bioRxiv preprint doi: https://doi.org/10.1101/630004; this version posted May 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



1 **TITLE**

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

4

5 **AUTHORS**

6 **Jessica L. O'Callaghan [a, b], Dana Willner [c, d], Melissa Buttini [e], Flavia Huygens [a,b], Elise S. Pelzer* [a, b]**

7

8 **AUTHOR AFFILIATIONS**

9 [a] School of Biomedical Sciences, Faculty of Health, Queensland University of Technology, Brisbane,

10 Queensland, Australia 4001

11 [b] Institute of Health and Biomedical Innovation, Faculty of Health, Queensland University of

12 Technology, Brisbane, Queensland, Australia 4001

13 [c] Australian Centre for Ecogenomics, University of Queensland, St Lucia, Queensland, Australia

14 [d] Department of Computer Science, College of William and Mary, Williamsburg, VA 23187, USA

15 [e] The Wesley Hospital, Auchenflower, Queensland, Australia, 4066

16 * *corresponding author*

17 Elise S. Pelzer
18 Queensland University of Technology
19 GPO Box 2434
20 Brisbane, QLD 4001
21 + 61 3138 0542
22 e.pelzer@qut.edu.au
23

24 **<u>Running title</u>**

25 Speciation of genital tract lactobacilli

26

27

28

bioRxiv preprint doi: https://doi.org/10.1101/630004; this version posted May 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



29

30

31 **ABSTRACT**

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

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

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

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

49 **KEYWORDS**

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

51

52 **IMPORTANCE**

bioRxiv preprint doi: https://doi.org/10.1101/630004; this version posted May 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



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

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

61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78

bioRxiv preprint doi: https://doi.org/10.1101/630004; this version posted May 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



79
80
81
82
83 **INTRODUCTION**

84 Molecular microbiology techniques have changed our ability to identify microbial communities,
85 revolutionizing the way we assess female genital tract microbiomes. In cultivation-dependent studies,
86 greater than 95% of the vaginal microbiota in healthy women was classified as lactobacilli. The advent
87 of cultivation-independent technology platforms has provided evidence to suggest that in up to two-
88 thirds of healthy women, the lactobacilli were co-aggregated with a diverse group of microbial
89 community members, and in some cases did not dominate (Fettweis et al., 2012; Klebanoff et al.,
90 1991). Lactobacilli establish niche dominance through co-aggregation, competitive inhibition,
91 production of metabolic acids, hydrogen peroxide, and antimicrobial components including
92 bacteriocins (Amabebe and Anumba, 2018). The discovery that lactobacilli do not dominate the genital
93 tract of all healthy women suggests that: there is redundancy in function and protection based on
94 community membership; and all lactobacilli may not provide the same level of protection in the genital
95 tract environment. This discovery casts doubt over the long-held view that a healthy female genital
96 tract is characterized by a *Lactobacillus* sp.-dominant microbiota. The ability to confidently assign
97 lower order taxonomic classification to lactobacilli is critical in advancing our understanding of the
98 protective role played by the various species within this genus in reproductive health. The objective
99 of this study was to examine the discriminatory power of current molecular microbiology techniques
100 for identification of genital tract lactobacilli.

101
102
103
104

bioRxiv preprint doi: https://doi.org/10.1101/630004; this version posted May 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



105

106

107

108

109 **METHODS**

110 **Patient cohort, sample collection and genomic DNA preparation**

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

114 **Details of ethical approval**

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

118 **Next-generation sequencing**

119 The 16S rRNA PCR assay was performed using the previously published primers, 803F (5′-TTA GAT ACC
120 CTG GTA GTC -3′) and 1392R (5′-ACG GGC GGT GTG TRC -3′) and PCR cycling conditions (Willner et al.,
121 2014). Fusion primers with 454 adaptor sequences were ligated to the 803F and 1392R primers to
122 amplify the V5 and V8 regions of the 16S rRNA gene (Willner et al., 2014). PCR reactions were
123 performed as previously described (Pelzer et al., 2018). The five frequently encountered genital tract
124 *Lactobacillus* sp. were aligned using the SILVA database to determine the degree of variation within
125 the V5-V8 regions of the 16S rRNA gene. The annealing site of the sequencing primers is marked on
126 the alignment (Figure 1a).

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

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

bioRxiv preprint doi: https://doi.org/10.1101/630004; this version posted May 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



130 annealing was confirmed using species-specific alignment of the five *Lactobacillus* sp. interrogated in
131 this study (Figure 1b).
132

133 **Taxonomic classification**

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

139 ***Lactobacillus* Phylogenetic Trees**

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

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

bioRxiv preprint doi: https://doi.org/10.1101/630004; this version posted May 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



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

157 **Hierarchical clustering**

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

164 **RESULTS**

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

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

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/630004; this version posted May 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



179 *acidophilus* exposed by qPCR (Table 3, Figure 2). The V5-V8 region of the 16S rRNA gene was not as
180 effective in distinguishing *L. acidophilus* from *L. gasseri* using the primers published by Ma *et al*.
181 (2013) due to high sequence homology between these two species in the V5-V8 variable region.

182 ***Lactobacillus* phylogeny**

183 The phylogenetic tree constructed using the full-length16S rRNA gene sequences of the *Lactobacillus*
184 sp. described in this study indicated that *L. acidophilus* and L. *crispatus* appear more closely related
185 than *L. acidophilus* and *L. gasseri*. When comparing only the V5-V8 region, however, *L. gasseri*
186 clusters closer to *L. acidophilus* (Figure 4). The heat map confirmed that *L. crispatus* and *L. iners*
187 dominated the microbial communities in samples analyzed in this study and samples were more
188 likely to cluster based on *Lactobacillus* community dominance, than the patient history
189 (dysmenorrhea or menorrhagia), the anatomical site of collection (endometrium or cervix), or the
190 analysis technique (qPCR or pyrosequencing).

191 **DISCUSSION**

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

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

bioRxiv preprint doi: https://doi.org/10.1101/630004; this version posted May 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



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

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

bioRxiv preprint doi: https://doi.org/10.1101/630004; this version posted May 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



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

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

241 **CONCLUSION**

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

248 **List of abbreviations**

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

bioRxiv preprint doi: https://doi.org/10.1101/630004; this version posted May 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



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

257

258

259 **ACKNOWLEDGEMENTS**

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

264 **CONTRIBUTION TO AUTHORSHIP**

265 JOC: designed and completed bioinformatics analyses, contributed to the analysis and interpretation
266 of the data, and contributed to the writing of the manuscript.
267 DW: designed and completed bioinformatics analyses, contributed to the analysis and interpretation
268 of the data, and contributed to the writing of the manuscript.
269 MB: conceived and designed the project, performed collection of clinical specimens and contributed
270 to the writing of the manuscript.
271 FH: designed and completed the qPCR experiments, contributed to the analysis and interpretation of
272 the qPCR data and contributed to the writing of the manuscript.
273 EP: conceived and designed the project, completed tissue processing, DNA extraction and 16S PCR
274 experiments, contributed to the analysis and interpretation of the data, and drafted significant parts
275 of the work.

276

bioRxiv preprint doi: https://doi.org/10.1101/630004; this version posted May 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



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 Franasiak, 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 Graspeuntner, 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 PackageR package version 2.5-2.
299 Klebanoff, S.J., Hillier, S.L., Eschenbach, D.A., Waltersdorph, 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-oopherectomy. 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.

bioRxiv preprint doi: https://doi.org/10.1101/630004; this version posted May 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



323 Onderdonk, A.B., Hecht, J.L., McElrath, T.F., Delaney, M.L., Allred, E.N., Leviton, A., 2008.
324 Colonization of second-trimester placenta parenchyma. American journal of obstetrics and
325 gynecology 199, 52.e51-52.e10.
326 Pelzer, E.S., Willner, D., Buttini, M., Huygens, F., 2018. A role for the endometrial microbiome in
327 dysfunctional menstrual bleeding. Antonie van Leeuwenhoek 111, 933-943.
328 Petricevic, L., Domig, K.J., Nierscher, F.J., Sandhofer, M.J., Fidesser, M., Krondorfer, I., Husslein, P.,
329 Kneifel, W., Kiss, H., 2014. Characterisation of the vaginal Lactobacillus microbiota associated with
330 preterm delivery. Scientific reports 4, 5136.
331 Pollock, J., Glendinning, L., Wisedchanwet, T., Watson, M., 2018. The Madness of Microbiome:
332 Attempting To Find Consensus "Best Practice" for 16S Microbiome Studies. Applied and
333 environmental microbiology 84.
334 Poretsky, R., Rodriguez, R.L., Luo, C., Tsementzi, D., Konstantinidis, K.T., 2014. Strengths and
335 limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community
336 dynamics. PloS one 9, e93827.
337 Romero, R., Hassan, S.S., Gajer, P., Tarca, A.L., Fadrosh, D.W., Bieda, J., Chaemsaithong, P., Miranda,
338 J., Chaiworapongsa, T., Ravel, J., 2014. The vaginal microbiota of pregnant women who subsequently
339 have spontaneous preterm labor and delivery and those with a normal delivery at term. Microbiome
340 2, 18.
341 Tao, X., Franasiak, J.M., Zhan, Y., Scott, R.T., Rajchel, J., Bedard, J., Newby, R., Scott, R.T., Treff, N.R.,
342 Chu, T., 2017. Characterizing the endometrial microbiome by analyzing the ultra-low bacteria from
343 embryo transfer catheter tips in IVF cycles: Next generation sequencing (NGS) analysis of the 16S
344 ribosomal gene. Human Microbiome Journal 3, 15-21.
345 Team, R.C., 2013. R: A language and environment for statistical computing. R Foundation for
346 Statistical Computing.
347 Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of
348 progressive multiple sequence alignment through sequence weighting, position-specific gap
349 penalties and weight matrix choice. Nucleic acids research 22, 4673-4680.
350 Van Der Pol, W.J., Kumar, R., Morrow, C.D., Blanchard, E.E., Taylor, C.M., Martin, D.H., Lefkowitz,
351 E.J., Muzny, C.A., 2019. In Silico and Experimental Evaluation of Primer Sets for Species-Level
352 Resolution of the Vaginal Microbiota Using 16S Ribosomal RNA Gene Sequencing. The Journal of
353 infectious diseases 219, 305-314.
354 Warnes, G., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T., Mächler, M.,
355 Magnusson, A., Möller, S., 2005. gplots: Various R programming tools for plotting data.
356 Willner, D., Low, S., Steen, J.A., George, N., Nimmo, G.R., Schembri, M.A., Hugenholtz, P., 2014.
357 Single clinical isolates from acute uncomplicated urinary tract infections are representative of
358 dominant in situ populations. mBio 5, e01064-01013.

359

360

361

362

363

364

365

366

bioRxiv preprint doi: https://doi.org/10.1101/630004; this version posted May 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



367

368

369

370 **TABLE/FIGURE CAPTION LIST**

371 **Table 1.** Sample abbreviations

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

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

374

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

377

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

380

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

383

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

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/630004; this version posted May 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



395

396

397

398 **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 |

399

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

| *Lactobacillus* species | Primer name | Primer sequence |
|---|---|---|
| *L. acidophilus* | LAA-1 | 5'CATCCAGTGCAAACCTAAGAG3' |
| | LAA-2 | 5'GATCCGCTTGCCTTCGCA3' |
| *L. crispatus* | LcrisF | 5'AGCGAGCGGAACTAACAGATTTAC3' |
| | LcrisR | 5'AGCTGATCATGCGATCTGCTT3' |
| *L. gasseri* | LgassF | 5'AGCGAGCTTGCCTAGATGAATTTG3' |
| | LgassR | 5'TCTTTTAAACTCTAGACATGCGTC3' |
| *L. jensenii* | LjensF | 5'AAGTCGAGCGAGCTTGCCTATAGA3' |
| | LjensR | 5'CTTCTTTCATGCGAAAGTAGC3' |
| *L. iners* | InersF | 5'GTCTGCCTTGAAGATCGG3' |
| | InersR | 5'ACAGTTGATAGGCATCATC3' |

401

402

403

404

405

406

bioRxiv preprint doi: https://doi.org/10.1101/630004; this version posted May 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



407

408

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

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

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

bioRxiv preprint doi: https://doi.org/10.1101/630004; this version posted May 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



426

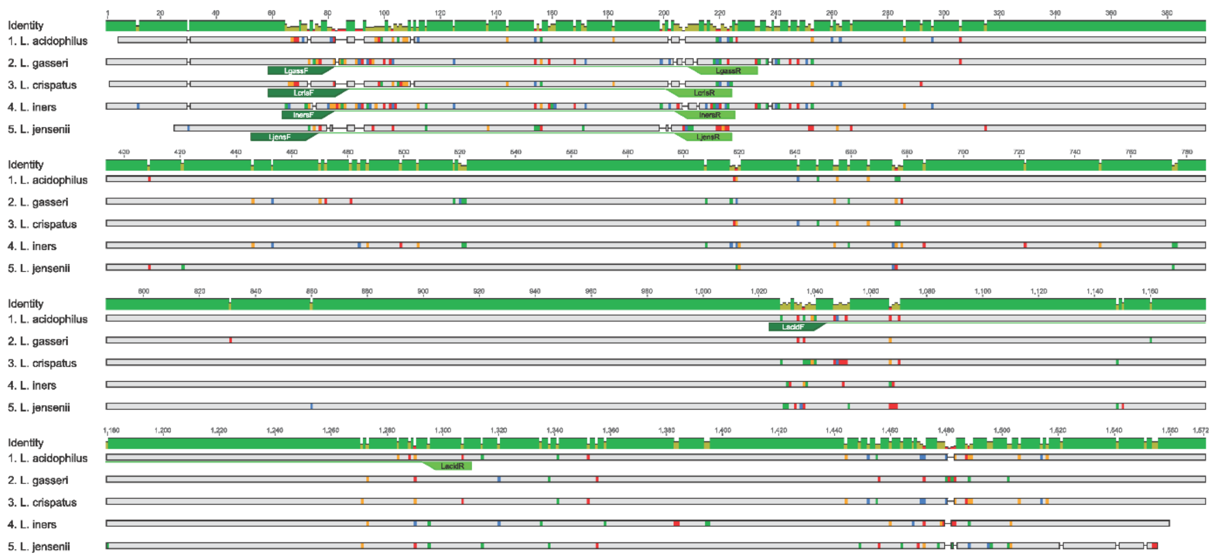
427

428

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

430 *Lactobacillus* species.

431




432

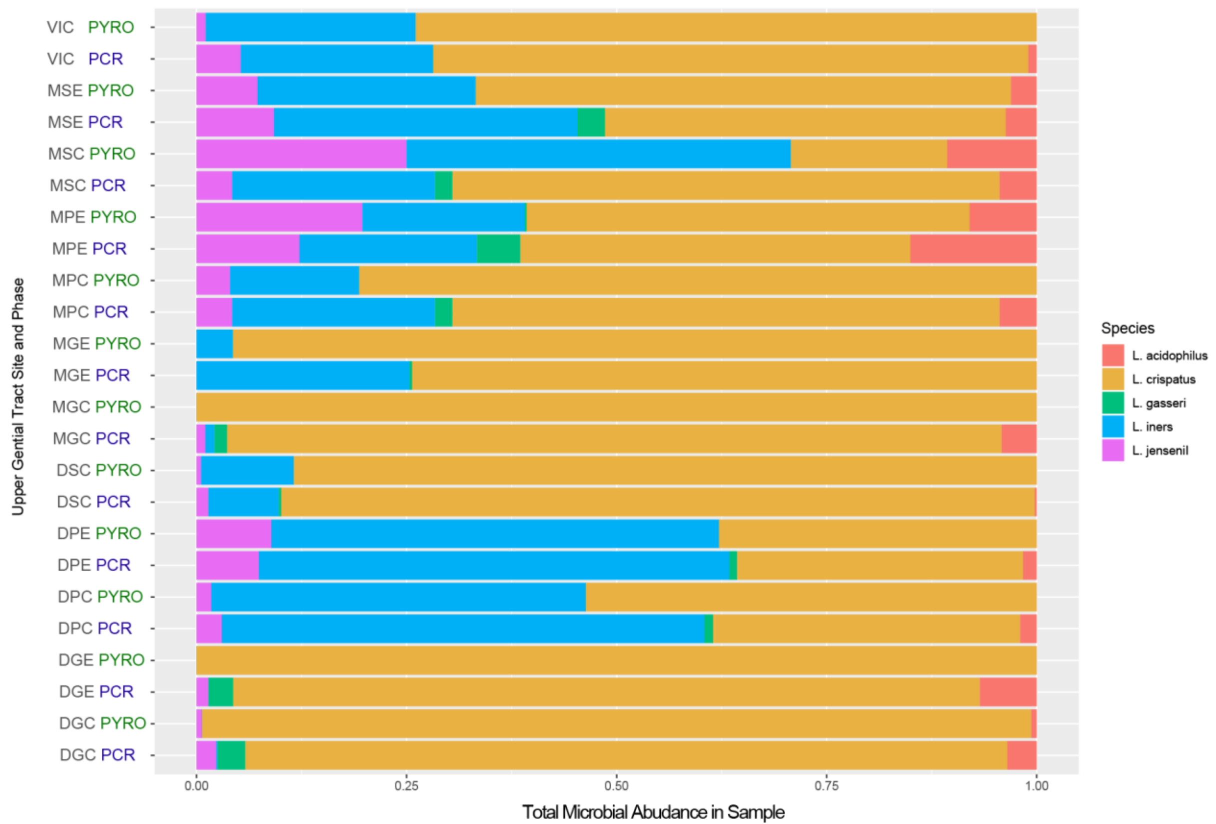
433 **Figure 1b :** Annealing site of species specific primers (Table 2) for the *Lactobacillus* species which

434 underwent qPCR.

435

bioRxiv preprint doi: https://doi.org/10.1101/630004; this version posted May 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

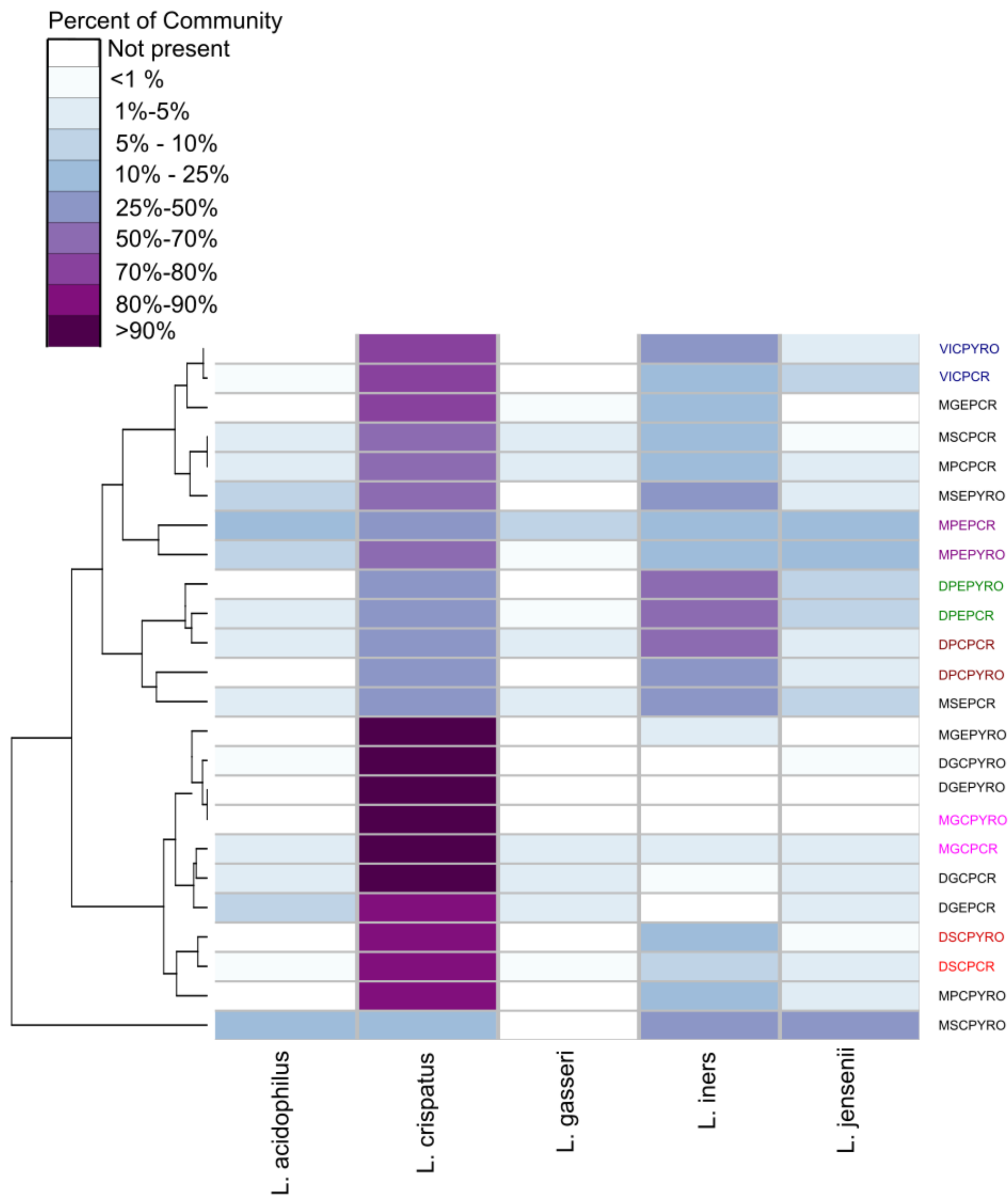





436
437
438 **Figure 2**: Comparison of relative microbial abundance at all sites and phases between 454
439 pyrosequencing of the V5 - V8 region and species specific *Lactobacillus* qPCR.
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455

bioRxiv preprint doi: https://doi.org/10.1101/630004; this version posted May 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.





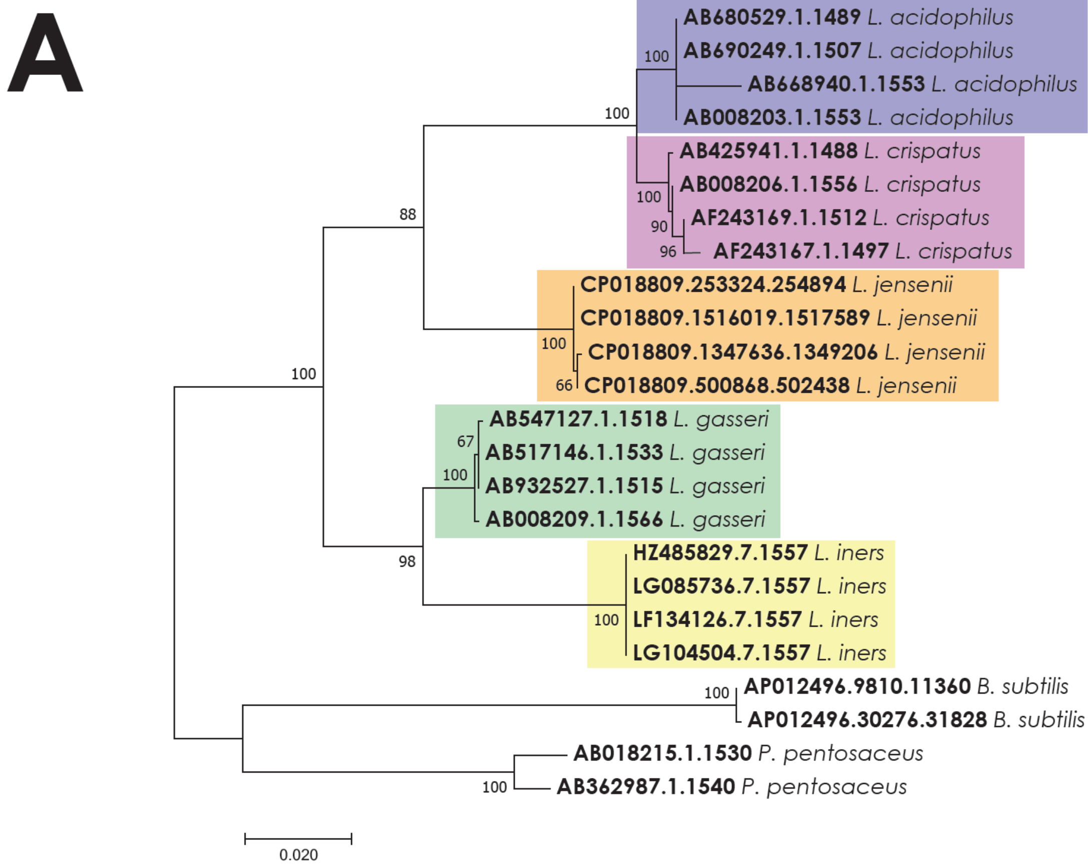

456

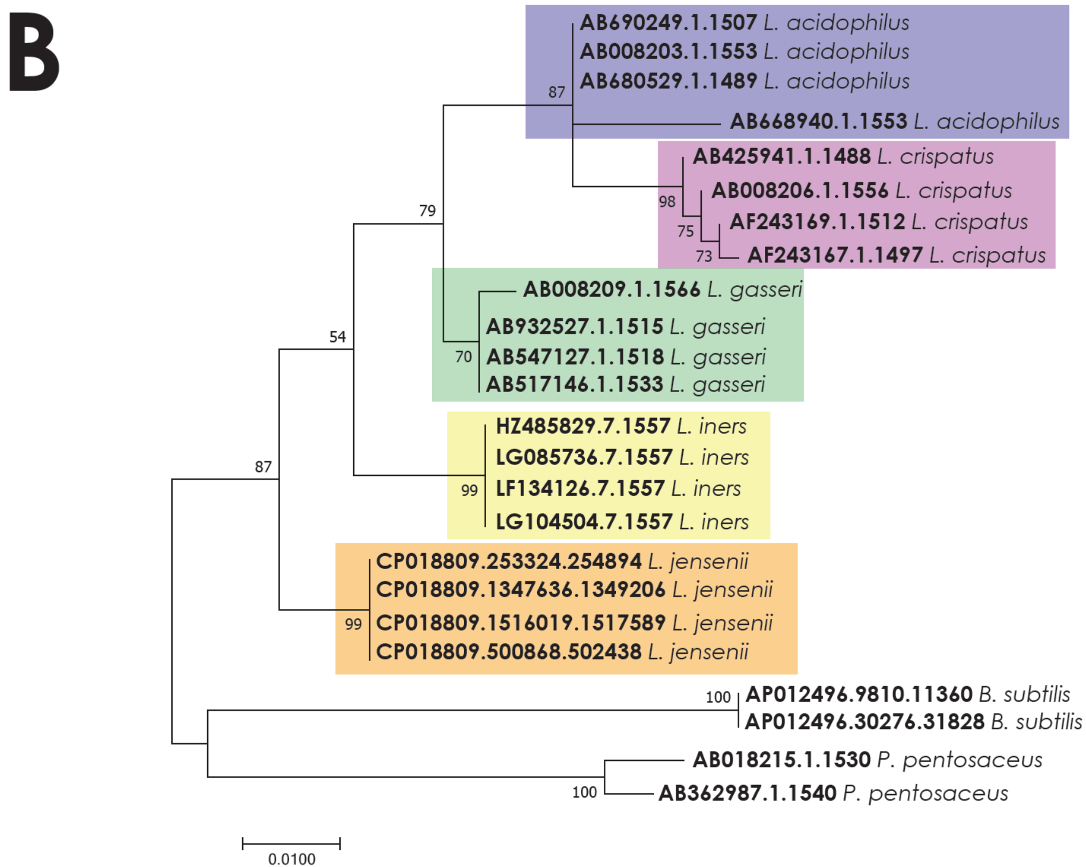
457 **Figure 3: Hierarchical clustering/distance ordination quantifying similarities between qPCR and**
458 **PYRO pairs**

459

bioRxiv preprint doi: https://doi.org/10.1101/630004; this version posted May 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.






460

bioRxiv preprint doi: https://doi.org/10.1101/630004; this version posted May 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



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

462 **(A) Full length**

463 **(B) V5-V8 region**

464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494

bioRxiv preprint doi: https://doi.org/10.1101/630004; this version posted May 7, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



495 **CONFLICT OF INTEREST STATEMENT**

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

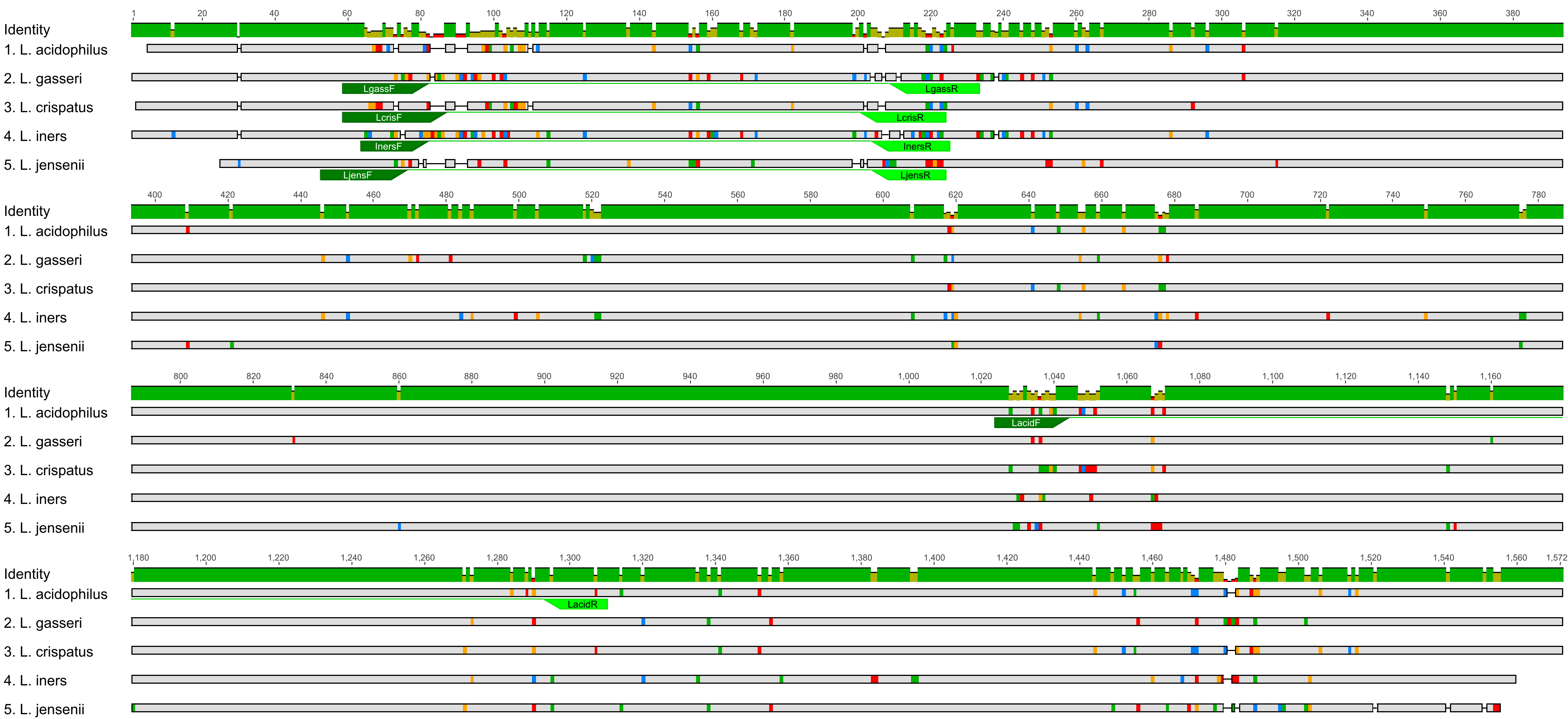
497 **FUNDING STATEMENT**

498 Funding for this project was awarded by the Wesley Research Institute (Grant number 2011-12). The

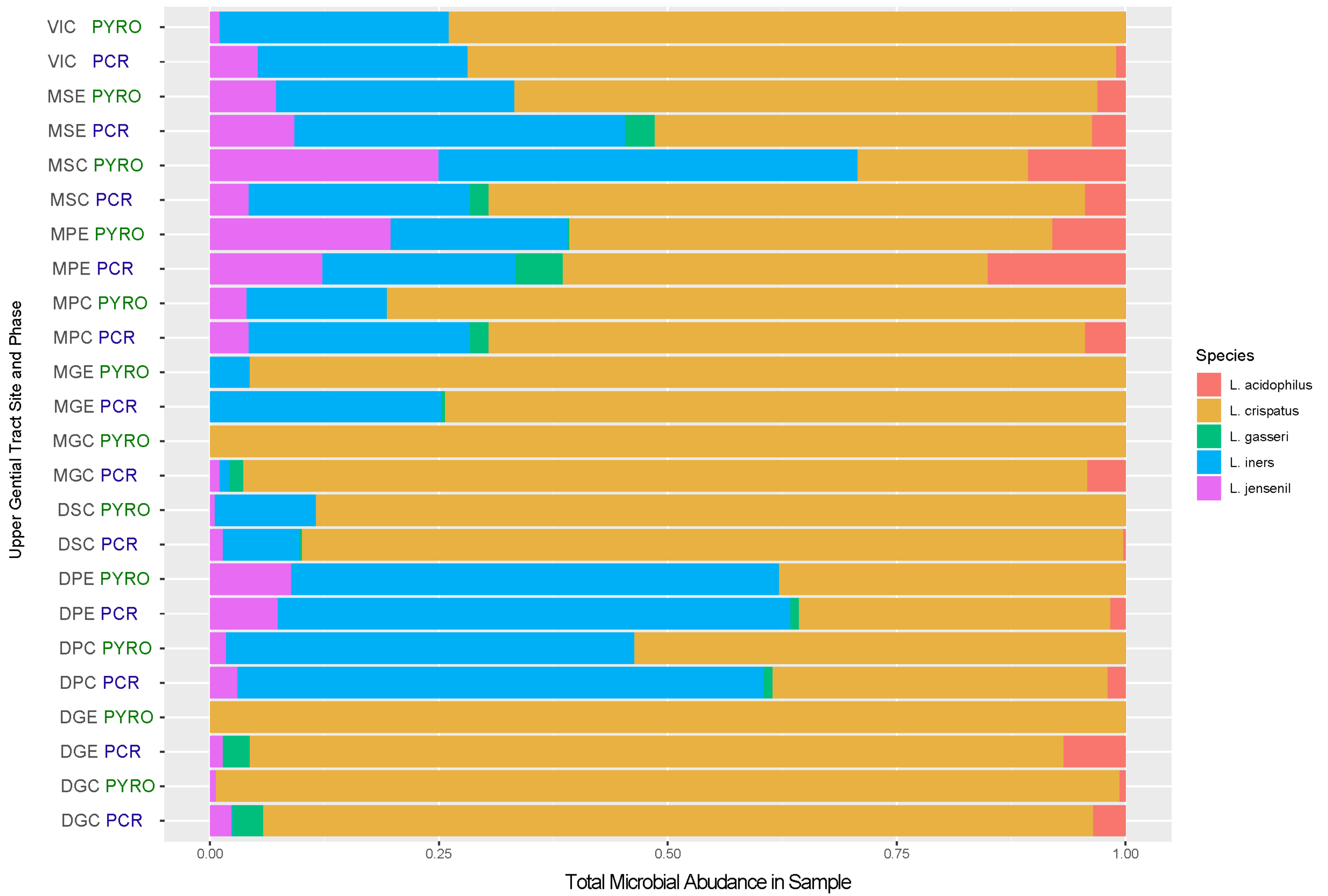
499 funding body played no role in conducting the research or preparing the manuscript.

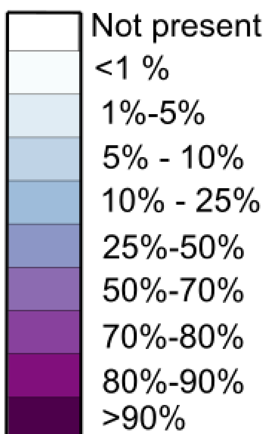
500

```
Consensus        GGATTAGATACCCTGGTAGTCCATGCCGTAAACGATGAGTGCTAAGTGTTGGGAGGTTTCCGCCTCTCAGTGCTGCAGCTAACGCATTAAGCACTCCGCCTGGGGAGTACGACCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCG
                   PyroFORWARD
Identity
1. L. acidophilus GGATTAGATACCCTGGTAGTCCATGCCGTAAACGATGAGTGCTAAGTGTTGGGAGGTTTCCGCCTCTCAGTGCTGCAGCTAACGCATTAAGCACTCCGCCTGGGGAGTACGACCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCG
2. L. gasseri     GGATTAGATACCCTGGTAATCCATGCCGTAAACGATGAGTGCTAAGTGTTGGGAGGTTTCCGCCTCTCAGTGCTGCAGCTAACGCATTAAGCACTCCGCCTGGGGAGTACGACCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCG
3. L. crispatus   GGATTAGATACCCTGGTAGTCCATGCCGTAAACGATGAGTGCTAAGTGTTGGGAGGTTTCCGCCTCTCAGTGCTGCAGCTAACGCATTAAGCACTCCGCCTGGGGAGTACGACCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCG
4. L. iners       GGATTAGATACCCTGGTAGTCCATGCCGTAAACGATGAGTGCTAAGTGTTGGGAGGTTTCCGCCTCTCAGTGCTGCAGCTAACGCATTAAGCACTCCGCCTGGGGAGTACGACCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCG
5. L. jensenii    GGATTAGATACCCTGGTAGTCCATGCCGTAAACGATGAGTGCTAAGTCTTGGGAGGTTTCCGCCTCTCAGTGCTGCAGCTAACGCATTAAGCACTCCGCCTGGGGAGTACGACCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCG

Consensus        GTGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCCAGTGCCAWCCTAAGAGATTAGGTGTTCCCTTCGGGGACRCTGAGACAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCG
Identity
1. L. acidophilus GTGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCTAGTGCAATCCGTAGAGATACGGAGTTCCCTTCGGGGACACTAAGACAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCG
2. L. gasseri     GTGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCCAGTGCAAACCTAAGAGATTAGGTGTTCCCTTCGGGGACGCTGAGACAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCG
3. L. crispatus   GTGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCTAGTGCCATTTGTAGAGATACAAAGTTCCCTTCGGGGACGCTAAGACAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCG
4. L. iners       GTGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCCATAGCCAGTCTAAGAGATTAGATGTTCCCTTCGGGGACTATGAGACAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCG
5. L. jensenii    GTGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCCTTTGACCACCTAAGAGATTAGGTTTTCCCTTCGGGGACAAAGAGACAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCG

Consensus        CAACGAGCGCAACCCTTGTCATTAGTTGCCAGCATTAAGTTGGGCACTCTAATGAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAGTCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAATGGACAGTACAACGAGAAG
Identity
1. L. acidophilus CAACGAGCGCAACCCTTGTCATTAGTTGCCAGCATTAAGTTGGGCACTCTAATGAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAGTCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAATGGACAGTACAACGAGGAG
2. L. gasseri     CAACGAGCGCAACCCTTGTCATTAGTTGCCATCATTAAGTTGGGCACTCTAATGAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAGTCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAATGGACGGTACAACGAGAAG
3. L. crispatus   CAACGAGCGCAACCCTTGTTATTAGTTGCCAGCATTAAGTTGGGCACTCTAATGAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAGTCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAATGGGCAGTACAACGAGAAG
4. L. iners       CAACGAGCGCAACCCTTGTCATTAGTTGCCAGCATTAAGTTGGGCACTCTAATGAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAGTCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAATGGACGGTACAACGAGAAG
5. L. jensenii    CAACGAGCGCAACCCTTGTTAATAGTTGCCAGCATTAAGTTGGGCACTCTATTGAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAGTCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAATGGGCAGTACAACGAGAAG

Consensus        CGARCCTGCGAAGGCAAGCGGATCTCTGAAAGCTGTTCTCAGTTCGGACTGCAGGCTGCAACTCGCCTGCACGAAGCTGGAATCGCTAGTAATCGCGGATCAGCACGCCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCA
                                                                                                                                                         PyroREVERSE
Identity
1. L. acidophilus CAAGCCTGCGAAGGCAAGCGAATCTCTTAAAGCTGTTCTCAGTTCGGACTGCAGTCTGCAACTCGACTGCACGAAGCTGGAATCGCTAGTAATCGCGGATCAGCACGCCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCA
2. L. gasseri     CGAACCTGCGAAGGCAAGCGGATCTCTGAAAGCCGTTCTCAGTTCGGACTGTAGGCTGCAACTCGCCTACACGAAGCTGGAATCGCTAGTAATCGCGGATCAGCACGCCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCA
3. L. crispatus   CGAGCCTGCGAAGGCAAGCGAATCTCTGAAAGCTGTTCTCAGTTCGGACTGCAGTCTGCAACTCGACTGCACGAAGCTGGAATCGCTAGTAATCGCGGATCAGCACGCCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCA
4. L. iners       CGACCCTGTGAAGGCAAGCGGATCTCTGAAAGCCGTTCTCAGTTCGGATTGCAGGCTGCAACTCGCCTGCATGAAGCTGGAATCGCTAGTAATCGCAAATCAGCACGTTGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCA
5. L. jensenii    CGAACCTGTGAAGGCAAGCGGATCTCTTAAAGCTGTTCTCAGTTCGGACTGTAGGCTGCAACTCGCCTACACGAAGCTGGAATCGCTAGTAATCGCGGATCAGCACGCCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCA
```



Identity
1. L. acidophilus
2. L. gasseri
3. L. crispatus
4. L. iners
5. L. jensenii
LgassF
LgassR
LcrisF
LcrisR
InersF
InersR
LjensF
LjensR
LacidF
LacidR


VIC PYRO
VIC PCR
MSE PYRO
MSE PCR
MSC PYRO
MSC PCR
MPE PYRO
MPE PCR
MPC PYRO
MPC PCR
MGE PYRO
MGE PCR
MGC PYRO
MGC PCR
DSC PYRO
DSC PCR
DPE PYRO
DPE PCR
DPC PYRO
DPC PCR
DGE PYRO
DGE PCR
DGC PYRO
DGC PCR
Upper Gential Tract Site and Phase
0.00
0.25
0.50
0.75
1.00
Total Microbial Abudance in Sample
Species
L. acidophilus
L. crispatus
L. gasseri
L. iners
L. jensenil


Percent of Community
Not present
<1 %
1%-5%
5% - 10%
10% - 25%
25%-50%
50%-70%
70%-80%
80%-90%
>90%
L. acidophilus
L. crispatus
L. gasseri
L. iners
L. jensenii
VICPYRO
VICPCR
MGEPCR
MSCPCR
MPCPCR
MSEPYRO
MPEPCR
MPEPYRO
DPEPYRO
DPEPCR
DPCPCR
DPCPYRO
MSEPCR
MGEPYRO
DGCPYRO
DGEPYRO
MGCPYRO
MGCPCR
DGCPCR
DGEPCR
DSCPYRO
DSCPCR
MPCPYRO
MSCPYRO

A

**AB680529.1.1489** *L. acidophilus*
**AB690249.1.1507** *L. acidophilus*
**AB668940.1.1553** *L. acidophilus*
**AB008203.1.1553** *L. acidophilus*
**AB425941.1.1488** *L. crispatus*
**AB008206.1.1556** *L. crispatus*
**AF243169.1.1512** *L. crispatus*
**AF243167.1.1497** *L. crispatus*
**CP018809.253324.254894** *L. jensenii*
**CP018809.1516019.1517589** *L. jensenii*
**CP018809.1347636.1349206** *L. jensenii*
**CP018809.500868.502438** *L. jensenii*
**AB547127.1.1518** *L. gasseri*
**AB517146.1.1533** *L. gasseri*
**AB932527.1.1515** *L. gasseri*
**AB008209.1.1566** *L. gasseri*
**HZ485829.7.1557** *L. iners*
**LG085736.7.1557** *L. iners*
**LF134126.7.1557** *L. iners*
**LG104504.7.1557** *L. iners*
**AP012496.9810.11360** *B. subtilis*
**AP012496.30276.31828** *B. subtilis*
**AB018215.1.1530** *P. pentosaceus*
**AB362987.1.1540** *P. pentosaceus*

100 100 100 90 96 88 100 66 100 67 100 98 100 100 100

0.020

B

**AB690249.1.1507** *L. acidophilus*
**AB008203.1.1553** *L. acidophilus*
**AB680529.1.1489** *L. acidophilus*
**AB668940.1.1553** *L. acidophilus*
**AB425941.1.1488** *L. crispatus*
**AB008206.1.1556** *L. crispatus*
**AF243169.1.1512** *L. crispatus*
**AF243167.1.1497** *L. crispatus*
**AB008209.1.1566** *L. gasseri*
**AB932527.1.1515** *L. gasseri*
**AB547127.1.1518** *L. gasseri*
**AB517146.1.1533** *L. gasseri*
**HZ485829.7.1557** *L. iners*
**LG085736.7.1557** *L. iners*
**LF134126.7.1557** *L. iners*
**LG104504.7.1557** *L. iners*
**CP018809.253324.254894** *L. jensenii*
**CP018809.1347636.1349206** *L. jensenii*
**CP018809.1516019.1517589** *L. jensenii*
**CP018809.500868.502438** *L. jensenii*
**AP012496.9810.11360** *B. subtilis*
**AP012496.30276.31828** *B. subtilis*
**AB018215.1.1530** *P. pentosaceus*
**AB362987.1.1540** *P. pentosaceus*

87 98 75 73 79 70 54 99 87 99 100 100

0.0100