SUPPORTING INFORMATION APPENDIX

Title: Temporal and Spatial Variation of the Human Microbiota During Pregnancy

Running head: Microbiota During Pregnancy

Authors: Daniel B. DiGiulio[a,b,c,1], Benjamin J. Callahan[a,d,1], Paul J. McMurdie[a,d], Elizabeth K. Costello[a,e], Deirdre J. Lyell[a,f], Anna Robaczewska[a,b,c], Christine L. Sun[a,e], Daniela S. Aliaga Goltsman[a,e], Ronald J. Wong[a,g], Gary M. Shaw[a,g], David K. Stevenson[a,g], Susan P. Holmes[a,d], David A. Relman[a,b,c,e,2]

Affiliations:
[a]March of Dimes Prematurity Research Center, Stanford University School of Medicine, Stanford, CA 94305; [b]Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305; [c]Veterans Affairs Palo Alto Health Care System, Palo Alto, CA 94304; [d]Department of Statistics, Stanford University, Stanford, CA 94305; [e]Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA 94305; [f]Department of Obstetrics and Gynecology, Stanford University School of Medicine, Stanford, CA 94305; and [g]Department of Pediatrics, Stanford University School of Medicine, Stanford, CA 94305

[1]D.B.D. and B.J.C. contributed equally to this work.

[2]To whom correspondence should be addressed. Email: relman@stanford.edu

Contents

Supporting Methods

Study population and sampling procedures

Pregnant women presenting to the obstetrical clinics of the Lucille Packard Children's Hospital at Stanford University for prenatal care were invited to participate in this study. The cohort of women in the present analysis was enrolled between November 2011 and December 2013. Inclusion criteria included age 18 years or older, ability to perform the study procedures, and ability to provide informed consent. Exclusion criteria included use of antibiotics or probiotics within the twelve weeks prior to enrollment, significant immunosuppression, and upon subsequent review, delivery of a baby with congenital defects. From the population of enrolled subjects, two case-control groups were selected for analysis. The first group consisted of 40 women from whom a discovery data set was generated; the second group consisted of nine women from whom a validation data set was generated. For both groups, cases consisted of women who delivered preterm, and controls were women who delivered at term. The study was approved by an Administrative Panel for the Protection of Human Subjects (Institutional Review Board) of Stanford University (IRB protocol #21956); all women provided written informed consent prior to completing an enrollment questionnaire and providing biological samples.

Specimens from the vagina, of stool, saliva, and from tooth/gum were self-collected by participants at weekly intervals from the time of study enrollment until delivery, and at monthly intervals from the time of delivery for up to 12 months. Saliva (2-5 mL volume) was collected in a sterile 50 mL conical collection tube (B-D Falcon, Franklin Lakes, NJ, USA). The other three body sites were sampled using sterile Catch-All™ Sample Collection Swabs (Epicentre Biotechnologies, Madison, WI, USA) at the 1) lateral vaginal wall, 2) rectal mucosa, and 3) molar tooth surfaces including along the gum-line. All clinical specimens were placed immediately after collection at -20°C until transport to the laboratory for storage at -80°C until processing. Information related to socioeconomic status, diet, medical conditions and symptoms, medication and dietary supplement use, and stress were captured in a detailed questionnaire completed by each study participant upon enrollment; any significant changes in these parameters were documented in a follow-up questionnaire completed at each subsequent clinical visit.

DNA extraction, 16S rDNA amplification and amplicon sequencing

Whole genomic DNA was extracted from each clinical specimen by means of the PowerSoil® DNA isolation kit (MO BIO laboratories, Carlsbad, CA, USA) according to the manufacturer's

protocol except for the inclusion of a 10 min incubation at 65°C immediately after the addition of solution C1. The 16S rRNA gene region that was targeted differed for each of the two sequencing approaches we used (454 pyrosequencing and Illumina HiSeq 2500). For pyrosequencing, the V3-V5 hyper-variable region of the 16S rRNA gene was amplified by PCR. The forward PCR primer (5′ **CGT ATC GCC TCC CTC GCG CCA TCA G**NN NNN NNN NNN NGC ACT CCT ACG GGA GGC AGC A 3′) was a 58-nucleotide (nt) fusion primer consisting of the 25-nt 454 Life Sciences primer 'A' sequence (designated by bold font), a unique 12-nt error-correcting Golay barcode to label each amplicon (designated by the N's) (1), a 2-nt linker (GC) immediately after the barcode, and the 19-nt broad-range bacterial primer 338F (designated by underlining). The 47-nt reverse primer (5′ **CTA TGC GCC TTG CCA GCC CGC TCA G**AA CCG TCA ATT CCT TTG AGT TT 3′) consisted of the 25-nt 454 Life Sciences primer 'B' sequence (designated by bold font), a 2-nt linker (AA), and the 20-nt broad-range bacterial primer 906R (designated by underlining).

For Illumina-based sequencing, the V4 hypervariable region of the 16S rRNA gene was amplified by PCR. The forward PCR primer (5′ **AAT GAT ACG GCG ACC ACC GAG ATC TAC ACG CT**N NNN NNN NNN NNT ATG GTA ATT GTG TGY CAG CMG CCG CGG TAA 3′) was a 74-nucleotide (nt) fusion primer consisting of the 32-nt Illumina adapter (designated by bold font), a unique 12-nt barcode to label each amplicon (designated by the N's), a 9-nt forward primer pad, a 2-nt linker (GT), and the 19-nt broad-range bacterial primer 515F (designated by underlining). The 56-nt reverse primer (5′ **CAA GCA GAA GAC GGC ATA CGA GAT** AGT CAG CCA GCC GGA CTA CNV GGG TWT CTA AT 3′) consisted of the 24-nt Illumina adapter (designated by bold font), a 10-nt reverse primer pad, a 2-nt reverse primer linker (CC), and the 20-nt broad-range bacterial primer 806R (designated by underlining).

For both pyrosequencing and Illumina-based sequencing, triplicate 25-µL PCR reactions were carried out using 1× HotMasterMix (5 PRIME, Gaithersburg, MD, USA), 0.4 µM concentrations of each commercially-synthesized primer, and 3 µL of prepared DNA template. Thermal cycling conditions consisted of an initial denaturing step of 94°C for 3 min, followed by 30 cycles of 94°C for 45 sec, 52°C for 30 sec, and 72°C for 90 sec, with a final extension step of 72°C for 10 min. Upon completion of the PCR reactions, the corresponding triplicate reaction mixtures were pooled and purified using the Ultra-clean-htp 96-well PCR clean-up kit (MO BIO laboratories, Carlsbad, CA, USA) according to the manufacturer's protocol. DNA concentrations from each triplicate pool were quantified using the Quant-iT™ High-Sensitivity dsDNA Assay Kit (Invitrogen, Carlsbad, CA, USA) and combined in equimolar ratios into a single tube. The

resulting amplicon mixture was concentrated by ethanol precipitation and re-suspended in 100 µL of molecular-biology-grade water (Life Technologies, Grand Island, NY, USA). The resuspended amplicon mixture was gel purified and recovered using a QIAquick gel extraction kit (Qiagen, Valencia, CA, USA). For PCR products originating from the first group of 40 women, pyrosequencing was carried out by the Genome Sequencing and Analysis Core Resource at Duke University (Durham, NC, USA) using the Roche 454 FLX instrument (454 Life Sciences, Branford, CT, USA) and titanium series reagents. The resulting sequences comprised the discovery data set. For PCR products originating from the second group of nine women, Illumina-based sequencing was carried out by the W. M. Keck Center for Comparative Functional Genomics at the University at Illinois, Urbana-Champaign (USA) using an Illumina HiSeq 2500 instrument (Illumina Inc., San Diego, CA, USA). The resulting sequences comprised the validation data set.

Sequence filtering, OTU clustering and chimera removal

Pyrosequencing and Illumina-based datasets were analyzed similarly, but separately. Quality-filtering of pyrosequencing reads was performed in a two-step process. First, de-multiplexing and initial quality processing of reads was performed using QIIME version 1.7 (http://qiime.org) and python scripts. To pass this step, sequences had to meet the following criteria: i) no mismatches to the PCR primer sequence or to the barcode tag; ii) no homopolymer runs >6 bases long; and iii) read length between 375 and 600 bases. Second, sequence reads that met these criteria were evaluated for sequence quality at read ends. The aim of this evaluation was to determine the optimal global trimming length, as global trimming to a fixed length enables alignment of sequences with no terminal gaps (2) . Based on these assessments, 350 bases was chosen as the global trimming length. Finally, trimmed reads were filtered for a maximum average expected error rate of 0.25 nucleotide per 350-base sequence read, a relatively stringent threshold (http://drive5.com/usearch/manual/fastq_choose_filter.html). From a total of 26,689,010 input sequences, this filtering approach yielded 19,306,851 high-quality 350-base reads for subsequent analysis. Raw Illumina read-pairs (2 x 250 bases) were quality filtered and merged using SeqPrep software version 1.1 (https://github.com/jstjohn/SeqPrep.git). Merged reads >230 bases and <270 bases in length were retained and de-multiplexed in QIIME, yielding 50,034,186 high-quality reads.

OTU clustering at a 97% sequence identity threshold was performed using the UPARSE algorithm based on a maximum parsimony model and implemented within the USEARCHv7 software package (http://www.drive5.com/usearch/). UPARSE leverages a ‘greedy’ clustering

algorithm that in benchmark tests against other commonly used OTU clustering approaches significantly reduces the likelihood of spurious OTUs (2) . The first step in a greedy clustering approach is to order sequence reads by abundance, with singletons temporarily discarded. Next, OTU centroids are created by considering sequence reads in order of decreasing abundance; a new OTU centroid is created if the sequence read being considered has a lower percentage identity with any existing OTU than the chosen threshold (97% for our analysis). The rationale for this approach is that high-abundance reads are much more likely than low-abundance reads to be correct amplicon sequences and, therefore, true biological sequences.

After OTU clustering, removal of chimeric sequences was performed in a stringent two-step process. First, *de novo* chimera filtering was performed to detect and remove chimeric models built from more abundant 'parent' reads. Next, reference-based chimera filtering was performed to remove remaining chimeric sequences that may have escaped *de novo* filtering (e.g., chimeras of parent sequences that are absent from the reads or are present at very low abundance). This step was performed using the UCHIME algorithm (3) informed by the ChimeraSlayer reference database (aka 'gold' database) developed as one of the microbiome utilities by the Broad Institute (http://microbiomeutil.sourceforge.net/).

Raw sequence data have been deposited at SRA (see SRP288562). The final OTU table and associated data are available as Datasets S1 and S2.

Sequence alignment and taxonomic assignment

One representative sequence from each OTU was used to generate a sequence alignment against the Greengenes core set (4) using PyNAST (5). Taxonomic assignments of OTU representatives were made using the Ribosomal Database Project (RDP) classifier version 2.2 with a minimum support threshold of 80% and the RDP taxonomic nomenclature (6).

Bioinformatics approach and statistical analysis

Statistical analyses were performed using 'R' language and environment (R 2014, http://www.r-project.org) version 3.1.1, and the add-on packages 'phyloseq' (7), 'ggplot2' (8), 'cluster' (9), 'nlme' (10), and 'DEseq2' (11). The code is publicly available: http://statweb.stanford.edu/~susan/papers/PNASRR.html.

Evaluating trends with gestational time

Because we consider the communities at each body site separately, the subject factor is the largest source of variation in sampled communities. Therefore, progressive changes in stability

and diversity over the course of pregnancy were evaluated by linear mixed-effects (LME) modeling using the 'nlme::lme' function in R. The subject was included as a random effect for both the intercept and the slope of the estimated fit. Significance of the fixed effect of gestational time (the subject-independent portion of the trend) was then evaluated by a t-test. We performed a Kruskal-Wallis test on the residuals grouped by subject. The absence of any small p-values shows that the subject effect is no longer present after the MLE fit is performed. We also performed additional Kruskal-Wallis tests on the residuals grouped by race, and by White/Non-white race, and again found no small p-values.

Beta-diversity is a function of distances between pairs of communities, and thus, the subject factor is not accounted for by adding the subject as a random-effect. Therefore, a permutation test was used to evaluate the significance of the linear trend estimated on the beta-diversities. The test statistic was the slope m of the simple linear fit of average pairwise distances to gestational weeks. The null distribution of this statistic $\{m^*\}$ was estimated from an ensemble of 1000 randomly time-reversed sets of samples. Each time-reversed set of samples was generated by reversing, with 50% probability, the gestational times associated with the samples of each subject. The two-tailed p-value was then calculated as normal, $p = \text{Prob}(|m^*| > |m|)$.

In all cases, the analysis of trends with gestational time was repeated with multiple measures of stability and diversity. For alpha-diversity, the Shannon Diversity Index (12) was supplemented with the Chao1 estimate of species richness (13), and the Simpson diversity index (14). For beta-diversity and instability, both of which are based on distances between communities, the phylogenetically-aware weighted UniFrac distance was supplemented with the non-phylogenetic Bray-Curtis (15) and Jensen-Shannon (16) dissimilarities. Also, since the p-values reported for these trends were almost all individually insignificant, no further correction for multiple testing was performed.

Evaluating differential abundance

DESeq2 (11) was used to perform two-class testing for differential relative abundance. Paired tests (by subject) were used when comparing pre- vs. post-delivery and early vs. late pregnancy. Unpaired tests were used when comparing term vs. preterm delivery. DESeq2 uses a negative-binomial model for count data that better reflects the overdispersion of OTU counts between biological replicates. The overdispersion parameter is estimated by sharing the information across OTUs, and is allowed to vary with mean abundance (11). DESeq2 has been shown to be efficient in testing microbiome data (17).

Prior to testing for differential abundance, an independent filter was used to exclude OTUs absent in >75% of samples (18). OTUs were considered significantly differentially abundant between classes if their adjusted p-value was below 0.1 (corresponding to a FDR <10% under the Benjamini-Hochberg correction) and if the estimated fold change was >1.5 or <1/1.5.

Clustering into community state types (CSTs)

CSTs are clusters of community samples with similar compositions of microbial taxa. The (dis)similarity between communities is quantified by the dissimilarity between communities' vectors of relative OTU abundances. There are many dissimilarity measures in common usage, but perhaps the most important distinction is between those measures which account for phylogenetic relatedness (e.g., UniFrac) and those that do not (e.g., Bray-Curtis). Typically the weighted UniFrac measure is a more robust choice for community clustering, but because of the unique structure of vaginal communities -- most are dominated by a single Lactobacillus strain, although it can be one of several species -- the Bray-Curtis distance was used to define the vaginal CSTs.

First, the Bray-Curtis distance between all samples was calculated. This distance matrix was 'denoised' by extracting the most significant PCoA eigenvectors. The partitioning around medoids algorithm (*pam* in R) was applied to these PCoA distances. The number of clusters (k=5) was determined from the gap statistic (19), a goodness of cluster measurement (Fig. S6). This clustering effectively separated vaginal communities into five distinct CSTs: four dominated by different *Lactobacillus* species and one CST with greater diversity and without a dominant *Lactobacillus* strain. These CSTs were analogous to those described previously (20, 21), and were named in accordance.

Estimating vaginal CST transition rates

The sampling regimen in this study was weekly, but as is typical with self-collection there was a significant amount of missing data, as well as variation in the day of the week on which sampling was performed. Therefore, before estimating vaginal CST transition rates, the data set was restricted to those pairs of consecutive samples that were collected 4-10 days apart. This set of 652 paired samples had time-separations of 4-10 days, mean of 6.96 days, and a 1st-quartile, 3rd-quartile and median of 7 days. The one-week transition rate was quantified as the maximum-likelihood estimate from this set of paired samples.

Analysis scripts and code

As a means of enhancing the reproducibility of our results, the reader will find all the code necessary to reproduce the analyses and figures presented in this article at <http://statweb.stanford.edu/~susan/papers/PNASRR.html>. This website includes pointers to the R markdown files, the output in html, and to the data in the form of OTU count tables. In addition, raw sequence data are deposited at SRA (SRP #pending), and OTU tables and associated data at PNAS as Datasets.

Supporting Discussion

Because vaginal communities exhibited inter-state transitions, we represented vaginal CST dynamics as a Markov chain. A Markov chain is a schematic representation of a multi-state system that undergoes transitions between states as a function only of the current state. Our model indicated that the four *Lactobacillus*-dominated CSTs (CSTs 1, 2, 3, and 5) were more stable (had higher self-transition probabilities) than the diverse CST (4). This is qualitatively similar to what Gajer *et al*. observed in non-pregnant women (22); however, stability of the *Lactobacillus*-dominated CSTs was higher in our cohort. We observed one-week self-transition probabilities of 97.9%, 97.6% and 87.5% for CST 1, 2 and 3, respectively; whereas Gajer and colleagues observed one-week self-transition probabilities of approximately 70.9%, 79.7% and 73.6% for these same CSTs.

The structure of the observed inter-CST transition patterns is of interest. CST 2 (*L. gasseri*-dominated) had the fewest connections. Indeed, in our cohort, CST 2 was not observed to be reachable from any other CST, and when CST 2 transitioned to another state, it transitioned to CST 1 (*L. crispatus*-dominated) only. Although precise characterization of CST 2 dynamics was limited by low numbers, these general features of CST 2 transitions – especially its few connections overall – were similar to prior observations (22).

In their recent longitudinal study, Romero and colleagues found no significant difference in taxon abundances or diversity of vaginal communities between term and preterm pregnancies (23). One possible reason for the contradictory results is the difference in racial composition between our study cohorts. Black women comprised a vast majority (79/90) in the study by Romero et al., but only a small fraction (2/40) in our study. Previous work has demonstrated significant differences between vaginal microbial communities associated with race (20). In particular, black race was associated with higher proportions of the diverse CSTs, which could interact with the association between such states and preterm birth observed here and in other studies. In addition, the study by Romero et al. included only those preterm births occurring prior

to 34 gestational weeks. Also, our cohort included some women whose labor onset was non-spontaneous (e.g., due to preterm premature rupture of membranes) (Table S2).

Another potentially important difference between our study and previous work, especially when considering patterns related to *Gardnerella* abundance, may be the use of different PCR primers (21, 22). A commonly used forward primer has up to three mismatches for many taxa within the phylum *Actinobacteria* which includes the genera *Gardnerella* and *Atopobium* (24). We used forward primer 338F, which provides a perfect sequence match to the vast majority of the *Actinobacteria*.

Both *Gardnerella* and *Ureaplasma* have been implicated by a large body of literature as having potential roles in the pathogenesis of preterm delivery. *Gardnerella vaginalis* is the bacterium that accounts for 'clue cells', which are epithelial cells with adherent gram-variable bacteria that are visualized on a wet mount of cervicovaginal fluid and are considered a *sine qua non* of bacterial vaginosis, a dysbiosis that is associated with an approximately two-fold increased risk of preterm delivery (25, 26). *Ureaplasma* species are the bacteria most commonly found to invade the human amniotic cavity and have been associated with preterm birth in culture-based and molecular studies (27-30). It is possible that absolute or relative abundance, and not merely presence/absence of bacterial taxa may have implications for whether a baby is born preterm; in addition, strain-level associations of *Gardnerella, Ureaplasma,* or of other species may be important. Furthermore, our data suggest that concurrent alterations in the abundances of more than one taxon (e.g., low *Lactobacillus* plus high *Gardnerella*) may play a previously under-recognized role.

Conclusion

We found the microbial communities of the vagina, distal gut, saliva and tooth/gum to be remarkably stable during pregnancy. A particular state type of the vaginal community, CST 4, exhibited strong associations with preterm birth, including dose-response and temporal associations. A high relative abundance of either Gardnerella or Ureaplasma within CST 4 further stratified preterm risk, a finding that was borne out in a small but separate second group of subjects. Most women, regardless of whether they delivered vaginally or by cesarean section, experienced a post-partum alteration of the vaginal microbiota that persisted in some cases for a year or more. This alteration may have clinical consequences, but requires further study.

Supporting Figures

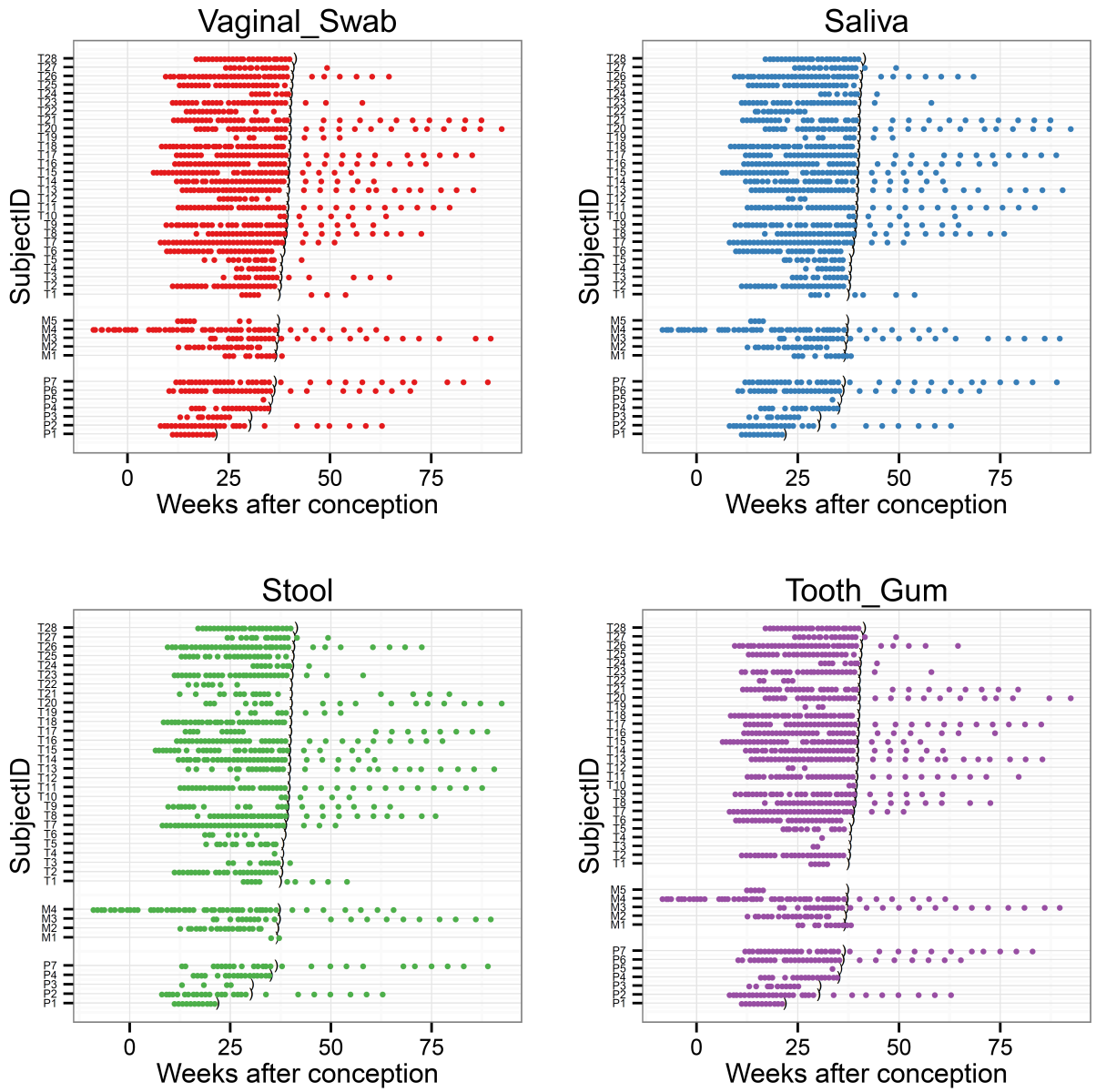

Fig. S1. Sampling time course inclusive of post-partum sampling.

The entire sampling course of the 40 women in the first subject group for each of the four body sites. Twenty-five subjects provided at least one sample post-delivery (delivery indicated by black cross).

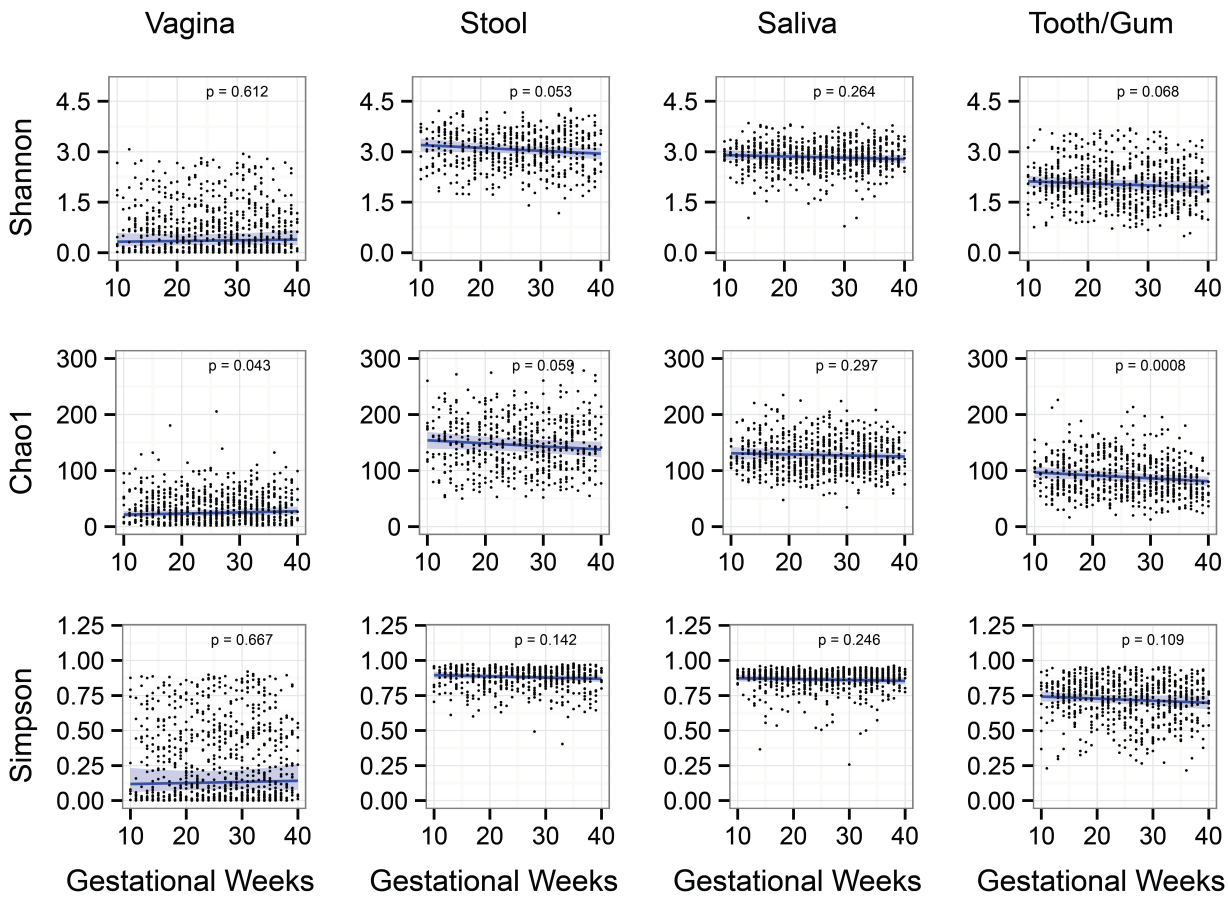

Fig. S2. Alpha diversity of human-associated microbial communities is unaffected by the progression of pregnancy.

Diversity (Shannon: top panels, Chao1: middle panels, Simpson: bottom panels) is plotted against gestational weeks for vaginal, stool, saliva and tooth/gum communities from the 40 women in the first subject group. Blue lines indicate the linear mixed-effects regression of diversity on time with grouping by subject (using the *lme* function in the *nlme* R package). Shading indicates the 95% confidence interval. All trends are insignificant ($p > 0.05$, t-test), except for the Tooth/Gum community when quantified by the Chao1 diversity measure. Taken together, these results indicate that gestational time does not substantially influence the diversity of these microbial communities.

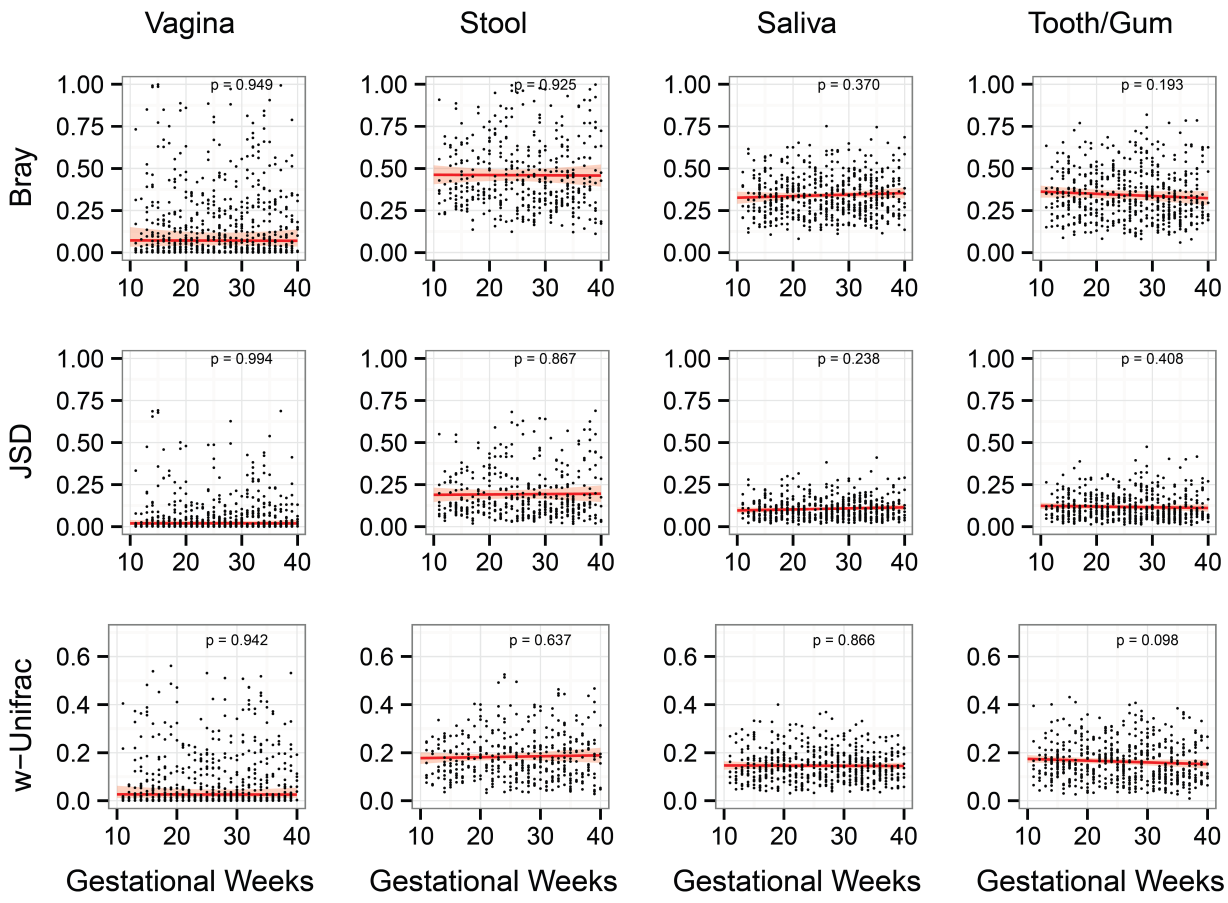

Fig. S3. Week-to-week stability of human-associated microbial communities is unaffected by the progression of pregnancy.

Distance (Bray-Curtis: top panels, Jensen-Shannon: middle panels, weighted UniFrac: bottom panels) between microbial communities from consecutive weeks is plotted against gestational weeks for the vagina, stool, saliva and tooth/gum communities from the 40 women in the first subject group. Red lines indicate the linear mixed-effects regression of diversity on time with grouping by subject (using the *lme* function in the *nlme* R package). Shading indicates the 95% confidence interval. All trends are insignificant ($p > 0.05$, t-test). Taken together, these results indicate that gestational time does not substantially influence the week-to-week stability of these microbial communities.

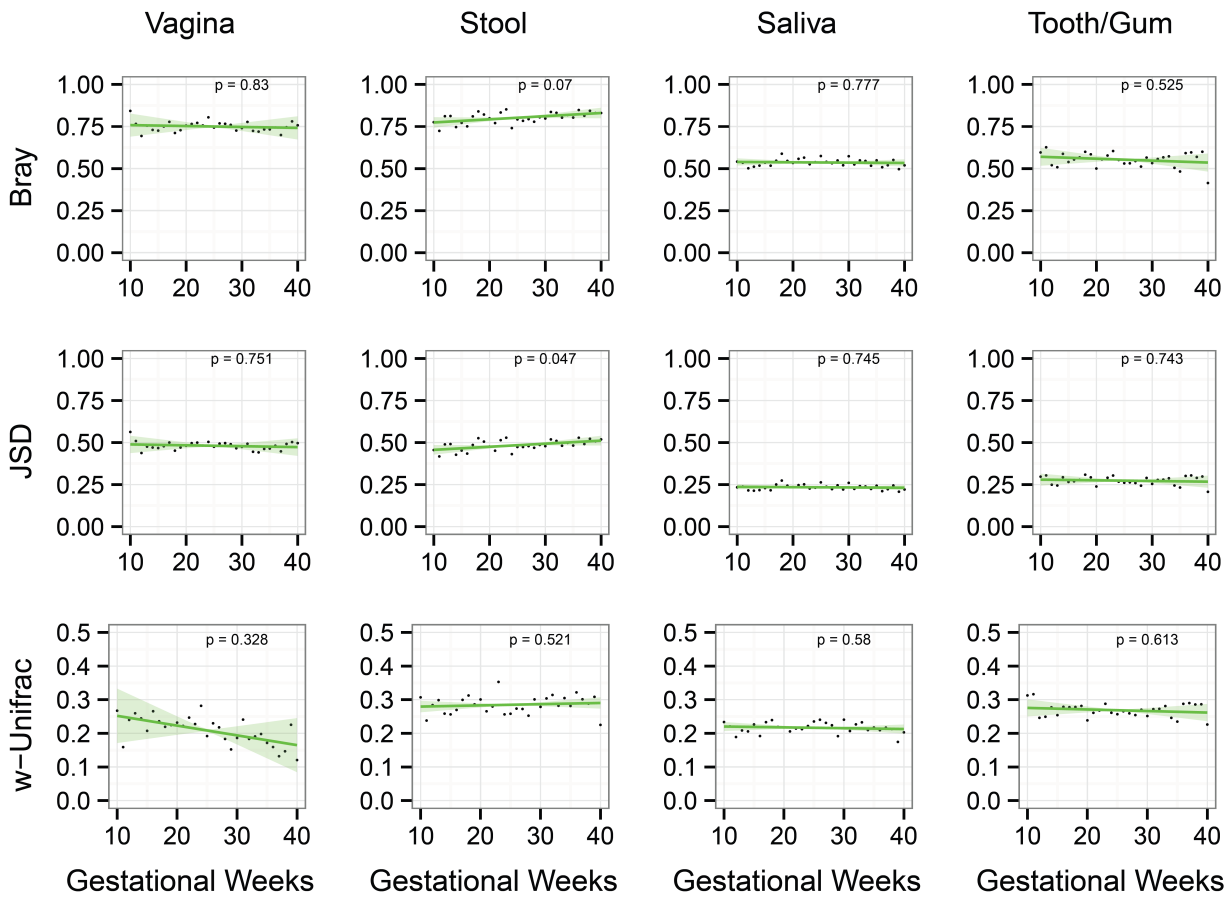

Fig. S4. Beta diversity of human-associated microbial communities is unaffected by the progression of pregnancy.

For the 40 women in the first subject group, beta-diversity was quantified by the average pairwise distance (Bray-Curtis: top panels, Jensen-Shannon: middle panels, weighted UniFrac: bottom panels) between communities of different subjects as a function of gestational weeks. The green line indicates the naive linear fit, while the shaded area indicates the uncertainty estimated from the ensemble of randomly time-reversed subjects (see Methods). All trends are insignificant ($p > 0.05$, permutation bootstrap).

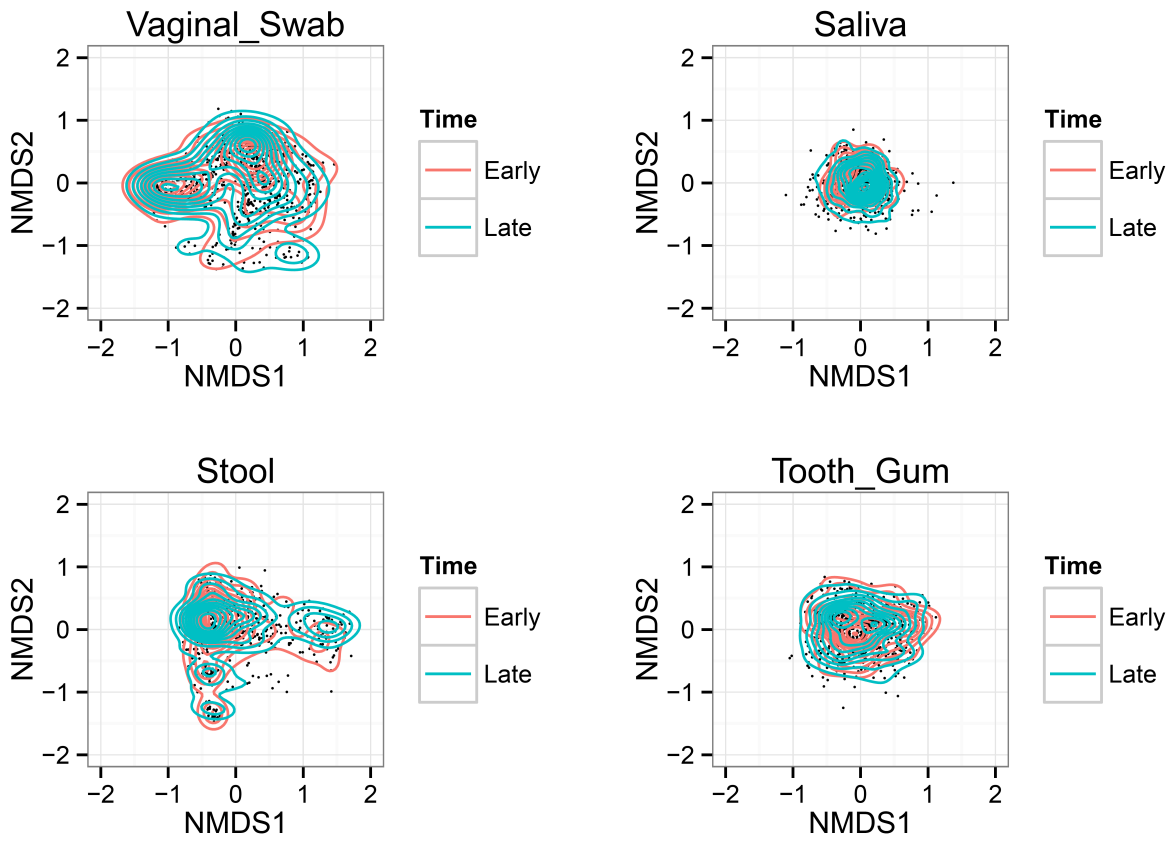

Fig. S5. Taxonomic (OTU) composition of communities from early and late pregnancy are similar.

An NMDS ordination on the Bray-Curtis distance was performed on the communities sampled from each body site for the 40 women belonging to the first subject group. Contour lines are drawn for the communities from early (gestational weeks 14-21) and late (gestational weeks 28-35) sampling time points. These contours strongly overlap, suggesting that the compositions of early and late pregnancy communities do not significantly differ.

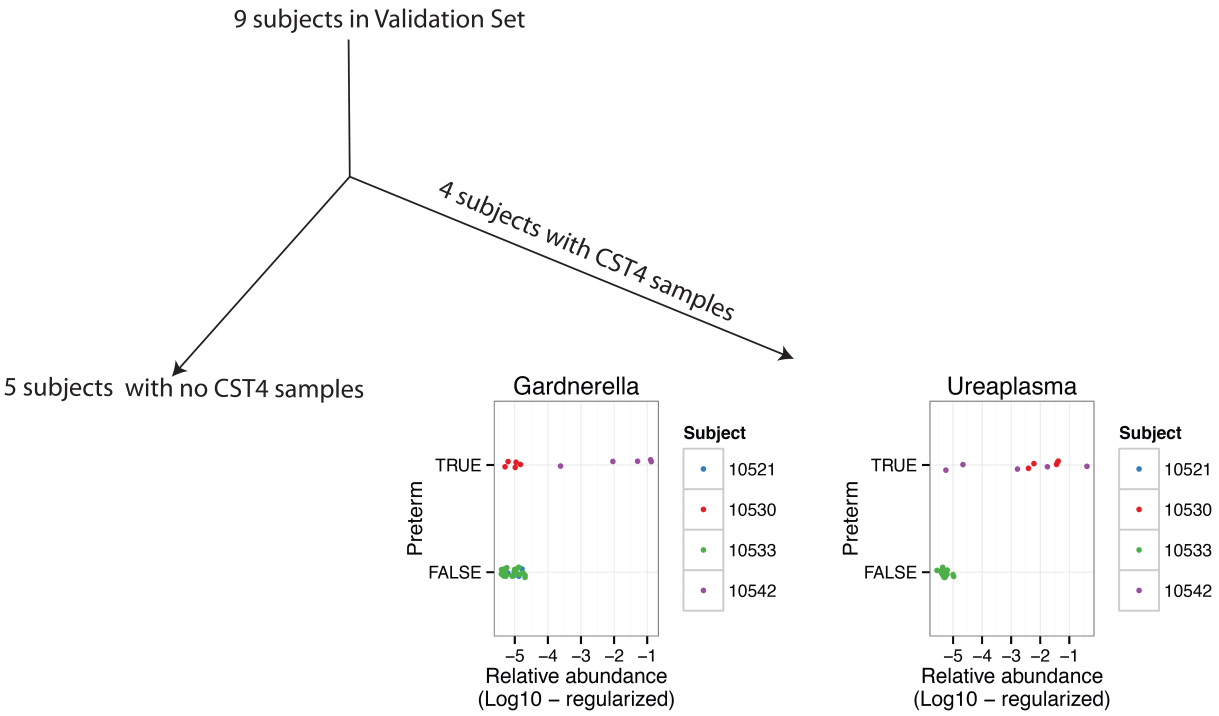

Fig. S6. Assessment of a high-risk vaginal community profile in the second subject group.

A high-risk vaginal community profile identified in the discovery data set, and consisting of a CST 4 vaginal community plus a high abundance of either *Gardnerella* or *Ureaplasma*, was evaluated in the test data set. The test data set consisted of Illumina-based sequences from 246 vaginal samples collected from the second subject group (n=9, four of whom delivered preterm). For this analysis, low *Lactobacillus* relative abundance (<75% *Lactobacillus*) was used as a proxy for CST4. All such “CST 4” samples were evaluated for *Gardnerella* and *Ureaplasma* abundance and plotted by the regularized log of relative abundances. Of the four subjects with CST 4 communities, the two subjects who delivered preterm had either high Ureaplasma abundance (subject 10530), or high Ureaplasma and high Gardnerella abundances (subject 10542). Thus, the combination of low *Lactobacillus* abundance plus high abundance of either *Gardnerella* or *Ureaplasma* was found only in subjects who deliver preterm.

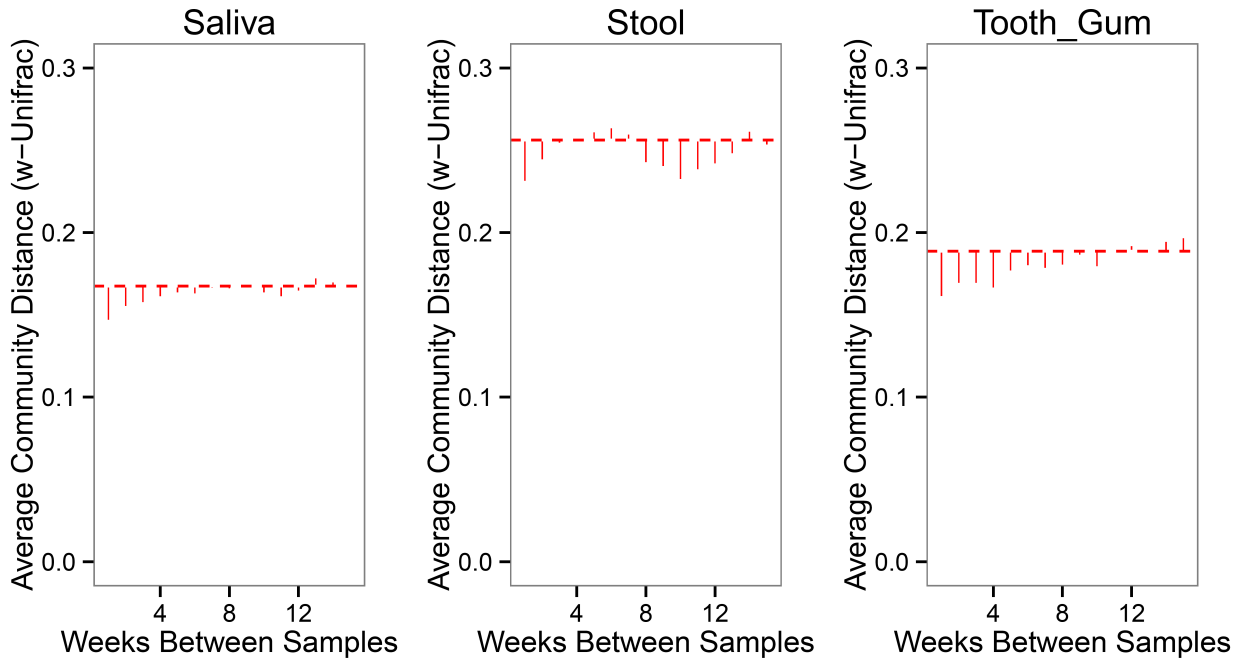

Fig. S7. Average distance between communities as a function of sampling interval.

The dashed line shows the average pair-wise weighted Unifrac distance between same-subject, pre-partum samples separated by more than a month, for the 40 women in first subject group. Bars show the deviation from this long-time average. Communities at all body sites were more similar when nearby in time. This increased similarity decays on roughly a month time-scale for the oral communities, while in the stool community it decays faster.

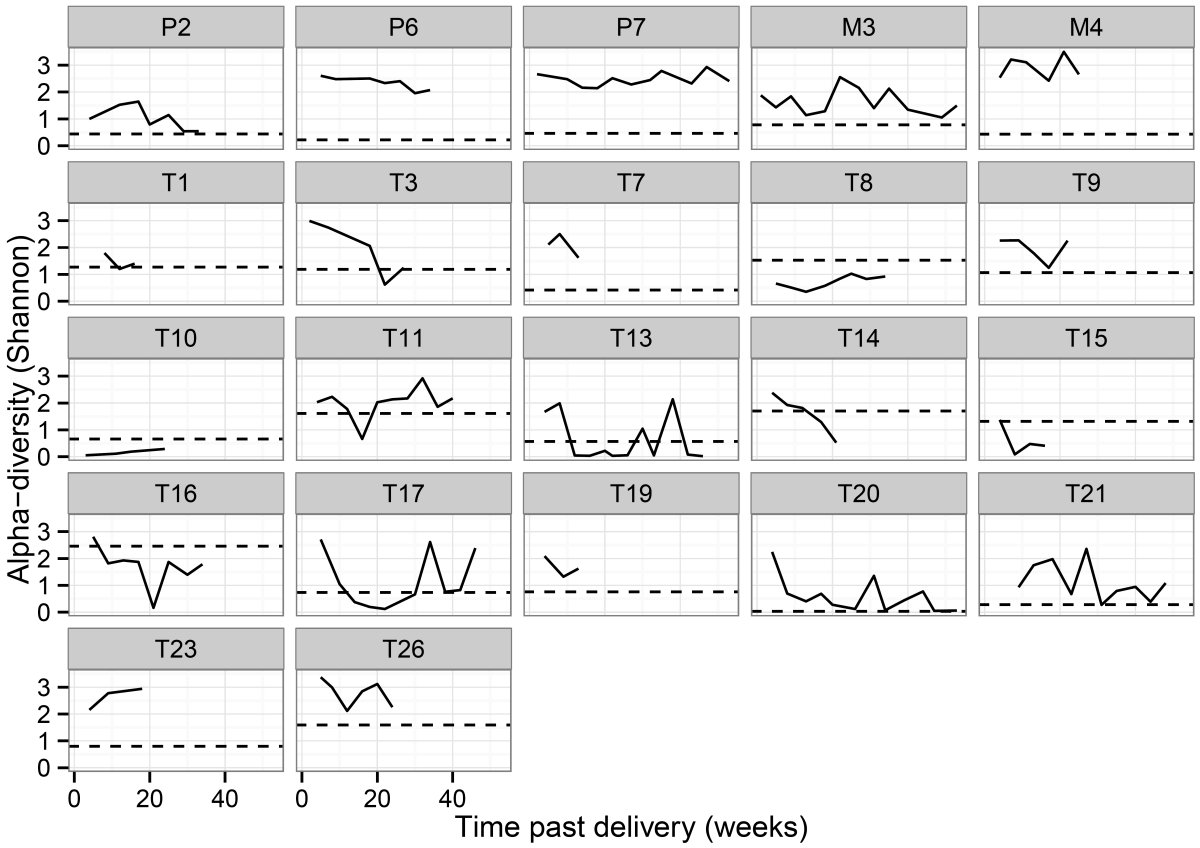

Fig. S8. Patterns of vaginal community alpha diversity after delivery.

Data are shown for each of the 22 subjects from the first subject group, who provided at least one post-delivery sample. Dashed lines indicate the average diversity in the vaginal community of that subject during her pregnancy.

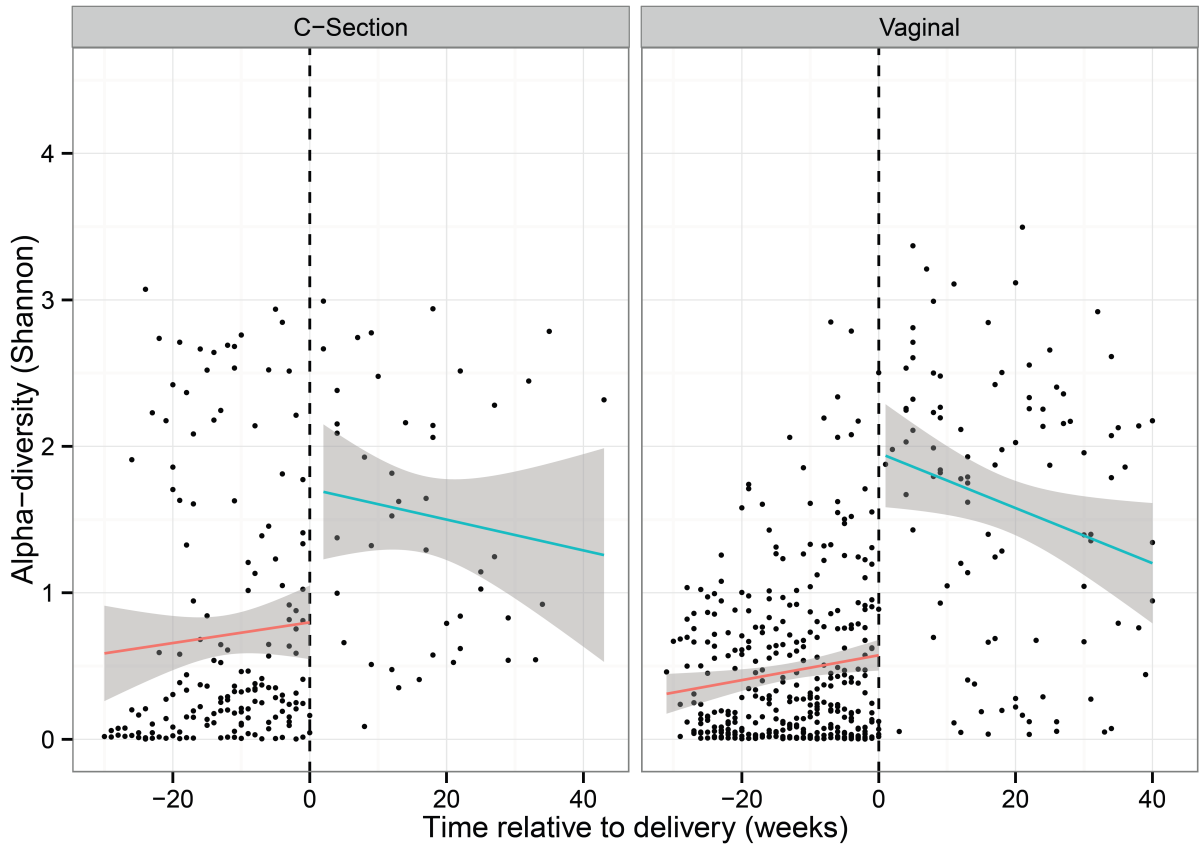

Fig. S9. Alpha diversity of the vaginal communities pre- and post-delivery for cesarean-section and vaginal deliveries.

Alpha diversity, as quantified by the Shannon index, is plotted versus the time relative to delivery at which the sample was collected for women in the first group of subjects (n=40), who provided at least one post-delivery sample. Linear fits are shown for the pre- (red) and post- (green) delivery samples, considered separately. There is a shift to higher diversity both for subjects who delivered via cesarean-section and those who delivered vaginally.

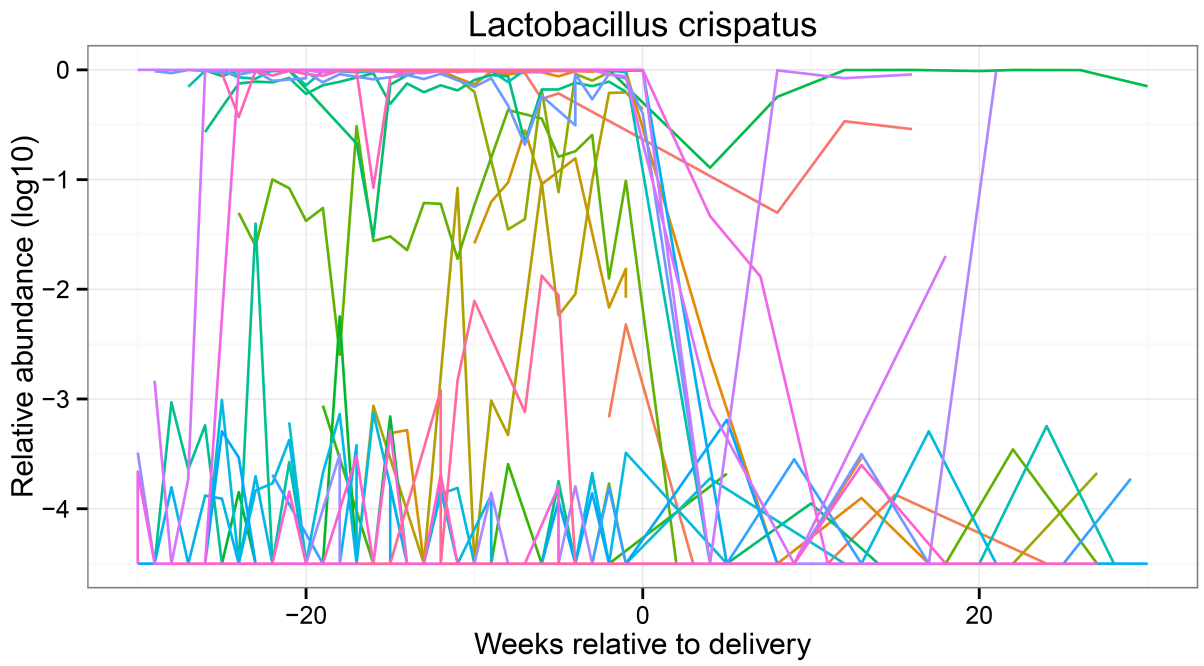

Fig. S10. Dynamics of *Lactobacillus crispatus* in the vaginal communities before and after delivery.

Time courses of the relative abundance of *L. crispatus* in the vaginal communities of the 25 subjects (indicated by color) from the first subject group (n=40), that provided at least one post-partum sample. *L. crispatus* is characteristic of the other taxa (Table S6) that were found to decrease significantly post-delivery. Most, but not all, women with abundant *L. crispatus* prior to delivery had abrupt decreases, which persisted in many cases for months.

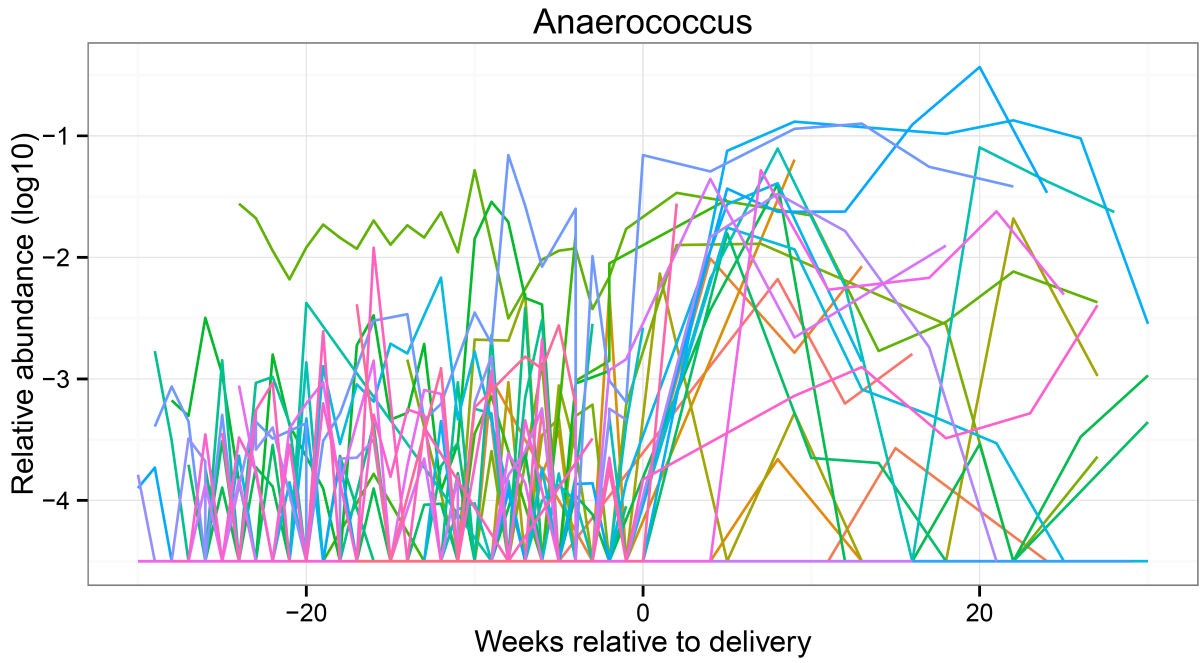

Fig. S11. Dynamics of *Anaerococcus* in the vaginal community before and after delivery.

Time courses of the relative abundance of *Anaerococcus* (GG: 362308) in the vaginal communities of the 25 women (indicated by color) from the first subject group (n=40), that provided at least one post-partum sample. *Anaerococcus* is characteristic of the other taxa (Table S6) that were found to increase significantly post-delivery. Most, but not all, women demonstrated significant increases in the proportion of *Anaerococcus* in their vaginal community from low to moderate relative abundances (0.1%-10%), which persisted for months.

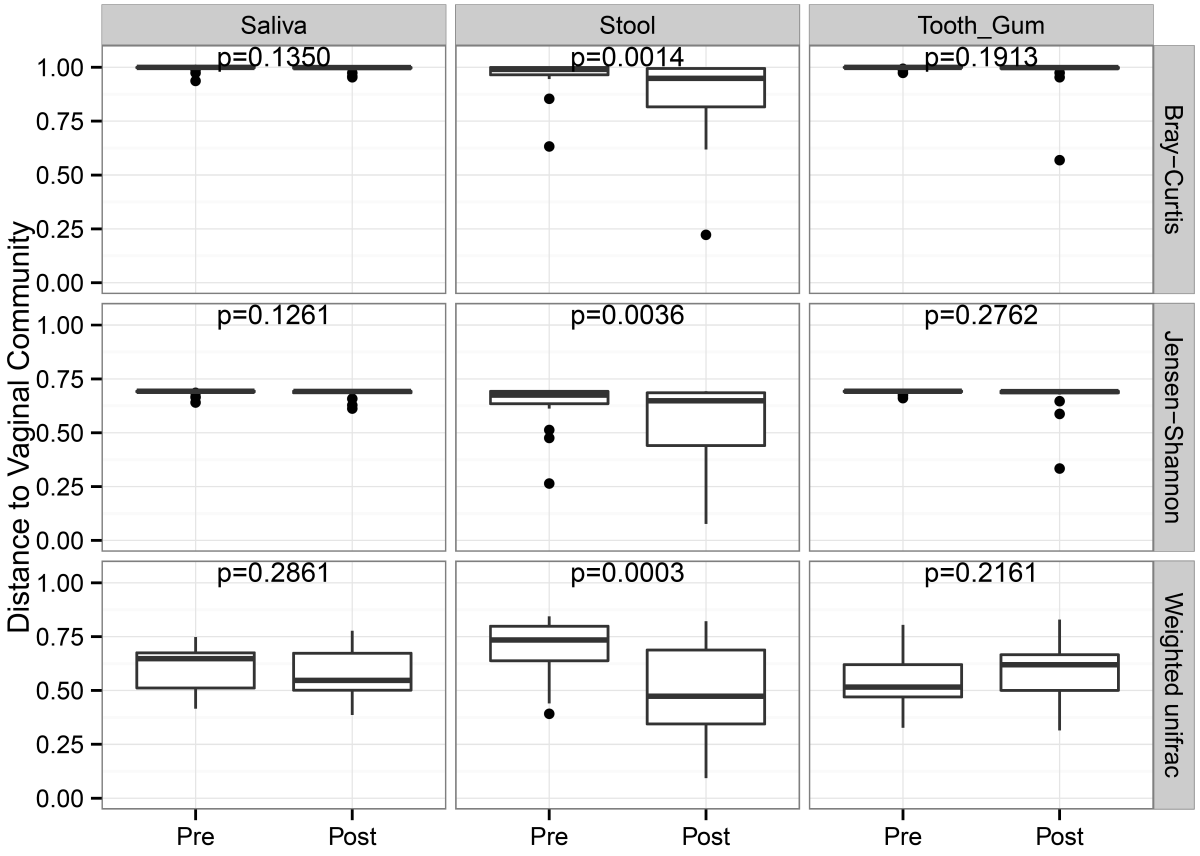

Fig. S12. Differences between the vaginal microbial community and the microbial communities at other body sites, pre- and post- delivery.

The distances between the vaginal community and the communities at the other body sites using the last sample prior to delivery and the first sample post-delivery are compared for the 25 women from the first subject group (n=40), who provided a post-delivery sample using the Bray-Curtis, Jensen-Shannon and weighted-UniFrac distance measures. The vaginal and stool communities became significantly more similar (less distant) after delivery using each distance measure, while there was no significant change in the distance between the vaginal community and the oral communities using any distance measure. P-values are from the Wilcoxon rank-sum test.

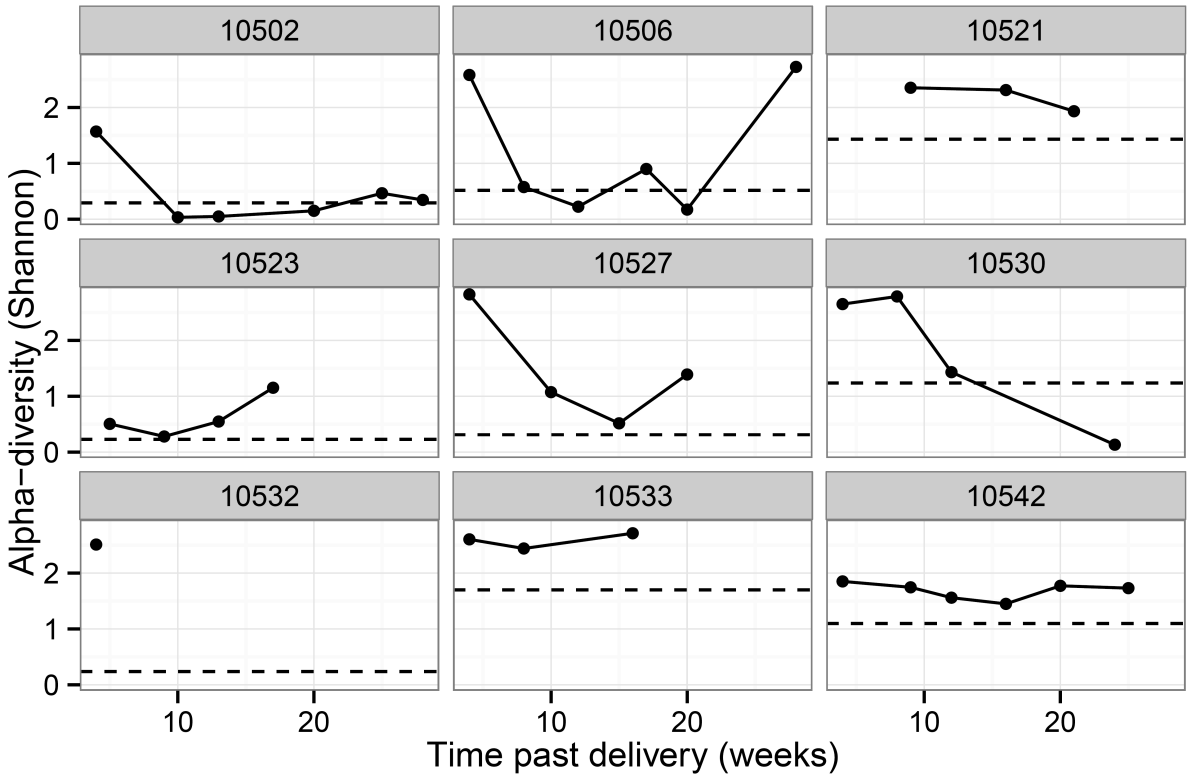

Fig. S13. Post-delivery patterns of vaginal community alpha-diversity in the second subject group (n=9).

Data are shown for all nine women comprising the second group of subjects. Dashed lines indicate the average diversity in the vaginal community of that subject during her pregnancy.

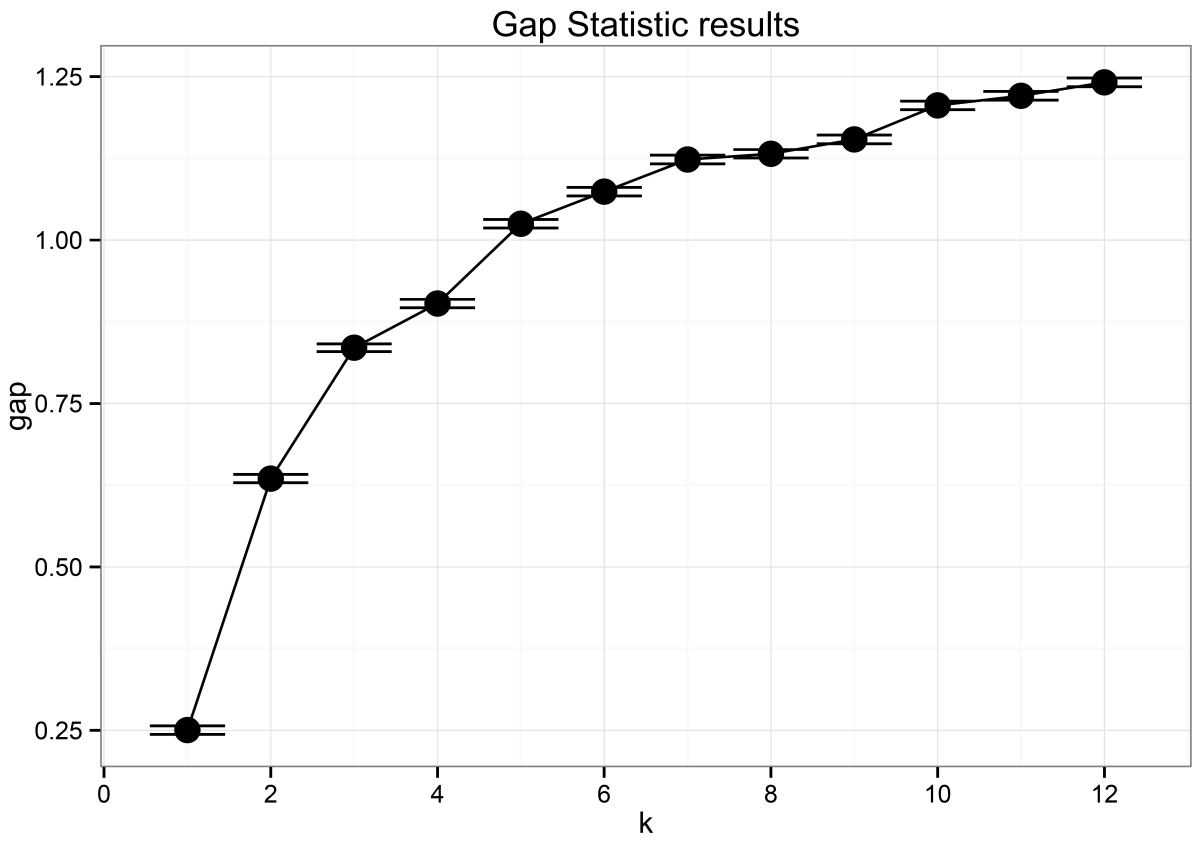

Fig. S14. The gap statistic between partitioning around medoids clustering with different cluster number (k).

The gap statistic was calculated between *pam* clusterings with different cluster numbers (see (19)). Five (clusters) was chosen as the optimal number. Analysis is based on vaginal samples from the 40 women in the first subject group.

Supporting Tables

Table S1. Demographic and clinical characteristics of the total study population (n=49).

Data are shown based on delivery outcome (term vs. preterm) according to whether subjects belonged to the first group (n=40; 11 of whom delivered preterm) or the second group (n=9; five of whom delivered preterm).

Variable	Term delivery		Preterm delivery	
	First Group (n=29)	Second Group (n=5)	First Group (n=11)	Second Group (n=4)
Mean maternal age (range)	30 (19-41)	31.8 (29-37)	28 (19-39)	34 (27-45)
Race:				
White	18 (62%)	4 (80%)	4 (36%)	3 (75%)
Asian	3 (10%)	1 (20%)	1 (9%)	1 (25%)
Pacific Islander	1 (3%)	0	2 (18%)	0
American Indian	1 (3%)	0	2 (18%)	0
Black	1 (3%)	0	1 (9%)	0
Others	5 (17%)	0	1 (9%)	0
Ethnicity:				
Non-hispanic	17 (59%)	4 (80%)	9 (82%)	3 (75%)
Hispanic	11 (38%)	1 (20%)	2 (18%)	1 (25%)
Declined to state	1 (3%)	0	0	0
Nulliparity	4 (14%)	5 (100%)	2 (18%)	2 (50%)
Mean BMI*, kg/m^2 (range)	26.7 (18.2-36.8)	22.5 (20.1-24.6)	31.9 (22-39.7)	27.2 (25.5-29.5)
Gestational Diabetes Mellitus	4 (14%)	1 (20%)	2 (18%)	2 (50%)
Antibiotic use during gestation	8 (28%)	0	4 (36%)	0

*BMI: Body mass index

Table S2. Characteristics of the 15 women who delivered before 37 weeks of gestation.

Subjects are ordered by increasing gestational day at delivery. None were diagnosed with chorioamnionitis.

Preterm Subject	Subject Group	Race	Hispanic Ethnicity	Prior Preterm Delivery	Gest. Weeks at Delivery	sPTB*	pPROM‡
10029	First	Pacific Islander	No	Yes	21 4/7	No	No
10542	Second	Asian (Thai)	No	Yes	23	SROM†	Yes
10031	First	White	Yes	No	29 6/7	No	No
19009	First	Black	No	No	30 1/7	Yes	No
10013	First	White	Yes	Yes	34 6/7	Yes	No
10532	Second	White	No	NA	34 6/7	No	No
10011	First	American Indian	No	Yes	35 2/7	Yes	No
10036	First	White	No	No	35 6/7	SROM†	Yes
10018	First	American Indian	No	Yes	36 1/7	NA	No
19007	First	Pacific Islander	No	Yes	36 2/7	Yes	No
10527	Second	White	No	NA	36 2/7	No	No
10530	Second	White	Yes	Yes	36 2/7	No	No
10055	First	Other	No	Yes	36 3/7	Yes	No
10014	First	Asian-unspecified	No	No	36 6/7	No	No
10101	First	White	No	Yes	36 6/7	SROM†	Yes

*sPTB: spontaneous preterm birth

†SROM: spontaneous rupture of membranes without contractions

‡pPROM: premature preterm rupture of membranes

Table S3. OTUs with significant changes in relative abundance between early and late pregnancy for the 40 women in the first group of subjects.

The relative abundances of all OTUs present in >25% of samples were compared between early and late pregnancy at each body site using DESeq, using a p_{adj}* cutoff of 0.1, and a fold-change threshold of 1.5x (see Methods). Very few (0/27 from vagina, 1/129 from saliva, 0/157 from stool, 2/98 from tooth and gums) OTUs were found to change significantly in abundance, and those that were identified changed to relatively small degrees. The "early" and "late" samples were the earliest and latest samples between gestational weeks 10-20 and 30-40, respectively, from the n=26/23/22/23 subjects with samples from the Vagina/Saliva/Stool/Tooth-Gum in those time ranges. The minimum, median and maximum gestational weeks for the samples in each body-site/time-class are listed: Vagina/Early (10,12.5,19), Vagina/Late (33,39,40), Saliva/Early (10,12,17), Saliva/Late (33,39,40), Stool/Early (10,13,19), Stool/Late (32,38,40), Tooth-Gum/Early (10,12,17), Tooth-Gum/Late (33,39,40).

Body Site	Taxon assigned	Greengenes ID	Fold-change	p_{adj}*
Saliva	*Leptotrichia* sp.	4430826	0.536x	1.31×10^{-4}
Tooth/Gum	*Rothia mucilaginosa*	4294457	1.597x	0.0983
Tooth/Gum	*Haemophilus* sp.	4469359	1.550x	0.0986

* p-value adjusted by the Benjamini-Hochberg method to control for false discovery rate (see (11))

Table S4. One-week transition rates between vaginal community state types (CSTs) for the 40 women in the first group of subjects.

The transition rates are the maximum likelihood estimate from all consecutively sampled pairs of vaginal samples separated by 4-10 days.

	CST1	CST2	CST3	CST4		CST5
CST1	0.979	0	0	0.016		0.004
CST2	0.024	0.976	0	0		0
CST3	0.010	0	0.875	0.082		0.034
CST4	0.084	0	0.179	0.684		0.053
CST5	0	0	0.138	0.031		0.831

Table S5. Association of vaginal CSTs with preterm birth in a group of 40 women.

The number of times each CST was sampled from each class of pregnancy is shown. Of the 40 subjects, 28 delivered at term (>37 gestational weeks), 7 delivered preterm (<37 gestational weeks) and 3 delivered very preterm (<32 gestational weeks). The 5 subjects (who delivered during the 37th gestational week, and whose deliveries were classified as 'marginal') were excluded from the comparison of women with term and preterm outcomes. The very-preterm counts in brackets are a subset of the preterm counts.

CST	Term	Preterm	Very Preterm	Marginal
1	232 (82%)	28 (9.9%)	[0]	23 (8.1%)
2	36 (75%)	0	[0]	12 (25%)
3	201 (80%)	10 (4%)	[10 (4%)]	39 (16%)
4	35 (32%)	68 (62%)	[33 (30%)]	7 (6%)
5	67 (96%)	1 (1.4%)	[0]	2 (2.8%)

Table S6. Vaginal OTUs with a significant shift in relative abundance at delivery.

The shifts between the last pre-delivery and first post-delivery vaginal samples (mean 37 days, range 14-92) in the 25 subjects from the first group of 40 women, who provided at least one post-delivery sample, were evaluated for all OTUs present in ≥25% of samples. *Lactobacillus* OTUs decreased significantly after delivery, while a mix of primarily anaerobic bacteria increased significantly in relative abundance.

Delivery Shift	Taxon	GreenGenes ID	Fold-Change	p_{adj}*
Increase	*Prevotella*	760967	6.4x	4.6×10^{-9}
Increase	*Anaerococcus*	362308	5.0x	4.6×10^{-9}
Increase	*Dialister*	753638	4.2x	3.0×10^{-8}
Increase	*Peptoniphilus*	654307	4.1x	1.3×10^{-7}
Increase	*Peptoniphilus*	4429335	2.5x	0.0013
Increase	*Prevotella*	1077373	2.6x	0.0022
Increase	*Porphyromonas*	495017	2.9x	0.0024
Increase	*Prevotella*	1066814	2.4x	0.0073
Increase	Order *Clostridiales*	259604	2.4x	0.0073
Increase	*Anaerococcus*	4476950	2.3x	0.0083
Increase	*Campylobacter*	253584	2.1x	0.020
Increase	*Mobiluncus*	104135	2.1x	0.022
Increase	*Anaerococcus*	1097113	1.8x	0.030
Increase	Order *Clostridiales*	755148	1.9x	0.046
Increase	*Peptoniphilus*	521795	1.8x	0.046
Decrease	*Lactobacillus crispatus*	4447432	0.0385	1.7×10^{-15}
Decrease	*L. gasseri*	4428313	0.1353	9.5×10^{-7}
Decrease	*L. reuteri-vaginalis*	137043	0.2273	9.5×10^{-7}
Decrease	*L. jensenii*	31171	0.1637	1.5×10^{-6}
Decrease	*L. iners*	129798	0.1565	4.4×10^{-6}
Decrease	*Actinomyces*	12564	0.3409	4.0×10^{-4}
Decrease	*Ureaplasma*	15806	0.2927	4.0×10^{-4}
Decrease	*Staphylococcus aureus*	4446058	0.4036	1.8×10^{-3}
Decrease	*Corynebacterium*	1000986	0.4559	4.0×10^{-3}

* p-value adjusted by the Benjamini-Hochberg method to control for false discovery rate (see (11))

Supporting References

1. Hamady M, Walker JJ, Harris JK, Gold NJ, & Knight R (2008) Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nature methods* 5(3):235-237.
2. Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature methods* 10(10):996-998.
3. Edgar RC, Haas BJ, Clemente JC, Quince C, & Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics (Oxford, England)* 27(16):2194-2200.
4. DeSantis TZ*, et al.* (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and environmental microbiology* 72(7):5069-5072.
5. Caporaso JG*, et al.* (2010) PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics (Oxford, England)* 26(2):266-267.
6. Wang Q, Garrity GM, Tiedje JM, & Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and environmental microbiology* 73(16):5261-5267.
7. McMurdie PJ & Holmes S (2013) phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PloS one* 8(4):e61217.
8. Wickham H (2009) *ggplot2: elegant graphics for data analysis* (Springer New York) p 213.
9. Maechler M, Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K. (2015) cluster: Cluster Analysis Basics and Extensions.R package version 2.0.0.
10. Pinheiro J BD, DebRoy S, Sarkar D and R Core Team (2015) nlme: Linear and Nonlinear Mixed Effects Models.R package version 3.1-119.
11. Love MI, Huber W, & Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* 15(12):550.
12. Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27:379-423 and 623-656.
13. Chao A (1984) Non-parametric estimation of the number of classes in a population. *Scand J Stat* 11:265-270.
14. Simpson EH (1949) Measurement of Diversity. *Nature* 163:688-688.
15. Bray JR & Curtis JT (1957) An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs* 27:326-349.

16. Lin J (1991) Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory* 37(1):145-151.
17. McMurdie PJ & Holmes S (2014) Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS computational biology* 10(4):e1003531.
18. Bourgon R, Gentleman R, & Huber W (2010) Independent filtering increases detection power for high-throughput experiments. *Proceedings of the National Academy of Sciences of the United States of America* 107(21):9546-9551.
19. Tibshirani R, Walter G, & Hastie T (2001) Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society (B)* 63:411-423.
20. Ravel J*, et al.* (2011) Vaginal microbiome of reproductive-age women. *Proceedings of the National Academy of Sciences of the United States of America* 108 Suppl 1:4680-4687.
21. Romero R*, et al.* (2014) The composition and stability of the vaginal microbiota of normal pregnant women is different from that of non-pregnant women. *Microbiome* 2(1):4.
22. Gajer P*, et al.* (2012) Temporal dynamics of the human vaginal microbiota. *Science translational medicine* 4(132):132ra152.
23. Romero R*, et al.* (2014) The vaginal microbiota of pregnant women who subsequently have spontaneous preterm labor and delivery and those with a normal delivery at term. *Microbiome* 2:18.
24. Frank JA*, et al.* (2008) Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. *Applied and environmental microbiology* 74(8):2461-2470.
25. Hillier SL*, et al.* (1995) Association between bacterial vaginosis and preterm delivery of a low-birth-weight infant. The Vaginal Infections and Prematurity Study Group. *The New England journal of medicine* 333(26):1737-1742.
26. Leitich H*, et al.* (2003) Bacterial vaginosis as a risk factor for preterm delivery: a meta-analysis. *American journal of obstetrics and gynecology* 189(1):139-147.
27. DiGiulio DB*, et al.* (2008) Microbial prevalence, diversity and abundance in amniotic fluid during preterm labor: a molecular and culture-based investigation. *PloS one* 3(8):e3056.
28. DiGiulio DB*, et al.* (2010) Prevalence and diversity of microbes in the amniotic fluid, the fetal inflammatory response, and pregnancy outcome in women with preterm pre-labor rupture of membranes. *American journal of reproductive immunology (New York, N.Y. : 1989)* 64(1):38-57.

29. Gardella C*, et al.* (2004) Identification and sequencing of bacterial rDNAs in culture-negative amniotic fluid from women in premature labor. *American journal of perinatology* 21(6):319-323.
30. Han YW, Shen T, Chung P, Buhimschi IA, & Buhimschi CS (2009) Uncultivated bacteria as etiologic agents of intra-amniotic inflammation leading to preterm birth. *Journal of clinical microbiology* 47(1):38-47.