

## Accepted Manuscript

Microbial profiling of cpn60 universal target sequences in artificial mixtures of vaginal bacteria sampled by nylon swabs or self-sampling devices under different storage conditions

John J. Schellenberg, Angela Yena Oh, Janet E. Hill



PII: S0167-7012(17)30064-7  
DOI: doi: [10.1016/j.mimet.2017.03.003](https://doi.org/10.1016/j.mimet.2017.03.003)  
Reference: MIMET 5130

To appear in: *Journal of Microbiological Methods*

Received date: 20 December 2016  
Revised date: 5 March 2017  
Accepted date: 6 March 2017

Please cite this article as: John J. Schellenberg, Angela Yena Oh, Janet E. Hill , Microbial profiling of cpn60 universal target sequences in artificial mixtures of vaginal bacteria sampled by nylon swabs or self-sampling devices under different storage conditions. The address for the corresponding author was captured as affiliation for all authors. Please check if appropriate. Mimet(2017), doi: [10.1016/j.mimet.2017.03.003](https://doi.org/10.1016/j.mimet.2017.03.003)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

©2017. This manuscript version is made available under the CC-BY-NC-ND 4.0 license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

# REVISED

## **Microbial profiling of *cpn60* universal target sequences in artificial mixtures of vaginal bacteria sampled by nylon swabs or self-sampling devices under different storage conditions**

John J. Schellenberg<sup>#</sup>, Angela Yena Oh<sup>#</sup>, Janet E. Hill<sup>\*</sup>

Department of Veterinary Microbiology, University of Saskatchewan, Saskatoon, Canada

Running head: Validating HerSwab<sup>TM</sup> for vaginal microbiome profiling

<sup>#</sup> These authors contributed equally to this work.

<sup>\*</sup> Corresponding Author

Western College of Veterinary Medicine, 52 Campus Dr., University of Saskatchewan,  
Saskatoon, CANADA S7N 5B4, Janet.Hill@usask.ca

## ABSTRACT

The vaginal microbiome is increasingly characterized by deep sequencing of universal genes. However, there are relatively few studies of how different specimen collection and sample storage and processing influence these molecular profiles. Here, we evaluate molecular microbial community profiles of samples collected using the HerSwab<sup>TM</sup> self-sampling device, compared to nylon swabs and under different storage conditions. In order to minimize technical variation, mixtures of 11 common vaginal bacteria in simulated vaginal fluid medium were sampled and DNA extracts prepared for massively parallel sequencing of the *cpn60* universal target (UT). Three artificial mixtures imitating commonly observed vaginal microbiome profiles were easily distinguished and proportion of sequence reads correlated with the estimated proportion of the organism added to the artificial mixtures. Our results indicate that *cpn60* UT amplicon sequencing quantifies the proportional abundance of member organisms in these artificial communities regardless of swab type or storage conditions, although some significant differences were observed between samples that were stored frozen and thawed prior to DNA extraction, compared to extractions from samples stored at room temperature for up to 7 days. Our results indicate that an on-the-market device developed for infectious disease diagnostics may be appropriate for vaginal microbiome profiling, an approach that is increasingly facilitated by rapidly dropping deep sequencing costs.

**Keywords:** vaginal self-sampling; HerSwab<sup>TM</sup>; quantitative microbiome profiling; *cpn60* universal target; hierarchical clustering; UniFrac; ALDEx2

## INTRODUCTION

The vaginal microbiome, defined as the community of microbes that colonize the female reproductive tract, plays a critical but still poorly understood role in reproductive health. The identification and quantification of bacterial types in the vaginal microbiome is increasingly measured by molecular census using clone libraries, high-throughput sequencing or quantitative PCR of phylogenetically informative genes (Hill et al., 2005, Schellenberg et al., 2011, van de Wijkert et al., 2014). A 552 bp portion of the *cpn60* gene (the universal target, UT) is an ideal molecular barcode since it is protein-coding and therefore more variable than 16S rRNA genes between closely related taxa (Goh et al., 1996, Hill et al., 2006), and has been shown to provide similar phylogenetic resolution to whole genome-based comparisons of closely related groups of organisms that cannot be easily distinguished with 16S rRNA gene sequences (Katyal et al., 2015, Links and Hill, 2012, Schellenberg et al., 2016, Verbeke et al., 2011). *cpn60*-based sequencing is particularly useful in the context of the vaginal microbiome since it can easily distinguish at least four subgroups of *Gardnerella vaginalis* that are likely different species (Schellenberg, et al., 2016). The normal range of microbiological profiles, temporal fluctuations and clinical manifestations of the vaginal microbiota in relation to mucosal inflammation, pre-term birth and increased acquisition of exogenous pathogens can now be defined in quantitative terms. However, the contribution of technical sources of variation in microbial community profile, due to swab type, sample quality, storage conditions, DNA extraction methods, sequencing platform, etc. have yet to be adequately distinguished from biological sources of variation, complicating the interpretation of various studies which use different combinations of methods. In particular, research attention has focused on the appropriateness and acceptability of self-collected vaginal swabs for various clinical and research applications, comparing favourably

with physician-collected specimens in terms of diagnostics and participant comfort (Arias et al., 2016, Lunny et al., 2015, Serlin et al., 2002, Shafer et al., 2003, Strauss et al., 2005).

In this study, we set out to evaluate the HerSwab<sup>TM</sup> (Eve Medical, Inc., Toronto, ON), a novel vaginal self-sampling device designed for infectious disease diagnostics, as a potential tool for vaginal microbiome profiling, defining proportional abundance of different organisms in this environment. Evaluation was based on a comparison of microbiome profiles generated from HerSwab<sup>TM</sup> collected samples to those created using an established protocol based on samples collected using nylon swabs. HerSwab<sup>TM</sup> devices have previously been shown to be effective and acceptable for diagnosis of sexually transmitted infections (Arias, et al., 2016). We demonstrate that the HerSwab<sup>TM</sup> collects a biological sample that is sufficient and suitable for downstream processes for *cpn60* UT deep sequencing analysis, including DNA extraction, indexing PCR and sequencing using the MiSeq platform. We also confirm that the proportion of reads per phylotype per sample from the deep sequencing results is significantly correlated with the estimated proportion of each organism in mixtures, further evidence that *cpn60* UT deep sequencing is a quantitative method in molecular microbial ecology.

## MATERIALS AND METHODS

### Culture models of vaginal micro-organisms

In order to standardize organism input, we made three mixtures with varying proportions of 11 vaginal bacteria from our culture collection (Table 1), representing simplified versions of commonly observed “community state types” (CST) in healthy Canadian women (Albert et al., 2015), including CST I (*Lactobacillus crispatus*-dominated), CST IV (mixed or intermediate) and CST V (*Gardnerella* subgroup A-dominated). *L. crispatus* ATCC 33820 and *L. jensenii*

ATCC 25258 (American Type Culture Collection, Manassas, VA) were revived on Mann-Rogosa-Sharpe Agar (EMD Chemicals Inc., Mississauga, ON). *L. iners* ATCC 13335 and all other organisms, isolates from our culture collection of vaginal micro-organisms (Schellenberg et al., 2012, Schellenberg, et al., 2016), were revived on Columbia Agar with 5% sheep's blood (BD Biosciences, Mississauga, ON), in anaerobic jars with GasPak anaerobic sachets (BD Biosciences, Mississauga, ON) at 37°C. The identities of the micro-organisms were confirmed based on known phenotype, Gram stain characteristics, and Sanger sequencing of the *cpn60* UT amplified with the same universal primers used for sequencing library preparation. *cpn60* sequences were compared to the cpnDB reference database to confirm species identity ([www.cpnadb.ca](http://www.cpnadb.ca)) (Hill et al., 2004). Micro-organisms growing on plates were resuspended with sterile swabs in 0.85% saline solution to McFarland standard ~4 and an aliquot of 100 µL transferred to 10 mL NYC III (ATCC # 1685) medium (per L: HEPES buffer, 4 g; proteose peptone, 15 g; sodium chloride, 5 g; yeast extract, 2 g, with 10% v/v deactivated horse serum and 10 mL filter-sterilized 50% w/v glucose solution added after autoclaving). After 24 h incubation at 37°C, cell concentration was estimated by transferring 100 µL of sample to duplicate wells of a clear-bottomed 96-well plate and optical density at 595 nm assessed in a Vmax microplate reader (Molecular Devices Inc., Sunnyvale, CA). Micro-organisms from plates or liquid cultures were mixed at different ratios (adjusted for optical density, see Table 1), centrifuged and resuspended in 2 mL Medium Simulating Vaginal Fluid or MSVF (per L: glucose, 10 g; bovine glycogen, 10 g; lactic acid, 2 g; acetic acid, 1 g; bovine serum albumin, 2 g; porcine stomach mucin, 0.5 g; urea, 0.25 g; sodium chloride, 3.5 g; potassium chloride, 1.5 g; Tween 80, 1.08 g; cysteine hydrochloride, 0.5 g; all from Sigma-Aldrich, Oakville, ON) (Geshnizgani and

Onderdonk, 1992, Juarez Tomas and Nader-Macias, 2007). Aliquots of un-inoculated MSVF were reserved for use as negative controls.

### **Sampling of artificial vaginal microbiome mixtures**

For each of the three CST (CST-I, -IV and -V), one set of triplicate nylon swabs (N = 3) and four sets of triplicate HerSwabs<sup>TM</sup> (N = 12) were immersed sequentially in vials containing the bacterial mixtures (Fig. 1A). One set of triplicate nylon swabs and one set of HerSwabs<sup>TM</sup> were stored immediately at -80°C (NY vs. HS). The three remaining sets of triplicate HerSwabs<sup>TM</sup> were stored at room temperature for 1, 4 and 7 days to simulate shipping of self-collected samples to a laboratory (D1 vs. D4 vs. D7). Swabs were weighed before and after sampling to quantify fluid retained in milligrams, with differences between nylon swabs (N=9) and HerSwabs<sup>TM</sup> (N=34, 2 values missing) assessed using the Wilcoxon rank sum test in R statistical software, v. 3.3.2 (R Core Team, 2015).

### **DNA extraction and estimation of total bacterial cells**

DNA was extracted from all swabs and un-inoculated MSVF using the MagMAX DNA extraction kit (Life Technologies, Burlington, ON), following the manufacturer's directions. DNA concentration and assessment of DNA quality was measured by the NanoDrop spectrophotometer (Thermo Fisher Scientific Inc., Burlington, ON) and the Qubit fluorometer with the high-sensitivity double-stranded DNA quantification kit (Life Technologies Inc., Burlington, ON). Since the degenerate *cpn60* primers produce a product too large for optimal SYBR green qPCR, estimation of total bacterial cells was carried out for all samples by quantitative polymerase chain reaction (qPCR) targeting the 16S rRNA gene in reference to a

standard curve, as previously described (Chaban et al., 2013). Briefly, previously published 16S rRNA primers (Lee et al., 1996) were applied in a qPCR assay against a standard curve composed of plasmids containing the V3 region of the 16S rRNA gene from an uncultured human fecal bacterium (Fernando et al., 2010) at concentrations of  $10^0$  to  $10^7$  copies/reaction. Reactions in duplicate consisted of 12.5  $\mu$ l 2 $\times$  SsoFast EvaGreen Supermix (BioRad, Mississauga, ON), 1  $\mu$ l forward primer (SRV3-1 : 5' - CGG YCC AGA CTC CTA C - 3'), 1  $\mu$ l reverse primer (SRV3-2 : 5' - TTA CCG CGG CTG CTG GCA C - 3', both primers 400 nM final concentration) and 2  $\mu$ l template DNA in a final volume of 25  $\mu$ l. No template controls were included with all qPCR runs. Reactions were amplified on a CFX96 real-time thermocycler (BioRad, Mississauga, ON), using the following program: 95°C for 3 min., followed by 40 cycles of 95°C for 15 sec, 62°C for 15 sec and 72°C for 15 sec, with fluorescence measured after each annealing step and in a subsequent dissociation curve consisting of 81 cycles at 0.5°C increments from 55°C to 95°C. qPCR results were analyzed using CFX Manager Software v3.1 (BioRad, Mississauga, ON).

### ***cpn60* universal target PCR and sequencing library preparation**

Procedures were adapted from the “16S Metagenomic Sequencing Library Preparation” manual (available at <http://www.illumina.com/content/dam/illumina-support/documents>) for amplification of the *cpn60* UT instead of the 16S rRNA gene. Initial amplification of the *cpn60* UT was carried out using a 1:3 molar ratio of degenerate *cpn60* primer pairs H279/H280:H1612/H1613, as previously described (Schellenberg et al., 2009), except that the base primers were modified with Illumina sequencing adaptors (in bold underline) at the 5' and 3' ends (Forward primer #1 – M279 : 5' - **TCG TCG GCA GCG TCA GAT GTG TAT AAG**

**AGA CAG** GAI III GCI GGI GAY GGI ACI ACI AC - 3'; Reverse primer #1 – M280 : 5' - **GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G** YKI YKI TCI CCR AAI CCI GGI GCY TT - 3' ; Forward primer #2 – M1612 : 5' - **TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG** GAI III GCI GGY GAC GGY ACS ACS AC – 3' ; Reverse primer #2 – M1613 : 5' **GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G** CGR CGR TCR CCG AAG CCS GGI GCC TT - 3'). DNA extract from swab samples and uninoculated MSVF (2 µl) was added to a 50 µl (final volume) PCR Master Mix, including (per reaction): 5 µl 10x PCR buffer, 2.5 µl 50 mmol l<sup>-1</sup> magnesium chloride, 1.25 µl dNTPs, 0.3 µl 10 mmol l<sup>-1</sup> forward primer #1 (M279), 0.3 µl 10 mmol l<sup>-1</sup> reverse primer #1 (M280), 0.9 µl 10 mmol l<sup>-1</sup> forward primer #2 (M1612), 0.9 µl 10 mmol l<sup>-1</sup> reverse primer #2 (M1613), and 0.5 µl 50 U µl<sup>-1</sup> Platinum Taq DNA Polymerase. Thermocycling parameters were 40 cycles of 95°C for 30 sec, 50°C for 30 sec, 72°C for 30 sec, with a final extension of 72°C for 7 min. No template controls were included in all runs. PCR products (~700bp) were gel-purified based on size in an ethidium bromide-free 1% agarose gel, using known migration of target band size against a multi-colour loading dye and subsequent extraction using the QIAEX II gel extraction kit (QIAGEN, Inc., Toronto, ON). Resulting extracts were used for attaching Nextera XT dual indices in a mixture containing 5 µl gel-purified amplicon, 5 µl each index primer, 25 µl 2x Kapa HiFi HotStart ReadyMix (KAPA Biosciences, Inc., Wilmington, MA) in a final volume of 50 µl per reaction. Thermocycling parameters were 8 cycles of 95°C for 30 sec, 55°C for 30 sec, 72°C for 30 sec, with a final extension of 72°C for 5 min. Indexed amplicons were purified with AMPure XP beads (Beckman Coulter, Inc., Mississauga, ON). Average size of the final amplicon libraries (649 bp) was calculated using the DNA High-Sensitivity chip on the 2100 Bioanalyzer (Agilent Technologies, Inc., Santa Clara, CA) and concentration in ng/µL assessed

using the Qubit fluorometer (Life Technologies, Inc., Burlington, ON). Molar concentration was calculated using the formula provided in the manual and all libraries normalized to 4 nmol l<sup>-1</sup> prior to pooling. Dilution, denaturing and loading of the Nano v2 500 cycle sequencing cartridge with pooled libraries was carried out and flow cell loaded onto the Illumina Miseq with 10% PhiX as directed by the manufacturer. In order to maximize coverage of the *cpn60* UT, the number of cycles was set at 400 for the first read and 100 for the second read. Our previous experience has shown that reads >150 bp provide adequate species identification in database comparisons (Schellenberg, et al., 2009).

### Deep sequencing analysis

Demultiplexed .fastq files were trimmed based on quality scores using Trimmomatic (www.usadellab.org) with a 4 base sliding window, minimum average quality of 15, LEADING and TRAILING both set to 3, and MINLEN set to 150. This eliminated the shorter R2 reads from the dataset, leaving one quality-trimmed .fastq file (R1) per sample. These sequences were mapped on to *cpn60* UT sequences of the 11 bacterial species comprising the synthetic samples, using Bowtie 2 (Langmead and Salzberg, 2012), to determine the numbers of reads in each library derived from each input organism. All subsequent analyses were carried out in R statistical software v.3.3.2 (R Core Team, 2015). For each sample, the proportion of sequencing reads mapped to each organism was visualized as a heatmap and hierarchical clustering of samples was performed based on the Jensen-Shannon distance matrix, as previously described (Albert, et al., 2015). In order to compare expected proportions of each organism based on volume with the proportion of sequencing reads for each organism, the adjusted volume ( $\mu\text{l}$  / OD<sub>595</sub>) of each organism added to the final mixture was plotted against sequencing reads and

assessed using Pearson's test in the *ggplot2* package. The extent of variability between CST (CST-I vs. CST-IV vs. CST-V) replicates (Replicate #1 vs. Replicate #2 vs. Replicate #3), swab type (HS vs. NY), days at room temperature (D1 vs. D4 vs. D7) and storage condition (HS/NY vs. D1/4/7) was assessed using two analytical packages: Unifrac (Lozupone et al., 2011) and ALDEx2 (Fernandes et al., 2013). The Unifrac algorithm calculates the overall phylogenetic distance between pairs of samples, based on presence and abundance of sequencing reads for each organism, resulting in a value on a scale of 0 to 1, where 0 indicates most similar and 1 indicates least similar. In this study, the weighted Unifrac metric (including abundance information) was applied to address the hypothesis that phylogenetic distance between pairs of samples is likely to be higher when collected by different devices or frozen and thawed prior to analysis. A boxplot of median UniFrac distances by CST, replicates and experimental conditions was created in R and compared using the Wilcoxon rank sum test. ALDEx2 was used in order to assess which specific elements being measured (in this case, abundance of sequencing reads for each organism) are significantly differentially abundant across tested conditions. In this study, ALDEx2 was applied in order to assess hypotheses that each specific organism is significantly differentially abundant across devices or storage conditions. ALDEx2 uses Wilcoxon and Kruskal-Wallis statistics to compare the distribution of taxa, while GLM and Welch's t statistics are used to compare means. Differences between and within conditions were plotted using ALDEx2 functions and organisms that were significantly different by all four statistical tests were coloured red, while those that were significantly different by one or two tests were coloured green.

## RESULTS

### Lower fluid retention by HerSwabs<sup>TM</sup>

Nylon swabs absorbed approximately three times greater sample mass when compared to HerSwabs™ (W=306,  $p<0.0001$  by Wilcoxon rank sum test, Fig. 1B, right). Although measurement of extracted DNA was below the spectrophotometric detection limit, high-sensitivity fluorometry showed that the amount of DNA extracted was also greater for nylon swabs, as expected (W=324,  $p<0.0001$  by Wilcoxon rank sum test, Fig. 1B, middle). Quantitative PCR showed that 16S rRNA gene copy numbers ranged from  $10^5$ - $10^7$  per swab, with significantly higher estimated copy number in nylon swabs (W=270,  $p=0.001$ , Fig. 1B, left), consistent with higher fluid retention. Despite low DNA concentrations in all swabs, amplicon of the expected size (~700 bp) was visible in the agarose gel after *cpn60* UT PCR, with sufficient concentration (>4 nM) after the indexing PCR step for loading onto the flow cell for sequencing. Extracts from un-inoculated MSVF and no template controls yielded no PCR product. These results indicate that HerSwabs™ provided sufficient material for downstream sequencing procedures.

### Sequencing results and detection of input organisms

The 500 sequencing cycles were allocated asymmetrically (400 cycles for R1 and 100 cycles for R2 rather than 250 for both reads) in order to maximize coverage of the *cpn60* UT sequence. This procedure is done routinely in our laboratory to facilitate *de novo* assembly of amplicon sequence data into full-length UT sequences. Although assembly was not performed in this study, generating 400 bp reads facilitated Bowtie2 mapping results. Approximately 1 million raw reads were obtained and subjected to quality trimming as described in the Methods. The inclusion of a minimum length requirement of 150 eliminated all R2 reads. Approximately 11% of raw R1 reads were removed for poor quality or inadequate length leaving 447,369 high quality

R1 reads for analysis, with an average of 9,518 reads per library (range 1,308-51,383) and an average read length of 354 bases. A total of 12,785 high-quality reads (2.8% of total) did not map to the database. Sequences corresponding to *Prevotella disiens* and *Peptoniphilus harei* were not detected in any sample. *Streptococcus anginosus* sequences were detected in 34/45 samples and *Prevotella bivia* was detected in 44/45. The other 8 input organisms were detected in all samples (Supplementary Table 1).

### **Samples cluster primarily by CST, not by swab type**

Hierarchical clustering demonstrated that deep sequencing clearly differentiates the three mixtures imitating CST-I, CST-IV and CST-V (Fig. 2A). HerSwab™ replicates stored at -80°C did not cluster separately from nylon swab samples, indicating that profiles generated from HerSwabs™ were highly similar to those from nylon swabs stored at -80°C. However, sub-clusters were observed for swabs stored at -80°C (NY and HS) and swabs stored at room temperature (D1, D4 and D7), in CST-IV and CST-V, but not CST-I. Most strikingly, in CST-IV and CST-V, *P. bivia* appears to be more abundant and *Gardnerella C* less abundant in -80°C samples compared to room temperature samples. For D1 samples, all three replicates clustered together in all three CST, while replicates in D4 samples did not cluster together and replicates in D7 samples only clustered together in CST-V.

### **Sequence abundance correlates with quantitative estimates of input organisms**

Percent sequencing reads attributed to each organism correlated strongly with estimated percent of adjusted volume added to each artificial mixture ( $R^2 = 0.95$ ,  $p < 0.0001$ )(Fig. 2B), indicating that deep sequencing provides quantitative estimates of the proportions of component

organisms. The correlation was similarly strong when the most abundant species, *L. crispatus* in CST I, was removed from the analysis ( $R^2 = 0.86$ ,  $p < 0.0001$ ). While input and output were similar overall, the intended proportion (adjusted volume) of *Gardnerella* A/B/C and *P. bivia* in CST-V was higher than the outcome proportion (sequencing reads), indicating the sequence-based profiles under-estimated the abundance of these organisms in CST-V samples.

### **Profile reproducibility and phylogenetic distance by UniFrac**

Phylogenetic profiles generated from HerSwabs<sup>TM</sup> were highly similar to those generated from standard nylon swabs, as calculated using the UniFrac algorithm (Fig. 3A). As expected, the greatest and least phylogenetic distance was observed between samples from different CST and different replicates, respectively ( $p < 0.0001$  by Wilcoxon rank sum test). Distance values across experimental conditions were significantly higher than values across replicates, but much closer to replicate values when considering values across CST. Further, most of the phylogenetic distance across conditions was observed when comparing -80°C samples to room temperature (RT) samples (Fig 3A, left panel), not when comparing -80°C samples to -80°C samples or RT to RT samples. Within -80°C comparisons and within RT comparisons were only marginally higher than within replicates, although the difference was still statistically significant ( $p = 0.02$  by Wilcoxon rank sum test).

### **Organism-level statistical comparisons by ALDEx2**

No differences in the abundance of specific organisms were observed between HerSwabs<sup>TM</sup> and nylon swabs, as assessed by the Kruskal-Wallis p-value generated by ALDEx2 (Fig. 3B, right panel). As expected, all p-values were significant ( $< 0.05$ ) when comparing CST,

while none were significant when comparing replicates (Fig. 3B, left panel). Although p-values were generally much lower than those for replicates when comparing across all conditions, across samples stored at -80°C, and across room temperature samples, none were statistically significant. However, a number of organisms were significantly different when comparing -80°C samples to room temperature samples (Fig. 3B, right panel). To visualize these differences in the context of differences observed between CST and replicates, ALDEx2 values for differences within and between comparators were plotted (Fig. 3C) and each assessed using four statistical tests (see Methods) corrected for multiple comparisons by the Benjamini-Hochberg test *post hoc*. Most differences in CST were statistically significant using all four tests (red circles), while none were significant when comparing replicates, across -80°C samples or across room temperature samples (small grey circles). A subset of organisms were significantly different by the GLM test alone or by the GLM and Welch's t tests only (green circles) when comparing room temperature samples with previously frozen samples (Fig. 3C, bottom right and Supplementary Table 2). However, these organisms were not significantly different in all ALDEx2 comparisons of fresh vs. frozen samples (Supplementary Table 2), indicating that the effect of freezing on organism abundance is inconsistent. Eight of the 9/161 comparisons across room temperature and previously frozen samples where a significant difference was indicated by both GLM and Welch's t test, were accounted for by *Prevotella bivia*, which was detected at lower levels in room temperature samples than frozen (Supplementary Table 2). The visually apparent differences in *Gardnerella vaginalis* C abundance in CST-IV and CST-V (Fig. 2) were not fully supported by ALDEx2 results, where differences were detected in only three comparisons between room temperature and previously frozen CST-IV samples and only by GLM. Overall, our results demonstrate that swab type and room temperature storage for 7 days do not

significantly influence *cpn60* UT sequence profiles, and that observed differences due to frozen storage were modest compared to overall differences between CST.

## DISCUSSION

Vaginal samples are routinely collected with different types of medical swabs used to prepare wet mounts and Gram stained smears for observation of clue cells and diagnosis of bacterial vaginosis, and increasingly for molecular diagnostics. Self-collected vaginal samples have been widely used for a wide range of both research and clinical applications (Arias, et al., 2016, Lunny, et al., 2015, Shafer, et al., 2003), increasing participant comfort and convenience and allowing for more frequent or routine sampling without the need of physician examinations. In this study, we performed an evaluation of the HerSwab<sup>TM</sup> device, consisting of a rubber-like head with soft-toothed bristles, housed in a graspable, moulded plastic sheath that is twistable at one end, unsheathing and re-sheathing the swab during sampling to minimize contamination with external surfaces. Notably, this design facilitates targeted cervical sampling for Pap smears and human papillomavirus screening, minimizing exposure to other vaginal surfaces. Once sheathed, the sample remains dried on the swab head during transport to the lab, providing a sample that is suitable for molecular analysis of component organisms.

Although HerSwabs<sup>TM</sup> absorbed less specimen than nylon swabs, as reflected in lower DNA concentrations in swab extracts and lower bacterial concentrations determined by quantitative PCR of 16S rRNA copy number, enough cells were collected by the HerSwabs<sup>TM</sup> to yield sufficient high-quality DNA for deep sequencing analysis in this study. Total 16S rRNA copy numbers for the synthetic samples overall were  $10^5$ - $10^7$  per swab, which is within the wide range of  $10^4$ - $10^9$  per swab that we have quantified in actual vaginal samples using the same method (Albert, et al., 2015). Although the total bacterial load of the synthetic samples was

similar to actual samples, they were less complex, containing only the 11 added species. Total community size or density may be an important parameter to evaluate, but this would require knowledge of the total sample weight or volume, which could be challenging with self-collected samples. In this study we were limited to reporting 16S rRNA copy numbers per swab. Knowledge of the actual sample mass or volume is likely to be challenging for self-collected samples, especially those like the HerSwab<sup>TM</sup> that produce dried samples.

All analytical techniques confirmed that CST was the primary factor differentiating samples, indicating that swab type or storage conditions did not affect our ability to classify profiles into CST categories (Fig. 2). Given the wide variation in the number of reads generated for each sample, we have reported species abundance in terms of the proportion of the specific sequence against total sequencing reads. Expression of microbiome profiles as proportional data is common, but the appropriateness of this approach remains an active topic of discussion in the literature. Analysis tools such as ALDEx2 (Fernandes, et al., 2013), utilized in this study, have been specifically developed to overcome challenges associated with compositional data. Replicates within conditions were also remarkably similar. Some differences were detected in the abundance of individual organisms when frozen samples were compared to those stored at room temperature (Fig. 3C, bottom right and Supplementary Table 2). However, the differences were not consistently associated with conditions. For example, there was a significant difference in abundance of *Prevotella bivia* in HerSwab<sup>TM</sup> stored at room temperature for one day and HerSwab<sup>TM</sup> stored frozen (D1 vs. HS), but no difference between *P. bivia* abundance between D7 HerSwab<sup>TM</sup> and HerSwab<sup>TM</sup> stored frozen (D7 vs. HS). These results illustrate the idiosyncratic effects of storage, perhaps particularly on rare low abundance organisms. A previous study has shown that frozen storage affected flow cytometric counts of bacteria eluted

from vaginal swabs (Schellenberg et al., 2008), and preferential lysis of some cell types during freeze-thaw cycles may explain observed differences in proportional abundance of bacterial phylotypes in deep sequencing read data. However, results of studies of the effects of storage conditions on microbiome profile results are mixed and difficult to compare since some investigators focus only on overall community structure and abundant organisms, rather than individual species (Dominianni et al., 2014, Song et al., 2016). In general, variation introduced by different preservation and storage methods does not overwhelm larger differences between sample types (Bai et al., 2012, Lauber et al., 2010), individuals sampled (Dominianni, et al., 2014), or in the case of this study, CST. We are unaware of any other studies evaluating room temperature storage of vaginal samples. Differences of borderline statistical significance we observed between samples stored frozen and at room temperature samples will require further elucidation.

Overall correlation between sequence read abundance and input amount of each bacterial species was strong (Fig. 2), confirming that read abundance is a good indicator of relative abundance in the original sample. This has previously been shown in *cpn60* based microbiome profiles of human vaginal (Chaban et al., 2014) and respiratory tract microbiomes (Chaban, et al., 2013), where organism abundance based on sequence read numbers was validated by direct species-specific qPCR of samples. We did observe some cases where sequence read abundance was lower than expected (*Gardnerella* A/B/C and *P. bivia* in CST V), which could be the result of PCR bias within that particular mixture of species. Failure to detect community members that were shown to be amplifiable using the universal PCR primers is not likely explained by differences in swab type, but rather by PCR bias or insufficient sequencing depth. Differences in the efficiency of PCR amplification of various targets using “universal” primers is an issue

regardless of the target, and is difficult to quantify since it is affected by the amounts and types of target sequences present in any sample (Hong et al., 2009). Biases also vary among targets and primer sets (Hill et al., 2010, Schellenberg, et al., 2009). Notably, although PCR bias may account for observed differences between actual community composition and proportional sequence read abundance, it would not affect the comparison of the sequence-based microbiome profiles to each other since samples were processed identically, ensuring that biases were uniformly imposed. Our results indicate that while not strictly quantitative, *cpn60* universal PCR and deep sequencing creates a microbiome profile that reflects the relative abundance of species in the original microbial community.

As sequencing costs are reduced and procedures streamlined, routine diagnostic applications for high-throughput sequencing of vaginal microbial communities, long predicted (Brotman and Ravel, 2008), are increasingly feasible. A key benefit of *cpn60* universal target (UT) sequencing is that it can reliably distinguish four different subgroups of *Gardnerella vaginalis*, likely different species, that cannot be reliably distinguished using 16S rRNA gene sequences (Schellenberg, et al., 2016). Only recently have studies begun to define the range of microbial communities commonly observed in healthy women and specific groups around the world, with preliminary information available about the demographic distribution of microbial communities, shifts over time, and correlations with immune parameters and other health outcomes (Brotman et al., 2014, Chaban, et al., 2014, Gautam et al., 2015, Ravel et al., 2013, Schellenberg et al., 2012, Schellenberg, et al., 2011). The potential to generate quantitative and statistically verifiable estimates for individual organisms in artificial mixed cultures is corroborated by several recent studies describing quantitative deep sequencing output for a variety of biological applications using the MiSeq platform (Greiff et al., 2014, Katsuoka et al.,

2014, Masser et al., 2013, Zheng et al., 2015). Combined with user-friendly devices for self-collection of samples, these technologies offer enormous potential for research and diagnostics for women's health.

In conclusion, we have demonstrated that HerSwabs<sup>TM</sup> provide reproducible, molecular profiles that are consistent with those from nylon swab samples in our *in vitro* model, and that minor variations in profiles were associated with storage conditions but not swab type. HerSwabs<sup>TM</sup> have been shown to be an appealing technology for women participating in research concerning screening of sexually transmitted diseases (Arias, et al., 2016), increasing the feasibility of frequent and/or longitudinal vaginal microbiome profiling for research, diagnostics and treatment monitoring (Ravel, et al., 2013, Santiago et al., 2012). Further studies with human participants will be required to confirm the effectiveness of HerSwab<sup>TM</sup> self-sampling in quantitative microbiome profiling.

## ACKNOWLEDGEMENTS

The authors are grateful to Champika Fernando and Aline Freitas for technical support. Funding was provided by an Engage grant to JEH from the Natural Sciences and Engineering Research Council of Canada. JJS was supported by fellowships from the Saskatchewan Health Research Foundation and the Canadian Institutes of Health Research. AYO was supported by an Undergraduate Student Research Award from the Natural Sciences and Engineering Research Council of Canada. The funders had no role in the design, execution or interpretation of the study results.

**CONFLICT OF INTEREST**

No conflict of interest declared.

ACCEPTED MANUSCRIPT

## REFERENCES

- Albert, A.Y., Chaban, B., Wagner, E.C., Schellenberg, J.J., Links, M.G., van Schalkwyk, J., Reid, G., Hemmingsen, S.M., Hill, J.E., 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.
- Arias, M., Jang, D., Gilchrist, J., Luinstra, K., Li, J., Smieja, M., Chernesky, M.A., 2016. Ease, comfort, and performance of the HerSwab vaginal self-sampling device for the detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae*. Sex. Transm. Dis. 43, 125-129.
- Bai, G., Gajer, P., Nandy, M., Ma, B., Yang, H., Sakamoto, J., Blanchard, M.H., Ravel, J., Brotman, R.M., 2012. Comparison of storage conditions for human vaginal microbiome studies. PLoS ONE. 7, e36934.
- Brotman, R.M., Ravel, J., 2008. Ready or not: the molecular diagnosis of bacterial vaginosis. Clin. Infect. Dis. 47, 44-46.
- Brotman, R.M., Shardell, M.D., Gajer, P., Fadrosch, D., Chang, K., Silver, M.I., Viscidi, R.P., Burke, A.E., Ravel, J., Gravitt, P.E., 2014. Association between the vaginal microbiota, menopause status, and signs of vulvovaginal atrophy. Menopause. 21, 450-458.
- Chaban, B., Albert, A., Links, M.G., Gardy, J., Tang, P., Hill, J.E., 2013. Characterization of the upper respiratory tract microbiomes of patients with pandemic H1N1 influenza. PLoS ONE. 8, e69559.

- Chaban, B., Links, M.G., Paramel Jayaprakash, T., Wagner, E.C., Bourque, D.K., Lohn, Z., Albert, A.Y.K., van Schalkwyk, J., Reid, G., Hemmingsen, S.M., Hill, J.E., Money, D.M., 2014. Characterization of the vaginal microbiota of healthy Canadian women through the menstrual cycle. *Microbiome*. 2, 23.
- Dominianni, C., Wu, J., Hayes, R.B., Ahn, J., 2014. Comparison of methods for fecal microbiome biospecimen collection. *BMC Microbiol.* 14, 103.
- Fernandes, A.D., Macklaim, J.M., Linn, T.G., Reid, G., Gloor, G.B., 2013. ANOVA-like differential expression (ALDEx) analysis for mixed population RNA-Seq. *PLoS ONE*. 8, e67019.
- Fernando, W.M.U., Hill, J.E., Zello, G.A., Tyler, R.T., Dahl, W.J., Van Kessel, A.G., 2010. Diets supplemented with chickpea or its main oligosaccharide component raffinose modify fecal microbial composition in healthy adults. *Beneficial Microbes*. 1, 197-207.
- Gautam, R., Borgdorff, H., Jespers, V., Francis, S.C., Verhelst, R., Mwaura, M., Delany-Moretlwe, S., Ndayisaba, G., Kyongo, J.K., Hardy, L., Menten, J., Crucitti, T., Tsivtsivadze, E., Schuren, F., van de Wijkert, J.H., Vaginal Biomarkers Study, G., 2015. Correlates of the molecular vaginal microbiota composition of African women. *BMC Infect. Dis.* 15, 86.
- Geshnizgani, A.M., Onderdonk, A.B., 1992. Defined medium simulating genital tract secretions for growth of vaginal microflora. *J. Clin. Microbiol.* 30, 1323-1326.

Goh, S.H., Potter, S., Wood, J.O., Hemmingsen, S.M., Reynolds, R.P., Chow, A.W., 1996.

HSP60 gene sequences as universal targets for microbial species identification: studies with coagulase-negative staphylococci. *J. Clin. Microbiol.* 34, 818-823.

Greiff, V., Menzel, U., Haessler, U., Cook, S.C., Friedensohn, S., Khan, T.A., Pogson, M., Hellmann, I., Reddy, S.T., 2014. Quantitative assessment of the robustness of next-generation sequencing of antibody variable gene repertoires from immunized mice. *BMC Immunol.* 15, 40.

Hill, J.E., Fernando, W.M.U., Zello, G.A., Tyler, R.T., Dahl, W.J., Van Kessel, A.G., 2010. Improvement of the representation of bifidobacteria in fecal microbiota metagenomic libraries by application of the *cpn60* universal primer cocktail. *Appl. Environ. Microbiol.* 76, 4550–4552.

Hill, J.E., Goh, S.H., Money, D.M., Doyle, M., Li, A., Crosby, W.L., Links, M., Leung, A., Chan, D., Hemmingsen, S.M., 2005. Characterization of vaginal microflora of healthy, nonpregnant women by chaperonin-60 sequence-based methods. *Am. J. Obstet. Gynecol.* 193, 682-692.

Hill, J.E., Penny, S.L., Crowell, K.G., Goh, S.H., Hemmingsen, S.M., 2004. *cpnDB*: a chaperonin sequence database. *Genome Res.* 14, 1669-1675.

Hill, J.E., Town, J.R., Hemmingsen, S.M., 2006. Improved template representation in *cpn60* PCR product libraries generated from complex templates by application of a specific mixture of PCR primers. *Environ. Microbiol.* 8, 741-746.

Hong, S., Bunge, J., Leslin, C., Jeon, S., Epstein, S.S., 2009. Polymerase chain reaction primers miss half of rRNA microbial diversity. *ISME J.* 3, 1365-1373.

Juarez Tomas, M.S., Nader-Macias, M.E., 2007. Effect of a medium simulating vaginal fluid on the growth and expression of beneficial characteristics of potentially probiotic lactobacilli. In: A.-M. Vilas (Ed.), *Communicating Current Research and Educational Topics and Trends in Applied Microbiology*, Formatex, Badajoz, Spain, pp. 732–740.

Katsuoka, F., Yokozawa, J., Tsuda, K., Ito, S., Pan, X., Nagasaki, M., Yasuda, J., Yamamoto, M., 2014. An efficient quantitation method of next-generation sequencing libraries by using MiSeq sequencer. *Anal. Biochem.* 466, 27-29.

Katyal, I., Chaban, B., Hill, J.E., 2015. Comparative genomics of *cpn60* defined *Enterococcus hirae* ecotypes and relationship of gene content differences to competitive fitness. *Microbial Ecol.* 72, 917-930.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nature Meth.* 9, 357-359.

Lauber, C.L., Zhou, N., Gordon, J.I., Knight, R., Fierer, N., 2010. Effect of storage conditions on the assessment of bacterial community structure in soil and human-associated samples. *FEMS Microbiol. Lett.* 307, 80-86.

Lee, D.H., Zo, Y.G., Kim, S.J., 1996. Nonradioactive method to study genetic profiles of natural bacterial communities by PCR-single-strand-conformation polymorphism. *Appl. Environ. Microbiol.* 62, 3112-3120.

Links, M.G., Hill, J.E., 2012. cpn60 universal target region as a barcode for bacteria, Canadian Society of Microbiologists 62nd Annual Meeting, Vancouver, BC.

Lozupone, C., Lladser, M.E., Knights, D., Stombaugh, J., Knight, R., 2011. UniFrac: an effective distance metric for microbial community comparison. *ISME J.* 5, 169-172.

Lunny, C., Taylor, D., Hoang, L., Wong, T., Gilbert, M., Lester, R., Krajden, M., Ogilvie, G., 2015. Self-collected versus clinician-collected sampling for *Chlamydia* and *Gonorrhea* screening: A systemic review and meta-analysis. *PLoS ONE.* 10, e0132776.

Masser, D.R., Berg, A.S., Freeman, W.M., 2013. Focused, high accuracy 5-methylcytosine quantitation with base resolution by benchtop next-generation sequencing. *Epigenetics Chromatin.* 6, 33.

R Core Team, 2015. *R: A Language and Environment for Statistical Computing*, Vienna, Austria

Ravel, J., Brotman, R.M., Gajer, P., Ma, B., Nandy, M., Fadrosh, D.W., Sakamoto, J., Koenig, S.S., Fu, L., Zhou, X., Hickey, R.J., Schwebke, J.R., Forney, L.J., 2013. Daily temporal dynamics of vaginal microbiota before, during and after episodes of bacterial vaginosis. *Microbiome.* 1, 29.

Santiago, G.L., Tency, I., Verstraelen, H., Verhelst, R., Trog, M., Temmerman, M., Vancoillie, L., Decat, E., Cools, P., Vaneechoutte, M., 2012. Longitudinal qPCR study of the dynamics of *L. crispatus*, *L. iners*, *A. vaginae*, (sialidase positive) *G. vaginalis*, and *P. bivia* in the vagina. PLoS ONE. 7, e45281.

Schellenberg, J., Blake Ball, T., Lane, M., Cheang, M., Plummer, F., 2008. Flow cytometric quantification of bacteria in vaginal swab samples self-collected by adolescents attending a gynecology clinic. J. Microbiol. Meth. 73, 216-226.

Schellenberg, J., Links, M.G., Hill, J.E., Dumonceaux, T.J., Peters, G.A., Tyler, S., Ball, B., Severini, A., Plummer, F.A., 2009. Pyrosequencing of the chaperonin-60 universal target as a tool for determining microbial community composition. Appl. Environ. Microbiol. 75, 2889-2898.

Schellenberg, J.J., Card, C.M., Ball, T.B., Mungai, J.N., Irungu, E., Kimani, J., Jaoko, W., Wachihi, C., Fowke, K.R., Plummer, F.A., 2012. Bacterial vaginosis, HIV serostatus and T-cell subset distribution in a cohort of East African commercial sex workers: retrospective analysis. AIDS. 26, 387-393.

Schellenberg, J.J., Dumonceaux, T.J., Hill, J.E., Kimani, J., Jaoko, W., Wachihi, C., Mungai, J.N., Lane, M., Fowke, K.R., Ball, T.B., Plummer, F.A., 2012. Selection, phenotyping and identification of acid and hydrogen peroxide producing bacteria from vaginal samples of Canadian and East African women. PLoS ONE. 7, e41217.

Schellenberg, J.J., Links, M.G., Hill, J.E., Dumonceaux, T.J., Kimani, J., Jaoko, W., Wachihi, C., Mungai, J.N., Peters, G.A., Tyler, S., Graham, M., Severini, A., Fowke, K.R., Ball, T.B., Plummer, F.A., 2011. Molecular definition of vaginal microbiota in East African commercial sex workers. *Appl. Environ. Microbiol.* 77, 4066-4074.

Schellenberg, J.J., Paramel Jayaprakash, T., Withana Gamage, N., Patterson, M.H., Vanechoutte, M., Hill, J.E., 2016. *Gardnerella vaginalis* subgroups defined by cpn60 sequencing and sialidase activity in isolates from Canada, Belgium and Kenya. *PLoS ONE*. 11, e0146510.

Serlin, M., Shafer, M.A., Tebb, K., Gyamfi, A.A., Moncada, J., Schachter, J., Wibbelsman, C., 2002. What sexually transmitted disease screening method does the adolescent prefer? Adolescents' attitudes toward first-void urine, self-collected vaginal swab, and pelvic examination. *Arch. Pediatr. Adolesc. Med.* 156, 588-591.

Shafer, M.A., Moncada, J., Boyer, C.B., Betsinger, K., Flinn, S.D., Schachter, J., 2003. Comparing first-void urine specimens, self-collected vaginal swabs, and endocervical specimens to detect *Chlamydia trachomatis* and *Neisseria gonorrhoeae* by a nucleic acid amplification test. *J. Clin. Microbiol.* 41, 4395-4399.

Song, S.J., Amir, A., Metcalf, J.L., Amato, K.R., Xu, Z.Z., Humphrey, G., Knight, R., 2016. Preservation methods differ in fecal microbiome stability, affecting suitability for field studies. *mSystems*. 1, e00021-00016.

Strauss, R.A., Eucker, B., Savitz, D.A., Thorp, J.M., Jr., 2005. Diagnosis of bacterial vaginosis from self-obtained vaginal swabs. *Infect. Dis. Obstet. Gynecol.* 13, 31-35.

van de Wijgert, J.H., Borgdorff, H., Verhelst, R., Crucitti, T., Francis, S., Verstraelen, H., Jespers, V., 2014. The vaginal microbiota: what have we learned after a decade of molecular characterization? *PLoS ONE*. 9, e105998.

Verbeke, T.J., Sparling, R., Hill, J.E., Links, M.G., Levin, D., Dumonceaux, T.J., 2011. Predicting relatedness of bacterial genomes using the chaperonin-60 universal target (*cpn60* UT): application to *Thermoanaerobacter* species. *Syst. Appl. Microbiol.* 34, 171-179.

Zheng, W., Tsompana, M., Ruscitto, A., Sharma, A., Genco, R., Sun, Y., Buck, M.J., 2015. An accurate and efficient experimental approach for characterization of the complex oral microbiota. *Microbiome*. 3, 48.

## FIGURE LEGENDS

**Fig. 1:** A. Experimental overview. Artificial mixtures of vaginal bacteria representing three community state types (CST-I, CST-IV, CST-V) were sampled with one set of triplicate nylon swabs (NY = 3) and one set of triplicate HerSwabs™ (HS = 3) and stored at -80°C, as well as three sets of triplicate HerSwabs™ stored at room temperature for 1, 4 and 7 days (D1 = 3, D4 = 3, D7 = 3). B. Nylon swabs absorbed significantly more sample than HerSwabs™, as indicated by greater sample mass (left), high concentration DNA extracts (middle) and higher 16S rRNA gene copy number (right). P values show significant differences by Wilcoxon rank sum test. Note that for HerSwabs™, N=34, due to two missing values.

**Fig. 2:** A. Hierarchical clustering of all samples based on Jensen-Shannon distance. Heatmap shows proportion of total quality trimmed reads for each organism for all samples. The bars allow for visual aid of the classification of each sample. Heatmap shows the proportional abundance of the organism (row) in each sample (column) in a red color gradient. RT = Room temperature. Expected results based on input amounts of each organism in each CST are indicated to the left of the main heatmap. B. Correlation of percent adjusted volume (volume adjusted for optical density of culture, see Methods) and percent sequencing reads, with significance assessed using Pearson's product-moment correlation in R.

**Fig. 3:** A. Profile reproducibility across replicates as measured by phylogenetic distance using the weighted UniFrac metric. Range of values is 0 to 1, representing the range of all inter-sample comparisons (0 is most similar, 1 is least similar). Statistical comparisons by Wilcoxon rank sum

test in R. B. Organism-level differences between sample sets using ALDEx2, showing plots of Kruskal-Wallis p-values, with significance level indicated by dashed grey line. C. Scatterplots of ALDEx2 between- and within-difference values for individual organisms across CST, replicates and experimental conditions. Organisms with significant p-values by all four statistical tests (see Methods) are shown as large orange circles, those significant by 1-2 statistical tests as large blue circles, and those significant in no tests as small grey circles. Dashed lines circumscribe significant differences (abline = 1, -1). RT = Room temperature, NY = nylon swab, HS = HerSwab<sup>TM</sup>, D1/4/7 = HerSwabs<sup>TM</sup> Day 1/4/7 (RT).

## SUPPORTING INFORMATION

**Supplementary Table 1.** Number of sequence reads mapped by Bowtie2 on to each species comprising the synthetic vaginal microbiome communities.

**Supplementary Table 2.** ALDEx2 results for comparison of abundance of each species across storage temperature. Corrected p values for GLM, Welch's t test, Wilcoxon and Kruskal-Wallis are shown. Significant values are highlighted in red.

**Table 1** : Proportion of vaginal micro-organisms in artificial mixtures imitating community state types

| Micro-organisms                     | CST-I             |           |       |                   | CST-IV    |       |                   | CST-V     |       |                   |
|-------------------------------------|-------------------|-----------|-------|-------------------|-----------|-------|-------------------|-----------|-------|-------------------|
|                                     | OD <sub>595</sub> | Vol. (μl) | % AV* | % TR <sup>†</sup> | Vol. (μl) | % AV* | % TR <sup>†</sup> | Vol. (μl) | % AV* | % TR <sup>†</sup> |
| <i>Lactobacillus crispatus</i> (LC) | 0.239             | 1701      | 94.3  | 87.4              | 292       | 26.9  | 16.1              | 11        | 0.8   | 2.0               |
| <i>Lactobacillus iners</i> (LI)     | 0.046             | 20        | 0.2   | 1.2               | 30        | 0.5   | 1.6               | 10        | 0.1   | 0.6               |
| <i>Lactobacillus jensenii</i> (LJ)  | 0.179             | 25        | 2.2   | 4.4               | 60        | 8.9   | 10.7              | 10        | 1.2   | 2.0               |
| <i>Gardnerella vaginalis</i> A (GA) | 0.179             | 20        | 0.8   | 2.3               | 239       | 16.4  | 24.4              | 561       | 30.6  | 41.5              |
| <i>Gardnerella vaginalis</i> B (GB) | 0.103             | 10        | 0.2   | 0.3               | 293       | 11.6  | 10.5              | 181       | 5.7   | 5.4               |
| <i>Gardnerella vaginalis</i> C (GC) | 0.069             | 32        | 1.0   | 2.9               | 330       | 16.4  | 28.0              | 522       | 20.6  | 27.7              |
| <i>Atopobium vaginae</i> (AV)       | 0.069             | 21        | 0.3   | 0.1               | 180       | 4.7   | 2.0               | 360       | 7.5   | 4.6               |
| <i>Peptoniphilus harei</i> (PH)     | 0.103             | 2         | <0.1  | --                | 18        | 0.7   | --                | 5         | 0.2   | --                |
| <i>Streptococcus anginosus</i> (SA) | 0.147             | 10        | 0.3   | <0.1              | 114       | 6.4   | 0.3               | 8         | 0.4   | 0.1               |
| <i>Prevotella bivia</i> (PB)        | 0.206             | 10        | 0.5   | 1.3               | 94        | 7.5   | 6.4               | 522       | 32.9  | 16.1              |
| <i>Prevotella disiens</i> (PD)      | 0.151             | 0.1       | <0.1  | --                | 0.1       | <0.1  | --                | 0.1       | <0.1  | --                |

\*Percent Adjusted Volume: The percent of each organism based on total volume adjusted per 0.1 OD<sub>595</sub>. <sup>†</sup>Percent Total Reads: The percent of deep sequencing reads mapped to the *cpn60* UT sequence for each organism.

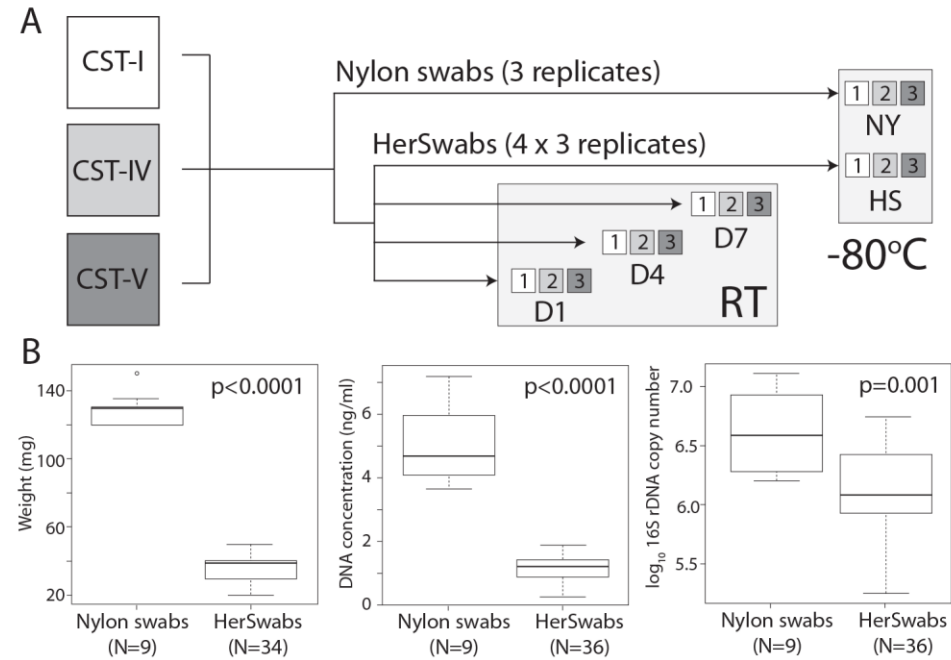


Fig. 1

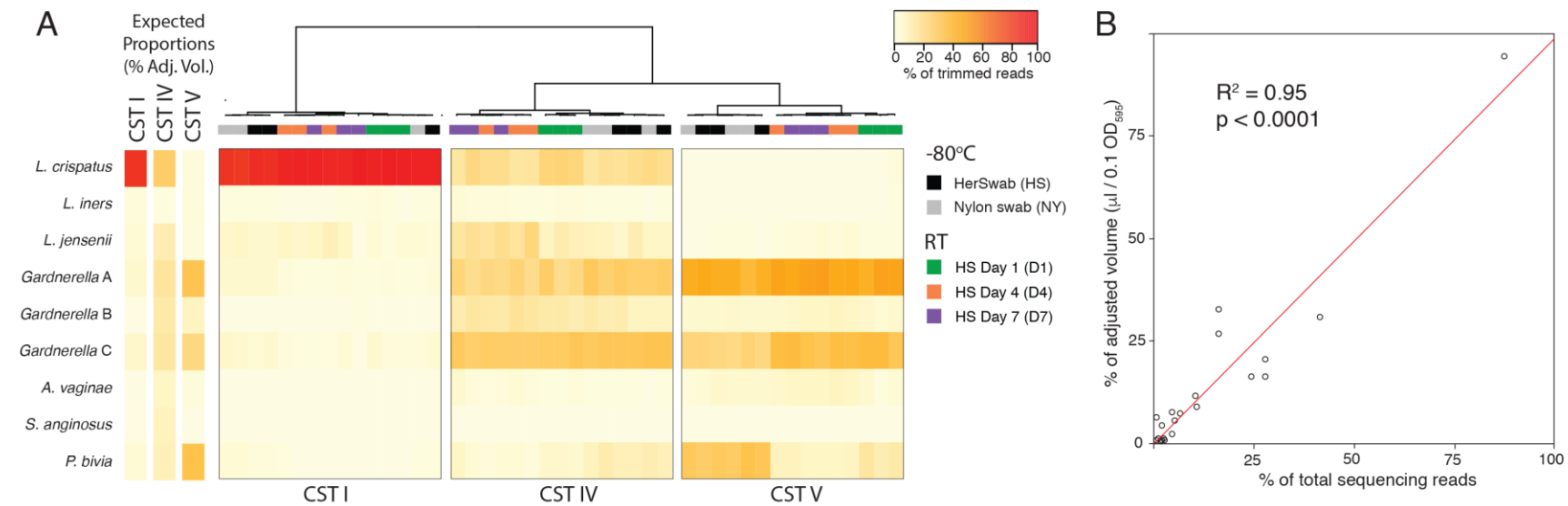


Fig. 2

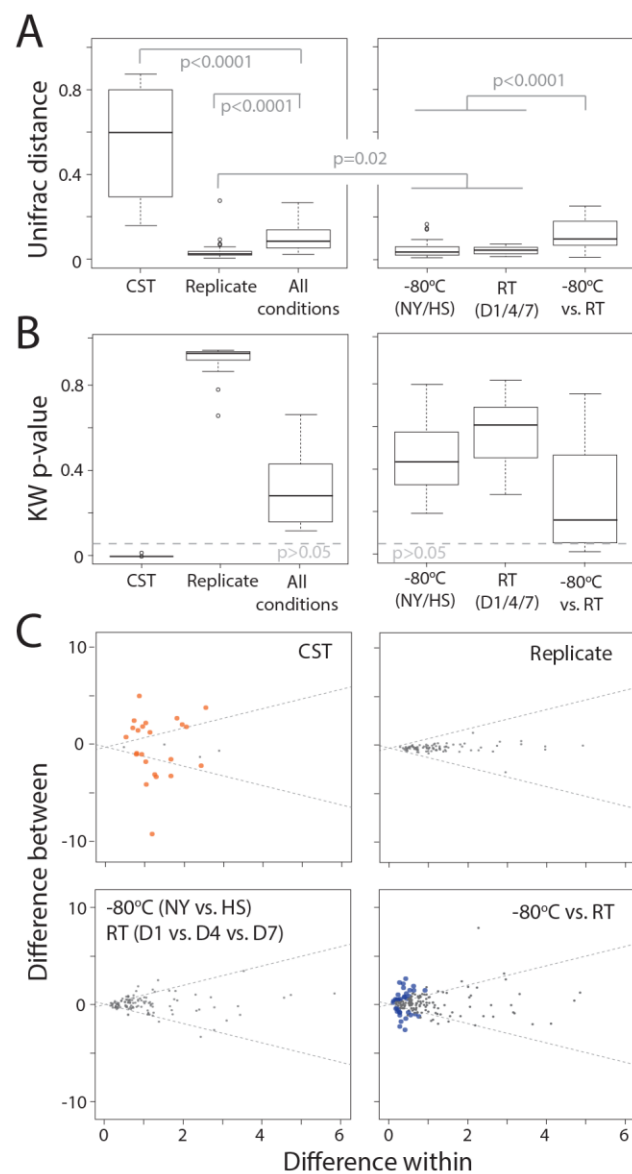


Fig. 3

### Highlights

- On-market vaginal self-sampling device provides sample for microbiome profiling
- Device allows room-temperature storage of samples for at least 7 days
- Minor variations in microbiome profiles due to storage conditions but not swab type
- Microbiome profiles based on cpn60 amplicon sequencing are highly reproducible
- Strong correlation between sequence read number and organism abundance