Trait-based community assembly and succession in the infant gut microbiome.

John Guittar, Ashley Shade, Elena Litchman

Tables and figures

Trait	Sources	Units
16S gene copies	8,9	No. in genome
Obligate anaerobe	9	0 or 1
B vitamins	5	No. pathways in genome
Forms biofilms	9	0 or 1
GC content	1,3	Percent (%)
Genes	6	No. in genome
Gram positive	9	0 or 1
IgA binding affinity	7	$\log ([IgA+]/[IgA-] + 1)$
Length	1,3,4	$\log (\mu m)$
Motility	3,4	NA
Sporulation	2,3,4	Continuous score (0 - 1)
Temperature optimum	3,1	°C
Width	1,3,4	$\log (\mu m)$

Table 1.

Sources of trait data used in this study.
[1] BacDive: Bacterial Diversity Metadatabase;
[2] Browne et al. 2016;
[3] IJSEM: Internat. Jo. of Syst. and Evol. Micro.;
[4] JGI: Joint Genome Institute;
[5] Magnusdottir et al. 2015;
[6] NCBI: the National Center for Biotech. Info.;
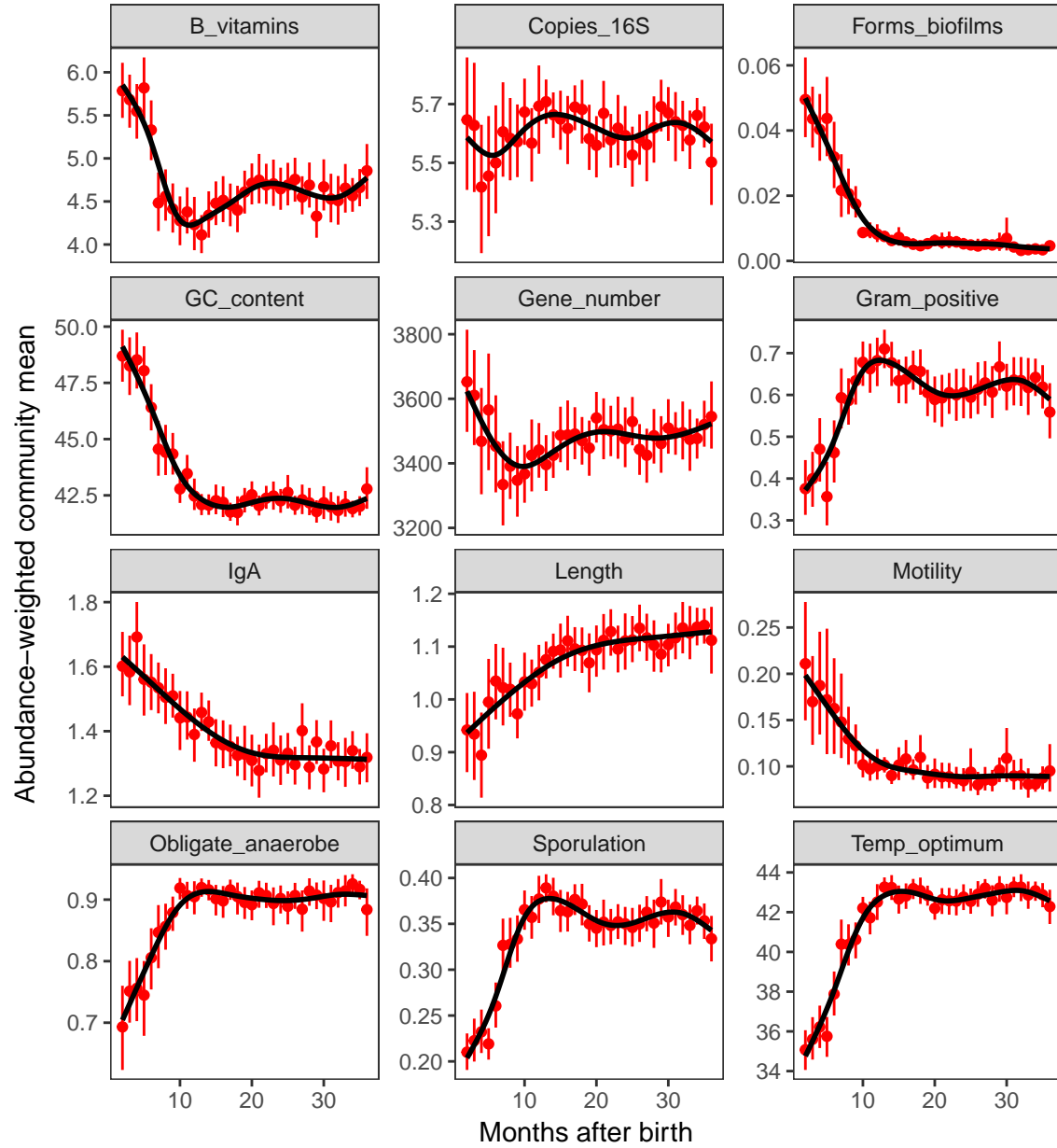[7] Palm et al. 2014;
[8] Stoddard et al. 2015;
[9] Ward et al. 2017.

Figure 1. OTU and family-level abundance patterns over time.

Temporal ranges of taxa over infant gut community development (left), ranked by peak abundance across infants. Peak abundance is calculated as the mean time of OTU presence, weighted by OTU abundance. Each colored row is an OTU, with their x-axis ranges reflecting the first and last times that their mean abundance surpassed 0.1% across all infants. Taxa that never attained 0.1% mean abundance are omitted. OTU colors reflect successional status; taxa were categorized as early (red) or late (green) successional if their global abundances trended significantly ($p < 0.05$) negative or positive over time based on linear regression; OTUs were categorizeds as mid-successional/no trend (blue) when they did not trend significantly over time. The center column of panels shows percent total abundance of OTUs from each successional category across all infants over time. The right column of panels shows proportions of total community abundance of the top five most abundant families for each successional group.

Figure 2. Abundance-weighted trait patterns over time.

Red circles show the average abundance-weighted mean of samples in that month (N = number of samples, and ranges from 26 - 55), and vertical red lines show 95% confidence intervals. Black lines show trends in community weighted trait means over time, calculated using generalized additive models.
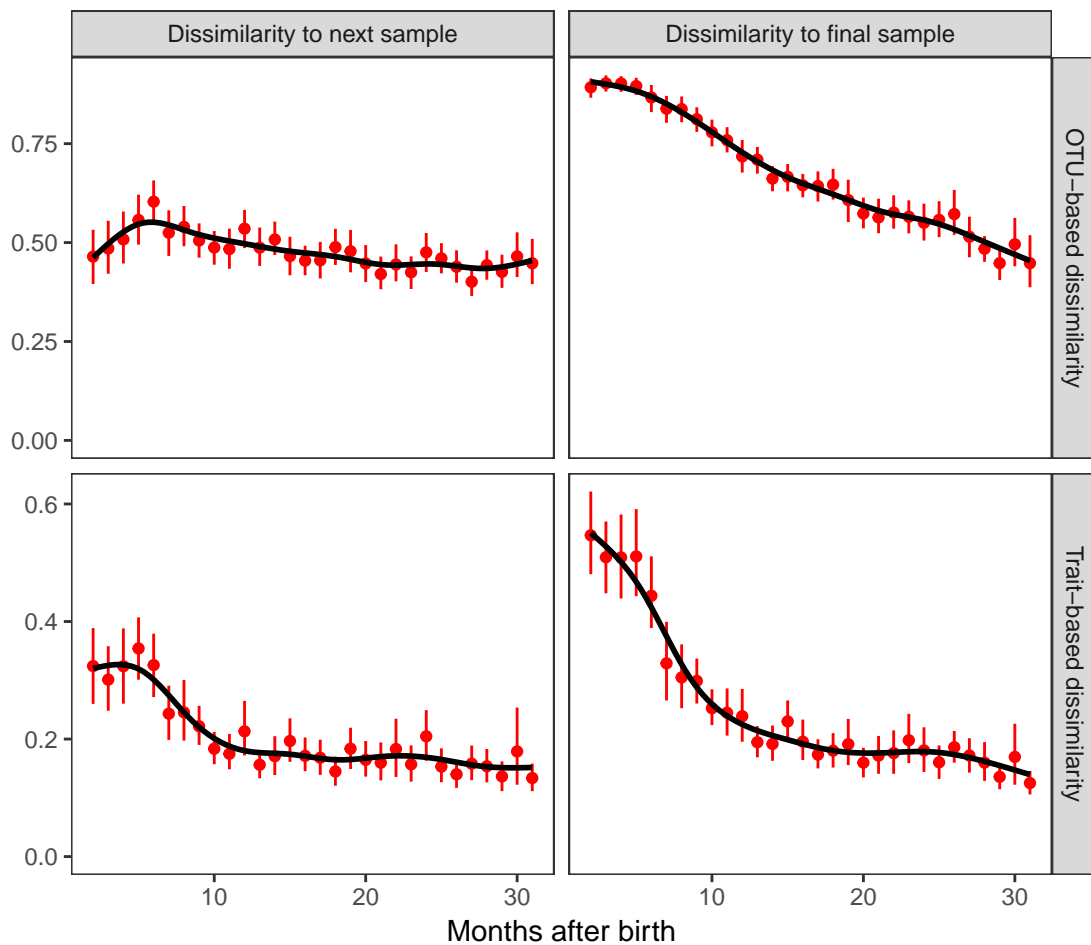
**Figure 3. Taxonomic and trait-based rates of community change within infants.**

Red circles show mean dissimilarities between among microbiome samples within infants over their development, grouped by month (N = number of sample comparisons, ranging from 26 - 55), and vertical red lines show 95% confidence intervals. At left, datashow changes in dissimilarity between subsequent samples taken from the same infants over time. At right, lines show dissimilarity between samples and the last sample taken for each infant(up to the 988 day cutoff used to standardize sampling range). Black lines show trends in community weighted trait means over time, calculated using generalized additive models.
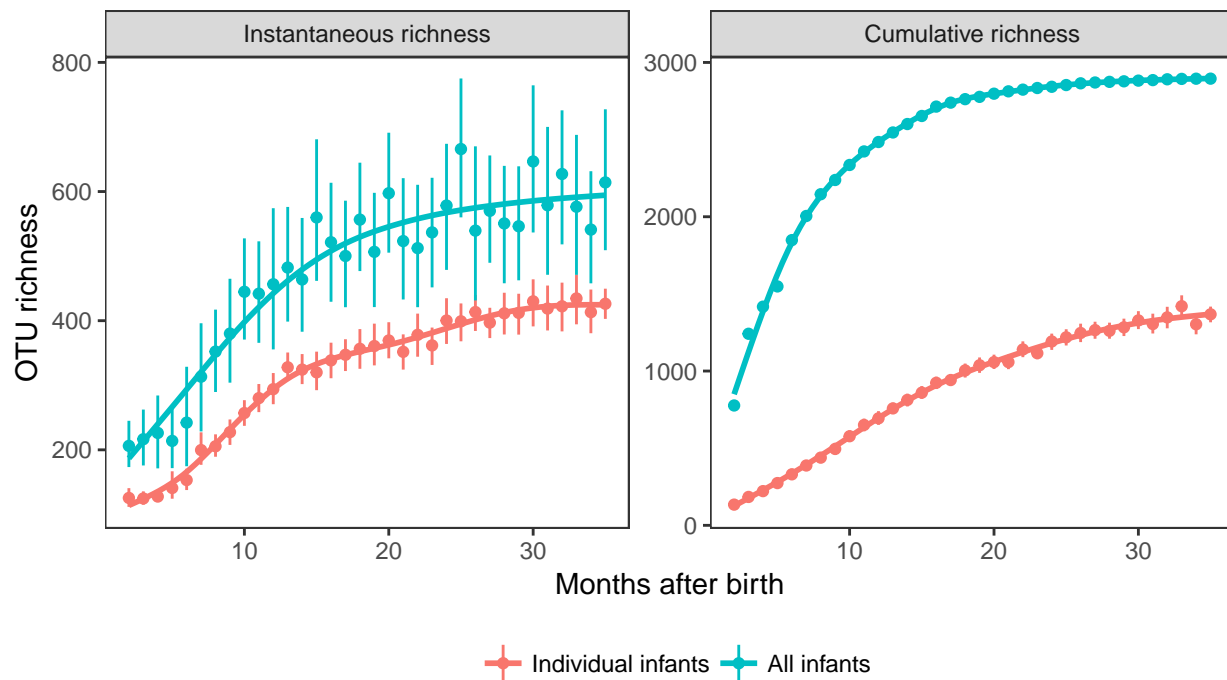
Figure 4. Patterns in OTU richness in individual infants over time.

Colored circles show mean OTU richness of infant microbiome samples over time, grouped by month. Left panel shows instantaneous richness within single infants (red) or across all infants (blue); right panel shows cumulative richness (i.e., the cumulative number of OTUs encountered by one (red) or all (blue) infants over time). Vertical red lines show 95% confidence intervals.
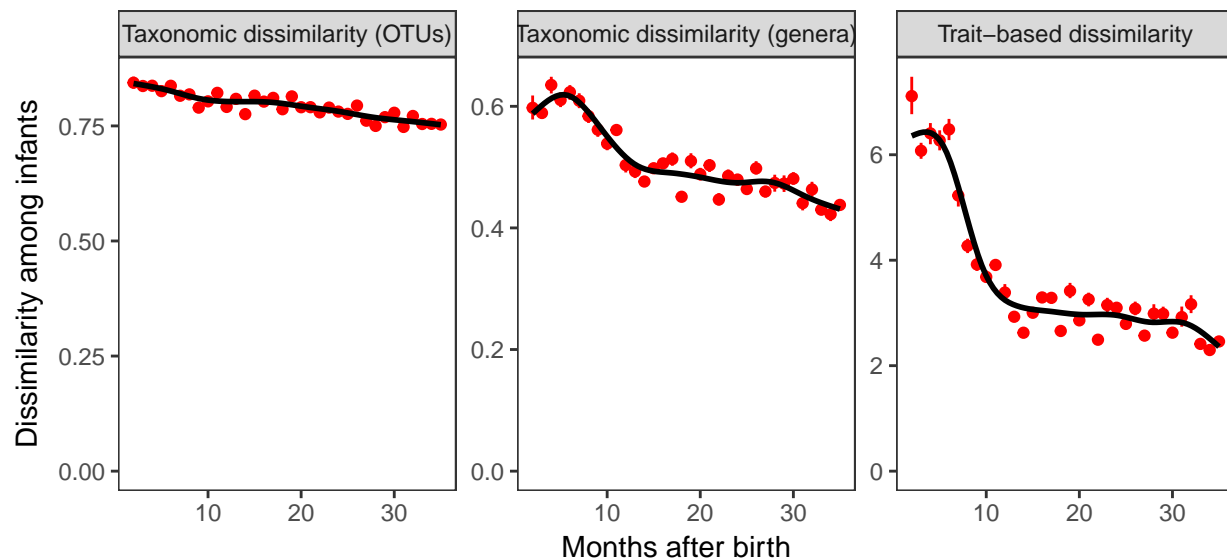
Figure 5. Mean dissimilarity in gut microbiomes across infants over early development.

Mean community dissimilarity among samples across all infants, grouped into one-month slices (excluding pairs of samples from the same infant). Taxonomic dissimilarity is quantified using Bray-Curtis dissimilarity using either OTUs (left) or genera (center) as the units of comparison; trait-based dissimilarity is calculated using Euclidean distance of all traits examined in this study, scaled so each trait has the potential for equal contribution to net dissimilarity.
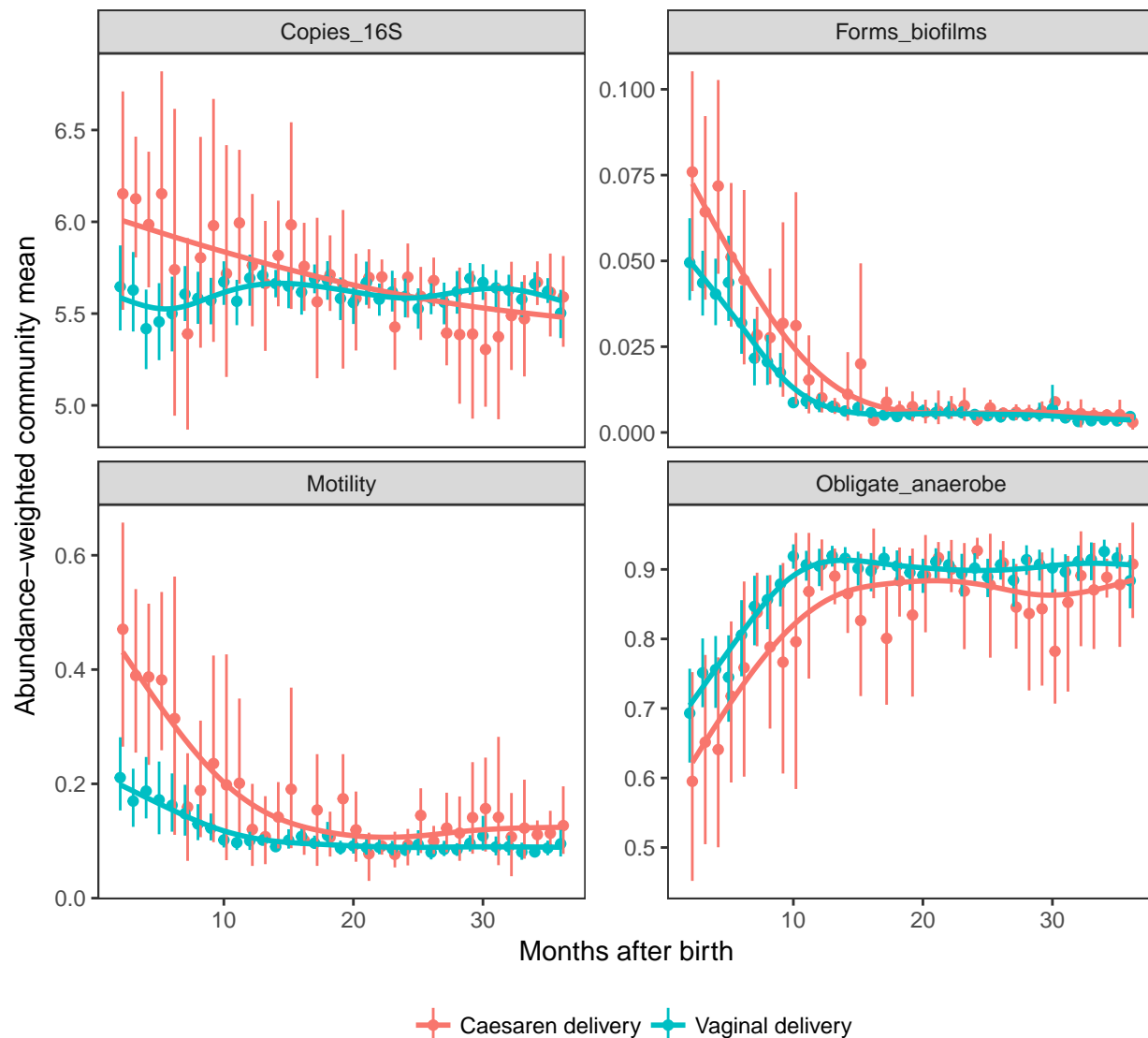
Figure 6. Abundance-weighted trait patterns over time, grouped by delivery mode.

Circles show the average abundance-weighted mean of samples in each month (N = number of samples, ranging from and ranges from three - eight samples from a total of five infants born by Caesaren, or 26 - 47 samples from a total of 51 infants born vaginally. Vertical red lines show 95% confidence intervals. Colored trendlines illustrate changes in community weighted trait means over time, calculated using generalized additive models. For comparisons of abundance-weighted patterns by delivery mode for all traits, see Fig. S9.

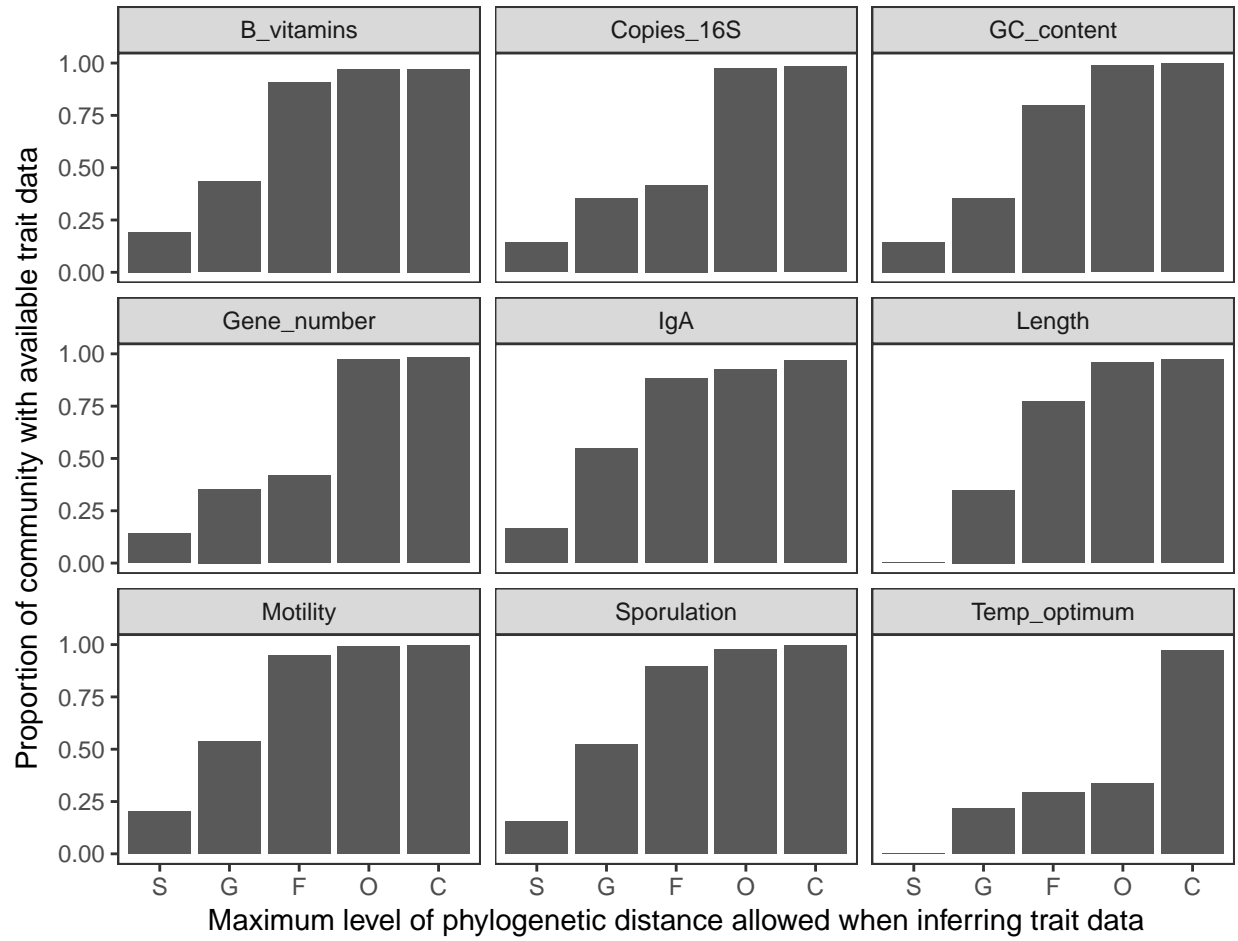# Supplementary tables and figures

Figure S1. Phylogenetic distances used when inferring traits

Unknown trait data were inferred by averaging trait values over descending subtrees in the greengenes phylogeny. That is, for each unknown trait value, the predicted value is set to the arithmetic average state of all tips descending from that node. On the x-axis, S = Species, G = Genus, F = Family, O = Order, C = Class. Missing trait data for three traits (biofilm forming, obgligate anaerobe status, and Gram positive status) were inferred in Ward et al. 2017, and are therefore not examined in this figure.
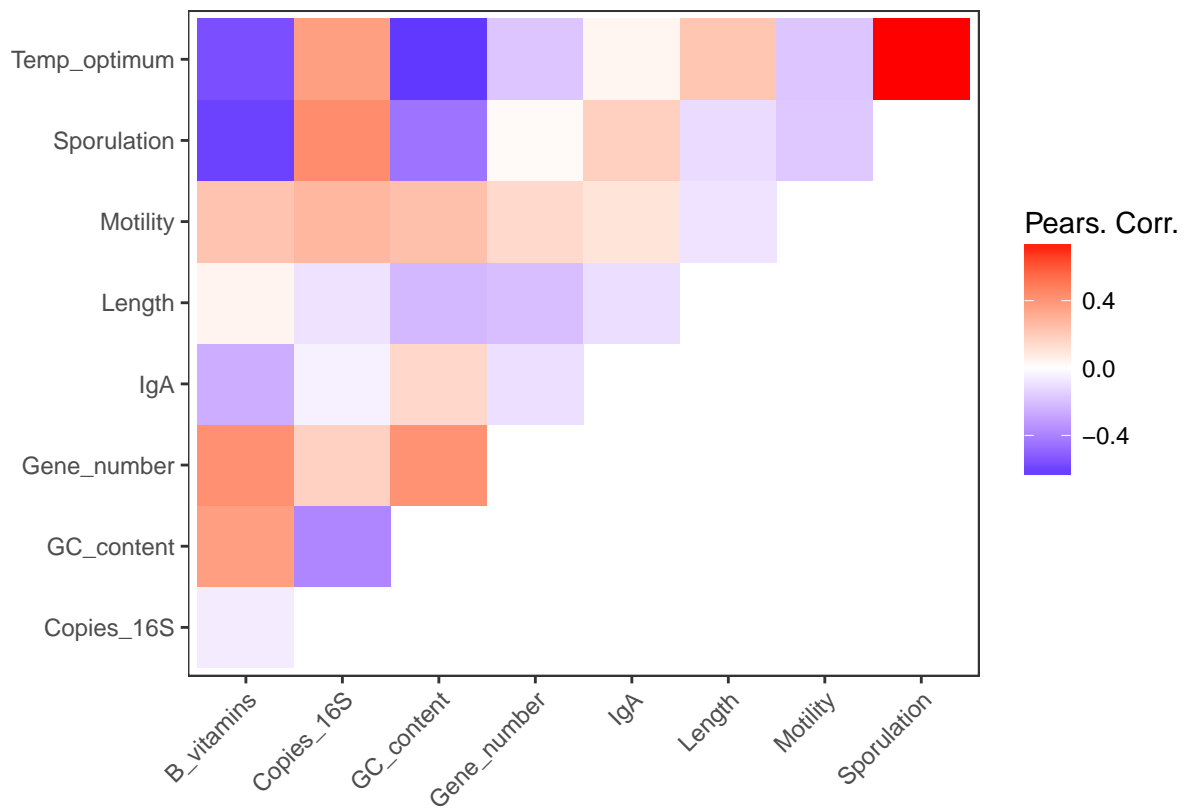
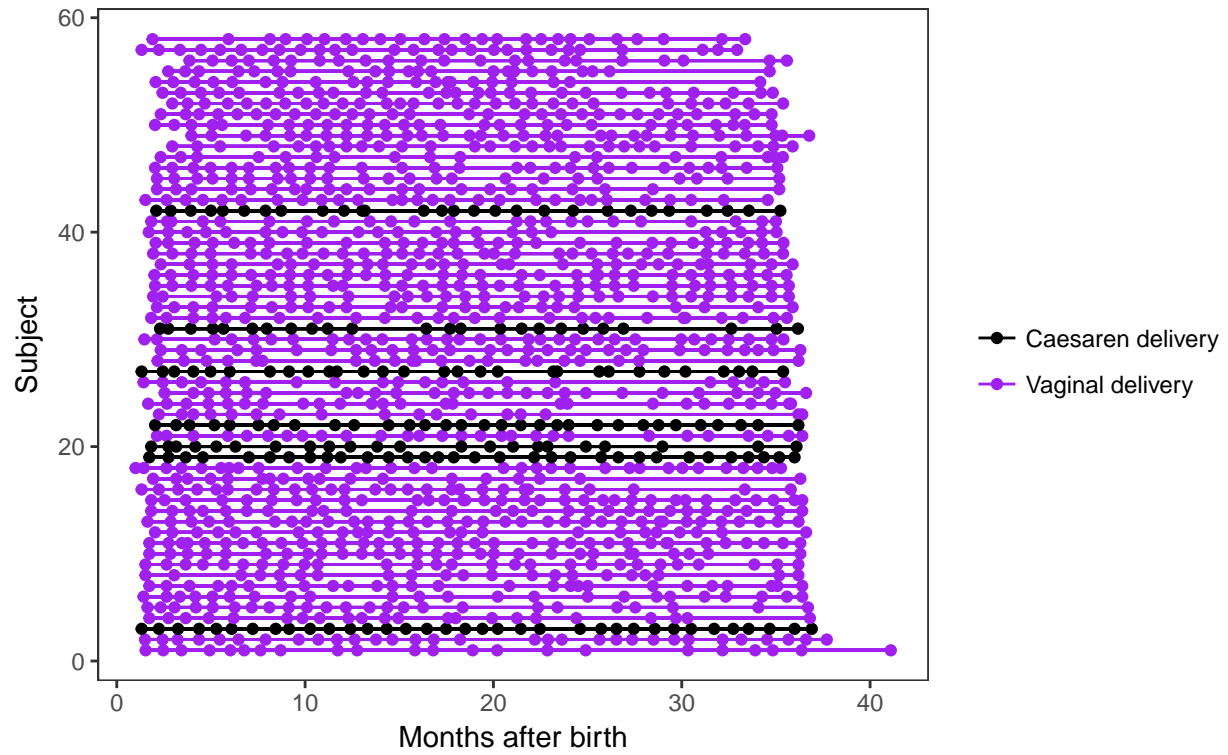Figure S2. Pearson correlation coefficients among traits.

Figure S3. Sampling times, sampling durations, and delivery modes of infants used in this study.

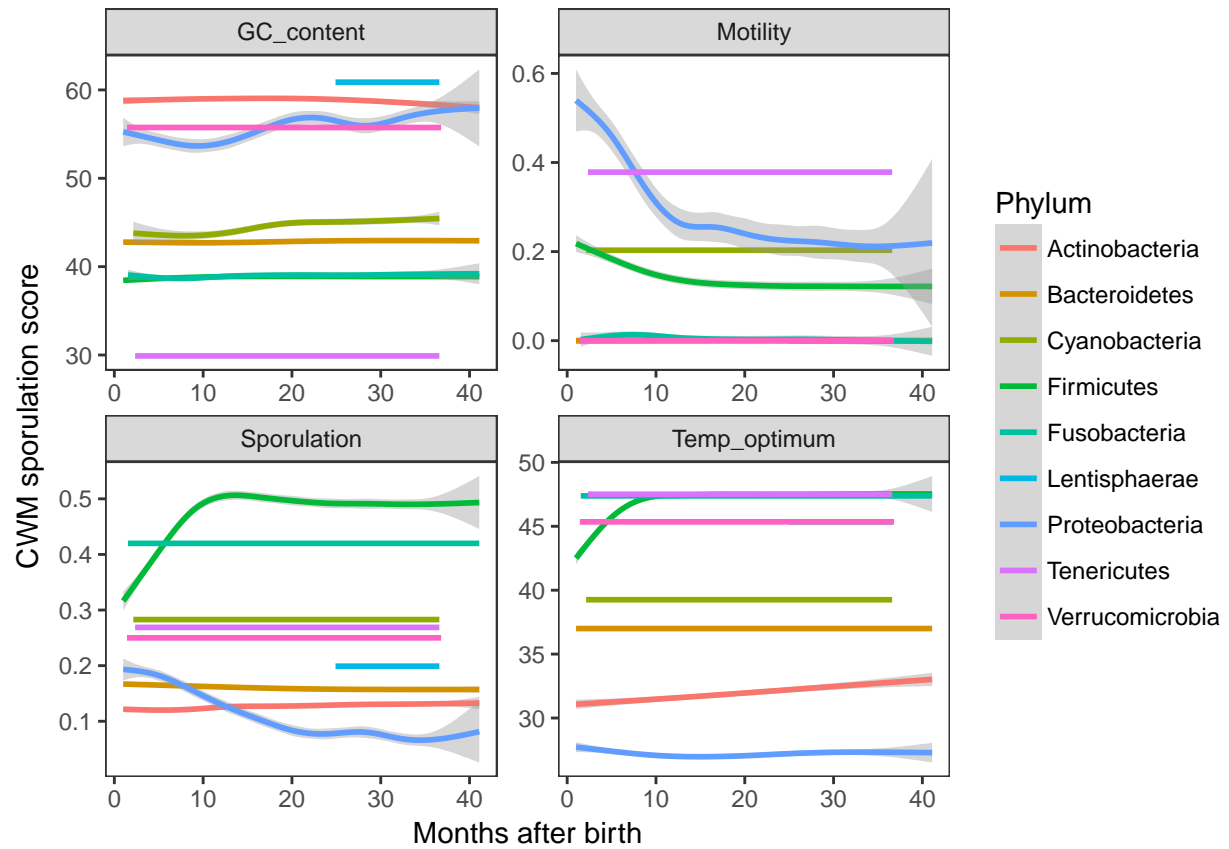Each horizontal line is a subject; each colored dot is a sample.

Figure S4. Changes in abundance-weighted mean trait scores over time, by Phylum.
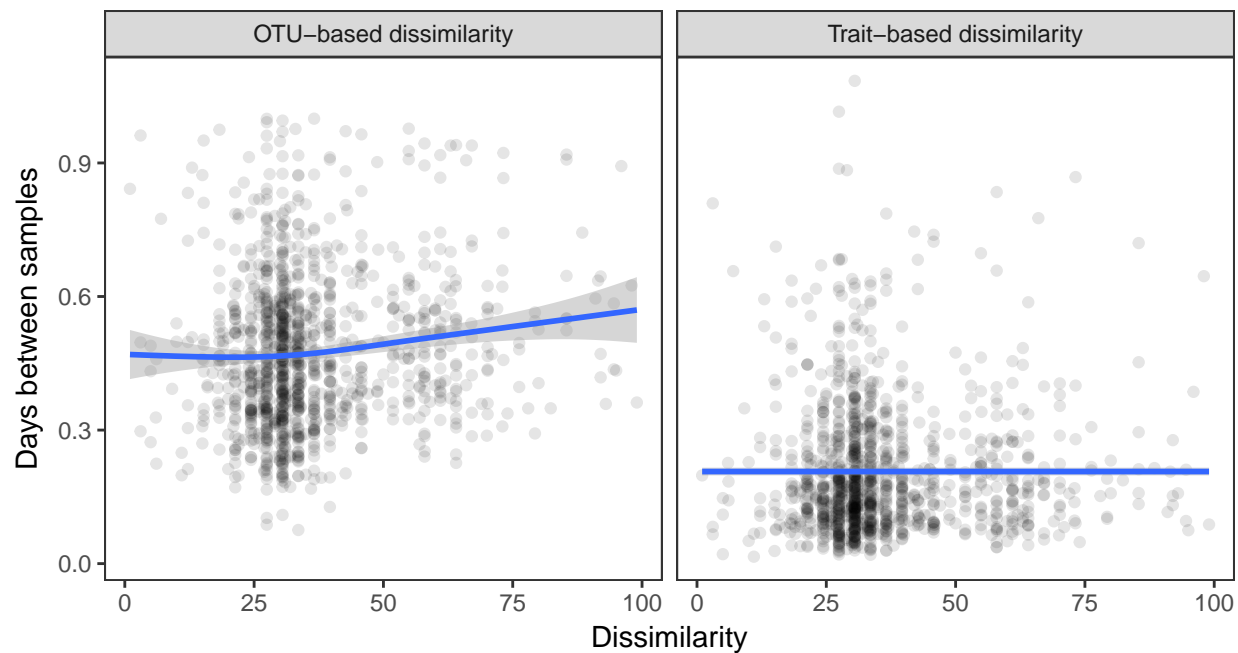
Figure S5. Dissimilarity in community composition plotted by the number of days between samples.

The Pearson correlation between OTU-based dissimilarity and the number of days between samples: r = 0.11 (N = 1118). The Pearson correlation between trait-based dissimilarity and the number of days between samples: r = 0.013 (N = 1118). Samples with more than 100 days between them (N = 30) were excluded.
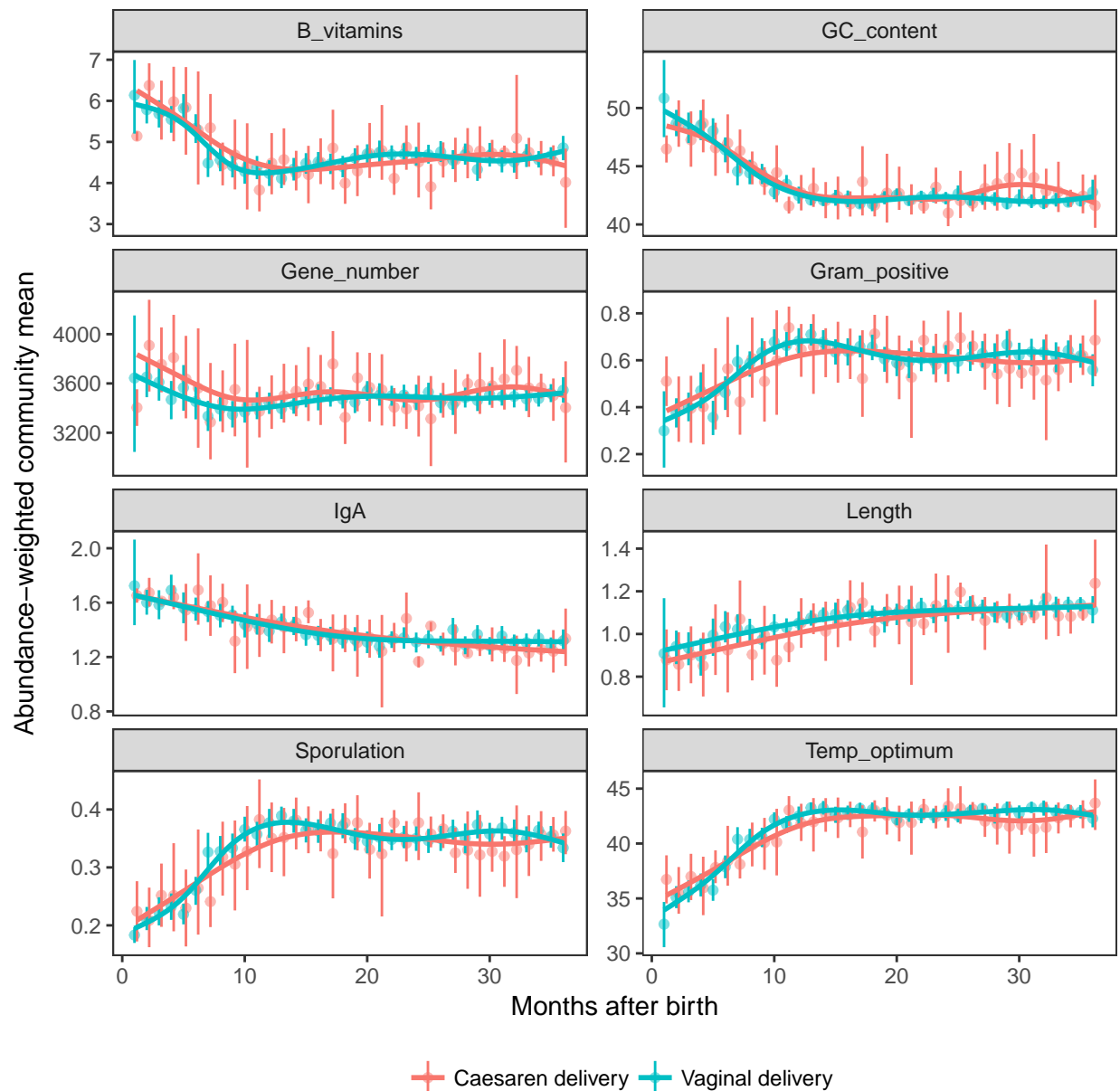
Figure S6. Trait-patterns by delivery mode

Each circle represents the CWM of one sample for one infant; red and black circles are from infants born vaginally (N = 54 infants) and by Caesaren section (N = 7 infants), respectively bla. Black lines show community weighted mean (CWM) trait values of gut microbiota over time across all infants, grouped by delivery mode. Taxa without trait information were necessarily omitted when calculating CWMs; see Table 1 for coverage statistics. For abundance-weighted patterns for all traits, split by delivery mode, see Fig. S9.
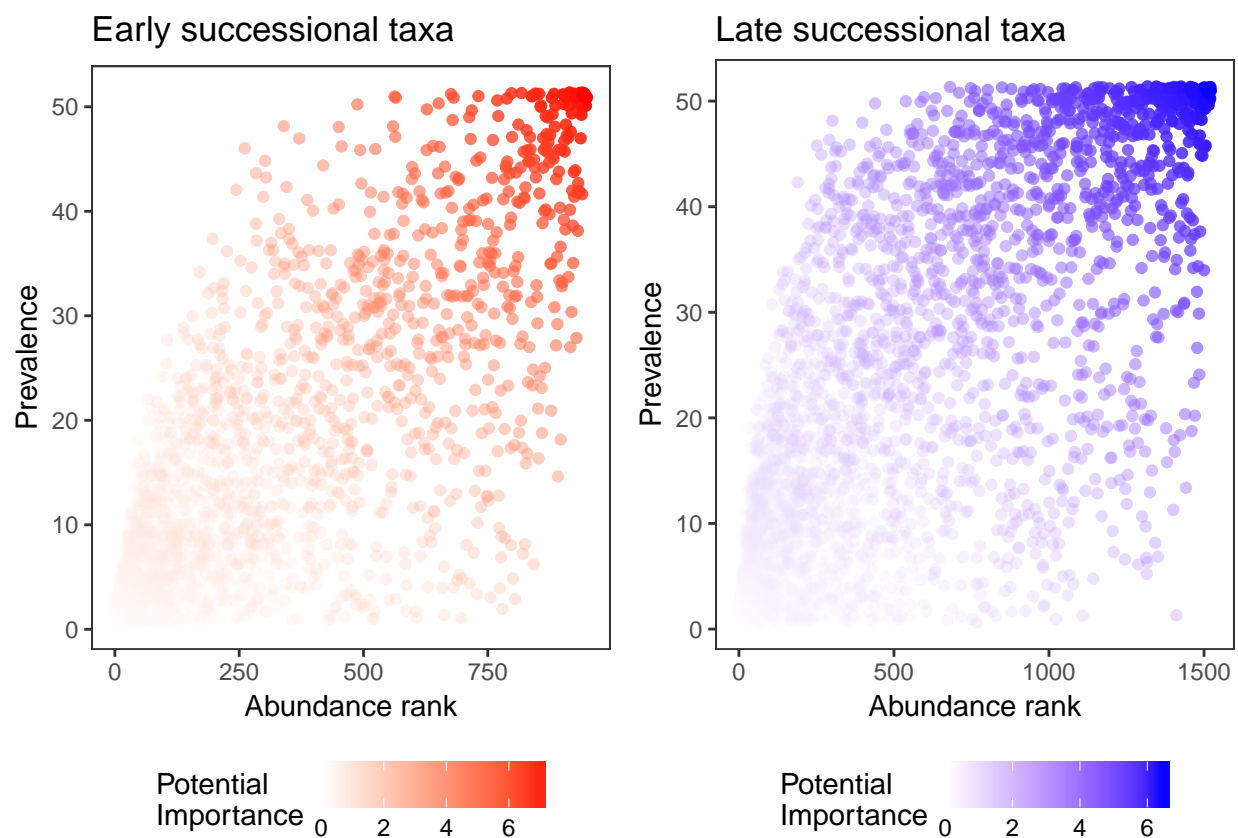
Figure S7. Taxa abundance by prevalence: a measure of importance.

Importance is calculated as the sum of prevalence (scaled) and abundance rank (scaled).

Family	Genus	Species	OTU	Importance
Bifidobacteriaceae	Bifidobacterium	longum	otu72820	7.31
Lachnospiraceae	unclassified	unclassified	otu2724175	7.30
Bacteroidaceae	Bacteroides	fragilis	otu4343627	7.30
Bacteroidaceae	Bacteroides	unclassified	otu3600504	7.29
Bacteroidaceae	Bacteroides	unclassified	otu4381553	7.29
Enterobacteriaceae	unclassified	unclassified	otu782953	7.28
Bifidobacteriaceae	Bifidobacterium	unclassified	otu825808	7.28
Bifidobacteriaceae	Bifidobacterium	adolescentis	otu4347159	7.26
Veillonellaceae	Veillonella	dispar	otu4388775	7.26
Veillonellaceae	Veillonella	dispar	otu4453501	7.25
Enterobacteriaceae	unclassified	unclassified	otu2119418	7.23
Pasteurellaceae	Haemophilus	parainfluenzae	otu4477696	7.23
Veillonellaceae	Veillonella	unclassified	otu4458959	7.22
Lachnospiraceae	unclassified	unclassified	otu289734	7.21
Ruminococcaceae	unclassified	unclassified	otu265871	7.20
Enterobacteriaceae	unclassified	unclassified	otu4425571	7.18
unclassified	unclassified	unclassified	otu2714942	7.16
Erysipelotrichaceae	unclassified	unclassified	otu4390365	7.16
Clostridiaceae	unclassified	unclassified	otu4434334	7.15
Streptococcaceae	Streptococcus	unclassified	otu4425214	7.13

Table S1. Early successional taxa, sorted by importance; importance is calculated as the sum of prevalence (scaled) and abundance rank (scaled).

Family	Genus	Species	OTU	Importance
Ruminococcaceae	unclassified	unclassified	otu265871	6.53
Lachnospiraceae	unclassified	unclassified	otu289734	6.53
Lachnospiraceae	unclassified	unclassified	otu2724175	6.52
Bacteroidaceae	Bacteroides	unclassified	otu3600504	6.52
Bacteroidaceae	Bacteroides	unclassified	otu4381553	6.52
Ruminococcaceae	Faecalibacterium	prausnitzii	otu199145	6.51
Lachnospiraceae	unclassified	unclassified	otu3134492	6.51
Verrucomicrobiaceae	Akkermansia	muciniphila	otu4306262	6.51
Bacteroidaceae	Bacteroides	fragilis	otu4343627	6.50
Lachnospiraceae	unclassified	unclassified	otu4472399	6.50
Bifidobacteriaceae	Bifidobacterium	longum	otu72820	6.50
Bifidobacteriaceae	Bifidobacterium	unclassified	otu825808	6.49
Bacteroidaceae	Bacteroides	uniformis	otu328617	6.48
Bacteroidaceae	Bacteroides	uniformis	otu3327894	6.48
Bacteroidaceae	Bacteroides	unclassified	otu4439360	6.48
Lachnospiraceae	Blautia	unclassified	otu292735	6.47
Bacteroidaceae	Bacteroides	ovatus	otu4256470	6.47
Bifidobacteriaceae	Bifidobacterium	adolescentis	otu4347159	6.47
Ruminococcaceae	Ruminococcus	unclassified	otu146554	6.46
Lachnospiraceae	unclassified	unclassified	otu197004	6.46

Table S2.   Late successional taxa, sorted by importance; importance is calculated as the sum of prevalence (scaled) and abundance rank (scaled).