# Title: Evidence that vertical transmission of the vaginal microbiota can persist into adolescence

Running title: Vertical transmission of the vaginal microbiota

**Authors:** Christine M. Bassis (cbassis@umich.edu)[1], Kaylie A. Bullock (bullockk@umich.edu)[1], Daniel E. Sack (sackd@umich.edu)[1], Katie Saund (katiephd@umich.edu)[2], Ali Pirani (apirani@med.umich.edu)[2], Evan S. Snitkin (esnitkin@med.umich.edu)[1,2], Veronica I. Alaniz (veronica.alaniz@ucdenver.edu)[3], Elisabeth H. Quint (equint@med.umich.edu)[3], Vincent B. Young (youngvi@med.umich.edu)[1,2] and Jason D. Bell (jasonbel@med.umich.edu)[3].

1. Division of Infectious Disease, Department of Internal Medicine
2. Department of Microbiology and Immunology
3. Department of Obstetrics and Gynecology, University of Michigan, Ann Arbor, MI 48109

Corresponding author: Christine M. Bassis (cbassis@umich.edu)

## Abstract

**Background:** Factors that influence vaginal microbiota composition, including its source, are not well understood. To determine if vaginal microbiota transmission from mother to daughter at birth influences the human vaginal microbiota composition in adolescence, we investigated the relationship between the vaginal microbiota of 13 mother/daughter pairs and the daughter's birth mode.

**Results:** Based on analysis of bacterial 16S rRNA gene sequences, the vaginal microbiotas of mother/daughter pairs were more similar to each other if the daughter was born by vaginal delivery rather than by C-section. Additionally, genome sequences from an important member of the vaginal microbiota, *Lactobacillus crispatus*, isolated from one mother/daughter pair in which the daughter was born by vaginal delivery, were highly similar.

**Conclusions:** Both community-level analysis and isolate genome sequence analysis provide evidence of birth-mode dependent transmission and persistence of at least some members of the vaginal microbiota.

## Importance

The composition of the human vaginal microbiota is related to many aspects of health from infection susceptibility to preterm birth. Our study provides evidence that transmission of vaginal bacteria from mother to daughter at birth is an important factor influencing vaginal microbiota composition into adolescence.

## Keywords

Vaginal microbiota, transmission, birth mode, 16S rRNA gene sequences, *Lactobacillus crispatus* genomics

## Background

The vaginal microbiota plays an important role in human health. The community structure of the vaginal microbiota has strong links to HIV susceptibility and preterm birth [1-3]. While there has been considerable effort in examining the role of the gastrointestinal microbiota in human health and disease (e.g. [4]), there have been fewer studies examining the microbial communities that inhabit the female reproductive tract. The composition of the vaginal microbiota is distinct from other body sites and contains types of bacteria that seem specific to the vagina [5]. For example, the vaginal microbiota is often dominated by specific types of *Lactobacillus,* most commonly *L. crispatus* and *L. iners* [6, 7]. Vaginal *Lactobacillus* sp. are thought to maintain dominance and inhibit colonization of other microbes through lactic acid production [8, 9].

Despite strong evidence that the vaginal microbiota can have significant impacts on a woman's reproductive tract health, the factors that influence the composition of the vaginal microbiota are not well understood. It is not known how this vagina-specific community is maintained from generation to generation. One possibility is that at least some members of the vaginal microbiota are transmitted from mother to daughter at birth and maintained in daughters through adolescence.

In healthy babies, the first large, direct exposure to microbes occurs at birth. Birth mode has been shown to influence the composition of the newborn microbiota (gut, skin, mouth), likely due to different bacterial exposure in vaginal delivery and C-sections [10, 11]. However, the effect of birth mode on the composition of the vaginal microbiota has not been investigated. In this study, we compared the vaginal microbiotas of 13

mother/daughter pairs and investigated the effect of birth mode on mother/daughter microbiota similarity. We also compared the genome sequences from *Lactobacillus crispatus* isolates from one mother/daughter pair. We hypothesized that the vaginal microbiota of mothers and daughters would be more similar if the daughter was born by vaginal delivery than by C-section.

## Methods

### Subject recruitment and sample collection

Mother/daughter pairs were recruited from the Pediatric and Adolescent Gynecology Clinic at the University of Michigan Health System in 2014 and 2015. Exclusions were pregnancy and age of less than 15 years. Written, informed consent was obtained and participants completed a baseline survey on their demographics and pertinent gynecologic and medical history. Vaginal samples were self-collected using a dual-headed swab (Starplex Scientific, S09D) at baseline and then weekly for 4 weeks. The baseline swab was obtained in the clinic, with immediate storage on ice and transfer to -80°C within a few hours. The subsequent swabs were returned via mail at ambient temperature. After the fifth swab was received and a completion incentive was mailed to the subject, the link between samples and subject names was destroyed, irreversibly de-identifying all samples. The study was approved by the University of Michigan IRB (HUM00086661).

### DNA isolation and 16S rRNA gene sequencing

One of the swab heads from each sample was clipped directly into the bead plate of a PowerMag Microbiome RNA/DNA Isolation Kit (Mo Bio Laboratories, Inc.). DNA isolation was performed according the manufacturer's instructions using an epMotion

5075 liquid handling system. The V4 region of the 16S rRNA gene was amplified from 1 or 7μl DNA and sequenced with a MiSeq (Illumina, San Diego, CA) using the 500 cycle MiSeq Reagent Kit, v. 2 (Illumina, catalog No. MS-102–2003) by the University of Michigan Microbial Systems Molecular Biology Laboratory as described previously [12]. The other swab head was used for cultivation or stored at -80°C.

**Bacterial community analysis**

The 16S rRNA gene sequences were processed using mothur v.1.36.1 and v.1.39.5 following the mothur MiSeq SOP [13, 14]. Details of the processing steps are available in mother.daughter_mothur.batch (https://github.com/cbassis/MotherDaughter_Vaginal_Microbiota.study). After sequence processing and alignment to the SILVA reference alignment (Release 102) [15], sequences were binned into operational taxonomic units (OTUs) based on 97% sequence similarity using the average neighbor method [16, 17]. Samples with fewer than 1000 sequences were excluded from the analysis. OTUs were classified to the genus level within mothur using a modified version of the Ribosomal Database Project (RDP) training set (version 9) [18, 19]. To further classify the *Lactobacillus* OTUs, representative sequences were analyzed using standard nucleotide BLAST for highly similar sequences (megablast) on the National Center for Biotechnology Information (NCBI) BLAST web page (https://blast.ncbi.nlm.nih.gov/Blast.cgi) [20]. OTU relative abundances were calculated and plotted in a heatmap. To compare bacterial communities between pairs, within pairs and within subjects, we calculated $\theta_{YC}$ distances (a metric that takes relative abundances of both shared and non-shared OTUs into account) [21]. A Kruskal-Wallis test with a Dunn's posttest or a Wilcoxon

(Mann-Whitney) test were used to determine if differences in $\theta_{YC}$ distances were statistically significant. Principal coordinates analysis (PCoA) was used to visualize the $\theta_{YC}$ distances between samples. R Studio (Version 1.1.456) with R (Version 3.5.1) was used for the statistical tests and plotting the heat map, box and whisker plots, and the ordination using the code available: https://github.com/cbassis/MotherDaughter_Vaginal_Microbiota.study/tree/master/R_code. Adobe Illustrator (CS6) was used for labeling and formatting figures.

**_Lactobacillus crispatus_ isolation**

For pair I, the second swab head from the freshly collected baseline vaginal sample was swabbed onto an MRS agar plate and incubated in an anaerobic chamber (Coy Laboratory Products) at 37°C. Individual isolates were identified via Sanger sequencing of the near-full length 16S rRNA gene.

**DNA isolation and genome sequencing**

Three *Lactobacillus crispatus* isolates from pair I, 2 from the mother and 1 from the daughter, were grown overnight in 1 ml liquid MRS in an anaerobic chamber (Coy Laboratory Products) at 37°C. Genomic DNA was isolated from the liquid cultures using the PowerMicrobiome™ RNA Isolation Kit (Mo Bio Laboratories, Inc.) without the DNase treatment. Genome sequencing was performed by the Microbial Systems Molecular Biology Laboratory at the University of Michigan using an Illumina Nextera™ sequencing kit and a MiSeq (Illumina, San Diego, CA).

**Genome sequence analysis**

Phylogenetic relationships between *L. crispatus* isolates from mother/daughter pair I and all *L. crispatus* strains with genome sequences available as fastq files from NCBI

on December 27th, 2018 were determined based on recombination-filtered single nucleotide polymorphisms (SNPs). Quality of reads was assessed with FastQC v0.11.3 [22], and Trimmomatic 0.36 [23] was used for trimming adapter sequences and low-quality bases. Variants were identified by (i) mapping filtered reads to reference genome sequence *L. crispatus* ST1 (SAMEA2272191) using the Burrows-Wheeler short-read aligner (bwa-0.7.17) [24, 25], (ii) discarding polymerase chain reaction duplicates with Picard (picard-tools-2.5.0) [26], and (iii) calling variants with SAMtools (samtools-1.2) and bcftools [27]. Variants were filtered from raw results using GATK 's (GenomeAnalysisTK-3.3-0) VariantFiltration (QUAL, >100; MQ, >50; >=10 reads supporting variant; and FQ, <0.025) [28]. In addition, a custom python script was used to filter out single-nucleotide variants that were (i) <5 base pairs (bp) in proximity to indels, (ii) fell under Phage and Repeat region of the reference genome (identified using Phaster [29] and Nucmer (MUMmer3.23) [30]), (iii) not present in the core genome, or (iv) in a recombinant region identified by Gubbins 2.3.1 [31]. A maximum likelihood tree was constructed in RAxML 8.2.8 [32] using a general-time reversible model of sequence evolution. Bootstrap analysis was performed with the number of bootstrap replicates determined using the bootstrap convergence test and the autoMRE convergence criteria (-N autoMRE). Bootstrap support values were overlaid on the best scoring tree identified during rapid bootstrap analysis (-f a). The final maximum likelihood tree was plotted and pairwise SNP distances were calculated in R Studio (Version 1.1.463) with R (Version 3.5.3):

https://github.com/cbassis/MotherDaughter_Vaginal_Microbiota.study/blob/master/R_co

de/Mother_Daughter_Figure_3_Genome_Tree_and_Genome_Analysis.Rmd. Adobe Illustrator (CS6) was used for labeling and formatting the figure.

**Calculation of doubling time estimate for vaginal *L. crispatus in vivo***

We used the number of SNPs between the pair I mother and daughter *L. crispatus* isolates to estimate the doubling time of vaginal *L. crispatus in vivo* if all SNPs in the recombination-filtered core genome were due to mutations acquired since the daughter's birth:

Doubling time=(mutation rate)(daughter's age)(genome length)/(# of mutations)

The mutation rate of *L. crispatus* is unknown, so for this estimate we used the published mutation rate of another *Lactobacillus, L. casei* Zhang, *in vitro,* without antibiotics ($1.0 \times 10^{-9}$ bp/generation) [33]. The pair I daughter's age in hours was: 175,200 hours =(20 years)(365 days/year)(24 hour/day). The average length of the recombination-filtered core genome (940,943 bp) was used for genome length. We assumed that the isolates arose from a common ancestor and that all mutations were non-convergent, so the number of mutations acquired would equal the number of SNPs between the mother's isolate and the daughter's isolate divided by 2.

**Results**

**Subject characteristics and sequencing results**

A total of 107 self-collected, vaginal swab samples were obtained from 26 subjects (13 mother/daughter pairs) (Table 1). Each subject returned 1-5 weekly samples (median=5 samples/subject, IQR=1). After sequence processing and exclusion of samples with fewer than 1000 sequences, a total of 2,336,437 high quality bacterial 16S rRNA gene sequences from 101 samples were analyzed with an average of 23,133

184 +/- 10,212 sequences per sample.

**Table 1. Subject Characteristics**

| | Mother (n=13) | Daughter (n=13) |
|---|---|---|
| | n (%) | n (%) |
| **Age, mean ± SD, years** | 47±6 | 17±2 |
| **Race: White (vs. Black, Asian, Hispanic, other)** | 12 (92) | 12 (92) |
| **Birth mode: Vaginal (vs. C-section)** | 10 (77) | 10 (77) |
| **Reproductive stage: Premenarchal** | 0 (0) | 1 (8) |
| **Reproductive stage: Reproductive** | 8 (62) | 12 (92) |
| **Reproductive stage: Postmenopausal** | 5 (38) | 0 (0) |

185

186 **An individual's vaginal microbiota is relatively stable over 4 weeks**

187 During the sampling period, the vaginal microbiota of each subject was relatively
188 stable. The high stability of the vaginal microbiota is apparent from the consistent within
189 subject community composition (Figure 1). For example, high relative abundances of
190 OTU1 (*L. crispatus*) and/or OTU2 (*L. iners*) persisted from week to week in many
191 subjects. Additionally, average $\theta_{YC}$ distances were significantly lower within subjects
192 than between subjects (Figure 2A) and samples clustered by subject in a PCoA based
193 on $\theta_{YC}$ distances (Supplemental Figure 1).

194 **Daughters born via vaginal delivery have greater microbiota similarity with their**
195 **mothers than those born via C-section**

196 To determine if mothers and their daughters had more similar vaginal microbiotas
197 than unrelated subjects, we compared the average $\theta_{YC}$ distances between all unrelated
198 subjects (between pairs) and the average $\theta_{YC}$ distances between mothers and their own
199 daughters (within pairs) (Figure 2A). There was a trend toward greater similarity (lower
200 $\theta_{YC}$ distances) within all mother/daughter pairs than between subjects in different

mother/daughter pairs. To determine if birth mode was related to vaginal microbiota similarity within mother/daughter pairs, we compared the average within pair $\theta_{YC}$ distances for pairs in which the daughter was born by vaginal delivery and by C-section (Figure 2B). The average within pair $\theta_{YC}$ distances were significantly lower for pairs in which the daughter was born by vaginal delivery compared to C-section (Fig. 2B). Therefore, the vaginal microbiotas of daughters born by vaginal delivery were significantly more similar to their mothers' than the daughters born by C-section were to their mothers' (Fig. 2B).

**_Lactobacillus crispatus_ isolates from mother/daughter pair I have highly similar genome sequences**

The birth mode-dependent similarity of the vaginal microbiotas of mothers and their daughters suggested that vaginal bacteria could be transmitted between generations at birth and persist into adolescence. However, it is possible that genetic or environmental factors shared by a mother and her daughter lead to acquisition of similar bacteria later, resulting in the *de novo* establishment of similar vaginal communities. To investigate the possibility of direct transmission and persistence of one member of the vaginal microbiota, we generated draft genome sequences of *Lactobacillus crispatus* strains isolated from the freshly collected second swab head of mother/daughter pair I. The draft genome sequences of these isolates were compared with publicly available *L. crispatus* genome sequences by constructing a maximum likelihood phylogenetic tree based on a recombination-filtered core genome alignment. Interestingly, the three strains of *L. crispatus* from mother/daughter pair I, UMP1M1, UMP1M2 and UMP1D1, were more similar to each other than to any of the other strains, including others

224 isolated from the female reproductive tract (Fig.3).

225 We also calculated the number of SNPs between our isolates using the
226 recombination-filtered core genome alignment. There were 11 recombination-filtered
227 SNPs between the 2 isolates from the mother (UMP1M1 and UMP1M2) and 25 and 16
228 recombination-filtered SNPs between the daughter's isolate (UMP1D1) and the 2 isolates
229 from the mother (UMP1M1 and UMP1M2, respectively).

230 **Estimate of *in vivo* doubling time for vaginal *L. crispatus***

231 To further investigate the plausibility that the *L. crispatus* strain isolated from
232 daughter I descended from a strain transmitted from her mother at birth, we estimated
233 the doubling time that would allow our isolates to acquire the observed number of SNPs
234 over 20 years. Based on the 25 SNPs between UMP1M1 and UMP1D1, the estimated
235 doubling time for *L. crispatus in vivo* would be 13.2 hours. Based on the 16 SNPs
236 between UMP1M2 and UMP1D1, the estimated doubling time would be 20.6 hours.

237 **Discussion**

238 Our study provides preliminary evidence that the vaginal microbiota is vertically
239 transmitted from mother to daughter at birth via vaginal delivery and persists into
240 adolescence. Because the daughters in our study were 15-21 years old, both
241 transmission and persistence were required to observe evidence of vertical
242 transmission. The first piece of evidence supporting vertical transmission is that the
243 vaginal microbiotas of mothers and their adolescent daughters were more similar if their
244 daughter was born by vaginal delivery rather than C-section. The second piece of
245 evidence supporting vertical transmission and persistence is that an important member
246 of the vaginal microbiota, *L. crispatus,* isolated from a vaginally-born, 20-year-old

daughter and her mother (pair I) had highly similar genome sequences.

Other studies have compared the vaginal microbiotas of mothers and daughters without detecting notable similarity between them [34, 35]. There are multiple reasons that high similarity between mothers and daughters was not observed in these studies. First, the effect of birth mode was not analyzed in these previous studies. In our study, there was a trend toward greater community similarity within mother/daughter pairs, regardless of birth mode, than between unrelated subjects in different mother/daughter pairs (Figure 2A). However, most (10/13) of the daughters in our study were born by vaginal delivery. Indeed, when mother/daughter pairs were separated by the daughter's birth mode, the vaginal microbiotas of mothers and their daughters were significantly different if the daughter was born by C- section. So, if many of the daughters in the other studies were born by C-section then high similarity between mothers and daughters would not be expected. With C- section rates of ~30% in the United States (study site for [35]) and ~36% in South Korea (study site for [34]) this is a possibility [36, 37]. Additionally, our study focused on adolescent daughters (age 15-21) while the other studies focused on either younger or older daughters. Since reproductive stage seems to influence the structure of the vaginal microbiota [38], differences in reproductive stage may contribute to differences in vaginal community composition between mothers and daughters. Finally, we used a different method of comparing the vaginal microbiotas of mothers and daughters. We calculated distances between mothers and daughter using $\theta_{YC}$, a metric that accounts for the relative abundances of shared and non-shared OTUs, while the other studies were based on community types [35] and Unifrac [34]. Although an overall community similarity was not observed in these

studies, specific community members (*Lactobacillus* and *Prevotella*) were identified as most heritable in one study [34].

Based on the number of SNPs observed between the mother and daughter *L. crispatus* isolates and published mutations rates for *L. casei* Zhang [33], we estimated that *L. crispatus* would have an *in vivo* doubling time of 13.2-20.6 hours, depending on the specific isolates compared. The doubling time estimates of 13.2 hours and 20.6 hours for *L. crispatus in vivo* are within the range estimated for other bacteria in their natural environments, including *Escherichia coli* (15 hours) and *Salmonella enterica* (25 hours) [39]. These doubling times are faster than the 4.1-5.6 days doubling times measured for *L. casei* Shirota in mouse intestines, where its growth rate was insufficient to maintain colonization [40]. Although the actual growth and mutation rates of *L. crispatus* in the human vagina have not been measured, we estimated reasonable *in vivo* doubling times for vaginal *L. crispatus* based on the observed number of SNPs between *L. crispatus* isolates from mother/daughter pair I, the age of daughter I and *L. casei* Zhang mutation rates. Considering the uncertainty in the estimates, transmission of *L. crispatus* from mother to daughter at birth followed by the accumulation of independent mutations during 20 years of persistence in the mother and daughter is a plausible explanation for the observed recombination-filtered SNPs. Future studies comparing genomes of *L. crispatus* isolates from more mother/daughter pairs with a variety of daughter ages are needed.

The 2 *L. crispatus* isolates from the mother had highly similar genomes, differing by only 11 recombination-filtered SNPs. A previous study also observed high similarity between the genomes of multiple vaginal *L. crispatus* isolates from one individual, noting

293 that they were indistinguishable [41]. Future investigations of *L. crispatus* genomic
294 variation within an individual may yield further insight on colonization and dynamics of the
295 vaginal microbiota.

296 Consistent with a previous study, *L. crispatus* isolates from the human vagina were
297 phylogenetically intermixed with isolates from the human urinary tract, including highly
298 similar vaginal (ERS1867668 (SAMEA104208650)) and bladder (ERS1867667
299 (SAMEA104208649)) isolates from the same subject (Figure 3) [42].

300 The health implications of vertical transmission of the vaginal microbiota are
301 unknown and were not addressed in this study. However, because vertical transmission
302 seems to be an important factor in determining the composition of the vaginal
303 microbiota there may be important consequences. Vertical transmission of the vaginal
304 microbiota may be one mechanism for maintaining human microbiota over generations
305 via a consistent and specific seeding of the newborn microbiota. Delivery mode is an
306 important factor in determining the early composition of the gut microbiota [43, 44] and
307 is a risk factor for development of immune-related disorders later in life [45]. This
308 suggests an important role for the mother's vaginal microbiota in seeding the infant and
309 setting the stage for development of the gut microbiota. Therefore, maintenance of the
310 vaginal microbiota between generations may be critical for gut microbiota development
311 in each generation.

312 Additionally, the vaginal microbiota plays an important if not well understood role
313 in reproductive health, with associations between vaginal microbiota composition and
314 infection susceptibility, BV and preterm birth [46-48]. Evidence from this study suggests
315 that transmission of microbes from mother to daughter at birth may influence the

316 composition of the daughter's microbiota later in life and may contribute to the
317 maintenance of specific members of the human vaginal microbiota over generations.
318 This study provides tantalizing evidence of vertical transmission of the vaginal
319 microbiota. However, this was a small study with only 13 mother/daughter pairs (92%
320 white) and 3/13 daughters born by C-section. Genomic analysis of isolates was limited
321 to one member of the vaginal microbiota from 1 mother/daughter pair. Future studies in
322 larger populations, including more racially diverse subjects, more daughters born by C-
323 section and analysis of more isolate genome sequences or metagenomes are required
324 to validate these findings.

325 **Conclusions**

326 At least some members of the vaginal microbiota are transmitted from mother to
327 daughter at birth and persist into adolescence. Birth-mode dependent vaginal
328 microbiota transmission is an important factor determining vaginal microbiota
329 composition.

330 **Figures**

331 **Figure 1. Vaginal bacterial community compositions of mother/daughter pairs.**
332 Relative abundances of OTUs in weekly vaginal swab samples from 13
333 mother/daughter pairs. Mother/daughter pairs were ordered by average within pair $\theta_{YC}$
334 distances, with the most similar pair (I) on top and the least similar pair (XIII) on the
335 bottom. OTUs with a minimum of 200 sequences in the dataset overall and present at a
336 relative abundance greater than 2% in at least 1 sample were included in the heat map.

337 **Figure 2. Average distances between vaginal bacterial communities.** A. Average
338 $\theta_{YC}$ distances between subjects from different mother/daughter pairs (between pairs),

between subjects within a mother/daughter pair (within pair) and between samples from the same subject (within subject). P-values for comparisons that were significantly different by Dunn's posttest are shown (Kruskal-Wallis p-value= 8.154e-10). B. Average $\theta_{YC}$ distances between subjects within a mother/daughter pair for daughters born by vaginal birth and by C-section. Wilcoxon (Mann-Whitney) test p-value is shown. In the box and whiskers plots, the median $\theta_{YC}$ distance is indicated by a line, values within the first to the third quartiles are inside the box and the whiskers extend to the smallest and largest values within 1.5x the interquartile range.

**Figure 3. Phylogenetic relationships between *L. crispatus* strains.** Maximum likelihood tree based on recombination-filtered SNP distances between *L. crispatus* genome sequences of isolates from mother/daughter pair I and other *L. crispatus* strains with publicly available genomes. Tip labels indicate *L. crispatus* strain names and NCBI BioSample identifiers. Bootstrap values were greater than or equal to 0.65.

**Supplemental Figure 1. Principal coordinates analysis (PCoA) of vaginal microbiota from 13 mother/daughter pairs.** The $\theta_{YC}$ distances between 101 vaginal microbiota samples are represented by PCoA. Samples from daughters are represented by triangles and samples from mothers by circles. Each mother/daughter pair is represented by a unique color. Biplot arrows represent the 3 OTUs most correlated with position on the PCoA plot.

**List of abbreviations**

C-section: Cesarean section

rRNA: ribosomal RNA

OTU: operational taxonomic unit

SNPs: single nucleotide polymorphisms

PCoA: principal coordinates analysis

## Declarations

- Ethics approval and consent to participate

All subjects provided written informed consent. The study was approved by the University of Michigan IRB (HUM00086661).

- Consent for publication

Not applicable.

- Availability of data and material

The raw sequence data generated in this study are available in the NCBI's SRA:

Bacterial 16S rRNA gene sequences: BioProject PRJNA547595

*L. crispatus* draft genome sequences: BioProject PRJNA547620

GitHub repository:

https://github.com/cbassis/MotherDaughter_Vaginal_Microbiota.study

This repository includes:

- the mothur batch file with steps used to process and analyze 16S rRNA gene sequences
- mothur output files used in final bacterial community analysis and figures
- R code for manuscript figures, statistics and genomic analysis

- Competing interests

The authors declare that they have no competing interests.

- Funding

Not applicable.

- Authors' contributions

CMB was involved in study design and planning, data analysis, figure preparation and manuscript writing. KAB was involved in subject recruitment, sample processing, isolation of *L. crispatus* genomic DNA for sequencing, data analysis and manuscript editing. DES was involved in subject recruitment, sample processing and manuscript editing. KS was involved in genomic data analysis, interpretation of genomic data, phylogenetic tree construction and manuscript editing. AP was involved in genomic data analysis and manuscript editing. ES was involved in genomic data analysis and interpretation and manuscript review. VIA was involved in subject recruitment and planning. EHQ was involved in study design and planning, subject recruitment and manuscript editing. VBY was involved in study design and manuscript editing. JDB was involved with study design and planning, subject recruitment and manuscript editing. All authors read and approved the final manuscript.

- Acknowledgements

We'd like to thank the participants who contributed samples. We'd also like to thank Megan Lemieur for help with study organization and Qualtrics survey implementation and the University of Michigan Microbial Systems Molecular Biology Laboratory for swab DNA isolation, bacterial community and genomic library preparation and sequencing. We'd also like to thank Ingrid Bergin and members of the Young lab for many discussions related to this project over the years. This research was also supported in part through computational resources and services provided by Advanced Research Computing at the University of Michigan, Ann Arbor.

## References


1. Callahan BJ, DiGiulio DB, Goltsman DSA, Sun CL, Costello EK, Jeganathan P, Biggio JR, Wong RJ, Druzin ML, Shaw GM, et al: **Replication and refinement of a vaginal microbial signature of preterm birth in two racially distinct cohorts of US women.** *Proceedings of the National Academy of Sciences* 2017.
2. Klatt NR, Cheu R, Birse K, Zevin AS, Perner M, Noël-Romas L, Grobler A, Westmacott G, Xie IY, Butler J, et al: **Vaginal bacteria modify HIV tenofovir microbicide efficacy in African women.** *Science (Wash DC)* 2017, **356:**938.
3. McClelland RS, Lingappa JR, Srinivasan S, Kinuthia J, John-Stewart GC, Jaoko W, Richardson BA, Yuhas K, Fiedler TL, Mandaliya KN, et al: **Evaluation of the association between the concentrations of key vaginal bacteria and the increased risk of HIV acquisition in African women from five cohorts: a nested case-control study.** *The Lancet Infectious Diseases* 2018, **18:**554-564.
4. Durack J, Lynch SV: **The gut microbiome: Relationships with disease and opportunities for therapy.** *The Journal of Experimental Medicine* 2018**:**jem.20180448.
5. **Structure, function and diversity of the healthy human microbiome.** *Nature* 2012, **486:**207-214.
6. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SSK, McCulle SL, Karlebach S, Gorle R, Russell J, Tacket CO, et al: **Vaginal microbiome of reproductive-age women.** *Proceedings of the National Academy of Sciences* 2011, **108:**4680-4687.
7. Bassis CM, Allsworth JE, Wahl HN, Sack DE, Young VB, Bell JD: **Effects of intrauterine contraception on the vaginal microbiota.** *Contraception* 2017, **96:**189-195.
8. O'Hanlon DE, Moench TR, Cone RA: **Vaginal pH and Microbicidal Lactic Acid When Lactobacilli Dominate the Microbiota.** *PLoS ONE* 2013, **8:**e80074.
9. Tachedjian G, O'Hanlon DE, Ravel J: **The implausible "in vivo" role of hydrogen peroxide as an antimicrobial factor produced by vaginal microbiota.** *Microbiome* 2018, **6:**29.
10. Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N, Knight R: **Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns.** *Proceedings of the National Academy of Sciences* 2010, **107:**11971-11975.
11. Madan JC, Hoen AG, Lundgren SN, et al.: **Association of cesarean delivery and formula supplementation with the intestinal microbiome of 6-week-old infants.** *JAMA Pediatrics* 2016, **170:**212-219.
12. Seekatz AM, Theriot CM, Molloy CT, Wozniak KL, Bergin IL, Young VB: **Fecal Microbiota Transplantation Eliminates *Clostridium difficile* in a Murine Model of Relapsing Disease.** *Infection and Immunity* 2015, **83:**3838-3846.
13. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD: **Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform.** *Appl Environ Microbiol* 2013, **79:**5112-5120.
14. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, et al: **Introducing mothur: Open-Source, Platform-**

**Independent, Community-Supported Software for Describing and Comparing Microbial Communities.** *Appl Environ Microbiol* 2009, **75:**7537-7541.

15. Schloss PD: **A High-Throughput DNA Sequence Aligner for Microbial Ecology Studies.** *PLoS ONE* 2009, **4:**e8230.
16. Westcott SL, Schloss PD: **De novo clustering methods outperform reference-based methods for assigning 16S rRNA gene sequences to operational taxonomic units.** *PeerJ* 2015, **3:**e1487.
17. Schloss PD, Westcott SL: **Assessing and Improving Methods Used in Operational Taxonomic Unit-Based Approaches for 16S rRNA Gene Sequence Analysis.** *Appl Environ Microbiol* 2011, **77:**3219-3226.
18. Wang Q, Garrity GM, Tiedje JM, Cole JR: **Naive Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy.** *Appl Environ Microbiol* 2007, **73:**5261-5267.
19. Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, Brown CT, Porras-Alfaro A, Kuske CR, Tiedje JM: **Ribosomal Database Project: data and tools for high throughput rRNA analysis.** *Nucleic Acids Res* 2014, **42:**D633-642.
20. Morgulis A, Coulouris G, Raytselis Y, Madden TL, Agarwala R, Schäffer AA: **Database indexing for production MegaBLAST searches.** *Bioinformatics (Oxf)* 2008, **24:**1757-1764.
21. Yue JC, Clayton MK: **A similarity measure based on species proportions.** *Communications in Statistics-Theory and Methods* 2005, **34:**2123-2131.
22. **FastQC** [http://www.bioinformatics.babraham.ac.uk/projects/fastqc/]
23. Bolger AM, Lohse M, Usadel B: **Trimmomatic: a flexible trimmer for Illumina sequence data.** *Bioinformatics (Oxf)* 2014, **30:**2114-2120.
24. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform.** *Bioinformatics (Oxf)* 2009, **25:**1754-1760.
25. **BWA** [http://bio-bwa.sourceforge.net]
26. **Picard** [https://broadinstitute.github.io/picard/]
27. **Calling SNPs/INDELs with SAMtools/BCFtools** [http://samtools.sourceforge.net/mpileup.shtml]
28. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA: **The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.** *Genome Res* 2010, **20:**1297-1303.
29. Pon A, Marcu A, Arndt D, Grant JR, Sajed T, Liang Y, Wishart DS: **PHASTER: a better, faster version of the PHAST phage search tool.** *Nucleic Acids Res* 2016, **44:**W16-W21.
30. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL: **Versatile and open software for comparing large genomes.** *Genome Biol* 2004, **5:**R12.
31. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill J, Harris SR: **Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins.** *Nucleic Acids Res* 2015, **43:**e15.
32. Stamatakis A: **RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies.** *Bioinformatics (Oxford, England)* 2014, **30:**1312-1313.

33. Wang J, Dong X, Shao Y, Guo H, Pan L, Hui W, Kwok L-Y, Zhang H, Zhang W: **Genome adaptive evolution of Lactobacillus casei under long-term antibiotic selection pressures.** *BMC Genomics* 2017, **18:**320.
34. Si J, You HJ, Yu J, Sung J, Ko G: ***Prevotella* as a Hub for Vaginal Microbiota under the Influence of Host Genetics and Their Association with Obesity.** *Cell Host & Microbe* 2016, **21:**97-105.
35. Hickey RJ, Zhou X, Settles ML, Erb J, Malone K, Hansmann MA, Shew ML, Van Der Pol B, Fortenberry JD, Forney LJ: **Vaginal Microbiota of Adolescent Girls Prior to the Onset of Menarche Resemble Those of Reproductive-Age Women.** *mBio* 2015, **6:**e00097-00015.
36. Kim SJ, Kim SJ, Han K-T, Park E-C: **Medical costs, Cesarean delivery rates, and length of stay in specialty hospitals vs. non-specialty hospitals in South Korea.** *PLOS ONE* 2017, **12:**e0188612.
37. Caughey AB, Cahill AG, Guise J-M, Rouse DJ: **Safe prevention of the primary cesarean delivery.** *American Journal of Obstetrics and Gynecology* 2014, **210:**179-193.
38. Brotman RM, Shardell MD, Gajer P, Fadrosh D, Chang K, Silver MI, Viscidi RP, Burke AE, Ravel J, Gravitt PE: **Association between the vaginal microbiota, menopause status, and signs of vulvovaginal atrophy.** *Menopause* 2018, **25:**1321-1330.
39. Gibson B, Wilson Daniel J, Feil E, Eyre-Walker A: **The distribution of bacterial doubling times in the wild.** *Proceedings of the Royal Society B: Biological Sciences* 2018, **285:**20180789.
40. Lee YK, Ho PS, Low CS, Arvilommi H, Salminen S: **Permanent Colonization by *Lactobacillus casei* Is Hindered by the Low Rate of Cell Division in Mouse Gut.** *Appl Environ Microbiol* 2004, **70:**670.
41. Abdelmaksoud AA, Koparde VN, Sheth NU, Serrano MG, Glascock AL, Fettweis JM, Strauss JF, 3rd, Buck GA, Jefferson KK: **Comparison of Lactobacillus crispatus isolates from Lactobacillus-dominated vaginal microbiomes with isolates from microbiomes containing bacterial vaginosis-associated bacteria.** *Microbiology* 2016, **162:**466-475.
42. Thomas-White K, Forster SC, Kumar N, Van Kuiken M, Putonti C, Stares MD, Hilt EE, Price TK, Wolfe AJ, Lawley TD: **Culturing of female bladder bacteria reveals an interconnected urogenital microbiota.** *Nat Commun* 2018, **9:**1557.
43. Wampach L, Heintz-Buschart A, Fritz JV, Ramiro-Garcia J, Habier J, Herold M, Narayanasamy S, Kaysen A, Hogan AH, Bindl L, et al: **Birth mode is associated with earliest strain-conferred gut microbiome functions and immunostimulatory potential.** *Nature Communications* 2018, **9:**5091.
44. Rutayisire E, Huang K, Liu Y, Tao F: **The mode of delivery affects the diversity and colonization pattern of the gut microbiota during the first year of infants' life: a systematic review.** *BMC Gastroenterology* 2016, **16:**86.
45. Tamburini S, Shen N, Wu HC, Clemente JC: **The microbiome in early life: implications for health outcomes.** *Nature Medicine* 2016, **22:**713.
46. Bayigga L, Kateete DP, Anderson DJ, Sekikubo M, Nakanjako D: **Diversity of vaginal microbiota in sub-Saharan Africa and its effects on HIV transmission and prevention.** *American Journal of Obstetrics and Gynecology* 2018.
47. Tamarelle J, Thiébaut ACM, de Barbeyrac B, Bébéar C, Ravel J, Delarocque-Astagneau E: **The vaginal microbiota and its association with human papillomavirus, Chlamydia**

**trachomatis, Neisseria gonorrhoeae and Mycoplasma genitalium infections: a systematic review and meta-analysis.** *Clinical Microbiology and Infection* 2019, **25:**35-47.

48. Callahan BJ, DiGiulio DB, Goltsman DSA, Sun CL, Costello EK, Jeganathan P, Biggio JR, Wong RJ, Druzin ML, Shaw GM, et al: **Replication and refinement of a vaginal microbial signature of preterm birth in two racially distinct cohorts of US women.** *Proceedings of the National Academy of Sciences* 2017, **114:**9966.

Figure 1

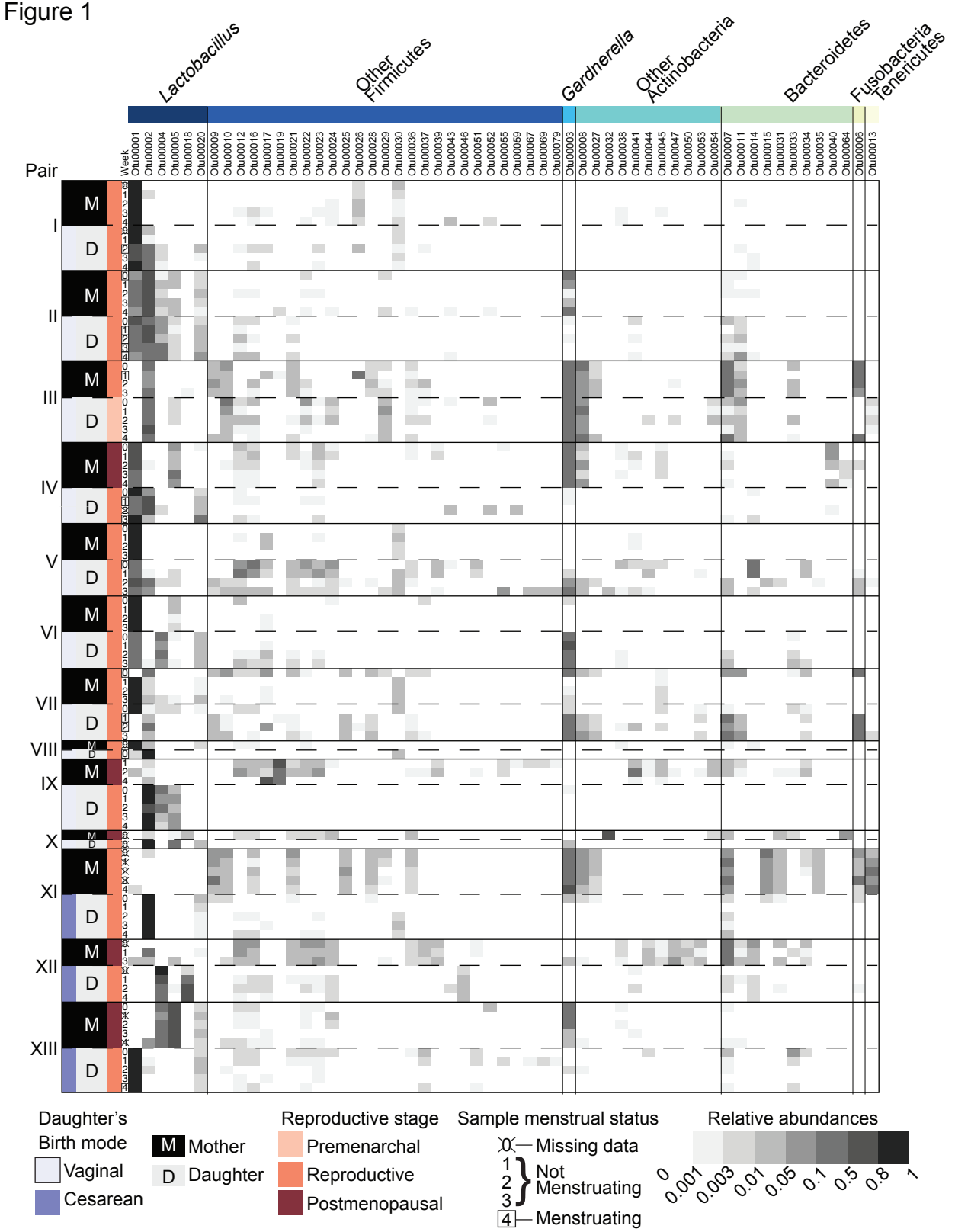
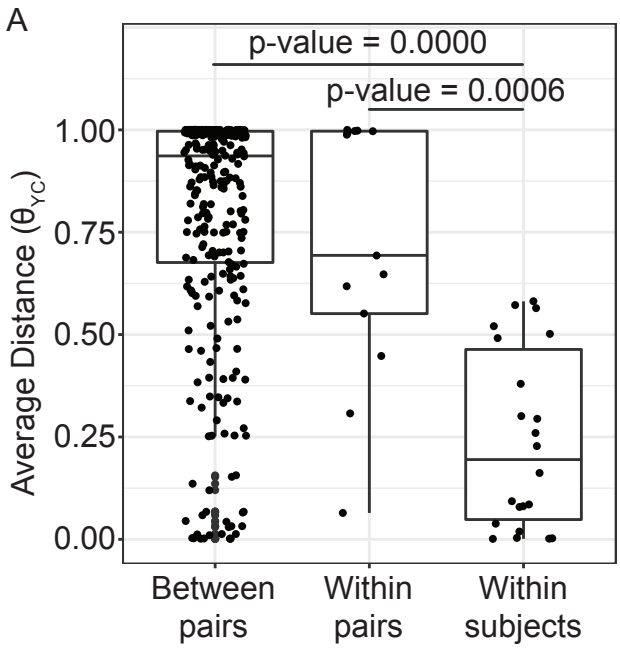

Lactobacillus
Other Firmicutes
Gardnerella
Other Actinobacteria
Bacteroidetes
Fusobacteria Tenericutes
Pair
Week
I
II
III
IV
V
VI
VII
VIII
IX
X
XI
XII
XIII
M
D
Daughter's Birth mode
Vaginal
Cesarean
Mother
Daughter
Reproductive stage
Premenarchal
Reproductive
Postmenopausal
Sample menstrual status
Missing data
Not Menstruating
Menstruating
Relative abundances
0
0.001
0.003
0.01
0.05
0.1
0.5
0.8
1

Figure 2


A
p-value = 0.0000
p-value = 0.0006
Average Distance ($\theta_{YC}$)
1.00
0.75
0.50
0.25
0.00
Between pairs
Within pairs
Within subjects

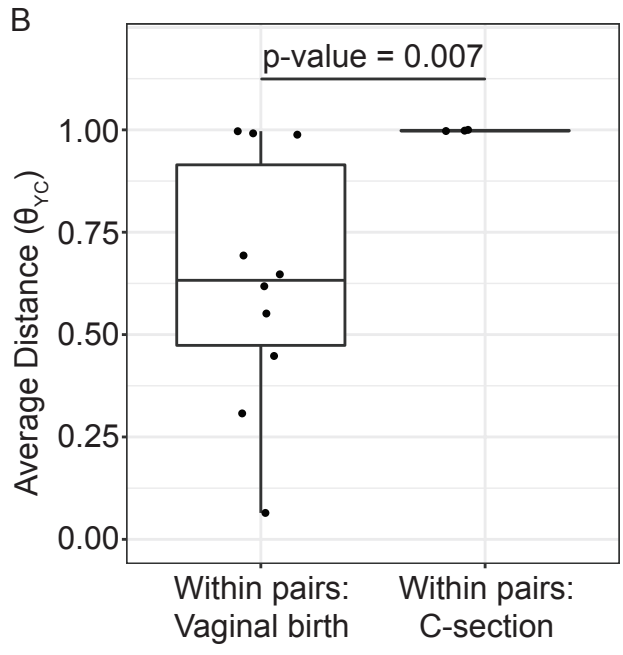
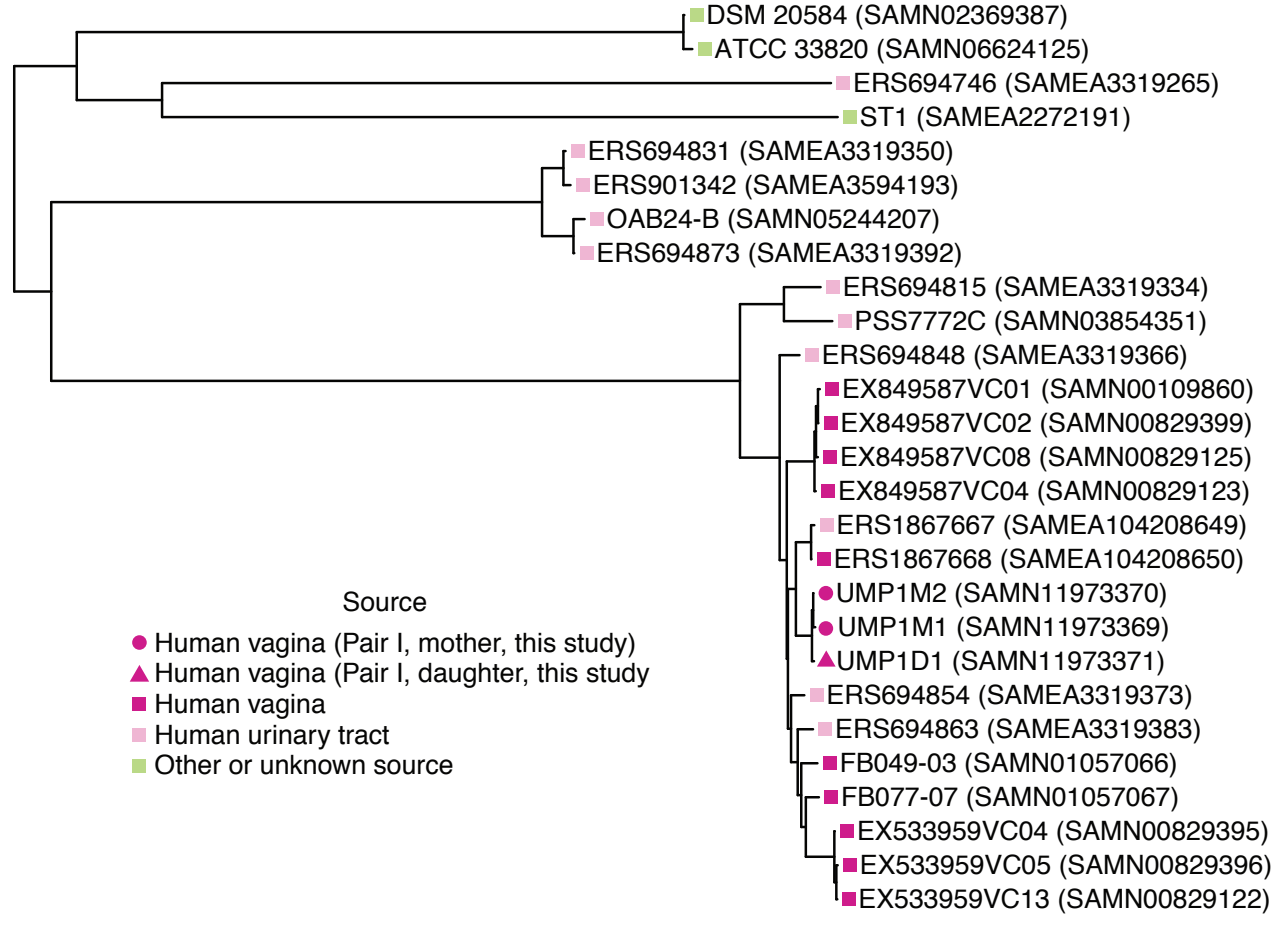

B
p-value = 0.007
Average Distance ($\theta_{YC}$)
1.00
0.75
0.50
0.25
0.00
Within pairs: Vaginal birth
Within pairs: C-section

Figure 3


DSM 20584 (SAMN02369387)
ATCC 33820 (SAMN06624125)
ERS694746 (SAMEA3319265)
ST1 (SAMEA2272191)
ERS694831 (SAMEA3319350)
ERS901342 (SAMEA3594193)
OAB24-B (SAMN05244207)
ERS694873 (SAMEA3319392)
ERS694815 (SAMEA3319334)
PSS7772C (SAMN03854351)
ERS694848 (SAMEA3319366)
EX849587VC01 (SAMN00109860)
EX849587VC02 (SAMN00829399)
EX849587VC08 (SAMN00829125)
EX849587VC04 (SAMN00829123)
ERS1867667 (SAMEA104208649)
ERS1867668 (SAMEA104208650)
UMP1M2 (SAMN11973370)
UMP1M1 (SAMN11973369)
UMP1D1 (SAMN11973371)
ERS694854 (SAMEA3319373)
ERS694863 (SAMEA3319383)
FB049-03 (SAMN01057066)
FB077-07 (SAMN01057067)
EX533959VC04 (SAMN00829395)
EX533959VC05 (SAMN00829396)
EX533959VC13 (SAMN00829122)
Source
Human vagina (Pair I, mother, this study)
Human vagina (Pair I, daughter, this study
Human vagina
Human urinary tract
Other or unknown source

Supplemental Figure 1