Supplemental Figures:

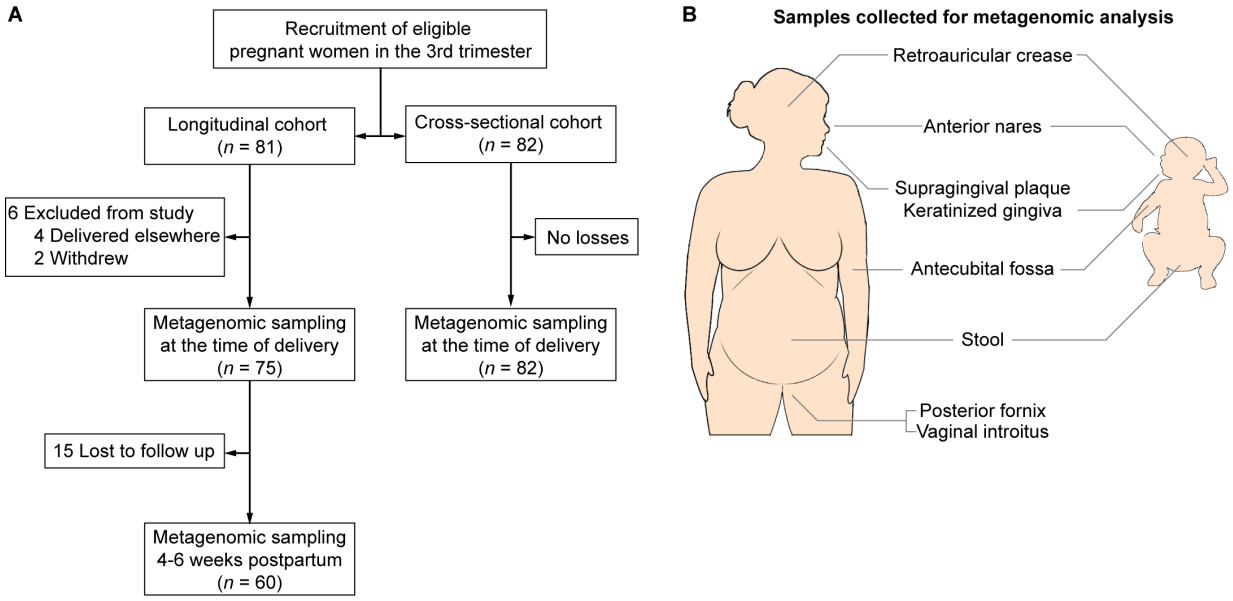

Fig. S1. Study Design Overview

(A) Flowchart of prospective cohort study detailing enrollment and incurred study losses. (B) Samples collected from mother and infant at each sampling time point.

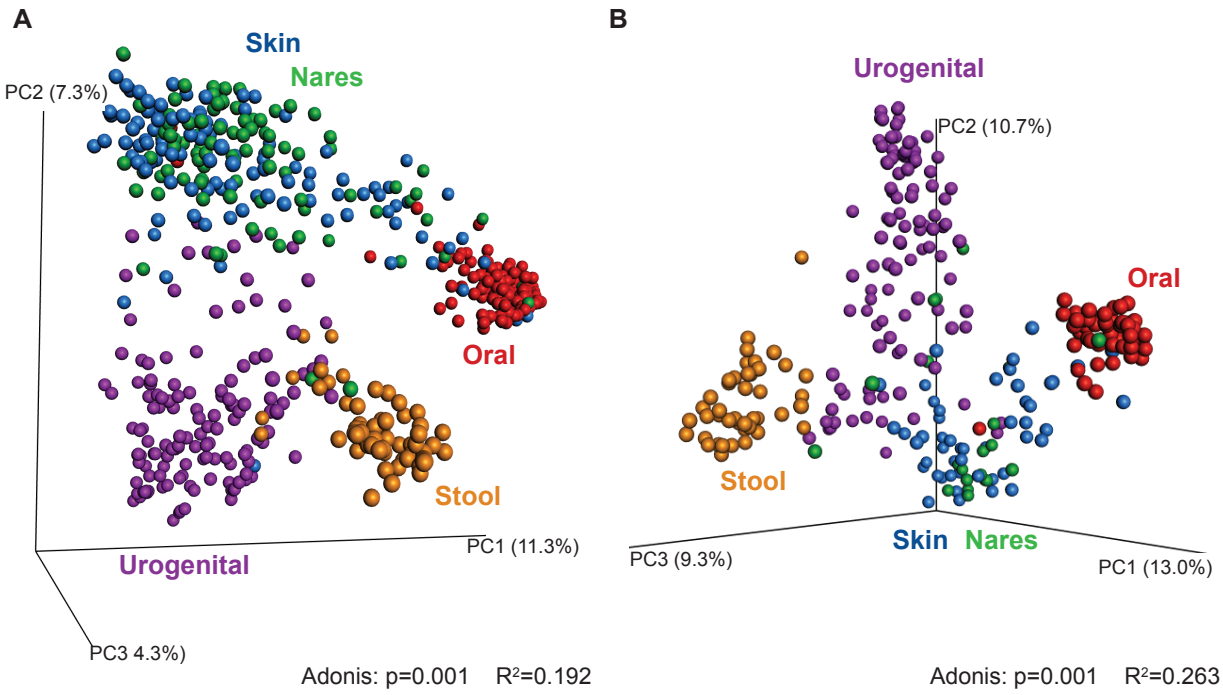

Fig. S2. Observable Body Site Specificity of the Maternal Microbiota

Principal Coordinate Analysis (PCoA) on unweighted UniFrac distances between maternal samples obtained at (A) delivery and at (B) 6 weeks postpartum. The percentage of the total variation explained by the plotted principal coordinate (PC) is indicated on the axis. Each point represents a single sample and its color indicates which body site from which it was sampled: Skin, blue; Nares, green; Oral cavity, red; Urogenital tract (vagina) purple; Stool, oral). The significance of clustering by virtue of body site was determined by Adonis with a $p<0.05$ considered significant.

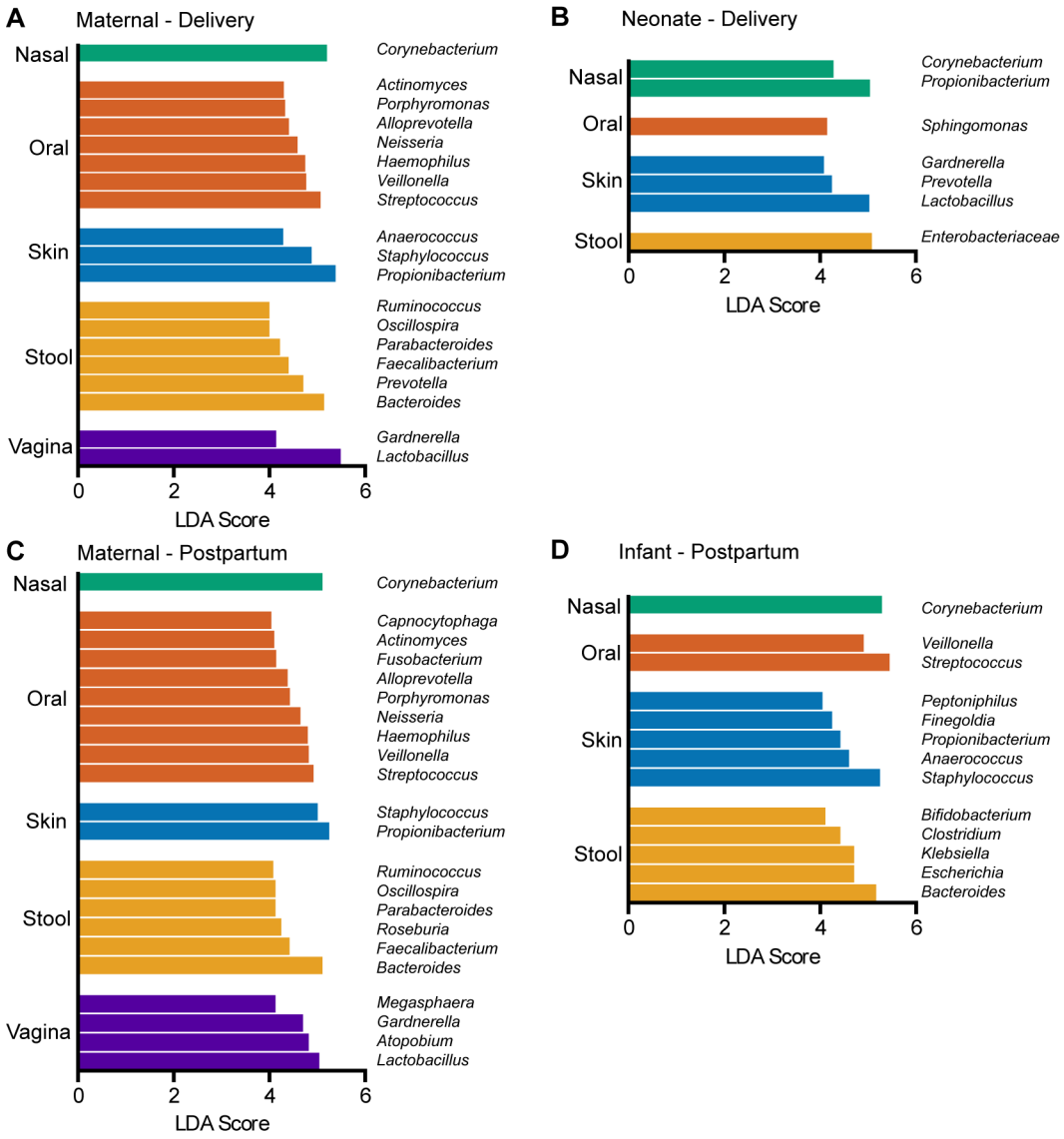

Fig. S3. Characteristics Taxa of Neonatal and Maternal Microbiota at Delivery and 6 Weeks determined by Linear Discriminant Analysis Effect Size (LEfSe)

Representative taxa for each body site (Linear Discriminant Analysis (LDA) Score >4, $p<0.05$) are shown. (A) Maternal samples taken at delivery. (B) Neonatal samples taken at delivery. (C) Maternal samples taken at 6 weeks postpartum (D) Infant samples taken at 6 weeks postpartum.

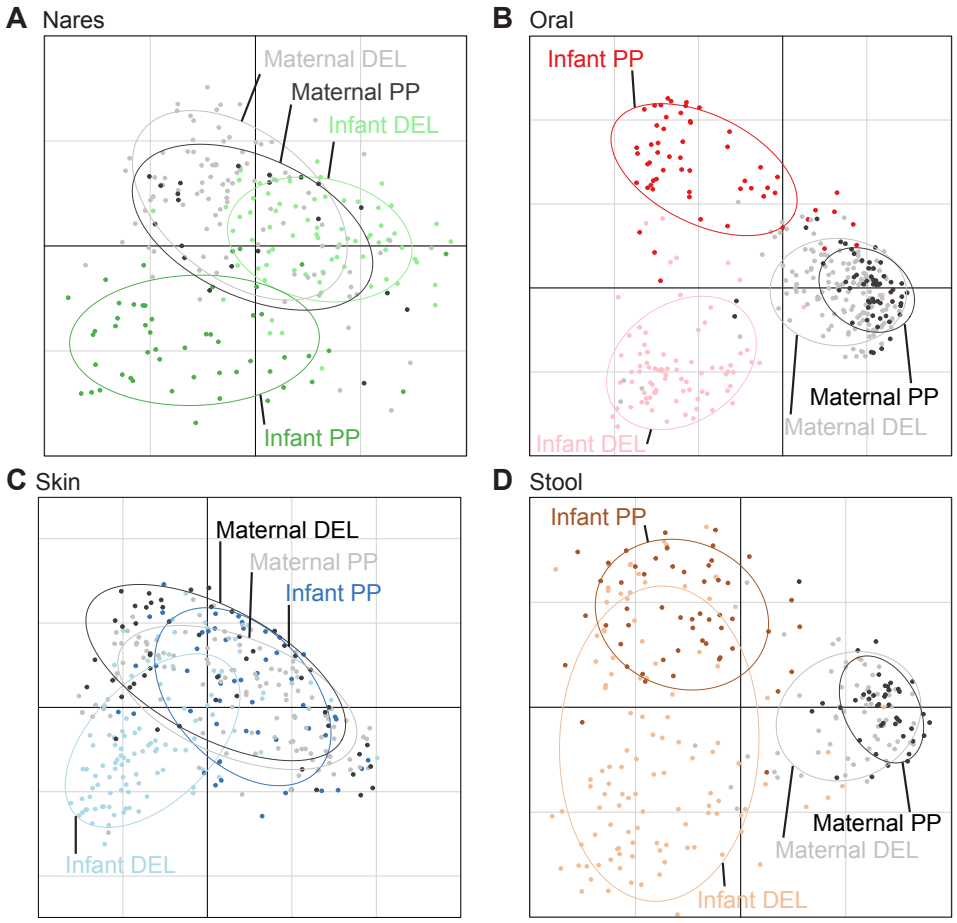

Fig. S4. Comparisons of the Neonatal and Maternal Microbiota Within Body Sites

PCoA on unweighted UniFrac distances shown for the nares (A), oral cavity (B), skin (C) and stool (D). Samples are differentially colored by age (maternal, gray; infant, corresponding color) and by time point (delivery (DEL), lighter shade; 6 weeks postnatal age (PP), darker shade). Ellipses represent 95% confidence intervals around the cluster centroid.

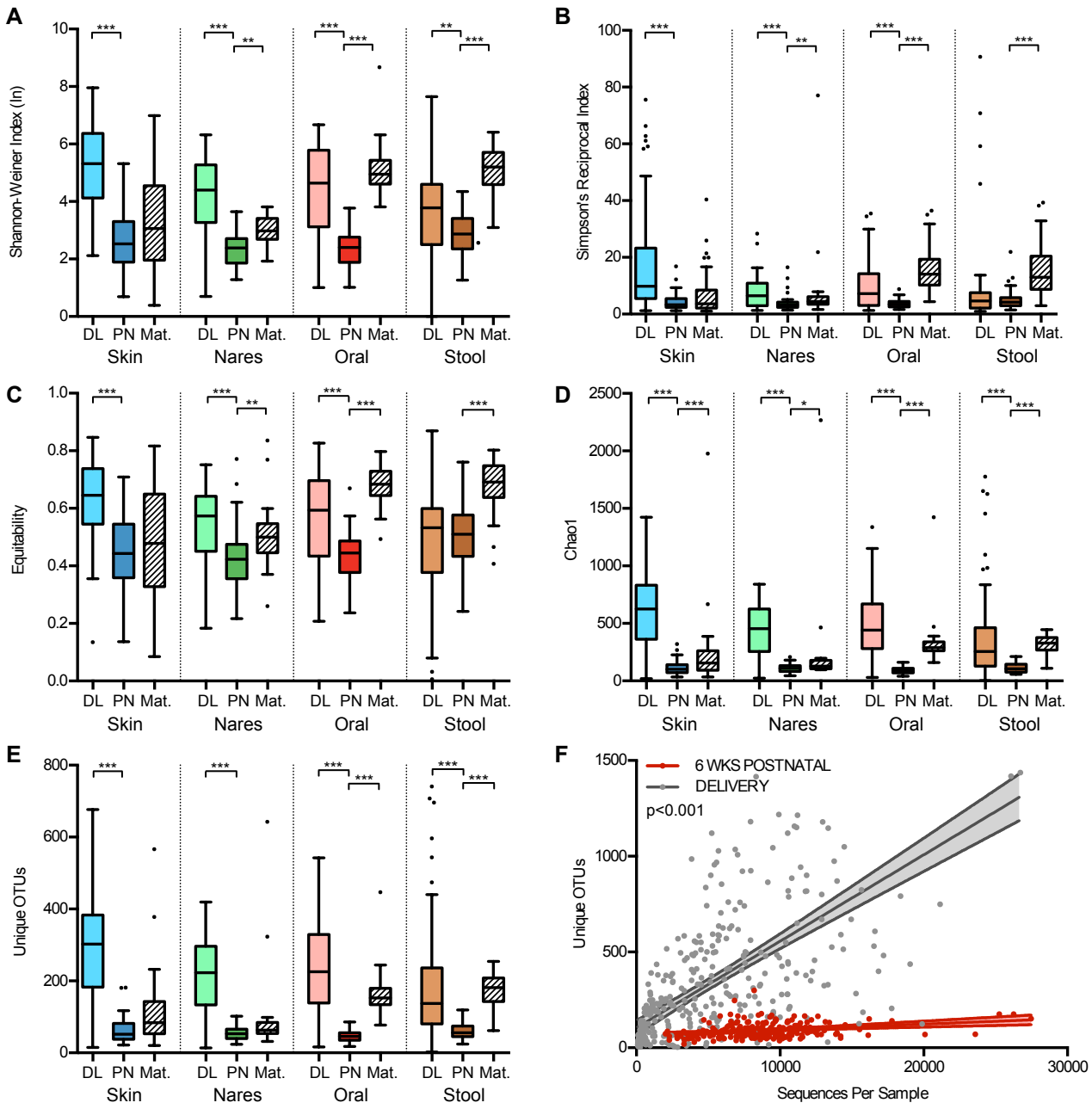

Fig. S5. Measurements of Alpha Diversity Across Body Site and Time

(A) Shannon-Weiner Index (B) Simpson's Reciprocal Index (C) Shannon's Equitability (D) Chao1 (E) Number of Unique OTUs per sample. Significant differences between groups determined by a Mann-Whitney U test (***p< 0.001, **p<0.01, *p<0.05). Box plots represent median and interquartile range with whiskers determined by Tukey's method. (F) Unique OTUs as a function of sequencing depth in infant samples is plotted with points colored by time point (Delivery, gray; 6 weeks postnatal age, red). A linear regression line with 95% confidence interval for each time point is plotted to demonstrate the linear trend. The significance of the differences between the regression slopes was determined by a Student's t-test.

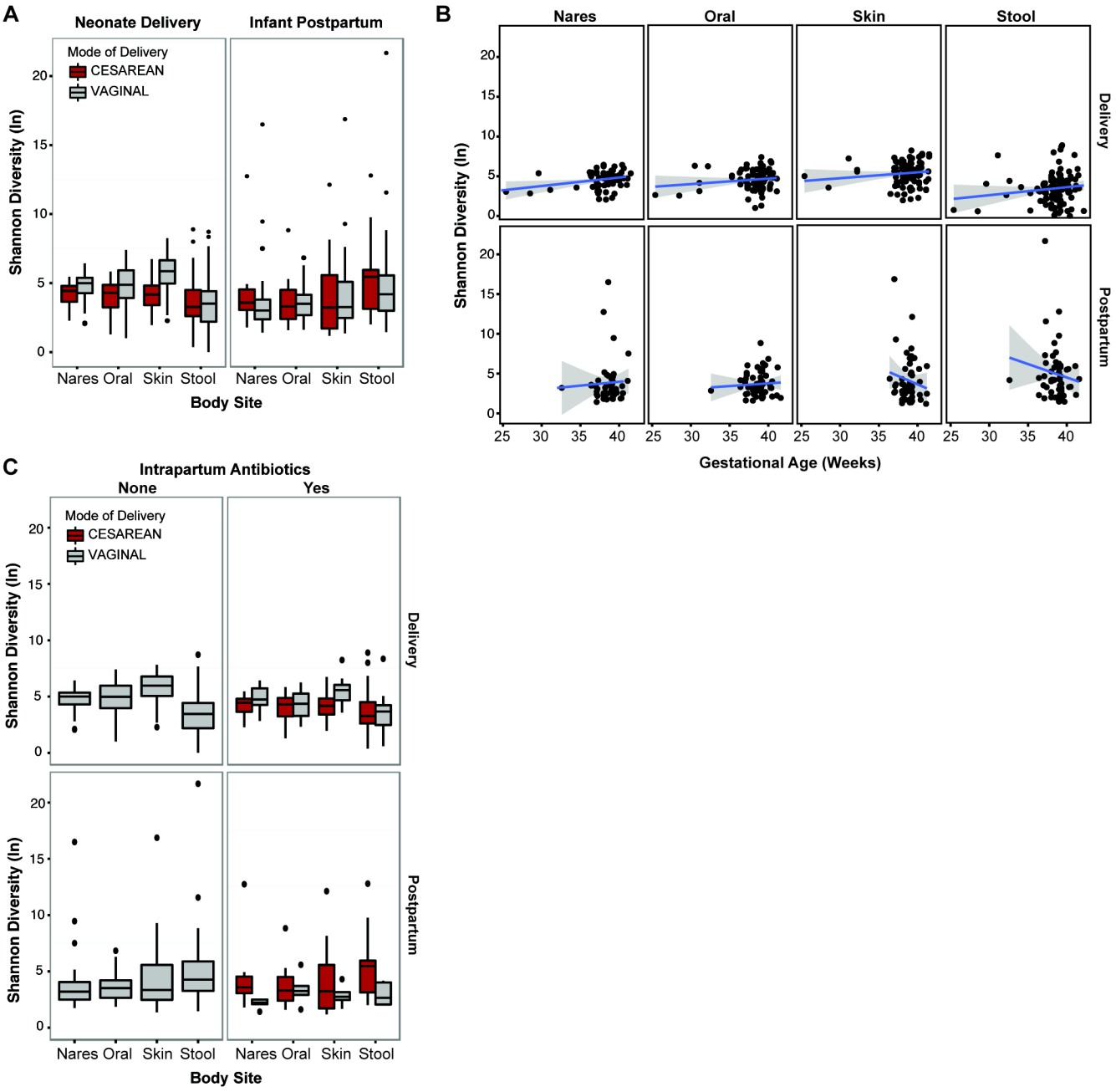

Fig. S6. Alpha Diversity Comparisons by Delivery Mode and Gestational Age (A) Shannon-Weiner index for the neonatal (delivery, left) and infant (6 weeks postpartum, right) microbiome is stratified by body site and by mode of delivery (Vaginal delivery (VB), gray; Cesarean delivery (CS), red). Significance was determined by Student's t-tests (***p< 0.001). (B) Correlations of alpha diversity measurements with gestational age, stratified by time point and body site (all regression coefficients, p > 0.05). (C) Alpha diversity measurements, stratified by intrapartum antibiotic use and mode of delivery. All neonates delivered by Cesarean received a single dose of antibiotics at the time of delivery. All box plots represent median and interquartile range with whiskers determined by Tukey's method.

Fig. S7. **Neonatal Microbiota at Delivery Classified at the Genus Level**

Heatmap indicating the relative abundance of each genera (row) found throughout the neonatal microbiome with each sample (column). Dendrograms indicate hierarchical complete linkage clustering on Bray-Curtis dissimilarity distances. Horizontal color bars indicate the body site (Skin, blue; Oral cavity, red; Nares, green; Meconium, orange) and the mode of delivery (Cesarean, red; Vaginal, grey) for each sample.

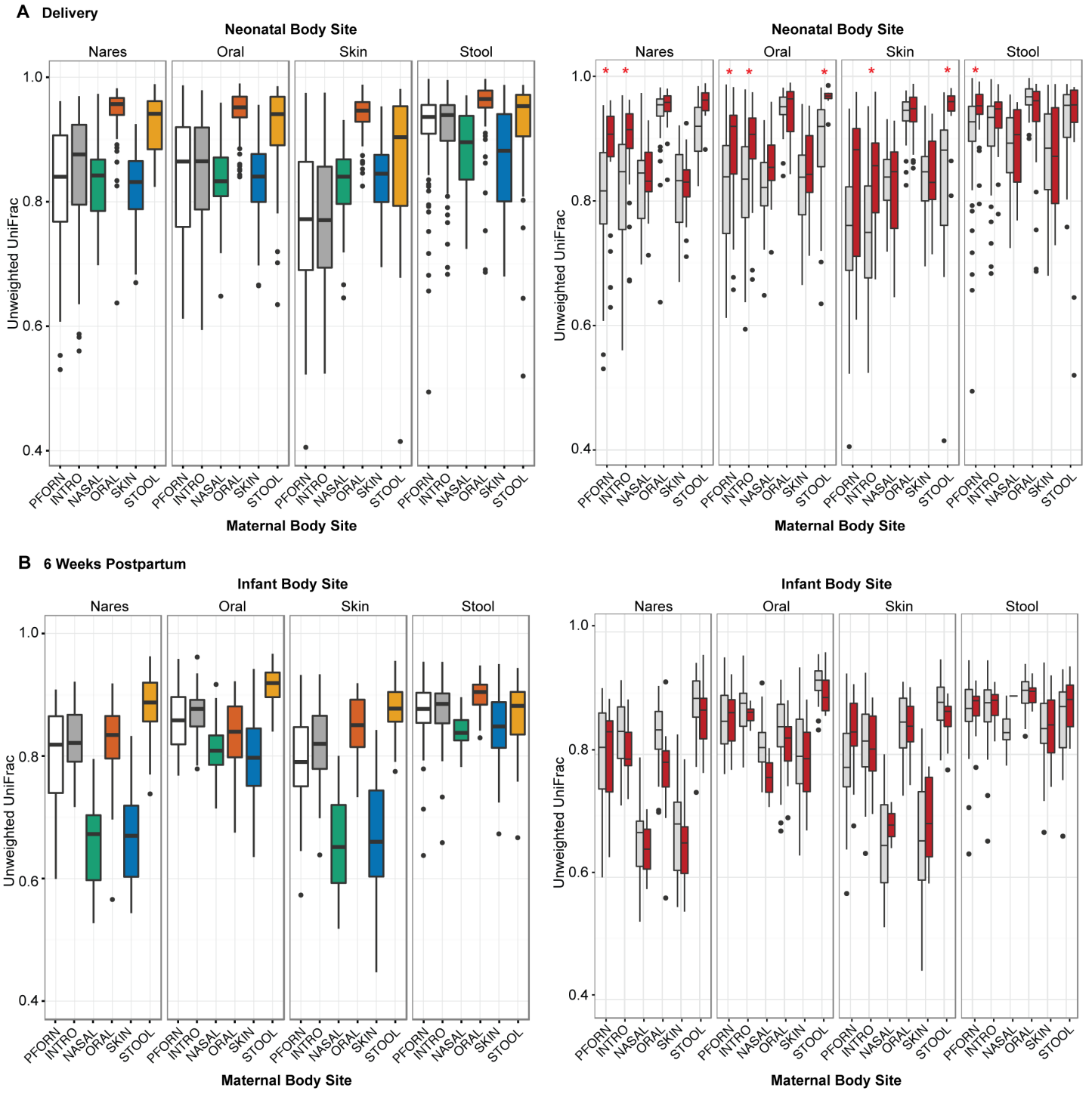

Fig. S8. Similarity of Microbiota between Neonate-Maternal Dyads Unweighted UniFrac distances between samples obtained from the same neonatal-mother dyad. Data are shown for samples obtained at delivery (A) and at 6 weeks (B). (*Left panels)* Distances are stratified by infant and maternal body site (x-axes) as indicated. (*Right panels*) Similar data as shown, but further stratified by mode of delivery (gray, vaginal delivery; red, Cesarean delivery). Box plots indicate median and interquartile range with whiskers determined by Tukey's method. Significance determined by Student's t-test, with Holm's correction for multiple hypothesis testing (*p<0.05). (PFORN: Vaginal posterior fornix, INTRO: Vaginal Introitus).

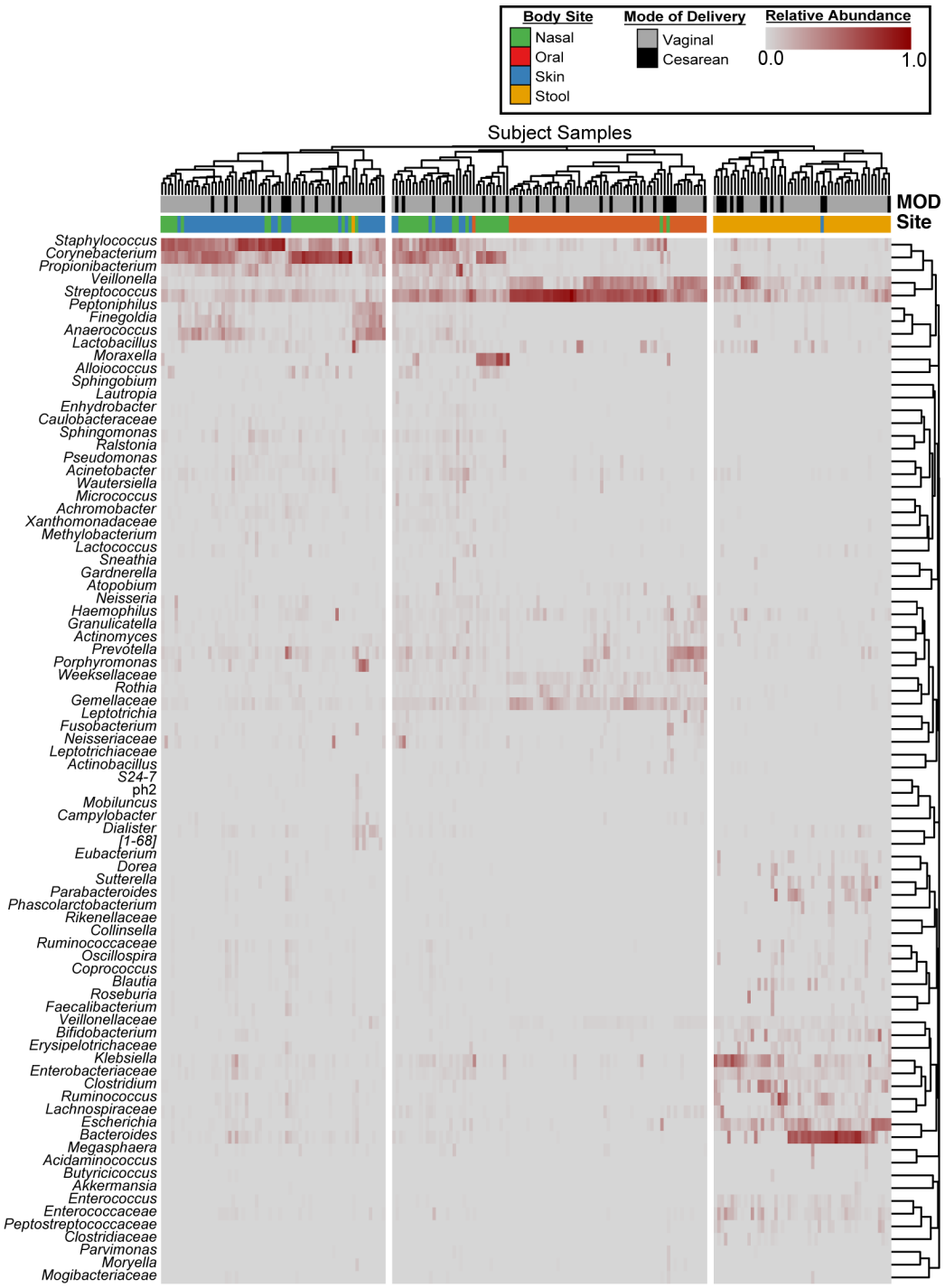

Fig. S9. Infant Microbiota at 6 Weeks Classified at the Genus Level

Heatmap indicating the relative abundance of each genera (row) found throughout the infant microbiome with each sample (column). Dendrograms indicate hierarchical complete linkage clustering on Bray-Curtis dissimilarity distances. Horizontal color bars indicate the body site (Skin, blue; Oral cavity, red; Nares, green; Meconium, orange) and the mode of delivery (Cesarean, red; Vaginal, grey) for each sample.

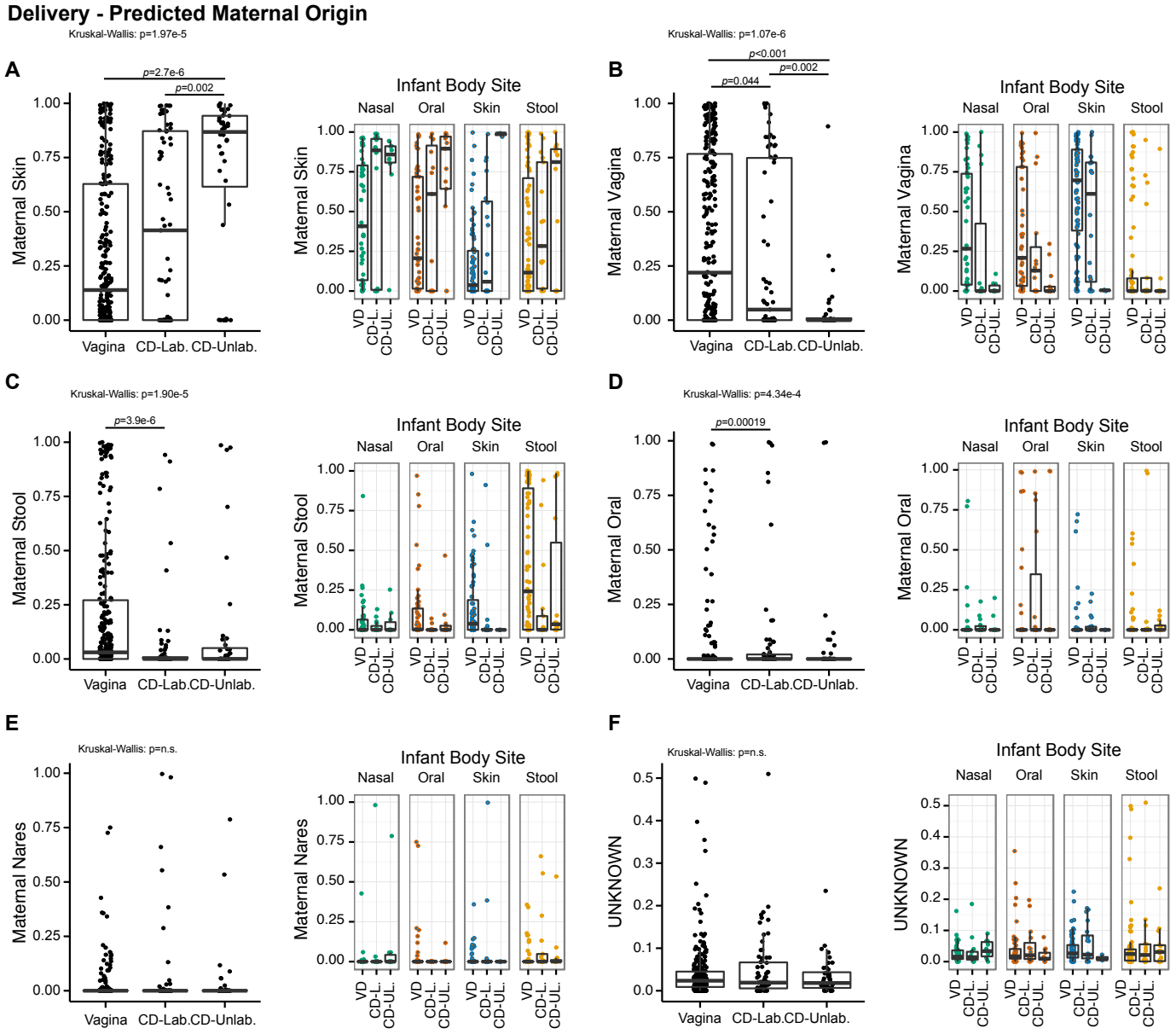

Fig. S10. Projected Taxa Similarities Among Maternal and Neonatal Microbial Communities at Delivery

The predicted proportion of OTUs within the neonatal microbiome (delivery) predicted to originate from the maternal microbiota, stratified by maternal body site (y-axes; (A) skin, (B) vagina, (C) stool, (D) oral, (E) nares and (F) unknown) and by mode of delivery (vagina, vaginal delivery; CD-Lab, Cesarean Labored; CD-Unlab., Cesarean Unlabored). Each point represents an individual sample, indicating the proportion of OTUs predicted to originate from the given maternal body site (y-axis). Boxplots represent the median and interquartile range with whiskers determined by Tukey's method. (*Left Panels)* Data plotted for all neonatal sites, stratified by mode of delivery. Significance between groups was determined by Kruskal-Wallis tests with post-hoc Dunn's non-parametric comparisons. (*Right panels)* Data stratified by neonatal body site (Nasal, green; Oral, red; Skin, blue; Stool, orange) and again by mode of delivery.

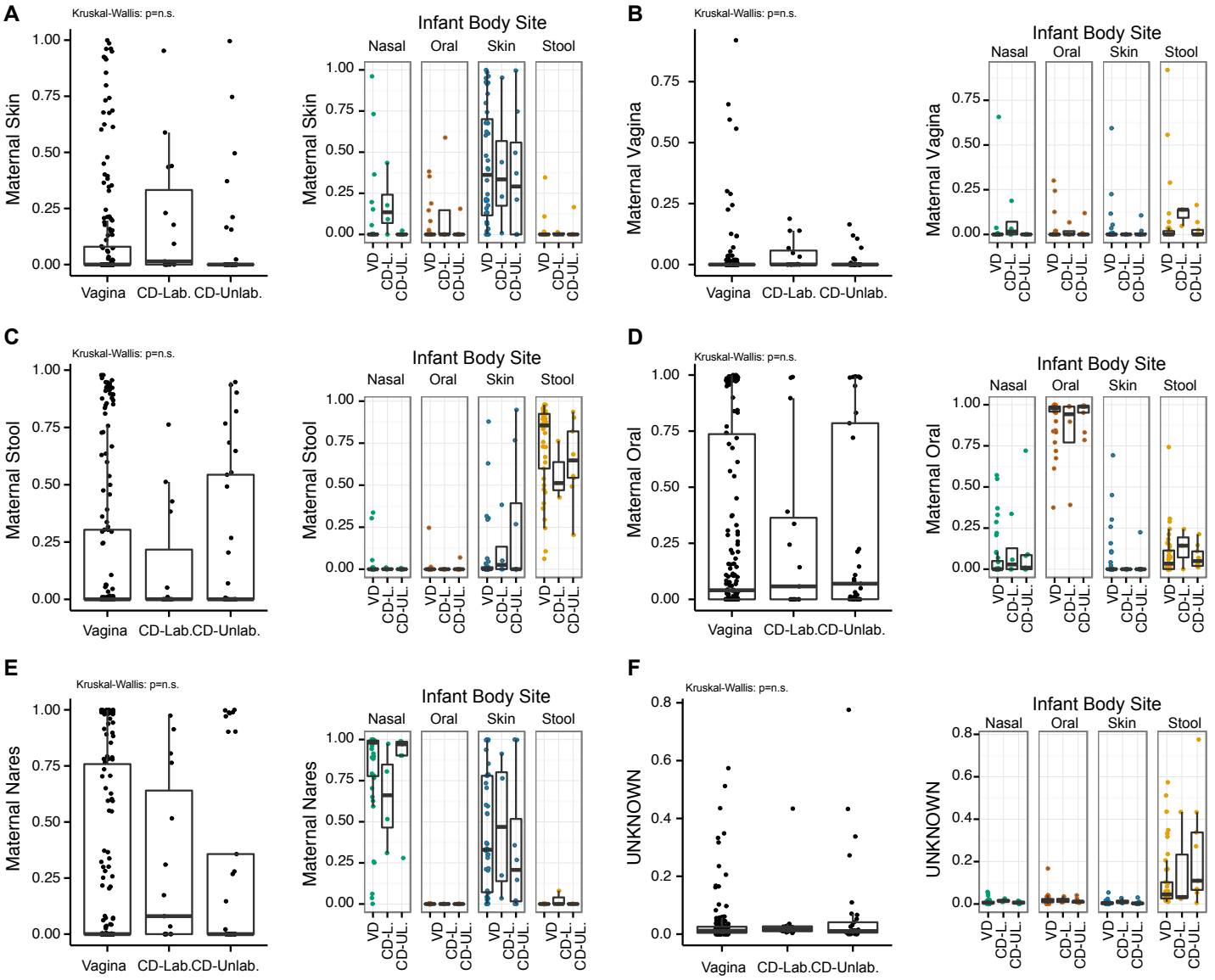

Fig. S11. Projected Taxa Similarities Among Maternal and Neonatal Microbial Communities at 6 Weeks

The predicted proportion of OTUs within the infant microbiome (6 weeks) predicted to originate from the maternal microbiota, stratified by maternal body site (y-axes; (A) skin, (B) vagina, (C) stool, (D) oral, (E) nares and (F) unknown) and by mode of delivery (vagina, vaginal delivery; CD-Lab, Cesarean Labored; CD-Unlab., Cesarean Unlabored). Each point represents an individual sample, indicating the proportion of OTUs predicted to originate from the given maternal body site (y-axis). Boxplots represent the median and interquartile range with whiskers determined by Tukey's method. (*Left Panels)* Data plotted for all infant sites, stratified by mode of delivery. Significance between groups was determined by Kruskal-Wallis tests with post-hoc Dunn's non-parametric comparisons. (*Right panels)* Data stratified by neonatal body site (Nasal, green; Oral, red; Skin, blue; Stool, orange) and again by mode of delivery.

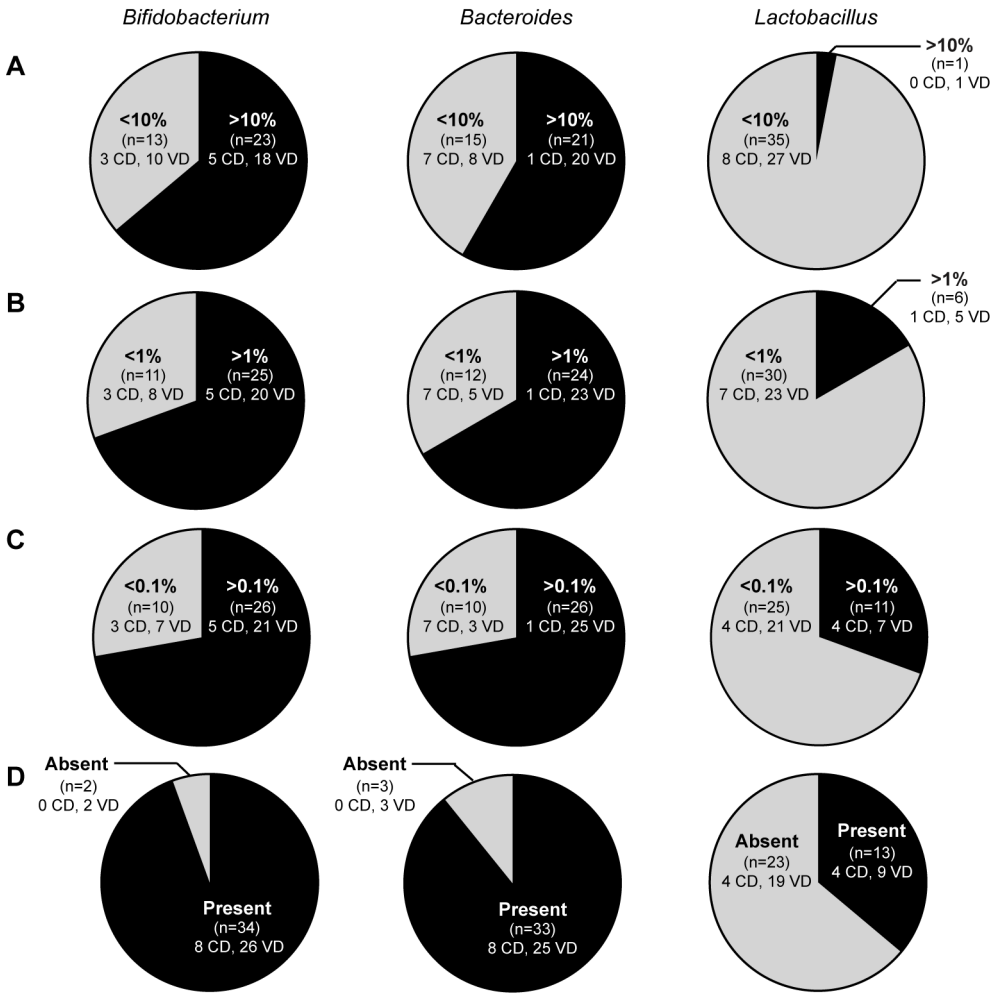

Fig. S12. Proportion of Infant Stool Samples containing *Bifidobacterium*, *Bacteroides* and *Lactobacillus* as Determined by Whole Genome Shotgun Sequencing

Pie charts represent the proportion of samples whose relative abundance of each taxa (columns) meets the indicated threshold (A, >10%; B: > 1%; C: >0.1%; C: absence (0%)). The number of vaginal or cesarean deliveries in each category are indicated in the text.

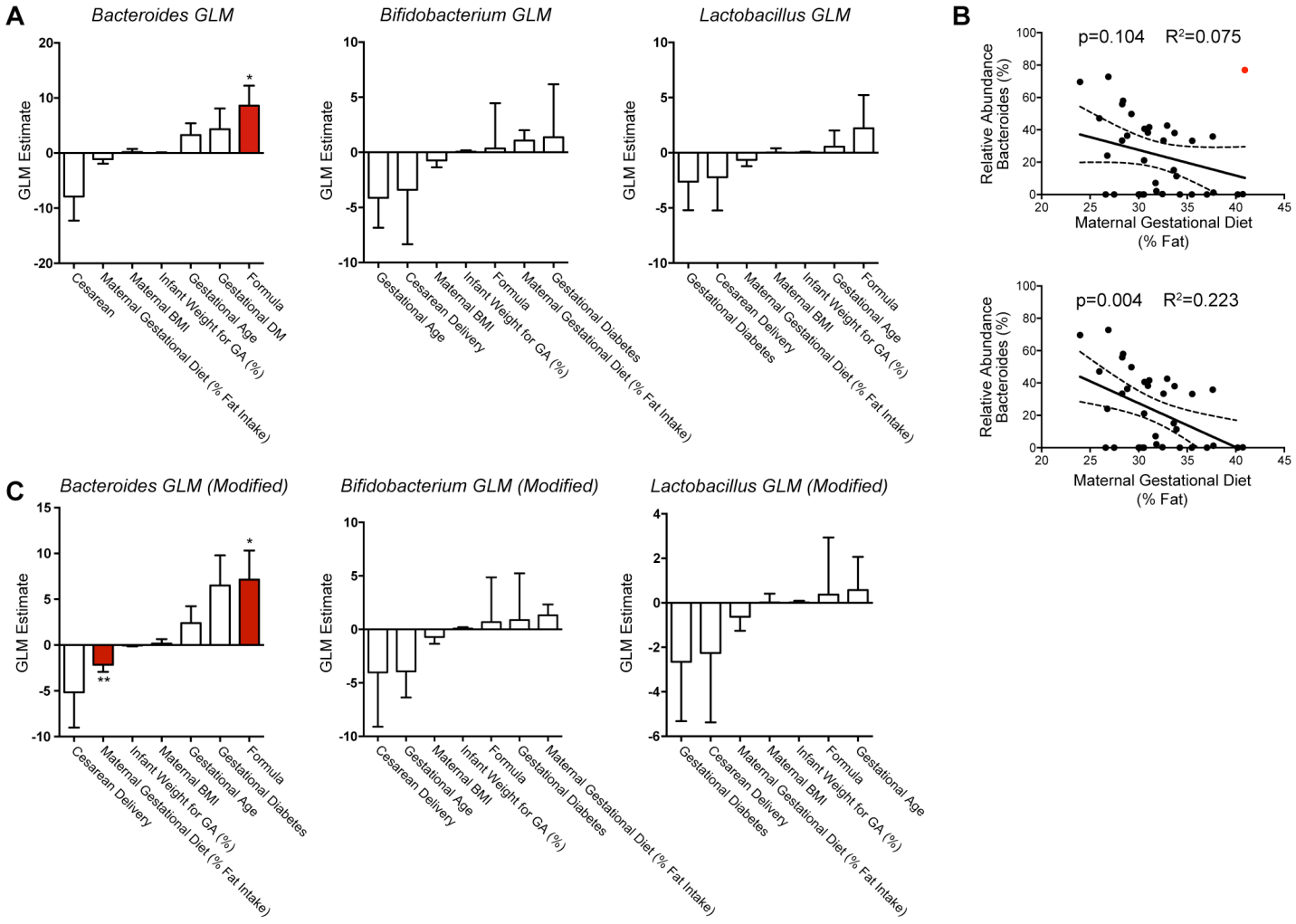

Figure S13: Clinical Factors Influencing the Relative Abundance of *Bacteroides*, *Bifidobacterium* and *Lactobacillus* in the Infant Stool at 6 Weeks (A) Generalized Linearized Model (GLM) estimates for breastfeeding status (exclusive human milk vs. any formula), mode of delivery (Cesarean vs. vaginal), maternal gestational diet (% fat intake in 3rd trimester), maternal pre-pregnancy BMI, infant weight for gestational age (GA), gestational age at delivery, and maternal gestational diabetes. Colored bars indicate factors with significant associations with decreased or increased amounts of the indicated taxa (* $p<0.05$, ** $p<0.01$). Error bars depict standard error. (B) Association of maternal gestational intake of fat (%) against the relative abundance of *Bacteroides* in the infant stool at 4-6 weeks of age. (*Top*) Data shown with a potential high-leverage outlier indicated in red. Solid line indicates a linear model fit ($p=0.104$, $R^2=0.075$) inclusive of all points with a 95% confidence interval (dashed lines). (*Bottom*) Linear model fit with potential outlier excluded ($p=0.004$, $R^2=0.223$). (C) GLM Estimates for indicated factors, but with the single potential outlier excluded.

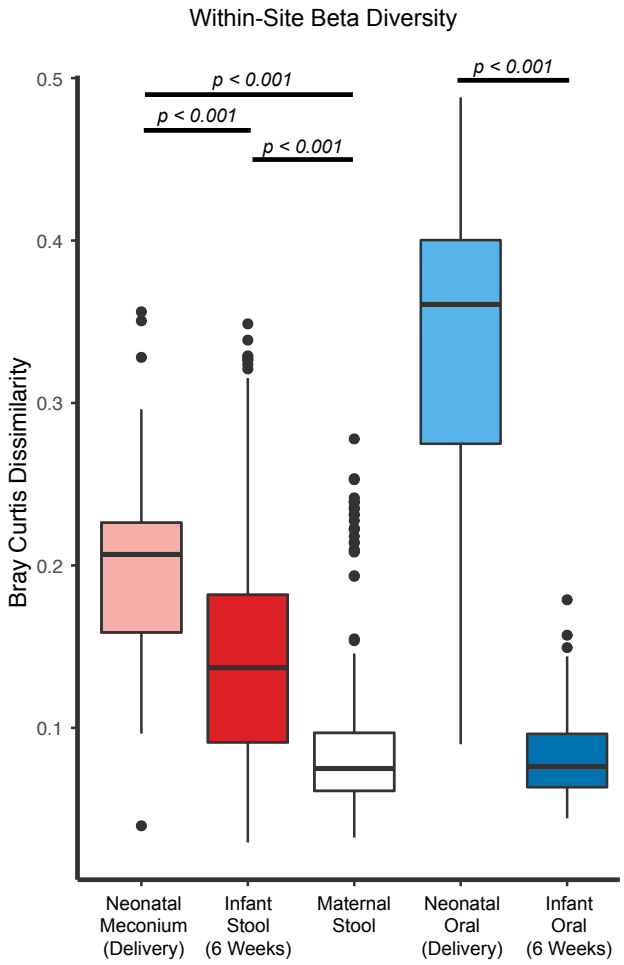

Fig. S14. Within Site Beta Diversity Comparisons of the Infant and Maternal Metagenomes

Bray-Curtis dissimilarity distances (beta diversity) within the same group (matched age, site and time point) for all sequenced metagenomes. Smaller values indicate a greater similarity within a group. Boxplots represent the median and interquartile range with whiskers created using Tukey's method. Statistical comparisons were made using pairwise t-tests with subsequent multiple hypothesis correction with Holm's method.

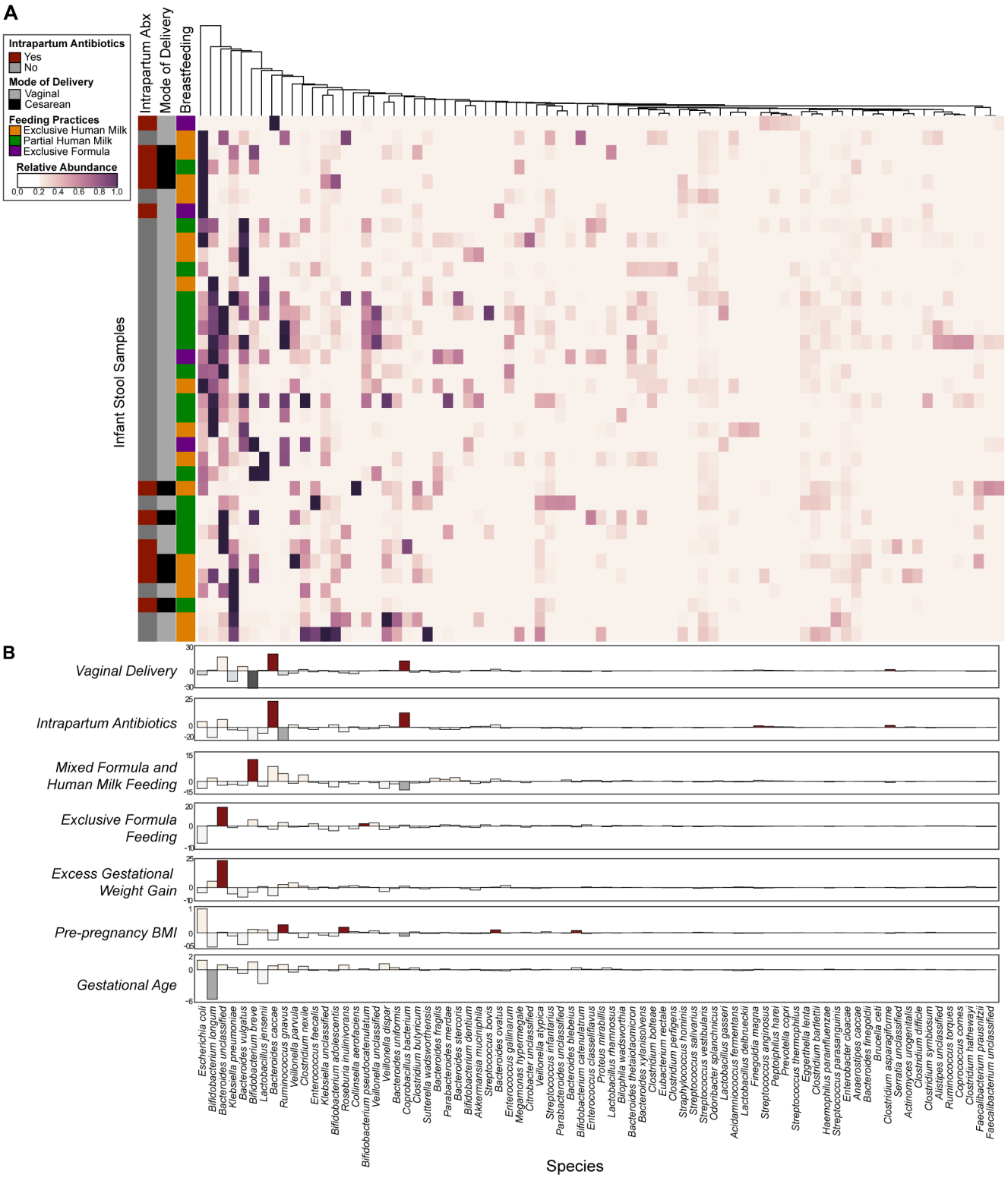

Fig. S15. Association of Infant Microbiota with Clinical Metadata using a Generalized Linear Model (A) Heatmap of the relative abundances of species found in infant stool samples as determined by WGS sequencing. The vertical bar indicates mode of delivery, breastfeeding practices and antibiotic usage. Dendrograms indicate hierarchical clustering on Euclidean distances using average linkage. (B) A generalized linear mixed model was fitted to identify taxa whose abundances differed significantly between individuals by

virtue of various clinical metadata. The strength of linear model predictions for each metadata category is represented by bar height. Significant correlations ($p<0.05$) are indicated by the darker shade (dark red or dark gray). (a) Taxa enriched in infants born vaginally (red, up) or by Cesarean (gray, down) (b) taxa enriched in infants exposed to intrapartum antibiotics (red, up) as opposed to no antibiotics (gray, down); (c) taxa enriched in partially breast fed (both human milk and formula) infants (red, up) as opposed to exclusive human milk only (gray, down); (d) taxa enriched in exclusively formula fed infants (red, up), as opposed to human milk only (gray, down); (e) taxa enriched in cases of excess maternal gestational weight gain (red, up), as opposed to normal weight gain (gray, down); (f,g) taxa positively (up, red) or negatively (down, gray) correlated with pre-pregnancy BMI or gestational age.

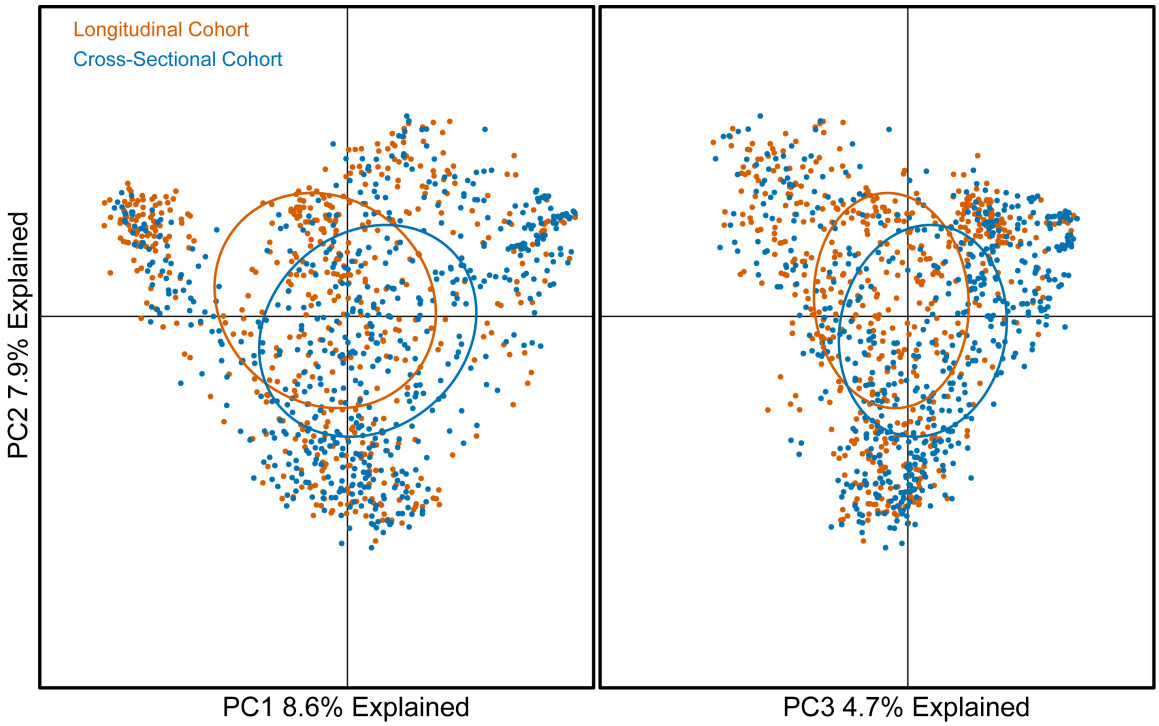

Fig. S16. Principal Coordinates Analysis of all Samples Stratified by Cohort.

The fisrt three principal coordinate (PC) axis shown representing 21.2% of the total variation within the dataset. Samples of all body sites are included, colored by the cohort from which it was derived (blue: Cross-sectional cohort; red, Longitudinal Cohort). Ellipses represent a 95% confidence interval around the centroid for each group.

Table Legends

Table S1. Sequencing Summary. Tabular data given for number of samples sequenced, OTUs identified and high quality sequencing reads retained after filtering (see methods). The number of sequences per sample is stratified by age, body site and time point.

Table S2. Cohort Demographics Stratified by Cohort. Characteristics of the cohort given, stratified by cohort. ± indicates standard deviation from mean where applicable. *Statistical comparisons between cohorts were determined with a chi-squared test for categorical variables or a Student's t-test for continuous variables.

Table S3. Cohort Demographics Stratified by Mode of Delivery. Characteristics of the cohort given, stratified by mode of delivery. ± indicates standard deviation from mean where applicable. *Statistical comparisons between mothers who delivered vaginally or by Cesarean were determined with a chi-squared test for categorical variables or a Student's t-test for continuous variables.

Table S4. Taxa Composition Summaries of the Neonatal and Maternal Microbiome at Delivery. Average relative abundances and standard deviation (SD) of each identified genera stratified by age neonate vs. mother) and body site. The indicator value indexes (IndVal) for each taxa are given. Higher values indicate a stronger association of a taxa with a given body site.

Table S5. Taxa Composition Summaries of the Infant and Maternal Microbiome at 6 weeks postpartum. Relative abundances of each identified genera stratified by age (infant vs. mother) and body site. The indicator value indexes (IndVal) for each taxa are given. Higher values indicate a stronger association of a taxa with a given body site.

Table S6. General Linear Model Estimates for Lactobacillus, Bacteroides and Bifidobacterium in the Infant Stool at 6 Weeks.

Table S7: KEGG Pathway Relative Abundances. Data for WGS shotgun sequencing of infant stool, oral and maternal stool at delivery and 6 weeks provided. KEGG pathways are given with its corresponding higher level classification (Module | KEGG Pathway Name).

Table S8: Infant Metadata Per Subject

Table S9: Maternal Metadata Per Subject

Table S10. Sequence Statistics by Sequencing Pool. Tabular data indicating the number of samples and average reads per sample for each sequencing pool after minimal filtering. The samples sequenced in each pool are also indicated, identified by study ID, age, body site, visit and cohort.

Table S11. Mapping Statistics of WGS data by HUMAaN. Tabular data indicating the total number and percent of paired reads remaining after filtering for human contamination and after mapping to a bacterial database using HUMAaN.

Table S12: Taxa (16S) Relative Abundances. Data for 16S rRNA gene sequencing of all maternal and infant samples provided, with OTUs identified down to the genus level.