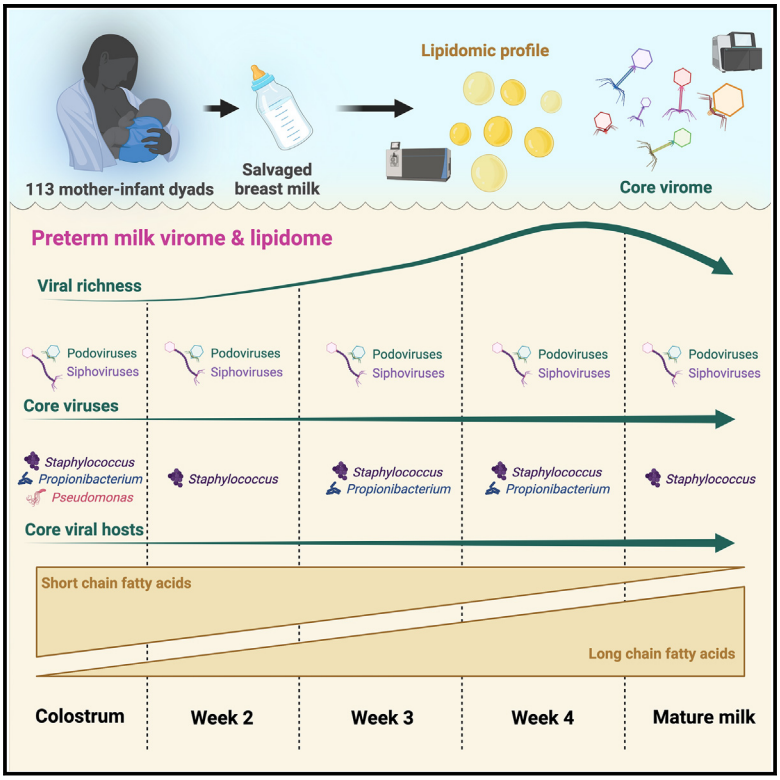
Cell Reports

Article

The core phageome and its interrelationship with preterm human milk lipids

Graphical abstract



Highlights

- Preterm human milk harbors core bacteriophage communities from the first week of lactation
- Preterm human milk bacteriophages mostly infect skin-commensal bacteria
- Core viruses and lipids/fatty acids are interrelated in preterm human milk (*in vivo*)
- Associations between fatty acid chain length and phage morphotype can be replicated *in vitro*

Authors

Wen C. Yew, Gregory R. Young, Andrew Nelson, ..., Janet E. Berrington, Nicholas D. Embleton, Darren L. Smith

Correspondence

darren.smith@northumbria.ac.uk

In brief

Yew et al. highlight that core viromes of preterm human milk develop over time and are dominated by bacteriophages infecting human skin commensals. Co-occurrence networks and *in vitro* association studies link phage morphotypes with fatty acid chain length. This highlights potential vehicles for vertical transmission of bacteriophages from mother to infant.

Yew et al., 2023, Cell Reports *42*, 113373
November 28, 2023
https://doi.org/10.1016/j.celrep.2023.113373

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

The core phageome and its interrelationship with preterm human milk lipids

Wen C. Yew,[1,6] Gregory R. Young,[1,2,6] Andrew Nelson,[1] William Cheung,[1] Christopher J. Stewart,[3] Simon H. Bridge,[1,3] Claire Granger,[3,4] Janet E. Berrington,[3,4] Nicholas D. Embleton,[3,5] and Darren L. Smith[1,2,7,*]

[1]Faculty of Health and Life Sciences, Northumbria University, Newcastle upon Tyne NE1 8ST, UK
[2]Hub for Biotechnology in the Built Environment, Northumbria University, Newcastle upon Tyne NE1 8ST, UK
[3]Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne NE1 7RU, UK
[4]Neonatal Medicine, Newcastle upon Tyne Hospitals National Health Service Foundation Trust, Newcastle upon Tyne NE7 7DN, UK
[5]Population Health Sciences Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne NE1 7RU, UK
[6]These authors contributed equally
[7]Lead contact
*Correspondence: darren.smith@northumbria.ac.uk
https://doi.org/10.1016/j.celrep.2023.113373

SUMMARY

Phages and lipids in human milk (HM) may benefit preterm infant health by preventing gastrointestinal pathobiont overgrowth and microbiome modulation. Lipid association may promote vertical transmission of phages to the infant. Despite this, interrelationships between lipids and phages are poorly characterized in preterm HM. Shotgun metagenomics and untargeted lipidomics of phage and lipid profiles from 99 preterm HM samples reveals that phages are abundant and prevalent from the first week and throughout the first 100 days of lactation. Phage-host richness of preterm HM increases longitudinally. Core phage communities characterized by *Staphylococcus*- and *Propionibacterium*-infecting phages are significantly correlated with long-chain fatty acid abundances over lactational age. We report here a phage-lipid interaction in preterm HM, highlighting the potential importance of phage carriage in preterm HM. These results reveal possible strategies for phage carriage in HM and their importance in early-life microbiota development.

INTRODUCTION

Very preterm infants born at <32 weeks gestation are at risk of complications such as necrotizing enterocolitis (NEC) and late-onset sepsis (LOS).[1,2] Microbial community development plays an important role in disease onset. Greater proportions of Proteobacteria,[3,4] as well as low bacterial[5,6] and viral,[7] community stability have been associated with NEC. However, it remains difficult to determine how these microbial communities establish or identify specific drivers of healthy microbiome development. This issue is further complicated by the substantial clinical intervention required while caring for preterm neonates compared to term infants.

Human milk (HM) provides the basis for the optimal diet in preterm infants[8,9] and reduced rates of NEC.[10–12] Many studies have sought to characterize the composition of HM, identifying a myriad of bioactive compounds, cells, and organisms including lipids,[13,14] oligosaccharides,[15,16] bacteria,[17,18] and proteins.[19,20] While providing nutrients essential to neonatal growth, HM exerts a major impact on the microbial communities in the gastrointestinal tract (GIT).[21] Together, these direct (nutritional) and indirect (microbial) influences of HM impact neonatal health and disease. While recent data improve our understanding of how the presence and abundance of oligosaccharides[15] and proteins[19] in HM shape the gut microbiota, few studies have identified how viruses found in HM can shape microbial populations in the neonatal GIT. This is particularly important for preterm infants, who are at the greatest risk of NEC and LOS.[1]

Bacteriophages (also known as phages) are viruses that infect bacteria and are ubiquitous in nature alongside their cellular hosts. Recent studies have identified phages in term HM[22,23] and suggest these may be vertically transmitted from mother to infant,[24] potentially shaping initial colonization of the infant gut.[25] Current data focus on mothers of healthy, term infants and may not be applicable to preterm infants, for whom practices such as reduced parental skin-to-skin contact,[17] antibiotics,[26] parenteral feeding, and reduced milk feeding[27] represent barriers to typical microbial assembly, in part by interfering with vertical microbiome transmission.

Lipids provide around 5% of the total composition by weight but around 50% of the available energy in HM.[28] Triglycerides represent a major component, but HM also contains glycoproteins such as mucin that are encased by the membrane of phospholipids, known as milk-fat globules (MFGs).[29] Mucins can bind bacteria and viruses and reduce their infectivity within the GIT of infants.[30–32] Previously, our group has isolated phages from the lipid fraction of donor HM,[33] suggesting that phages could be associated with the membrane of MFGs and that this may serve as protection for phages during transmission from mother to infant. Although substantial work has already investigated lipids in HM, mechanisms of action and interactions *in vivo* are poorly understood. One reason may be the lower resolution

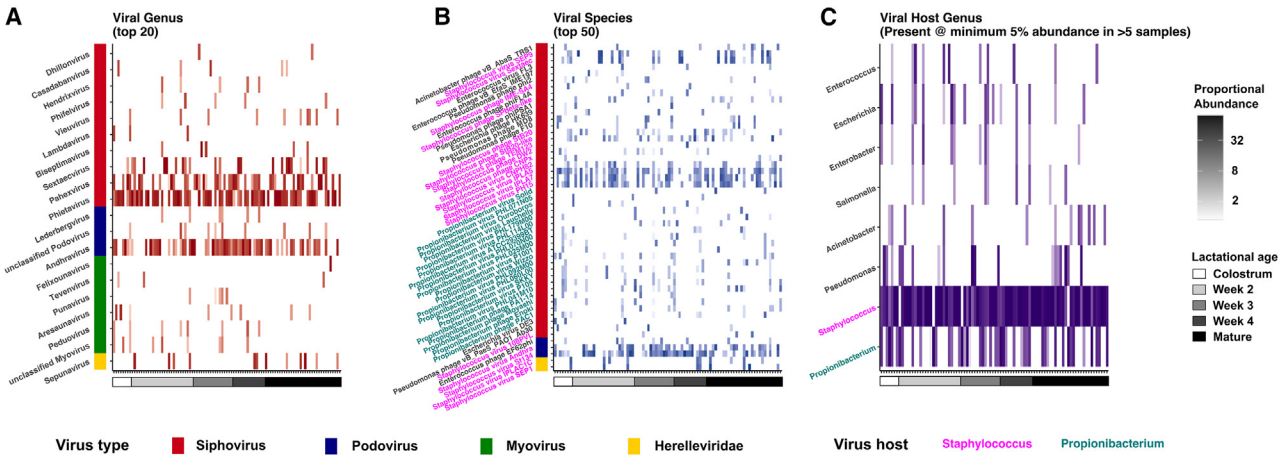


Figure 1. dsDNA phage community compositions in preterm HM during the first 100 days post-delivery

Intensity of heatmap tile colors indicates proportional abundance of taxonomic features, arranged on the y axis of each panel. The genus (A), species (B), and viral-host (C) levels of taxonomic composition of HM samples are shown. Samples are arranged by lactational age on the x axis of each panel. Virus types (previously families) of genus and species annotations are color-coded on the y axis. Viral species names (B) are colored by their inferred host specificity: *Staphylococcus*, pink; *Propionibacterium*, green.

and sensitivity than current lipidomic workflows allow. Furthermore, the fatty acid compositions of preterm HM are different from that of term HM.[34,35] With advances in gas chromatography, Moltó-Puigmartí and colleagues reported significantly higher levels of lipids in preterm HM than those in term HM.[13] High concentrations of lipids in preterm HM may be useful to premature neonates who require enhanced caloric intake[36] and may increase carriage of viable phages to the GIT, thereby contributing to health and disease.

The majority of phages previously reported in the human gut are double-stranded DNA (dsDNA) viruses[37–39]; this study therefore aimed to characterize the dsDNA phage community composition and lipidome of preterm HM across different lactational ages using a validated method for isolating and characterizing phages in the lipid fraction.[33] We applied shotgun metagenomics and untargeted lipidomics to identify the core dsDNA virome and lipidome of preterm HM from a large cohort (n = 113) of preterm infants from a single neonatal intensive care unit (NICU) and explore the associations between the preterm HM phageome and lipidome.

RESULTS

A total of 118 preterm HM samples were successfully salvaged from 113 participants who delivered preterm neonates at the Royal Victoria Infirmary (Newcastle upon Tyne, UK).[40] The expressed preterm HM was collected over the first 100 days post-delivery. After dsDNA virome sequencing, 19 HM samples were removed from analysis following quality control (QC) filtering (as defined in the STAR Methods), yielding 99 preterm HM samples for further analysis. No viral contigs were identified in over half (56%) of the sequenced colostrum samples. Colostrum samples that were salvaged earliest failed to yield viral contigs (median = 4 days, interquartile range [IQR] 3.75–5.25 days). After the first week of lactation, viral contigs were consistently observed in HM samples (Table S1). No significant differences in the delivery mode, gestational age, or birth weight at delivery of neonates were observed between sampling numbers and lactational ages (all $p > 0.05$; Table S1).

Phages are the predominant dsDNA viruses in preterm HM

Siphoviruses were the most abundant dsDNA virus type in preterm HM and were well dispersed throughout the sampling period (median prop. abund. = 80%; IQR = 50.0%–97.9%). followed by Podoviruses (median prop. abund. = 7.9%; IQR = 0.0–28.7%) and Myoviruses (median prop. abund. = 0.0; IQR = 0.0–4.3). Eukaryotic viruses represented 0.4% of the total dsDNA viral contigs observed (Figure S1). Phage-host taxonomic specificity was inferred from nucleotide similarity to reference sequences. Phages mostly infected *Staphylococcus* (41 species) and *Propionibacterium* (52 species) throughout the sampling period. Further phages were observed with host specificity for *Enterobacteria/Klebsiella* (23 species), *Escherichia* (20 species), *Pseudomonas* (18 species), and *Enterococcus* (8 species) (Figure 1).

There is a core phageome in preterm HM

dsDNA phages from the traditional Sipho- and Podo-virus classifications made up the core community of dsDNA viruses in preterm HM. Relationships between distribution and abundance of phage taxa were interrogated across lactational age via linear regression at genus, species, and phage-host taxonomic levels. At all taxonomic levels, significant positive relationships were observed between taxonomic distribution and abundance (p [Bonferroni corrected] < 0.001, R2 > 0.5; Figure S2). Highly abundant phages (>50%) were also highly prevalent (>50%) and therefore classified as core virome members at each lactational age (Table 1). Phages putatively infecting *Staphylococcus* were identified as core in all lactational ages and were the only core phages identified overall (i.e., across all lactational ages analyzed together). Although

Table 1. The significance and goodness of fit presented in linear regression models by comparing viral taxonomic distribution and abundance across each lactational age

	Lactation age group					
	Overall	Colostrum	Week 2	Week 3	Week 4	Mature (>1 month)
Linear model						
Genus (R2, p)	0.92, 3.2×10^{-15}	0.91, 1.4×10^{-7}	0.91, 3.2×10^{-15}	0.88, 3.2×10^{-12}	0.94, 4.8×10^{-10}	0.86, 4.8×10^{-11}
Species (R2, p)	0.84, 3.2×10^{-15}	0.57, 3.2×10^{-12}	0.81, 3.2×10^{-15}	0.69, 3.2×10^{-15}	0.73, 3.2×10^{-15}	0.72, 3.2×10^{-15}
Host (R2, p)	0.90, 3.2×10^{-15}	0.94, 6.4×10^{-7}	0.91, 3.2×10^{-9}	0.85, 3.2×10^{-7}	0.88, 1.6×10^{-4}	0.94, $<9.6 \times 10^{-9}$
Core features (50% distribution)						
Genus	*Andhravirus*, *Phietavirus*	*Andhravirus*, *Pahexvirus*, *Phietavirus*	*Andhravirus*, *Phietavirus*	*Andhravirus*, *Phietavirus*, *Pahexvirus*	*Andhravirus*, *Phietavirus*, *Pahexvirus*	*Andhravirus*, *Phietavirus*
Species	*Staphylococcus virus St134*	*Staphylococcus virus St134*, *Staphylococcus virus IPLA5*	*Staphylococcus virus IPLA5*, *Staphylococcus virus St134*, *Staphylococcus virus PH15*	*Staphylococcus virus St134*, *Staphylococcus virus Andhra*	*Staphylococcus virus St134*	–
Host	*Staphylococcus*	*Staphylococcus*, *Pseudomonas*, *Propionibacterium*	*Staphylococcus*	*Staphylococcus*, *Propionibacterium*	*Staphylococcus*, *Propionibacterium*	*Staphylococcus*

All relationships are significant following Bonferroni correction, enabling identification of core viral taxa (present in >50% of all preterm HM samples) across lactational age.

most *Staphylococcus*-infecting phages were Siphoviruses (36 viral species), *Staphylococcus virus St134*, which is a persistent core virus throughout the first month of life, is a Podovirus in the *Andhravirus* genus. No *Staphylococcus*-infecting phages were Myoviruses. Likewise, no *Propionibacterium* phages were Myoviruses. All 53 species of *Propionibacterium* phages identified in this study were Siphoviruses. The majority of Myovirus phages observed in this study consisted of phages putatively infecting commonly observed pathobionts in the GIT of preterm infants such as *Escherichia*, *Enterobacter*, and *Klebsiella* as well as the commensal *Lactobacillus*. Six phages putatively infecting *Pseudomonas* were identified as core in colostrum. These six *Phietavirus* phages were all Siphoviruses but were not identified as core in any subsequent lactational ages.

Preterm HM phage communities expand over lactation

Patterns of rarefied richness at the phage genus and species levels generally increased between colostrum and week 3 before dropping at the end of the first month of lactation (Figure 2). Reductions in phage richness after the first month of lactation were significant at the genus ($p = 0.04$) and host levels ($p = 0.01$; Table S2). While only trends were observed at the phage genus and species levels, significant associations were observed between viral host richness and lactational age ($p = 0.02$). In line with phage genus and species richness, the predicted number of viral hosts infected by phages isolated from HM increased over the first 3 weeks of lactation before falling after 4 weeks of lactation ($p = 0.01$; Figure 2). Shannon diversity of phages identified in HM remained consistent over lactation. Gestational age and delivery mode were included in general mixed linear models, but neither were significantly associated with phage richness or diversity of preterm HM (Tables S2 and S3). Phage community compositions remained consistent across lactation (Figure 2; Table S4). Colostrum and mature milk were both defined by low phage richness and consistently clustered together at the genus and species levels. At the viral host level, colostrum clustered separately from HM viral compositions at all other lactational ages, but these community differences were not significant (Figure 2; Table S4).

Increased phage richness contrasted by consistent Shannon diversity and community composition across lactational age suggested an increase in dominance of single-phage species. *MaAsLin2* was used to identify significant associations between lactational age and abundance of individual phage species. Gestational ages at birth and delivery mode were included in *MaAsLin2* models as covariates. *Staphylococcus virus St134* was associated with lactational age, being significantly more abundant in week 4 HM than colostrum ($p < 0.001$, $q = 0.04$). Like phage richness, proportional abundance of *Staphylococcus virus St134* in HM dropped substantially following the first month of life. Levels of *Staphylococcus virus St134* in mature milk were similar to those observed in colostrum (Figure S3). These results contrast with levels of the inferred bacterial host observed in samples, where proportional abundance of *Staphylococcus* spp. was not significantly altered. Moreover, proportional abundance of virus and host appear to mirror each other, suggesting a Lotka-Voltera community dynamic (Figure S3).

There is a core lipidome in preterm HM

A variety of different lipid entities (1,059) were observed across 3 lipid categories and 18 lipid classes. Positive ion mode identified more lipid entities (648) than the negative ion mode (411). Fatty acids (FAs) made up >25% of the total lipid content of preterm HM but were identified almost exclusively in the positive ion

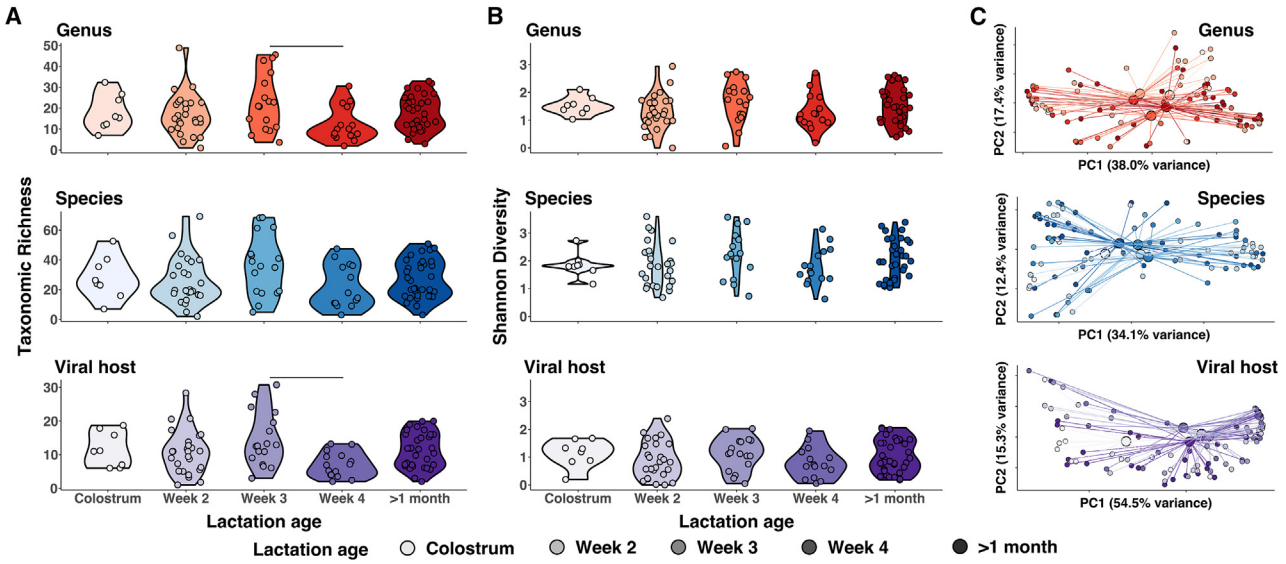


Figure 2. Characterising the longitudinal changes in dsDNA viral community composition in preterm human milk

(A–C) Rarefied dsDNA phage taxonomic richness (A), Shannon diversity (B), and Bray-Curtis dissimilarity (C) between preterm HM samples across lactation. Each point represents an individual sample and is colored by the taxonomic level at which analyses were performed (red, genus; blue, species; purple, viral host). Density clouds (A–C) are widest where the greatest proportions of samples reside and extend to the full range of data. Significant differences in alpha diversity metrics between lactational ages ($p < 0.05$) are highlighted by solid lines between density clouds and corrected for multiple hypothesis testing by Tukey's method. (D) Points in ordinations are linked to a centroid, illustrating the average community in preterm HM samples at that specific lactational age. Increasing lactational age is illustrated in (A)–(D) by increasing intensity of color.

mode. Only 4 FAs were identified in the negative ion mode, which also failed to classify lipid esterification (Figure S5). In the positive ion mode, triglyceride FAs were the most abundant FA type, accounting for >60% of the total FA content of preterm HM across lactational age.

Long-chain FAs (LCFAs) and polyunsaturated FAs became more proportionally abundant than short-chain or saturated FAs with increasing lactational age (Figure 3). Phospholipids associated with cellular membranes were a dominant lipid category, making up ~42% of the total lipid composition of preterm HM. Dominant phospholipid classes identified in both ion modes included the common mammalian and bacterial cell membrane component[41] phosphatidylethanolamine (8% proportional abundance) and the mammalian cell membrane component[42] phosphatidylcholine (9%). Other mammalian lipids identified included 3 cardiolipin entities (74C) and very-long-chain (VLC) sphingomyelins, which are important for neonatal brain development and neonatal growth[43,44] (Figure S4).

Longitudinal changes in preterm HM lipidome are related to phage carriage

Phage communities and lipid compositions in preterm HM are linked to lactational age. Compared to mature milk, 5 different FA types and 22 individual lipid entities were differentially abundant in HM expressed within the first month of lactation (Figure 4A). *MaAsLin2* was used to identify significantly altered abundance of lipids and associated phages across lactational ages ($p < 0.05$, $q < 0.1$). Short-chain saturated triglycerides were significantly more abundant in colostrum than mature milk ($p = 0.002$, $q = 0.08$). Specifically, 4-carbon-chain triglycerides were more abundant ($p < 0.001$, $q = 0.04$). Conversely, LC polyunsaturated FAs ($p = 0.002$, $q = 0.09$) and VLC monounsaturated triglycerides ($p \leq 0.001$, $q = 0.007$) were significantly less abundant in colostrum than mature milk. Triglycerides with 16- ($p \leq 0.001$, $q = 0.07$) and 18-carbon-length tails ($p \leq 0.001$, $q = 0.07$) were significantly more abundant in mature HM than colostrum (Figure 4A). At the end of the first month of life, lipid compositions had evolved from those of colostrum, with greater proportions of LCFAs observed. Indeed, LC polyunsaturated diglycerides ($p = 0.002$, $q = 0.09$) and 16-carbon triglycerides ($p \leq 0.001$, $q = 0.09$) were significantly more abundant in week 4 HM than mature milk. The increase in LCFAs correlated with increased proportions of *Staphylococcus virus St134* ($p \leq 0.001$, $q = 0.04$).

Associations between lipid and phage co-occurrence were observed in HM samples ($R^2 > 0.4$). Sparse partial least-squares discriminant analysis (sPLS-DA) models were built to identify phage and lipid features associated with lactational age. Proportional abundance of VLC polyunsaturated FAs as well as LC poly-/monounsaturated FAs were positively associated with relative abundance of dsDNA viral types dominated by *Staphylococcus*- and *Propionibacterium*-infecting phages, namely Siphoviruses and Podoviruses (Figure 4B). In contrast, Siphovirus and Podovirus phages were negatively associated with short- and medium-chain saturated triglycerides (Figure 4B). When the whole lipidome was correlated with viral taxa at the species level, LC triglycerides (16C) were negatively associated with a group of Siphoviruses putatively infecting *Staphylococcus* and *Propionibacterium* (Figure 4C). These same phages were negatively associated with a VLC (40C) sphingomyelin. Short-chain

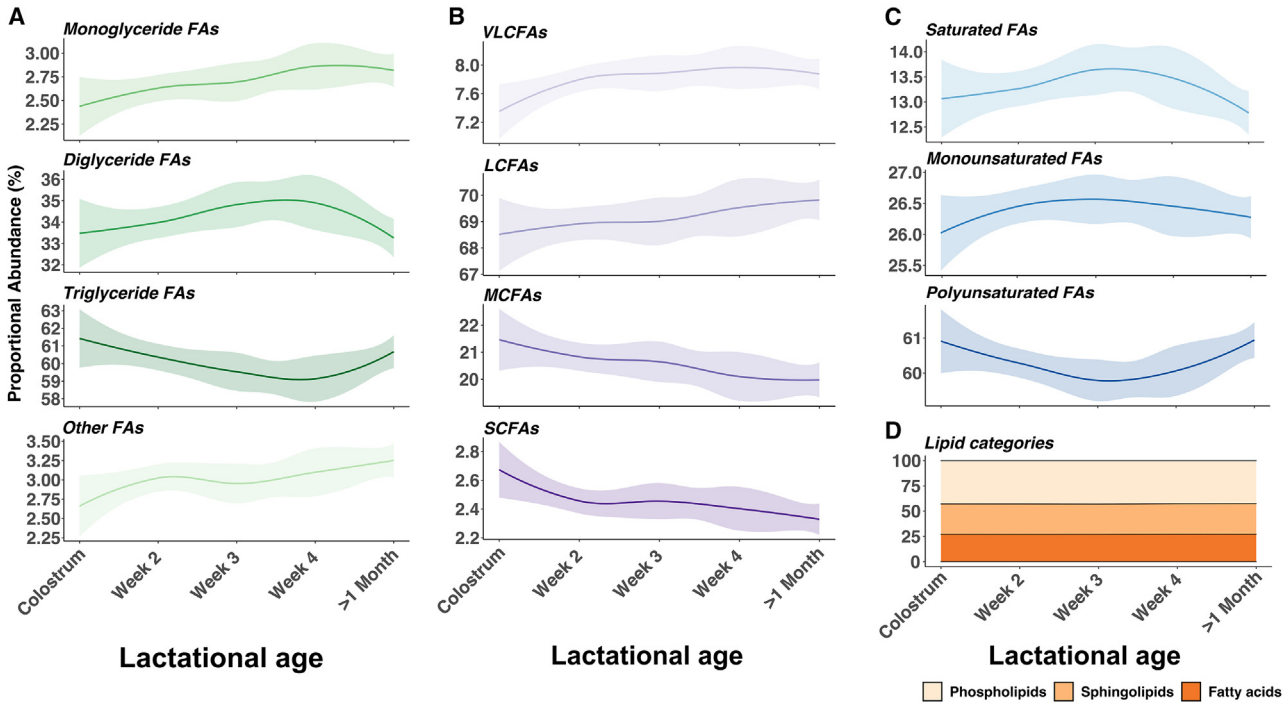


Figure 3. Mapping the longitudinal changes of preterm human milk fatty acid and lipid composition

Composition of fatty acid esterification (A), chain length (B) saturation (C), and wider lipid category (D) abundances observed in all preterm HM samples across lactational age. Lines in the fatty acid images (A–C) illustrate the average abundance of each fatty acid type. Lines are colored by stratification (esterification, greens; chain length, purples; saturation, blues). Shaded areas around the lines represent the 95% confidence interval at each lactational age. Each bar in the lipid categories image (D) represents a distinct lipid category and includes lipids identified using both positive and negative mass spectrometric ion mode (FA, fatty acid; SCFAs, short-chain fatty acids; MCFAs, medium-chain fatty acids; LCFAs, long-chain fatty acids; VLCFAs, very-long-chain fatty acids).

(4C) triglycerides were associated with phages putatively infecting *Escherichia* phages.

Siphoviruses were the most connected phage morphotype within the computed networks and were frequently associated with a bacterial cell membrane component, phosphatidylethanolamine (16C). Higher connectedness of Siphoviruses and lipids may be because phage tail length is associated with lipid binding in HM or because we observed more Siphoviruses than other phage types in this study. Nonetheless, multiple correlations between Siphoviruses and lipid entities, specifically sphingolipids and phospholipids, were observed (Figure 4C).

Phage-lipid associations are quantifiable *in vitro*

Plaque-forming units of siphoviral, but not myoviral, coliphages are reduced in increased concentrations of VLCFAs. We sought to validate the associations observed between phage abundance and milk lipids in co-occurrence networks by incubating pure viral lysates of sipho- and myoviral morphotypes with increasing concentrations of linoleic acid. Linoleic acid is a VLCFA (18C) reported in preterm HM at an approximate concentration of ~0.006 g/mL.[45] We quantified active populations of the *Escherichia coli*-infecting phages T4 (Myovirus) and λ (Siphovirus) by counting plaque-forming units when incubated with their bacterial hosts in the presence of serial dilutions of linoleic acid running from log units above to log units below physiological concentrations.[42]

T4 phage was not associated with lipid concentration, with no significant change in phage infectivity observed with an increasing concentration of linoleic acid ($p = 0.66$). However, plaque-forming units of λ phage were significantly associated with linoleic acid concentrations ($p = 0.02$) (Figure 4D). Significant differences in plaque-forming units were observed between no linoleic acid control (mean = 3.4×10^9, SD = 7.8×10^8) and linoleic acid concentrations well above physiological levels (0.09 g/mL: mean = 5.4×10^8, SD = 2.6×10^8; $p = 0.0003$) but also at physiological levels (0.009 g/mL: mean = 1.06×10^9, SD = 6.0×10^7; $p = 0.0005$) and even at 10-fold lower concentrations than physiological levels (0.0009 g/mL: mean = 1.80×10^9, SD = 2.1×10^8; $p = 0.003$).

DISCUSSION

Our findings identify that dsDNA viral richness in preterm HM increases significantly over the first month of life. Among the dsDNA virome of preterm infants, we define a core phageome, dominated by Siphovirus phages, targeting bacteria previously associated with preterm infants. This occurs alongside a developing lipid profile over the first 100 days of maternal lactation. Importantly, we demonstrate a developing core phageome and its interrelationship with preterm HM lipids of different FA chain lengths. This study presents primary data highlighting lipids as a vehicle for vertical transmission of a core phageome.

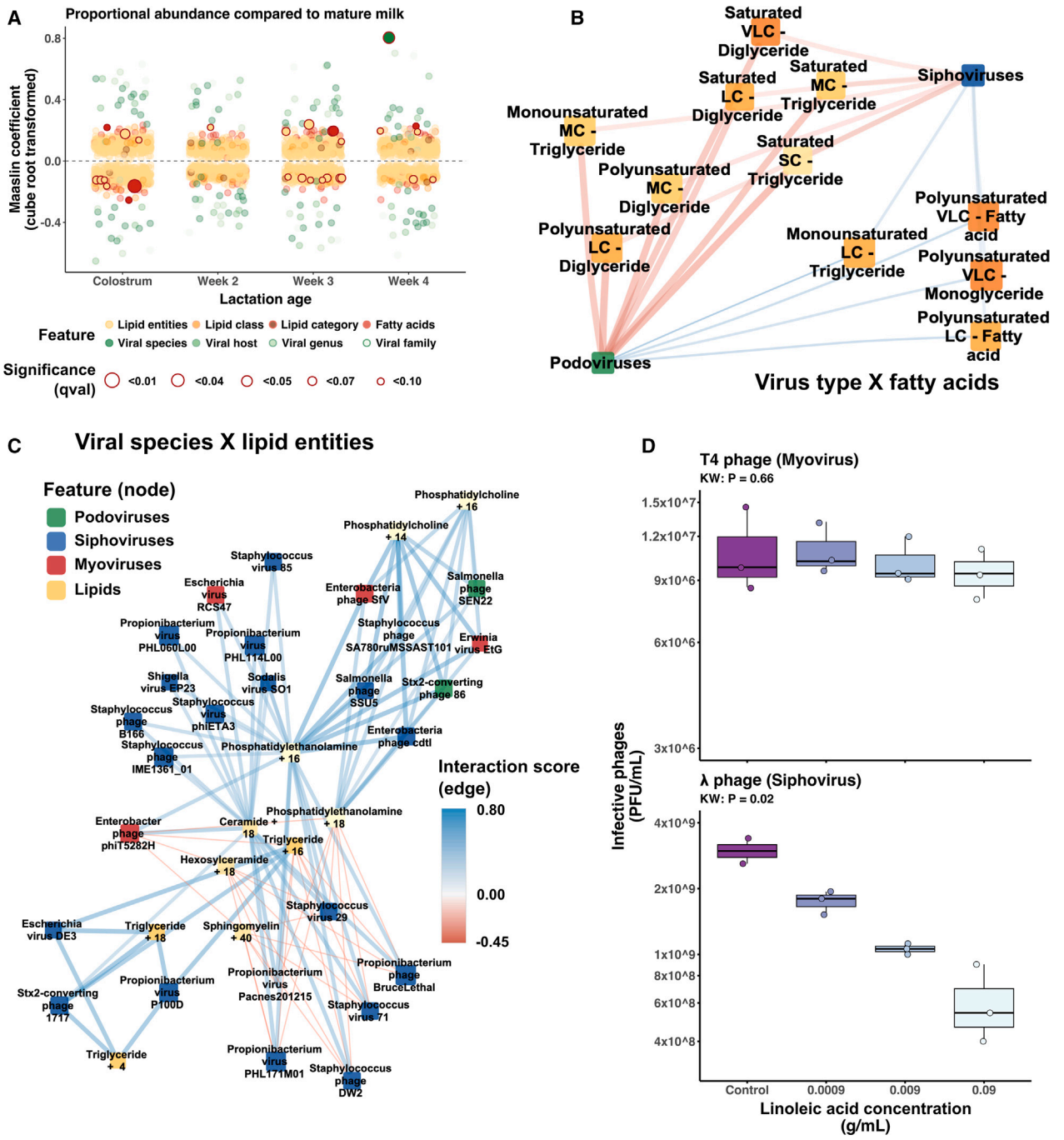


Figure 4. Associations between lipid and dsDNA viral compositions

(A) Differential lipid (yellows) or viral (greens) features are plotted as a single point according to lactational age (using >1 month lactation as a comparator group). Points above the dashed line were more abundant, while those below the dashed line were less abundant, at each lactational age. Features exhibiting significantly differential abundance at any lactational age are outlined red. Size of each point illustrates the significance level.

(B and C) Association networks illustrate co-occurrence of fatty acids and dsDNA viruses in preterm HM at virus-type (B) level as well as total lipids and viral species (C). Networks are self-organized with nodes representing lipids (yellow) and viral (colored by family) features. Nodes are connected by edges where associations are stronger than R2 = 0.4 and are colored by correlation (red, negatively associated; blue, positively associated). Weight of lines indicate the strength and number of connections between nodes.

(D) *In vitro* analyses explored the influence of lipid concentrations on phage infectivity, highlighting a reduction in plaque-forming units of λ phage (Siphovirus), but not T4 phage (Myovirus), with increasing concentrations of linoleic acid. Each point represents a single assay (n = 3 replicates per variable) (SC, short chain; MC, medium chain; LC, long chain; VLC, very long chain; KW, Kruskal-Wallis test).

Our results show that patterns of dsDNA viral richness in preterm HM oppose those seen previously in the term infant gut, with viral richness of preterm HM increasing with lactational age, while gut viral richness decreases.[7,38,39] Previous studies have described a reduction in viral diversity in the gut of healthy and preterm infants following birth but focused on stool samples, not HM.[7,37–39] Previous studies have also made specific attempts to explore communities of RNA and single-stranded DNA (ssDNA) viruses as well, which this study did not. No studies, to our knowledge, have explored the viral richness of preterm HM across different lactational ages.

A core phageome is observable within 1 week of lactation. The inability to identify phage contigs from very early preterm HM suggests that these initial colostrum samples may not harbor defined dsDNA phage communities, though RNA and ssDNA viruses may be present at this time. As lactational age increases, phages were more readily observable in HM, suggesting that dsDNA viruses move from transiently present to active community members as microbial communities within the milk are established. dsDNA phages were consistently observed in the core virome across lactation. Myo-, Sipho-, and Podoviruses were commonly identified in previous studies investigating the term HM virome[22,46,47]; however, these studies observed dominance of Podo- and Myoviruses with fewer Siphovirus phages than observed in these preterm HM samples. These results suggest that prematurity may play a role in HM virome composition.

Unlike term HM, Siphovirus phages were the dominant viral morphotype in preterm HM, as identified in this study. These primarily infected *Staphylococcus* and *Propionibacterium* spp. of bacteria, where both bacterial genera are commonly isolated skin commensals.[48,49] The increasing abundance of Podovirus phages, targeting *Staphylococcus epidermidis,* and Siphovirus phages, targeting *Staphylococcus* spp. and *Propionibacterium* spp., throughout lactation suggests a link between skin commensal bacteria and lactational age of the expressed milk phageome. In preterm infants, direct skin-skin contact (kangaroo mother care) may be limited in frequency and duration, and direct feeding at the breast tends not to occur before 32–34 weeks corrected gestation. Almost all HM provided to these infants is expressed and may be then stored refrigerated, frozen, or both.[40] In this study, samples were salvaged from feeding tube systems that deliver milk to the stomach of a preterm infant unable to feed directly at the breast. These important differences in feeding practices may explain the altered viromes of preterm compared to term HM.

Viruses infecting *Staphylococcus*, particularly *Staphylococcus virus St134*, were consistently found in preterm HM. *Staphylococcus* were the only dsDNA viral hosts identified across all lactational ages. Proportional abundance of these core *Staphylococcus*-infecting bacteriophages increased longitudinally. *Staphylococcus virus St134* is a persistent member of the core milk virome from delivery until the end of the first month of lactation. *Staphylococcus virus St134* and *Staphylococcus virus Andhra* (another highly prevalent and abundant dsDNA viral species observed here) form a homogeneous sub-cluster of *Staphylococcus*-infecting dsDNA Podovirus phages.[50] These phages were also highly abundant in a study by Dinleyici and colleagues investigating HM viromes from mothers delivering at term.[46] They are known to infect *Staphylococcus epidermidis*,[50] a common skin commensal,[51,52] providing further evidence of links between microbes on maternal skin and those in HM, even when direct breastfeeding is not possible.

Lipid compositions of preterm HM show longitudinal variance as FA chain lengths increase with lactational age. Previous studies have focused on comparing HM compositions expressed following preterm and term delivery,[13,14,34,35,53] whereas this study explores longitudinal development of the preterm HM lipidome with such sensitivity through the increased resolution of the mass spectrometry now available. This study also associates the interrelationship of the lipidome during lactation to the developing core virome of preterm HM.

The HM investigated in this study exhibited high proportions of SC saturated triglycerides in colostrum and increasing FA chain length with lactational age. These results contrast with those of previous studies, which observed reduced LCFA content with advancing lactational age.[13,14] The study by Moltó-Puigmartí and colleagues recruited 43 infants from European cohorts, but important differences exist between their and our study groups. HM analyzed in this study was most similar in gestational age to ten neonates classified as ‘‘very preterm’’ in the Moltó--Puigmartí study. Our results are from substantially more infant/mother pairs (n = 99) at around 1-week-greater gestational ages (mean gestational age [GA]: 26.7 vs. 27.7 weeks). Given the variability of HM composition over the first month post-delivery, these results suggest gestational age may also be critical in determining longitudinal variance in HM compositions. Analytical differences may also explain the contrasting results between studies. Moltó-Puigmartí and colleagues used FA methyl ester derivatization for gas chromatography-mass spectrometry (GCMS) analysis of HM FA profiles. This method requires post-derivatization normalization.[54] By contrast, this study employs liquid chromatography-tandem mass spectrometry (LC/MS-MS)-based analysis for untargeted profiling of global HM lipid profiles. This allows finer separation and greater lipid coverage compared to GC-based methods.

This study was not designed to assess changing lipid composition in line with nutritional demands of preterm infants; however, the longitudinal variance in preterm HM composition observed here provides further evidence of highly abundant SCFAs in colostrum that may be of importance to the neonate as an important utilizable energy source.[55] Alexandre-Gouabau and colleagues demonstrated that preterm HM rich in saturated FAs and low in LC triglycerides was associated with faster growth in a cohort of 26 preterm infant-mother dyads fed exclusively on HM.[43] Increasing lactational age was associated with greater relative abundance of LC polyunsaturated FAs in this study. LF polyunsaturated FAs in HM are likely to be beneficial to preterm infant brain and retina development.[56] Contrary to the results observed here, a study by Luukkainen and colleagues found LC polyunsaturated FAs to decrease after 1 month of lactation, leading them to recommend that donor milk provided to preterm infants be sourced solely from mothers who had recently delivered. Those results were based on GCMS data and may highlight the importance of the analytical methodology when comparing study results. This is further illustrated by the fact that no single ion mode was able to characterize the entire

lipidome of HM investigated here. Specifically, FA esterification was not detectable using the negative ion mode; therefore, parsing of FA into mono-, di-, and triglycerides was possible only in positive-ion-mode MS. These results are consistent with those of previous studies[57] and remain an important consideration for future studies exploring the lipid composition of HM.

Duranti and colleagues found *Bifidobacterium* prophages in HM samples and recipient's stool in term infants.[24] We find no evidence of *Bifidobacterium* phages in preterm HM, even though neonates in this cohort are routinely supplemented with *Bifidobacterium*-containing probiotics. While RNA and ssDNA phages may be overlooked due to the methods applied here, other important differences must also be considered. The mother-infant pairs recruited here are premature, unlike those in the study by Duranti, and existing data show that *Bifidobacterium* are rarely prevalent in the first few weeks in preterm infants unless provided as probiotic supplements.[58,59] This may impact viral composition, as abundance of *Bifidobacterium* as a host is lower in samples from infants born at earlier gestational ages.[60] Furthermore, taxonomic classifications made in this study were by alignment to the NCBI viral database. This contains only 29 phages infecting *Bifidobacterium,* with only 1 infecting *Bifidobacterium bifidum.* The remaining 28 were isolated from honeybee GITs.[61] Future studies may aim to expand the knowledge of phages infecting *Bifidobacterium* and their role in shaping the development of the preterm infant gut microbiota.

The establishment of gut microbial communities in the preterm infant is notoriously unstable.[62,63] Unlike term infants, there are many factors that prevent typical microbial colonization observed in preterm infants.[64,65] Development of stable communities, rich in strict anaerobes such as *Bifidobacterium* spp., are associated with positive health outcomes.[66,67] How this stability develops remains difficult to fully understand, especially for very-low-birth-weight (VLBW) preterm infants.[17] Microbial compositions of HM fed to VLBW infants is difficult to study due to the importance of the milk to the infant itself, meaning only small sample volumes remaining in feeding systems can be accessed.[8,9] HM promotes growth, engraftment of lactic acid bacteria, and progression to stable bacterial communities associated with greater health outcomes. We here provide further evidence for how bioactive constituents in preterm HM, vertically transmitted from mother to child, modulate microbial engraftment. This is the first study to illustrate a core phage community in preterm HM over the first month of lactation. We show that this core community is associated with lactational changes in lipid profiles. Combined, increased abundances of skin-commensal-infecting dsDNA bacteriophages and decreased abundance of SCFAs (<6C) suggest an interaction between carriage of specific phage groups and lipid compositions. *In vitro* studies found *E. coli* phage λ (a Siphovirus) were significantly less infective in a dose-dependent manner when incubated with linoleic acid (a VLC polyunsaturated FA). This phenomenon, however, was not observed for *E. coli* phage T4 (a Myovirus), which corroborates what is detailed in the network association seen in Figure 4B. We suggest here that certain phages may be carried on specific lipids and offer a novel delivery system for bacteriophages. Mechanisms of interaction between phages and lipids are unknown and may be due to specific phage evolution or may be circumstantial. Nonetheless, it is of practical importance to feeding infants in a NICU. As an example, HM lipid delivery to neonates is reduced using horizontally placed feeding syringes compared to vertically placed ones. The longitudinal patterns and relationship between HM phages and lipid composition observed here may be due to active phage-lipid association or may simply be due to the fact that phages infecting specific bacterial hosts, such as *Staphylococcus* and *Propionibacterim,* increase in abundance as the HM microbiome develops to harbor more of those bacteria. On the contrary, lipids in HM may provide a selective advantage for adsorption and infection of phages. A recent review highlighted the importance of lipids as functional and signaling components in the human gut microbiota.[68] We propose that phages carrying defined surface motifs specific for lipid-rich bacterial cell surfaces may associate with other charged lipid molecules in HM.

Future work should consider the probable evolved interrelationship between specific lipids and viral carriage. They should also note that phage binding to lipids may protect phage particles against low pH in the stomach. This will provide essential insight to understand how phages identified in HM might influence the developing infant gut microbiota.

Limitations of the study

This study was informed by previous works exploring viromes of the human and infant GIT observing dsDNA phages to be the predominant viruses.[37–39] The method employed here specifically focused on identifying dsDNA phages, omitting a reverse transcription and multiple displacement amplification step prior to metavirome library preparation. As such, RNA or ssDNA viruses may be under-represented in the data.

Due to the high-throughput metavirome approaches employed here, no attempt to isolate viable phages from preterm HM was undertaken as part of this study. Information regarding bacteriophage taxonomy, tail morphology, and host-infective range were inferred from nucleotide similarity of assembled viral contigs to the NCBI viral database. Characterization of phage communities observed is therefore focused on viral contigs with high nucleotide similarity to viruses in the NCBI viral database, which may fail to annotate previously uncharacterized viral entities. Unclassified virus annotation and host prediction using predicted protein approaches may improve phage recovery in future studies of preterm infant samples. Furthermore, while some database entries are annotated with definite viral tail morphologies (as identified by transmission electron microscopy [TEM]) or host range (as identified by host-susceptibility tests), this may not be true for all entries in the NCBI virus database.

Previous studies have shown that maternal factors can influence the composition of expressed milk.[69] The data presented here are all collected from mothers of infants within a single NICU in North East England. No data were collected on maternal diet, body mass, or other demographic information. The results of this study may therefore not be replicated in cohorts of mothers delivering preterm in different regions or with different environmental exposures. Future studies may validate these data.

To eliminate the risk of reducing volumes of HM available to the neonates enrolled in this study, all preterm HM samples characterized are salvaged from feeding apparatus following

completion of feeds. Because of this, preterm HM compositions reported are representative of what is actually fed to the infant, not necessarily the composition of freshly expressed preterm HM. Nonetheless, this represents an important advance in understanding how bioactives in preterm HM impact on neonatal health and disease.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Participant cohort
  - Lactational ages
- METHOD DETAILS
  - Sample salvaging, storage and processing
  - Phage isolation, DNA extraction and shotgun metagenomics
  - Bioinformatics processing
  - Bacterial community analyses from metagenome data
  - Lipid extraction, lipidomic data alignment and peak table generation
  - In-vitro assessment of lipid on phage infectivity
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Statistical analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2023.113373.

ACKNOWLEDGMENTS

This project is funded by Action Medical Research (AMR) for children (grant reference no. GN2730). We also thank the research nurses, clinical staff, and all participant families at the Royal Victoria Infirmary, Newcastle upon Tyne, UK, for their willingness to allow the use of residual samples.

AUTHOR CONTRIBUTIONS

W.Y. performed laboratory experiments and read-based viral sequence processing. G.Y. performed assembly-based viral sequence processing and statistical analysis on phageome and lipidome. A.N. prepared and sequenced the DNA libraries. W.C. performed lipidomics laboratory experiments. J.B., C.G., and N.E. enrolled participants in the clinical study and arranged sampling. D.S., J.B., N.E., S.H.B., and A.N. conceived the grant application submitted to Action Medical Research. W.Y., G.Y., and D.S. wrote the manuscript. All authors reviewed and approved the manuscript before submission.

DECLARATION OF INTERESTS

N.E. and J.B. declare research funding paid to their employing institution from Prolacta Biosciences US, Neokare, and Danone Early Life Nutrition, and both declare lecture honoraria from Nestle Nutrition Institute.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: March 24, 2023
Revised: August 9, 2023
Accepted: October 18, 2023

REFERENCES

1. Rose, A.T., and Patel, R.M. (2018). A critical analysis of risk factors for necrotizing enterocolitis. Semin. Fetal Neonatal Med. *23*, 374–379. https://doi.org/10.1016/j.siny.2018.07.005.
2. Israel, E.J. (1994). Neonatal Necrotizing Enterocolitis, a Disease of the Immature Intestinal Mucosal Barrier. Acta Paediatr. *396*, 27–32. https://doi.org/10.1111/j.1651-2227.1994.tb13238.x.
3. Warner, B.B., Deych, E., Zhou, Y., Hall-Moore, C., Weinstock, G.M., Sodergren, E., Shaikh, N., Hoffmann, J.A., Linneman, L.A., Hamvas, A., et al. (2016). Gut bacteria dysbiosis and necrotising enterocolitis in very low birthweight infants: A prospective case-control study. Lancet *387*, 1928–1936. https://doi.org/10.1016/S0140-6736(16)00081-7.
4. Pammi, M., Cope, J., Tarr, P.I., Warner, B.B., Morrow, A.L., Mai, V., Gregory, K.E., Kroll, J.S., McMurtry, V., Ferris, M.J., et al. (2017). Intestinal dysbiosis in preterm infants preceding necrotizing enterocolitis: a systematic review and meta-analysis. Microbiome *5*, 31. https://doi.org/10.1186/s40168-017-0248-8.
5. Stewart, C.J., Embleton, N.D., Marrs, E.C.L., Smith, D.P., Nelson, A., Abdulkadir, B., Skeath, T., Petrosino, J.F., Perry, J.D., Berrington, J.E., and Cummings, S.P. (2016). Temporal bacterial and metabolic development of the preterm gut reveals specific signatures in health and disease. Microbiome *4*, 67. https://doi.org/10.1186/s40168-016-0216-8.
6. Yee, A.L., Miller, E., Dishaw, L.J., Gordon, J.M., Ji, M., Dutra, S., Ho, T.T.B., Gilbert, J.A., and Groer, M. (2019). Longitudinal Microbiome Composition and Stability Correlate with Increased Weight and Length of Very-Low-Birth-Weight Infants. mSystems *4*, e00229-18. https://doi.org/10.1128/mSystems.00229-18.
7. Kaelin, E.A., Rodriguez, C., Hall-Moore, C., Hoffmann, J.A., Linneman, L.A., Ndao, I.M., Warner, B.B., Tarr, P.I., Holtz, L.R., and Lim, E.S. (2022). Longitudinal gut virome analysis identifies specific viral signatures that precede necrotizing enterocolitis onset in preterm infants. Nat. Microbiol. *7*, 653–662. https://doi.org/10.1038/s41564-022-01096-x.
8. Walker, A. (2010). Breast Milk as the Gold Standard for Protective Nutrients. J. Pediatr. *156*, S3–S7. https://doi.org/10.1016/j.jpeds.2009.11.021.
9. Corpeleijn, W.E., Kouwenhoven, S.M.P., Paap, M.C., van Vliet, I., Scheerder, I., Muizer, Y., Helder, O.K., van Goudoever, J.B., and Vermeulen, M.J. (2012). Intake of Own Mother's Milk during the First Days of Life Is Associated with Decreased Morbidity and Mortality in Very Low Birth Weight Infants during the First 60 Days of Life. Neonatology *102*, 276–281. https://doi.org/10.1159/000341335.
10. Embleton, N., and Cleminson, J. (2017). Randomized trial of exclusive human milk versus preterm formula diets in extremely premature infants. Acta Paediatr. *106*, 1538. https://doi.org/10.1111/apa.13820.
11. Sullivan, S., Schanler, R.J., Kim, J.H., Patel, A.L., Trawöger, R., Kiechl-Kohlendorfer, U., Chan, G.M., Blanco, C.L., Abrams, S., Cotten, C.M., et al. (2010). An exclusively human milk-based diet is associated with a lower rate of necrotizing enterocolitis than a diet of human milk and bovine milk-based products. J. Pediatr. *156*, 562–567.e1. https://doi.org/10.1016/j.jpeds.2009.10.040.
12. Herrmann, K., and Carroll, K. (2014). An Exclusively Human Milk Diet Reduces Necrotizing Enterocolitis. Breastfeed. Med. *9*, 184–190. https://doi.org/10.1089/bfm.2013.0121.
13. Moltó-Puigmartí, C., Castellote, A.I., Carbonell-Estrany, X., and López-Sabater, M.C. (2011). Differences in fat content and fatty acid proportions

among colostrum, transitional, and mature milk from women delivering very preterm, preterm, and term infants. Clin. Nutr. *30*, 116–123. https://doi.org/10.1016/j.clnu.2010.07.013.

14. Bitman, J., Wood, L., Hamosh, M., Hamosh, P., and Mehta, N.R. (1983). Comparison of the lipid composition of breast milk from mothers of term and preterm infants. Am. J. Clin. Nutr. *38*, 300–312. https://doi.org/10.1093/ajcn/38.2.300.
15. Masi, A.C., Embleton, N.D., Lamb, C.A., Young, G., Granger, C.L., Najera, J., Smith, D.P., Hoffman, K.L., Petrosino, J.F., Bode, L., et al. (2021). Human milk oligosaccharide DSLNT and gut microbiome in preterm infants predicts necrotising enterocolitis. Gut *70*, 2273–2282. https://doi.org/10.1136/gutjnl-2020-322771.
16. Masi, A.C., and Stewart, C.J. (2022). Untangling human milk oligosaccharides and infant gut microbiome. iScience *25*, 103542. https://doi.org/10.1016/j.isci.2021.103542.
17. Young, G.R., van der Gast, C.J., Smith, D.L., Berrington, J.E., Embleton, N.D., and Lanyon, C. (2020). Acquisition and Development of the Extremely Preterm Infant Microbiota Across Multiple Anatomical Sites. J. Pediatr. Gastroenterol. Nutr. *70*, 12–19. https://doi.org/10.1097/MPG.0000000000002549.
18. Cabrera-Rubio, R., Collado, M.C., Laitinen, K., Salminen, S., Isolauri, E., and Mira, A. (2012). The human milk microbiome changes over lactation and is shaped by maternal weight and mode of delivery. Am. J. Clin. Nutr. *96*, 544–551. https://doi.org/10.3945/ajcn.112.037382.
19. Young, G., Berrington, J.E., Cummings, S., Dorling, J., Ewer, A.K., Frau, A., Lett, L., Probert, C., Juszczak, E., Kirby, J., et al. (2023). Mechanisms affecting the gut of preterm infants in enteral feeding trials: a nested cohort within a randomised controlled trial of lactoferrin. Arch. Dis. Child. Fetal Neonatal Ed. *108*, 272–279. https://doi.org/10.1136/archdischild-2022-324477.
20. Embleton, N.D., Berrington, J.E., McGuire, W., Stewart, C.J., and Cummings, S.P. (2013). Lactoferrin: Antimicrobial activity and therapeutic potential. Semin. Fetal Neonatal Med. *18*, 143–149.
21. Stewart, C.J., Ajami, N.J., O'Brien, J.L., Hutchinson, D.S., Smith, D.P., Wong, M.C., Ross, M.C., Lloyd, R.E., Doddapaneni, H., Metcalf, G.A., et al. (2018). Temporal development of the gut microbiome in early childhood from the TEDDY study. Nature *562*, 583–588. https://doi.org/10.1038/s41586-018-0617-x.
22. Pannaraj, P.S., Ly, M., Cerini, C., Saavedra, M., Aldrovandi, G.M., Saboory, A.A., Johnson, K.M., and Pride, D.T. (2018). Shared and Distinct Features of Human Milk and Infant Stool Viromes. Front. Microbiol. *9*, 1162. https://doi.org/10.3389/fmicb.2018.01162.
23. Breitbart, M., Haynes, M., Kelley, S., Angly, F., Edwards, R.A., Felts, B., Mahaffy, J.M., Mueller, J., Nulton, J., Rayhawk, S., et al. (2008). Viral diversity and dynamics in an infant gut. Res. Microbiol. *159*, 367–373. https://doi.org/10.1016/j.resmic.2008.04.006.
24. Duranti, S., Lugli, G.A., Mancabelli, L., Armanini, F., Turroni, F., James, K., Ferretti, P., Gorfer, V., Ferrario, C., Milani, C., et al. (2017). Maternal inheritance of bifidobacterial communities and bifidophages in infants through vertical transmission. Microbiome *5*, 66. https://doi.org/10.1186/s40168-017-0282-6.
25. Liang, G., Zhao, C., Zhang, H., Mattei, L., Sherrill-Mix, S., Bittinger, K., Kessler, L.R., Wu, G.D., Baldassano, R.N., DeRusso, P., et al. (2020). The stepwise assembly of the neonatal virome is modulated by breastfeeding. Nature *581*, 470–474. https://doi.org/10.1038/s41586-020-2192-1.
26. Gasparrini, A.J., Wang, B., Sun, X., Kennedy, E.A., Hernandez-Leyva, A., Ndao, I.M., Tarr, P.I., Warner, B.B., and Dantas, G. (2019). Persistent metagenomic signatures of early-life hospitalization and antibiotic treatment in the infant gut microbiota and resistome. Nat. Microbiol. *4*, 2285–2297. https://doi.org/10.1038/s41564-019-0550-2.
27. Gregory, K.E., Samuel, B.S., Houghteling, P., Shan, G., Ausubel, F.M., Sadreyev, R.I., and Walker, W.A. (2016). Influence of maternal breast milk ingestion on acquisition of the intestinal microbiome in preterm infants. Microbiome *4*, 68. https://doi.org/10.1186/s40168-016-0214-x.
28. Jensen, R.G. (1999). Lipids in human milk. Lipids *34*, 1243–1271. https://doi.org/10.1007/s11745-999-0477-2.
29. Delplanque, B., and Baudry, C. (2015). Dairy fat in infant formula. Eur. J. Lipid Sci. Technol. *117*, 1487–1488. https://doi.org/10.1002/ejlt.201500352.
30. Mthembu, Y., Lotz, Z., Tyler, M., de Beer, C., Rodrigues, J., Schoeman, L., and Mall, A.S. (2014). Purified Human Breast Milk MUC1 and MUC4 Inhibit Human Immunodeficiency Virus. Neonatology *105*, 211–217. https://doi.org/10.1159/000357201.
31. Pricope-Ciolacu, L., Nicolau, A.I., Wagner, M., and Rychli, K. (2013). The effect of milk components and storage conditions on the virulence of Listeria monocytogenes as determined by a Caco-2 cell assay. Int. J. Food Microbiol. *166*, 59–64. https://doi.org/10.1016/j.ijfoodmicro.2013.05.027.
32. Liu, B., and Newburg, D.S. (2013). Human Milk Glycoproteins Protect Infants Against Human Pathogens. Breastfeed. Med. *8*, 354–362. https://doi.org/10.1089/bfm.2013.0016.
33. Young, G.R., Yew, W.C., Nelson, A., Bridge, S.H., Berrington, J.E., Embleton, N.D., and Smith, D.L. (2022). Optimisation and Application of a Novel Method to Identify Bacteriophages in Maternal Milk and Infant Stool Identifies Host-Phage Communities Within Preterm Infant Gut. Front. Pediatr. *10*, 856520.
34. Kovács, A., Funke, S., Marosvölgyi, T., Burus, I., and Decsi, T. (2005). Fatty Acids in Early Human Milk after Preterm and Full-Term Delivery. J. Pediatr. Gastroenterol. Nutr. *41*, 454–459.
35. Al-Tamer, Y.Y., and Mahmood, A.A. (2004). Fatty-acid composition of the colostrum and serum of fullterm and preterm delivering Iraqi mothers. Eur. J. Clin. Nutr. *58*, 1119–1124. https://doi.org/10.1038/sj.ejcn.1601939.
36. Underwood, M.A. (2013). Human Milk for the Premature Infant. Pediatr. Clin. North Am. *60*, 189–207. https://doi.org/10.1016/j.pcl.2012.09.008.
37. Walters, W.A., Granados, A.C., Ley, C., Federman, S., Stryke, D., Santos, Y., Haggerty, T., Sotomayor-Gonzalez, A., Servellita, V., Ley, R.E., et al. (2023). Longitudinal comparison of the developing gut virome in infants and their mothers. Cell Host Microbe *31*, 187–198.e3. https://doi.org/10.1016/j.chom.2023.01.003.
38. Gregory, A.C., Zablocki, O., Zayed, A.A., Howell, A., Bolduc, B., and Sullivan, M.B. (2020). The Gut Virome Database Reveals Age-Dependent Patterns of Virome Diversity in the Human Gut. Cell Host Microbe *28*, 724–740.e8. https://doi.org/10.1016/j.chom.2020.08.003.
39. Lim, E.S., Zhou, Y., Zhao, G., Bauer, I.K., Droit, L., Ndao, I.M., Warner, B.B., Tarr, P.I., Wang, D., and Holtz, L.R. (2015). Early life dynamics of the human gut virome and bacterial microbiome in infants. Nat. Med. *21*, 1228–1234. https://doi.org/10.1038/nm.3950.
40. Granger, C.L., Lamb, C.A., Embleton, N.D., Beck, L.C., Masi, A.C., Palmer, J.M., Stewart, C.J., and Berrington, J.E. (2022). Secretory immunoglobulin A in preterm infants: determination of normal values in breast milk and stool. Pediatr. Res. *92*, 979–986. https://doi.org/10.1038/s41390-021-01930-8.
41. Cho, G., Lee, E., and Kim, J. (2021). Structural insights into phosphatidylethanolamine formation in bacterial membrane biogenesis. Sci. Rep. *11*, 5785. https://doi.org/10.1038/s41598-021-85195-5.
42. van der Veen, J.N., Kennelly, J.P., Wan, S., Vance, J.E., Vance, D.E., and Jacobs, R.L. (2017). The critical role of phosphatidylcholine and phosphatidylethanolamine metabolism in health and disease. Biochim. Biophys. Acta. Biomembr. *1859*, 1558–1572. https://doi.org/10.1016/j.bbamem.2017.04.006.
43. Alexandre-Gouabau, M.-C., Moyon, T., Cariou, V., Antignac, J.-P., Qannari, E.M., Croyal, M., Soumah, M., Guitton, Y., David-Sochard, A., Billard, H., et al. (2018). Breast Milk Lipidome Is Associated with Early Growth Trajectory in Preterm Infants. Nutrients *10*, 164. https://doi.org/10.3390/nu10020164.
44. Ingvordsen Lindahl, I.E., Artegoitia, V.M., Downey, E., O'Mahony, J.A., O'Shea, C.-A., Ryan, C.A., Kelly, A.L., Bertram, H.C., and Sundekilde,

U.K. (2019). Quantification of Human Milk Phospholipids: the Effect of Gestational and Lactational Age on Phospholipid Composition. Nutrients *11*, 222. https://doi.org/10.3390/nu11020222.

45. Ramiro-Cortijo, D., Singh, P., Liu, Y., Medina-Morales, E., Yakah, W., Freedman, S.D., and Martin, C.R. (2020). Breast Milk Lipids and Fatty Acids in Regulating Neonatal Intestinal Development and Protecting against Intestinal Injury. Nutrients *12*, 534. https://doi.org/10.3390/nu12020534.

46. Dinleyici, M., Pérez-Brocal, V., Arslanoglu, S., Aydemir, O., Sevuk Ozumut, S., Tekin, N., Vandenplas, Y., Moya, A., and Dinleyici, E.C. (2021). Human Milk Virome Analysis: Changing Pattern Regarding Mode of Delivery, Birth Weight, and Lactational Stage. Nutr *13*, 1779. https://doi.org/10.3390/nu13061779.

47. Rabia, M., B., R.J., I., W.L., A., H.L., Ruth, N., Dorothy, M.-N., Elizabeth, M.-O., R., B.E., Soren, G., Ednah, O., et al. (2021). Breast Milk Virome and Bacterial Microbiome Resilience in Kenyan Women Living with HIV. mSystems *6*, e01079-20. https://doi.org/10.1128/mSystems.01079-20.

48. Byrd, A.L., Belkaid, Y., and Segre, J.A. (2018). The human skin microbiome. Nat. Rev. Microbiol. *16*, 143–155. https://doi.org/10.1038/nrmicro.2017.157.

49. Grice, E.A., and Segre, J.A. (2011). The skin microbiome. Nat. Rev. Microbiol. *9*, 244–253. https://doi.org/10.1038/nrmicro2537.

50. Oliveira, H., Sampaio, M., Melo, L.D.R., Dias, O., Pope, W.H., Hatfull, G.F., and Azeredo, J. (2019). Staphylococci phages display vast genomic diversity and evolutionary relationships. BMC Genom. *20*, 357. https://doi.org/10.1186/s12864-019-5647-8.

51. Conlan, S., Mijares, L.A., NISC Comparative Sequencing Program; Becker, J., Blakesley, R.W., Bouffard, G.G., Brooks, S., Coleman, H., Gupta, J., Gurson, N., et al. (2012). Staphylococcus epidermidis pan-genome sequence analysis reveals diversity of skin commensal and hospital infection-associated isolates. Genome Biol. *13*, R64. https://doi.org/10.1186/gb-2012-13-7-r64.

52. Brown, M.M., and Horswill, A.R. (2020). Staphylococcus epidermidis—Skin friend or foe? PLoS Pathog. *16*, e1009026.

53. Genzel-Boroviczény, O., Wahle, J., and Koletzko, B. (1997). Fatty acid composition of human milk during the 1st month after term and preterm delivery. Eur. J. Pediatr. *156*, 142–147. https://doi.org/10.1007/s004310050573.

54. Skogerson, K., Wohlgemuth, G., Barupal, D.K., and Fiehn, O. (2011). The volatile compound BinBase mass spectral database. BMC Bioinf. *12*, 321. https://doi.org/10.1186/1471-2105-12-321.

55. Aggett, P.J., Haschke, F., Heine, W., Hernell, O., Koletzko, B., Launiala, K., Rey, J., Rubino, A., Schöch, G., Senterre, J., et al. (1991). Comment on the Content and Composition of Lipids in Infant Formulas. Acta Paediatr. *80*, 887–896. https://doi.org/10.1111/j.1651-2227.1991.tb11969.x.

56. Luukkainen, P., Salo, M.K., and Nikkari, T. (1994). Changes in the Fatty Acid Composition of Preterm and Term Human Milk from 1 Week to 6 Months of Lactation. J. Pediatr. Gastroenterol. Nutr. *18*, 355–360.

57. Apffel, A., Zhao, L., and Sartain, M.J. (2021). A Novel Solid Phase Extraction Sample Preparation Method for Lipidomic Analysis of Human Plasma Using Liquid Chromatography/Mass Spectrometry. Metabolites *11*, 294. https://doi.org/10.3390/metabo11050294.

58. Samara, J., Moossavi, S., Alshaikh, B., Ortega, V.A., Pettersen, V.K., Ferdous, T., Hoops, S.L., Soraisham, A., Vayalumkal, J., Dersch-Mills, D., et al. (2022). Supplementation with a probiotic mixture accelerates gut microbiome maturation and reduces intestinal inflammation in extremely preterm infants. Cell Host Microbe *30*, 696–711.e5. https://doi.org/10.1016/j.chom.2022.04.005.

59. Ennis, D., Moriel, N., and Yassour, M. (2022). Preemies going pro: How probiotic treatment matures the microbiome of extreme premature infants. Cell Host Microbe *30*, 599–600. https://doi.org/10.1016/j.chom.2022.04.011.

60. Alcon-Giner, C., Dalby, M.J., Caim, S., Ketskemety, J., Shaw, A., Sim, K., Lawson, M.A.E., Kiu, R., Leclaire, C., Chalklen, L., et al. (2020). Microbiota Supplementation with Bifidobacterium and Lactobacillus Modifies the Preterm Infant Gut Microbiota and Metabolome: An Observational Study. Cell Rep. Med. *1*, 100077. https://doi.org/10.1016/j.xcrm.2020.100077.

61. Lugli, G.A., Fontana, F., Tarracchini, C., Mancabelli, L., Milani, C., Turroni, F., and Ventura, M. (2022). Exploring the biodiversity of Bifidobacterium asteroides among honey bee microbiomes. Environ. Microbiol. *24*, 5666–5679. https://doi.org/10.1111/1462-2920.16223.

62. Enav, H., Bäckhed, F., and Ley, R.E. (2022). The developing infant gut microbiome: A strain-level view. Cell Host Microbe *30*, 627–638. https://doi.org/10.1016/j.chom.2022.04.009.

63. Koenig, J.E., Spor, A., Scalfone, N., Fricker, A.D., Stombaugh, J., Knight, R., Angenent, L.T., and Ley, R.E. (2011). Succession of microbial consortia in the developing infant gut microbiome. Proc. Natl. Acad. Sci. *108*, 4578–4585. https://doi.org/10.1073/pnas.1000081107.

64. Brooks, B., Firek, B.A., Miller, C.S., Sharon, I., Thomas, B.C., Baker, R., Morowitz, M.J., and Banfield, J.F. (2014). Microbes in the neonatal intensive care unit resemble those found in the gut of premature infants. Microbiome *2*, 1–16. https://doi.org/10.1186/2049-2618-2-1.

65. Brooks, B., Olm, M.R., Firek, B.A., Baker, R., Thomas, B.C., Morowitz, M.J., and Banfield, J.F. (2017). Strain-resolved analysis of hospital rooms and infants reveals overlap between the human and room microbiome. Nat. Commun. *8*, 1814. https://doi.org/10.1038/s41467-017-02018-w.

66. Stewart, C.J., Embleton, N.D., Marrs, E.C.L., Smith, D.P., Fofanova, T., Nelson, A., Skeath, T., Perry, J.D., Petrosino, J.F., Berrington, J.E., and Cummings, S.P. (2017). Longitudinal development of the gut microbiome and metabolome in preterm neonates with late onset sepsis and healthy controls. Microbiome *5*, 75. https://doi.org/10.1186/s40168-017-0295-1.

67. Beck, L.C., Masi, A.C., Young, G.R., Vatanen, T., Lamb, C.A., Smith, R., Coxhead, J., Butler, A., Marsland, B.J., Embleton, N.D., et al. (2022). Strain-specific impacts of probiotics are a significant driver of gut microbiome development in very preterm infants. Nat. Microbiol. *7*, 1525–1535. https://doi.org/10.1038/s41564-022-01213-w.

68. Brown, E.M., Clardy, J., and Xavier, R.J. (2023). Gut microbiome lipid metabolism and its impact on host physiology. Cell Host Microbe *31*, 173–186. https://doi.org/10.1016/j.chom.2023.01.009.

69. Kumar, H., du Toit, E., Kulkarni, A., Aakko, J., Linderborg, K.M., Zhang, Y., Nicol, M.P., Isolauri, E., Yang, B., Collado, M.C., and Salminen, S. (2016). Distinct Patterns in Human Milk Microbiota and Fatty Acid Profiles Across Specific Geographic Locations. Front. Microbiol. *7*, 1619.

70. Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics *34*, i884–i890. https://doi.org/10.1093/bioinformatics/bty560.

71. McIver, L.J., Abu-Ali, G., Franzosa, E.A., Schwager, R., Morgan, X.C., Waldron, L., Segata, N., and Huttenhower, C. (2018). bioBakery: a meta'omic analysis environment. Bioinformatics *34*, 1235–1237. https://doi.org/10.1093/bioinformatics/btx754.

72. Huson, D.H., Auch, A.F., Qi, J., and Schuster, S.C. (2007). MEGAN analysis of metagenomic data. Genome Res. *17*, 377–386. https://doi.org/10.1101/gr.5969107.

73. Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics *31*, 1674–1676. https://doi.org/10.1093/bioinformatics/btv033.

74. Li, H. (2013). Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM. https://arxiv.org/abs/1303.3997.

75. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

76. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

77. Guo, J., Bolduc, B., Zayed, A.A., Varsani, A., Dominguez-Huerta, G., Delmont, T.O., Pratama, A.A., Gazitúa, M.C., Vik, D., Sullivan, M.B., and Roux, S. (2021). VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. Microbiome *9*, 37. https://doi.org/10.1186/s40168-020-00990-y.

78. Nayfach, S., Camargo, A.P., Schulz, F., Eloe-Fadrosh, E., Roux, S., and Kyrpides, N.C. (2021). CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nat. Biotechnol. *39*, 578–585. https://doi.org/10.1038/s41587-020-00774-7.

79. Wood, D.E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. Genome Biol. *20*, 257. https://doi.org/10.1186/s13059-019-1891-0.

80. Lu, J., Breitwieser, F.P., Thielen, P., and Salzberg, S.L. (2017). Bracken: Estimating Species Abundance in Metagenomics Data (PeerJ).

81. Davis, N.M., Proctor, D.M., Holmes, S.P., Relman, D.A., and Callahan, B.J. (2018). Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome *6*, 226. https://doi.org/10.1186/s40168-018-0605-2.

82. Blanco-Miguez, A., Beghini, F., Cumbo, F., McIver, L.J., Thompson, K.N., Zolfo, M., Manghi, P., Dubois, L., Huang, K.D., Thomas, A.M., et al. (2022). Extending and Improving Metagenomic Taxonomic Profiling with Uncharacterized Species with MetaPhlAn 4. Preprint at bioRxiv. https://doi.org/10.1101/2022.08.22.504593.

83. Oksanen, J., Guillaume Blanchet, F., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Henry, M., Stevens, H., et al. (2015). Vegan: Community Ecology Package.

84. Brooks, M.E., Kristensen, K., van Benthem, K.J., Magnusson, A., Berg, C.W., and Bolker, B.M. (2017). glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. R J *9*, 379–400.

85. Searle, S.R., Speed, F.M., and Milliken, G.A. (1980). Population Marginal Means in the Linear Model: An Alternative to Least Squares Means. Am. Stat. *34*, 216–221. https://doi.org/10.1080/00031305.1980.10483031.

86. Mallick, H., Rahnavard, A., McIver, L.J., Ma, S., Zhang, Y., Nguyen, L.H., Tickle, T.L., Weingart, G., Ren, B., Schwager, E.H., et al. (2021). Multivariable association discovery in population-scale meta-omics studies. PLoS Comput. Biol. *17*, e1009442.

87. Rohart, F., Gautier, B., Singh, A., and Lê Cao, K.A. (2017). mixOmics: An R package for 'omics feature selection and multiple data integration. PLoS Comput. Biol. *13*, e1005752. https://doi.org/10.1371/journal.pcbi.1005752.

88. Kim, D., Hofstaedter, C.E., Zhao, C., Mattei, L., Tanes, C., Clarke, E., Lauder, A., Sherrill-Mix, S., Chehoud, C., Kelsen, J., et al. (2017). Optimizing methods and dodging pitfalls in microbiome research. Microbiome *5*, 52. https://doi.org/10.1186/s40168-017-0267-5.

89. Shkoporov, A.N., and Hill, C. (2019). Bacteriophages of the Human Gut: The Known Unknown of the Microbiome. Cell Host Microbe *25*, 195–209. https://doi.org/10.1016/j.chom.2019.01.017.

90. Brewster, J.D., and Paul, M. (2016). Short communication: Improved method for centrifugal recovery of bacteria from raw milk applied to sensitive real-time quantitative PCR detection of Salmonella spp.1. J. Dairy Sci. *99*, 3375–3379. https://doi.org/10.3168/jds.2015-9655.

91. Luo, J., Wang, Z., Li, Y., Chen, C., Ren, F., and Guo, H. (2019). The simulated in vitro infant gastrointestinal digestion of droplets covered with milk fat globule membrane polar lipids concentrate. J. Dairy Sci. *102*, 2879–2889. https://doi.org/10.3168/jds.2018-15044.

92. Bharti, R., and Grimm, D.G. (2021). Current challenges and best-practice protocols for microbiome analysis. Brief. Bioinform. *22*, 178–193. https://doi.org/10.1093/bib/bbz155.

93. Bushnell, B. (2014). BBTools. https://sourceforge.net/projects/bbmap/.

94. Vorkas, P.A., Isaac, G., Anwar, M.A., Davies, A.H., Want, E.J., Nicholson, J.K., and Holmes, E. (2015). Untargeted UPLC-MS Profiling Pipeline to Expand Tissue Metabolome Coverage: Application to Cardiovascular Disease. Anal. Chem. *87*, 4184–4193. https://doi.org/10.1021/ac503775m.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
Escherichia coli	DSMZ	Catalog: 4230
Phage T4	DSMZ	Catalog: 4505
Phage Lambda	DSMZ	Catalog: 4499
Biological samples		
Preterm human milk	RVI NICU	N/A
Chemicals, peptides, and recombinant proteins		
Digestive enzymes - pepsin	Sigma-Aldrich	Catalog: P7000
Digestive enzymes - lipase	Sigma-Aldrich	Catalog: 62305
Linoleic acid	Sigma-Aldrich	Catalog: 62230-5ML-F
Critical commercial assays		
Norgen Phage DNA Isolation Kit	Norgen Biotek Corp.	Catalog: 46850
Beckman Coulter™ Agencourt AMPure XP beads	Thermo Fisher Scientific	Catalog: 10453438
Qubit™ dsDNA HS Assay Kit	Thermo Fisher Scientific	Catalog: Q32851
Nextera XT DNA Library Preparation Kit	Illumina	Catalog: FC-131-1096
NextSeq 500/550 Kit v2.5 (300 cycles)	Illumina	Catalog: 20024908
Deposited data		
Raw sequencing data	This paper	ENA: PRJEB58774 https://www.ebi.ac.uk/ena/browser/view/PRJEB58774
Viral reference database	NCBI	https://ftp.ncbi.nlm.nih.gov/refseq/release/viral/ accessed 5 May 2022
Software and algorithms		
BBTools	Bushnell, 2017	https://jgi.doe.gov/data-and-tools/software-tools/bbtools/
Fastp	Chen et al.[70]	https://github.com/OpenGene/fastp
bioBakery	McIver et al.[71]	http://huttenhower.sph.harvard.edu/biobakery
MEGAN Community Edition	Huson et al.[72]	https://github.com/danielhuson/megan-ce
MEGAHIT	Li et al.[73]	https://github.com/voutcn/megahit
BWA-MEM	Li[74]	https://arxiv.org/abs/1303.3997
SAMtools	Li et al.[75]	http://samtools.sourceforge.net
BEDTools	Quinlan and Hall[76]	http://code.google.com/p/bedtools
VirSorter2	Guo et al.[77]	https://bitbucket.org/MAVERICLab/virsorter2
CheckV	Nayfach et al.[78]	https://bitbucket.org/berkeleylab/CheckV
Kraken 2	Wood et al.[79]	https://doi.org/10.5281/zenodo.3520272
Bracken	Lu et al.[80]	http://ccb.jhu.edu/software/bracken/
Decontam	Davis et al.[81]	https://github.com/benjjneb/decontam
MetaPhlAn 4	Blanco-Miguez et al.[82]	https://huttenhower.sph.harvard.edu/metaphlan/
Vegan	Oksanen et al.[83]	https://github.com/vegandevs/vegan/
glmmTMB	Brooks et al.[84]	https://cran.r-project.org/web/packages/glmmTMB/index.html
emmeans	Searle et al.[85]	https://cran.r-project.org/web/packages/emmeans/index.html
MaAsLin	Mallick et al.[86]	http://huttenhower.sph.harvard.edu/maaslin2
mixOmics	Rohart et al.[87]	www.mixOmics.org
Opticlust	Westcott et al., 2017	https://github.com/SchlossLab/Westcott_OptiClust_mSphere_2017

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to Darren Smith (darren.smith@northumbria.ac.uk).

Materials availability

This study did not generate unique reagents.

Data and code availability

- All original code has been deposited at ENA (PRJEB58774) and is publicly available as of the date of publication. Lipidomic data are available upon request. DOIs are listed in the key resources table.
- This publication did not generate any original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Participant cohort

This study involved a total of 113 participants who delivered very preterm infants (weeks of gestation $\leq$32) and were enrolled in the Royal Victoria Infirmary Neonatal Intensive Care Unit (NICU) Supporting Enhanced Research in Vulnerable Infants (SERVIS) project at Newcastle upon Tyne, UK. Demographic information including delivery mode, gestational age (weeks), infant weights and lactational ages of HM are listed in Table S1. The expressed very preterm human milk (HM) was collected after written consent, using a developed standard operating procedure.[40] The study was performed with ethical permission granted by the National Research Ethics Service (NRES) Committee North East – Newcastle and North Tyneside 2 (10/H0908/39).

Lactational ages

Lactational ages of preterm HM were defined by the number of weeks of post-delivery that salvaged milk was fed to neonates.[37] Colostrum was defined as milk samples expressed and fed to neonates during the first week of life. Mature milk was defined as milk samples expressed later than one month following delivery. One HM sample per patient, per lactational age was included in analysis (Figure S5).

METHOD DETAILS

Sample salvaging, storage and processing

Preterm HM was salvaged from residual in feeding tubes, labeled with a study ID number and date, and stored at -20° for as short a time as possible before transfer (usually within one week) to -80°C storage until the day of laboratory processing. In the early days of expressing, milk is usually given to the baby on the day it is expressed. In later days milk may be stored frozen before administration as close as possible to expression day. Day of receipt to the baby is the day used as a proxy of lactational age in this study. All laboratory work was conducted in a Class II biosafety cabinet using nuclease-free or sterilised consumables and equipment to reduce the impacts of contamination.[88] HM was defrosted in the fridge and further sub-sampled to two tubes each with a volume of 200 μL for shotgun metagenomics of free and inducible phages and one tube with a volume of 100 μL for lipidomics to prevent multiple freeze-thaw cycles.

Phage isolation, DNA extraction and shotgun metagenomics

Both lytic and lysogenic phages were isolated from each sample using our previously published dsDNA phage isolation protocol that was optimised for low-volume milk samples.[33] The protocol was designed to specifically characterise observable dsDNA viral communities[89] by omitting steps such as cesium chloride density gradient ultracentrifugation which fails to recover enveloped viruses and those with atypical buoyant densities or multiple displacement amplification with φ29 DNA polymerase which preferentially amplifies small circular ssDNA viruses. Here, an intense agitation step was performed on the samples to disrupt the binding of microbes on the membrane of MFG[90] prior to centrifugal separation of milk fat and skim milk. Later, a lipid hydrolysis step was performed on the separated milk fat using digestive enzymes similar to those in infant guts.[91] Both steps were aimed to maximise recovery of phages and microbes from the milk fat. Norfloxacin was then used to induce lysogenic phages from the pelleted microbial cells. Isolated lytic and lysogenic phage suspensions from both milk lipid and skim milk fractions were pooled for DNA extraction in duplicate, using Norgen Phage DNA Isolation Kit (Norgen Biotek Corp., Canada) and were then pooled into a single tube for DNA concentrating and purifying using Agencourt AMPure XP beads (Beckman Coulter Genomics, USA). The concentration of phage DNA was determined using Qubit 2.0 Fluorometer (Thermo Fisher Scientific, USA) and dsDNA libraries were then prepared using Nextera XT DNA Library Prep Kit. Isolation kit and library preparation controls were included in sequencing runs. Sequencing was performed by NUOMICS (Northumbria University, Newcastle) on the Illumina NextSeq (Illumina Inc., USA), using a V2.5 300 cycle chemistry.

Bioinformatics processing

Samples yielded a median 5.62 x 10^6 sequence reads across all 118 samples with a viral enrichment score of 1.4 (median; Figure S6A). QC filtering reduced the total library size to 3.48 x 10^6. QC filtering invoked *FastP*[70] with a 5-nucleotide sliding window and a minimum Q score of 30. Any reads <100 nucleotides or mapping to human reads were also removed using *Kneaddata* within the *biobakery* environment.[71] To improve taxonomy classification of the phage communities in preterm HM, both reference-based and assembly-based analyses were performed on the filtered clean reads.[92] The former analysis generated viral composition profiles using merged and unmerged paired reads and unpaired reads, whilst the later generated taxonomy classification and functional gene prediction using assembled contigs.

Reference-based viral composition analysis

Cleaned paired reads were merged using *bbmerge.sh* in BBTools.[93] The resulting merged and unmerged paired reads as well as the unpaired reads were concatenated and aligned against NCBI viral reference database (accessed 5 May 2022; https://ftp.ncbi.nlm.nih.gov/refseq/release/viral/) using the *blastn* option in BLAST (evalue 10^{-5}, best-hit overhang 0.1, best-hit score edge 0.1). BLAST output reads were uploaded to MEGAN Community Edition[72] and taxonomy assignment was performed at the family, genus and species levels using a modified lowest common ancestor (LCA) algorithm (minimum score 50, E-value <10^{-5}, minimum percent identity 60, minimum support 2). Information of the viral hosts were inferred from the genus assignment level.

Assembly-based virome analysis

Contiguous sequences were assembled from QC filtered reads using the default kmer steps in Megahit,[73] yielding 4.78 x 10^5 total contigs. *Bwa mem*,[74] *samtools*,[75] *bamtools* and *bedtools*[76] were used to cull contigs with <2x coverage leaving 4.39 x 10^5 contigs (N50 = 1291; Figure S6B). Any viral contigs with $\geq$2x coverage were identified using *Virsorter2*.[77] Completeness of viral assembled genomes was quantified using CheckV.[78] Taxonomic classification of $\geq$2x coverage viral contigs was performed with Kraken2[79] by alignment to pre-built viral databases (accessed 4 January 2022; https://benlangmead.github.io/aws-indexes/k2). A total of 3,208 viral contigs were classifiable by alignment to the NCBI viral db with a median 1,315 reads mapped to reference sequences per sample (IQR 443–5,093). This illustrated sufficient sampling depth to characterise HM viromes as demonstrated by asymptotic rarefaction curves in Figure S6C. Per-sample abundance of each taxonomically classified viral contig was calculated using Bracken.[80] Potential contaminant viral taxa were identified and removed from read and assembly-based feature tables by comparing sample and negative control communities using the prevalence method with a 50% confidence cut-off in Decontam.[81] Following removal of potentially contaminant taxa, significantly more viral contigs were observed in HM samples than negative controls (KW: p = 0.0006; Figure S6D) and sample compositions were significantly different from controls (ANOSIM: p = 0.01, R^2 = 0.42; Figure S6E). Many viral contigs remained unclassifiable after alignment to the viral reference database (median 64%, IQR 36–82%; Figure S6F). Viral tail morphology and host-range were inferred by nucleotide similarity of assembled viral contigs to pre-annotated,NCBI virus database.

Bacterial community analyses from metagenome data

Taxonomic annotation of bacterial sequences was performed using MetaPhlAn v.4.0.3[82] with default settings.

Lipid extraction, lipidomic data alignment and peak table generation

Lipidomic analysis was performed by Northumbria University Omics Research Facility. Lipids were extracted from 100 μL HM using a method adopted from Vorkas et al.[94] Briefly, 1mL of Dichloromethane/Methanol (3:1 v/v) were added to the samples before vortexing for 30 s and sonicating in an ice water bath for 15 min. Samples were then centrifuged at 15K rpm at 10°C for 15 min. Resulting supernatants were collected and dried in a Speed Vac at 45°C for 2 h or until the extraction buffer fully evaporated. Dried supernatants were then stored at −20°C until further analysis.

Dried supernatants were thawed and reconstituted in 100 μL of 1:1:2 water:ACN:IPA. A volume of 80 μL was transferred to an autosampler vial and profiled on a Vanquish ultra-high pressure chromatography system with a Waters Acquity UPLC CSH C18 column (Waters Corporation, US) connected to an IDX high-res mass spectrometer (Thermo Fisher Scientific, US). Data were acquired in both positive and negative ion modes using the AcquireX acquisition DDA workflow (Thermo Fisher Scientific, US). Lipid classifications were achieved using the Thermo Scientific Lipid search version 4 workflow (Thermo Fisher Scientific, US) using the high energy collision database (HCD), a retention time of 0.1 min, parent ion mass tolerance of 5 ppm, and product ion of 10 ppm. Lipid IDs were matched using the lipid search *in silico* MS2 level library.

A volume of 20 μL from each sample was pooled to create QC for AcquireX data-dependent acquisition template generation. Extraction blanks were included as part of the sample analysis and any features identified in these blanks were excluded from chemical analysis. Further details of sample preparation as well as in depth chromatography and mass spectrometry parameters are included in Supplementary Methods 1.

In-vitro assessment of lipid on phage infectivity

Siphoviruses and Podoviruses (but not Myoviruses) were significantly positively correlated with LCFA (Figures 4B and 4C). We further performed an *in vitro* experiment to examine if linoleic acid affects the infectivity of two viable and morphologically different *Escherichia coli* phages T4 (Myovirus) and λ (Siphovirus). Linoleic acid is a VLC polyunsaturated FA (18C) reported in preterm HM with a concentration range of 0.001–0.006 g/mL.[45] We incubated 100 μL of known titers of T4 (4.0 X 10^8) and λ (5.0 X 10^{10}) phage suspensions with 900 μL serial dilutions of linoleic acid (concentration between 0.0009 and 0.09 g/mL) at 37°C for 1 h. To control for dilution

effects of adding linoleic acids at different concentrations to phage suspensions, equal volumes of 1 x PBS (900 μL) were added to phage suspensions and incubated under the same conditions. Following incubation, phage, lipid suspensions were centrifuged at 5K rcf and at 4°C for 10 min to fractionise and separate linoleic acid. Plaque assays using *Escherichia coli* as the infectivity host (n = 3 per linoleic acid concentration) were performed to count the plaque-forming units (PFU) of both T4 and λ phages upon removal of linoleic acid.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis

All statistical analyses were performed in RStudio (RStudio Team, 2020). Fisher's exact test and Kruskal-Wallis rank-sum test were used to compare means of categorical and continuous clinical variables, respectively, between lactational ages.

Phage richness and diversity analysis

To capture the full richness and diversity of phages in preterm HM, read-based taxonomic compositions were used to calculate alpha and beta diversity metrics[92] using the vegan R package for community ecology.[83] Within-sample alpha diversity was calculated as rarefied taxonomic richness and Shannon diversity. Generalised mixed linear models (glmms) were built using glmmTMB[84] to assess the associations between alpha diversity metrics, gestational age, delivery mode and lactational age (all included as fixed effects). Least-square means (LMS) of alpha diversity metrics were compared using emmeans[85] when glmms identified a significant impact of the lactational age. Multiple pairwise comparisons were corrected using Tukey's method. In addition, weighted Bray-Curtis dissimilarity was calculated to determine beta-diversity between samples. Adonis PERMANOVA was used to assess the association between sample composition, gestational age, delivery mode and lactational age. The impact of each variable was assessed marginally in a model containing all other variables.

Core and differential phage community analysis

A greater confidence in taxonomic classification is achievable through assembly-based methods.[92] We therefore used ≥2x coverage viral contigs to identify core viromes and differential viral features across each lactational age. Linear regression models were built to determine the relationships between distribution and abundance. Following Bonferroni correction for multiple hypothesis testing, significant relationships ($p < 0.05$, $R^2 > 0.4$) were deemed indicative of true viral core communities. Any viral features observed in >50% of samples were considered core. MaAsLin2[86] was used to identify significantly differential features of the virome between different lactational ages (default threshold of $q = 0.25$). Lactational and gestational ages were included as continuous fixed effects with mature milk, and vaginal delivery acting as reference categories, respectively. Delivery mode was included as a categorical fixed effect.

Lipid-phage co-occurrence networks

Feature abundances (virome) and intensities (lipids) were cumulative-sum scaled before combining lipid feature data with family, species and viral-host level taxonomic data. MixOmics[87] was used to build sPLS regressions linking mixed viral and lipidomic features with lactational ages. Correlations stronger than $R^2 = 0.4$ between lipid and phage feature abundance at each different lactational age were used to build co-occurrence networks to visualise relationships between HM lipids and phage.