenome-assembled genome, and the *Shuttleworthia satelles* draft genome NZ_ACIP00000000.

		"*Ca.* Lachnocurva vaginae" MAG ID						
		Y2694_MAG_6	Y2624_MAG_5	Y3255_MAG_4	Y2337_MAG_3	Y2266_MAG_2	Y3207_MAG_1	UAB071
"*Ca.* Lachnocurva vaginae" MAG ID	Y2624_MAG_5	99.2	–	–	–	–	–	–
	Y3255_MAG_4	99.0	99.01	–	–	–	–	–
	Y2337_MAG_3	98.95	98.99	98.97	–	–	–	–
	Y2266_MAG_2	99.06	99.01	98.93	99.05	–	–	–
	Y3207_MAG_1	98.98	98.95	99.07	98.99	99.06	–	–
	UAB071	99.15	99.00	98.92	99.09	99.13	98.98	–
	Shuttleworthia satelles	80.80	81.90	81.51	81.63	81.30	81.86	81.85

estimated to be 99.1% complete and 0% contaminated at the order level and 90.9% complete and 0% contaminated at the family level. The relatively low completion estimate at the family level was due to 22 missing marker genes out of 312 (**Supplemental Table 2**). Mean coverage of the cMAG assembly by long reads was 124X (range: [26–285]). The "*Ca.* Lachnocurva vaginae" MAGs constructed using the short-read Illumina HiSeq 4000 metagenomic sequence data were 1.48–1.62 Mb in size (mean: 1.57 Mb) with a number of contigs ranging from 26 to 152 (mean N50: 99,943 bp) and 31.6% GC (**Table 3**). Genome
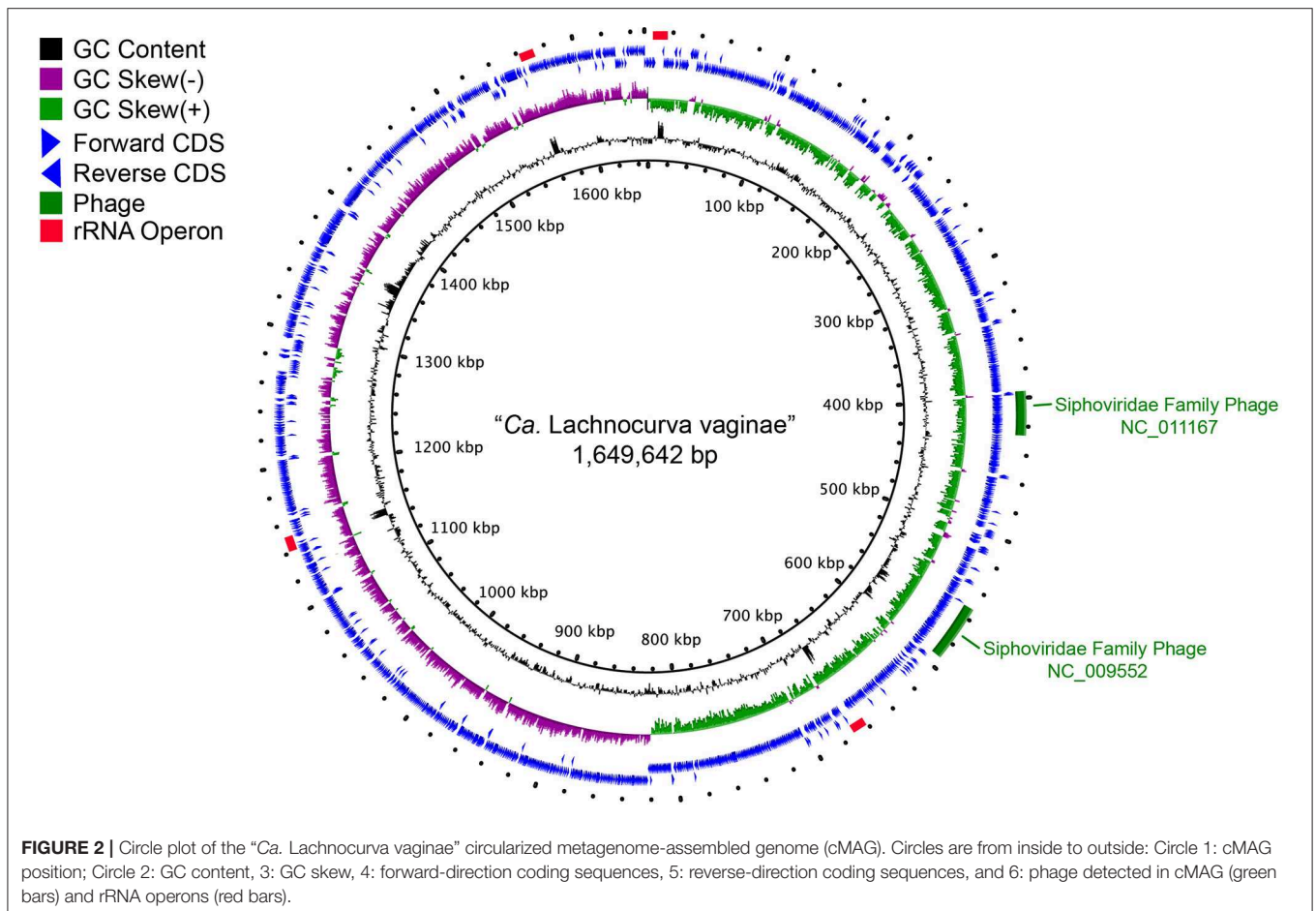
annotation indicated the presence of four complete rRNA operon copies (**Figure 2**) and 42 tRNA genes in the cMAG. As expected, partial rRNA operons were identified in the Illumina-based MAGs (**Supplemental Image 1**).

Genomic Features of "*Ca.* Lachnocurva vaginae"

Metabolic modeling and reconstruction of the "*Ca.* Lachnocurva vaginae" cMAG contained 122 genes, 480

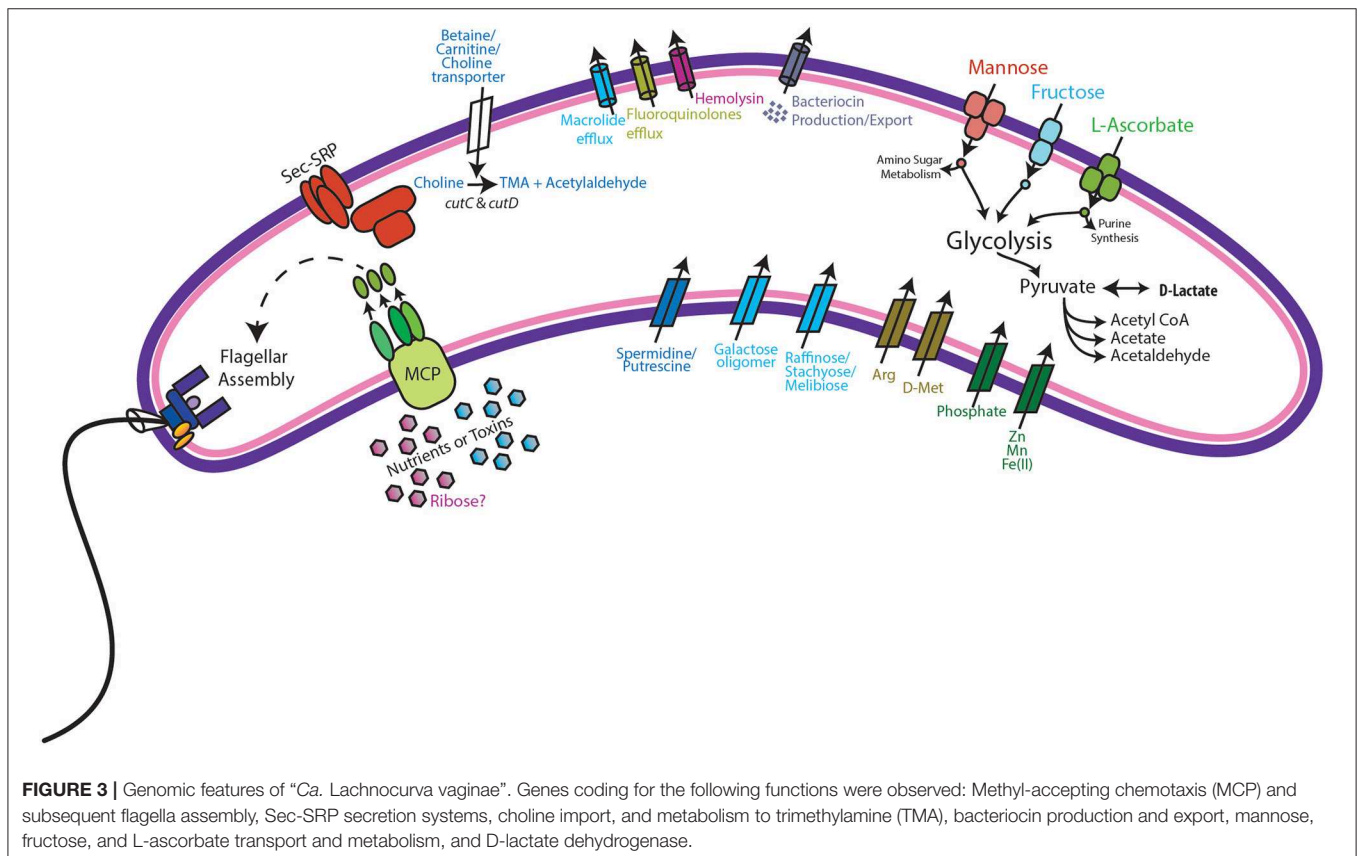**TABLE 3** | "*Ca.* Lachnocurva vaginae" MAG assembly characteristics.

"*Ca.* Lachnocurva vaginae" MAG ID	No. CDS	Total Length (Mb)	N50	%GC	No. Contig	No. tRNAs	No. 5S rRNA	No. 16S rRNA	No. 23S rRNA
UAB071	1,578	1.64	–	31.8	1	42	4	4	4
Y3207_MAG_1	1,463	1.60	42,408	31.6	152	41	1	1	1
Y2266_MAG_2	1,322	1.48	297,822	31.3	26	22	1	1	0
Y2337_MAG_3	1,492	1.62	60,576	31.6	143	42	0	0	0
Y3255_MAG_4	1,433	1.59	80,762	31.5	49	38	1	1	0
Y2624_MAG_5	1,459	1.61	168,048	31.8	31	43	2	2	2
Y2694_MAG_6	1,379	1.52	156,872	31.5	30	28	1	1	0

FIGURE 2 | Circle plot of the "*Ca.* Lachnocurva vaginae" circularized metagenome-assembled genome (cMAG). Circles are from inside to outside: Circle 1: cMAG position; Circle 2: GC content, 3: GC skew, 4: forward-direction coding sequences, 5: reverse-direction coding sequences, and 6: phage detected in cMAG (green bars) and rRNA operons (red bars).

reactions, 581 metabolites, and no mass imbalance was found. Of the 1,578 genes detected, 255 genes had best blast matches of 80% sequence identity covering > 80% of the query gene. Of these, 165 matches were within the order *Clostridiales*, and 90 were not (**Supplemental Image 2**). Genes encoding transporters for mannose, fructose, and L-ascorbate were identified in all MAGs (**Figure 3** and **Supplemental Table 1, LCVA_199-201, LCVA_1491-1494**), as were complete pathways for glycolysis, pyruvate oxidation, and the non-oxidative phase of the pentose phosphate pathway. Mannose was noticeably absent, or below the limit of detection, in the metabolome of all the samples, but present in other BV-like samples that did not contain a high relative abundance of the candidate species (**Figure 4**). We found "*Ca.* Lachnocurva vaginae" to have the genetic capability to uptake and metabolize mannose (**LCVA_1184**), suggesting mannose could also be a carbohydrate source for the candidate species (**Figure 3**).

The D-lactate dehydrogenase gene (**LCVA_41**), but not an L-lactate dehydrogenase gene was also observed in all MAGs. This result was unexpected, as the production of D-lactate in the vaginal environment is thought to be a key and somewhat unique feature of certain *Lactobacillus* spp. (Witkin et al., 2013; France et al., 2016). Metabolomic data for each sample were explored for the presence of lactate to provide evidence supporting this genomic finding. While lactate was observed in most samples, its abundance was substantially lower than that from representative samples that were dominated by *Lactobacillus crispatus*, a known D-lactate producer (**Figure 4**). Thus, either the enzyme exhibits lower activity in "*Ca.* Lachnocurva vaginae" or is instead used in the reverse reaction to consume *Lactobacillus* spp.-produced D-lactate to produce pyruvate. This is a strategy described for another vaginal anaerobic Gram-negative coccus, *Veillonella parvula*, which is able to grow on lactate as the sole source of carbon (Gronow et al., 2010), and would possibly contribute to an increase in vaginal pH, thereby creating a more favorable growth environment. However, for lactate to convert to pyruvate in anaerobic bacteria, an electron-bifurcating complex with an electron transfer flavoprotein alpha and beta subunit (EtfAB) would be necessary to make the reaction favorable (Weghoff et al., 2015). An EtfAB homolog was not observed in the "*Ca.* Lachnocurva vaginae" cMAG, suggesting that another mechanism may be involved. Instead, succinate was the most abundant short chain fatty acid in the communities from which "*Ca.* Lachnocurva vaginae" MAGs originated (**Figure 4**). Not surprisingly, succinate has been associated with bacterial vaginosis, a condition in which "*Ca.* Lachnocurva vaginae" is often found (Srinivasan et al., 2015). Interestingly, all "*Ca.* Lachnocurva vaginae" MAGs lacked genes for all steps of the tricarboxylic acid cycle, indicating it is unable to produce succinate via this pathway. It is possible that "*Ca.* Lachnocurva vaginae" produces metabolite(s) other species in the community can convert into succinate or produces it via a yet to be
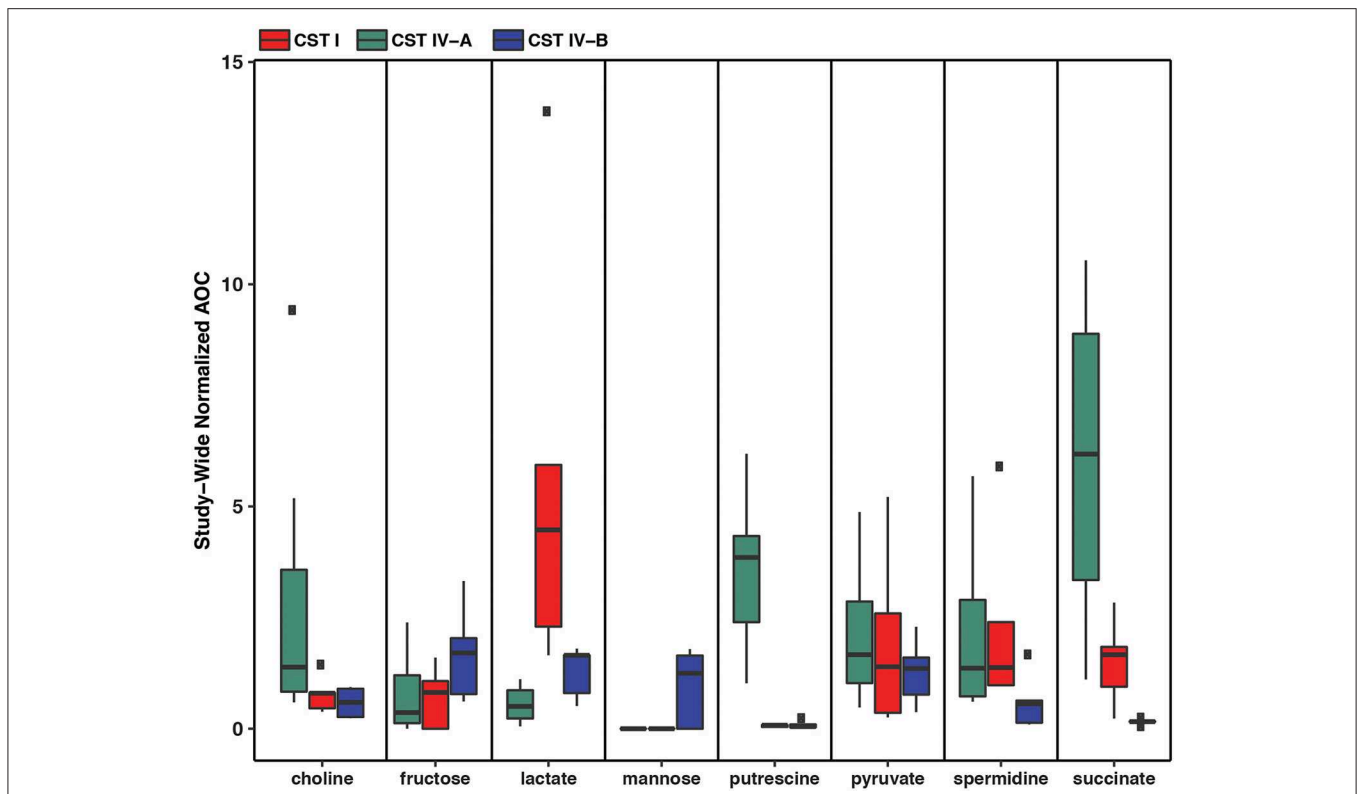

FIGURE 3 | Genomic features of "*Ca.* Lachnocurva vaginae". Genes coding for the following functions were observed: Methyl-accepting chemotaxis (MCP) and subsequent flagella assembly, Sec-SRP secretion systems, choline import, and metabolism to trimethylamine (TMA), bacteriocin production and export, mannose, fructose, and L-ascorbate transport and metabolism, and D-lactate dehydrogenase.

FIGURE 4 | Metabolomic signatures for biochemicals of interest in vaginal community state type (CST) IV-A (high abundance of "*Ca.* Lachnocurva vaginae"), CST I (dominated by *Lactobacillus crispatus*), and CST IV-B (high abundance of *Gardnerella vaginalis*).
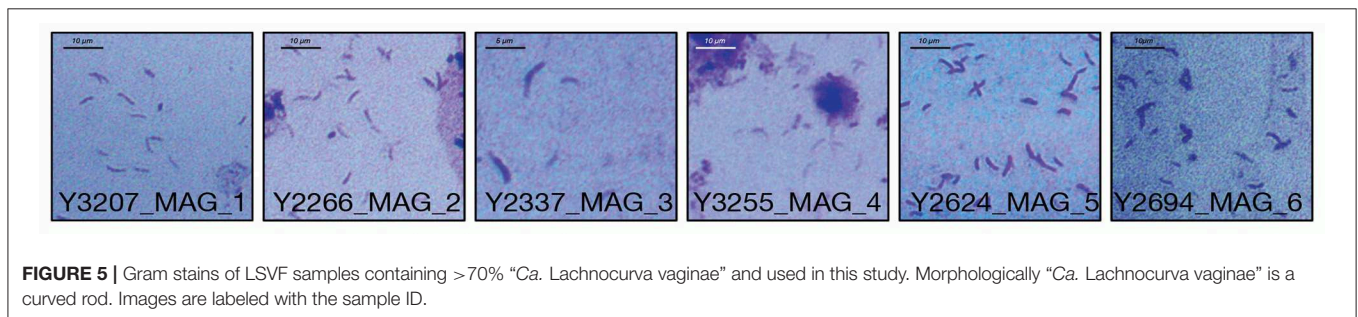

FIGURE 5 | Gram stains of LSVF samples containing >70% "*Ca.* Lachnocurva vaginae" and used in this study. Morphologically "*Ca.* Lachnocurva vaginae" is a curved rod. Images are labeled with the sample ID.

determined pathway. While "*Ca.* Lachnocurva vaginae" does not seem to be able to produce succinate, it has the genetic machinery to produce acetate from pyruvate (i.e., *ackA* and *pta* **LCVA_939** and **LCVA_1006**). We were, however, unable to detect acetate in the metabolome through the methods used, as it is too small of a molecule.

A full suite of genes required for flagella assembly was observed in all MAGs, as well as genes required for methyl-accepting chemotaxis (**LCVA_205, LCVA_992, LCVA_1053, LCVA_1122)** and for the downstream signal process that mediates flagellar response, *cheY* (**LCVA_362**, **LCVA_688**) (Bren and Eisenbach, 2000). Flagella have yet to be visually observed on "*Ca.* Lachnocurva vaginae" and were not observed in the source samples used for this study (**Figure 5**).

We found the cMAG and some MAGs to also include genes that encode protection against antibiotics including the *tetO* gene (**LCVA_1310**, also observed in MAG Y3255) and drug efflux pumps for macrolides (*macB* **LCVA_500**, also observed in MAGs Y2337 and Y2266) and fluoroquinolones (**LCVA_1202**, present in all MAGs). A gene for hemolysin, *tlyA,* was also observed in all MAGs (**LCVA_1031**), suggesting direct interaction between "*Ca.* Lachnocurva vaginae" and host tissues.

Additionally, choline transporters, as well as the *cutC* and *cutD* genes were detected in all MAGs (**LCVA_1194** and **LCVA_1193**, respectively). The *cut* genes metabolize choline and produce trimethylamine (TMA) (Martinez-Del Campo et al., 2015). Aside from transport of exogenous choline, another potential source of choline would be via phospholipase D

hydrolysis of phosphotidyl choline into choline as seen in gut bacteria (Chittim et al., 2019), however no genes encoding such activity were found in the cMAG of "*Ca.* Lachnocurva vaginae". Comparing the translated *cutC* amino acid sequence from "*Ca.* Lachnocurva vaginae" to those reported by Martinez-Del Campo et al. (2015), we observed three of the five active sites conserved in "*Ca.* Lachnocurva vaginae" (Cys489, Glu491, Gly821, data not shown) indicating that it is likely functional. TMA is one of the substances believed to be responsible for the fishy odor associated with BV (Brand and Galask, 1986; Wolrath et al., 2002), however it is rarely detected in metabolomics analyses as it is highly volatile, unless it is performed on freshly collected samples (Wolrath et al., 2002).

Two intact bacteriophages were detected in the cMAG (**Figure 2**). Partial matches to the same phages were observed in all MAGs (**Supplemental Image 1**). Best BLAST hits indicated that both bacteriophages belong to the Siphoviridae family of double-stranded viruses (NC_009552, NC_011167) which can exhibit both lytic and lysogenic phases. Bacteriophages of this family have previously been reported in the vaginal species *Lactobacillus jensenii* (Martin et al., 2010). Genomic islands were detected at multiple sites (**Supplemental Table 1** and **Figure 2**). Analysis of proteins encoded in the largest island (65.5 kb, coordinates 1,331,959–1,391,037, **LCVA_1280-1330**) showed similarity to proteins from *Shuttleworthia satelles*, and other taxa from the *Clostridiales* (**Supplemental Table 3** and **Supplemental Image 2**) and several transposons and integrases, as well as tetracycline resistance proteins (TetO, **LCVA_1310**), and ABC transporters. We ruled out the presence of genomic islands due to mis-assembly by analyzing the long-read coverage spanning the 5′ and 3′ ends of genomic islands. More than 100 long PacBio reads of a means size of 9 kb span the junctions of all detected genomic islands. Further, mean coverage of the genomic islands was similar to that of the mean cMAG coverage of 124X. Portions of these islands were observed in the other MAGs assembled in this study (**Supplemental Image 1** and **Supplemental Table 1**). The stringency of the mapping would not recover reads to regions of sequence diversity, thus these results may indicate this "*Ca.* Lachnocurva vaginae" likely contains multiple regions of genetic fluidity or diversity. However, the missing regions in MAGs may also be an artifact of metagenomic assembly.

CONCLUSION

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 assembles 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 environment and its relation to women's health. We have shown that "*Ca.* Lachnocurva vaginae" has the genomic potential for motility and chemotaxis, and is likely capable of resisting several antibiotics via drug efflux systems. Our analysis indicates this candidate species may contribute to the fishy 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 microbiome.

DATA AVAILABILITY STATEMENT

Sequence data have been submitted to NCBI under BioProject PRJNA562728.

ETHICS STATEMENT

Samples used in this study were archived and de-identified cervicovaginal lavages and swabs. The samples were originally collected after obtaining informed consent by all participants, who also provided consent for storage of the samples and use in future research studies related to women's health. The original study was approved by the University of Maryland School of Medicine Institutional Review Board.

AUTHOR CONTRIBUTIONS

EM and CR performed the extraction and library preparations from the raw samples. JH, AM, LT, MF, and BM performed bioinformatics analyses to generate the MAGs. JH, RB, and JR designed the study. JH, MF, BM, and JR executed the study, contributed to the analyses. JH, RB, MF, BM, and JR wrote and edited the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcimb. 2020.00117/full#supplementary-material

Supplemental Data Sheet 1 | Metagenomic library preparation modifications based on input amounts of DNA.

Supplemental Data Sheet 2 | Processing scripts for analysis.

Supplemental Data Sheet 3 | Taxonomic and phylogenetic placement of "Ca. Lachnocurva vaginae" (UAB071) from EZBioCloud TrueBacID system.

Supplemental Table 1 | List of genes annotated in the "Ca. Lachnocurva vaginae" cMAG. Excludes rRNA and tRNA.

Supplemental Table 2 | List of marker genes used in Family-level CheckM analysis and presence (1) or absence (0) in the "Ca. Lachnocurva vaginae" cMAG.

Supplemental Table 3 | Top 3 best BLAST matches to the genes of the "Ca. Lachnocurva vaginae" cMAG.

Supplemental Image 1 | Visualization of "Ca. Lachnocurva vaginae" whole genome alignments of cMAG and other MAGs.

Supplemental Image 2 | Circle plot of "Ca. Lachnocurva vaginae" cMAG with best BLAST hits indicated as either Clostridiales or not (Ring 5). If hit was within Clostridiales and >80% identical to cMAG, the specific taxon of close relatives was indicated in Ring 4. Included BLAST hits covered at least 80% of the query gene.

REFERENCES

Alikhan, N. F., Petty, N. K., Ben Zakour, N. L., and Beatson, S. A. (2011). BLAST ring image generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12:402. doi: 10.1186/1471-2164-12-402

Allsworth, J. E., and Peipert, J. F. (2007). Prevalence of bacterial vaginosis: 2001-2004 national health and nutrition examination survey data. *Obstet. Gynecol.* 109, 114–120. doi: 10.1097/01.AOG.0000247627.84791.91

Amsel, R., Totten, P. A., Spiegel, C. A., Chen, K. C., Eschenbach, D., and Holmes, K. K. (1983). Nonspecific vaginitis. Diagnostic criteria and microbial and epidemiologic associations. *Am. J. Med.* 74, 14–22. doi: 10.1016/0002-9343(83)91112-9

Ankenbrand, M. J., Hohlfeld, S., Hackl, T., and Förster, F. (2017). AliTV—interactive visualization of whole comparisons. *PeerJ. Comput. Sci.* 3:e116. doi: 10.7717/peerj-cs.116

Arkin, A. P., Cottingham, R. W., Henry, C. S., Harris, N. L., Stevens, R. L., Maslov, S., et al. (2018). KBase: the United states department of energy systems biology knowledgebase. *Nat. Biotechnol.* 36, 566–569. doi: 10.1038/nbt.4163

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., et al. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44, W16–W21. doi: 10.1093/nar/gkw387

Atashili, J., Poole, C., Ndumbe, P. M., Adimora, A. A., and Smith, J. S. (2008). Bacterial vaginosis and HIV acquisition: a meta-analysis of published studies. *AIDS* 22, 1493–1501. doi: 10.1097/QAD.0b013e3283021a37

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Bertelli, C., Laird, M. R., Williams, K. P., Simon Fraser University Research Computing, Group, Lau, B. Y., Hoad, G., et al. (2017). IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res.* 45, W30–W35. doi: 10.1093/nar/gkx343

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bradshaw, C. S., Morton, A. N., Hocking, J., Garland, S. M., Morris, M. B., Moss, L. M., et al. (2006). High recurrence rates of bacterial vaginosis over the course of 12 months after oral metronidazole therapy and factors associated with recurrence. *J. Infect. Dis.* 193, 1478–1486. doi: 10.1086/503780

Brand, J. M., and Galask, R. P. (1986). Trimethylamine: the substance mainly responsible for the fishy odor often associated with bacterial vaginosis. *Obstet. Gynecol.* 68, 682–685.

Bren, A., and Eisenbach, M. (2000). How signals are heard during bacterial chemotaxis: protein-protein interactions in sensory signal propagation. *J. Bacteriol.* 182, 6865–6873. doi: 10.1128/JB.182.24.6865-6873.2000

Cherpes, T. L., Meyn, L. A., Krohn, M. A., Lurie, J. G., and Hillier, S. L. (2003). Association between acquisition of herpes simplex virus type 2 in women and bacterial vaginosis. *Clin. Infect. Dis.* 37, 319–325. doi: 10.1086/375819

Chittim, C. L., Martinez Del Campo, A., and Balskus, E. P. (2019). Gut bacterial phospholipase Ds support disease-associated metabolism by generating choline. *Nat. Microbiol.* 4, 155–163. doi: 10.1038/s41564-018-0294-4

Cohen, C. R., Lingappa, J. R., Baeten, J. M., Ngayo, M. O., Spiegel, C. A., Hong, T., et al. (2012). Bacterial vaginosis associated with increased risk of female-to-male HIV-1 transmission: a prospective cohort analysis among African couples. *PLoS Med.* 9:e1001251. doi: 10.1371/journal.pmed.1001251

Domingo, M. C., Huletsky, A., Boissinot, M., Helie, M. C., Bernal, A., Bernard, K. A., et al. (2009). Clostridium lavalense sp. nov., a glycopeptide-resistant species isolated from human faeces. *Int. J. Syst. Evol. Microbiol.* 59, 498–503. doi: 10.1099/ijs.0.001958-0

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Fettweis, J. M., Serrano, M. G., Brooks, J. P., Edwards, D. J., Girerd, P. H., Parikh, H. I., et al. (2019). The vaginal microbiome and preterm birth. *Nat. Med.* 25, 1012–1021. doi: 10.1038/s41591-019-0450-2

France, M. T., Mendes-Soares, H., and Forney, L. J. (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. doi: 10.1128/AEM.02385-16

Fredricks, D. N., Fiedler, T. L., and Marrazzo, J. M. (2005). Molecular identification of bacteria associated with bacterial vaginosis. *N. Engl. J. Med.* 353, 1899–1911. doi: 10.1056/NEJMoa043802

Galens, K., Orvis, J., Daugherty, S., Creasy, H. H., Angiuoli, S., White, O., et al. (2011). The IGS standard operating procedure for automated prokaryotic annotation. *Stand. Genomic Sci.* 4, 244–251. doi: 10.4056/sigs.1223234

Gronow, S., Welnitz, S., Lapidus, A., Nolan, M., Ivanova, N., Glavina Del Rio, T., et al. (2010). Complete genome sequence of *Veillonella parvula* type strain (Te3). *Stand. Genomic Sci.* 2, 57–65. doi: 10.4056/sigs.521107

Holm, J. B., Humphrys, M., Robinson, C. K., Settles, M. L., Ott, S., Fu, L., et al. (2019). Ultra-high throughput multiplexing and sequencing of >500 bp amplicon regions on the Illumina HiSeq 2500 platform. *mSystems.* 4:e00029-19. doi: 10.1128/mSystems.00029-19

Jain, C., Rodriguez, R. L., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9:5114. doi: 10.1038/s41467-018-07641-9

Kanehisa, M., Sato, Y., and Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* 428, 726–731. doi: 10.1016/j.jmb.2015.11.006

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199

Klebanoff, M. A., Schwebke, J. R., Zhang, J., Nansel, T. R., Yu, K. F., and Andrews, W. W. (2004). Vulvovaginal symptoms in women with bacterial vaginosis. *Obstet. Gynecol.* 104, 267–272. doi: 10.1097/01.AOG.0000134783.98382.b0

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27, 722–736. doi: 10.1101/gr.215087.116

Koumans, E. H., Sternberg, M., Bruce, C., Mcquillan, G., Kendrick, J., Sutton, M., et al. (2007). The prevalence of bacterial vaginosis in the United States, 2001-2004; associations with symptoms, sexual behaviors, and reproductive health. *Sex. Transm. Dis.* 34, 864–869. doi: 10.1097/OLQ.0b013e318074e565

Lamont, R. F., Sobel, J. D., Akins, R. A., Hassan, S. S., Chaiworapongsa, T., Kusanovic, J. P., et al. (2011). The vaginal microbiome: new information about genital tract flora using molecular based techniques. *BJOG* 118, 533–549. doi: 10.1111/j.1471-0528.2010.02840.x

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Leitich, H., Bodner-Adler, B., Brunbauer, M., Kaider, A., Egarter, C., and Husslein, P. (2003). Bacterial vaginosis as a risk factor for preterm delivery: a meta-analysis. *Am. J. Obstet. Gynecol.* 189, 139–147. doi: 10.1067/mob.2003.339

Lennard, K., Dabee, S., Barnabas, S. L., Havyarimana, E., Blakney, A., Jaumdally, S. Z., et al. (2018). Microbial composition predicts genital tract inflammation and persistent bacterial vaginosis in South African adolescent females. *Infect. Immun.* 86:e00410–17. doi: 10.1128/IAI.00410-17

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Martin, R., Escobedo, S., and Suarez, J. E. (2010). Induction, structural characterization, and genome sequence of Lv1, a prophage from a human vaginal Lactobacillus jensenii strain. *Int. Microbiol.* 13, 113–121. doi: 10.2436/20.1501.01.116

Martinez-Del Campo, A., Bodea, S., Hamer, H. A., Marks, J. A., Haiser, H. J., Turnbaugh, P. J., et al. (2015). Characterization and detection of a widely distributed gene cluster that predicts anaerobic choline utilization by human gut bacteria. *MBio* 6:e00042–15. doi: 10.1128/mBio.00042-15

Mckinnon, L. R., Achilles, S. L., Bradshaw, C. S., Burgener, A., Crucitti, T., Fredricks, D. N., et al. (2019). The evolving facets of bacterial vaginosis: implications for HIV transmission. *AIDS Res. Hum. Retroviruses.* 35, 219–228. doi: 10.1089/aid.2018.0304

Muzny, C. A., Sunesara, I. R., Griswold, M. E., Kumar, R., Lefkowitz, E. J., Mena, L. A., et al. (2014). Association between BVAB1 and high Nugent scores among women with bacterial vaginosis. *Diagn. Microbiol. Infect. Dis.* 80, 321–323. doi: 10.1016/j.diagmicrobio.2014.09.008

Nelson, T. M., Borgogna, J. C., Michalek, R. D., Roberts, D. W., Rath, J. M., Glover, E. D., et al. (2018). Cigarette smoking is associated with an altered vaginal tract metabolomic profile. *Sci. Rep.* 8:852. doi: 10.1038/s41598-017-14943-3

Ness, R. B., Hillier, S. L., Kip, K. E., Soper, D. E., Stamm, C. A., Mcgregor, J. A., et al. (2004). Bacterial vaginosis and risk of pelvic inflammatory disease. *Obstet. Gynecol.* 104, 761–769. doi: 10.1097/01.AOG.0000139512.37582.17

Ness, R. B., Kip, K. E., Soper, D. E., Hillier, S., Stamm, C. A., Sweet, R. L., et al. (2005). Bacterial vaginosis (BV) and the risk of incident gonococcal or chlamydial genital infection in a predominantly black population. *Sex. Transm. Dis.* 32, 413–417. doi: 10.1097/01.olq.0000154493.87451.8d

Nugent, R. P., Krohn, M. A., and Hillier, S. L. (1991). Reliability of diagnosing bacterial vaginosis is improved by a standardized method of gram stain interpretation. *J. Clin. Microbiol.* 29, 297–301. doi: 10.1128/JCM.29.2.297-301.1991

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114

Petrova, M. I., van den Broek, M., Balzarini, J., Vanderleyden, J., and Lebeer, S. (2013). Vaginal microbiota and its role in HIV transmission and infection. *FEMS Microbiol. Rev.* 37, 762–792. doi: 10.1111/1574-6976.12029

Price, M. N., Dehal, P. S., and Arkin, A. P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* 26, 1641–1650. doi: 10.1093/molbev/msp077

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS ONE.* 5:e9490. doi: 10.1371/journal.pone.0009490

Quinlan, A. R. (2014). BEDTools: the Swiss-Army tool for genome feature analysis. *Curr. Protoc. Bioinformatics 47,* 11–34. doi: 10.1002/0471250953.bi11 12s47

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033

Ravel, J., Brotman, R. M., Gajer, P., Ma, B., Nandy, M., Fadrosh, D. W., et al. (2013). Daily temporal dynamics of vaginal microbiota before, during and after episodes of bacterial vaginosis. *Microbiome.* 1:29. doi: 10.1186/2049-2618-1-29

Schwebke, J. R., and Desmond, R. A. (2007). A randomized trial of the duration of therapy with metronidazole plus or minus azithromycin for treatment of symptomatic bacterial vaginosis. *Clin. Infect. Dis.* 44, 213–219. doi: 10.1086/509577

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Srinivasan, S., Morgan, M. T., Fiedler, T. L., Djukovic, D., Hoffman, N. G., Raftery, D., et al. (2015). Metabolic signatures of bacterial vaginosis. *MBio* 6:e00204–15. doi: 10.1128/mBio.00204-15

Srinivasan, S., Morgan, M. T., Liu, C., Matsen, F. A., Hoffman, N. G., Fiedler, T. L., et al. (2013). More than meets the eye: associations of vaginal bacteria with gram stain morphotypes using molecular phylogenetic analysis. *PLoS ONE* 8:e78633. doi: 10.1371/journal.pone.0078633

Taha, T. E., Hoover, D. R., Dallabetta, G. A., Kumwenda, N. I., Mtimavalye, L. A., Yang, L. P., et al. (1998). Bacterial vaginosis and disturbances of vaginal flora: association with increased acquisition of HIV. *AIDS* 12, 1699–1706. doi: 10.1097/00002030-199813000-00019

Weghoff, M. C., Bertsch, J., and Muller, V. (2015). A novel mode of lactate metabolism in strictly anaerobic bacteria. *Environ. Microbiol.* 17, 670–677. doi: 10.1111/1462-2920.12493

Wickham, H. (2009). *Ggplot2 : Elegant Graphics for Data Analysis.* New York, NY: Springer.

Witkin, S. S., Mendes-Soares, H., Linhares, I. M., Jayaram, A., Ledger, W. J., and Forney, L. J. (2013). Influence of vaginal bacteria and D- and L-lactic acid isomers on vaginal extracellular matrix metalloproteinase inducer: implications for protection against upper genital tract infections. *mBio* 4:e00460–13. doi: 10.1128/mBio.00460-13

Wolrath, H., Borén, H., Hallén, A., and Forsum, U. (2002). Trimethylamine content in vaginal secretion and its relation to bacterial vaginosis. *APMIS* 110, 819–824. doi: 10.1034/j.1600-0463.2002.1101108.x

Yoon, S. H., Ha, S. M., Kwon, S., Lim, J., Kim, Y., Seo, H., et al. (2017). Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int. J. Syst. Evol. Microbiol.* 67, 1613–1617. doi: 10.1099/ijsem.0.001755

Yutin, N., and Galperin, M. Y. (2013). A genomic update on clostridial phylogeny: Gram-negative spore formers and other misplaced clostridia. *Environ. Microbiol.* 15, 2631–2641. doi: 10.1111/1462-2920.12173

Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J., and Wishart, D. S. (2011). PHAST: a fast phage search tool. *Nucleic Acids Res.* 39, W347–W352. doi: 10.1093/nar/gkr485