Text S1: Supplemental Methods and Results

Vaginal microbiota of adolescent girls prior to the onset of menarche resemble those of reproductive-age women

Roxana J. Hickey[1,2], Xia Zhou[1,†], Matthew L. Settles[2], Julie Erb[3], Kristin Malone[3], Melanie A. Hansmann[3], Marcia L. Shew[4], Barbara Van Der Pol[5], J. Dennis Fortenberry[4], and Larry J. Forney[1,2]

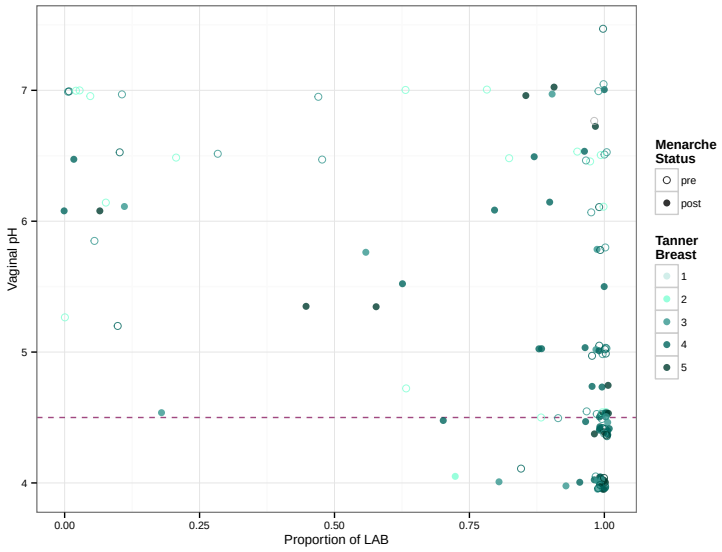
[1]Department of Biological Sciences, University of Idaho, Moscow, ID
[2]Institute for Bioinformatics and Evolutionary Studies, University of Idaho, Moscow, ID
[3]FemCare Division, Procter & Gamble, Cincinnati, OH
[4]Department of Pediatrics, Indiana University School of Medicine, Indianapolis, IN
[5]Department of Medicine, University of Alabama at Birmingham, Birmingham, AL
[†]Present address: Vista Encology Inc., Olympia, WA

Table of Contents

1 Estimation of 16S rRNA gene copy number in low-pH and high-pH vaginal microbiota

1.1 Pan-bacterial 16S rRNA qPCR assay

We observed many vaginal microbiota samples from girls that had a high relative abundance of lactobacilli along with a vaginal pH higher than what is typically considered healthy in adults (i.e., ≤ 4.5). We hypothesized this might be due to lower total numbers of bacteria, and thus lower levels of lactic acid production, which contributes to the low vaginal pH seen in most healthy women. We performed a coarse test of this hypothesis using the bacterial BactQuant qPCR assay, developed by Liu et al. (C. M. Liu, M. Aziz, S. Kachur, P. R. Hsueh, Y. T. Huang, P. Keim and L. B. Price, BMC Microbiol. 12:1, 2012), to estimate the number of 16S rRNA gene copies present in vaginal samples from girls with high proportions of lactobacilli (> 0.75) and either low (< 5.0, n=62) or high (≥ 5.0, n=40) vaginal pH. Twelve genomic DNA samples from each of the low and high groups were selected by a random sampling function in R. DNA samples had been stored in AE buffer (Qiagen, Venlo, Netherlands) at -20°C since the time of extraction and purification, ranging from December 2010 to August 2012 (the qPCR was done in June 2014). Each qPCR reaction consisted of 5.0 µL of 2X qPCR SuperMix (Invitrogen, Carlsbad, CA, USA), 0.45 µL of 40 µM forward primer, 0.45 µL of 40 µM reverse primer, 0.113 µL of 20 µM 6-FAM-labeled TaqMan probe (Applied Biosystems, Foster City, CA, USA), 2.99 µL molecular-grade water, and 1 µL genomic DNA template, for a total reaction volume of 10 µl. The primer and probe sequences were as described by Liu et al.:



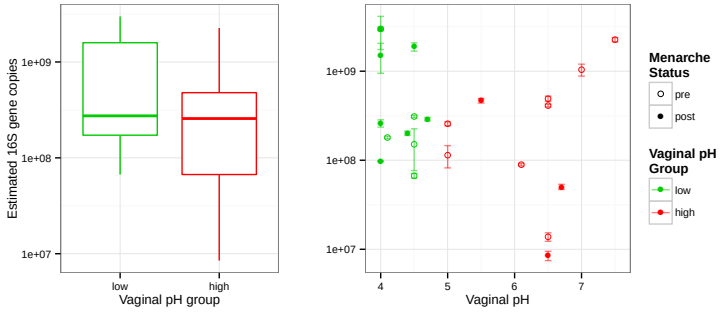
Relationship between proportion of lactic acid bacteria and pH in vaginal samples collected from girls. 197 vaginal microbiota samples from perimenarcheal girls are plotted to show the relationship between the proportion of lactic acid bacteria (LAB; includes *Lactobacillus*, *Streptococcus*, *Aerococcus* and *Facklamia*) on the x-axis and vaginal pH on the y-axis. Points are slightly jittered to decrease crowding around similar values. The magenta dashed line at vaginal pH 4.5 represents the upper range of the traditional 'hallmark' healthy vaginal pH of 4.0-4.5.

Primer/probe	Sequence	*E. coli* region (16S rRNA)
Forward primer	5'-CCTACGGG**D**GGC**W**GCA-3'	341-356
Reverse primer	5'-GGACTAC**HV**GGGT**M**TCTAATC-3'	786-806
Probe	(6FAM) 5'-CAGCAGCCGCGGTA-3' (MGBNFQ)	519-532

where bold font (in the primer sequences) denotes degenerate bases. Samples, along with an in-run standard curve (103-109 in 10-fold serial dilutions) and no-template controls, were processed in triplicate. Amplification and real-time fluorescence detections were performed on the AB StepOnePlus Real-Time PCR System (Applied Biosystems) with the following PCR conditions: 3 minutes at 50°C for UNG treatment, 10 minutes at 95°C for *Taq* activation, 15 seconds at 95°C for denaturation, and 1 minute at 60°C for annealing and extension, repeated over 40 cycles. Cycle threshold values (i.e., Ct value) were determined, and 16S rRNA gene copy number estimates generated from the standard curve, using StepOne Software v2.1 (Applied Biosystems).

1.2 Comparison of 16S rRNA gene copies in low-pH and high-pH vaginal microbiota

Results of the pan-bacterial 16S rRNA qPCR assay are shown in the figure at right. Out of 24 samples, 23 were successfully amplified. Estimated 16S rRNA copy numbers ranged from 6.7E+07 to 3.0E+09 copies per µL (mean 9.1E+08 copies per µL) in the low-pH samples, and 8.5E+06 to 2.3E+09 copies per µL (mean 4.7E+08 copies per µL) in the high-pH samples. These were compared using a one-tailed t-test (H_0: low-pH 16S rRNA copies per µl > high-pH 16S rRNA copies/µl) and found to be not significantly different (p=0.14). However, the distribution of the low-pH group appears to be skewed higher than in the high-pH group, providing at least some indication that bacterial counts may indeed be higher in low-pH microbiota, even if the difference is not statistically significant in the subset of samples we tested. Linear modeling also did not detect any statistically significant associations of 16S rRNA copy number or pH with individual *Lactobacillus* spp. present in the communities (data not shown).



Estimated number of 16S rRNA gene copies in low-pH vs. high-pH vaginal microbiota samples from girls. The left plot shows box plots for the estimated number of gene copies for the 'low' (green) and 'high' (red) groups. The right plot shows the vaginal pH (x-axis) and estimated number of gene copies (y-axis) for each sample with standard error bars. Open circles represent premenarcheal samples and filled circles represent postmenarcheal samples.

Although we failed to detect a significant difference in 16S rRNA copies (as a rough proxy for bacterial cells) in relation to vaginal pH, we should not rule out the possibility based on this analysis. We note that factors such as sample quantity and quality may have posed significant limitations to the qPCR assay, particularly since it was performed well after genomic DNA had been extracted and archived. Furthermore, the amount of material collected on the vaginal swabs was not strictly controlled for, other than rotating each swab in the vaginal introitus three times during collection. Differences in the amount of vaginal secretions or sloughed epithelial cells collected on the swabs could greatly impact the amount of DNA recovered. This is why we typically analyze proportions rather than absolute numbers of bacterial taxa. A more thorough evaluation of bacterial counts during puberty would ideally utilize fresh samples and some means of controlling for sample quantity (e.g., weighing swab material, quantifying vaginal secretions, or determining the proportion of bacterial cells or DNA relative to human).

2 Genomic DNA extraction and 16S rRNA pyrosequencing

2.1 Genomic DNA extraction and purification

Genomic DNA was extracted from vaginal and vulvar swabs as previously described (S. Yuan, D. B. Cohen, J. Ravel, Z. Abdo and L. J. Forney, PLoS ONE 7:e33865, 2012) by cell lysis using enzymatic and bead-beating treatments followed by purification using QIAamp DNA Mini Kit (Qiagen). Briefly, vaginal swabs suspended in Amies medium (Copan Diagnostics, Murrietta, CA, USA) were thawed on ice and vigorously vortexed for 5 minutes to dislodge and resuspend cells. A 500 µL aliquot was transferred to a clean sterile beading-beating tube (MP Biomedicals, Santa Ana, CA, USA) and kept on ice. A lytic enzyme cocktail was prepared at the time of extraction and added to each sample as follows: 50 µL of 10 mg/mL lysozyme, 6 µL of 25,000 U/mL mutanolysin (Sigma-Aldrich, St. Louis, MO, USA), 3 µL of 4,000 U/mL lysostaphin in sodium acetate (Sigma-Aldrich), and 41 µL TE50 buffer for a final volume of 100 µL per sample. Samples were digested by incubation at 37°C for 60 minutes in a dry heat block. 750 mg sterile 0.1 mm diameter zirconia/silica beads (Biospec Products, Bartlesville, OK, USA) were added to each digested sample. Bead-beating was performed for 1 minute at 36 oscillations per second (2,100 rpm) with the use of a Mini-Beadbeater-96 (Biospec Products). Following cell disruption, the tubes were centrifuged at 1,200 rpm for 1 minute. Aliquots of crude lysate from each sample were transferred to new sterile microcentrifuge tubes, and 50 µL proteinase K (20 mg/mL [>600 mAU/mL]) and 500 µL Qiagen buffer AL were added. Samples were mixed by pulse-vortexing for 15 seconds and then incubated at 56°C for 30 minutes. After this step, 50 µL 3 mol/L sodium acetate (pH 5.5) was added, followed by 500 µLL 100% ethanol at each sample. Vortexing was repeated for an additional 15 seconds before briefly centrifuging. From this point onward, purification of genomic DNA was done with QIAamp DNA Mini Kits following the manufacturer's instructions.

2.2 Pyrosequencing V1-V3 regions of 16S rRNA genes

To characterize the composition and structure of bacterial communities in vaginal and vulvar samples, we sequenced the V1-V3 regions of 16S rRNA genes amplified from each sample. The amplicons were obtained by PCR using primers that flanked hypervariable regions 1 and 3 of bacterial 16S rRNA genes (*Escherichia coli* positions 27-534). The sequences of the primers used were as follows:

Primer	Sequence
454_27F-YM	5'-CCTATCCCCTGTGTGCCTTGGCAGTCTCAGTC**AGAGTTTGATYMTGGCTCAG**-3'
454_27F-Bif	5'-CCTATCCCCTGTGTGCCTTGGCAGTCTCAGTC**AGGGTTCGATTCTGGCTCAG**-3'
454_27F-Bor	5'-CCTATCCCCTGTGTGCCTTGGCAGTCTCAGTC**AGAGTTTGATCCTGGCTTAG**-3'
454_27F-Chl	5'-CCTATCCCCTGTGTGCCTTGGCAGTCTCAGTC**AGAATTTGATCTTGGTTCAG**-3'
454_534R	5'-CCATCTCATCCCTGCGTGTCTCCGACTCAG*NNNNNNNN*TC**ATTACCGCGGCTGCTGGCA**-3'

where the underlined sequences are Roche 454 fusion adapters B and A in 27F and 534R, respectively, and the bold font denotes the universal 16S rRNA primers 27F and 534R. The four 27F primers were combined in equal amounts and designated 27F*. The 534R primer included a unique sequence tag to barcode each of the samples

denoted by the 8 italicized Ns. This allowed us to sequence the amplicons from all samples simultaneously, and afterwards assign each sequence to the sample they were obtained from. Each PCR contained 34.4 µl of molecular-grade water, 5.0 µl of 10X buffer (Applied Biosystems), 6.0 µl of 25 mM $MgCl_2$ (Applied Biosystems), 0.4 µl of 25 mM dNTP (Amersham Bioscience, Amersham, UK), 0.5 µl of 20 µM forward primer 454_27F*, and 0.5 µl of 20 µM reverse primer 454_534R, 0.2 µl of 5 U/µl Taq DNA polymerase (Applied Biosystems), and 1.0 µl of DNA template, in a total volume of 50 µl. Amplification of fragments was done using an initial denaturation step at 94°C for 4 min, followed by 30 cycles of denaturation at 94°C for 1 min, annealing at 55°C for 1 min, and an extension at 72°C for 2 min. A final extension step of 10 minutes at 72°C was done.

Concentrations of amplicons were estimated with the use of a fluorometric Picogreen assay on a SpectraMax GeminiXPS 96-well plate reader (Molecular Devices, Sunnyvale, CA, USA), and roughly equal amounts (~100 ng) were mixed in a single microfuge tube. Amplification primers and reaction buffer were removed by processing the amplicon mixture with the Agencourt AMPure Kit (Beckman Coulter, Brea, CA, USA). To determine the final quality we amplified the resulting amplicon pool with 454 adapter-specific primers in order to mimic the emulsion PCR (emPCR) process and processed the PCR product on a DNA 1000 chip for the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). The final amplicon pool was deemed acceptable only if no short fragments were identified after PCR; otherwise the procedure was repeated again. The cleaned amplicon pool was then quantified using the KAPA 454 library quantification kit (KAPA Biosciences, Wilmington, MA, USA) and the Applied Biosystems StepOne plus real-time PCR system (Applied Biosystems). Emulsion PCRs were performed on each quantified pool and final sequences were obtained using a Roche 454 GS-FLX+ pyrosequencer (Roche 454 Life Sciences, Branford, CT, USA).

2.3 DNA sequence data analysis and taxonomic classification

Raw, unclipped DNA sequence reads were cleaned, assigned, and filtered in the following manner. Raw SFF files were read directly in R using the R package rSFFreader (`http://www.bioconductor.org/packages/release/bioc/html/rSFFreader.html`). Full-length (unclipped) sequence reads were used for the identification of Roche 454 adapters, barcodes and amplicon primers sequence using Cross Match (v1.080806; parameters: minimum matches 8, minimum score 16) from the Phred/Phrap/Consed application suite. Cross Match alignment information was then read into R and processed to identify alignment quality, directionality, barcode assignment, and sequence quality clip points. Base-quality clipping was done with the use of the Lucy application (v1.20p; parameters: maximum average error 0.002, maximum error at ends 0.002), and the clipped reads were aligned with the SILVA bacterial sequence database with the use of mothur v1.27 (P. D. Schloss, S. L. Westcott, T. Ryabin, J. R. Hall, M. Hartmann, E. B. Hollister, et al., Appl. Environ. Microbiol. 75:7537–7541, 2009). Alignment end points were identified and used in subsequent filtering. Sequence reads were filtered to only those that met the following criteria: (a) sequences at least 100 bp in length; (b) maximum Hamming distance of barcode = 1; (c) maximum number of matching error to forward primer sequences = 2; (d) < 2 ambiguous bases (Ns); (e) < 10bp homopolymer run in sequence; (f) alignment to the SILVA bacterial database (`http://www.arb-silva.de/`) was within 75 bp of the expected alignment start position (507 bp); and (g) read alignment started within the first 5 bp and extended through read to within the final 5 bp. Each partial 16S rRNA gene sequence was classified with the use of the Ribosomal Database Project (RDP) Naïve Bayesian

Classifier (Q. Wang, G. M. Garrity, J. M. Tiedje and J. R. Cole, Appl. Environ. Microbiol. 73:5261–5267, 2007). Reads were assigned to the first RDP level with a bootstrap score ≥ 50.

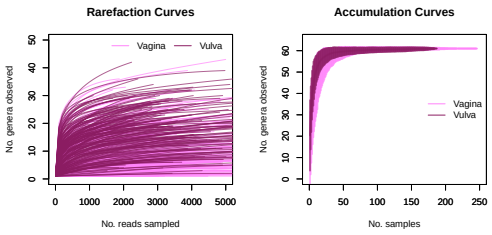
In this study, reads assigned to *Lactobacillus*, *Streptococcus* or *Gardnerella* were further assigned to the species level. Reference 16S sequences from the target genera were first extracted from the Patric database of bacterial genomes (`http://patricbrc.org`). These were aligned to each other using mothur and the SILVA database to generate a species reference alignment. Reads matching the target genus were extracted from the larger dataset and clustered using CD-HIT (W. Li and A. Godzik, Bioinformatics 22:1658-1659, 2006; L. Fu, B. Niu, Z. Zhu, S. Wu and W. Li, Bioinformatics 28:3150-3152, 2012) at 99.5% identity to reduce redundancy in the reads as well as the overall size of the dataset. Cluster representatives were then aligned to the SILVA database using mothur, and the species reference alignment was merged into the read alignment. Pairwise distances for the alignments were calculated using mothur and loaded into R for clustering and species identification. The pairwise distance matrix was clustered using hierarchical clustering with average linkage using the R package flashClust (P. Langfelder and S. Horvath, J. Stat. Software 46:1-17, 2012). Cluster modules were identified using the function cutreeDynamic from the WGCNA package (P. Langfelder and S. Horvath, BMC Bioinform. 9:559, 2008), varying the deepSplit parameter from 1 to 4. The final deepSplit parameter was selected based upon visual inspection of the hierarchical tree and module designations. Each read was then assigned to a cluster module based on its cluster representative's module assignment. If the module contained species representative sequences, the module was assigned this species name; otherwise the module was assigned with an ambiguous species identifier.

Downstream analysis of taxonomic composition was performed in R as described in the main text. Analysis scripts are available on GitHub.

3 Community richness and diversity analyses

3.1 Rarefaction analysis

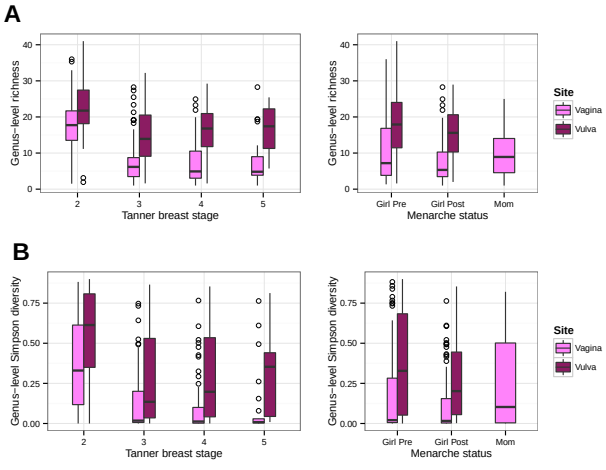
We performed analyses of community richness and diversity to qualitatively assess how they varied in relation to pubertal development and vaginal pH. These analyses are detailed on GitHub at `https://github.com/roxanahickey/adolescent/blob/master/06-post-review.md`. Because there was considerable variability in the sequence read count among samples, it was necessary to perform a rarefaction analysis and randomly subsample counts at the same depth across all samples prior to calculating richness and diversity. Rarefaction curves were generated using the 'rarecurve' function in the vegan R package. From this plot we determined that subsampling at 2,000 observations per sample would sufficiently detect the observed genera without having to disregard a significant proportion of our samples. We excluded 39 samples with less than 2,000 sequence reads from further analysis. We also generated genus accumulation curves using the 'specaccum' function in vegan. These curves indicate that most genera in our dataset could be observed from 50 or more samples, well below the number of samples we had available. A new genus abundance matrix was generated by subsampling at a depth of 2,000 observations using the 'rrarefy' function in vegan.



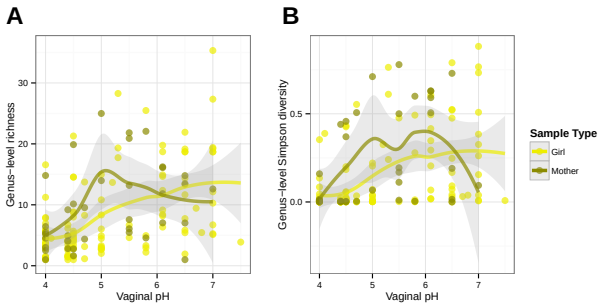
Genus rarefaction and accumulation curves. On the left, rarefaction curves are shown for 456 individual vagina (light pink) and vulva (dark magenta) samples. Genus accumulation curves are plotted on the right.

3.2 Genus richness and diversity

A considerable caveat to our analysis is that any taxa that could not be classified at the genus level were combined into an 'Other' category and treated as a single genus. Estimates of richness and diversity are therefore likely to be quite conservative based on these numbers. For this reason we elected not to perform quantitative analyses of diversity; an OTU analysis would perhaps be better suited for that purpose. Richness was calculated using the 'rarefy' function on the subsampled data. Shannon and Simpson diversity indices were calculated using the 'diversity' function. Because Shannon's index is more sensitive to large differences in richness, we chose to favor Simpson's index. Genus richness and diversity varied in both the vagina and vulva across Tanner breast stages and menarcheal stages as shown in the box plots in the figure at right.



Trends in genus-level richness and Simpson's diversity index. 181 vaginal, 190 vulvar, and 44 vaginal samples are represented. (A) Genus richness in relation to Tanner breast stage (girls only) on left; richness in relation to menarche status (girls and mothers) on right. (B) Simpson's diversity in relation to Tanner breast stage on left, and menarche status on right.



Genus-level richness, Simpson's diversity index and vaginal pH. Each point represents a vaginal sample from either a girl or a mother. Vaginal pH was recorded for 120 samples from 24 girls and 37 samples from 21 mothers. (A) Genus-level richness plotted against vaginal pH. A locally weighted scatterplot smoothing (LOESS) function was applied separately to girl and mother data points, with 95% confidence intervals indicated by the light grey regions. (B) Genus-level Simpson's diversity index plotted against vaginal pH with LOESS curves for girls and mothers.

As perhaps expected, the vulva generally exhibited higher richness and diversity than the vagina but still experienced downward trends with progressive Tanner stages. Differences between premenarche and postmenarche were less pronounced. Richness and diversity of the vaginal microbiota varied with respect to vaginal pH as shown in the figure at left. The positive relationship between increasing richness or diversity and vaginal pH is interesting, but we caution against extrapolating biologically meaningful conclusions from these data for reasons described above.