

SUPPLEMENT

Supplementary Materials and Methods

Stool sample Collection and Sequencing of Bacterial 16S rRNA gene

Mothers enrolled in the VDAART clinical trial were given a stool sample collection kit when their infants were between 3 and 6 months of age. One to two days prior to their clinical appointment, mothers were asked to collect ½ teaspoon of their infant's stool from a dirty diaper using a tongue depressor, and to store the sample in the freezer. Stool samples were not collected if the infant had taken antibiotics within the past 7 days. A total of 333 infant stool samples were collected. DNA extractions were performed on stool samples, and the bacterial rRNA 16S gene (V3 to V5 hyper-variable regions (V3-V5)) were amplified. Pyrosequencing data (Roche 454 Titanium) were produced by the Genome Center (TGI) at Washington University in St. Louis, MO. Quality filtering, trimming, and chimera checking were performed as previously described.⁽²¹⁾ In brief, the protocol allows one mismatch in the primer and zero mismatches in the barcode. Chimera slayer software was used to remove chimeric reads.⁽²²⁾ All samples that passed quality control screening (n=325) were classified from phylum to genus level at a confidence threshold of 0.5 using the Ribosomal Database Project (RDP) Naive Bayesian Classifier version 2.2, training set 6.⁽²³⁾ The reads whose taxonomic assignments were lower than a 0.5 confidence threshold were assigned to the unclassified group. All samples were scaled to 1,000 reads prior to analysis.

Vitamin D Assessment in Cord Blood

Liquid chromatography-mass spectrometry (LC-MS/MS) was performed to measure 25(OH)Vitamin D on 261 infants' umbilical cord blood samples, collected . Calibration was

performed using control solutions (1–200 ng/ml) diluted from standard stock solutions provided by Calbiochem (San Diego, CA). Samples from study subjects were prepared and analyzed through a turbulent flow LC system (Cohesive Technologies, Franklin, MA) followed by traditional laminar flow chromatography. The analysis was performed using a TSQ Quantum Ultra triple mass-spectrometer (Thermo Finnigan Corp., San Jose, CA).

Hierarchical (Tree) Structure of Microbiome We computed a taxonomic tree structure for each infant's gut microbiome sample (down to the genus level) with the HMPtrees package in R statistical Software. Comparison of taxonomic trees using this method accounts for the hierarchical nature of the data, and allows for visualization of microbial community differences at multiple taxonomic levels (phylum, class, order, family, and genus). Known taxonomic relationships from the Ribosomal Database Project were used to construct taxonomic trees, and tree validity was assessed for each stool sample. Infant characteristics showing the strongest and most consistent associations with overall diversity (C-section, Breastfeeding) were used as a basis of comparison for taxonomic tree structure. For these comparisons (based on object oriented data analysis (OODA))²⁴, maximum likelihood estimation was used to derive a central or modal tree for one group (i.e. infants delivered by C-section) vs. another (i.e. infants delivered vaginally). After computing and plotting the MLE trees for each group, we used the HMPtrees package to determine if the central trees computed for each group were different from an overall central/modal tree (p values were computed to reject the null hypothesis that central trees from each group were the same as a grand central tree across all subjects). Plots of central trees at the family level (rather than genus level, for ease of visibility) were color coded, with blue representing the branches with the highest confidence, and red representing branches with the lowest confidence level. Results of this analysis are shown in a supplementary file.

Supplementary Results

Hierarchical (Tree) Structure of Microbiome. A check of tree validity demonstrated a valid hierarchical tree structure for each infant's stool sample. Each infant's microbiome tree structure served as input for calculating overall central or modal hierarchical trees. Based on a pairwise likelihood ratio test (LRT) comparing the central, modal tree for C-section (vs. the central tree for vaginal delivery) we rejected the null hypothesis that the tree distribution parameters are the same regardless of mode of delivery ($p=0.018$, for 1000 bootstraps). Similarly, in a LRT comparing central, hierarchical taxonomic trees of breast fed vs. formula fed infants (or infants breast fed < 6 mths), we also detected a difference in tree structure that had borderline statistical significance ($p=0.06$). Plots of the central trees revealed differences in the individual branches that are likely responsible for the overall differences in tree structure (Figure 4). For instance, in the central tree plot for infants born by C-section, the confidence value for the tree branch representing the Bacteroidetes family is much less than it is for the vaginal delivery tree (where this branch of the tree is featured predominantly). Conversely, tree branches for *Fusobacteriaceae*, *Clostridiaceae*, and *Peptostreptococcaceae* are featured at a greater confidence level in the central tree for C-section delivery. Although the root tree branches are essentially the same for comparison of the central tree plots by breastfeeding, bacterial taxonomic families in the leaves of the tree have different levels of confidence by breastfeeding. For instance, the central tree structure for infants breast fed at least 6 months shows greater confidence in *Lactobacillus*, *Fusobacteriaceae*, and *Enterobacteriaceae* and lower confidence in the *Rikenellaceae* branch. Hierarchical tree comparisons for race/ethnicity (African Americans vs. all other races) and cord blood vitamin D (> 30ng/ml vs < or = 30 ng/ml) were not statistically significant ($p>0.2$).

Supplemental Tables

Supplemental Table 1. Bacterial Co-abundance Factor Analysis (of Top 25 Taxa)

	Bacterial Co-abundance Factor Loadings			
	<u>Factor 1</u>	<u>Factor 2</u>	<u>Factor 3</u>	<u>Factor 4</u>
<i>Bacteroides</i>	-0.31	-0.28	0.70	-0.22
<i>Escherichia Shigella</i>	-0.23	-0.28	-0.44	-0.07
<i>Bifidobacterium</i>	-0.28	-0.17	-0.47	0.00
<i>Veillonella</i>	-0.08	0.08	0.15	0.66
<i>Blautia XIV</i>	0.23	0.02	0.06	-0.08
<i>Enterococcus</i>	-0.02	0.05	-0.29	-0.11
<i>Klebsiella</i>	-0.05	0.46	-0.01	0.02
<i>Streptococcus</i>	-0.09	0.00	-0.20	0.18
<i>Lachnospiraceae Dorea</i>	0.59	-0.08	-0.14	-0.23
<i>Lachnospir. Coprococcus</i>	0.41	0.00	0.06	0.24
<i>Lachnospiraceae U.</i>	0.82	-0.02	-0.06	-0.09
<i>Parabacteroides</i>	-0.05	-0.17	0.28	-0.23
<i>Collinsella</i>	-0.03	0.05	0.08	-0.23
<i>Lactobacillus</i>	-0.20	-0.05	-0.12	0.15
<i>Clostridium</i>	0.03	0.12	-0.07	0.36
<i>Coprobacillus</i>	0.16	-0.02	-0.14	0.02
<i>Sporacetigenium</i>	0.14	0.06	-0.13	0.08
<i>Enterobacteriaceae U.</i>	-0.04	0.86	-0.06	-0.05
<i>Enterobacter</i>	-0.07	0.76	0.00	-0.17
<i>Holdemania</i>	0.29	-0.01	-0.03	-0.10
<i>Megasphaera</i>	-0.09	0.06	0.06	-0.10
<i>Eubacterium</i>	0.21	-0.03	0.08	0.06
<i>Akkermansia</i>	0.06	0.07	0.00	-0.12
<i>Clostridiales U.</i>	0.65	-0.02	0.07	0.01
<i>Lactococcus</i>	0.04	0.05	-0.03	-0.15

U. denotes "Unclassified" at the genus level

Supplemental Table 2a. Taxa Associated with Exclusive Breastfeeding in Multivariate Analysis*

Taxonomic Level	Feature	Value	Coefficient	N	N not 0	Q Value
Order	Clostridiales	exclus_brstfed	-0.0336	325	262	2.71E-05
Genus	Eubacterium	exclus_brstfed	-0.05168	325	211	9.54E-05
Genus	Lachnospiraceae Dorea	exclus_brstfed	-0.04496	325	188	9.54E-05
Genus	Lachnospiraceae Blautia	exclus_brstfed	-0.09946	325	247	9.54E-05
Family	Lachnospiraceae	exclus_brstfed	-0.08523	325	295	9.54E-05
Genus	Escherichia	exclus_brstfed	0.101134	325	301	9.54E-05
Genus	Lactococcus	exclus_brstfed	-0.00823	325	157	0.002101
Genus	Coprococcus	exclus_brstfed	-0.03186	325	205	0.002101
Family	Peptostreptococcaceae	exclus_brstfed	-0.02355	325	206	0.004193
Genus	Shigella	exclus_brstfed	0.001246	325	9	0.004898
Genus	Collinsella	exclus_brstfed	-0.05468	325	115	0.006761
Genus	Ruminococcus	exclus_brstfed	-0.03676	325	220	0.00682
Genus	Staphylococcus	exclus_brstfed	0.004348	325	76	0.013895
Genus	Actinobaculum	exclus_brstfed	0.001051	325	2	0.037126
Family	Pasteurellaceae	exclus_brstfed	0.005379	325	62	0.037333
Family	Ruminococcaceae	exclus_brstfed	-0.04783	325	257	0.048881
Genus	Ruminococcus	exclus_brstfed	-0.07437	325	282	0.048881
Family	Erysipelotrichaceae	exclus_brstfed	-0.03576	325	229	0.048881

*Analysis adjusted for C-section, race, gender, maternal antibiotics during labor, gestational age, treatment group

Supplemental Table 2b. Taxa Associated with Breastfeeding with formula supplementation in Multivariate Analysis*

Taxonomic Level	Feature	Value	Co-efficient	N	N not 0	Q Value
Genus	Megasphaera	Breastfed +formula	0.004281	325	157	0.125102
Genus	Clostridium	Breastfed +formula	-0.00348	325	117	0.31282
Genus	Achromobacter	Breastfed +formula	0.000444	325	2	0.31282
Genus	Pediococcus	Breastfed +formula	0.000985	325	4	0.406929
Family	Aerococcaceae	Breastfed +formula	-0.00146	325	29	0.410452
Genus	Sutterella	Breastfed +formula	0.016912	325	89	0.410452
Family	Stenotrophomonas	Breastfed +formula	0.002109	325	36	0.423436
Family	Lactobacillaceae	Breastfed +formula	0.000775	325	2	0.452165
Genus	Escherichia	Breastfed +formula	-0.03608	325	301	0.452165

Supplemental Table 3. Taxa Associated with C-section in Multivariate Analysis*

Taxonomic Level	Feature	Value	Coefficient	N	N not 0	Q Value
Genus	Bacteroides	Csection (Yes)	-0.18421	325	267	8.34E-05
Genus	Parabacteroides	Csection (Yes)	-0.05381	325	149	0.104271

*Analysis adjusted for breastfeeding, race, gender, maternal antibiotics during labor, gestational age, treatment group

Supplemental Table 4. Taxa Associated with Race

Taxonomic Level	Feature	Value	Coefficient	N	N not 0	Q Value
Genus	Megasphaera	Caucasian (vs. African Amer)	-0.00768	325	157	0.006
Genus	Megasphaera	African Amer(vs all other races)	0.0065	325	157	0.005
Genus	Lactococcus	African Amer (vs. all other races)	0.0064	325	157	0.014

*Analysis adjusted for breastfeeding, C-section, gender, maternal antibiotics during labor, gestational age, treatment group

Supplemental Table 5. Taxa Associated with Cord Blood Vitamin D (ng/ml)

Taxonomic Level	Feature	Value	Coefficient	N	N not 0	Q Value
Genus	Lactococcus	Cord Blood Vit D (ng/ml)	-0.00032	325	157	0.03
Genus	Saccharomonospora	Cord Blood Vit D (ng/ml)	1.86E-05	325	1	0.03
Family	Methylobacteriaceae	Cord Blood Vit D (ng/ml)	1.86E-05	325	1	0.03
Genus	Lachnobacterium	Cord Blood Vit D (ng/ml)	4.93E-05	325	4	0.039

*Analysis adjusted for race, breastfeeding, C-section, gender, maternal antibiotics during labor, gestational age, treatment group

Supplemental Table 6. Top 10 taxa in infant gut microbiome associated with dog ownership*

Taxonomic Level	Feature	Value	Coefficient	N	N not 0	Q Value
genus	Butyricoccus	Dog (Yes)	0.002921	325	8	0.246027
family	Bradyrhizobiaceae	Dog (Yes)	0.000477	325	4	0.731713
genus	Gluconacetobacter	Dog (Yes)	0.001198	325	5	0.748448
genus	Erwinia	Dog (Yes)	-0.00449	325	134	0.748448
family	Erysipelotrichaceae	Dog (Yes)	0.0005	325	1	0.748448
genus	Yersinia	Dog (Yes)	0.000177	325	1	0.748448
family	Pseudomonadaceae	Dog (Yes)	0.000614	325	1	0.748448
family	Aerococcaceae	Dog (Yes)	0.001445	325	29	0.748448
genus	Clostridium	Dog (Yes)	0.010475	325	238	0.748448
genus	Polaromonas	Dog (Yes)	0.000184	325	1	0.759363

*Analysis adjusted for race, breastfeeding, C-section, gender, maternal antibiotics during labor, gestational age, treatment group

Supplemental Table 7. Top 10 taxa in infant gut microbiome associated with cat ownership*

Taxonomic Level	Feature	Value	Coefficient	N	N not 0	Q Value
family	Porphyromonadaceae	Cat (Yes)	0.00722	325	3	0.476653
genus	Hydrogenophaga	Cat (Yes)	0.000362	325	1	0.476653
genus	Alicyclobacillus	Cat (Yes)	0.000282	325	1	0.476653
genus	Psychrobacter	Cat (Yes)	0.001777	325	1	0.476653
genus	Leucobacter	Cat (Yes)	0.0003	325	1	0.476653
genus	Janthinobacterium	Cat (Yes)	0.000327	325	1	0.476653
genus	Yersinia	Cat (Yes)	0.0003	325	1	0.476653
family	Pseudomonadaceae	Cat (Yes)	0.001039	325	1	0.476653
family	Lachnospiraceae	Cat (Yes)	0.050055	325	295	0.61248
genus	Dysgonomonas	Cat (Yes)	0.016777	325	35	0.850989

*Analysis adjusted for race, breastfeeding, C-section, gender, maternal antibiotics during labor,

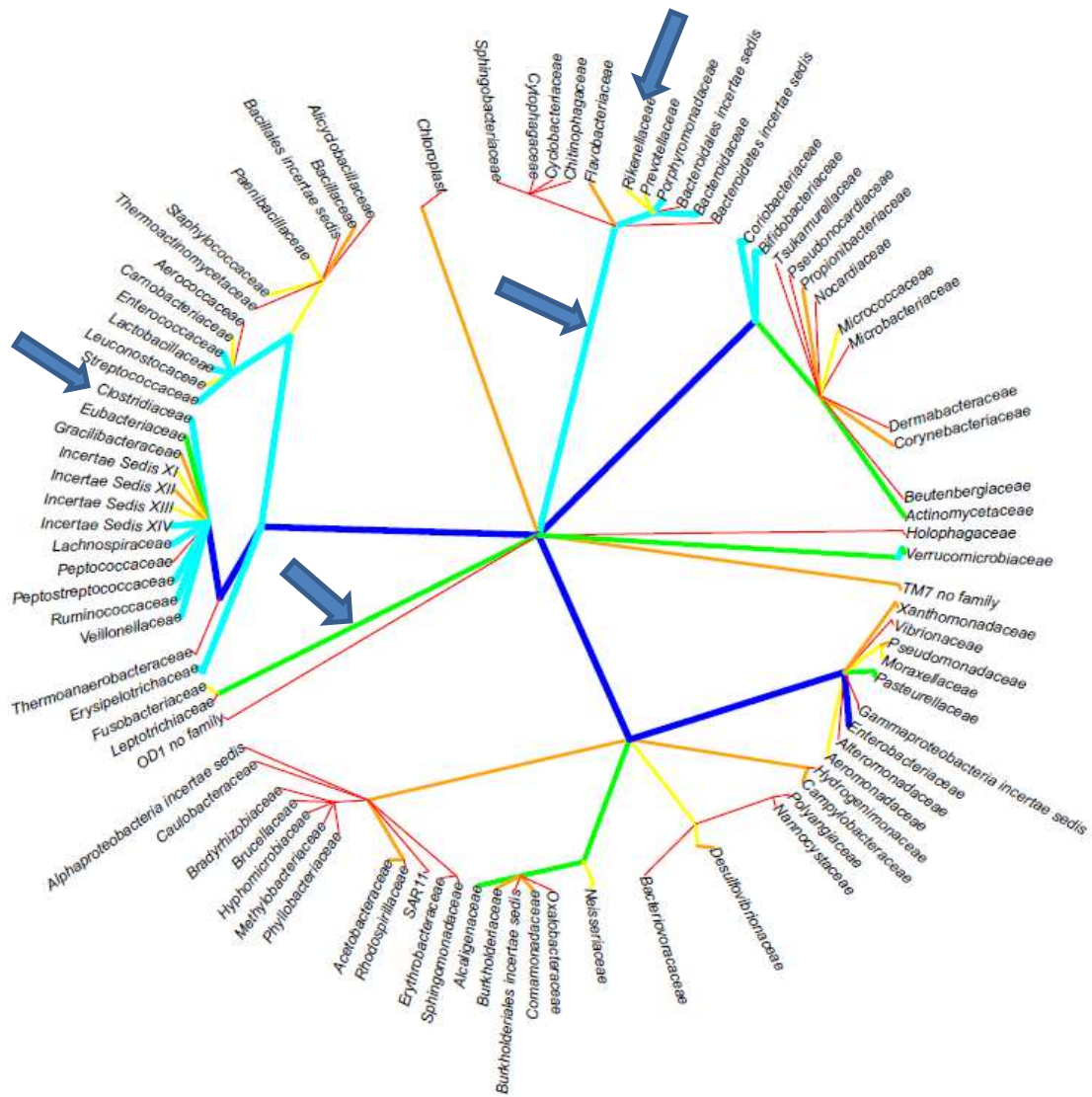
	Overall Diversity (Shannon Diversity Index)					
	Fecal Flora Sampling at 12-16 weeks (N=74)		Fecal Flora Sampling at >16weeks and < 20 weeks (N= 77)		Fecal Flora Sampling at > or = 20 weeks (N=154)	
	Beta	(P value)	Beta	(P value)	Beta	(P value)
Child's Race						
Caucasian*	-0.27	0.07	-0.20	0.25	-0.23	0.005
Breast Fed (1 st 6 months)	-0.27	0.03	-0.20	0.13	-0.21	0.004
C-section (Y/N)	0.11	0.34	0.11	0.34	0.09	0.20

gestational age, treatment group

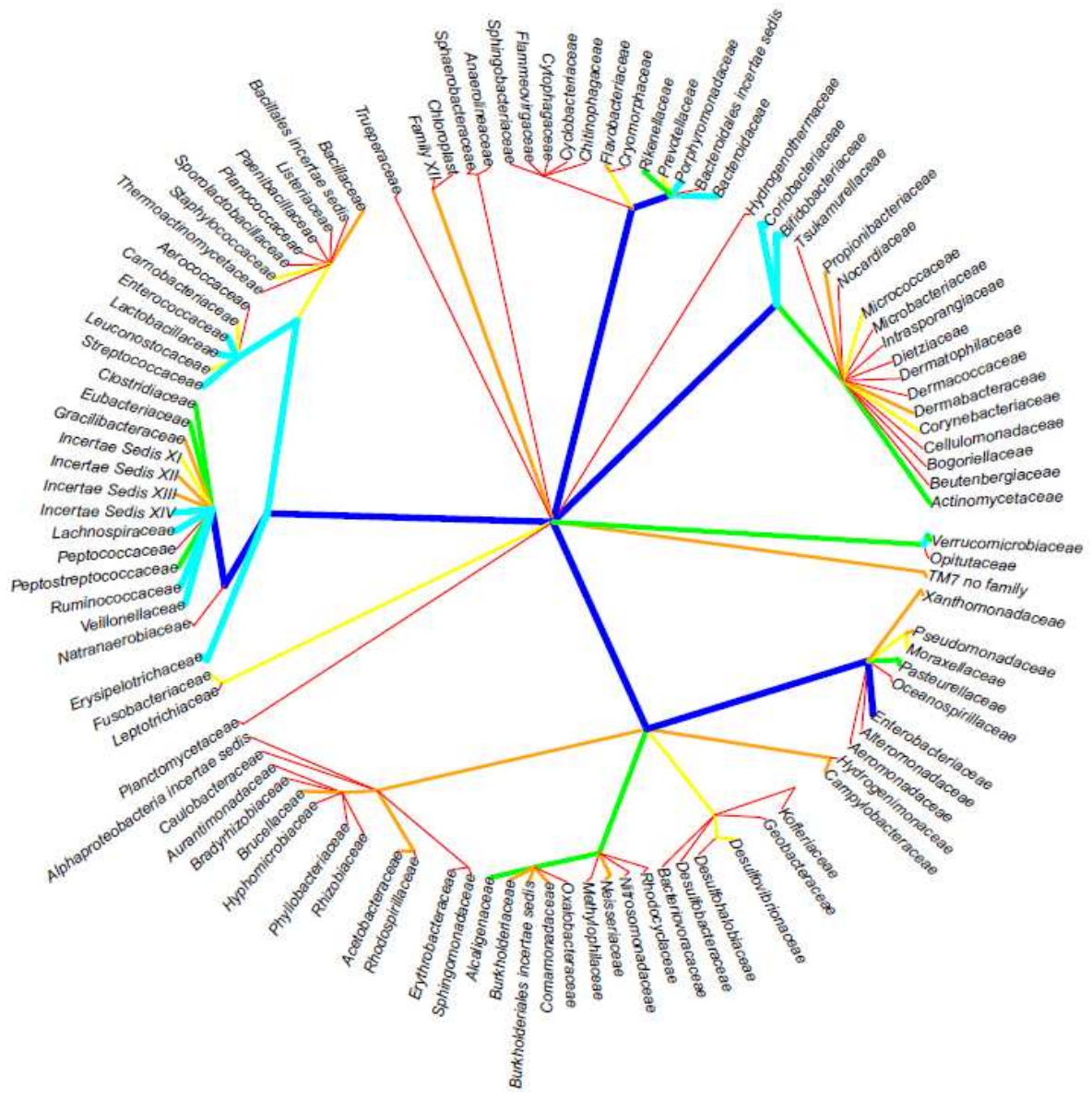
Supplemental Table 8. Relationships between infant predictors (race, Breastfeeding and c-section) by age at fecal flora sampling

Supplemental Figures

Figure S1 (a + b). Central HMP Tree structure (family level) for (a) C-section (b) Vaginal Delivery; differences in hierarchical tree branches (above a confidence value of 10) with C-section are shown with blue arrows, $p=0.018$ for comparison between two trees

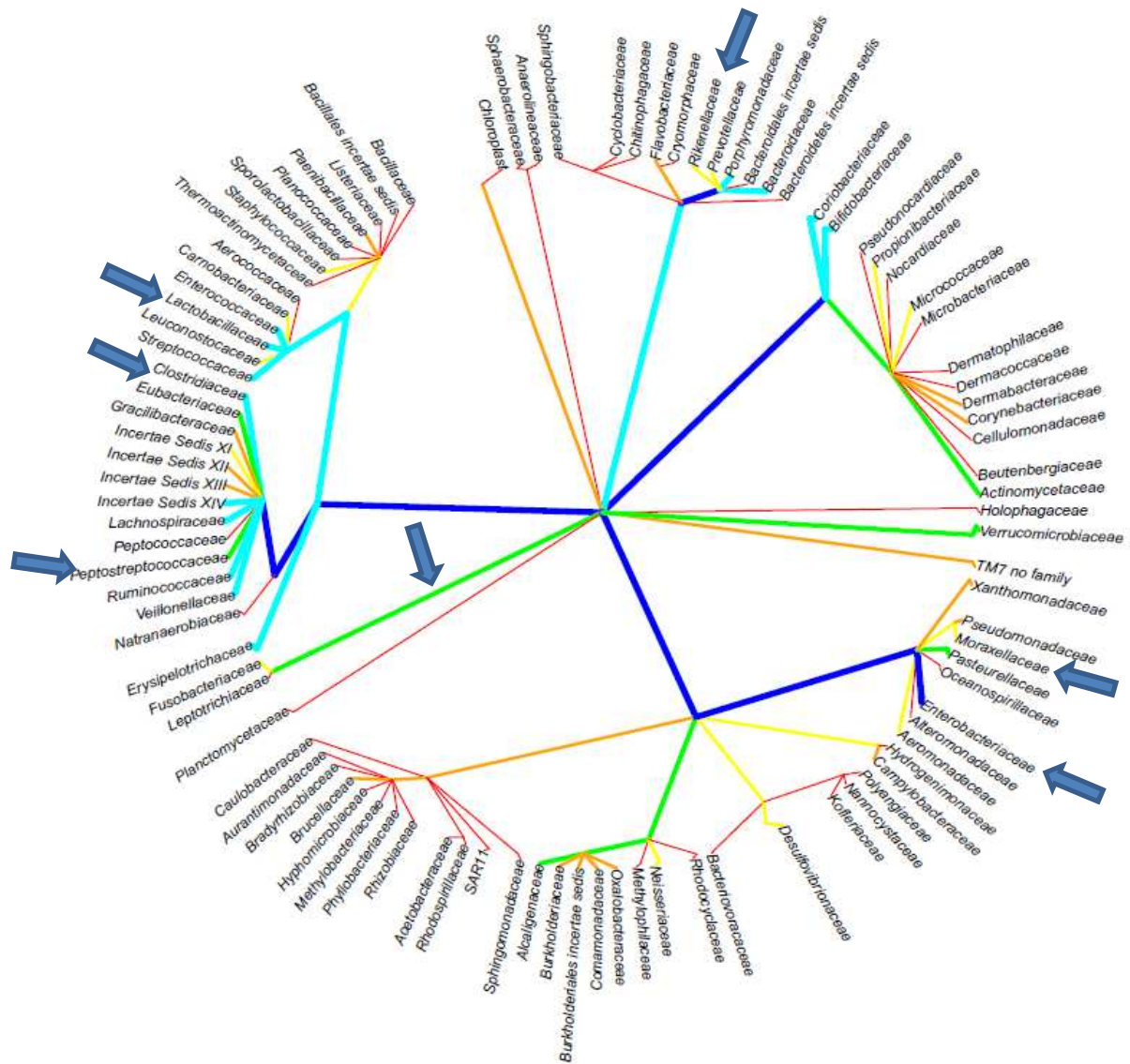


a) C-section

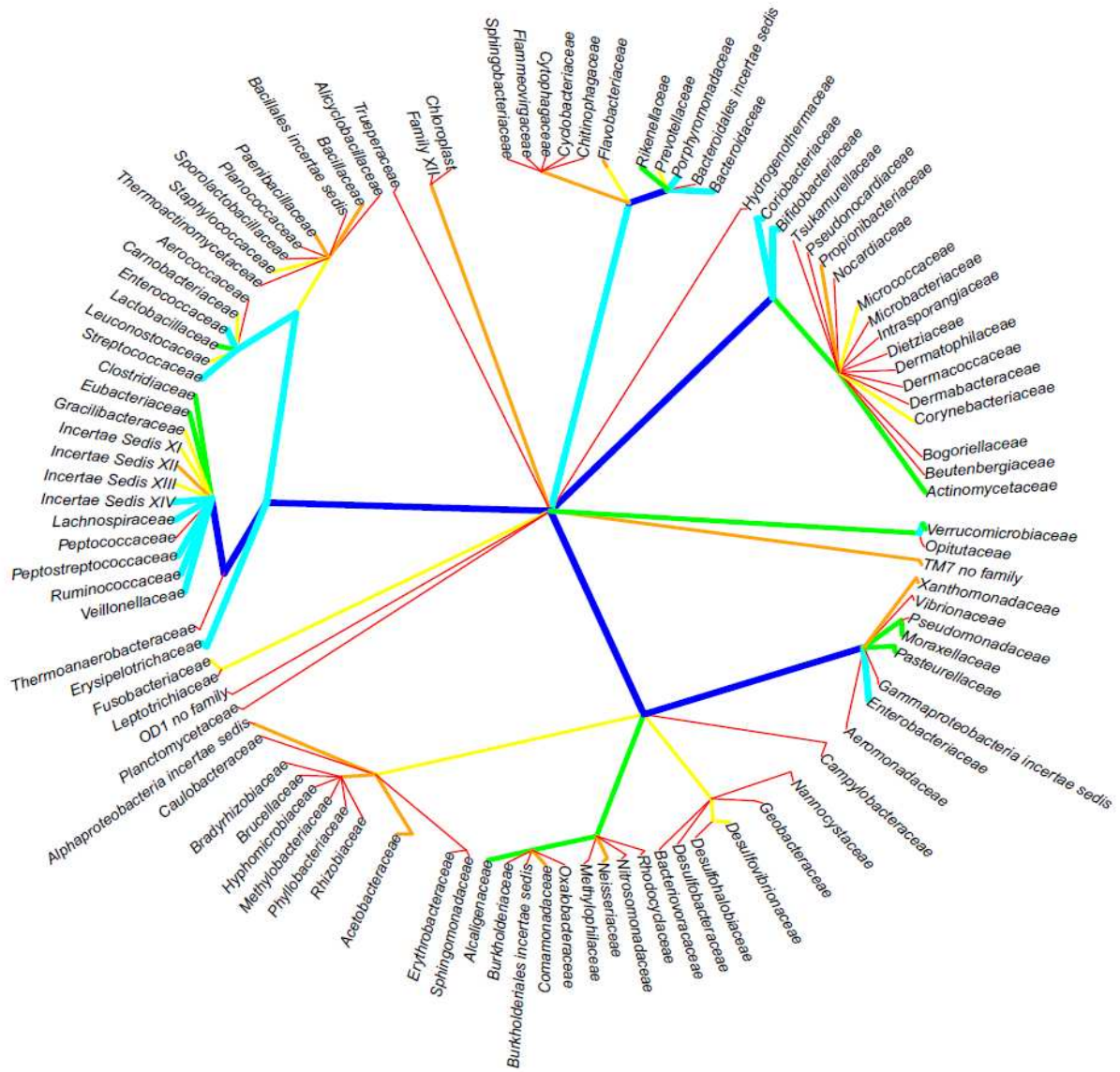


b) Vaginal Delivery

Figure S2 (a + b) Central HMP Tree structure (family level) for (a) Breast fed > or = 6 months (b) Breastfed < 6 months or formula fed; differences (above a confidence value of 10) in hierarchical tree with breastfeeding are shown with blue arrows, p=0.066 for comparison between two trees



a) Breast fed > or = 6 months



(b) Breastfed < 6 months or formula fed

Legend for Colored Lines in Figure 4:

Confidence Value	
●	10000-1000
●	1000-100
●	100-10
●	10-1
●	1-0.1
●	0.1-0