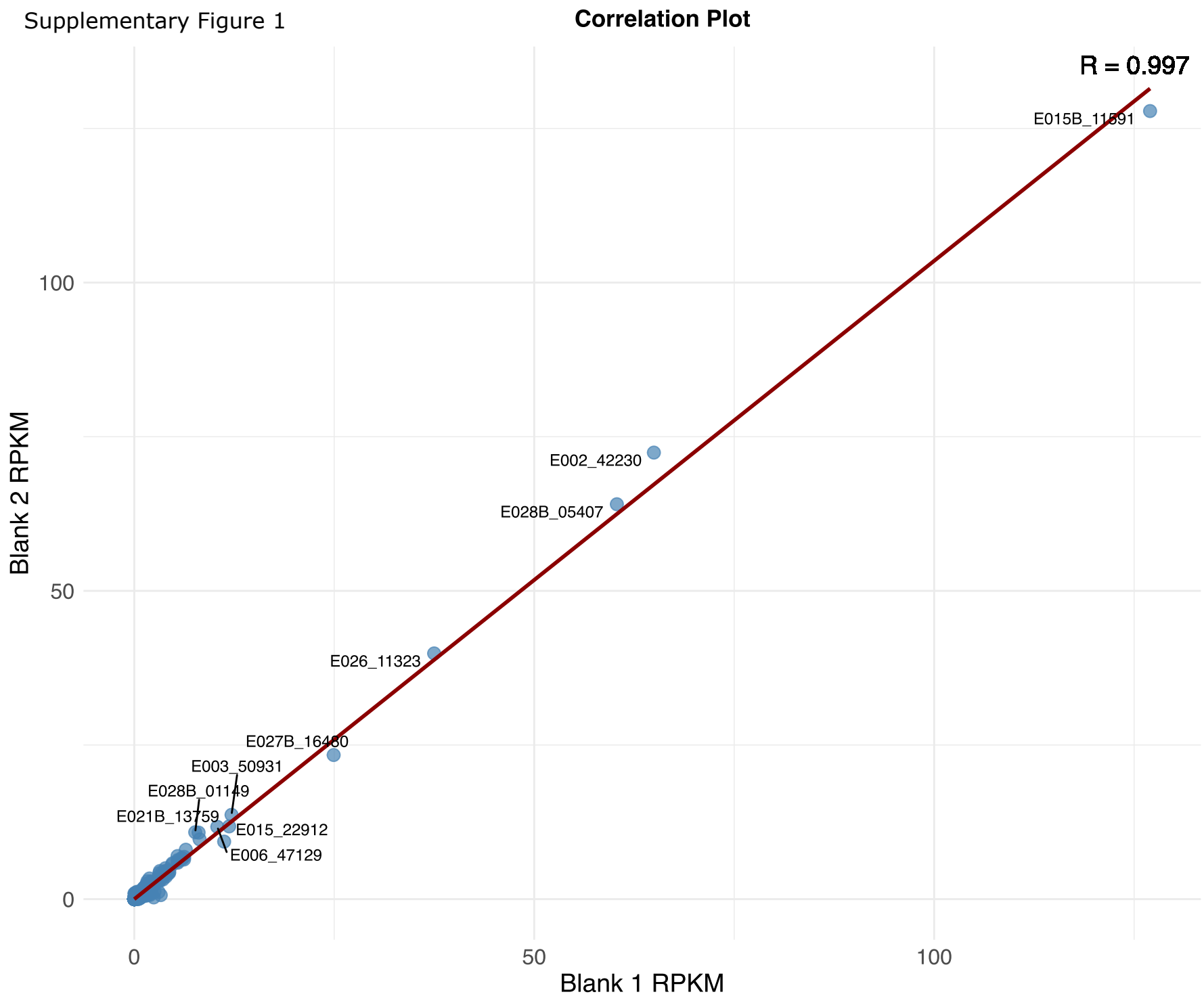
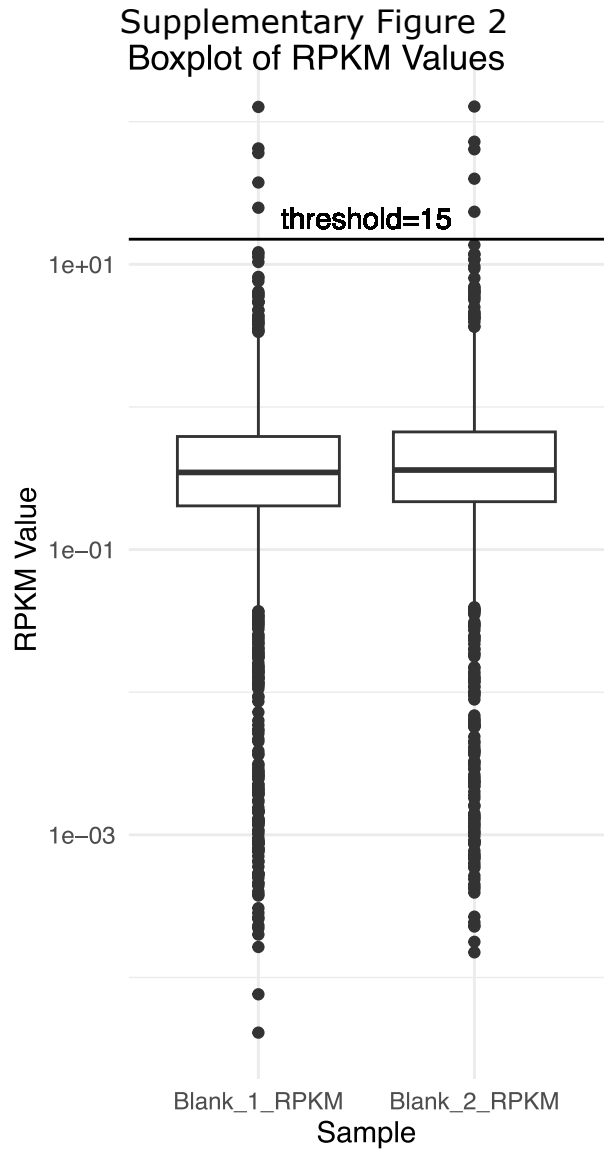
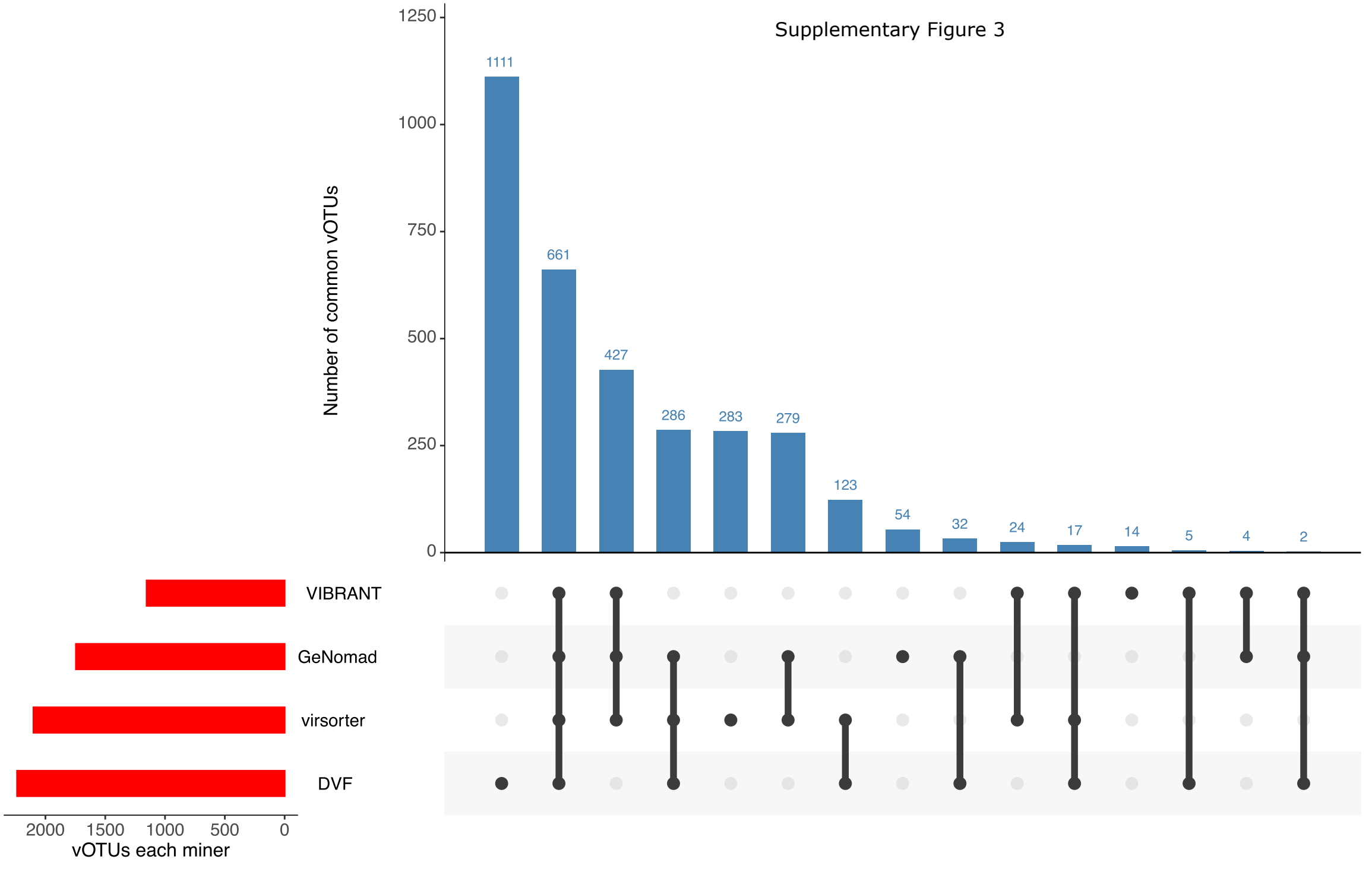
Supplementary Figure 1



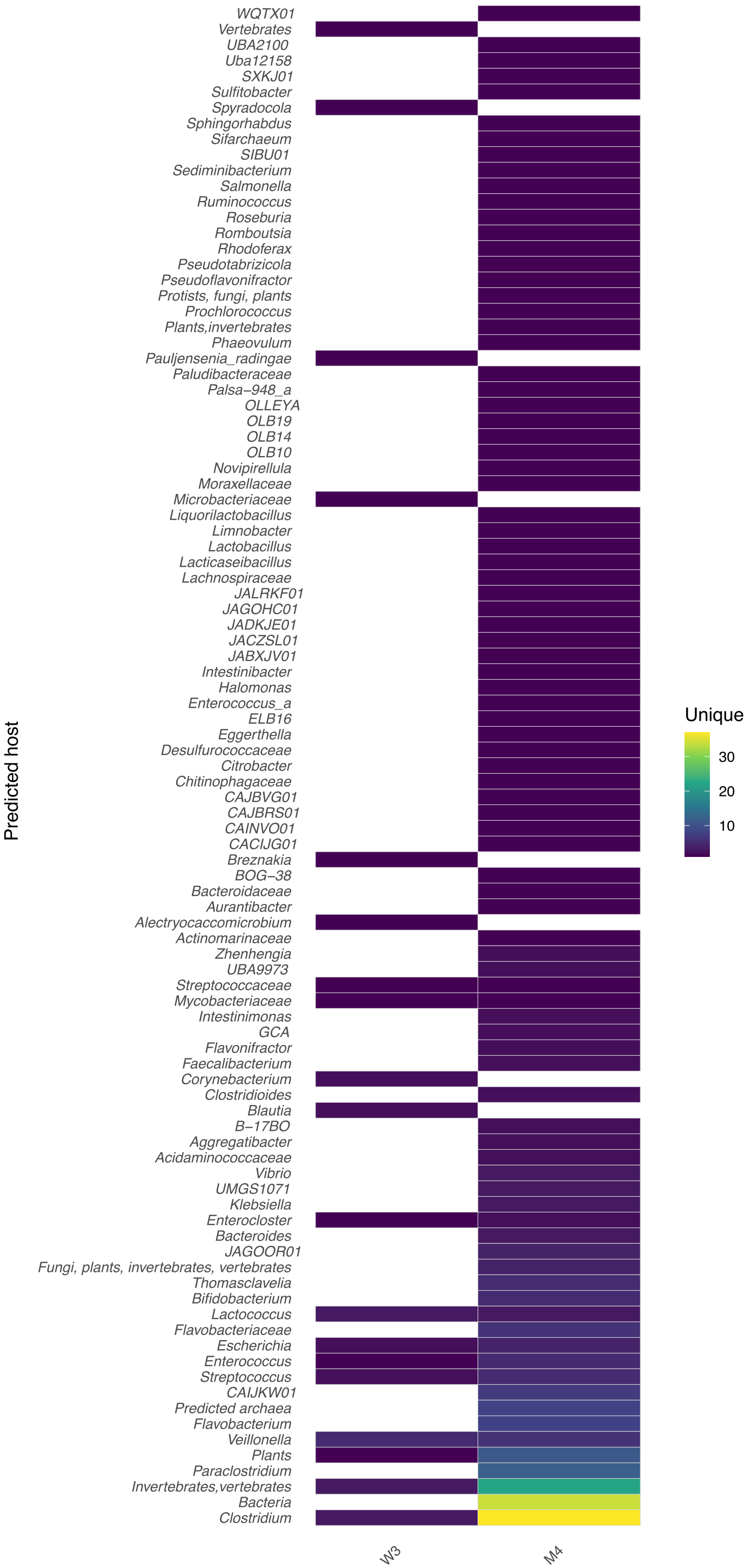

Supplementary Figure 2
Boxplot of RPKM Values
threshold=15
1e+01
1e−01
1e−03
RPKM Value
Blank_1_RPKM
Blank_2_RPKM
Sample

Supplementary Figure 3

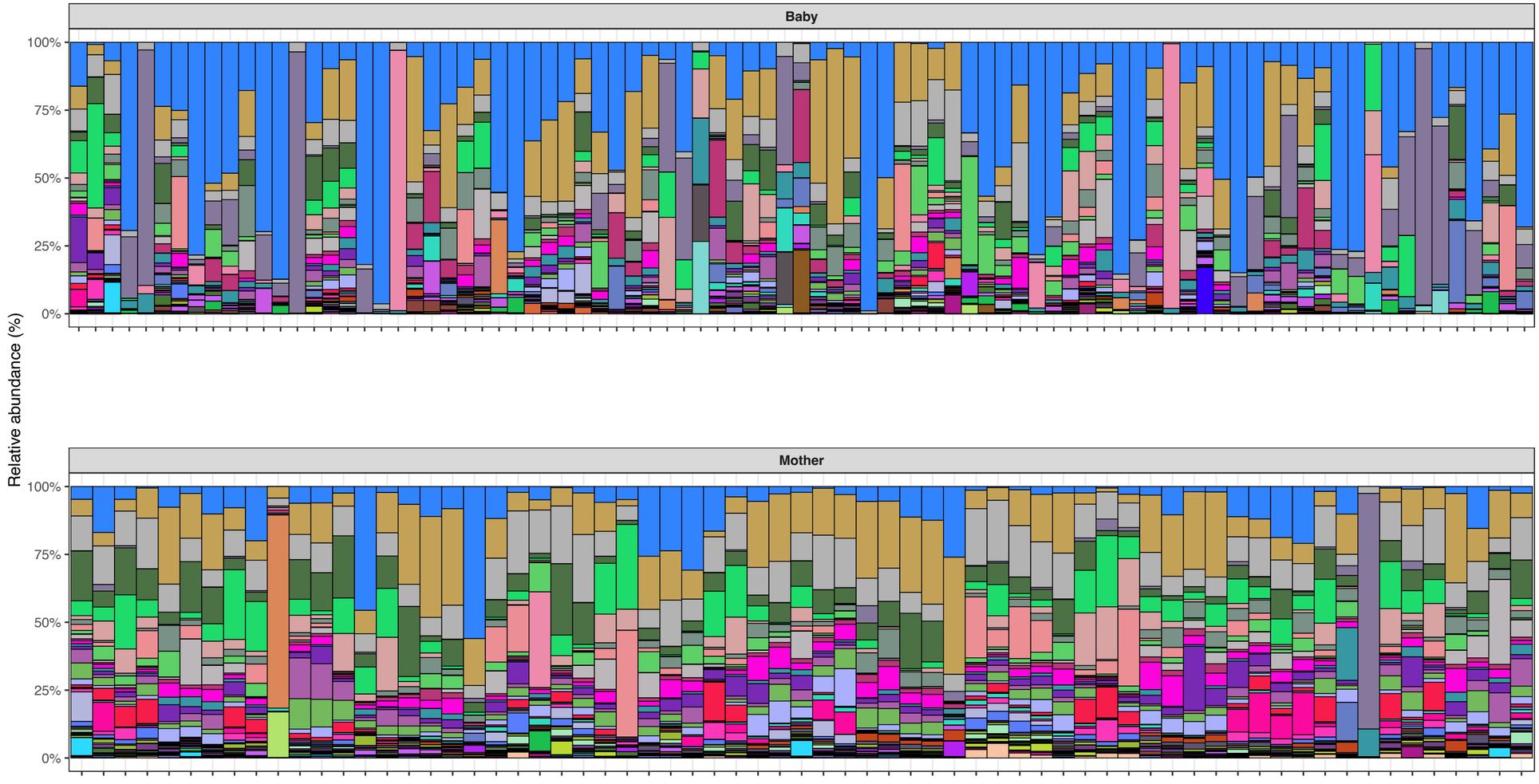


Supplementary Figure S4

Infant virome hosts at early timepoints (W3 & M4)



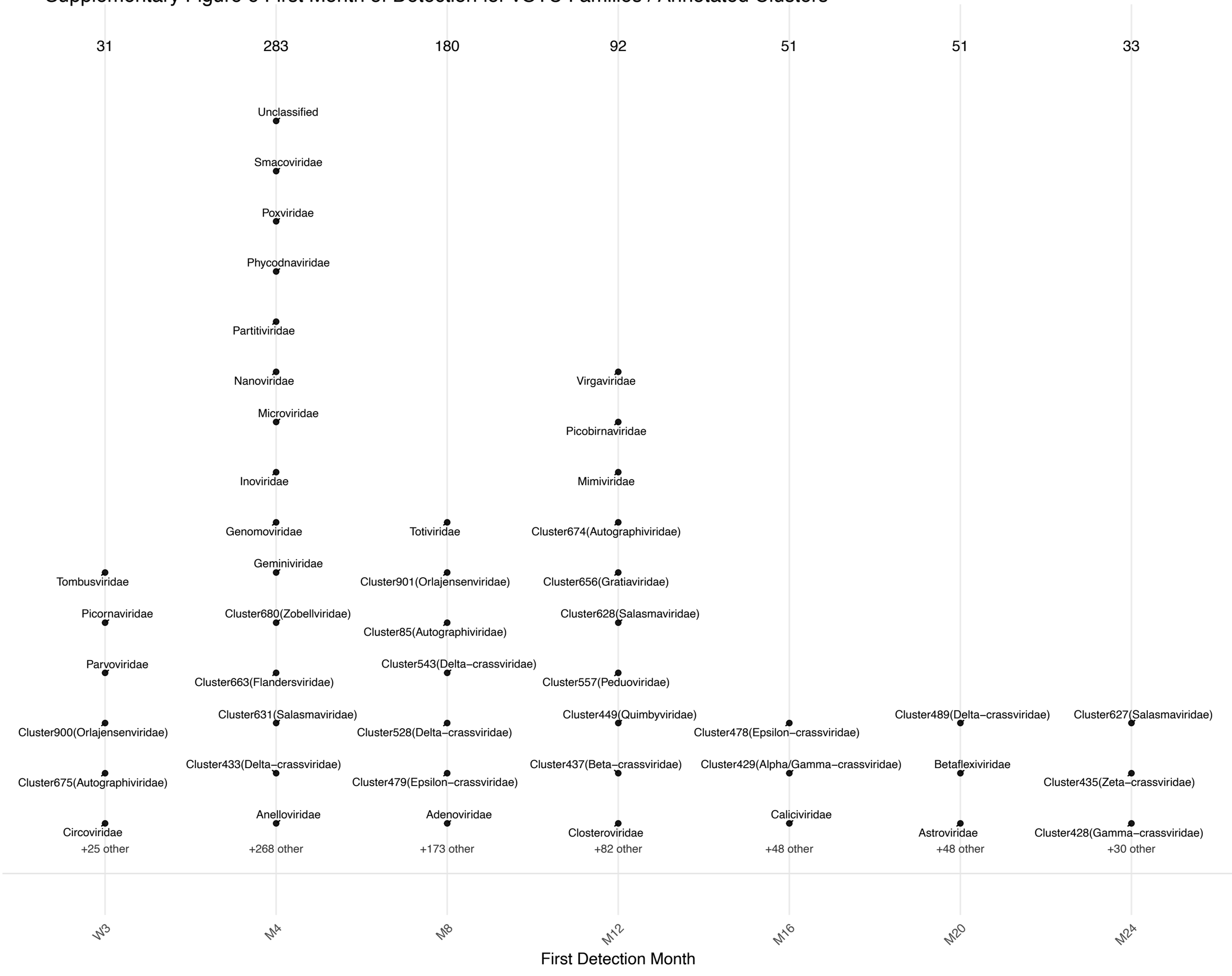
Supplementary Figure 5 :Top 50 genera in Baby vs Mother samples (Bracken, 100 % stacked)

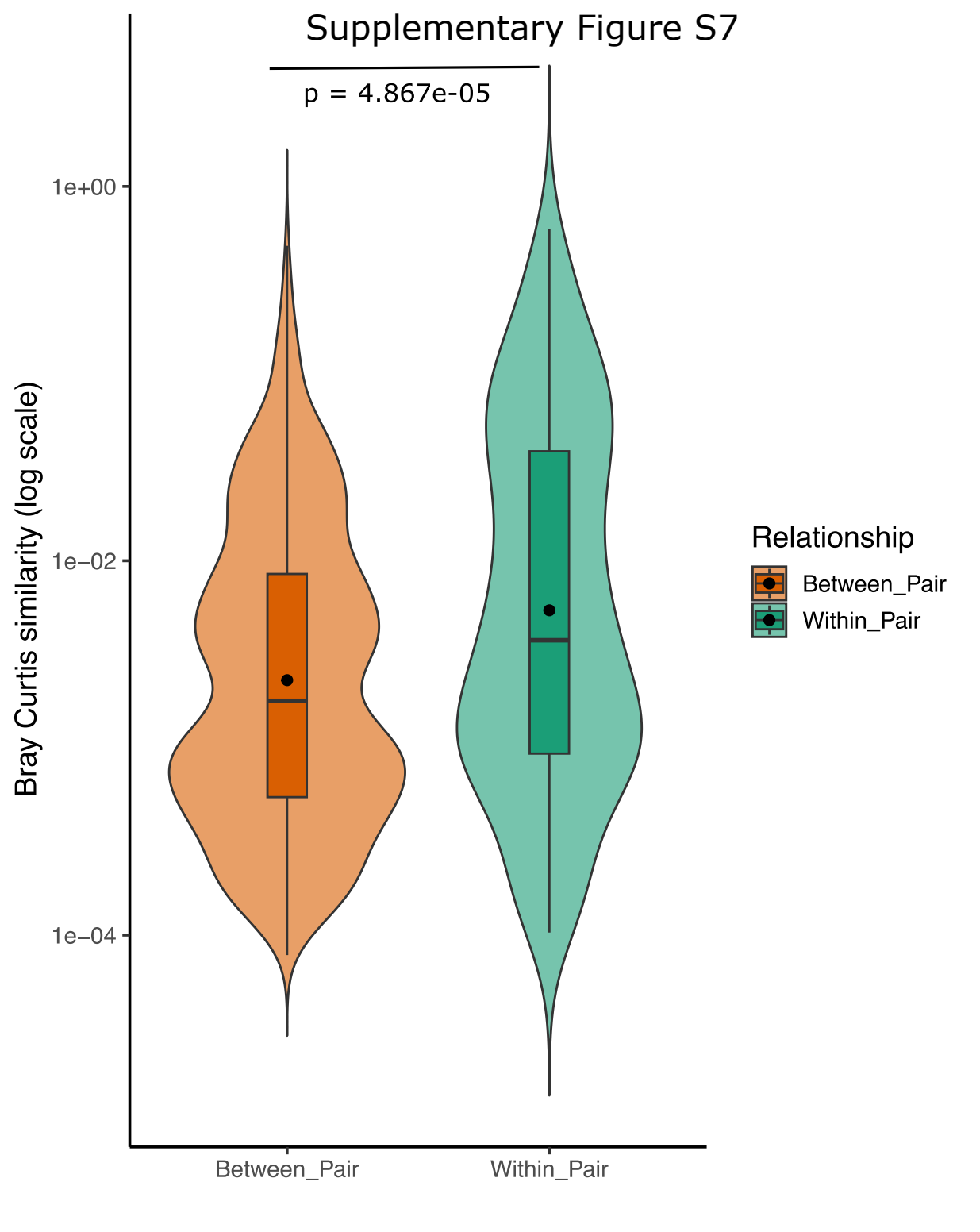


Genus

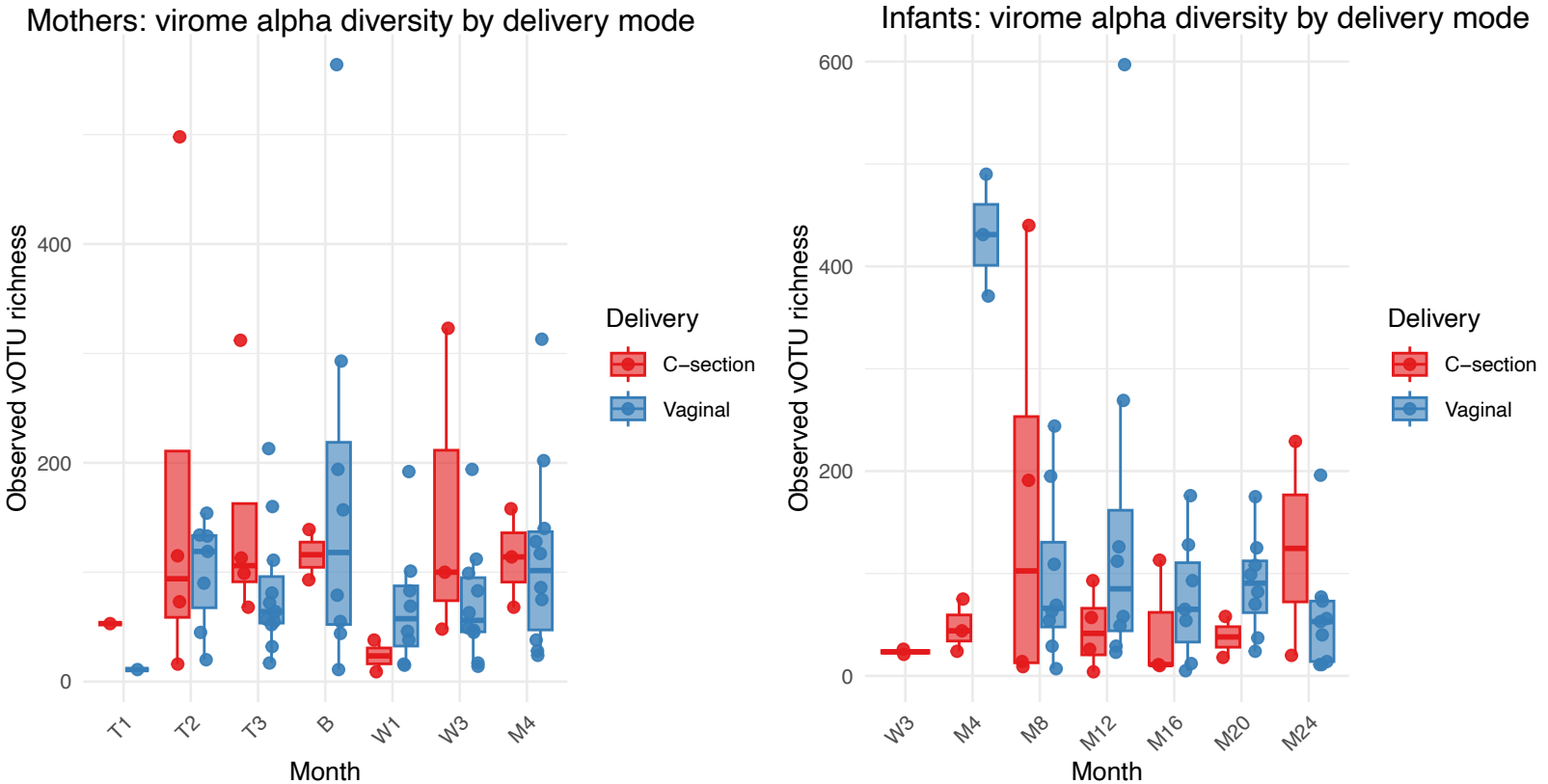
- Bifidobacterium
- Blautia
- Other
- Escherichia
- Faecalibacterium
- Bacteroides
- Phocaeicola
- Anaerostipes
- Akkermansia
- Collinsella
- Enterococcus
- Agathobacter
- Mediterraneibacter
- Anaerobutyricum
- Streptococcus
- Ruminococcus
- Roseburia
- Coprococcus
- Dorea
- Alistipes
- Veillonella
- Klebsiella
- Clostridium
- Methanobrevibacter
- Parabacteroides
- Segatella
- Vescimonas
- Thomasclavelia
- Faecalibacillus
- Eubacterium
- Ruthenibacterium
- Simiaoa
- uncultured
- Enterocloster
- Flavonifractor
- Eggerthella
- Lacticaseibacillus
- Sellimonas
- Actinomyces
- Staphylococcus
- Lachnospira
- Hungatella
- Dialister
- Schaalia
- Subdoligranulum
- Intestinibacter
- Phascolarctobacterium
- Vipetofemvirus
- Carjivirus
- Burzaovirus
- Sepunavirus

Supplementary Figure 6 First Month of Detection for vOTU Families / Annotated Clusters

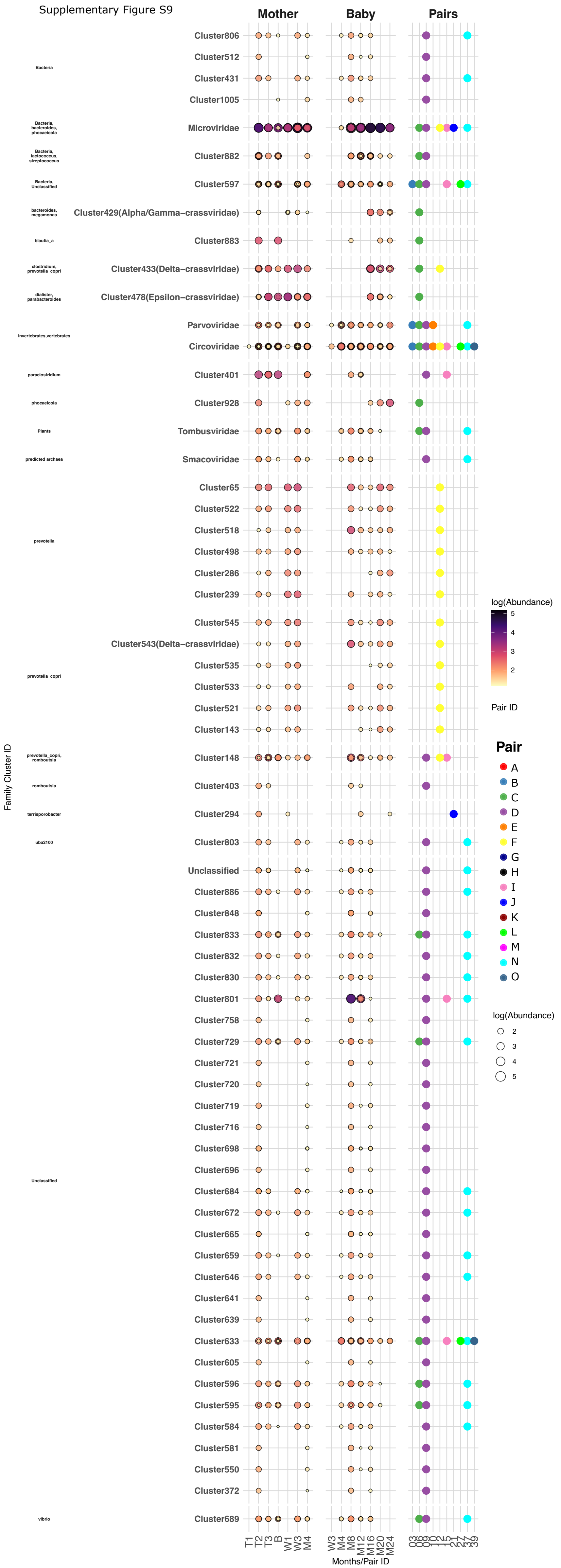



Supplementary Figure S7
p = 4.867e-05
1e+00
1e−02
1e−04
Bray Curtis similarity (log scale)
Between_Pair
Within_Pair
Relationship
Between_Pair
Within_Pair

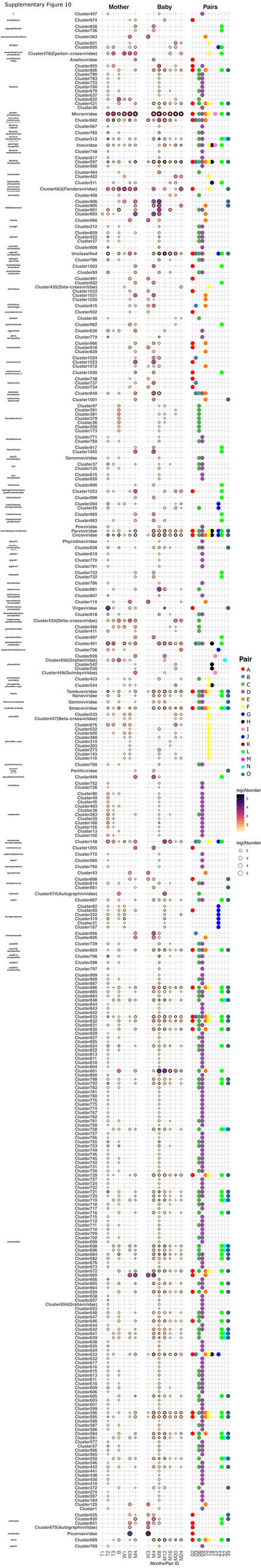
Supplementary Figure S8



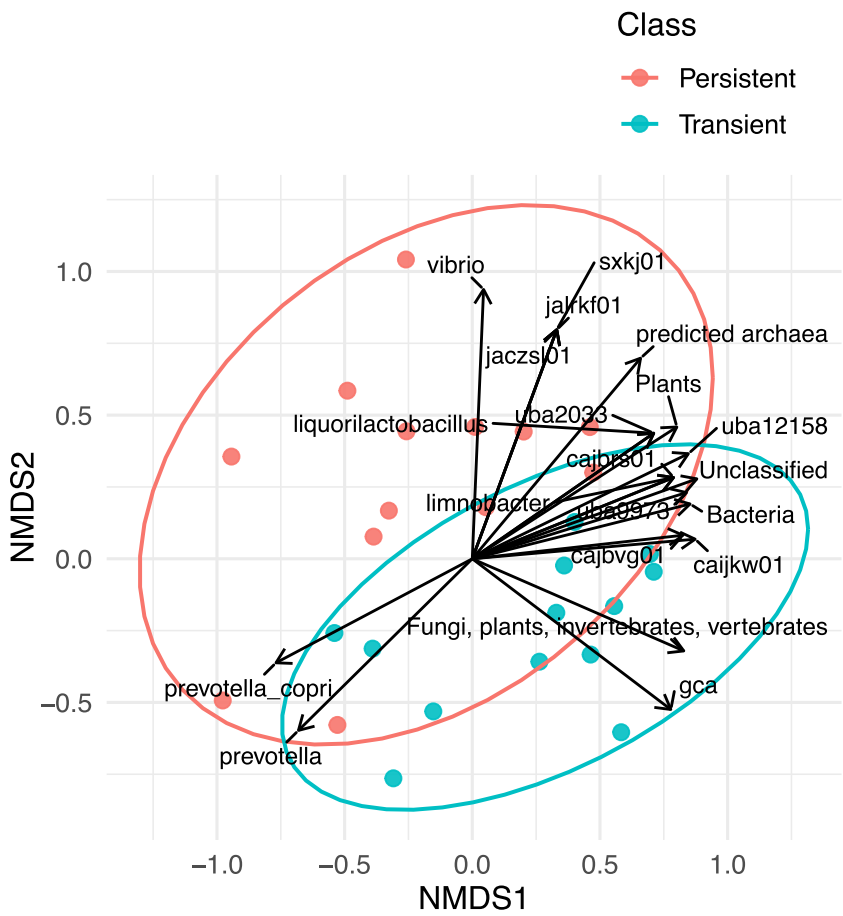
Supplementary Figure S9

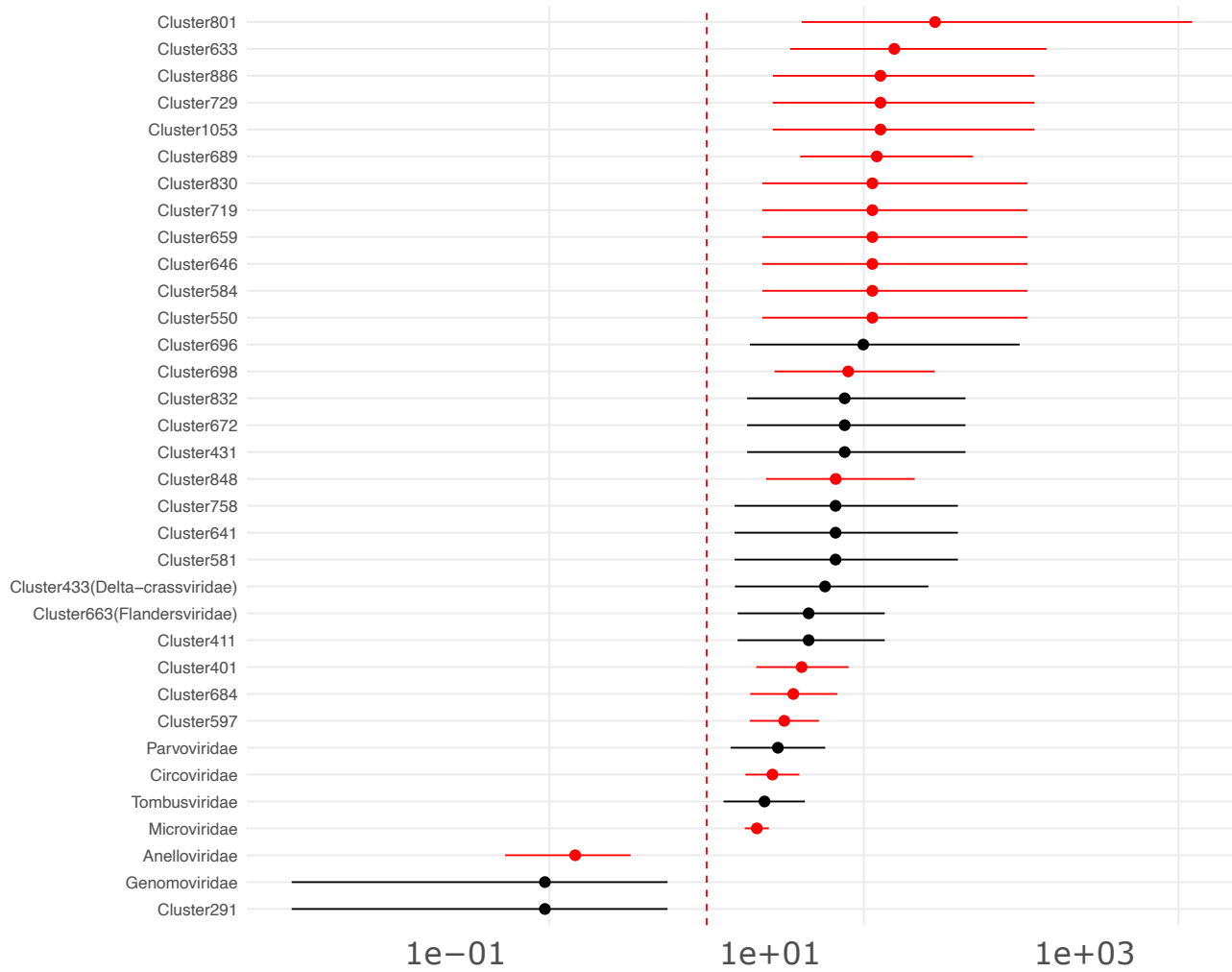


Supplementary Figure 10

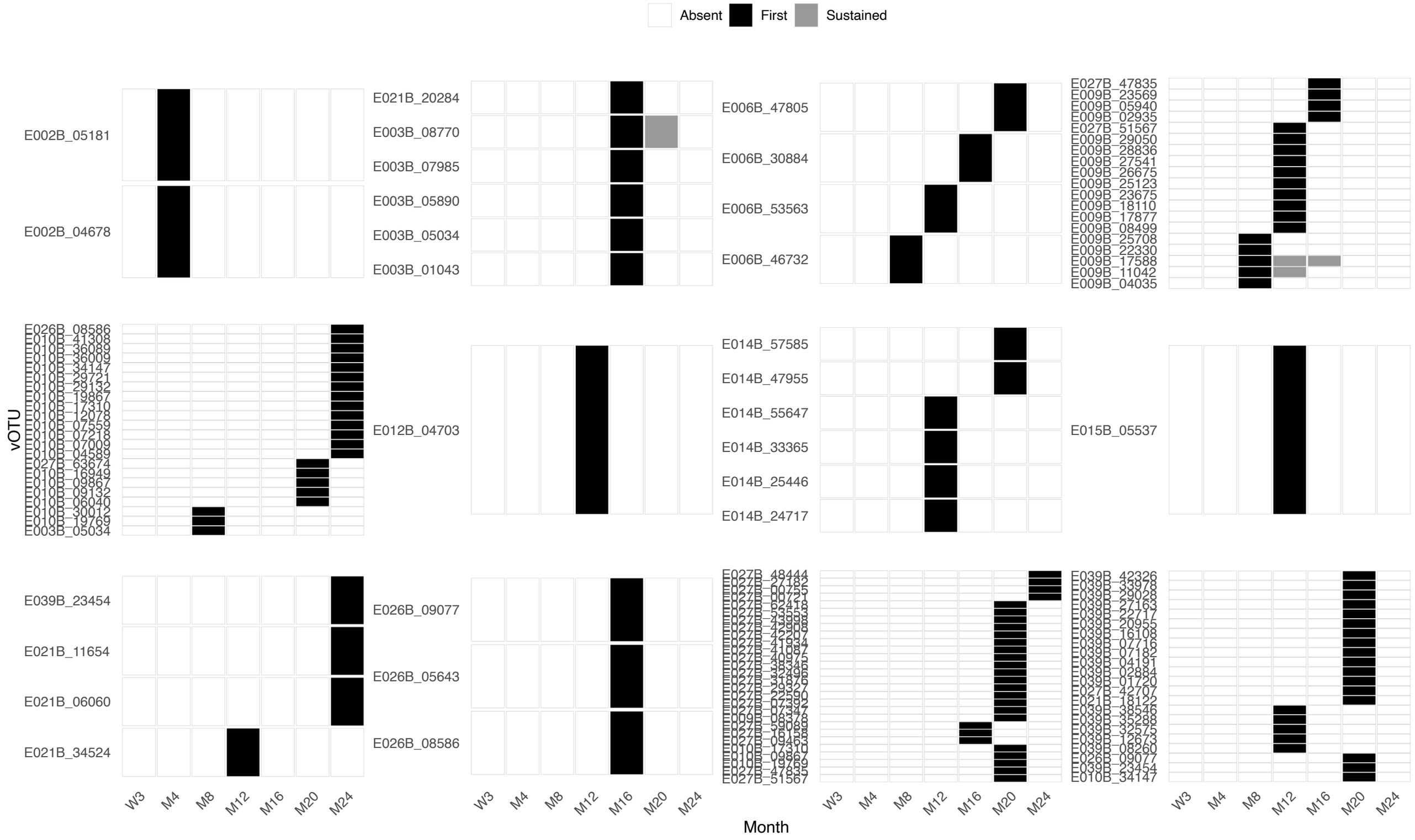


Supplementary_Figure11

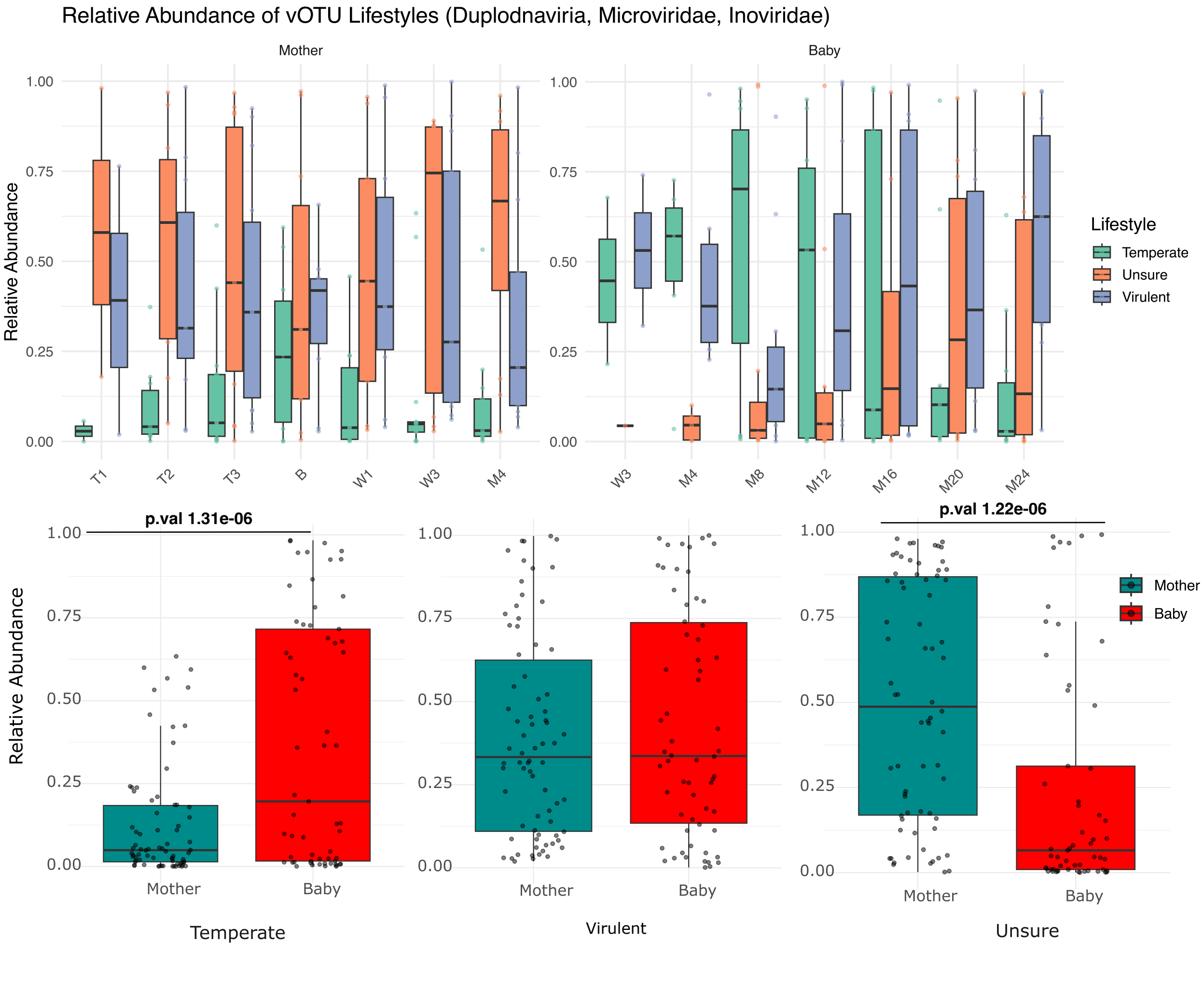




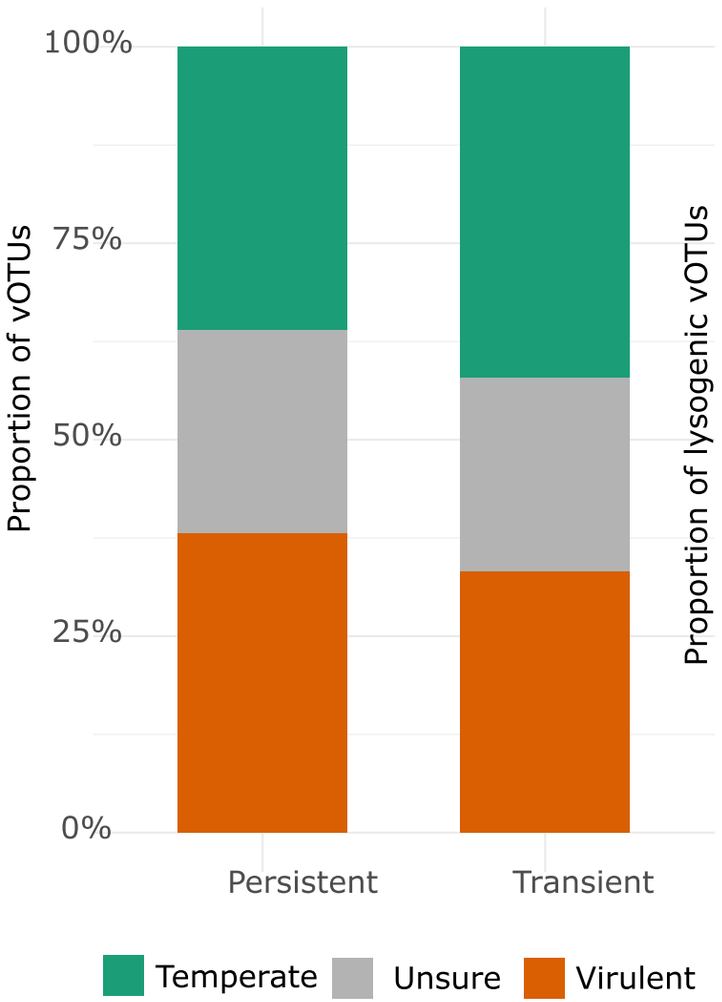
Supplementary Figure 12 :Anelloviridae vOTUs per Infant: First Detection (black) and Sustained Presence (grey)

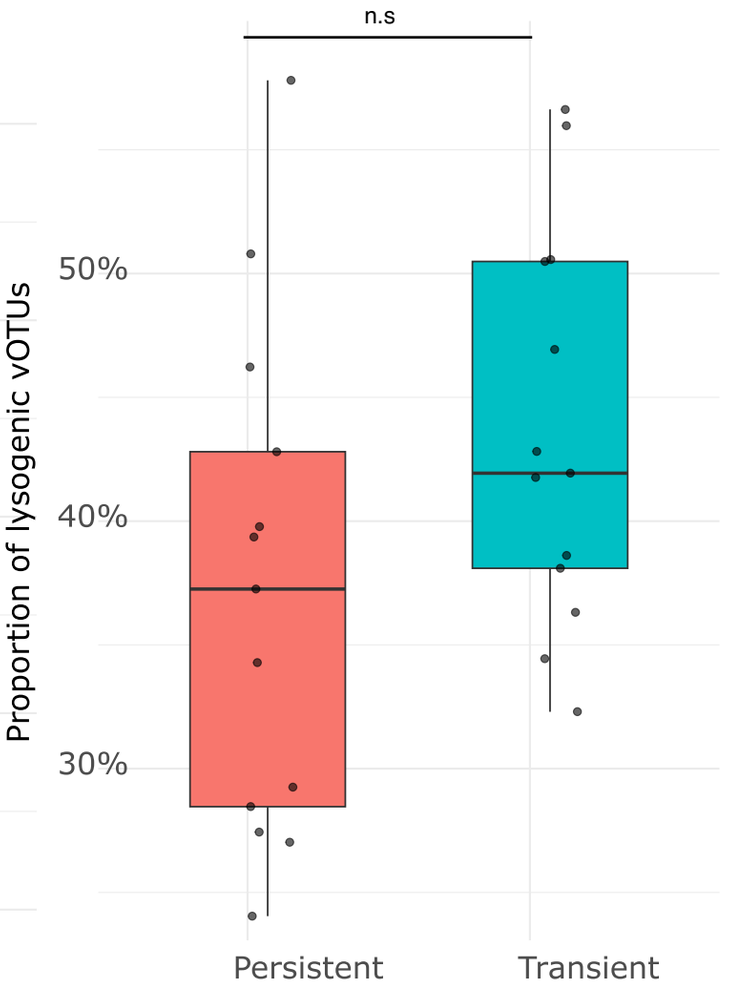


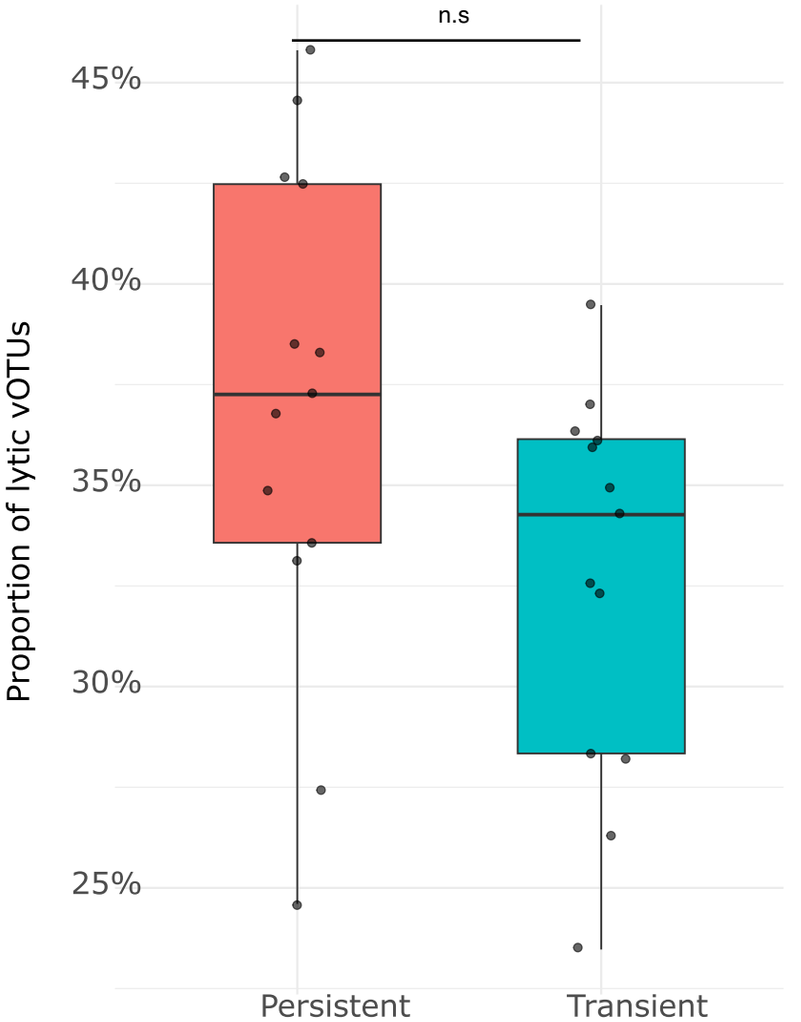
Supplementary Figure 13



Supplementary Figure 14







SUPPLEMENTARY FIGURE CAPTIONS

Supplementary Figure 1: Correlation of vOTU abundances (RPKM) between the two blank controls. Each point represents a vOTU detected in both blanks, with selected high abundance vOTUs labelled. The strong correlation (Pearson's R = 0.997) supports their use for identifying contaminants introduced during sample processing. vOTUs that were highly abundant in blanks and detected in biological samples were considered likely processing contaminants and removed from the dataset.

Supplementary Figure 2: Boxplots showing the distribution of vOTUs RPKM values across two negative control (blank) libraries. Each point represents a vOTU detected in the blanks. The horizontal line indicates the selected signal-to-noise threshold of 15 RPKM, above which vOTUs were considered reliably detectable. vOTUs with abundances consistently below this threshold across all samples were classified as low abundance and excluded from downstream analysis.

Supplementary Figure 3: Upset plot showing the overlap of viral contigs identified by four independent virus detection tools geNomad, VirSorter2, VIBRANT, and DeepVirFinder after dereplication. Bars indicate the number of viral contigs shared across different combinations of tools, with counts shown above each intersection. Horizontal bars represent the total number of viral contigs detected by each individual tool. Additional 209 vOTUs refined by phables assembly is not shown here.

Supplementary Figure 4 : Heatmap showing the number of unique vOTUs detected in infant early life gut samples at week 3 (W3) and month 4 (M4), grouped by predicted bacterial host taxa. A total of 636 vOTUs were identified across these early-life time points, with predicted hosts spanning 81 bacterial taxa, including members of *Bifidobacteriaceae*, *Enterobacteriaceae*, *Bacteroidaceae*, *Clostridiaceae*, *Peptostreptococcaceae*, and *Ruminococcaceae*. Colour intensity reflects the number of unique vOTUs associated with each predicted host.

Supplementary Figure 5: Stacked bar plots showing the relative abundance of the top 50 bacterial genera in infant (top) and maternal (bottom) samples, profiled using Bracken-based taxonomic assignment at genus level. Each bar represents an individual sample, and colours denote bacterial genera, with remaining low-abundance taxa grouped as "Other." Early infant samples specifically M4 samples support the presence of host genera (for instance; *Bifidobacterium* and *Escherichia*) for which the vOTUs were seen in our cohort given in Supplementary Figure 4.

Supplementary Figure 6: Plot showing the first time point of detection for viral families and annotated viral clusters across infant gut samples, spanning week 3 (W3) to month 24 (M24). Each point denotes a known viral family positioned by the earliest sampling month in which it was detected and other denotes the number of family level clusters. Numbers on the top indicate newly detected families per time point, revealing a rapid expansion of virome diversity with 283 new families/clusters detected in M4 followed by stabilisation later in life.

Supplementary Figure 7: Violin plots showing distribution of Bray Curtis similarity between pairs and within pairs (*Wilcoxon rank sum test; p = 4.867e-05*). The dot within the boxplot depicts the mean similarity score.

Supplementary Figure 8: Boxplots showing observed vOTU richness in maternal (left) and infant (right) gut viromes across sampling time points, classified by delivery mode vaginal delivery (blue) versus C-section (red). Individual points represent samples, The bimodal distribution of alpha diversity of infant's virome at M4 is solely explained by the delivery mode, with vaginally delivered infants consistently exhibited higher alpha diversity compared to those delivered via C-section, however statistical testing was underpowered due to limited sample size (n = 6).

Supplementary Figure 9: Bubble plot showing the log-transformed abundance (colour and size) of strongly shared persistent vOTUs grouped at the family rank, across timepoints, for mothers (left) and infants (middle). Unclassified viruses were indicated with family-level cluster numbers. Pair-specific colour coding on the right denotes in which mother-infant pairs these vOTUs were detected. Taxonomic information of predicted host is provided in small font either at the ICTV broad host range level or at the genera level where available. When host assignment was not possible using our methods "Unclassified" was used.

Supplementary Figure 10: Bubble plot showing the log-transformed abundance (colour and size) of all shared vOTUs except the strongly shared vOTUs shown above, grouped at the family rank, across timepoints, for mothers (left) and infants (middle). Unclassified viruses were indicated with family-level cluster numbers. Pair-specific colour coding on the right denotes in which mother-infant pairs these vOTUs were detected. Taxonomic information of predicted host is provided in small font either at the ICTV broad host range level or at the genera level where available. When host assignment was not possible using our methods "Unclassified" was used.

Supplementary Figure 11: Forest plot of vOTU's predicted host family associations with persistence versus transience (Fisher's OR, log scale, FDR < 0.05) with 95% confidence intervals are shown.

Supplementary Figure 12: Heat map showing the transient appearance and immediate clearance of vOTUs from *Anelloviridae* family in infants across each pair.

Supplementary Figure 13: Top panel showing relative abundance of phage lifestyle (Temperate, Virulent, Unsure) across months for mothers (left) and infants (right). Each point represents one sample. Bottom panel showing comparison of the relative abundance of each phage lifestyle category between maternal and infant viromes pooled across time points. Each point represents a sample. Differences between mothers and infants were assessed using the linear mixed-effects models with Pair as a random effect (Relative Abundance ~ Type * Viral Lifestyle + (1|Pair)); p values Temperate = 1.31e-06, Virulent = 0.45 (n.s), Unsure = 1.22e-06

Supplementary Figure 14: Stacked bar plot showing the proportion of vOTUs classified as temperate, virulent, or of uncertain lifestyle within persistent and transient virome categories(left). Boxplots comparing the proportion of temperate (middle) and obligate virulent (right) vOTUs between persistent and transient viromes. Statistical comparisons were performed using the Wilcoxon rank-sum test, with non-significant differences indicated (n.s.).