TITLE

Amplification of the V5 – V8 region of the 16S rRNA gene effectively speciates medically important genital tract *Lactobacillus* species in the upper female genital tract

AUTHORS

Jessica L. O’Callaghan^{a, b}, Dana Willner^{c, d}, Melissa Buttini^e, Flavia Huygens^{a, b}, Elise S. Pelzer*^{a, b}

AUTHOR AFFILIATIONS

^a School of Biomedical Sciences, Faculty of Health, Queensland University of Technology, Brisbane, Queensland, Australia 4001

^b Institute of Health and Biomedical Innovation, Faculty of Health, Queensland University of Technology, Brisbane, Queensland, Australia 4001

^c Australian Centre for Ecogenomics, University of Queensland, St Lucia, Queensland, Australia

^d Department of Computer Science, College of William and Mary, Williamsburg, VA 23187, USA

^e The Wesley Hospital, Auchenflower, Queensland, Australia, 4066

** corresponding author*

Elise S. Pelzer
Queensland University of Technology
GPO Box 2434
Brisbane, QLD 4001
+ 61 3138 0542
e.pelzer@qut.edu.au

Running title

Speciation of genital tract lactobacilli

ABSTRACT

Background: The endometrial cavity is an upper genital tract site largely heralded as sterile, however, advances in culture-independent, next generation sequencing technology have revealed that this site harbours a rich microbial community which includes multiple *Lactobacillus* species. These bacteria are considered to be the most common non-pathogenic genital tract commensals. Next-generation sequencing of the female lower genital tract has revealed significant variation amongst microbial community composition with respect to *Lactobacillus* sp. in samples collected from healthy and diseased women. The aim of this study was to evaluate the ability of the 16S rRNA gene to characterize genital tract lactobacilli to species-level taxonomy.

Methods: Samples were interrogated for the presence of microbial DNA using two-step next generation sequencing technology to exploit the V5–V8 regions of the 16S rRNA gene and compared to standard speciation using qPCR.

Results: The V5-V8 region of the 16S rRNA gene has sufficient sequence variation within frequently encountered genital tract lactobacilli to allow accurate determination of relative abundance within the community, and speciation for several key community members without completing additional experimentation.

Conclusions: Next-generation sequencing of clinical genital tract isolates is an effective method for high throughput identification to species-level of key *Lactobacillus* sp.

KEYWORDS

Lactobacillus sp.; genital tract; 16S rRNA; pyrosequencing; qPCR; speciation

IMPORTANCE

Human microbiome experiments, including the low biomass organs such as the upper genital tract, require the development of consensus protocols to ensure accurate comparison between such studies and our data forms an important foundation for future protocols.

This paper provides evidence to support the selection of the V5-V8 regions of the 16S rRNA gene improved *Lactobacillus* speciation using next generation sequencing technology. The choice of variable region for broad-range amplification in microbiome studies is important due to preferential primer binding associated with some genera based on nucleotide sequence patterns. By utilising the V5-V8 region, multiple species of *Lactobacillus* can be characterised with relative confidence.

79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104

INTRODUCTION

Molecular microbiology techniques have changed our ability to identify microbial communities, revolutionizing the way we assess female genital tract microbiomes. In cultivation-dependent studies, greater than 95% of the vaginal microbiota in healthy women was classified as lactobacilli. The advent of cultivation-independent technology platforms has provided evidence to suggest that in up to two-thirds of healthy women, the lactobacilli were co-aggregated with a diverse group of microbial community members, and in some cases did not dominate (Fettweis et al., 2012; Klebanoff et al., 1991). Lactobacilli establish niche dominance through co-aggregation, competitive inhibition, production of metabolic acids, hydrogen peroxide, and antimicrobial components including bacteriocins (Amabebe and Anumba, 2018). The discovery that lactobacilli do not dominate the genital tract of all healthy women suggests that: there is redundancy in function and protection based on community membership; and all lactobacilli may not provide the same level of protection in the genital tract environment. This discovery casts doubt over the long-held view that a healthy female genital tract is characterized by a *Lactobacillus* sp.-dominant microbiota. The ability to confidently assign lower order taxonomic classification to lactobacilli is critical in advancing our understanding of the protective role played by the various species within this genus in reproductive health. The objective of this study was to examine the discriminatory power of current molecular microbiology techniques for identification of genital tract lactobacilli.

METHODS

Patient cohort, sample collection and genomic DNA preparation

Clinical sample cohorts were constructed as previously described (Pelzer et al., 2018). Genomic DNA was extracted from individual samples prior to pooling using a modified protocol with the Qiagen QiAMP Mini DNA extraction kit (Qiagen, Australia) as previously described (Pelzer et al., 2018).

Details of ethical approval

All patients recruited for this study provided written informed consent. Ethical approval was obtained from the review boards of UnitingCare Health, Human Research Ethics Committee and Queensland University of Technology Human Ethics Committee.

Next-generation sequencing

The 16S rRNA PCR assay was performed using the previously published primers, 803F (5'-TTA GAT ACC CTG GTA GTC -3') and 1392R (5'-ACG GGC GGT GTG TRC -3') and PCR cycling conditions (Willner et al., 2014). Fusion primers with 454 adaptor sequences were ligated to the 803F and 1392R primers to amplify the V5 and V8 regions of the 16S rRNA gene (Willner et al., 2014). PCR reactions were performed as previously described (Pelzer et al., 2018). The five frequently encountered genital tract *Lactobacillus* sp. were aligned using the SILVA database to determine the degree of variation within the V5-V8 regions of the 16S rRNA gene. The annealing site of the sequencing primers is marked on the alignment (Figure 1a).

***Lactobacillus* sp.-specific quantitative real-time PCR**

Quantitative real-time PCR assays were performed using previously published primer pairs (Table 1) and cycling conditions. A standard curve was generated using *L. gasseri* ATCC strain 19992. Primer

annealing was confirmed using species-specific alignment of the five *Lactobacillus* sp. interrogated in this study (Figure 1b).

Taxonomic classification

Sequence clustering and operational taxonomic unit (OTU) selection was performed using a modified version of CD-HIT-OTU-454 which does not remove singleton clusters (Liu et al., 2011). Taxonomy was assigned to representative sequences by comparison to the latest build of the Greengenes database using BLAST, and OTU tables were constructed from the output using a custom Perl script (McDonald et al., 2012).

***Lactobacillus* Phylogenetic Trees**

Full-length 16S rRNA sequences for *Lactobacillus* spp. (accession numbers: AB680529.1, AB690249.1, AB668940.1.1, AB008203.1.1, AB425941.1.1, AB008206.1, AF243169.1, AF243167.1, CP018809.253324, CP018809.1516019, CP018809.1347636, CP018809.500868, AB547127.1, AB517146.1, AB932527.1, AB008209.1, HZ485829.7, LG085736.7, LF134126.7, LG104504.7), *Pediococcus pentosaceus* (accession numbers: AB018215.1 and AB362987.1), and *Bacillus subtilis* (accession numbers: AP012496.9810 and AP012496.30276) were downloaded from the SILVA database using the web interface (www.arb-silva.de). Sequences were aligned using ClustalW (Thompson et al., 1994) with the default settings. MEGA7 (Kumar et al., 2016) was used to generate the best-known maximum likelihood (ML) tree using a Jukes-Cantor model and 1000 bootstrapping iterations. The ML tree was visualised and edited within FigTree (<http://tree.bio.ed.ac.uk>) and Adobe Illustrator (Figure 4).

To generate the V5-V8 region phylogenetic tree the same full length 16S rRNA sequences from above were imported into Geneious along with two *Escherichia coli* sequences downloaded from SILVA (accession number: AB045730.1 and AB045731.1). Sequences were aligned using standard Geneious alignment and trimmed to include the variable regions V5-V8 (nucleotides 751 – 1300 for the *E.coli*

sequence). Trimmed sequences were then aligned using ClustalW with default settings and imported into MEGA7 and a tree was constructed and edited same as above (Figure 4).

Hierarchical clustering

A dissimilarity matrix was generated based on the relative abundances of *Lactobacillus spp.* in the pyrosequenced and qPCR analysed samples using the vegdist function in the vegan package in R with the Bray-Curtis dissimilarity metric (J. et al., 2018). Hierarchical clustering was performed using the hclust function in R with 'average' linkage (UPGMA) (Team, 2013). Clustering and relative abundances were visualized in a heatmap with associated dendrogram using the heatmap.2 function from the R package ggplots (Warnes et al., 2005).

RESULTS

NGS resolution of genital tract *Lactobacillus* sp. OTUs to genus and species

Ten OTUs were attributed to *Lactobacillus* sp. (*Lactobacillus* sp. genus level (n = 2), *L. crispatus* (n = 2), *L. iners* (n = 3), *L. intestinalis* (n = 1), *L. jensenii* (n = 1) and *L. vaginalis* (n = 1)). The majority of *Lactobacillus* sp. OTUs (8/10) were resolved to the genus and species level exploiting the V5-V8 regions of the 16S rRNA gene.

***Lactobacillus* species-specific quantitative real-time PCR assay comparison to NGS output**

The species-specific quantitative real-time PCR assays confirmed the identity and relative abundance of *L. crispatus*, *L. jensenii* and *L. iners* in clinical genital tract samples. Two of the species that underwent pyrosequencing identification were not included due to low abundance (*L. intestinalis* and *L. vaginalis*).

The abundance of the five commonly encountered species (*L. crispatus*, *L. gasseri*, *L. iners*, *L. jensenii* and *L. acidophilus*) were then compared between the qPCR and the 454 pyrosequencing (Figure 2). *L. crispatus* dominated the *Lactobacillus* community in most samples. All samples displayed similar abundance profiles with enriched lower abundance species including *L. jensenii*, *L. iners* and *L.*

acidophilus exposed by qPCR (Table 3, Figure 2). The V5-V8 region of the 16S rRNA gene was not as effective in distinguishing *L. acidophilus* from *L. gasseri* using the primers published by Ma *et al.* (2013) due to high sequence homology between these two species in the V5-V8 variable region.

***Lactobacillus* phylogeny**

The phylogenetic tree constructed using the full-length 16S rRNA gene sequences of the *Lactobacillus* sp. described in this study indicated that *L. acidophilus* and *L. crispatus* appear more closely related than *L. acidophilus* and *L. gasseri*. When comparing only the V5-V8 region, however, *L. gasseri* clusters closer to *L. acidophilus* (Figure 4). The heat map confirmed that *L. crispatus* and *L. iners* dominated the microbial communities in samples analyzed in this study and samples were more likely to cluster based on *Lactobacillus* community dominance, than the patient history (dysmenorrhea or menorrhagia), the anatomical site of collection (endometrium or cervix), or the analysis technique (qPCR or pyrosequencing).

DISCUSSION

Sequencing of the V5-V8 region of the 16S rRNA gene improves the discriminatory power for speciation of dominant genital tract lactobacilli. This study examined the different bacterial communities within the upper genital tract of women, reporting changes in the bacterial community composition of lactobacilli. Consistent with previous studies, *L. crispatus* and *L. iners* were the most abundant lactobacilli in the samples tested in this study.

Sequencing technologies frequently often only report the presence of lactobacilli at genus-level. Studies exploiting some regions of the 16S rRNA gene fail to discriminate lactobacilli beyond higher order taxonomic classification due to limited sequence variation. Therefore, some studies have reported that lactobacilli as a genera are: positively correlated with healthy pregnancy outcomes including successful implantation and delivery at term; and form abundant community members in cases of adverse pregnancy outcomes including recurrent implantation failure and preterm birth (Franasiak *et al.*, 2016; Moreno *et al.*, 2016; Onderdonk *et al.*, 2008; Tao *et al.*, 2017). Our research

design enabled us to overcome the shortfalls commonly associated with genus-level identification. Similar results can be observed in molecular studies characterising the female genital tract when multiple variable regions of the 16S rRNA gene were sequenced (Fettweis et al., 2012; Graspeuntner et al., 2018; Madhivanan et al., 2014; Miles et al., 2017). Van Der Pol *et al.* (Van Der Pol et al., 2019) reported that the choice of 16S rRNA reference sequence database and sample sequence clustering parameters are equally as important as the choice of variable region for amplification characterising microbial community members to lower orders.

There is no doubt that sequencing the conserved 16S rRNA gene has improved our understanding of extant biodiversity in human microbial communities and is critical for understanding the impact of low-abundance community members on health and disease. However, there is no consensus best practice for microbiome studies, and significant variability exists between sample collection and storage methods, DNA extraction, universal primer selection, and sequencing platform and data analysis software (Pollock et al., 2018). Characterization of microbial communities using the 16S rRNA gene have been hampered by inherent differences generated in community profiles when sequencing different hypervariable regions, short read lengths, and taxonomic classification difficulties due to limited resolution for closely related species (Poretsky et al., 2014). Sequencing technologies have been used to interrogate the genital tract microbial community in reproductive-aged women but most fail to resolve the isolates to species-level. Consequently, more recent efforts have focused on sequencing multiple variable regions of the gene with amalgamation of all data into a single profile (Fuks et al., 2018). Very current research focuses on removing bias associated with sequencing component variable regions by using full-length gene sequencing (Callahan et al., 2016). The need to characterise the full-length 16S rRNA gene is further required as exhibited by the change in *L. gasseri* clustering when comparing the full-length gene to the V5-V8 region. Within this study, these species were not able to be distinguished from each other using this region alone.

The significance of our research is highlighted by studies confirming that *L. iners* does not protect against preterm birth and is frequently reported as an abundant community member in women with bacterial vaginosis (Madhivanan et al., 2014; Petricevic et al., 2014). Further, significant differences between lactobacilli in term compared to preterm deliveries have not been reported for all studies (Amabebe and Anumba, 2018; Romero et al., 2014). Within our study we are able to identify the bacteria to a species level using pyrosequencing reads with relative confidence. One limitation of this study is the relatively small sample size.

Collectively our research confirms what other studies have shown, that health and disease may depend on species and strain-level differences for prominent community members at a given anatomical niche (Kraal et al., 2014). It is clear that additional discriminatory power is required to resolve lower order classifications using current sequencing methods. This current study confirms that speciation of key genital tract *Lactobacillus* sp., capable of modulating reproductive health is possible when the appropriate region of the 16S rRNA gene is interrogated.

CONCLUSION

Studies characterizing microbial communities in the female genital tract report inconsistent results when assessing dysbiosis as a cause of reproductive pathology. Our work provides evidence for the impact of primer selection on evaluating the biological significance of shifts in community taxa. Careful experimental design should include a comparative analysis of microbial community profiling data generated by interrogation of multiple variable regions to the 16S rRNA gene to ensure that species abundance and diversity are accurately reflected.

List of abbreviations

ATCC: American type culture collection; DGC: dysmenorrhea progestin effect endocervix; DGE: dysmenorrhea progestin effect endometrium; DNA: deoxyribonucleic acid; DPC: dysmenorrhea proliferative endocervix; DPE: dysmenorrhea proliferative endometrium; DSC: dysmenorrhea

secretory endocervix; DSE: dysmenorrhea secretory endometrium; HRM: high resolution melt; MGC: menorrhagia progestin effect endocervix; MGE: menorrhagia progestin effect endometrium; MPC: menorrhagia proliferative endocervix; MPE: menorrhagia proliferative endometrium; MSC: menorrhagia secretory endocervix; MSE: menorrhagia secretory endometrium; OTU: operational taxonomic unit; rRNA ribosomal ribonucleic acid; VIC: virgo intacta

ACKNOWLEDGEMENTS

The authors wish to thank Wesley Hospital theatre staff who facilitated collection of the genital tract samples. We wish to acknowledge the Australian Centre for Ecogenomics for 454 pyrosequencing and Professor Philip Hugenholtz and Dr Fiona May. This work was performed in the John and Wendy Thorsen Women's Health Laboratory.

CONTRIBUTION TO AUTHORSHIP

JOC: designed and completed bioinformatics analyses, contributed to the analysis and interpretation of the data, and contributed to the writing of the manuscript.

DW: designed and completed bioinformatics analyses, contributed to the analysis and interpretation of the data, and contributed to the writing of the manuscript.

MB: conceived and designed the project, performed collection of clinical specimens and contributed to the writing of the manuscript.

FH: designed and completed the qPCR experiments, contributed to the analysis and interpretation of the qPCR data and contributed to the writing of the manuscript.

EP: conceived and designed the project, completed tissue processing, DNA extraction and 16S PCR experiments, contributed to the analysis and interpretation of the data, and drafted significant parts of the work.

277

278

279 REFERENCES

- 280 Amabebe, E., Anumba, D.O.C., 2018. The Vaginal Microenvironment: The Physiologic Role of
281 Lactobacilli. *Frontiers in medicine* 5, 181.
- 282 Callahan, B.J., Sankaran, K., Fukuyama, J.A., McMurdie, P.J., Holmes, S.P., 2016. Bioconductor
283 Workflow for Microbiome Data Analysis: from raw reads to community analyses. *F1000Research* 5,
284 1492.
- 285 Fettweis, J.M., Serrano, M.G., Sheth, N.U., Mayer, C.M., Glascock, A.L., Brooks, J.P., Jefferson, K.K.,
286 Buck, G.A., 2012. Species-level classification of the vaginal microbiome. *BMC genomics* 13 Suppl 8,
287 S17.
- 288 Franasia, J.M., Werner, M.D., Juneau, C.R., Tao, X., Landis, J., Zhan, Y., Treff, N.R., Scott, R.T., 2016.
289 Endometrial microbiome at the time of embryo transfer: next-generation sequencing of the 16S
290 ribosomal subunit. *Journal of assisted reproduction and genetics* 33, 129-136.
- 291 Fuks, G., Elgart, M., Amir, A., Zeisel, A., Turnbaugh, P.J., Soen, Y., Shental, N., 2018. Combining 16S
292 rRNA gene variable regions enables high-resolution microbial community profiling. *Microbiome* 6,
293 17.
- 294 Grasseuntner, S., Loeper, N., Kunzel, S., Baines, J.F., Rupp, J., 2018. Selection of validated
295 hypervariable regions is crucial in 16S-based microbiota studies of the female genital tract. *Scientific*
296 *reports* 8, 9678.
- 297 J., O., FG., B., M., F., R., K., P., L., D., M., PR., M., RB., O.H., GL., S., P., S., MHH., S., E., E.S., H., W., 2018.
298 *vegan: Community Ecology Package* R package version 2.5-2.
- 299 Klebanoff, S.J., Hillier, S.L., Eschenbach, D.A., Waltersdorff, A.M., 1991. Control of the microbial
300 flora of the vagina by H₂O₂-generating lactobacilli. *The Journal of infectious diseases* 164, 94-100.
- 301 Kraal, L., Abubucker, S., Kota, K., Fischbach, M.A., Mitreva, M., 2014. The prevalence of species and
302 strains in the human microbiome: a resource for experimental efforts. *PloS one* 9, e97279.
- 303 Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version
304 7.0 for Bigger Datasets. *Mol Biol Evol* 33, 1870-1874.
- 305 Liu, B., Gibbons, T., Ghodsi, M., Treangen, T., Pop, M., 2011. Accurate and fast estimation of
306 taxonomic profiles from metagenomic shotgun sequences. *BMC genomics* 12 Suppl 2, S4.
- 307 Ma, L., Lv, Z., Su, J., Wang, J., Yan, D., Wei, J., Pei, S., 2013. Consistent condom use increases the
308 colonization of *Lactobacillus crispatus* in the vagina. *PloS one* 8, e70716-e70716.
- 309 Madhivanan, P., Raphael, E., Rumphs, A., Krupp, K., Ravi, K., Srinivas, V., Arun, A., Reingold, A.L.,
310 Klausner, J.D., Riley, L.W., 2014. Characterization of culturable vaginal *Lactobacillus* species among
311 women with and without bacterial vaginosis from the United States and India: a cross-sectional
312 study. *Journal of medical microbiology* 63, 931-935.
- 313 McDonald, D., Price, M.N., Goodrich, J., Nawrocki, E.P., DeSantis, T.Z., Probst, A., Andersen, G.L.,
314 Knight, R., Hugenholtz, P., 2012. An improved Greengenes taxonomy with explicit ranks for
315 ecological and evolutionary analyses of bacteria and archaea. *The ISME journal* 6, 610-618.
- 316 Miles, S.M., Hardy, B.L., Merrell, D.S., 2017. Investigation of the microbiota of the reproductive tract
317 in women undergoing a total hysterectomy and bilateral salpingo-oophorectomy. *Fertility and*
318 *sterility* 107, 813-820.e811.
- 319 Moreno, I., Codoner, F.M., Vilella, F., Valbuena, D., Martinez-Blanch, J.F., Jimenez-Almazan, J.,
320 Alonso, R., Alama, P., Remohi, J., Pellicer, A., Ramon, D., Simon, C., 2016. Evidence that the
321 endometrial microbiota has an effect on implantation success or failure. *American journal of*
322 *obstetrics and gynecology* 215, 684-703.

Onderdonk, A.B., Hecht, J.L., McElrath, T.F., Delaney, M.L., Allred, E.N., Leviton, A., 2008. Colonization of second-trimester placenta parenchyma. *American journal of obstetrics and gynecology* 199, 52.e51-52.e10.

Pelzer, E.S., Willner, D., Buttini, M., Huygens, F., 2018. A role for the endometrial microbiome in dysfunctional menstrual bleeding. *Antonie van Leeuwenhoek* 111, 933-943.

Petricevic, L., Domig, K.J., Nierscher, F.J., Sandhofer, M.J., Fidesser, M., Krondorfer, I., Husslein, P., Kneifel, W., Kiss, H., 2014. Characterisation of the vaginal *Lactobacillus* microbiota associated with preterm delivery. *Scientific reports* 4, 5136.

Pollock, J., Glendinning, L., Wisedchanwet, T., Watson, M., 2018. The Madness of Microbiome: Attempting To Find Consensus "Best Practice" for 16S Microbiome Studies. *Applied and environmental microbiology* 84.

Poretsky, R., Rodriguez, R.L., Luo, C., Tsementzi, D., Konstantinidis, K.T., 2014. Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. *PloS one* 9, e93827.

Romero, R., Hassan, S.S., Gajer, P., Tarca, A.L., Fadrosh, D.W., Bieda, J., Chaemsathong, P., Miranda, J., Chaiworapongsa, T., Ravel, J., 2014. The vaginal microbiota of pregnant women who subsequently have spontaneous preterm labor and delivery and those with a normal delivery at term. *Microbiome* 2, 18.

Tao, X., Franasiak, J.M., Zhan, Y., Scott, R.T., Rajchel, J., Bedard, J., Newby, R., Scott, R.T., Treff, N.R., Chu, T., 2017. Characterizing the endometrial microbiome by analyzing the ultra-low bacteria from embryo transfer catheter tips in IVF cycles: Next generation sequencing (NGS) analysis of the 16S ribosomal gene. *Human Microbiome Journal* 3, 15-21.

Team, R.C., 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research* 22, 4673-4680.

Van Der Pol, W.J., Kumar, R., Morrow, C.D., Blanchard, E.E., Taylor, C.M., Martin, D.H., Lefkowitz, E.J., Muzny, C.A., 2019. In Silico and Experimental Evaluation of Primer Sets for Species-Level Resolution of the Vaginal Microbiota Using 16S Ribosomal RNA Gene Sequencing. *The Journal of infectious diseases* 219, 305-314.

Warnes, G., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T., Mächler, M., Magnusson, A., Möller, S., 2005. gplots: Various R programming tools for plotting data.

Willner, D., Low, S., Steen, J.A., George, N., Nimmo, G.R., Schembri, M.A., Hugenholtz, P., 2014. Single clinical isolates from acute uncomplicated urinary tract infections are representative of dominant in situ populations. *mBio* 5, e01064-01013.

TABLE/FIGURE CAPTION LIST

Table 1. Sample abbreviations

Table 2. 16S rRNA *Lactobacillus*-species specific primers (Ma et al., 2013)

Table 3: Unique to species SNPs from the 16S rRNA gene variable regions.

Figure 1a: Annealing site of 454 pyrosequencing 16S rRNA primers to V5-V8 region in the *Lactobacillus* species.

Figure 1b: Annealing site of species specific primers (Table 2) in the *Lactobacillus* species which underwent qPCR.

Figure 2: Comparison of relative microbial abundance all sites and phases between 454 pyrosequencing of the V5 - V8 region and species specific *Lactobacillus* qPCR.

Figure 3: Hierarchical clustering/distance ordination quantifying similarities between qPCR and PYRO pairs

Lactobacillus relative abundance distribution and hierarchical clustering in pyrosequenced and qPCR analysed samples. The heatmap shows relative abundances of detected *Lactobacillus* species in each sample, with columns organised by relative positions in the dendrogram. Some subsets of paired pyrosequenced and qPCR samples have been coloured to highlight clustering.

Figure 4a: Full length and region specific 16S rRNA gene phylogeny for key genital tract lactobacilli

ML tree derived from full-length and trimmed (V5-V8 region) 16S rRNA sequences of the multiple species of *Lactobacillus* identified by pyrosequencing and 16S qPCR. Accession numbers for reference sequences are provided in the methods. *Pediococcus pentasaecus* and *Bacillus subtilis* were included as outgroups. Branch support values are based on 1000 bootstrap repetitions.

Table 1. Sample abbreviations

Sample Abbreviations	Full Sample Name
DGC	dysmenorrhea progestin effect endocervix
DSC	dysmenorrhea secretory endometrium
MGC	menorrhagia progestin effect endocervix
DPC	dysmenorrhea proliferative endocervix
DGE	dysmenorrhea progestin effect endometrium;
DPE	dysmenorrhea proliferative endometrium
MGE	menorrhagia progestin effect endometrium
MSE	menorrhagia secretory endometrium
MSC	menorrhagia secretory endocervix
MPE	menorrhagia proliferative endometrium
MPC	menorrhagia proliferative endocervix
VIC	virgo intacta

Table 2. 16S rRNA *Lactobacillus*-species specific primers (Ma et al., 2013)

<i>Lactobacillus</i> species	Primer name	Primer sequence
<i>L. acidophilus</i>	LAA-1	5'CATCCAGTGCAAACCTAAGAG3'
	LAA-2	5'GATCCGCTTGCCTTCGCA3'
<i>L. crispatus</i>	LcrisF	5'AGCGAGCGGAACTAACAGATTTAC3'
	LcrisR	5'AGCTGATCATGCGATCTGCTT3'
<i>L. gasseri</i>	LgassF	5'AGCGAGCTTGCCTAGATGAATTTG3'
	LgassR	5'TCTTTTAACTCTAGACATGCGTC3'
<i>L. jensenii</i>	LjensF	5'AAGTCGAGCGAGCTTGCCTATAGA3'
	LjensR	5'CTTCTTTCATGCGAAAGTAGC3'
<i>L. iners</i>	InersF	5'GTCTGCCTTGAAGATCGG3'
	InersR	5'ACAGTTGATAGGCATCATC3'

Table 3. Unique to species SNPs from the 16S rRNA gene variable regions.

Species	16S rRNA gene variable Region								
	V1	V2	V3	V4	V5	V6	V7	V8	V9
<i>L. iners</i>	5	17	0	2	0	0	1	2	4
<i>L. gasseri</i>	4	0	2	0	0	0	1	0	1
<i>L. acidophilus</i>	2	4	0	3	2	0	0	2	0
<i>L. crispatus</i>	2	1	0	1	3	2	0	0	0
<i>L. jensenii</i>	6	6	0	1	1	5	1	0	5

Figure 2: Comparison of relative microbial abundance at all sites and phases between pyrosequencing of the V5 - V8 region and species specific *Lactobacillus* qPCR.

Figure 3: Hierarchical clustering/distance ordination quantifying similarities between qPCR and PYRO pairs

A

B

Figure 4: 16S rRNA gene phylogeny for key genital tract lactobacilli

(A) Full length

(B) V5-V8 region

495 **CONFLICT OF INTEREST STATEMENT**

496 The authors declare that there are no conflicts of interest.

497 **FUNDING STATEMENT**

498 Funding for this project was awarded by the Wesley Research Institute (Grant number 2011-12). The
499 funding body played no role in conducting the research or preparing the manuscript.

500

Percent of Community