Supplementary Information

Early life dynamics of the human gut virome and bacterial microbiome in infants

Efrem S. Lim[1,2], Yanjiao Zhou[3,4], Guoyan Zhao[1], Irma K. Bauer[3], Lindsay Droit[1,2], I. Malick Ndao[3], Barbara B. Warner[3], Phillip I. Tarr[1,3], David Wang[1,2], Lori R. Holtz[3]

[1]Department of Molecular Microbiology, Washington University School of Medicine, St Louis, Missouri

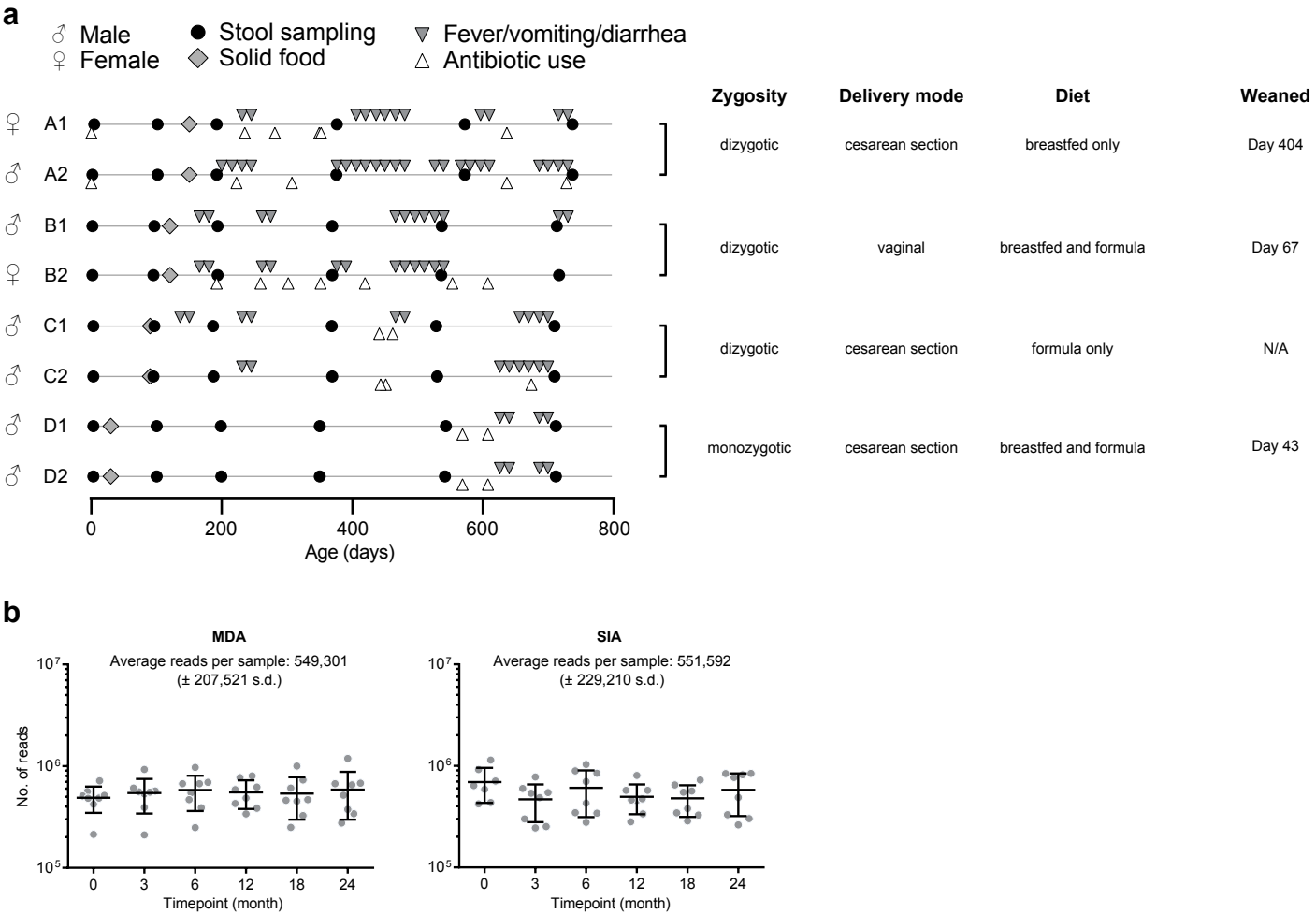
[2] Department of Pathology & Immunology, Washington University School of Medicine, St Louis, Missouri

[3]Department of Pediatrics, Washington University School of Medicine, St Louis, Missouri

[4]Current address: The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut; Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts

Address correspondence to: L.R.H. (Holtz_L@kids.wustl.edu), D.W. (davewang@wusm.wustl.edu)

Supplementary Figure 1

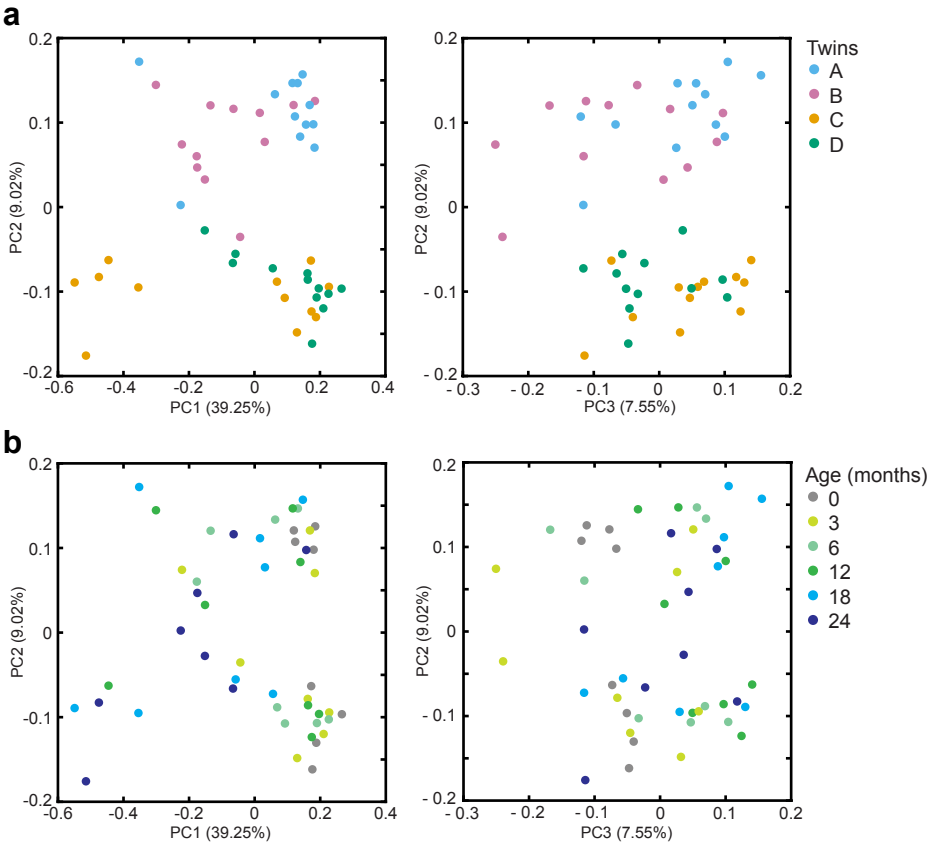


Supplementary Figure 1. Infant cohort description and sequencing depth.

(**a**) Timeline represents stools collected from 8 healthy (4 twin pairs) infants, as indicated by closed circles. Grey diamonds indicate time when solid food was introduced. Reported cases of fever, vomiting or diarrhea are indicated with grey inverted triangles. Reported antibiotic use is indicated with a white triangle.

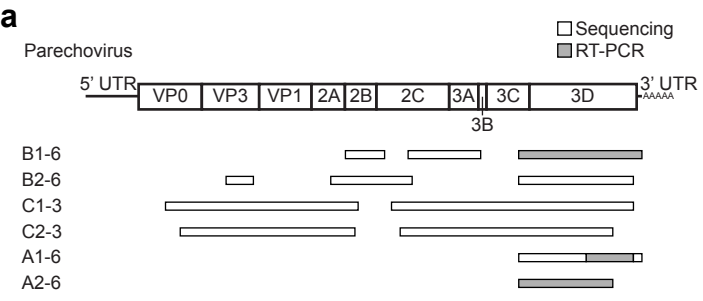
(**b**) Number of sequencing reads obtained for each sample for the MDA method (left) and SIA method (right). SIA, sequence independent DNA and RNA amplification; MDA, multiple displacement amplification; s.d., standard deviation.

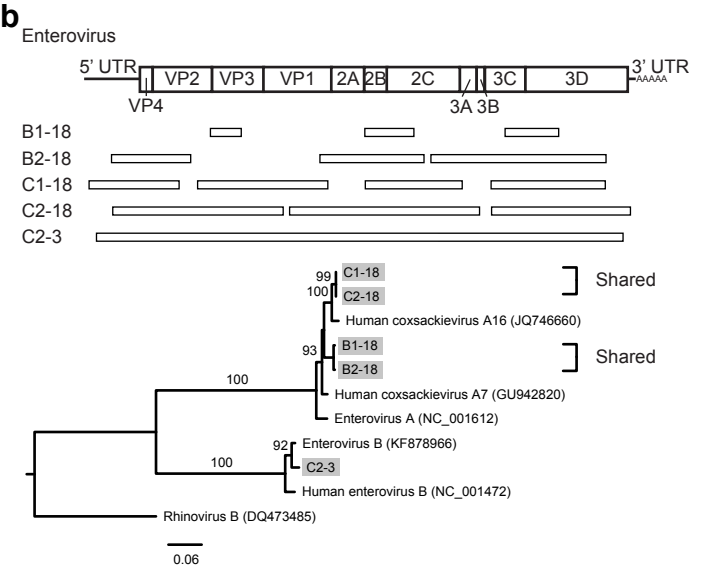
Supplementary Figure 2

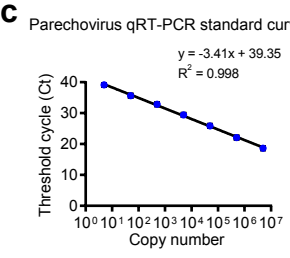


Supplementary Figure 2. Principal coordinate analysis (PCoA) of viromes. PCoA analysis of unweighted bray-curtis distance of virome communities (eukaryotic viruses and bacteriophages; genera) (*n* = 48 sampling time points). The variance explained by each PC is indicated on the axis. Specimens from the same twin pairs are indicated in (**a**), and age from 0 – 24 months is shown in the color gradient in (**b**).

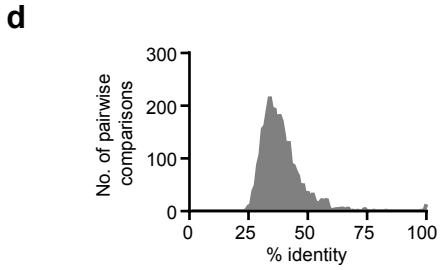
Supplementary Figure 3

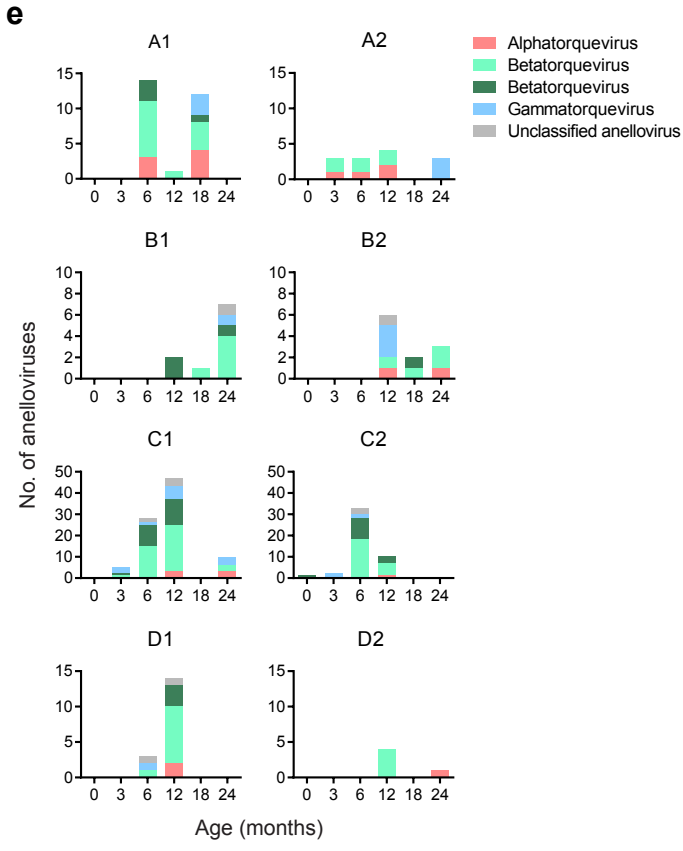






Sample	Sequencing reads	qRT-PCR Ct	qRT-PCR copy no./15mg fecal specimen
B1-6	210	25.69	10,110
B2-6	5,299	25.43	12,058
C1-3	307	24.17	28,191
C2-3	1,245	22.75	73,826
A1-6	31	27.97	2,178
A2-6	0	29.59	730
D2-12	0	29.70	675
D1-12	0	35.09	18





Supplementary Figure 3. Characterization of eukaryotic RNA and DNA viruses.

(**a**) Contigs mapped to the parechovirus genome are shown. Contigs assembled from sequencing reads are indicated in white, and sequence regions obtained from RT-PCR fragments are indicated in grey.

(**b**) Contigs mapped to enterovirus genome are shown. Phylogenetic analysis of enterovirus amino acid sequences from concatenated 2BC and 3CD regions. Maximum likelihood analyses was performed, bootstrap support are indicated above branches.

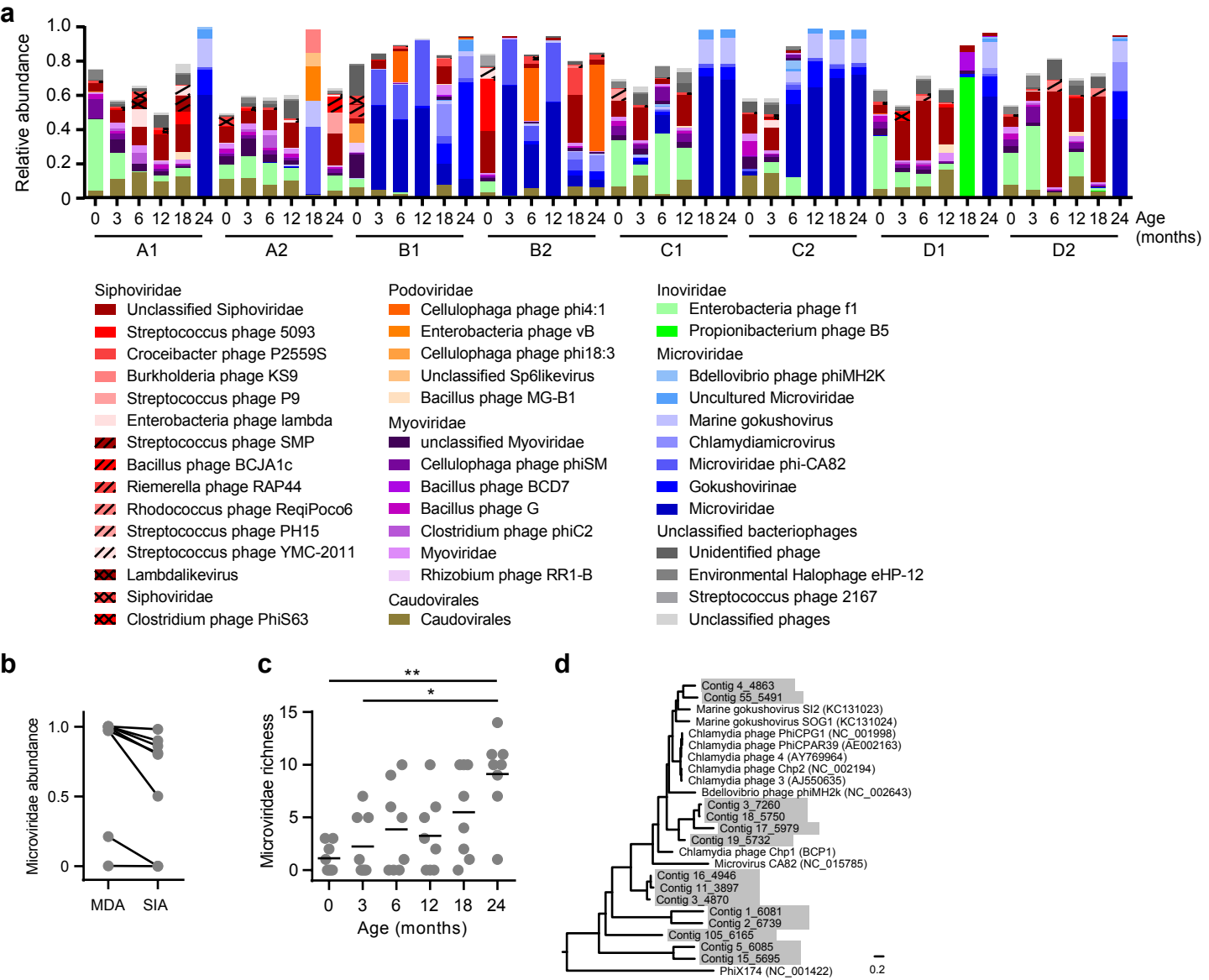
(**c**) Standard curve of the qRT-PCR assay for human parechovirus was performed in triplicate. Error bars indicate the standard deviation. Best-fit line, equation and R-squared values are shown. Comparison of the number of parechovirus sequencing reads to qRT-PCR assay threshold cycle (Ct) for parechovirus-positive samples and the viral load measurement (copy number/15mg of fecal specimen) are shown on the right.

(**d**) Histogram plot of the pairwise % identity comparisons of the anellovirus genome contigs.

(**e**) Distribution of anelloviruses from each infant is shown. The genera distribution is colored as indicated.

(**f**) Concordance between PCR assay and in silico sequence mapping was assessed using a 2x2 contingency table. The PCR assay was based on screening results of an alphatorquevirus, betatorquevirus and gammatorquevirus.

Supplementary Figure 4



Supplementary Figure 4. Analysis of the bacteriophage community.

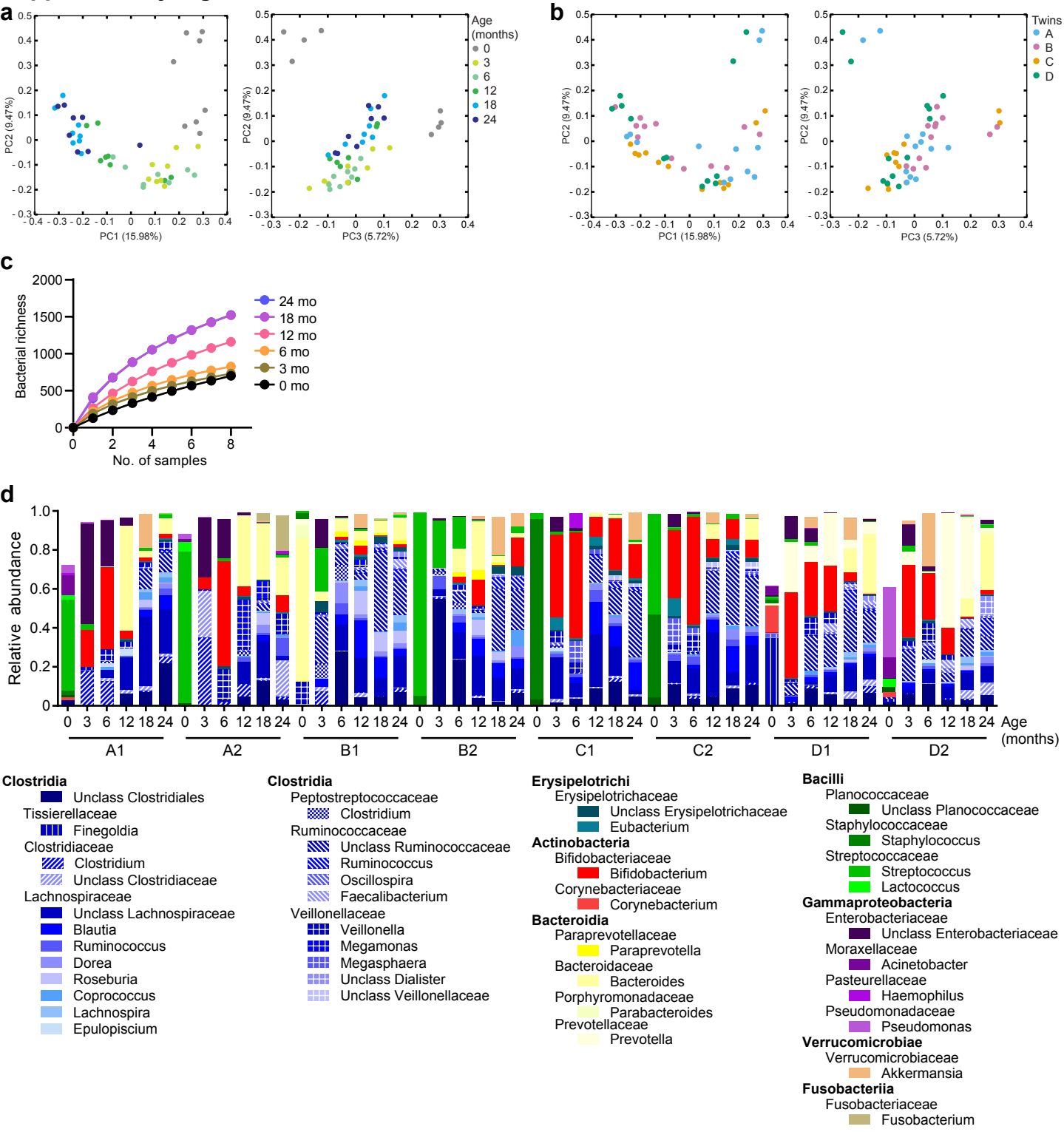
(**a**) Relative abundance of bacteriophages genera. Only bacteriophage taxa with greater than 0.05 relative abundance is shown.

(**b**) Comparison of Microviridae abundance of the MDA and SIA methods from specimens at 24 months age is shown (*n* = 8 infants). The abundance observed for each sample is indicated with a line connecting the paired method.

(**c**) Richness of Microviridae bacteriophage species is shown (*n* = 8 infants). Statistical significance was assessed by Wilcoxon test (paired, non-parametric); * P = 0.01 - 0.05, ** P <0.01.

(**d**) Maximum likelihood phylogenetic analyses of the Microviridae bacteriophages identified in this study are shown. Phylogenetic relationships were inferred from the amino acid sequence alignment of the major capsid protein.

Supplementary Figure 5



Supplementary Figure 5. Analysis of the bacterial community.

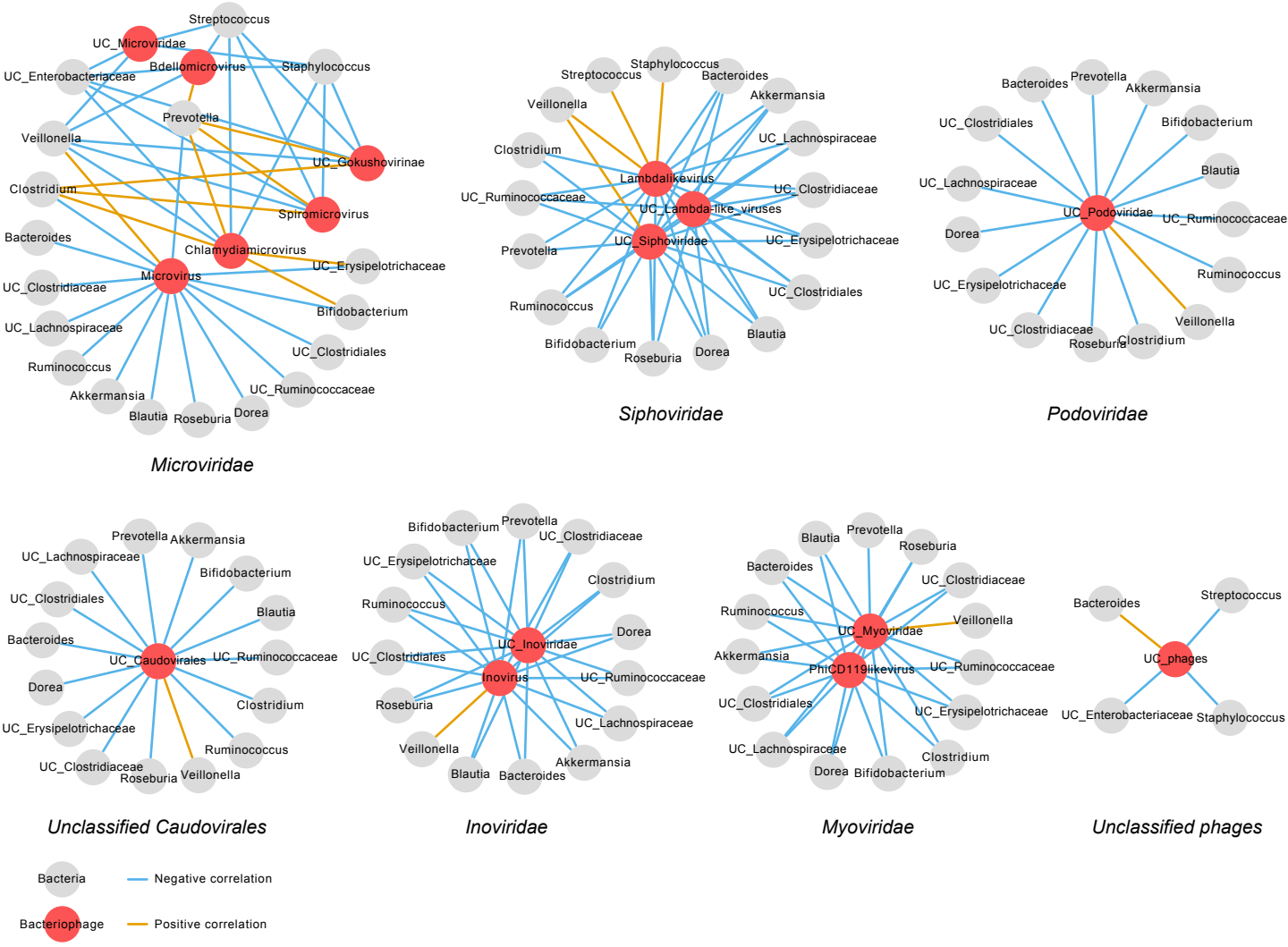
(**a**) PCoA plots of bacterial unifrac distance matrices (*n* = 48 sampling time points). The variance explained by each PC is indicated on the axis. Color gradient represents community progression with age.

(**b**) PCoA plots of unifrac distance matrices (*n* = 48 sampling time points). Twin pairs are colored as indicated.

(**c**) Rarefaction curves of bacterial richness (OTUs) versus an increasing number of specimen subsamplings with replacement are shown. The curves represent the average of 500 iterations at each depth of samples.

(**d**) Relative abundance of the 40 most abundant bacterial genera is shown.

Supplementary Figure 6



Supplementary Figure 6. Bacteriophage-bacteria network analysis.
Network between bacteriophage genera (red node) and bacteria genera (grey node) is plotted by indicated bacteriophage family. Linear mixed model was applied to identify significant changes in relative abundance over time. Negative correlations are shown with blue lines and positive correlations are shown in orange lines. UC, unclassified.