Resolution and Cooccurrence Patterns of *Gardnerella leopoldii*, *G. swidsinskii*, *G. plotii*, and *G. vaginalis* within the Vaginal Microbiome

Janet E. Hill,^a Arianne Y. K. Albert,^b the VOGUE Research Group

^aDepartment of Veterinary Microbiology, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

^bWomen's Health Research Institute, BC Women's Hospital and Health Centre, Vancouver, British Columbia, Canada

ABSTRACT *Gardnerella vaginalis* is a hallmark of vaginal dysbiosis, but it is found in the microbiomes of women with and without vaginal symptoms and those who do not have microbiologically defined dysbiosis. *G. vaginalis* encompasses diverse taxa differing in attributes that are potentially important for virulence, and there is evidence that clades or subgroups within the species are differentially associated with clinical outcomes. The *G. vaginalis* species description was recently emended, and three new species within the genus were defined (*G. leopoldii*, *G. swidsinskii*, and *G. plotii*). 16S rRNA sequences for the four *Gardnerella* species are all >98.5% identical, and no signature sequences differentiate them. We demonstrated that *Gardnerella* species can be resolved using partial chaperonin 60 (cpn60) sequences, with pairwise percent identities of 87.1 to 97.8% among the type strains. Pairwise cooccurrence patterns of *Gardnerella* spp. in the vaginal microbiomes of 413 reproductive aged Canadian women were investigated, and several significant cooccurrences of species were identified. Abundance of *G. vaginalis* and *G. swidsinskii* was associated with vaginal symptoms of abnormal odor and discharge. cpn60 barcode sequencing can provide a rapid assessment of the relative abundance of *Gardnerella* spp. in microbiome samples, providing a powerful method of elucidating associations between these diverse organisms and clinical outcomes. Researchers should consider using cpn60 instead of 16S RNA for better resolution of these important organisms.

KEYWORDS *Gardnerella*, microbiome, vaginal microbiome

Since its original isolation from human vaginal samples in 1953 (1), the species that eventually became known as *Gardnerella vaginalis* has been strongly associated with vaginal dysbiosis and negative reproductive outcomes (2). Evaluation of the abundance of *Gardnerella* morphotypes in Gram-stained vaginal smears is a key factor in calculation of the Nugent score (3) and thus in the microbiological definition of bacterial vaginosis (BV). Understanding of its role in the vaginal microbiome has been complicated due to phenotypic diversity within the taxon and its perplexing presence, sometimes in high numbers, in the vaginal microbiomes of women without any signs or symptoms of dysbiosis (2). It can also often be detected at low levels in microbiomes dominated by other species, including *Lactobacillus* spp.

Over the past decades, several classification schemes have been developed in an attempt to delineate subgroups within *Gardnerella vaginalis*. These have included both biotyping schemes based on a set of biochemical test results (4, 5) and molecular methods based on amplification and restriction digestion of 16S rRNA gene sequences (6). In 2005, four clusters of *G. vaginalis*-like sequences were observed in vaginal microbiome profiles based on sequencing of amplified chaperonin 60 (cpn60) barcode sequences (7). More recently, four subgroups or clades of *G. vaginalis* were defined

Citation Hill JE, Albert AYK, the VOGUE Research Group. 2019. Resolution and cooccurrence patterns of *Gardnerella leopoldii*, *G. swidsinskii*, *G. plotii*, and *G. vaginalis* within the vaginal microbiome. Infect Immun 87:e00532-19. <https://doi.org/10.1128/IAI.00532-19>.

Editor Denise Monack, Stanford University

Copyright © 2019 American Society for Microbiology. All Rights Reserved.

Address correspondence to Janet E. Hill, Janet.Hill@usask.ca.

Received 11 July 2019

Returned for modification 3 September 2019

Accepted 9 September 2019

Accepted manuscript posted online 16 September 2019

Published 18 November 2019

based either on the cpn60 barcode sequence (8–10) or on the concatenated sequences of 473 genes common to a set of 17 *G. vaginalis* isolates (11). When we compared these two approaches directly, they were consistent with each other, with cpn60-defined subgroups A to D corresponding to clades 4, 2, 1, and 3, respectively (8). Based on the results of this comparison, we suggested that pairwise average nucleotide identity (ANI) values between whole-genome sequences of representative isolates were consistent with their definition as separate species (12) but that additional phenotypic characteristics that differentiate the subgroups should be identified (8).

In 2019, Vaneechoutte et al. (13) formally proposed the emendment of the species *G. vaginalis* and defined three new species: *G. piovii*, *G. swidsinskii*, and *G. leopoldii*, based on comparison of whole-genome sequences, biochemical properties, and matrix-assisted laser desorption ionization–time of flight (MALDI-TOF) mass spectrometry analysis. The species could not be delineated using full-length 16S rRNA gene sequences (13). In addition to these four species, the authors also defined nine additional “genome species” based on whole-genome sequence comparisons. These additional genome species were not named or formally described, presumably due to a lack of sufficient numbers of isolates to make a strong case for their designation.

Given the apparent significance of *Gardnerella* spp. in the vaginal microbiome, resolution of these species in metagenomic samples and association of their presence and abundance with clinical outcomes is critical. We have already demonstrated that amplification and sequencing of the cpn60 barcode can be used to resolve cpn60-defined subgroups A to D in vaginal samples and demonstrated that their differential abundance can be used to reveal previously unreported microbial community compositions (14). The objectives of the current study were to determine if cpn60 barcode sequences could differentiate the newly defined species and genome species of *Gardnerella*, to investigate their distribution in a collection of 417 previously sequenced vaginal microbiome profiles, and to identify associations of *Gardnerella* spp. with vaginal symptoms. Our results confirm the discriminatory power of the cpn60 barcode sequence and reveal significant cooccurrences of *Gardnerella* spp. in the vaginal microbiome that have implications for diagnostics for women’s health and for our understanding of vaginal microbial ecology.

RESULTS AND DISCUSSION

Resolution of *Gardnerella* spp. based on cpn60 universal target sequences. The four named *Gardnerella* spp. were resolved in the cpn60 phylogenetic tree based on an alignment of the 552-bp “universal target” sequence barcode (Fig. 1), with good bootstrap support for nodes separating the species. *G. vaginalis* (genome species 1) corresponds to the previously described subgroup C/clade 1. This observation is consistent with previous demonstrations that cpn60 barcode sequences are generally excellent predictors of whole-genome sequence relationships among closely related bacteria (15, 16).

G. swidsinskii (genome species 6) and *G. leopoldii* (genome species 5) share a node in the tree but are separated clearly with good bootstrap support. These two species (represented by strains AMD and 5-1) were previously grouped together within cpn60 subgroup A/clade 4 based on the selection of isolates available for analysis at the time. Vaneechoutte et al. (13) noted in their description of the species definitions that *G. swidsinskii* and *G. leopoldii* could be distinguished based on ANI, DNA-DNA hybridization (DDH), and MALDI-TOF profiles.

G. piovii (genome species 4) corresponds to subgroup B/clade 2 and can be distinguished from other species based on ANI, DDH, MALDI-TOF mass spectrometry, and a positive sialidase test. One *G. piovii* isolate (JCP8151A) clustered with genome species 3 in the tree. In the original description of the cpn60 subgroups (10), what is designated genome species 3 by Vaneechoutte et al. (13) was also included in subgroup B. It is noteworthy that several isolates examined by Schellenberg et al. (8) with cpn60 sequences identical to that of strain 00703C2mash (genome species 3) were also

FIG 1 Phylogenetic relationships of *Gardnerella* spp. based on an alignment of the 552-bp cpn60 barcode sequence. Type strains are shown in red, and representatives of the other 9 genome species designated by Vaneechoutte et al. (13) are in blue. Bootstrap values are indicated at the major nodes. The tree is rooted with *Alloscardovia omnicolens*. cpn60 subgroups A to D (10) and clades 1 to 4 (11) are labeled as subgroup/clade based on sequences common to previous studies.

sialidase positive. The characterization of additional isolates of genome species 3 and *G. piovii* will be necessary to determine if these groups should be combined.

Subgroup D/clade 3 was the most diverse in previous descriptions, and so it is not surprising to find that isolates previously identified as subgroup D (strains 101, 1500E, 6119V5, and 00703Dmash) are separated into three genome species: genome species 8, 9, and 10. Complete characterization and possible naming of these three genome species, along with genome species 2, 7, and 11 to 13, will require analysis of additional isolates to establish differentiation by whole-genome sequence and additional phenotypic characteristics.

Pairwise nucleotide sequence identities for the type strains of *Gardnerella* spp. and representatives of the other nine genome species were calculated from the aligned sequences (see Table S1 in the supplemental material). Identities among the four type strains were from 87.1% (*G. leopoldii* versus *G. piovii*) to 97.8% (*G. leopoldii* versus *G. swidsinskii*). When representatives of the other nine genome species were included,

percent identities for the 552-bp cpn60 barcode sequence ranged from 84.2% (genome species 7 versus genome species 2 or *G. piovii*) to 99.4% (genome species 9 versus genome species 10). No isolates had identical cpn60 barcode sequences. Interspecific cpn60 barcode sequence identities are known to vary widely among bacterial genera, so this range was not unexpected (17). To investigate the resolving power of the 5' and 3' ends of the barcode sequence that would be determined using routine next-generation sequencing protocols, we truncated the alignment to examine 250 bp of either end of the barcode sequence. Average pairwise identities were 88.2% (range, 83.2 to 99.6) and 91.3% (range, 84.4 to 99.2) for the 5' and 3' ends of the barcode sequence, respectively (Table S1). None of the species were identical. These identities cover the same range as observed for the entire barcode sequence, as is expected given the uniform distribution of sequence variation along its length (9).

Classification of *Gardnerella* sequence variants. One of the major advantages of use of the cpn60 barcode sequence for taxonomic profiling of microbial communities is the ability to achieve species level classification of sequence reads or assembled operational taxonomic unit (OTU) sequences routinely. It was this resolution that led to the identification of previously undescribed microbial community compositions in the human vaginal microbiome, based on the detection of subspecies level sequences within *Gardnerella* (14). Elucidation of the role of genomically and phenotypically distinct *Gardnerella* lineages in the vaginal microbiome and determination of their association with clinical outcomes require determining their distribution in clinical cohorts. Accomplishing this on a large scale requires culture-independent techniques. While whole-genome shotgun metagenomics might provide resolution of *Gardnerella* spp., this approach requires orders of magnitude more sequencing effort and much more complex bioinformatics than amplicon sequencing. Based on the successful resolution of 13 genome species of *Gardnerella* described above, we next sought to discover if they could be reliably detected and quantified in cpn60 amplicon sequence-based microbiome profiles.

cpn60 barcode sequence data were available from 417 previously characterized vaginal samples from nonpregnant, reproductive aged Canadian women. For the purposes of the current study, exact sequence variants were identified using DADA2 and a truncation length of 250 bp and variants were compared to the cpnDB_nr reference database (17) to identify the nearest database neighbor.

Most (301/413) of the samples for which at least 500 reads were available contained some *Gardnerella*-like sequence variants, and variants corresponding to all 13 *Gardnerella* spp. and genome species were detected. The median sequence identity of variants to reference sequences was 98.4%. Sample prevalence and proportional abundance ranged widely among species (Table S2). For example, 68.4% (206/301) of *Gardnerella*-positive samples contained *G. vaginalis* and 49% (148/301) contained *G. swidsinskii*, but eight genome species (2 and 7 to 13) were detected in $\leq 10\%$ of samples. The prevalence and abundance patterns are generally consistent with previous descriptions of vaginal microbiomes based on cpn60 barcode sequencing (14, 18–21) or clade-specific quantitative real-time PCR (22, 23). There were 60 samples with at least 50% of their read counts accounted for by *Gardnerella* spp. and 30 samples with at least 75% *Gardnerella* (Table S2).

The number of *Gardnerella* spp. detected per *Gardnerella*-positive sample ranged from 1 (109/301 [36.2%]) to 10 (3/301 [1%]), although the majority (184/301 [61.1%]) contained one or two species (Fig. 2). Overall, multiple *Gardnerella* spp. were detected in 63.8% of samples, consistent with a previous report of multiple *Gardnerella* clades in 70% of samples from women with BV (22). The prevalence and proportional abundance of *Gardnerella* spp. in samples with $>50\%$ *Gardnerella* ($n = 60$) are shown in Fig. 3. In addition to the four named species, genome species 3 was detected frequently and in relatively high proportional abundance, in striking contrast to the rarely detected genome species 2 and 7 to 13 (Fig. 3 and Table S2).

FIG 2 Number of *Gardnerella* spp. detected per sample ($n = 301$).

The prevalence and abundance patterns we observed in these samples mirror the isolate and whole-genome sequence collection used to provide evidence for the emendment of *Gardnerella* and the designation of the new species (13). Based on our experience, there is no obvious bias in PCR amplification of *Gardnerella* lineages, and multiple representatives of all previously defined cpn60 subgroups were readily amplified using cpn60 “universal” PCR primers (8). Furthermore, we have shown a strong correlation between *Gardnerella* cpn60 sequence read counts in amplicon-based microbiome profiles and abundances determined by *Gardnerella*-specific quantitative real-time PCR (18). Thus, it seems that some *Gardnerella* species are actually less prevalent and do not achieve proportional dominance in the populations of women we have examined to date. Elucidating the ecological mechanisms responsible for this differential “success” of *Gardnerella* spp. will require further focused study.

Cooccurrence of *Gardnerella* spp. Given the frequency with which women are colonized by multiple species of *Gardnerella*, we were interested to determine if there are any consistent patterns of cooccurrence among species in the 301 *Gardnerella*-positive samples. Closely related species occupying similar environmental niches might be expected to cooccur more frequently, and depending on resource levels, they might also compete with each other. Raw correlations of read counts are not recommended for assessing cooccurrence, as they are biased in the context of the compositional nature of amplicon sequence analysis (24–26). Methods based on presence/absence, such as Jaccard’s index, can be informative in this context and perform better than raw correlations (27). To determine whether taxa cooccur more or less often than expected by chance, a reasonable null model for Jaccard’s index is required. Traditional null models of cooccurrence have used randomizations and simulations but have been shown to be biased under many circumstances (28). Therefore, we used a probabilistic null model of cooccurrence (29), which performs well for microbial sequencing data (27). In addition to cooccurrence analysis using presence/absence, we also investigated proportionality of species (25) on the center-log ratio (CLR)-transformed read counts using the Propr package (30). The results were very similar, with *G. vaginalis* and *G. swidsinskii* showing clustering by proportionality, as well as *G. piovii* and genome species 3. However, as there is currently no agreed upon hypothesis testing method for proportionality, we used the presence/absence data to determine significant cooccurrences.

Significant cooccurrences were observed in the 301 *Gardnerella*-positive samples for several pairs of species, but there were also many cases where species cooccurred only randomly (Fig. 4A and Table S3). Among the most frequently detected species (*G. vaginalis*, *G. swidsinskii*, *G. leopoldii*, *G. piovii*, and genome species 3), the smallest pairwise Jaccard distances (i.e., the most samples in common) were observed for *G.*

FIG 3 Proportional abundances of 13 *Gardnerella* spp. in vaginal microbiomes of women with $\geq 50\%$ *Gardnerella* sequence reads ($n = 60$). Red circles indicate the proportional abundance of each species according to scale on the left; blue circles represent the proportion of reads identified as non-*Gardnerella*.

FIG 4 (A) Significance of pairwise cooccurrences of *Gardnerella* spp. in 413 vaginal samples determined by Jaccard index of similarity (J) calculation. P values are indicated by color according to the key. Species were considered present if the raw number of sequence reads in the sample was ≥ 10 ; otherwise, the species was marked as absent. P values were Benjamini-Hochberg corrected to a false-discovery rate of 0.05. (B) Hierarchical clustering of species based on Jaccard distances ($1 - J$), using complete linkage.

vaginalis and *G. swidsinskii* and for *G. piotii* and genome species 3 (Fig. 4B). *G. leopoldii* and *G. swidsinskii* did not occur together more often than expected by chance, which is of note as both were previously labeled as subgroup A based on *cpn60* sequences and whole-genome sequence comparisons (8). The differentiating features of these two species detected by MALDI-TOF mass spectrometry (13) may be associated with their occupation of distinct niches. This suggests that the new labeling is indeed useful for understanding differences in distributions at this deeper level. None of the species pairs had fewer cooccurrences than expected by chance, suggesting that competitive exclusion may not be important for describing their relative distributions. Conclusions regarding the rarely detected genome species are limited, since very few samples were positive and thus chances of observing cooccurrence were correspondingly low.

Association of *Gardnerella* spp. with BV status and vaginal symptoms. To understand how resolution of different *Gardnerella* spp. may inform clinically important outcomes, we compared the relative abundances of the more frequently occurring species (*G. vaginalis*, *G. swidsinskii*, *G. leopoldii*, *G. piovii*, and genome species 3) among groups based on clinical Nugent scores (negative, intermediate, and BV) and self-reported symptoms in the 2 weeks prior to the swab collection (odor, irritation, and discharge). There was a significant association between Nugent category and relative abundance of *G. vaginalis*, *G. swidsinskii*, and *G. piovii* (Table 1). Genome species 3 was marginally associated with Nugent category, but none of the pairwise comparisons was significant after *P* value adjustment. The relationship between *Gardnerella* abundance and Nugent score is not surprising, as the presence of *Gardnerella* “morphotypes” on Gram-stained slides of vaginal specimens is part of the calculation of the clinical score (3). The lack of association of *G. leopoldii* with Nugent category is interesting in that this species was previously investigated together with *G. swidsinskii* as cpn60 subgroup A/clade 4, which was found to be associated with Nugent category (14, 23, 31). These species are very closely related phylogenetically (Fig. 1) and were resolved by Vaneechoutte et al. (13) only by MALDI-TOF mass spectrometry and whole-genome comparison, so the specific factors responsible for their apparently different relationships with the microbiological definition of BV remain to be identified.

Phenotypic diversity has long been considered a possible explanation for the detection of *Gardnerella* in women without vaginal symptoms; however, attempts to identify associations between particular biotypes and clinical status have yielded inconsistent and often contradictory results (32–36). The major limitation of investigations relying on phenotypic characterization of isolates is that they focus only on the most readily culturable isolates from individual specimens (often only one isolate per specimen), which is inadequate since women are usually colonized by multiple species of *Gardnerella*. We observed strong relationships between abnormal odor and discharge with higher relative abundance of *G. vaginalis* and *G. swidsinskii* but not with the other three species, although there is a marginal relationship between discharge and genome species 3 (*P* = 0.02) (Table 1). *G. vaginalis* and *G. swidsinskii* cooccurred more often than expected by chance and also showed proportionality, suggesting that they are correlated in abundance. Therefore, we cannot be sure if it is just one species or both that is associated with vaginal symptoms. Sialidase activity defines *G. piovii* (13) and is also observed for genome species 3 isolates (8); however, these species were not associated with discharge, nor were they the most strongly associated with a BV diagnosis by Nugent score. This is likely due to the polymicrobial nature of BV and the fact that many other BV-associated bacteria produce hydrolytic enzymes that may contribute to symptoms (37–39). The lack of association of sialidase-positive *Gardnerella* spp. with symptoms is also consistent with the suggestion that some types of *Gardnerella* may be important for “stage-setting”: establishing an anaerobic environment and initial adhesion to the vaginal epithelium that lead to abundant growth of other BV associated organisms and the development of multispecies biofilms (reviewed in reference 40). In these primary colonizers, hydrolytic enzymes and cholesterol-dependent cytotoxin (vaginolysin) may be more important in preparing the microbiome for secondary expansion of populations of BV-associated bacteria rather than acting as specific virulence factors affecting the host.

Considering *Gardnerella* as a monolithic taxon in vaginal microbiome studies (due to the ubiquitous application of 16S rRNA gene sequencing in microbiome profiling) has limited progress in understanding the link between vaginal microbiota and clinical outcomes and the development of improved diagnostics for women’s health. Using approaches such as the one presented here will be critical in further investigation of the intriguing association of *G. piovii* (subgroup B/clade 2) with “intermediate” microbiota, which has been observed independently using cpn60 barcode sequencing and clade-specific PCR (14, 22). It remains to be determined if *Gardnerella* species that do not regularly achieve numerical dominance in the microbiome contribute to establishment

TABLE 1 CLR-transformed relative abundance by clinical variables^a

Variable	<i>G. vaginalis</i>			<i>G. swidsinskii</i>			<i>G. leopoldii</i>			<i>G. plovii</i>			Genome species 3		
	Median (IQR)	P value		Median (IQR)	P value		Median (IQR)	P value		Median (IQR)	P value		Median (IQR)	P value	
Nugent category		<0.0001			0.0006										
Negative (<i>n</i> = 289)	0.21 (−0.13 to 6.36)	a		0.09 (−0.26 to 5.22)	a		−0.06 (−0.29 to 0.25)	a		−0.06 (−0.31 to 0.14)	a		−0.07 (−0.28 to 0.23)	a	0.03
Intermediate (<i>n</i> = 42)	7.6 (0.15 to 10.01)	b		0.11 (−0.29 to 8.63)	a, b		−0.06 (−0.24 to 0.44)	a		0.21 (−0.14 to 7.72)	b		0.07 (−0.22 to 4.62)	a	
BV (<i>n</i> = 64)	9.22 (6.0 to 11.77)	b		4.79 (−0.10 to 12.84)	b		0.002 (−0.28 to 7.64)	a		0.16 (−0.16 to 6.32)	b		0.04 (−0.26 to 6.61)	a	
Odor		0.001			0.001			0.11			0.22				0.15
No (<i>n</i> = 360)	0.35 (−0.09 to 4.77)			0.10 (−0.25 to 5.89)			−0.06 (−0.29 to 0.28)			−0.02 (−0.29 to 0.24)			−0.06 (−0.27 to 0.31)		
Yes (<i>n</i> = 33)	7.56 (4.77 to 10.77)			5.14 (0.03 to 12.80)			0.12 (−0.23 to 8.94)			0.05 (−0.20 to 4.99)			0.03 (−0.24 to 7.56)		
Irritation		0.13			0.20			0.97			0.58				0.25
No (<i>n</i> = 330)	0.46 (−0.10 to 8.66)			0.12 (−0.25 to 6.26)			−0.05 (−0.30 to 0.30)			−0.02 (−0.29 to 0.25)			−0.07 (−0.27 to 0.34)		
Yes (<i>n</i> = 63)	5.20 (0.02 to 9.11)			0.20 (−0.14 to 8.75)			−0.10 (−0.23 to 0.28)			−0.03 (−0.21 to 0.29)			0.006 (−0.25 to 5.10)		
Discharge		<0.0001			0.001			0.79			0.71				0.02
No (<i>n</i> = 331)	0.27 (−0.11 to 8.12)			0.09 (−0.25 to 5.65)			−0.05 (−0.28 to 0.30)			−0.02 (−0.28 to 0.26)			−0.06 (−0.28 to 0.28)		
Yes (<i>n</i> = 62)	7.52 (1.55 to 9.83)			2.44 (−0.05 to 11.98)			−0.10 (−0.29 to 0.28)			−0.05 (−0.23 to 0.35)			0.02 (−0.22 to 6.58)		

^aP values are from Kruskal-Wallis tests. Pairwise comparison significance for Nugent category is indicated using letters (a and b), where shared letters indicate no significant difference between those groups in pairwise Dunn tests with Benjamini-Hochberg *P* value adjustment. The center-log-ratio (CLR) transform is a ratio relative to the geometric mean log₂ relative abundance of all taxa (including non-*Gardnerella*); therefore, negative values indicate relative abundance less than the mean log₂ relative abundance of all taxa. Using log base 2 results in values that are in units of 2-fold increases, or halvings. A value of 0 indicates relative abundance that is exactly equal to the geometric mean abundance of all taxa. A value of 1 indicates a doubling relative to the mean, a value of 7 is 2⁷ greater relative abundance compared to the mean, and a value of −1 indicates a halving relative to the mean. Because the transformation is compositional in nature, the numbers do not reflect differences in absolute read counts but are rather in reference to proportionality of abundances, IQR, interquartile range.

and maintenance of dysbiosis. Robust and simple classification of *Gardnerella* isolates based on cpn60 barcode sequences will facilitate further characterization of these organisms within the new taxonomic framework and lead to identification of phenotypic features of the species that determine their ecological roles in the vaginal microbiome. In the clinical context, assessing longitudinal shifts in *Gardnerella* species abundance will also be important to evaluate natural or posttreatment changes. In future cohort studies, application of cpn60 barcode sequencing will provide new insight as to whether *Gardnerella* species diversity and differential distribution are an explanation for issues such as treatment failure and recurrent vaginal dysbiosis (23, 41) or for the failure of antimicrobial treatment in the prevention of preterm birth despite a strong association between vaginal dysbiosis and preterm delivery (42).

Conclusions. Our results provide a clear demonstration of the utility of cpn60 barcode sequencing for rapid, high-throughput determination of *Gardnerella* species abundance and distribution in the context of the entire vaginal microbiome with minimal sequencing effort. Furthermore, we demonstrated that 250 bp from either end of the barcode sequence is sufficient for discrimination of these species. Multiple *Gardnerella* species were detected in most women, although their proportional abundances varied widely. Significant cooccurrences were observed for *G. vaginalis* and *G. swidsinskii* and for *G. piovii* and genome species 3, whereas *G. swidsinskii* and *G. leopoldii* cooccurred less often than expected by chance. There was a significant association between Nugent category and relative abundances of *G. vaginalis*, *G. swidsinskii*, and *G. piovii* but not *G. leopoldii*. There were strong relationships between self-reported symptoms of abnormal odor and vaginal discharge with higher relative abundances of *G. vaginalis* and *G. swidsinskii*, but not with the other species.

MATERIALS AND METHODS

***Gardnerella* cpn60 sequence analysis.** cpn60 universal target sequences from 52 *Gardnerella* isolates representing the four named species (*G. vaginalis*, *G. piovii*, *G. swidsinskii*, and *G. leopoldii*) and nine additional genome species described by Vaneechoutte et al. (13) were retrieved from cpnDB (www.cpnDB.ca) (17, 43) and aligned using CLUSTALw. Interspecies nucleotide sequence similarities were calculated using Dnadist within PHYLIP (44). A bootstrapped phylogenetic tree was calculated using Seqboot, Dnadist (maximum likelihood option), and Neighbor. The type strain of *Alloscardovia omnicolens* (DSM 21503) was included as an outgroup. The consensus tree was computed with *consense* and branch lengths applied with *Fitch*. The tree was visualized using *FigTree* (v1.4.2).

Identification of *Gardnerella* spp. in vaginal microbiomes of Canadian women. Data from previously conducted studies by the VOGUE Research Group of the vaginal microbiome composition of nonpregnant, reproductive aged Canadian women were available for analysis. Nugent scores, cpn60 barcode sequence data, and self-reported symptoms were collected from healthy women (who were HIV negative and had not used any oral or intravaginal antibiotic or antifungal treatments in the past 4 weeks [$n = 310$]), women living with HIV ($n = 54$), and women who had at least four self-identified episodes of vulvovaginitis in the past 12 months ($n = 53$). Methods used for DNA extraction from vaginal swabs, cpn60 barcode PCR using universal primers, library preparation, and sequencing of amplicons are described in reference 14. Amplification primer sequences were removed using Cutadapt, followed by quality trimming with Trimmomatic (quality cutoff, 30; minimum length, 150). Quality-trimmed reads were loaded into QIIME2 (45) for sequence variant calling and read frequency calculation with DADA2 (46) and a truncation length of 250. For taxonomic identification, variant sequences were compared to the cpnDB_nr reference database (version 20190305, downloaded from www.cpnDB.ca) using watered-BLAST (47). The reference database includes representative sequences of each of the four named *Gardnerella* species and the nine additional genome species defined by Vaneechoutte et al. (13).

Cooccurrence and cluster analysis. For the following analyses, 4 samples with fewer than 500 sequence reads were removed from the collection of 417, leaving 413 samples. The average number of reads per sample was 8,669 (median, 5,103; range, 575 to 69,188). We determined pairwise cooccurrences using the presence or absence of each *Gardnerella* species or genome species in each sample. Presence was indicated if the raw number of sequence reads in the sample was ≥ 10 ; otherwise, the species was marked as absent. We calculated a Jaccard index of similarity (J) (48) for each pairwise combination and compared these to a probabilistic null model of species cooccurrence that takes into account observed frequencies to determine significance (27, 29) (File S1). P values were Benjamini-Hochberg (BH) corrected to a false-discovery rate of 0.05 (49). Jaccard dissimilarities ($1 - J$) were used to cluster the species using complete-linkage hierarchical clustering implemented in the *vegan* package for R (50).

Comparisons of clinical data and relative abundance. Relative abundances were calculated using the center-log-ratio (CLR) transformation as implemented in the ALDEx2 package (51) and Propr package (18). We compared CLR abundances among categories of clinical variables using Kruskal-Wallis tests. Significant omnibus tests were followed up with Dunn's *post hoc* tests with BH adjusted P values. For this

analysis, the sample size was reduced to 395 for which we had concurrent Nugent scoring data and 393 for which we had self-reported symptom data.

Ethics approval and consent to participate. Studies from which data were accessed for this substudy were approved by the University of British Columbia Children's & Women's Research Ethics Board (certificate numbers H10-02535, H11-00119, and H11-01912).

Availability of data and materials. The data sets supporting the results described in this article are available in the NCBI repository (BioProject accession numbers [PRJNA362575](#), [PRJNA278895](#), and [PRJNA528096](#)). R code for the cooccurrence analysis and data table are provided as supplemental material.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/IAI.00532-19>.

SUPPLEMENTAL FILE 1, XLSX file, 0.01 MB.

SUPPLEMENTAL FILE 2, XLSX file, 0.04 MB.

SUPPLEMENTAL FILE 3, XLSX file, 0.1 MB.

SUPPLEMENTAL FILE 4, PDF file, 0.03 MB.

ACKNOWLEDGMENTS

The VOGUE Research Group is Deborah Money, Alan Bocking, Sean Hemmingsen, Janet Hill, Gregor Reid, Tim Dumonceaux, Gregory Gloor, Matthew Links, Kieran O'Doherty, Patrick Tang, Julianne van Schalkwyk, and Mark Yudin.

We declare that we have no competing interests.

Financial support was provided by a joint Canadian Institutes of Health Research (CIHR) Emerging Team Grant and a Genome British Columbia (GBC) grant (reference number 108030) awarded to the VOGUE Research Group and by an NSERC Discovery Grant to J.E.H.

J.E.H. and A.Y.K.A. conceived the study, conducted the analysis, and wrote the paper. The VOGUE Research Group provided access to data for analysis and edited the paper. All authors read and approved the final manuscript.

REFERENCES

- Leopold S. 1953. Heretofore undescribed organism isolated from the genitourinary system. *U S Armed Forces Med J* 4:263–266.
- Schellenberg JJ, Patterson MH, Hill JE. 2017. *Gardnerella vaginalis* diversity and ecology in relation to vaginal symptoms. *Res Microbiol* 168: 837–844. <https://doi.org/10.1016/j.resmic.2017.02.011>.
- Nugent RP, Krohn MA, Hillier SL. 1991. Reliability of diagnosing bacterial vaginosis is improved by a standardized method of gram stain interpretation. *J Clin Microbiol* 29:297–301.
- Piot P, Van Dyck E, Peeters M, Hale J, Totten PA, Holmes KK. 1984. Biotypes of *Gardnerella vaginalis*. *J Clin Microbiol* 20:677–679.
- Benito R, Vazquez JA, Berron S, Fenoll A, Saez-Neito JA. 1986. A modified scheme for biotyping *Gardnerella vaginalis*. *J Med Microbiol* 21:357–359. <https://doi.org/10.1099/00222615-21-4-357>.
- Ingianni A, Petruzzelli S, Morandotti G, Pompei R. 1997. Genotypic differentiation of *Gardnerella vaginalis* by amplified ribosomal DNA restriction analysis (ARDRA). *FEMS Immunol Med Microbiol* 18:61–66. <https://doi.org/10.1111/j.1574-695X.1997.tb01028.x>.
- Hill JE, Goh SH, Money DM, Doyle M, Li A, Crosby WL, Links M, Leung A, Chan D, Hemmingsen SM. 2005. Characterization of vaginal microflora of healthy, nonpregnant women by chaperonin-60 sequence-based methods. *Am J Obstet Gynecol* 193:682–692. <https://doi.org/10.1016/j.jajog.2005.02.094>.
- Schellenberg JJ, Paramel Jayaprakash T, Withana Gamage N, Patterson MH, Vaneechoutte M, Hill JE. 2016. *Gardnerella vaginalis* subgroups defined by cpn60 sequencing and sialidase activity in isolates from Canada, Belgium and Kenya. *PLoS One* 11:e0146510. <https://doi.org/10.1371/journal.pone.0146510>.
- Links MG, Dumonceaux TJ, Hemmingsen SM, Hill JE. 2012. The chaperonin-60 universal target is a barcode for bacteria that enables *de novo* assembly of metagenomic sequence data. *PLoS One* 7:e49755. <https://doi.org/10.1371/journal.pone.0049755>.
- Paramel Jayaprakash T, Schellenberg JJ, Hill JE. 2012. Resolution and characterization of distinct cpn60-based subgroups of *Gardnerella vaginalis* in the vaginal microbiota. *PLoS One* 7:e43009. <https://doi.org/10.1371/journal.pone.0043009>.
- Ahmed A, Earl J, Retchless A, Hillier SL, Rabe LK, Cherpes TL, Powell E, Janto B, Eutsey R, Hillier NL, Boissy R, Dahlgren ME, Hall BG, Costerton JW, Post JC, Hu FZ, Ehrlich GD. 2012. Comparative genomic analyses of 17 clinical isolates of *Gardnerella vaginalis* provide evidence of multiple genetically isolated clades consistent with subspeciation into genovars. *J Bacteriol* 194:3922–3937. <https://doi.org/10.1128/JB.00056-12>.
- Richter M, Rosselló-Móra R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A* 106: 19126–19131. <https://doi.org/10.1073/pnas.0906412106>.
- Vaneechoutte M, Guschin A, Van Simaey L, Gansemans Y, Van Nieuwerburgh F, Cools P. 2019. Emended description of *Gardnerella vaginalis* and description of *Gardnerella leopoldii* sp. nov., *Gardnerella piovii* sp. nov. and *Gardnerella swidsinskii* sp. nov., with delineation of 13 genomic species within the genus *Gardnerella*. *Int J Syst Evol Microbiol* 69: 679–687. <https://doi.org/10.1099/ijsem.0.003200>.
- Albert AY, Chaban B, Wagner EC, Schellenberg JJ, Links MG, van Schalkwyk J, Reid G, Hemmingsen SM, Hill JE, Money D. 2015. A study of the vaginal microbiome in healthy Canadian women utilizing cpn60-based molecular profiling reveals distinct *Gardnerella* subgroup community state types. *PLoS One* 10:e0135620. <https://doi.org/10.1371/journal.pone.0135620>.
- Verbeke TJ, Sparling R, Hill JE, Links MG, Levin D, Dumonceaux TJ. 2011. Predicting relatedness of bacterial genomes using the chaperonin-60 universal target (cpn60 UT): application to *Thermoanaerobacter* species. *Syst Appl Microbiol* 34:171–179. <https://doi.org/10.1016/j.syapm.2010.11.019>.
- Katyal I, Chaban B, Hill JE. 2016. Comparative genomics of cpn60 defined *Enterococcus hirae* ecotypes and relationship of gene content differences to competitive fitness. *Microb Ecol* 72:917–930. <https://doi.org/10.1007/s00248-015-0708-2>.
- Vancuren SJ, Hill JE. 2019. Update on cpnDB: a reference database of

- chaperonin sequences. Database (Oxford) 2019:baz033. <https://doi.org/10.1093/database/baz033>.
18. Chaban B, Links MG, Jayaprakash T, Wagner EC, Bourque DK, Lohn Z, Albert AY, van Schalkwyk J, Reid G, Hemmingsen SM, Hill JE, Money DM. 2014. Characterization of the vaginal microbiota of healthy Canadian women through the menstrual cycle. *Microbiome* 2:23. <https://doi.org/10.1186/2049-2618-2-23>.
 19. Freitas AC, Chaban B, Bocking A, Rocco M, Yang S, Hill JE, Money DM. 2017. The vaginal microbiome of healthy pregnant women is less rich and diverse with lower prevalence of Mollicutes compared to healthy non-pregnant women. *Sci Rep* 7:9212. <https://doi.org/10.1038/s41598-017-07790-9>.
 20. Freitas AC, Bocking A, Hill JE, Money DM, VOGUE Research Group. 2018. Increased richness and diversity of the vaginal microbiota and spontaneous preterm birth. *Microbiome* 6:117. <https://doi.org/10.1186/s40168-018-0502-8>.
 21. Schellenberg JJ, Links MG, Hill JE, Dumonceaux TJ, Kimani J, Jaoko W, Wachihi C, Mungai JN, Peters GA, Tyler S, Graham M, Severini A, Fowke KR, Ball TB, Plummer FA. 2011. Molecular definition of vaginal microbiota in East African commercial sex workers. *Appl Environ Microbiol* 77:4066–4074. <https://doi.org/10.1128/AEM.02943-10>.
 22. Balashov SV, Mordechai E, Adelson ME, Gyax SE. 2014. Identification, quantification and subtyping of *Gardnerella vaginalis* in noncultured clinical vaginal samples by quantitative PCR. *J Med Microbiol* 63:162–175. <https://doi.org/10.1099/jmm.0.066407-0>.
 23. Hilbert DW, Schuyler JA, Adelson ME, Mordechai E, Sobel JD, Gyax SE. 2017. *Gardnerella vaginalis* population dynamics in bacterial vaginosis. *Eur J Clin Microbiol Infect Dis* 36:1269. <https://doi.org/10.1007/s10096-017-2933-8>.
 24. Gloor GB, Reid G. 2016. Compositional analysis: a valid approach to analyze microbiome high-throughput sequencing data. *Can J Microbiol* 62:692–703. <https://doi.org/10.1139/cjm-2015-0821>.
 25. Lovell D, Pawlowsky-Glahn V, Egozcue JJ, Marguerat S, Bähler J. 2015. Proportionality: a valid alternative to correlation for relative data. *PLoS Comput Biol* 11:e1004075. <https://doi.org/10.1371/journal.pcbi.1004075>.
 26. Erb I, Notredame C. 2016. How should we measure proportionality on relative gene expression data? *Theory Biosci* 135:21–36. <https://doi.org/10.1007/s12064-015-0220-8>.
 27. Mainali KP, Bewick S, Thielen P, Mehoke T, Breitwieser FP, Paudel S, Adhikari A, Wolfe J, Slud EV, Karig D, Fagan WF. 2017. Statistical analysis of co-occurrence patterns in microbial presence-absence datasets. *PLoS One* 12:e0187132. <https://doi.org/10.1371/journal.pone.0187132>.
 28. Ulrich W, Almeida-Neto M, Gotelli NJ. 2009. A consumer's guide to nestedness analysis. *Oikos* 118:3–17. <https://doi.org/10.1111/j.1600-0706.2008.17053.x>.
 29. Veech JA. 2013. A probabilistic model for analysing species co-occurrence. *Glob Ecol Biogeogr* 22:252–260. <https://doi.org/10.1111/j.1466-8238.2012.00789.x>.
 30. Quinn TP, Richardson MF, Lovell D, Crowley TM. 2017. propr: an R-package for identifying proportionally abundant features using compositional data analysis. *Sci Rep* 7:16252. <https://doi.org/10.1038/s41598-017-16520-0>.
 31. Shipitsyna E, Krysanova A, Khayrullina G, Shalepo K, Savicheva A, Guschin A, Unemo M. 2019. Quantitation of all four *Gardnerella vaginalis* clades detects abnormal vaginal microbiota characteristic of bacterial vaginosis more accurately than putative *G. vaginalis* sialidase A gene count. *Mol Diagn Ther* 23:139–147. <https://doi.org/10.1007/s40291-019-00382-5>.
 32. Numanovic F, Hukic M, Nurkic M, Gegic M, Delibegovic Z, Imamovic A, Pasic S. 2008. Importance of isolation and biotypization of *Gardnerella vaginalis* in diagnosis of bacterial vaginosis. *Bosn J Basic Med Sci* 8:270–276. <https://doi.org/10.17305/bjbm.2008.2932>.
 33. Briselden AM, Hillier SL. 1990. Longitudinal study of the biotypes of *Gardnerella vaginalis*. *J Clin Microbiol* 28:2761–2764.
 34. Aroutcheva AA, Simoes JA, Behbakht K, Faro S. 2001. *Gardnerella vaginalis* isolated from patients with bacterial vaginosis and from patients with healthy vaginal ecosystems. *Clin Infect Dis* 33:1022–1027. <https://doi.org/10.1086/323030>.
 35. Tosun I, Alpay Karaoğlu S, Ciftçi H, Buruk CK, Aydın F, Kiliç AO, Ertürk M. 2007. Biotypes and antibiotic resistance patterns of *Gardnerella vaginalis* strains isolated from healthy women and women with bacterial vaginosis. *Mikrobiyol Bul* 41:21–27. (In Turkish.)
 36. Pleckaityte M, Janulaitiene M, Lasickiene R, Zvirbliene A. 2012. Genetic and biochemical diversity of *Gardnerella vaginalis* strains isolated from women with bacterial vaginosis. *FEMS Immunol Med Microbiol* 65:69–77. <https://doi.org/10.1111/j.1574-695X.2012.00940.x>.
 37. Briselden AM, Moncla BJ, Stevens CE, Hillier SL. 1992. Sialidases (neuraminidases) in bacterial vaginosis and bacterial vaginosis-associated microflora. *J Clin Microbiol* 30:663–666.
 38. Wiggins R, Hicks SJ, Soothill PW, Millar MR, Corfield AP. 2001. Mucins and sialidases: their role in the pathogenesis of sexually transmitted infections in the female genital tract. *Sex Transm Infect* 77:402–408. <https://doi.org/10.1136/sti.77.6.402>.
 39. Robertson AM, Wiggins R, Horner PJ, Greenwood R, Crowley T, Fernandes A, Berry M, Corfield AP. 2005. A novel bacterial mucinase, glycosulfatase, is associated with bacterial vaginosis. *J Clin Microbiol* 43:5504–5508. <https://doi.org/10.1128/JCM.43.11.5504-5508.2005>.
 40. Hardy L, Cerca N, Jespers V, Vaneechoutte M, Crucitti T. 2017. Bacterial biofilms in the vagina. *Res Microbiol* 168:865–874. <https://doi.org/10.1016/j.resmic.2017.02.001>.
 41. Bradshaw CS, Morton AN, Hocking J, Garland SM, Morris MB, Moss LM, Horvath LB, Kuzevska I, Fairley CK. 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. <https://doi.org/10.1086/503780>.
 42. Nygren P, Fu R, Freeman M, Bougatsos C, Klebanoff M, Guise JM, US Preventive Services Task Force. 2008. Evidence on the benefits and harms of screening and treating pregnant women who are asymptomatic for bacterial vaginosis: an update review for the U.S. Preventive Services Task Force. *Ann Intern Med* 148:220–233. <https://doi.org/10.7326/0003-4819-148-3-200802050-00008>.
 43. Hill JE, Penny SL, Crowell KG, Goh SH, Hemmingsen SM. 2004. cpnDB: a chaperonin sequence database. *Genome Res* 14:1669–1675. <https://doi.org/10.1101/gr.2649204>.
 44. Felsenstein J. 1989. PHYLIP—phylogeny inference package (version 3.2). *Cladistics* 5:164–166.
 45. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7:335–336. <https://doi.org/10.1038/nmeth.f.303>.
 46. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods* 13:581–583. <https://doi.org/10.1038/nmeth.3869>.
 47. Schellenberg J, Links MG, Hill JE, Dumonceaux TJ, Peters GA, Tyler S, Ball B, Severini A, Plummer FA. 2009. Pyrosequencing of the chaperonin-60 universal target as a tool for determining microbial community composition. *Appl Environ Microbiol* 75:2889–2898. <https://doi.org/10.1128/AEM.01640-08>.
 48. Janson S, Vegelius J. 1981. Measures of ecological association. *Oecologia* 49:371–376. <https://doi.org/10.1007/BF00347601>.
 49. Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol* 57:289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
 50. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Henry M, Stevens H, Wagner H. 2012. vegan: Community Ecology Package. R package version 2.0–10. <http://CRAN.R-project.org/package=vegan>.
 51. Fernandes AD, Reid JN, Macklaim JM, McMurrough TA, Edgell DR, Gloor GB. 2014. Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome* 2:15. <https://doi.org/10.1186/2049-2618-2-15>.