

Preferential sharing of virus populations between mothers and infants in early life

Revathy Krishnamurthy

Quadram Institute Bioscience

Ryan Cook

Quadram Institute Bioscience

Alise Ponsero

Quadram Institute Bioscience

Leon Mijic

University of Birmingham

Sumeet Tiwari

Quadram Institute Bioscience

Judit Talas

Quadram Institute Bioscience

Xena Dyball

Quadram Institute Bioscience

Hannah Pye

Quadram Institute Bioscience

Andrea Telatin

Quadram Institute Bioscience

David Baker

Quadram Institute Bioscience

Sarah Philips

Quadram Institute Bioscience

Rachel Watt

Quadram Institute Bioscience

Sarah Hughes

Quadram Institute Bioscience

George Savva

Quadram Institute Bioscience

Antonieta Hayhoe

Quadram Institute Bioscience

Lindsay Hall

Quadram Institute Bioscience

Evelien Adriaenssens

evelien.adriaenssens@quadram.ac.uk

Quadram Institute Bioscience

Article

Keywords:

Posted Date: August 26th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9843599/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Maternal contributions to the establishment of the early infant microbiome play a crucial role in shaping the infant's microbiome for life. Yet this maternal contribution in shaping the infant's virome is not well-understood. Here, we investigated the extracellular viral fraction of the human faecal microbiome, including DNA and RNA viruses, using longitudinal samples collected from 15 healthy infant-mother pairs from pregnancy till first two years of infant development. By employing an individual mother-infant-centric approach combined with microdiversity analyses, we show that the virome of each mother-infant pair is highly distinct. Sharing of viruses within pairs is more often transient than persistent. Nonetheless, we show that a subset of vOTUs is persistently shared between mother and infants. We observed that vOTUs persistent in mothers and detected in babies undergo stronger positive selection. Interestingly, vOTUs belonging to the Class *Microviricetes* were significantly enriched among persistent viruses, whereas *Anelloviridae* were enriched among transient ones. Further, we show preferential sharing of specific viral families between pairs across the persistent and transient maternal virome pool. Our findings reveal that mother–infant virome sharing is not random but structured by specific transient and persistent viral families.

Introduction

The establishment of the gut microbiome, including the virome, during the first 1,000 days of life is crucial for healthy infant growth and development^{1–3}. Extensive microbial colonisation of newborns begins immediately after birth and is primarily shaped by delivery mode, maternal microbiome, breast feeding, antibiotic exposure, gestational age and various environmental exposure^{4–6}.

The human gut virome, dominated by bacteriophages (phages), harbors an estimated $10^9 - 10^{10}$ virus like particles (VLPs) per gram of faeces^{7–10}. Bacteriophages play a pivotal role in maintaining gut ecosystem homeostasis by modulating bacterial populations through complex interaction dynamics^{11–15}. The gut virome shapes mucosal immunity^{16,17} and its perturbation is linked to chronic diseases^{18–20}. Vertical transmission of specific viruses, including a few bacteriophages and certain pathogenic viruses from mother to infant has been shown^{21–23}. However, accumulating evidence suggest that an infant's virome is highly distinct from that of their mother with rapid compositional shifts during the first year of life due to changes in lifestyle and environment^{24–28}. The infant gut virome is comprised of passive temperate phages that are integral part of their bacterial host and spontaneously induced temperate phages along with virulent phages and other eukaryotic viruses forming the extracellular virome fraction²⁹. Despite its important role in regulating the microbiome, much less is understood about how the extracellular gut virome is formed and shaped during early infant development through maternal influence.

In this study, we investigated the extracellular faecal virome consisting of both DNA and RNA viruses of healthy mothers during pregnancy and in the first four months postpartum and that of their infants

during the first two years of life, using shotgun metagenomics combined with virome composition and microdiversity analysis. We show that the infant virome is shaped by a dynamic influx of unique, transiently present viruses over time, alongside a smaller but consistent set of strongly shared, persistent viruses shared with the mother. The extent of virome sharing between infants and mothers varied widely across the cohort, revealing heterogeneity in sharing patterns across early virome establishment. Further, we see preferential sharing of certain viral families associated with specific hosts between baby-mother pairs across the persistent and transient maternal virome pool.

Results

The human gut extracellular virome is predominantly composed of members of the realm *Duplodnaviria*

We identified 3,531 viral operational taxonomic units (vOTUs) each represented by its largest genome, treated as sequence defined ecological units rather than formal ICTV species) across all mother-infant pairs. The abundance distribution was highly skewed toward members of the former realm “*Monodnaviria*” (*Efunaviria*, *Floreoviria* and *Volvereviria*, but not *Pleomoviria*) and *Riboviria*, reflecting known amplification biases in the sequencing protocols^{30–32}. Median genome (or segment) lengths of vOTUs differed by realm with 36.9 kb for *Duplodnaviria* (dsDNA), 5.1 kb for the former “*Monodnaviria*” (ssDNA) and 2.6 kb for *Riboviria* (Fig 1B). Unclassified viruses and *Varidnaviria* showed median lengths of approximately 10 kbp, which was the minimum length cutoff applied during quality control steps to retain sufficient quality vOTUs.

Altogether, 69% of vOTUs belonged to the class *Caudoviricetes* of tailed bacterial and archaeal viruses (Fig 1C). Out of these 2,443 *Caudoviricetes* vOTUs, only 84 could be assigned to previously known viral families, and the rest were categorised into 933 unknown family level clusters (see methods), mostly linked to bacterial hosts (Figure 2A). *Riboviria* comprised 3% of vOTUs (12 families, across eight orders including unclassified families and orders), predicted to infect a broad range of animal, plant, fungal and bacterial hosts (Figure 2B). The ssDNA realms *Volvereviria* (343 vOTUs), *Floreoviria* (195 vOTUs) and *Efunaviria* (103 vOTUs) accounted for 18% of vOTUs (10 families across eight orders including unclassified families within unassigned orders), with members predicted to infect a broad range of hosts including invertebrates, vertebrates, plants and fungi (Figure 2C). Members of the realm *Varidnaviria* were rarely observed (6 vOTUs), while 10% of vOTUs were unclassified at the realm rank, representing 237 unknown family level clusters (Supplementary Table S2). Only 15.2% of the vOTUs matched at the species level in the latest gut-virome databases³³ and 35% of the vOTUs were not represented based on the normalised AAI defined by UHGV³⁴. Interestingly, ~1.35% of vOTUs in our cohort were predicted to utilise alternate reassigned stop codons with UGA for tryptophan and TAG for glutamine.

The largest observed changes in extracellular virome composition occurred during month 4 in infants

Across all infants, we detected 33 viral families/clusters as early as W3 (with the caveat that we only sampled two infants at this time point), followed by a sharp increase with 331 new families at M4, 188 at M8, and 92 at M12. Thereafter, the number of families declined, with only 49, 54, and 34 new families appearing at M16, M20, and M24, respectively. The overall model predicted observed richness was not significantly different between time points after accounting for delivery mode and repeated sampling (negative-binomial mixed-effect model, Type III Wald $\chi^2(6) = 10.514$, $p=0.105$). Although the unadjusted pairwise M4 - W3 comparison was significant (emmeans Wald $z = 2.04$, unadjusted $p = 0.04$), this difference did not hold after correcting for multiple comparisons (Tukey-adjusted, $p=0.385$). Similarly, the overall months effect on Shannon diversity did not reach significance (Type III ANOVA, $F_{6,55}=2.102$, $p = 0.068$). Month 4 marginal means were elevated relative to all other months in the unadjusted comparison (linear mixed-effects model, t -test $p < 0.05$), but not significant after post hoc adjustments except for M4–M16 comparison (Tukey-adjusted, $p=0.032$). We observed that C-section infants appeared to exhibit lower virome diversity, especially in M4 (observed vOTU richness (median 44; range 24–75) and Shannon diversity (median 1.60; range 1.02–1.77)), compared to vaginally delivered infants ((observed richness: median 431, range 371–490; Shannon diversity: median 4.79, range 4.45–5.12)) similar to the findings of McCann *et al*, 2018³⁵. However, delivery mode was not significantly associated with observed richness and Shannon diversity across the longitudinal dataset (observed: negative-binomial mixed-effects model: expected-count ratio = 0.82, 95% CI: 0.33–2.04, $p = 0.672$; Shannon: linear mixed-effects model: estimate = -0.473 , 95% CI: -1.102 to 0.155 , $p = 0.145$). Thus, the delivery mode associated pattern is presented as mere exploratory observation here (Fig. 3A, Fig. 3E and Supplementary Fig. 8).

Alongside dsDNA phage-rich groups, families containing eukaryote-infecting viruses like *Picornaviridae*, *Tombusviridae*, *Circoviridae* and *Parvoviridae* were already detectable at W3 infant samples. A total of 636 vOTUs from early life of W3 and M4 samples in the infant gut were found to infect 81 predicted hosts including

Bifidobacteriaceae, *Enterobacteriaceae*, *Bacteroidaceae*, *Clostridiaceae*, *Peptostreptococcaceae* and *Ruminococcaceae* (Supplementary Figures 4 & 5; Supplementary Table 2). Later in infancy from M12 to M24, diverse crAss-like phages from the families *Crevaviridae*, *Intestiviridae*, *Steigviridae* and *Suoliviridae* were detectable in this cohort. Additional eukaryotic viral families, including *Caliciviridae* (M16) and *Astroviridae* (M20), were also observed in the second year of life (Supplementary figure 6).

We observed a significant difference in beta diversity between baby-mother pairs, supported by PERMANOVA ($R^2 = 0.27478$, $p = 0.001$). This result was further supported by the observation that virome compositions clustered closely within same baby-mother pairs (Figure 3B and Supplementary Figure 7) and were significantly more similar within pairs compared to unrelated pairs (Median Bray–Curtis similarity 0.00113 vs 0.00062; Wilcoxon rank sum test with continuity correction; $p = 4.867e-05$). Further, NMDS showed overlapping but distinct virome clusters for infants and mothers indicating different viral community structures (PERMANOVA $R^2 = 0.02057$, $p = 0.001$) driving their separation (Fig 3B).

We observed modest changes in the infant virome composition during M4-M24 after accounting for repeated sampling of the same infants (PERMANOVA $R^2 = 0.087$, $p = 0.024$). However, none of the 15 exploratory pairwise comparisons after FDR correction were significant, showing that the overall virome composition transition could not be attributed to any of the specific months temporally. Bray-Curtis dissimilarity distances of each sample to its groupspatial median revealed variability within months, but beta dispersion showed no significant differences (PERMDISP, $p = 0.129$, Fig 3E), however, we acknowledge the small sample size ($n=6$) of month 4. Taken together, these results suggest modest temporal variation in the infant virome composition, but a distinct transition point and subsequent stabilisation could not be established in our cohort. (Fig 3C and 3E, Supplementary Figure 4).

The alpha diversity of the maternal virome remains consistent across gestation and postpartum (Observed: negative-binomial mixed-effect model, Type III Wald χ^2 (6) = 8.627, $p = 0.196$; Shannon: linear mixed-effects model, Type III F (6, 58.947) = 1.044, $p = 0.407$; Fig 3A). Moreover, maternal samples did not show any significant compositional shifts in virome over time (PERMANOVA, $R^2 = 0.03976$, $p = 0.399$) and pairwise adonis2 comparisons between months did not reveal significant shifts in average community composition (Fig 3F), suggesting that the overall mother virome is stable over the period of gestation till postpartum. Still, we observed that the virome profiles were more dispersed at the third trimester than at birth (PERMDISP, $F_{5,66} = 3.0223$, $p = 0.001$; Tukey adjusted $p = 0.0075$), indicating significant difference in the virome spread between T3 and birth (Fig 3D).

Early life is dominated by influx of unique vOTUs

We used presence/absence metrics based on vOTU abundance (Supplementary table S4) and divided the vOTUs into eight major categories (Methods), based on vOTUs persistence, transience and pair specific sharing between infants and mothers (Figure 4A). The early life virome increased over time dominated with unique vOTUs, along with maternally derived viruses although the proportion of shared vOTUs varied highly between each infant-mother pair (Figure 4B).

Infants showed higher counts of transient unique vOTUs than persistent unique vOTUs (Wilcoxon, p -value = 0.0002348), in contrast to mothers, suggesting a highly dynamic and developing virome during first two years of life (Figure 4B and 4C). Furthermore, the number of persistent unique vOTUs were significantly higher in mothers compared to infants (Wilcoxon, p -value = 0.002131), presumably due to the long-term accumulation of viruses in adult gut (Figure 4C). Across the cohort, infant virome contained exclusively 166 family-level clusters while the mother virome comprised 306 unique family-level clusters. Yet 549 vOTUs were shared between both, with some vOTUs being shared across multiple pairs. Of these, 100 vOTUs were persistent in mothers since the gestation and persistent in the respective infants as well, across 10 pairs. Also, 145 more vOTUs persistent in mother were detected in the respective infants over time, suggesting putative maternal contribution in the, co-sharing and colonisation of vOTUs in infants. (Supplementary figure 9 & 10).

Additionally, we observed transiently shared vOTUs, especially when samples were collected concurrently, suggesting briefly shared viral signatures. Even within this healthy cohort, we observed 32

potentially pathogenic vOTUs from families previously implicated in disease but known to colonise asymptomatically in healthy infants^{36–38}. Of these, vOTUs belonging to families *Picornaviridae*, *Parvoviridae*, and *Poxviridae* were shared transiently with mothers in W3 samples of pair A, E and N respectively. For instance, in sample A, we detected a contig from the family *Picornaviridae* in W3 samples of both baby and mother with high abundance (1e+05 RPKM in mother and 3e+05 RPKM in baby), indicating concurrent presence in both individuals at this timepoint. Further, vOTUs belonging to families *Adenoviridae* and *Astroviridae*, were unique to infants and appeared only after 8 months, in agreement with prior observations by Liang et al³⁹.

Persistent maternal viruses shared with infants exhibit strong signatures of positive selection, independent of viral lifestyle or host

Despite similar overall levels of microdiversity (π)⁴⁰ between maternal and infant viromes (Wilcoxon test, $p = 0.3107$), we found that the evolutionary dynamics of vOTUs differed remarkably between different persistence categories. Unique transient vOTUs were excluded from comparative analyses, as their limited detection renders them unsuitable for comparison with persistent or shared categories. Across each sample pair, 50 – 80% of detected vOTUs carried at least one gene under positive selection. To better understand the drivers of persistence, we examined shared and persistent subclasses of vOTUs defined above, for their enrichment for positive selection. The strongest evidence for positive selection was found in vOTUs that were persistently detected in mothers but only occasionally present in infants, compared to all other persistent subclasses (shared but non-persistent vOTUs - Fisher's OR = 2.06, FDR = 0.0308; unique persistent in baby - Fisher's OR = 2.76, FDR = 0.0028; unique persistent in mother - Fisher's OR = 2.57, FDR = 0.0028; Figure 5A, Supplementary table S7). Those vOTUs persistent in infants and detected in mothers were also enriched for positive selection compared to unique-persistent subclasses (Fisher's OR = 1.89 for unique persistent in baby and 1.76 unique persistent in mother, FDR = 0.0308), though they did not differ significantly from other persistent subclasses (Fisher's OR = 0.69 – 1.41, FDR > 0.05).

Next, we examined whether vOTU persistence was explained by the predicted host or viral family. Several predicted bacterial host families including

Vibrionaceae, *Bacteroidaceae*, *Streptococcaceae*, *Acidaminococcaceae* were enriched for persistent vOTUs (Fishers OR, FDR, $p_{\text{adj}} < 0.05$, Supplementary Figure 11), at the cohort level. Interestingly, vOTUs belonging to the *Microviricetes* (formerly family *Microviridae*), *Circoviridae* and several unclassified clusters of *Caudoviricetes* family were the most significantly enriched among persistent viruses compared to the transient ones (Figure 5B; Supplementary S6), whereas *Anelloviridae* vOTUs were enriched among transient ones (Fisher's OR = 0.146, FDR = 8.53×10^{-7} ; Figure 5B). We detected 108 vOTUs from the *Anelloviridae* family, with at least one in 80% of infant samples and most of them were transient and cleared within the next sampling timepoint, while only 4 vOTUs were identified in 2 mother samples (Supplementary Figure 12). Given the much more stringent clustering threshold for vOTUs than anellovirus species⁴¹, the vOTUs likely reflect strain-level patterns that cannot be extrapolated to the

species level. Temperate or virulent phages were equally present in the persistent and transient groups of the virome, suggesting that phage lifestyle does not impact persistence, even though the relative abundance of temperate phages was higher in infants than in mothers (Wald test; p -value = 1.31×10^{-6}) (Supplementary Figures 13 and 14). Altogether, our results reveal that persistent viruses in mothers - particularly those occasionally transmitted to infants - are under stronger adaptive pressure, presumably correlated with a selective advantage for maintaining long-term presence in the gut ecosystem.

Preferential sharing of maternal viruses with infants is associated with specific host and viral families

Next, we examined if viruses in mothers were preferentially shared with infants from both persistent and transient virome pool across all the pairs. We found significant enrichment and depletion patterns at both the viral family/cluster and predicted host levels. At the viral family level, vOTUs from families including *Tombusviridae*, *Circoviridae*, *Smacoviridae*, *Parvoviridae* and several unclassified clusters of *Caudoviricetes* like Cluster689 predicted to infect *Vibrio*, Cluster882 predicted to infect *Lactococcus*, and Cluster597 were strongly enriched in the shared category, whereas others such as *Picobirnaviridae*, Cluster148 and Cluster 631(*Salasmaviridae*) were depleted (FDR < 0.05; Fig. 5C; Supplementary table S8).

Parallel independent host-prediction-based analyses identified corresponding signals for vOTUs infecting hosts like, *Vibrio*, archaea, *Lactococcus*, and some unclassified hosts which were enriched for sharing. In contrast, we observed 131 vOTUs associated with the host *Romboutsia* in 9 mothers and only 19 of them were shared with 6 infants. Similarly, *Bacteroidaceae*, *Clostridiaceae*, and *Peptostreptococcaceae* were significantly underrepresented genera of predicted hosts among shared maternal vOTUs (FDR < 0.05; Supplementary Fig. 6C and Supplementary table S8). Strikingly, vOTUs associated with Co-Abundance Gene groups assemblies (CAGs) of *Clostridia* were observed only in maternal pools but never in infants, although Metagenome Assembled Genomes (MAGs) of this group were found in infants in later infancy, resulting in no detectable sharing between mothers and infants (FDR < 0.05) across all the samples. Among strongly shared vOTUs, we observed that 4 *Bifidobacteriaceae*-targeting dsDNA phages from Cluster 893 in pair E and Cluster 901 in pair D were persistent in mothers since gestation and were later detected in infants up until M8. Across strongly shared vOTUs that were persistent in both mother and infants, several *Caudoviricetes* phage families targeting a wide diversity of bacterial taxa were represented, alongside eukaryotic virus families with known vertical transmission potential such as *Parvoviridae* detected in 50% of pairs and *Circoviridae* detected in 90% of pairs (Supplementary Figure 10). Taken together, these results demonstrate that maternal–infant virome sharing is not random but structured by both host and viral lineage, with specific families and clusters preferentially shared, while others remain restricted to mothers.

Discussion

In this study, we characterized the faecal virome of 15 mother-infant pairs across 137 samples, spanning from the first trimester through the infant's first two years of life. By employing a VLP-enriched

metagenomics approach that captures both DNA and RNA viruses, representing free extracellular virions, we provide a holistic view of virome in the early life.

As early as three weeks of life (W3), the infant virome already exhibited considerable diversity, spanning viruses from three realms, including bacteriophages predicted to infect early bacterial colonizers of the infant gut such as *Escherichia*, *Lactococcus*, *Bifidobacterium* which are known to play foundational roles in shaping early microbial communities^{42,43}. Alongside these phages, we detected eukaryotic viruses belonging to families such as *Tombusviridae*, *Picornaviridae*, *Circoviridae* and *Parvoviridae*, highlighting that virome assembly begins almost immediately in infants, with contributions from both bacterial and eukaryotic viruses^{36,39}. Our results on M4 largest observed infant extracellular virome richness aligns well with the BaBel cohort study, additionally supported by Liang et al^{36,44}, and contradicts the observation of increased viral diversity in infants over time by Lou et al⁴⁵. A transient window of immune vulnerability in infants has been described for specific infections such as RSV between three and six months of age⁴⁶, when maternal antibodies decline^{47,48} while infant immunity continues to develop⁴⁹; however, whether this affects gut virome development at M4 remains unknown.

Previous studies have defined the personal persistent virome (PPV) based on either high abundance across individual samples; ~92% of reads in 50% of timepoints⁵⁰ or simply presence in > 75% of samples per individual²⁸. In our study, we used a stricter, pair-centric definition, looking for individual vOTUs that were present across multiple timepoints in both the mother and the infant to define the shared virome. With this approach, we found hundreds of persistent vOTUs in mothers being shared with infants and persistent. By examining the viromes of mother and infants together, we highlight the influence of maternal virome including dsDNA, ssDNA and RNA viruses detectable in the infant gut, showing potential maternal impact on early life virome assembly, with both transient and persistent sharing. Our observations on maternal persistent viruses that were detected in infants exhibited stronger signatures of positive selection, matches well with previous observations of adaptive pressures on diversity-generating retroelements in the infant gut⁵¹, opening new avenues to uncover the genetic basis of viral persistence and transmittance. Our data suggest that viruses across different realms, detected during the gestational period in mother faecal samples, were clearly detectable in infants, as early as W3 and M4, with the majority of them being bacteriophages. With this approach, we have shown that persistent vOTUs in mothers were shared with infants in some cases, but that this persistence did not always translate to sharing or long-term retention in infants within our study period. We observed several vOTUs that were persistent in mothers, but either absent in infants or detected transiently without persistence.

Vertical transmission of bifidophages from mother to infant has been documented in earlier studies including the recent demonstration by Yew et al and others, through breast milk lipids^{22,28,45,52}. While the exact directionality of viral sharing in our cohort remains difficult to establish only a few maternal vOTUs associated with *Bifidobacterium* and *Bacteroides* were shared with infants. Contrastingly, vOTUs associated with uncultured *Clostridia* were not shared with infants. This could be because of a lower rate of spontaneous induction of *Clostridia*-associated phages in infants compared to mothers, or an inability

of these phages to engraft in the infant microbiome. Our findings suggest a potential maternal influence on the persistence of certain vOTUs like phages of *Bacteroidaceae*, also to be interpreted with caution for the directionality. While the persistence of vOTUs itself could not be explained by the viral lifestyle or host composition, the family *Microviridae* (now class *Microviricetes*) was found to be significantly associated with persistent virome and *Anelloviridae* with transient virome, which are known to be rapidly cleared in the infants^{36,37}. Although ssDNA viruses are generally prone to amplification bias, our finding of persistent microviruses aligns with previous reports⁵⁰, whereas *Anelloviridae*, also a family consisting of circular ssDNA viruses, exhibited the opposite trend. This suggests that specific viral traits or host associations may influence persistence or transience^{39,53}.

In this cohort, the infant virome contained numerous transient, unique vOTUs, predominantly bacteriophages that correspond well with earlier studies^{13,28,36,54,55} compared to the shared or persistent virome. In contrast, the mother virome showed a more stable composition with a significantly elevated persistent virome compared with infants, suggesting the long-term diet, host immune landscape and environmental exposures could play a crucial role in the stable establishment and maintenance of the virome in adulthood^{4,28,36}. The overall composition of the maternal gut bacteriome landscape changes in pregnancy were noted in previous studies²⁸. Surprisingly, the maternal gut extracellular virome appeared largely stable from pregnancy to postpartum in our cohort. The observed changes in extracellular virome community dispersion in our cohort during third trimester and delivery suggest subtle shifts in virome stability that may reflect physiological or immunological transitions occurring around the time of birth^{28,56}.

Previous studies have shown that the choice of processing method—metagenome versus virome—can significantly influence how viral community structure and dynamics are interpreted⁵⁷. In our study, we employed a VLP-enriched virome approach, which is effective in capturing free viral particles but likely underrepresents prophages integrated within bacterial genomes. Still, we were able to show the trend observed in previous studies^{28,36} that the infant virome is highly abundant in temperate phages over virulent phages. The observed abundance of ssDNA and RNA viruses in our dataset could be partly explained by the amplification steps employed during cDNA conversion. Thirdly, we may have missed some members of these realms due to our assembly thresholds 1.5 kb, while a 10 kb cutoff for unclassified sequences may have led to the omission of fragmented or novel viral genomes. In addition, our study, based on a single-centre cohort with only 15 baby mother pairs and unequal sampling constrains the broader generalisability and our ability to examine feeding, delivery mode, maternal lifestyle, socioeconomic status and other environmental factors affecting these results. Further investigation in larger, well-powered cohorts is needed to confirm our observations. The 15-RPKM presence threshold and > 50% persistence threshold was empirically selected for this cohort, and alternative thresholds may produce different results. Maternal–infant sharing was assessed at the vOTU level and therefore cannot confirm strain identity, directionality or transmission, while the microdiversity analysis identified within-vOTU gene-level variation rather than strain-resolved differences. Finally, we have shown the preferential sharing of maternal virome to infants was associated with specific viral

families including phages predicted to infect *Lactococcus*, *Escherichia*, *Vibrio* and some eukaryotic viruses; however, these host predictions rely on specific bioinformatics tools with their own limitations and should be interpreted with appropriate caution.

Together, our findings demonstrate that maternal–infant virome sharing is highly heterogeneous across pairs. Infant viromes were dominated by influx of transient viruses, yet a smaller but consistent subset of persistent vOTUs strongly shared with mother during early life. Notably, *Microviridae* were strongly enriched among persistent vOTUs, whereas *Anelloviridae*, consistent with their rapid clearance in infancy, were predominantly transient. Persistent maternal viruses shared with infants also exhibited pronounced signatures of positive selection, independent of lifestyle or host range, suggesting adaptation during early colonisation. Nevertheless, only a subset of persistent maternal vOTUs seen in infants may indicate that early virome assembly is highly structured and governed by additional determinants, including infant immune selection in the gut, changes in the bacterial community or other determinants. Finally, preferential sharing of specific viral families linked to predicted hosts indicates that maternal–infant virome exchange is non-random and shaped by both microbial host context and viral lineage.

Methods

Sample cohort description

The samples used in this study were obtained from the Pregnancy and Early life study – PEARL⁵⁸ in which a healthy cohort of mothers were recruited from the maternity unit of Norwich and Norfolk University Hospital foundation Trust (NNUH). We collected stool samples of mothers during the three trimesters of pregnancy (T1, T2 and T3), at birth (B), at weeks 1 and 3 (W1, W3) and four months after birth (M4) as part of the Pregnancy and Early Life study⁵⁸. Infant faecal samples were collected at birth, weeks 1 and 3 (W1, W3) and up to 24 months after birth (M4, M8, M12, M16, M20 and M24), however, due to the COVID-19 pandemic, not all samples were available for all participants. We selected 137 stool samples belonging to 15 mother-infant pairs, specifically excluding infant samples collected at birth and W1 due to low biomass (Figure 1A, Supplementary Table 1 and 8), for VLP separation and total nucleic acid extraction according to the NetoVir protocol^{59,60}. Of the 15 infants, 11 were delivered vaginally and four by either planned or emergency caesarean section. All infants, except A and D for whom feeding metadata were unavailable, were breastfed at some point during the first four months of life.

Enrichment of VLP from faecal samples and total nucleic acid extraction

About 50 mg of faecal sample was mixed with 500 µL of sterile PBS at a ratio of 1:10 [w/v] and homogenized by vortexing for one minute. The sample was then centrifuged at 17,000 x g for 3 minutes at 4°C, after which the supernatant containing VLPs was transferred to a fresh 1.5 mL tube and centrifuged again under the same conditions. The supernatant was filtered sequentially through 0.8 µm and 0.45 µm PES membrane filter units into sterile 1.5 mL tubes.

A mixture of 7 μL of 20x homemade DNase buffer (1M Tris, 100 mM CaCl_2 and 30 mM MgCl_2 , pH=8), 2 μL of TURBO DNase (Invitrogen, Carlsbad, USA), and 13 μL of TURBO DNase 10X Reaction Buffer was added to 130 μL of filtrate, and the mixture was incubated at 37°C for two hours in a heat block. The reaction was stopped with the addition of 1 μL of 0.5M EDTA. The eDNA removed VLPs were further processed for total nucleic acid extraction using the Qiagen QIAamp Viral RNA Mini Kit (QIAGEN, Venlo, The Netherlands) without carrier RNA. A 140 μL aliquot of DNase-treated VLP filtrate was combined with 560 μL AVL buffer in an RNase-free microcentrifuge tube, mixed by pulse vortexing, and incubated at room temperature for 10 minutes. Following a brief centrifugation, 560 μL of pure ethanol was added, mixed by vortexing, and centrifuged again. The sample mixture was applied to a spin column in 630 μL aliquots and centrifuged at 6,000 x g for 1 minute, followed by a second loading and centrifugation step. The column was then washed with 500 μL of AW1 buffer, centrifuged at 6,000 x g for one minute, transferred to a fresh collection tube, and washed with 500 μL of AW2 buffer. This was followed by centrifugation at 17,000 x g for 3 minutes. The column was then transferred to a fresh RNase-free microcentrifuge tube and centrifuged again at 17,000 x g for one minute to ensure dryness. Nucleic acids were eluted by adding 60 μL of AVE buffer to the column, incubating at room temperature for 1 minute, and centrifuging at 6,000 x g for 1 minute. The extracted total nucleic acids were stored at -80°C in aliquots to prevent freeze-thaw cycles.

Conversion of RNA to cDNA

For the conversion of RNA to dsDNA, cDNA synthesis was carried out using the Sigma-Aldrich WTA2 Complete Whole Transcriptome Amplification Kit (Sigma-Aldrich, Burlington, USA) with a modified protocol. An aliquot of 2.82 μL of above isolated Total Nucleic Acid was combined with 0.5 μL of Library Synthesis Solution in a 0.2 mL PCR tube, mixed by pipetting, and briefly spun down. The reaction was incubated in a thermocycler at 95°C for two minutes, followed by cooling to 18°C. To the cooled mixture, 0.78 μL of nuclease-free water, 0.5 μL of Library Synthesis Buffer, and 0.4 μL of Library Synthesis Enzyme were added. After mixing and brief centrifugation, the sample was incubated in a thermocycler under a stepped temperature profile of 18°C for 10 minutes, 25°C for 10 minutes, 37°C for 30 minutes, 42°C for 10 minutes, and 70°C for 20 minutes, followed by a hold at 4°C. The prepared 5 μL DNA library was then supplemented with 60.2 μL of nuclease-free water, 1.5 μL of WTA2 dNTP mix, 7.5 μL of Amplification Mix, and 0.75 μL of Amplification Enzyme. Amplification was performed in a thermocycler with an initial denaturation at 94°C for two minutes, followed by 17 cycles of 94°C for 30 seconds and 70°C for five minutes. The reaction was held at 4°C before the PCR product was purified using QIAquick PCR Purification Kit (QIAGEN, Venlo, The Netherlands). The amplified dsDNA was quantified using fluorimetry with the Qubit dsDNA broad range assay (Invitrogen, Carlsbad, USA) and stored at -20°C. Nucleic acid extraction and downstream library preparation were randomised.

Library preparation and sequencing

The libraries were prepared using Illumina DNA Prep Kit (FLEX) for tagmentation of the DNA and barcoding of the tagmented samples, followed by quantification using GloMax Promega for viable

libraries, with 10 ng of DNA as input. Sequencing was performed using an Illumina NovaSeq X platform with 150 bp paired-end chemistry (Illumina®, San Diego, USA). On average, 60.466 ± 20.101 million paired end virome reads were generated per sample. Additionally, we sequenced two blank samples as process control to delineate the signal from the contaminant VLP noise throughout the sequencing process.

Read quality was assessed using Fastp v0.23.2 with default parameters, including a minimum sequence read quality score of Phred >30. Host-derived reads were eliminated by mapping against the appropriate reference genome (GRCh38.p13) using kraken v2.1.0, and only unmapped reads were retained for downstream analysis. We analysed a median of 58,657,612 paired raw reads (Inter Quartile Range (IQR) = 46,581,503 – 73,576,984) per sample. The raw reads can be accessed via accession number at ENA accession number PRJEB106949.

Metagenome sequencing and MAG assembly

Nucleic acid from stool samples were extracted using the Promega Maxwell RSC Pure Food GMO kit and Promega Maxwell® RSC 48 instrument (Promega, Madison, USA). Libraries were prepared by the inhouse sequencing team, and included DNA normalisation, tagmentation, PCR barcoding, quantification, pooling, and quality control, prior to sequencing on the Illumina iSeq100 platform by Genewiz. Data was cleaned by removing adaptors and trimming reads, and filtered for quality using Fastp (V.0.20.0) (-q 20), before reads mapping against a human reference database (human_20200311) were removed using Kraken2 (V.2.1.1+galaxy0). Quality controlled and- dehumanised reads were assembled using MEGAHIT v1.2.9⁶¹ and the quality of the assembly was checked with QUAST v5.3.0⁶². The assembled contigs were then binned using MetaWRAP v1.3.2⁶³ with three different binners, MetaBAT2 v2.12.1⁶⁴, MaxBin2 v2.2.6⁶⁵ and CONCOCT v1.0.0⁶⁶ the quality of the bins was then assessed using checkM v1.1.3⁶⁷. MetaWRAP then compared the quality of the generated bins and chose the highest quality version of each bin for continued analysis. Quality thresholds for the bins were set at >50% completion and <10% contamination for medium-quality draft bins based on Minimum information about Metagenome assembled genome⁶⁸. The chosen bins were then used for bin reassembly with MetaWRAP where reads were mapped on to the bins with BWA-MEM v 0.7.17⁶⁹ and reassembled with SPAdes v3.13.0⁷⁰ using the mapped reads. The quality of the reassembled bins was then assessed with checkM and compared with the original bins; the higher quality bins were chosen for continued analysis.

The taxonomy of the assembled bins was assessed using GTDB-Tk v2.4.0⁷¹ with the GTDB R220 database and the determined taxonomy was then used to add the species name to the bin filename as a prefix for clearer tracking.

Assembly and mining viral sequences

For assembly, short reads were reconstructed into viral genomes by grouping reads from each individual sample and performing a co-assembly in order to improve the assembly quality. Reads were assembled

using MEGAHIT v1.1.1-2⁶¹ with parameters including $-k\text{-min } 21$, $-k\text{-max } 149$, $-k\text{-step } 24$, and $-\text{min-contig-len } 1500$ to minimize spurious viral contigs. We used Phables v1.3.2⁷², a graph-based assembly refining tool, to enhance the quality of contigs and we saw ~1.54% increase in the quality.

Viral contigs were identified in assemblies using four tools including geNomad v1.8.1⁷³, VirSorter2 v2.2.4⁷⁴, VIBRANT v1.2.1⁷⁵, and DeepVirFinder v1.0⁷⁶ using the default parameters and output from the DeepVirFinder was filtered for contigs with adjusted p-value threshold of <0.05 . The predicted virus contigs of all tools were combined and dereplicated at 100% to generate a non-redundant contig database for quality assessment.

Quality assessment of retrieved viral sequences was performed using CheckV v1.0.1⁷⁷, using the end-to-end command, and results from the quality_summary.tsv output table were parsed to retain medium and above quality genomes. Contigs were categorized into five groups based on completeness: complete (100%), high-quality ($\geq 90\%$), medium-quality (50–90%), low-quality ($<50\%$), or not determined.

Predicted viruses were clustered following the Minimum Information on Uncultivated Virus Genome (MIUViG) guidelines⁷⁸, with genomes collapsed into single representatives at the species level for tailed dsDNA bacterial and archaeal viruses (95% average nucleotide identity (ANI) over an 85% alignment fraction (AF) relative to the shorter sequence). In brief, a BLAST v2.10.0+⁷⁹ database was created from the viral sequences obtained in the previous step, and self-BLAST was performed. The anicalc utility from CheckV v 1.0.1 was then used to compute pairwise ANI, while aniclust was applied to parse ANI results and generate clusters based on a 95% ANI threshold and an 85% coverage threshold.

Further, vOTUs with a completeness score of at least 50% (medium-quality and above) were retained for downstream analyses. “Not determined” vOTUs were scrutinized further based on genome length range known for each realm recorded in the ICTV⁸⁰ database (≥ 10 kb for *Duplodnaviria*, *Varidnaviria* and Unclassified; ≥ 1500 bp for *Riboviria* and ≥ 648 bp for *Monodnaviria*), as they potentially represented novel viruses distantly related to known references. Family-level taxonomic assignments were retained where supported, but vOTU clustering was not used to infer formal ICTV species identity.

Abundance estimation, decontamination, and denoising

To evaluate the relative abundance of the vOTUs identified above, trimmed reads from each sample were individually mapped against the vOTU reference database using minimap2 v2.8⁸¹ in combination with SAMtools v1.4.1⁸² to produce sorted BAM files. Further, coverM v0.7.0⁸³ contig function was used to generate a RPKM (reads per kilobase million) counts table for all samples. To minimize erroneous mapping, we set $-\text{min-read-percent-identity}$ to 0.90, $-\text{min-read-aligned-percent}$ to 0.75 and $-\text{min-covered-fraction}$ 0.75 for all the samples⁸⁴, except for blank where $-\text{min-covered-fraction}$ was set to 0. To identify and remove potential contaminants and low-abundance vOTUs, vOTUs that were highly abundant in blanks and samples were considered contaminants likely to be introduced during sample processing and were removed. Additionally, vOTUs with RPKM values like those in the negative controls across all

samples were considered low abundance. vOTUs that had low abundance values in all samples below our chosen signal-to-noise threshold of 15 RPKM were excluded from the analyses as they were not reliably detectable. The 15-RPKM signal-to-noise threshold was selected empirically from the negative controls, in which background vOTU signals were below 15 RPKM, while the five vOTUs exceeding this value in blanks were removed as potential contaminants (Supplementary Figures 1 & 2). After quality control, assembly, viral mining, dereplication of the viral contigs to MIUViG standards, checkV quality filtering and denoising (Supplementary Figures 1-3), our final dataset consists of 3531 vOTUs.

Taxonomic Assignment and other functional annotations

We functionally annotated vOTUs using Pharokka v 1.7.4^{85,85-95}, which predicted coding sequences (CDS) and assigned putative functions. Pharokka was run with -g prodigal-gv v2.11.0 to account for alternative stop codons in viral genomes^{73,96}, and with --meta and --split to efficiently annotate large vOTU datasets. The annotate module in geNomad v1.8.1 was used to obtain taxonomic classifications from realm to family levels. For genus and sub-family level classification of vOTUs, we employed vConTACT2 0.11.3⁹⁷, using the amino acid files from Pharokka and a custom CDS-to-contig mapping file. vConTACT2 was run with the INPHARED⁹⁰ database, with taxonomy assignments refined using graphanalyzer.py⁹⁸. Unclassified vOTUs, particularly those from *Duplodnaviria*, were further examined by grouping based on shared realm-level taxonomy and clustering by proteome similarity using VipTreeGen v1.1.3^{99,100}. A Newick tree was generated, and clades were defined by splitting the tree at branch lengths consistent with known viral family divergence (e.g., ~0.04 – 0.05) using the reference genome downloaded from VMR 39. These interim clades were used for downstream ecological analyses to avoid oversimplified “Unclassified” categories. Contigs of vOTUs were matched against published viral catalogues such as the Unified Human Gut Virome Catalog and AVrC^{34,55} to identify previously reported sequences and associated vOTU taxonomy information. Outputs from all the tools were integrated to produce consolidated taxonomic assignments. Because the initial analyses and tool-based taxonomic assignments were performed on earlier versions of the ICTV taxonomy, a final manual update was performed after the release of the 2026 ICTV taxonomy using Master Species List MSL41.v1, using the tab “Taxa Renamed or Abolished”. Major changes were replacement of the realm “*Monodnaviria*” with realms *Floreoviria*, *Volvereviria* and *Efunaviria* with the underlying kingdoms and phyla unchanged and the promotion of the family “*Microviridae*” to the class *Microviricetes* with in our analyses the underlying taxa now unassigned. Finally, computational prediction of host at family level was done using iPHOP¹⁰¹ v1.3.3 using minimum confidence score 90.0. Further, a total of 842 vOTUs matched in Metagenome Assembled Genomes (MAGs) at $\geq 95\%$ identity over $\geq 85\%$ length. These were annotated as ‘temperate’ phages and host information derived from the MAG annotation. Viral lifestyle for rest of the vOTUS were derived using PHATYP¹⁰² using score ≥ 0.7 .

Persistence and sharing of vOTU classification

We used presence/absence metrics based on abundance to classify each vOTU, within each mother–infant pair along two axes: 1) Sharing within mother- infant pair 2) persistence within a baby and/or mother sample. For the sharing axis, a vOTU was classified as shared when it was detected at least once in both the mother and infant. A vOTU was classified persistent in mother or baby when it was present in atleast half of the sampled time points, else transient.

For mothers, samples collected at trimester 1 (T1), trimester 2 (T2), trimester 3 (T3) and postpartum M4 were treated as separate time points as each sample was obtained approximately 3 months apart. Samples collected at birth (B), week1 (W1) and week3 (W3) were grouped together as single timepoint given their close proximity, that could artificially inflate the persistence. No grouping was done for infant samples. Baby- mother pairs with insufficient temporal resolution to define persistence, especially pairs G and M, had only one baby sample each, were excluded from the persistence analysis.

For each infant/mother pair, combining the sharing and persistence axes produced eight mutually exclusive vOTU categories: (1) shared and persistent in both mother and infant; (2) shared, persistent in the infant and transient in the mother; (3) shared, persistent in the mother and transient in the infant; (4) Shared vOTUs that were not persistent in either were categorised as transiently shared between baby and mothers; shared and transient in both; (5) infant-unique and persistent; (6) infant-unique and transient; (7) mother-unique and persistent; and (8) mother-unique and transient. Because assignments were pair-specific, a vOTU could be shared in one pair yet unique in another and vice versa.

Natural selection (pN/pS; the ratio of non-synonymous to synonymous substitution rates) and nucleotide diversity (π) metrics were quantified using the MetaPop software package⁴⁰ to evaluate the microdiversity of vOTUs. Briefly, reads were mapped to dereplicated vOTUs, and the resulting alignment were filtered using minimum global nucleotide identity of 90% (`--id_min 90`), calculated over the full read length (`--global`), and a minimum alignment length of 75 bp (`--min_len 75`). A vOTU was retained for within-sample microdiversity analysis when at least 75% of its genome was covered and its truncated average depth was at least 10×. SNPs were first identified globally by pooling per-position base coverage across all samples. Alternate alleles representing at least 1% of pooled coverage and supported by at least four reads were retained as true variants, using a minimum base-quality threshold of PHRED 30 (`--min_qual 30`). For local, per-sample microdiversity analyses, the globally identified SNP set was restricted to positions independently detected in each sample; positions detected only in other samples were therefore excluded. Coverage at variant loci was then proportionally subsampled to 10× (`--subsample_size 10`). Where multiple SNPs occurred within the same codon, their co-occurrence was resolved by direct inspection of the aligned reads. These calls were then used to calculate per-gene nucleotide diversity and pN/pS. Genes were called under positive selection if $pN/pS > 1$ or ∞ . Infinite values, caused by the absence of synonymous polymorphisms ($pS = 0$), were retained, and NA treated as non-selected. Shared or unique category subclasses were assigned from the table created above.

For each vOTU within each subclass, a binary flag recorded whether ≥ 1 gene was under positive selection. Fisher's exact tests were applied to 2×2 tables (selected vs. non-selected vOTUs) for all

subclass pairs, and multiple comparisons were controlled using the Benjamini–Hochberg FDR. Results were visualised as a heatmap reporting $\log_2$ (odds ratio), the OR value, and FDR significance.

To evaluate maternal influence on infant virome composition, maternal vOTUs were first divided into maternal-persistent and maternal-transient pools (as defined above). For each viral family/cluster, 2×2 contingency tables were constructed comparing the number of maternal vOTUs within that family that were shared with infants versus not shared, relative to the rest of the maternal virome. Fisher's exact tests were applied to estimate odds ratios with 95% confidence intervals, and p-values were adjusted for multiple testing using the Benjamini–Hochberg procedure. Significant enrichment or depletion was defined at FDR < 0.05. Results were visualized in forest plots displaying odds ratios on a log scale, with significantly enriched or depleted taxa highlighted. Infinite odds ratios, reflecting cases where vOTUs were exclusively shared or exclusively not shared, were retained for reporting but capped in figures for readability. The same approach was used for host genera.

Statistical methods and data visualisation

All statistical analyses were performed in R v4.3.2. Alpha diversity was analysed using mixed-effects models to account for repeated measurements within pairs. Observed vOTU richness was modelled with a negative binomial GLMM, and Shannon diversity with a linear LMM, with month and delivery mode included as a fixed effect and pair as a random intercept (Diversity metrics ~ Months + (Delivery == "delivery mode") + (1 | Pair)) for infants and (Diversity metrics ~ Months + (1 | Pair)) for mothers. Overall month effects were assessed using Type III omnibus tests. Post-hoc comparisons between months were performed using estimated marginal means with Tukey adjustment for multiple comparisons. Pairwise differences between months were assessed using Tukey-adjusted posthoc tests. A two-sided Wilcoxon rank-sum test with continuity correction was used for pairwise comparisons, with multiple testing corrected by the Benjamini–Hochberg method where appropriate. Beta-diversity (Bray Curtis), PERMANOVA using the `adonis2` function and non-metric multidimensional scaling (NMDS) ordinations, `envfit` were computed with the `vegan` v2.6.8 package. Repeated longitudinal sampling was accounted by employing, 999 permutations restricted within the mother–infant pair identifier (Pair). The primary temporal analysis included M4–M24 for infants and T2–M4 for mother; W3 samples in infants and T1 samples in mothers were excluded because only two samples were available. Exploratory all-pairwise PERMANOVA comparisons were performed using the same within-pair permutation restriction, using `pairwise.adonis2` package and p-values were adjusted across all 15 comparisons using the Benjamini–Hochberg FDR method. Fisher's exact tests were used to assess enrichment or depletion where required, with false discovery rate (FDR) correction applied to account for multiple testing. Data handling and summarisation were performed with `dplyr` v1.1.4. Figures, including bar plots, box plots, line plots, and scatter plots, were generated using `tidyverse` v2.0.0, `ggplot2` v3.5.1, and `ggpubr` v0.6.0. Where necessary, figures were assembled into final panels using Inkscape.

Ethical information and data availability

Samples used in this study were obtained from the PEARL study (IRAS ID: 241880), which received a favourable opinion from the London–Dulwich Research Ethics Committee (REC reference: 18/LO/1703) and approval from the Health Research Authority (HRA). The study was also reviewed and registered with the Quadram Institute Human Research Governance Committee.

The PEARL study was designed and conducted in accordance with the principles of the International Conference on Harmonisation Good Clinical Practice (ICH-GCP) guidelines and in compliance with all relevant UK legislation. Written informed consent was obtained from all participants prior to inclusion in the study. The collection, storage, and use of human tissue samples were carried out in accordance with the requirements of the Human Tissue Act 2004, under the oversight of the Human Tissue Authority (HTA). All the raw sequencing data is available through ENA accession ID PRJEB123118. Data required to reproduce this study is available through GitHub (<https://github.com/Revathylcosahedron/pearl-viromics/>).

Declarations

Ethical information and data availability

The study was conducted in accordance with the Declaration of Helsinki. Written informed consent was obtained from all participating mothers, including consent from the parent or legal guardian for participation of their infants. Samples used in this study were obtained from the PEARL study (IRAS ID: 241880), which received a favourable opinion from the London–Dulwich Research Ethics Committee (REC reference: 18/LO/1703) and approval from the Health Research Authority (HRA). The study was also reviewed and registered with the Quadram Institute Human Research Governance Committee.

The PEARL study was designed and conducted in accordance with the principles of the International Conference on Harmonisation Good Clinical Practice (ICH-GCP) guidelines and in compliance with all relevant UK legislation. Written informed consent was obtained from all participants prior to inclusion in the study. The collection, storage, and use of human tissue samples were carried out in accordance with the requirements of the Human Tissue Act 2004, under the oversight of the Human Tissue Authority (HTA). All the raw sequencing data is available through ENA accession ID PRJEB123118. Data required to reproduce this study is available through GitHub (<https://github.com/Revathylcosahedron/pearl-viromics/>).

Competing interests

The authors declare no competing financial or non-financial interests.

Acknowledgements

R.K., A.J.P., S.K.T., J.T., A.T., S.P., R.W., S.H., G.M.S., A.H., L.J.H. and E.M.A. were funded by the BBSRC Institute Strategic Programme Food Microbiome and Health BB/X011054/1 and its constituent projects BBS/E/QU/230001A, BBS/E/QU/230001B, BBS/E/QU/230001C and BBS/E/QU/230001D; R.K. and E.M.A.

were further funded by the BBSRC Institute Strategic Programme Microbes and Food Safety BB/X011011/1 and its constituent projects BBS/E/QU/230001A, BBS/E/QU/230001B, and BBS/E/QU/230001C; and G.M.S. and D.B. by the BBSRC Core Capability Grant BB/CCG2260/1. R.C. and E.M.A. were funded by a BBSRC grant BB/W015706/1. H.V.P. and E.M.A were funded by the Medical Research Council grant MR/W031205/1. X.D. is supported by the BBSRC Norwich Research Park Doctoral Training Partnership BB/T008717/1. This computation infrastructure was in part provided by MRC CLIMB BIG DATA grant MR/T030062/1. L.J.H. was funded by a Wellcome Trust Investigator Award no. 220876/Z/20/Z, and BBSRC Institute Strategic Programme, Gut Microbes and Health BB/R012490/1, and its constituent projects BBS/E/F/000PR10353 and BBS/E/F/ 000PR10356. L.M. is supported by a Colin Charles Frederick Blake University of Birmingham PhD Studentship with L.J.H. as supervisor. This research was funded in part by the Wellcome Trust 220876/Z/20/Z. For the purpose of open access, the author has applied a CC BY public copyright licence to any author accepted manuscript version arising from this submission. We would like to acknowledge Dr. Oliver Charity for his contributions in the early stages of this study.

References

1. Arrieta, M.-C. *et al.* Associations between infant fungal and bacterial dysbiosis and childhood atopic wheeze in a nonindustrialized setting. *J. Allergy Clin. Immunol.* **142**, 424-434.e10 (2018).
2. Vatanen, T. *et al.* The human gut microbiome in early-onset type 1 diabetes from the TEDDY study. *Nature* **562**, 589–594 (2018).
3. Robertson, R. C., Manges, A. R., Finlay, B. B. & Prendergast, A. J. The Human Microbiome and Child Growth - First 1000 Days and Beyond. *Trends Microbiol.* **27**, 131–147 (2019).
4. Ferretti, P. *et al.* Mother-to-Infant Microbial Transmission from Different Body Sites Shapes the Developing Infant Gut Microbiome. *Cell Host Microbe* **24**, 133-145.e5 (2018).
5. Samarra, A. *et al.* Breastfeeding and early Bifidobacterium-driven microbial colonization shape the infant gut resistome. *Nat. Commun.* **16**, 6099 (2025).
6. Sawhney, S. Influence of Environmental Gradients on Genomic Variation in Pediatric Commensals and Pathogens. *Arts Sci. Electron. Theses Diss.* <https://doi.org/https://doi.org/10.7936/gsxj-6r74> (2023) doi:<https://doi.org/10.7936/gsxj-6r74>.
7. Lepage, P. *et al.* Dysbiosis in inflammatory bowel disease: a role for bacteriophages? *Gut* **57**, 424–425 (2008).
8. Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473**, 174–180 (2011).
9. Hoyles, L. *et al.* Characterization of virus-like particles associated with the human faecal and caecal microbiota. *Res. Microbiol.* **165**, 803–812 (2014).
10. Shkoporov, A. N. *et al.* Reproducible protocols for metagenomic analysis of human faecal phageomes. *Microbiome* **6**, 68 (2018).
11. Shkoporov, A. N., Turkington, C. J. & Hill, C. Mutualistic interplay between bacteriophages and bacteria in the human gut. *Nat. Rev. Microbiol.* **20**, 737–749 (2022).

12. Shkoporov, A. N. & Hill, C. Bacteriophages of the Human Gut: The “Known Unknown” of the Microbiome. *Cell Host Microbe* **25**, 195–209 (2019).
13. Tisza, M. J. et al. Longitudinal phage–bacteria dynamics in the early life gut microbiome. *Nat. Microbiol.* **10**, 420–430 (2025).
14. Zuppi, M., Hendrickson, H. L., O’Sullivan, J. M. & Vatanen, T. Phages in the Gut Ecosystem. *Front. Cell. Infect. Microbiol.* **11**, (2022).
15. Minot, S. et al. Rapid evolution of the human gut virome. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 12450–12455 (2013).
16. Jansen, D. & Matthijnssens, J. The Emerging Role of the Gut Virome in Health and Inflammatory Bowel Disease: Challenges, Covariates and a Viral Imbalance. *Viruses* **15**, 173 (2023).
17. Barr, J. J. A bacteriophages journey through the human body. *Immunol. Rev.* **279**, 106–122 (2017).
18. Cook, R. A. et al. Bacteriophage diversity declines with COPD severity in the respiratory microbiome. 2025.06.16.659870 Preprint at <https://doi.org/10.1101/2025.06.16.659870> (2025).
19. Bikel, S. et al. Gut dsDNA virome shows diversity and richness alterations associated with childhood obesity and metabolic syndrome. *iScience* **24**, (2021).
20. Norman, J. M. et al. Disease-specific Alterations in the Enteric Virome in Inflammatory Bowel Disease. *Cell* **160**, 447–460 (2015).
21. Gay, N. J. et al. Age specific antibody prevalence to parvovirus B19: how many women are infected in pregnancy. *Commun. Dis. Rep. CDR Rev.* **4**, R104-7 (1994).
22. Duranti, S. et al. Maternal inheritance of bifidobacterial communities and bifidophages in infants through vertical transmission. *Microbiome* **5**, 66 (2017).
23. Aronoff, D. M., Correa, H., Rogers, L. M., Arav-Boger, R. & Alcendor, D. J. Placental pericytes and cytomegalovirus infectivity: Implications for HCMV placental pathology and congenital disease. *Am. J. Reprod. Immunol.* **78**, e12728 (2017).
24. Reyes, A. et al. Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* **466**, 334–338 (2010).
25. Maqsood, R. et al. Discordant transmission of bacteria and viruses from mothers to babies at birth. *Microbiome* **7**, 156 (2019).
26. Liang, G. & Bushman, F. D. The human virome: assembly, composition and host interactions. *Nat. Rev. Microbiol.* **19**, 514–527 (2021).
27. Walters, W. A. et al. Longitudinal comparison of the developing gut virome in infants and their mothers. *Cell Host Microbe* **31**, 187-198.e3 (2023).
28. Garmaeva, S. et al. Transmission and dynamics of mother-infant gut viruses during pregnancy and early life. *Nat. Commun.* **15**, 1945 (2024).
29. Redgwell, T. A. et al. Prophages in the infant gut are pervasively induced and may modulate the functionality of their hosts. *Npj Biofilms Microbiomes* **11**, 1–15 (2025).

30. Yilmaz, S., Allgaier, M. & Hugenholtz, P. Multiple displacement amplification compromises quantitative analysis of metagenomes. *Nat. Methods* **7**, 943–944 (2010).
31. d’Humières, C. et al. A simple, reproducible and cost-effective procedure to analyse gut phageome: from phage isolation to bioinformatic approach. *Sci. Rep.* **9**, 11331 (2019).
32. Džunková, M., Moraru, C. & Anantharaman, K. Editorial: Advances in viromics: new tools, challenges, and data towards characterizing human and environmental viromes. *Front. Microbiol.* **14**, (2023).
33. Galperina, A. et al. The Aggregated Gut Viral Catalogue (AVrC): A Unified Resource for Exploring the Viral Diversity of the Human Gut. 2024.06.24.600367 Preprint at <https://doi.org/10.1101/2024.06.24.600367> (2024).
34. Almeida, A. et al. A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nat. Biotechnol.* **39**, 105–114 (2021).
35. McCann, A. et al. Viromes of one year old infants reveal the impact of birth mode on microbiome diversity. *PeerJ* **6**, e4694 (2018).
36. Beller, L. et al. The virota and its transkingdom interactions in the healthy infant gut. *Proc. Natl. Acad. Sci.* **119**, e2114619119 (2022).
37. Lim, E. S. et al. Early life dynamics of the human gut virome and bacterial microbiome in infants. *Nat. Med.* **21**, 1228–1234 (2015).
38. Kapusinszky, B., Minor, P. & Delwart, E. Nearly constant shedding of diverse enteric viruses by two healthy infants. *J. Clin. Microbiol.* **50**, 3427–3434 (2012).
39. Liang, G. et al. The stepwise assembly of the neonatal virome is modulated by breastfeeding. *Nature* **581**, 470–474 (2020).
40. Gregory, A. C. et al. MetaPop: a pipeline for macro- and microdiversity analyses and visualization of microbial and viral metagenome-derived populations. *Microbiome* **10**, 49 (2022).
41. Kraberger, S. et al. ICTV Virus Taxonomy Profile: Anelloviridae 2026. *J. Gen. Virol.* **107**, 002222 (2026).
42. Milani, C. et al. The First Microbial Colonizers of the Human Gut: Composition, Activities, and Health Implications of the Infant Gut Microbiota. *Microbiol. Mol. Biol. Rev. MMBR* **81**, e00036-17 (2017).
43. Zhernakova, A., Yassour, M., Hall, L. J. & Collado, M. C. Unlocking the power of human milk and infant feeding: Understanding how nutrition and early microbiota interaction shapes health programming. *Cell Host Microbe* **33**, 820–835 (2025).
44. Liang, G. et al. The stepwise assembly of the neonatal virome is modulated by breastfeeding. *Nature* **581**, 470–474 (2020).
45. Lou, Y. C. et al. Infant gut DNA bacteriophage strain persistence during the first 3 years of life. *Cell Host Microbe* **32**, 35-47.e6 (2024).
46. Coindy, E. L. et al. Antibody-mediated protection against respiratory syncytial virus in children. *Eur. Respir. Rev.* **33**, (2024).

47. Pasittungkul, S. *et al.* High seroprevalence of antibodies against human respiratory syncytial virus and evidence of respiratory syncytial virus reinfection in young children in Thailand. *Int. J. Infect. Dis.* **125**, 177–183 (2022).
48. Ochola, R. *et al.* The level and duration of RSV-specific maternal IgG in infants in Kilifi Kenya. *PloS One* **4**, e8088 (2009).
49. Gleeson, M. & Cripps, A. W. Development of mucosal immunity in the first year of life and relationship to sudden infant death syndrome. *FEMS Immunol. Med. Microbiol.* **42**, 21–33 (2004).
50. Shkoporov, A. N. *et al.* The Human Gut Virome Is Highly Diverse, Stable, and Individual Specific. *Cell Host Microbe* **26**, 527–541.e5 (2019).
51. Macadangdang, B. R. *et al.* Targeted protein evolution in the gut microbiome by diversity-generating retroelements. *Science* **390**, eadv2111 (2025).
52. Yew, W. C. *et al.* The core phageome and its interrelationship with preterm human milk lipids. *Cell Rep.* **42**, 113373 (2023).
53. Tisza, M. J. *et al.* Longitudinal phage–bacteria dynamics in the early life gut microbiome. *Nat. Microbiol.* **10**, 420–430 (2025).
54. Fernández-Pato, A. *et al.* Early-life development of the gut virome and plasmidome: A longitudinal study in cesarean-born infants. *Cell Rep.* **44**, 115731 (2025).
55. Galperina, A. *et al.* The Aggregated Gut Viral Catalogue (AVrC): A unified resource for exploring the viral diversity of the human gut. *PLOS Comput. Biol.* **21**, e1012268 (2025).
56. Koren, O. *et al.* Host remodeling of the gut microbiome and metabolic changes during pregnancy. *Cell* **150**, 470–480 (2012).
57. Kosmopoulos, J. C., Klier, K. M., Langwig, M. V., Tran, P. Q. & Anantharaman, K. Viromes vs. mixed community metagenomes: choice of method dictates interpretation of viral community ecology. *Microbiome* **12**, 195 (2024).
58. Phillips, S. *et al.* The Pregnancy and EARly Life study (PEARL) - a longitudinal study to understand how gut microbes contribute to maintaining health during pregnancy and early life. *BMC Pediatr.* **21**, 357 (2021).
59. Conceição-Neto, N., Yinda, K. C., Van Ranst, M. & Matthijnssens, J. NetoVIR: Modular Approach to Customize Sample Preparation Procedures for Viral Metagenomics. *Methods Mol. Biol. Clifton NJ* **1838**, 85–95 (2018).
60. Conceição-Neto, N. *et al.* Modular approach to customise sample preparation procedures for viral metagenomics: a reproducible protocol for virome analysis. *Sci. Rep.* **5**, 16532 (2015).
61. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
62. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).

63. Uritskiy, G. V., DiRuggiero, J. & Taylor, J. MetaWRAP-a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* **6**, 158 (2018).
64. Kang, D. D. et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).
65. Wu, Y.-W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605–607 (2016).
66. Alneberg, J. et al. Binning metagenomic contigs by coverage and composition. *Nat. Methods* **11**, 1144–1146 (2014).
67. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
68. Bowers, R. M. et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* **35**, 725–731 (2017).
69. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
70. Bankevich, A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol. J. Comput. Mol. Cell Biol.* **19**, 455–477 (2012).
71. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* **38**, 5315–5316 (2022).
72. Mallawaarachchi, V. et al. Phables: from fragmented assemblies to high-quality bacteriophage genomes. *Bioinformatics* **39**, btad586 (2023).
73. Camargo, A. P. et al. Identification of mobile genetic elements with geNomad. *Nat. Biotechnol.* **42**, 1303–1312 (2024).
74. Guo, J. et al. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* **9**, 37 (2021).
75. Kieft, K., Zhou, Z. & Anantharaman, K. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* **8**, 90 (2020).
76. Ren, J. et al. Identifying viruses from metagenomic data using deep learning. *Quant. Biol.* **8**, 64–77 (2020).
77. Nayfach, S. et al. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat. Biotechnol.* **39**, 578–585 (2021).
78. Roux, S. et al. Minimum Information about an Uncultivated Virus Genome (MIUViG). *Nat. Biotechnol.* **37**, 29–37 (2019).
79. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).

80. Turner, D. *et al.* Abolishment of morphology-based taxa and change to binomial species names: 2022 taxonomy update of the ICTV bacterial viruses subcommittee. *Arch. Virol.* **168**, 74 (2023).
81. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
82. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
83. Aroney, S. T. N. *et al.* CoverM: Read alignment statistics for metagenomics. *Bioinformatics* btaf147 (2025) doi:10.1093/bioinformatics/btaf147.
84. Roux, S., Emerson, J. B., Eloë-Fadrosch, E. A. & Sullivan, M. B. Benchmarking viromics: an in silico evaluation of metagenome-enabled estimates of viral community composition and diversity. *PeerJ* **5**, e3817 (2017).
85. Alcock, B. P. *et al.* CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **48**, D517–D525 (2020).
86. Bland, C. *et al.* CRISPR Recognition Tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* **8**, 209 (2007).
87. Bouras, G. *et al.* Pharokka: a fast scalable bacteriophage annotation tool. *Bioinformatics* **39**, btac776 (2023).
88. Chan, P. P., Lin, B. Y., Mak, A. J. & Lowe, T. M. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* **49**, 9077–9096 (2021).
89. Chen, L. *et al.* VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res.* **33**, D325–D328 (2005).
90. Cook, R. *et al.* INfrastructure for a PHAge REference Database: Identification of Large-Scale Biases in the Current Collection of Cultured Phage Genomes. *PHAGE* **2**, 214–223 (2021).
91. Larralde M. Pyrodigal: Python bindings and interface to Prodigal, an efficient method for gene prediction in prokaryotes. *J. Open Source Softw.* **7**, 4296 (2022).
92. Larralde M and Camargo A.(2023). Pyrodigal-gv: A Pyrodigal extension to predict genes in giant viruses and viruses with alternative genetic code. <https://github.com/althonos/pyrodigal-gv>.
93. Larralde M, M. & Zeller, G. PyHMMER: a Python library binding to HMMER for efficient sequence analysis. *Bioinformatics* **39**, btad214 (2023).
94. Laslett, D. & Canback, B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* **32**, 11–16 (2004).
95. Shimoyama, Y. pyCirclize: Circular visualization in Python. (2022).
96. Cook, R. *et al.* Driving through stop signs: predicting stop codon reassignment improves functional annotation of bacteriophages. *ISME Commun.* **4**, ycae079 (2024).
97. Bin Jang, H. *et al.* Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat. Biotechnol.* **37**, 632–639 (2019).

98. Pandolfo, M., Telatin, A., Lazzari, G., Adriaenssens, E. M. & Vitulo, N. MetaPhage: an Automated Pipeline for Analyzing, Annotating, and Classifying Bacteriophages in Metagenomics Sequencing Data. *mSystems* **7**, e0074122 (2022).

99. Nishimura, Y. et al. ViPTree: the viral proteomic tree server. *Bioinformatics* **33**, 2379–2380 (2017).

100. Rohwer, F. & Edwards, R. The Phage Proteomic Tree: a Genome-Based Taxonomy for Phage. *J. Bacteriol.* **184**, 4529–4535 (2002).

101. Roux, S. et al. iPHoP: An integrated machine learning framework to maximize host prediction for metagenome-derived viruses of archaea and bacteria. *PLOS Biol.* **21**, e3002083 (2023).

102. Shang, J., Tang, X. & Sun, Y. PhaTYP: predicting the lifestyle for bacteriophages using BERT. *Brief. Bioinform.* **24**, bbac487 (2023).

Figures

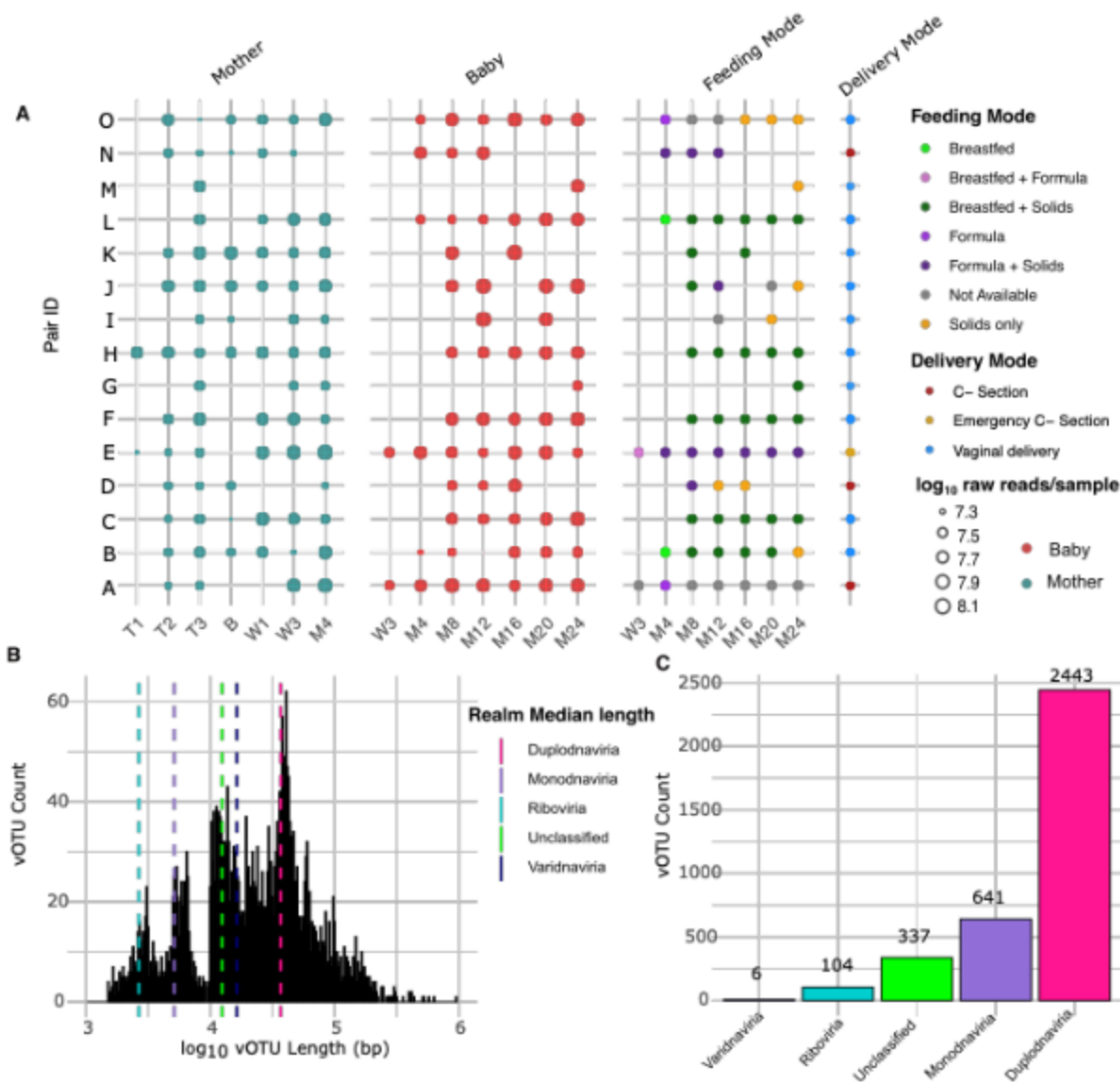


Figure 1

(A) Bubble plot showing $\log_{10}$ transformed raw read counts per sample (dot size) across all baby–mother pairs (y-axis) and timepoints (x-axis). Dot colour indicates feeding mode for babies and delivery mode. (B) Distribution of vOTU contig lengths ($\log_{10}$ bp), with median lengths for each realm. (C) Bar plot of vOTU count per viral realm.

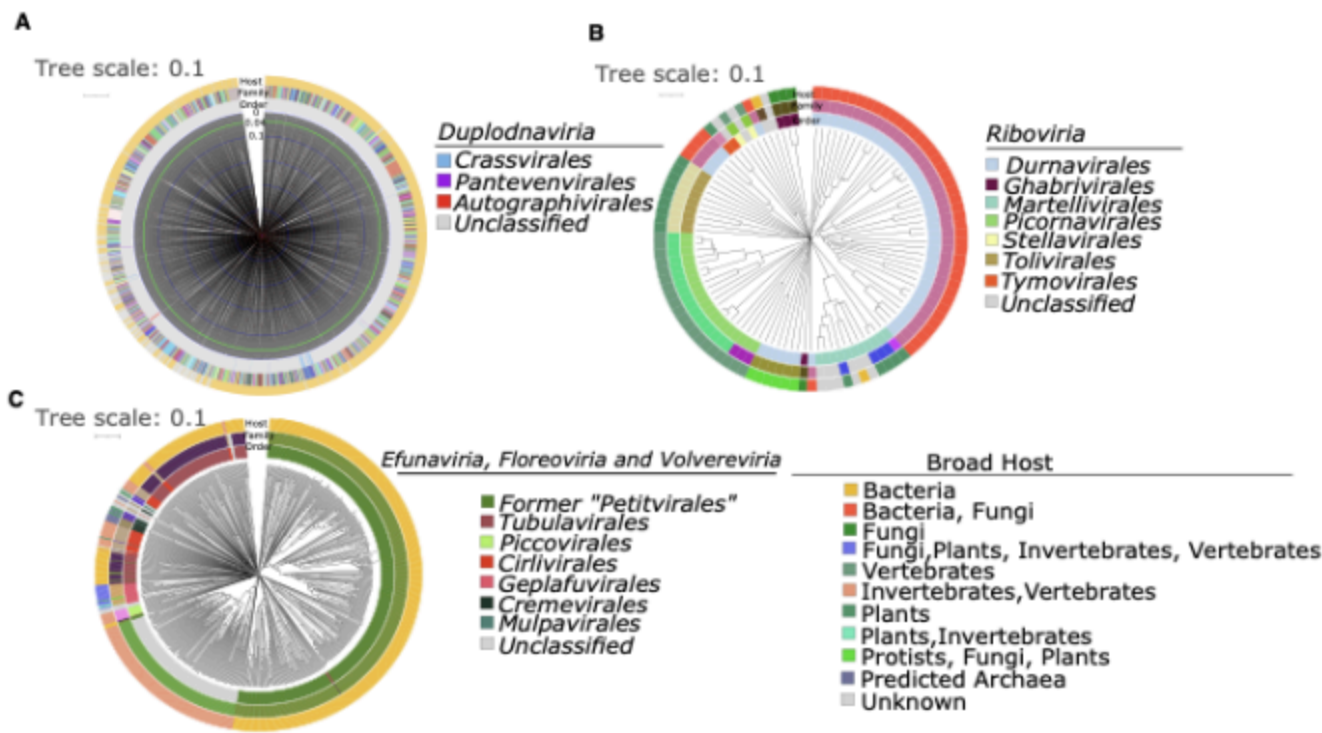


Figure 2

Virus Proteomic trees of PEARL vOTUs. (A) Duplodnaviria (dsDNA viruses), (B) Riboviria (RNA viruses), (C) Volvereviria, Floreoviria and Efunaviria ssDNA viruses. Dendrograms represent the pairwise distance between the proteomes as calculated using tblastx by VipTree. Coloured rings indicate predicted taxonomic order (inner ring), family/family-rank clusters (middle ring) and broad host range (outer ring).

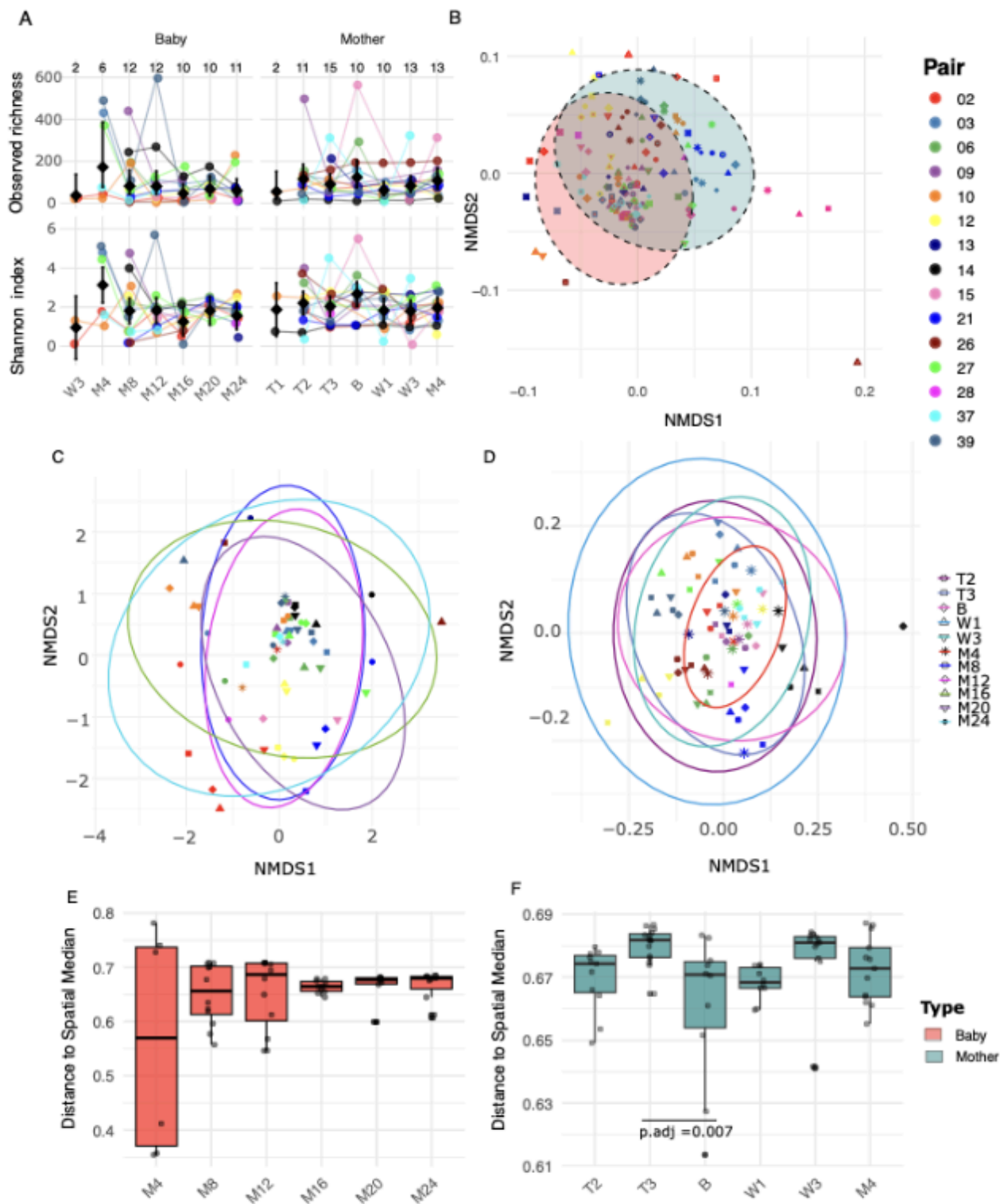


Figure 3

(A) Alpha diversity across months in infants (left) and mothers (right), showing observed richness (top) and Shannon diversity (bottom). Sample size of each time points are indicated above. Black diamonds and error bars represent model-estimated marginal means and their 95% confidence intervals; infant models were adjusted for delivery mode. (B) Global NMDS based on Bray-Curtis distances colored by baby-mother pairs show distinct clustering (NMDS stress = 0.203), with overlap between infant and maternal viromes; Pair:

PERMANOVA, $R^2 = 0.27478$, $p = 0.001$; Type: PERMANOVA $R^2 = 0.02057$, $p = 0.001$. (C) NMDS plots with ellipses for each month in infants except M4; (NMDS stress = 0.168), (D) NMDS plots with ellipses for each month in mothers;(NMDS stress = 0.228). (E) Beta-dispersion of distance to centroid over time in infants. (F) Beta-dispersion of distance to centroid over time in mothers; PERMDISP $p = 0.001$; Tukey HSD T3 vs B, FDR=0.007.

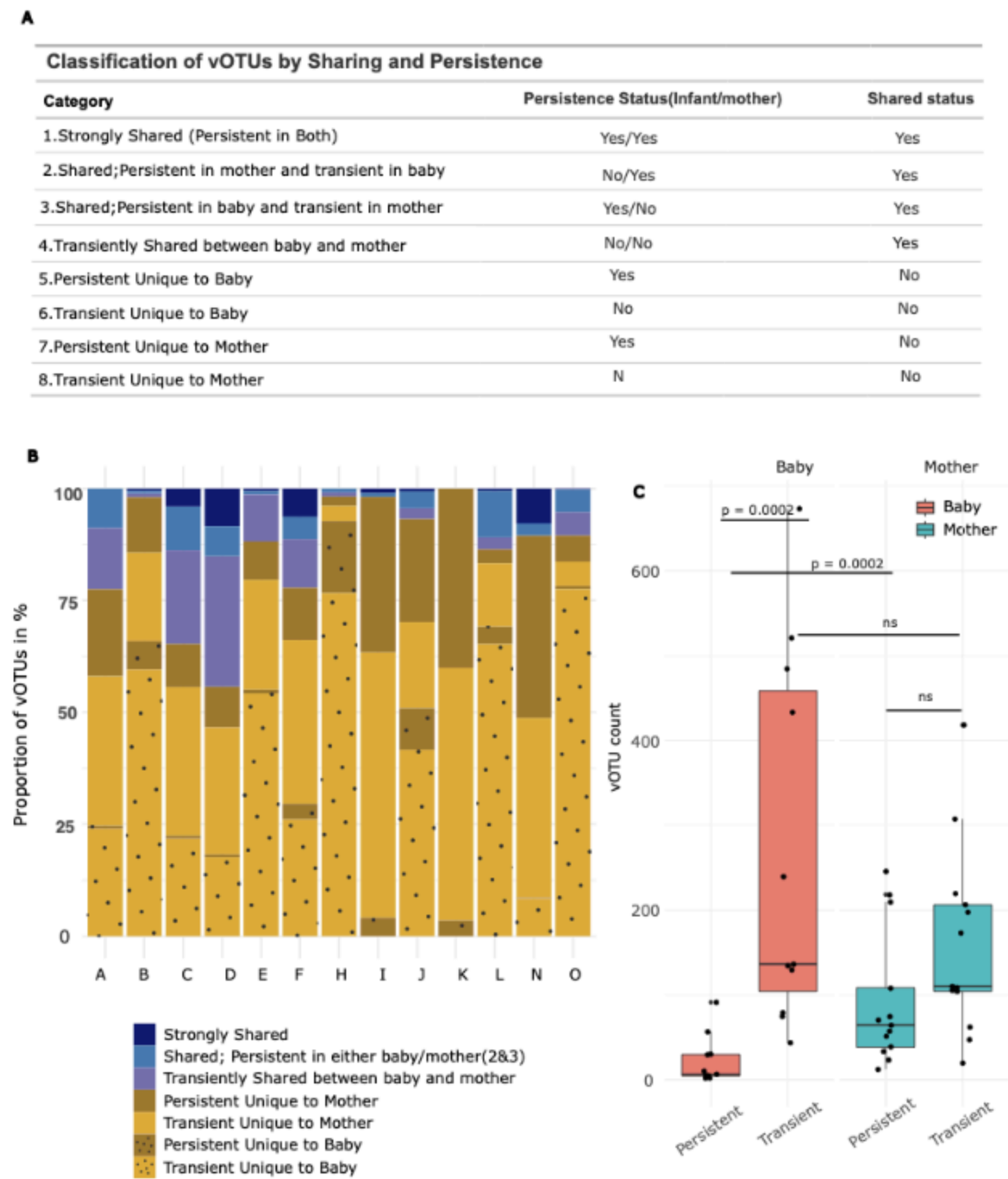


Figure 4

(A) Schematic illustration of categories of vOTUs based on the persistence and sharing between baby and mother. Two pairs, G & M were omitted from the analysis as there was just one infant sample available. (B) Proportional distribution of vOTUs in each category. The proportion of vOTU sharing varied widely between each pair, for instance in pair D ~44% of vOTUs were shared between baby and mother while pair K showed no overlap; For this figure categories 2 & 3 are combined. (C) Distribution of unique vOTUs in infants and mothers among persistent and transient categories; Infant samples that had just two baby samples each were omitted from this analysis; Two-sided Wilcoxon rank sum test was used to report the difference.

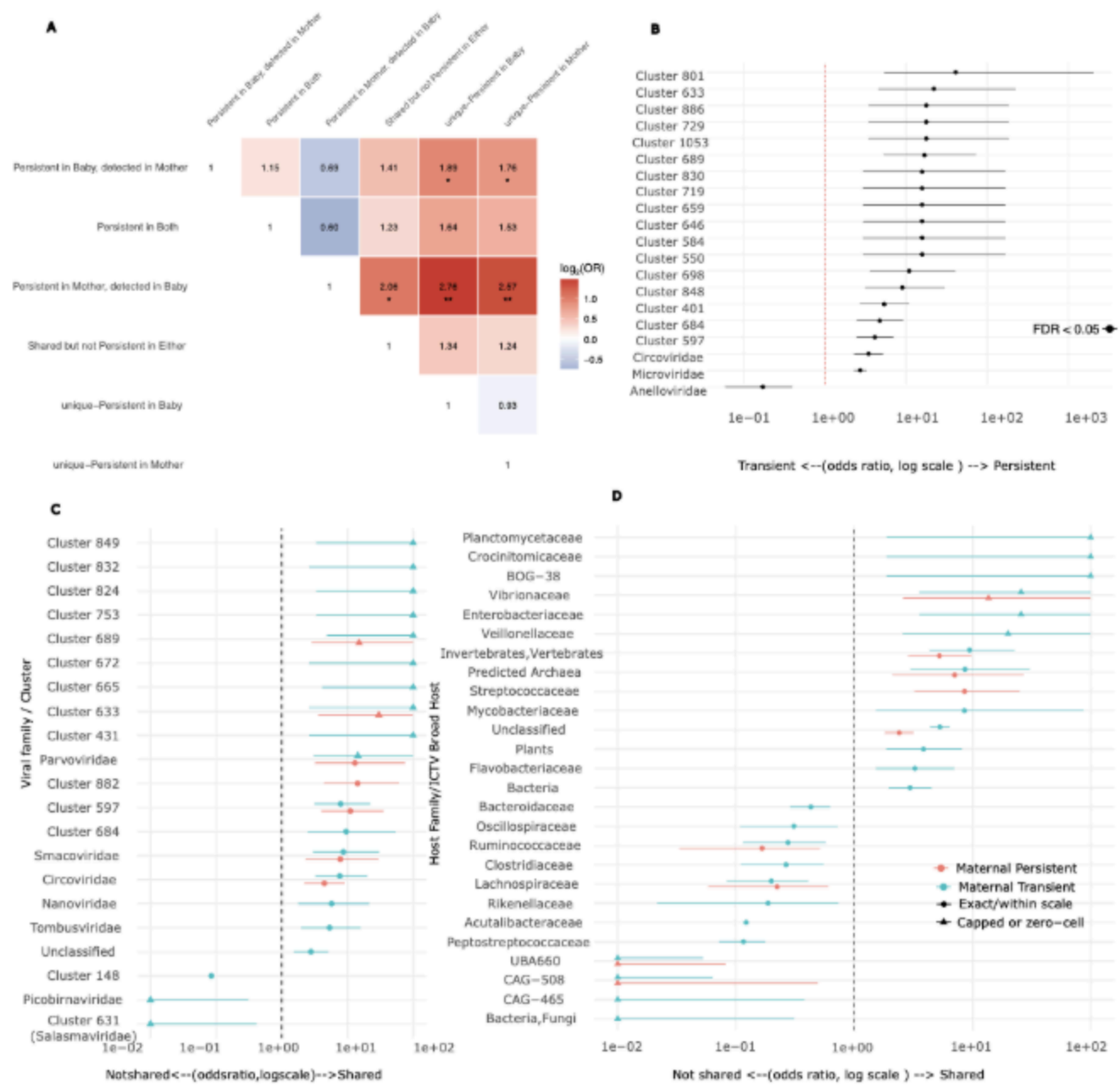


Figure 5

A) Odds-ratio matrix comparing the prevalence of positive selection across vOTU persistence subclasses. Each cell shows the odds of positive selection of vOTUs in the row class relative to the column class (numeric value = OR; tile colour = log₂ (OR); diagonal fixed at 1). Asterisks mark FDR-adjusted significance (*

$p \leq 0.05$; ** $p \leq 0.01$). (B) Forest plot of viral family associations with persistence versus transience (Fisher's OR, log scale) with 95% confidence intervals are shown. Note: the family Microviridae is now the class Microviricetes (C) Forest plot of viral families enriched/depleted (Fisher's OR, FDR <0.05) for sharing in maternal persistent and maternal transient subclass. (D) Forest plot of predicted host families enriched/depleted (Fisher's OR, FDR <0.05) for sharing in maternal persistent and maternal transient subclass.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [PEARLSupplementary.pdf](#)
- [Supplementarytablesrevised.xlsx](#)