

Washington University School of Medicine

Digital Commons@Becker

---

Open Access Publications

---

12-10-2019

## Discordant transmission of bacteria and viruses from mothers to babies at birth

Rabia Maqsood

Rachel Rodgers

Cynthia Rodriguez

Scott A Handley

I. Malick Ndao

*See next page for additional authors*

Follow this and additional works at: [https://digitalcommons.wustl.edu/open\\_access\\_pubs](https://digitalcommons.wustl.edu/open_access_pubs)

---

---

**Authors**

Rabia Maqsood, Rachel Rodgers, Cynthia Rodriguez, Scott A Handley, I. Malick Ndao, Phillip I Tarr, Barbara B Warner, Efrem S Lim, and Lori R Holtz


---

RESEARCH

Open Access

# Discordant transmission of bacteria and viruses from mothers to babies at birth



Rabia Maqsood<sup>1,2†</sup>, Rachel Rodgers<sup>3†</sup>, Cynthia Rodriguez<sup>3</sup>, Scott A. Handley<sup>4</sup>, I. Malick Ndao<sup>3</sup>, Phillip I. Tarr<sup>3,5</sup>, Barbara B. Warner<sup>3</sup>, Efrem S. Lim<sup>1,2\*</sup> and Lori R. Holtz<sup>3\*</sup> 

## Abstract

**Background:** The earliest microbial colonizers of the human gut can have life-long consequences for their hosts. Precisely how the neonatal gut bacterial microbiome and virome are initially populated is not well understood. To better understand how the maternal gut microbiome influences acquisition of the infant gut microbiome, we studied the early life bacterial microbiomes and viromes of 28 infant twin pairs and their mothers.

**Results:** Infant bacterial and viral communities more closely resemble those of their related co-twin than unrelated infants. We found that 63% of an infant's bacterial microbiome can be traced to their mother's gut microbiota. In contrast, only 15% of their viral communities are acquired from their mother. Delivery route did not determine how much of the bacterial microbiome or virome was shared from mother to infant. However, bacteria-bacteriophage interactions were altered by delivery route.

**Conclusions:** The maternal gut microbiome significantly influences infant gut microbiome acquisition. Vertical transmission of the bacterial microbiome is substantially higher compared to vertical transmission of the virome. However, the degree of similarity between the maternal and infant gut bacterial microbiome and virome did not vary by delivery route. The greater similarity of the bacterial microbiome and virome between twin pairs than unrelated twins may reflect a shared environmental exposure. Thus, differences of the inter-generation transmissibility at birth between the major kingdoms of microbes indicate that the foundation of these microbial communities are shaped by different rules.

**Keywords:** Virome, Microbiome, Transmission

## Background

The gut undergoes a profound ecological transition as the infant leaves the near sterile or sterile womb and becomes home to diverse microbial populations. This once in a lifetime event has lasting effects on the infant by shaping growth [1], conferring resistance to infection [2], calibrating inflammation [3], and programming immune function [4]. Remarkably, however, we know very little about how and when the neonatal gut bacterial microbiome and virome are acquired.

Histological evidence [5] and sequencing data [6, 7] suggest that the infant might encounter microbes in the

womb, thereby starting colonization in that venue. However, several recent publications find no convincing evidence of bacterial or viral nucleic acid in human amniotic fluid or placenta [8–11], calling into question the in utero colonization hypothesis. Several studies have demonstrated striking differences between the maternal and infant gut bacterial microbiome communities, but also provide evidence of shared bacterial species among related mothers and infants [12–15].

During the first days of life, portions of the infant gut microbiome can be traced to multiple maternal sources including the vaginal, skin, oral, and gut microbiomes. However, of these sources, colonization by bacteria that are found in the mothers' gut is the predominant driver of long-term persistence [16]. Indeed, seeding by vertically transmitted bacteria (such as *Bifidobacterium*, *Ruminococcus*, *Coprococcus* species) causes transcriptionally active colonization in the infant gut [13].

\* Correspondence: [Efrem.Lim@asu.edu](mailto:Efrem.Lim@asu.edu); [loriholtz@wustl.edu](mailto:loriholtz@wustl.edu)

<sup>†</sup>Rabia Maqsood and Rachel Rodgers contributed equally to this work.

<sup>1</sup>School of Life Sciences, Arizona State University, Tempe, AZ 85287, USA

<sup>3</sup>Department of Pediatrics, Washington University School of Medicine, St. Louis, MO 63110, USA

Full list of author information is available at the end of the article



Various factors such as birth route and food source may affect transmission to her infant. The effect of birth route on the similarity of newborn and maternal bacterial gut microbiome composition is conflicted [17, 18]. While diet is known to impact bacterial microbiome composition over time, interestingly the bacterial gut microbiomes of breastfed vaginally delivered newborns do not differ significantly from those receiving formula [17, 18].

Considerably less is known about early colonization of the infant gut with viruses and if vertical transmission plays a role [19]. Previously, we described a high diversity of bacteriophages in stools from infants during their first 96 h of life [20]. The source of these bacteriophages is unknown; our data suggest mother-to-infant transmission could play a significant role. For instance, vertical transmission of bifidobacteria bacteriophages can be traced to breastmilk [21]. Further, at 1 year of age, vaginally delivered infants have increased gut viral diversity than infants delivered by C-section [22], suggesting that other factors may also influence infant virome acquisition.

Here, we sought to understand how microbes initially colonize the human infant gut. To do so, we ask if earliest in life stools (obtained at 3–93 h of age) from 28 term twin pairs share their mothers' stool virome and bacterial microbiome. This cohort additionally offers a unique opportunity to determine the consistency of vertical transmission for co-twins. Such data will greatly increase our understanding of how microbes initially occupy the infant gut.

## Results

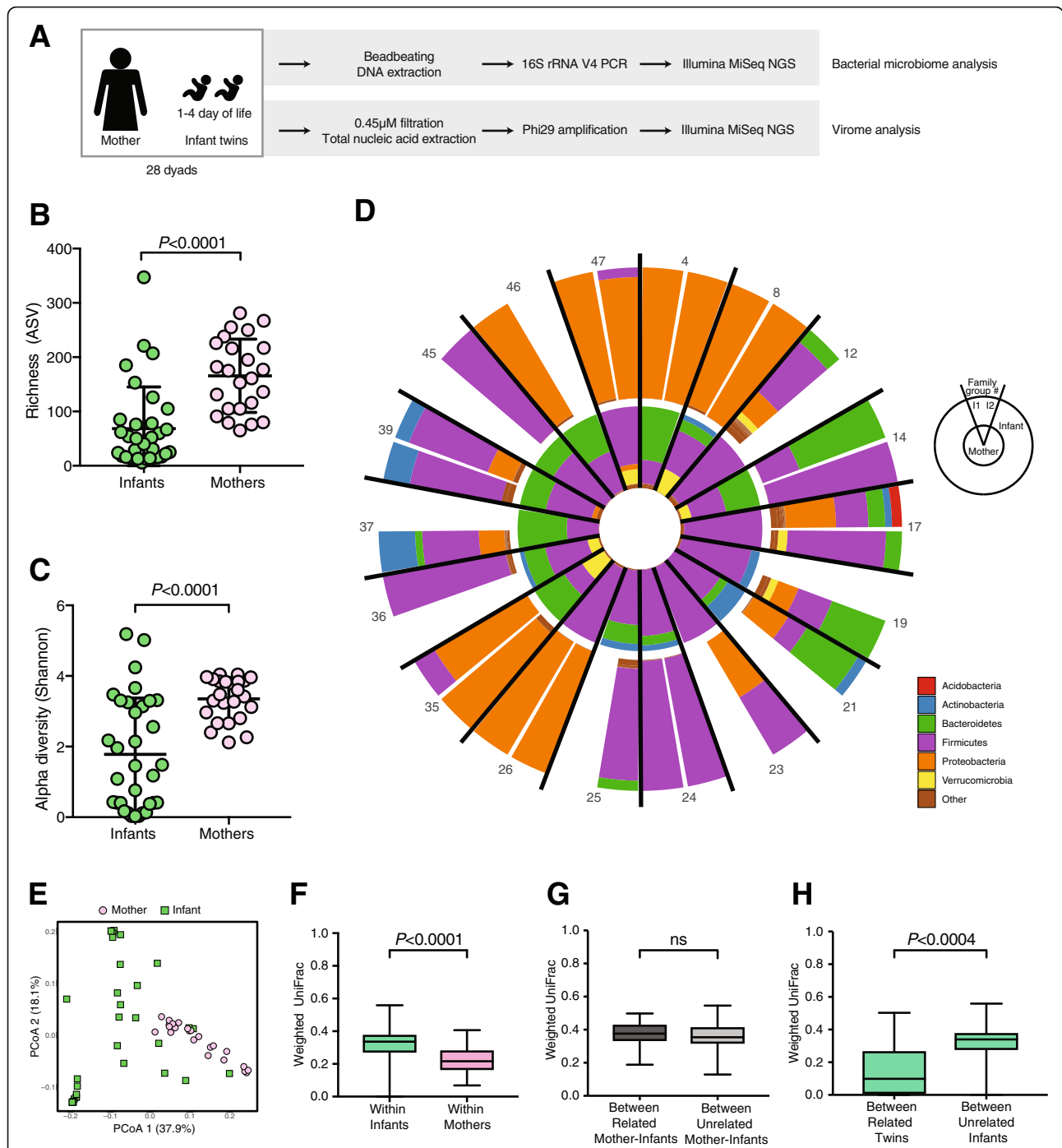
### Gut bacterial microbiome of infant twin pairs and their mothers

We performed 16S rRNA gene sequencing of stools from 56 infants (28 twin pairs, collected on or before the fourth day of life) and their 28 mothers (Fig. 1a). Of the 84 samples, five samples failed to amplify, four were excluded due to low (< 1000) read counts, and 18 generated sequences indistinguishable from buffer control samples (see the "Methods") and were also excluded from analysis (Additional file 6: Table S1). Sequences from the remaining 24 maternal and 33 infant samples correspond to 2615 different bacterial taxa. Both richness and Shannon diversity of infant bacterial gut communities were significantly less than in their mothers' stools (both  $p < 0.001$ ) (Fig. 1b and c). Proteobacteria dominate several infant bacterial communities, but were much less abundant in the maternal samples (Fig. 1d). Principal coordinate analysis (PCoA) of weighted UniFrac distances demonstrate distinct and tight clustering of those from maternal samples and a more diffuse distribution of infant samples (Fig. 1e). Group significance

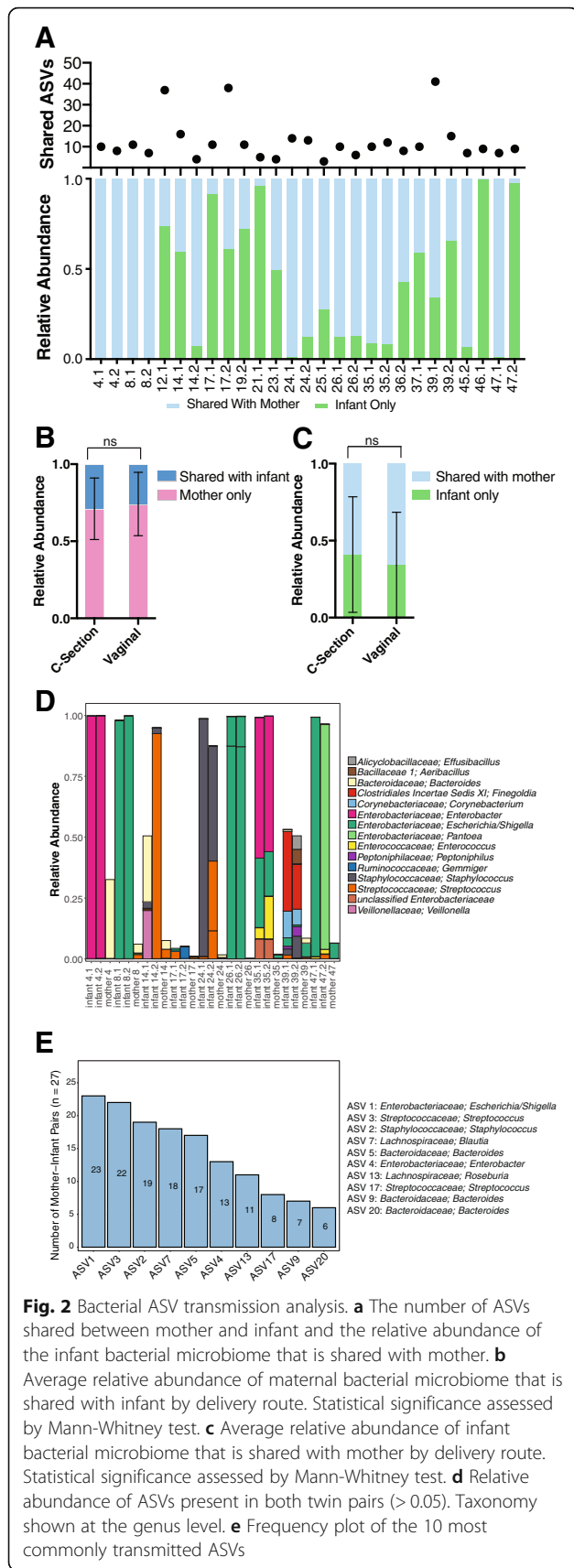
testing with ADONIS confirmed that the clustering of samples by source (mother or infant) was statistically significant and accounts for approximately 23% of the variation in weighted UniFrac distances ( $p < 0.001$ ,  $R^2 = 0.2347$ ). To determine if other factors might be associated with specific microbial communities, we performed a multivariate analyses and random forest classification; both methods demonstrate that mothers and infants have significant differences between their bacterial microbiome (Additional file 1: Figure S1B–E). Race was identified as a significant determinant of bacterial community composition by multivariate analysis, but not by random forest (Additional file 1: Figure S1B). Therefore, age of host (i.e., adult vs. infant) was the only factor that separated the bacterial communities in this cohort.

Infant bacterial communities and the relationship to maternal communities were further characterized by pairwise weighted UniFrac distances. Similar to the PCoA data (Fig. 1e), the pairwise distance between all infant bacterial communities was significantly greater ( $p < 0.001$ ) than the mean distance between all maternal communities (Fig. 1f). To test if the infant gut bacterial communities more closely resembled those of their own mother than all unrelated mothers, we measured the distances of infant bacterial community structure to their cognate mother versus the distances to all mothers in the study. Interestingly, the distance of an infant to their own mother was no different than to an unrelated mother (Fig. 1g). However, the mean distance between co-twin bacterial communities was significantly less ( $p < 0.001$ ) than to unrelated infants (Fig. 1h). The distance between co-twins did not differ with zygosity (Additional file 1: Figure S1F), suggesting that age and environmental exposures are the more important determinants of the bacterial communities.

To develop a more detailed view of the relationship of related mother-infant bacterial communities, we identified bacterial taxa shared between the communities of mothers and their related infant(s). On average, a related mother and infant share 12 bacterial taxa which accounted for 63% of the total reads in the infant's sample (Fig. 2a). We then asked if mothers share differing amounts of their bacterial communities with their infants depending on the route of delivery. The average relative abundance of reads from shared taxa within the mother did not vary by route of delivery (Fig. 2b). We next examined if the amount of the infants' bacterial microbiome that was shared with their mother differed when comparing infants born vaginally versus by C-section and found no influence of route of delivery (Fig. 2c). We next asked what bacterial taxa were shared between twins and what is the relative of abundance of these bacterial taxa in both the infant and maternal stools. In general, while the bacterial taxa present in



**Fig. 1** Bacterial microbiota analysis of mothers and infants. **a** Overview of study design. **b** Richness of bacterial amplicon sequence variants (ASV) in infants and mothers. Statistical significance assessed by Mann-Whitney test. **c** Alpha diversity of bacterial ASV in mothers and infants. Statistical significance assessed by Mann-Whitney test. **d** Relative abundance of bacteria at phylum level for mothers and infants; 11 (infant twin 1) and 12 (infant twin 2). **e** PCoA plot of weighted UniFrac distances. **f** Weighted UniFrac pairwise comparison within infants and within mothers. Statistical significance assessed by Mann-Whitney test. **g** Weighted UniFrac pairwise comparison between related mother-infant ( $n = 27$  pairs) and between unrelated mother-infant ( $n = 765$  pairs). Statistical significance assessed by Mann-Whitney test. **h** Weighted UniFrac pairwise comparison between co-twins ( $n = 11$  twin pairs) and between unrelated infants ( $n = 517$  pairs). Statistical significance assessed by Mann-Whitney test



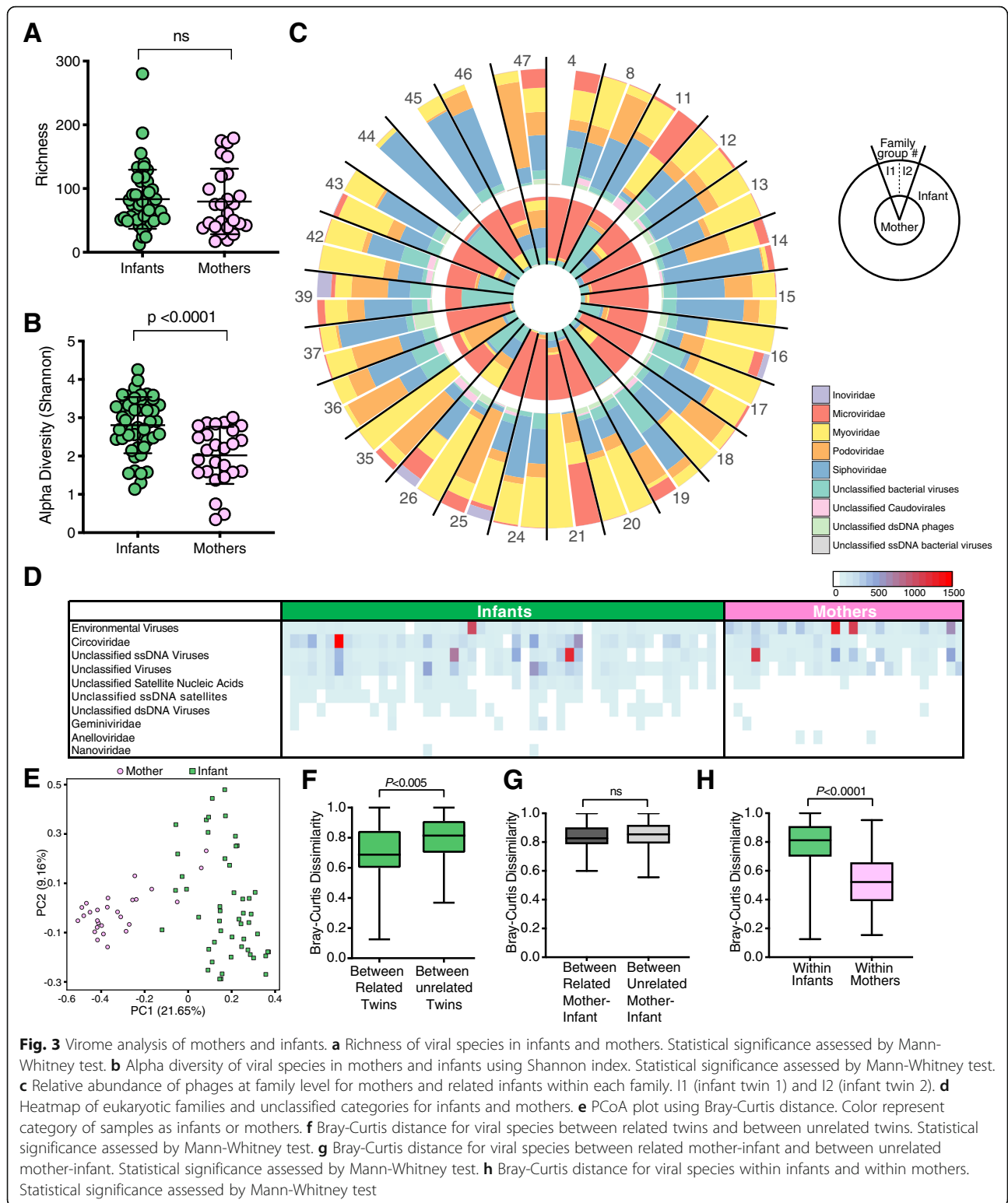
**Fig. 2** Bacterial ASV transmission analysis. **a** The number of ASVs shared between mother and infant and the relative abundance of the infant bacterial microbiome that is shared with mother. **b** Average relative abundance of maternal bacterial microbiome that is shared with infant by delivery route. Statistical significance assessed by Mann-Whitney test. **c** Average relative abundance of infant bacterial microbiome that is shared with mother by delivery route. Statistical significance assessed by Mann-Whitney test. **d** Relative abundance of ASVs present in both twin pairs (>0.05). Taxonomy shown at the genus level. **e** Frequency plot of the 10 most commonly transmitted ASVs

both twins had a high relative abundance in the infant, they represented only a small fraction of the maternal bacterial community (Fig. 2d and Additional file 7: Table S2). We additionally examined if there were ASVs which were frequently shared by mother-infant pairs across all the families (Fig. 2e). One ASV belonging to the *Escherichia/Shigella* genus was shared in 23 of the 27 mother-infant pairs (Additional file 8: Table S3). The relative abundance of these 10 commonly shared ASVs was again generally less in the maternal samples than the infant samples (Additional file 2: Figure S2). These data collectively indicate that the mother shares a relatively small proportion of bacterial community with her infants, that these shared taxa make significant contributions to the infant’s bacterial microbiome, and that the amount either shared (mother) or received (infant) does not vary by delivery route.

**Gut virome of infant twin pairs and their mothers**

To assess early life mother-infant virome transmission, we performed metagenomic sequencing of the same fecal specimens from 56 healthy infants (28 twin pairs, collected on or before the fourth day of life) and their 28 healthy mothers. Seven of the 84 samples were not entered into analysis: five did not amplify, one had insufficient volume after filtering, and one had low sequencing reads (3768) (Additional file 6: Table S1). This resulted in a virome dataset for 50 infants and 27 mothers, i.e., 26 families with mother and at least one infant. There was no significant difference between the sequence depth of infants (mean 131,846 ± 105,022) and maternal samples (mean 103,040 ± 39,986) ( $p = 0.271$ ) (Additional file 3: Figure S3A). The mean infant and maternal gut virome species richness were similar (83 vs 78 species, respectively) ( $P = 0.6091$ ) (Fig. 3a), but the gut virome Shannon diversity was significantly greater in infants than in mothers ( $p < 0.0001$ ) (Fig. 3b). The most abundant bacteriophages in the maternal gut virome were from the *Microviridae* family, whereas infant gut viromes consisted predominantly of bacteriophages from the *Myoviridae*, *Podoviridae*, and *Siphoviridae* families (Fig. 3c). Circoviruses (*Circoviridae* family) and unclassified viruses (within the “ssDNA viruses” and “environmental viruses” taxonomic group classifications) were frequently detected in both mother and infant samples (Fig. 3d). Taken together, these data suggest considerable age-dependent virome differences (i.e., whether the specimen is from an adult or infant) that might confound assessments of familial relationships between related mother-infant pairs.

To resolve this, we first compared virome diversity by measuring the unweighted Bray-Curtis distance between gut virome communities. PCoA indicated that the maternal gut viromes differ from infant gut viromes



(Fig. 3e). To further understand additional factors that might influence the composition of the infant gut virome, we performed multivariate analyses of the cohort metadata (e.g., delivery route, feeding type, delivery site)

and found that age (i.e., maternal or infant origin of stool) was significantly associated with differences in the virome (Additional file 3: Figures S3B–D). This was corroborated by building classifier models with random



forest that accurately distinguished the virome of mothers from infants (AUC = 0.996) (Additional file 3: Figure S3E). Maternal BMI (pre-pregnancy) was also identified as a determinant by the multivariate analyses but was not a significant factor in random forest (Additional file 3: Figure S3B).

Because mother-infant dyads shared a substantial subset of their bacterial microbiome, we hypothesized that they would also harbor relatively similar viromes. Consistent with this hypothesis, we found that related infants shared a more similar gut virome to their co-twin than to unrelated infants (Fig. 3f,  $p < 0.005$ ). This co-twin virome similarity was not influenced by their zygosity (Additional file 3: Figure S3F). However, infant gut viromes were highly dissimilar from their own mother. In fact, the ecological distance between an infant and their own mother was the same as it was to unrelated mothers (Fig. 3g,  $p = 0.1400$ ). This might be explained in part by the significantly lower virome inter-personal variation among mothers as compared to infants (Fig. 3h,  $p < 0.0001$ ).

These diversity measurements do not account for differences in virome abundance (i.e., unweighted distances). Hence, we assessed virome transmission in terms of the shared viral abundance. For example, a large proportion of the infant virome might be established from transmission of a relatively small number of viruses from the mother. To quantify the proportion of the infant virome shared with their mother, we mapped the infants' sequencing reads to viral contigs assembled from the infant(s) and their mothers' virome. On average, only 15.0% of the viruses present in the infant stool could be mapped back to the sequences in their mother's stool (Fig. 4a). Conversely, 19.8% of the mother's virome was shared with her infant (Fig. 4b). The proportion of the viruses shared between mothers and their infants was not influenced by delivery route (Fig. 4c and d). Further, the number of viruses transmitted from mother-to-infant was independent of the number of viruses in the mother's virome (Fig. 4e, slope  $m = 0.127$ ). Thus, unlike the bacterial microbiome, infants shared only a limited proportion (approximately 15%) of their gut virome with their mother. These results were also validated by analyses of bacteriophage contigs identified by VirSorter (Additional file 4: Figure S4).

#### Route of delivery influences transkingdom interactions between bacteria and bacteriophage

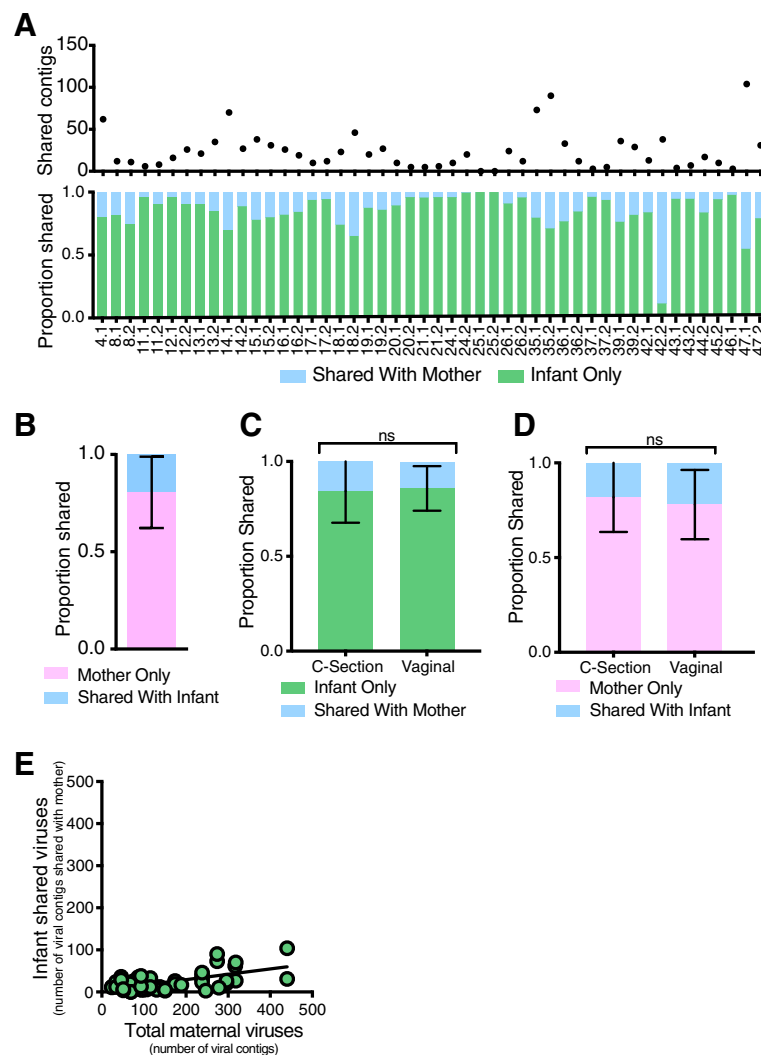
After demonstrating that only a subset of the virome and bacterial microbiome is shared between mothers and infants, we sought clinical factors that might influence their transkingdom interactions. Hierarchical clustering indicated that the gut bacterial microbiomes of infants born by vaginal delivery (*Ruminococcaceae*,

*Lachnospiraceae*, *Bacteroidales*, and *Lactobacillales* amplicon sequence variants (ASV)) were strongly correlated with bacteriophages of the *Microviridae*, *Myoviridae*, *Siphoviridae*, and *Podoviridae* families (Fig. 5). In contrast, microbiome interactions of infants born by C-section delivery were discordant from infants born by vaginal delivery. First, bacteriophages also detected in the stools of infants born by C-section correlated with different bacteria than the same bacteriophages from infants delivered vaginally. In children born by C-section, *Microviridae*, *Myoviridae*, *Siphoviridae*, and *Podoviridae* bacteriophages were strongly correlated with *Staphylococcaceae*, *Bacteroidaceae*, *Lachnospiraceae*, and other *Ruminococcaceae* ASVs. For example, Parabacteroides phage YZ-2015b from *Microviridae* family correlated with *Bacteroidales* ASVs (*Bacteroides uniformis* and *Parabacteroides distasonis*) in infants born through vaginal delivery, but correlated with other *Bacteroidales* ASVs (*Bacteriodes ovatus*) in infants born through C-section (Fig. 5, compare pink outline; Additional file 5: Figure S5A). Second, microbiome interactions in stool were also absent from infants born by C-section compared to those born vaginally (Fig. 5, compare green outline), indicating a transkingdom signature which was unique to each delivery route. This was corroborated by Pearson's correlations between bacterial ASVs- and VirSorter-identified contigs (Additional file 5: Figure S5B). Therefore, this suggests that delivery route might influence interactions between the bacterial microbiome and virome.

#### Discussion

The dynamic neonatal bacterial microbiome and virome likely influence or even define the more stable adult bacterial microbiome and virome. It is thought that proper assembly of these communities promotes health. However, precisely how the neonatal gut bacterial microbiome and virome are initially populated is not well understood. Here, we find that twins and their mothers have distinct gut virome and bacterial microbiomes. This finding is consistent with studies showing bacterial microbiome and virome composition is age dependent [23, 24]. However, despite having unique communities, mothers and their infants do share some components, predominantly among the bacterial communities. Interestingly the proportion and specific taxa that are shared differed within and between families. It is important to note, that while the proportion of the bacterial microbiome that a mother shares with her infant is relatively small, these bacteria do account for over half of the infant's bacterial microbiome. In our study, the average of maternal-matched bacterial microbiome is 63%, similar to the ranges (30–70%) of other cohorts [16, 18, 25, 26]. We do note that 18 infants stool bacterial microbiomes





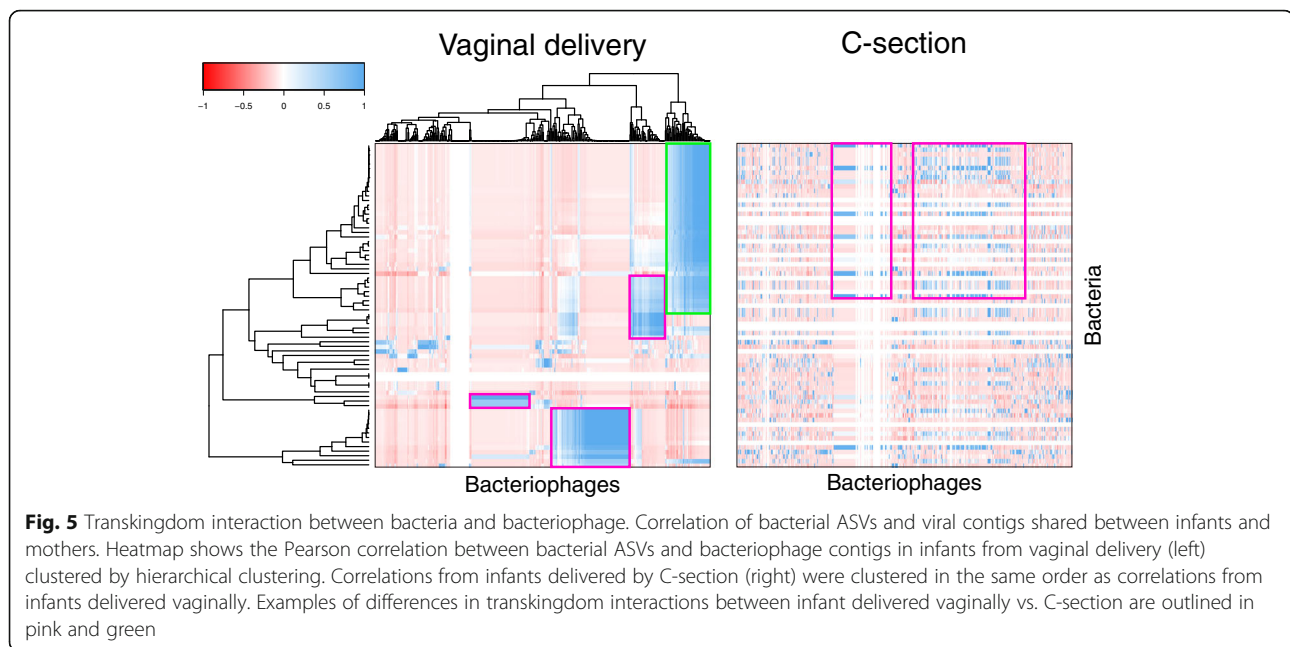
**Fig. 4** Viral contig transmission analysis. **a** The number of contigs shared between mother and infant and the proportion of infant virome that is shared with mother or is infant only. **b** Average contig proportion of mother virome that is shared with infant or is mother only. **c** Average contig proportion of infant virome that is shared with mother by delivery route. Statistical significance assessed by Mann-Whitney test. **d** Average contig proportion of maternal virome that is shared with infant by delivery route. Statistical significance assessed by Mann-Whitney test. **e** Plot of the number of infant shared viral contigs vs. the total number of maternal contigs. Linear regression line fit to data

were indistinguishable from buffer background, and if we interpret these as having no maternal transmission, 63% may be an overestimate. Additionally, 16S rRNA gene sequencing may not always have sufficient resolution to distinguish different strains which would also lead to an overestimation of shared bacteria. However, together, these studies emphasize that the maternal gut microbiome can be a substantial, though not the exclusive, source of an infant's early bacterial microbiome.

Our virome data are unique, and offer substantial contrast to the bacterial microbiome. Our finding of only 15% of the infant's virome in the cognate newborn's mother's stool strongly suggest that infant gut viruses are likely from other habitats (e.g., mother's skin,

breastmilk) or the inanimate environment. Second, it is possible that the early window in this study (1–4 days of life) is when bacterial colonization is still underway. This builds on the longitudinal study of Ferretti et al., which demonstrated that maternal gut-sourced bacteria increasingly dominate the infant microbiome by 4 months [16]. In this model, it is plausible that “maternally sourced” viruses that can replicate in “maternally sourced” bacteria would become enriched weeks to months after bacterial colonization. This would be consistent with the shift in bacteriophage community previously reported to occur from 0 to 2 years [20].

This twin study also offers unique data to compare vertical transmission and co-twin relationships. The



greater similarity of the bacterial microbiome and virome between twin pairs than unrelated twins may reflect commonality of a shared environmental exposure. Shared exposure from a maternal site other than stool, such as maternal breastmilk, might be another source. Although breastfeeding did not significantly influence mother-to-infant bacterial transmission, the viromes of breastfed infants are more similar to their mother than were those of formula fed infants (Additional file 1: Figure S1G and Additional file 3: Figure S3G). Given the limitations of this cohort, future studies will be needed to clarify the role of diet. Nonetheless, this highlights that the dynamics of the acquisition of the human bacterial microbiome and virome clearly differ.

The neonatal gut virome consists of largely bacteriophage with only a small portion being eukaryotic viruses. Of note, we focused on the DNA virome as our previous study of the infant virome [20] (those four twin pairs are also included here) found a sparse RNA virus component early in life. As with the neonatal bacterial microbiome, the virome varies across individuals. In contrast to what has been described in adult twins, bacteriophage diversity does not correlate with bacterial diversity [27]. Additionally, the discrepancy in proportion of overlap between the maternal and infant virome and bacterial microbiome suggests that bacterial microbiome and virome are acquired from different sources.

Previous studies of the effect of birth route on the newborn gut bacterial microbiome have offered conflicting findings. While we found bacterial communities differed significantly between infants born vaginally versus those born by C-section, we were unable to find

discriminating ASVs for delivery route. The shared proportion of the infant and maternal bacterial microbiome is not significantly different by delivery route. Furthermore, we found no difference between the Bray-Curtis dissimilarity of infant gut virome by delivery route, and we found no viral contig that discriminated between infants born vaginally versus by C-Section. Additionally, the proportion of infant and maternal virome that is shared is not significantly different between infants born vaginally versus by C-section. However, when we examined transkingdom interactions, we found that this relationship did differ by delivery route. This transkingdom interaction might reconcile why some studies show delivery route is an important factor, while others do not.

## Conclusions

In summary, our data demonstrates that while mothers and infants do share some of their gut microbial communities, the degree of sharing is more substantial for bacteria than for virus. Additionally, the proportion and specific taxa shared varies between and even within families. Lastly, bacterial-bacteriophage interactions differed depending on delivery route. Therefore, our data depict that the inter-generation transmissibility at birth diverges between the major kingdoms of microbes suggesting that the founding sources of these communities may differ.

## Methods

### Subjects

This study was approved by the Human Research Protection Office of Washington University School of

Medicine in St. Louis and Missouri-Baptist Medical Center institutional review board. We obtained written informed consent from women pregnant with twin gestations were consented to collect stool specimens around the time of birth from the mother and her twins. Stools from the mothers were collected at home or in the hospital; infant stools were collected in the hospital. Home-produced samples were couriered to the laboratory in insulated envelopes containing frozen ( $-20^{\circ}\text{C}$ ) freezer packs and stored at  $-80^{\circ}\text{C}$  until analysis. Twenty-eight families were selected in which maternal stools were collected peripartum and there was adequate sample from early in life from the twins. Metadata collected included mode of delivery, feeding content, maternal age, maternal weight and weight gain, and maternal race (Table 1 and Additional file 6: Table S1). This birth cohort has been described in various detail in prior publications [20, 23, 28–33].

#### Bacterial 16S rRNA gene sequencing

One hundred milligrams of stool was disrupted by bead beating and DNA was extracted using QIAamp DNA Stool Mini Kit on a QIAcube automated DNA extraction unit. In parallel, four buffer-only controls were disrupted by bead beating and similarly extracted to serve as extraction negative controls. PCR was performed using Golay-barcoded primers specific for the V4 region (F515/R806). Reactions were held at  $94^{\circ}\text{C}$  for 2 min to denature the DNA, with amplification proceeding for 40 cycles at  $94^{\circ}\text{C}$  for 15 s,  $50^{\circ}\text{C}$  for 30 s, and  $68^{\circ}\text{C}$  for 30 s; a final extension of 2 min at  $68^{\circ}\text{C}$ . Stool and buffer samples, as well as 4 water negative controls, were amplified in triplicate, combined, and cleaned using Agencourt Ampure XP beads (Beckman-Coulter). Equimolar libraries were pooled and sequenced using an Illumina MiSeq sequencer ( $2 \times 250$  v2 kit) at the Center for Genome Sciences & Systems Biology at Washington University.

**Table 1** Cohort demographics

Infant age at time of stool (h) (median, IQR, range)	37.2 (16.7, 54.4), 3.3 to 92.5
Maternal stool time from birth (h) (median, IQR, range)	108 (36, 390), –360 to 1680
C-section no. (%)	35 (62.5%)
Feeding (between birth and sampling) no. (%)	
Breastmilk	8 (14.3%)
Formula	16 (28.6%)
Mix	32 (57.1%)
Monozygotic no. (%)	30 (53.6%)

#### Bacterial 16S rRNA gene analysis

16S rRNA amplicon sequences were demultiplexed using QIIME (Quantitative Insights Into Microbial Ecology, v 1.8.0) [34]. Read quality control and the resolution of amplicon sequence variants (ASVs) using the forward-reads only was performed with the DADA2 R package [35]. Taxonomy was assigned to the ASVs from the DADA2-formatted training files derived from the Ribosomal Database Project's Training Set 16 [36]. ASVs were aligned using an R implementation of MSA [37] and arranged into a maximum likelihood phylogeny (GTR model with optimization of the proportion of invariable sites and the gamma rate parameter) using the phangorn package [38]. The ASV counts, taxonomy assignments, phylogeny, and sample metadata were combined to generate a PhyloSeq object for further analyses [39]. Five samples failed to amplify. Four samples with less than 1000 reads were excluded from analysis. To compare the bacterial community structure between the stool samples and controls we generated a multidimensional scaling (MDS) plot of the unweighted UniFrac distance with normal confidence ellipses. Eighteen samples encapsulated by the confidence ellipse of the buffer control samples were removed from further analysis (Additional file 1: Figure S1A). The Decontam R package "prevalence" method at a threshold of 0.25 was used to identify and remove remaining contaminating sequences [40]. Further ecological analyses were performed in R.

If a mother and one of her related infants had an identical ASV, then it was assumed to be shared between mother and infant; otherwise, it was exclusive to mothers or infants. Only families that had a sample from mother and at least one of her infants were used to determine potential mother to infant transmission.

#### Virome sequencing

Fecal specimens (approximately 100 mg) were diluted 1:6 with phosphate-buffered saline (PBS) and filtered through a  $0.45\text{-}\mu\text{m}$ -pore-size membrane. Total nucleic acid was extracted from the filtrate using COBAS Ampliprep (Roche). In parallel, PBS was filtered and extracted to serve as extraction reagent only control. Total nucleic acid from stool and negative controls was amplified with Phi29 polymerase (GenomiPhi V2 kit, GE Healthcare) according to the manufacturer's instructions and used for Nextera DNA library construction (Illumina). Additionally, to evaluate the level of specimen cross-contamination that might occur after library construction [41], a uniquely indexed library of cDNA derived from the nematode Orsay virus RNA1 segment [42] was included in the pool for each sequencing run. Libraries were purified and size-selected using Agencourt Ampure XP beads (Beckman-Coulter),

followed by quantification on a 2100 Bioanalyzer (Agilent Technologies).

#### **Virome analysis (read-based)**

To control for sample extraction and cross-sample contamination, we included two types of negative controls: buffer-only controls that underwent the same extraction process, and PBS spiked with Orsay virus (a nodavirus of nematodes). Illumina MiSeq sequencing reads ( $2 \times 250$  bp) were processed through the VirusSeeker workflow (version 0.063) [43]. In brief, VirusSeeker follows a hierarchical workflow to perform a quality filter, subtract host sequences (human), identify candidate viral sequences through curated virus-only databases, then search sequentially against comprehensive databases to remove potential false positives (i.e., candidate viral sequences that have higher similarity to non-viral sequences). VirusSeeker takes a multi-step approach to remove bacterial sequences. Candidate viral sequences are queried against NCBI bacteria reference genomes to remove bacterial sequences. Second, VirusSeeker also performs three sequential BLAST searches (MegaBLAST against the complete NCBI NT database, BLASTn against the complete NT database, and BLASTx against the complete NCBI NR database) to remove potential false-positive viral sequences that have higher similarity to non-viral sequences (bacterial, fungal, etc.).

Contaminants were identified by decontam (version 1.0.0) [40] in a 2-step process, which uses a prevalence-based method to determine contaminant taxa. First, samples were compared to “Orsay” controls in decontam at a threshold of 0.1, and identified contaminant taxa were removed. Next, buffer controls were compared to samples using decontam at a stricter threshold of 0.5, and contaminants removed likewise. The resulting quality-controlled sequencing reads of each sample were then normalized to 31,000. Virus species with fewer than three reads were masked.

Ecological analyses (richness, alpha diversity—Shannon Index) were performed using Vegan R package (version 2.5-2) [44]. The beta diversity and PCoA plots were obtained by using QIIME [34] using the Bray-Curtis distance on a presence-absence species data matrix. Results were plotted in GraphPad Prism (version 8) and the statistical significance was assessed using a non-parametric Mann-Whitney  $U$  test. For the familial relative abundance plot, only paired data of one mother and at least one infant per family were used.

#### **Metadata**

##### ***MaAslin***

MaAslin, a multivariate linear modeling tool, was utilized in finding the association between study metadata

and relative abundance [45]. The significance of each association is given via a  $q$ -value; any relationship with a  $q$ -value of less than 0.05 is shown. For the metadata categories exclusive to infants such as, delivery route, feeding type, and zygosity, only the infant data was used. All sample data was used for the other general categories such as, mother/infant, delivery site, pre-pregnancy BMI [category], and race.

##### ***Random forest***

RandomForest was used to classify metadata categories (version 4.6-14) [46]. Five hundred trees were built using randomForest and the Out-Of-Bag samples were used to test the classifier and obtain their error rates. The ROC curves and AUC measures for the classifiers were made by a R package, pROC, which plots pseudo-probabilities from classification tree votes for OOB samples (version 1.13.0) [47]. Similar to MaAslin metadata analysis, for the metadata categories exclusive to infants, such as, delivery route, feeding type, and zygosity, only the infant data was used. For the other general categories such as mother/infant, delivery site, pre-pregnancy BMI [category], and race, all sample data was used.

#### **Virome contig analyses**

##### ***Contigs databases***

Using IDBA\_UD, the QC reads of each sample were used to create initial raw contigs (version 1.1.0) [48]. These contigs were then filtered using bbtools [31] to remove contigs with length 500 bp and de-duplicated at minidentity  $\geq 99$ . The contigs created from buffer samples were queried against NT database using megablast and any contig with a coverage  $\geq 95$  and percent identity  $\geq 97$ , was designated as a true contaminant and used to create a database of contaminant contigs. Megablast was used to query sample contigs against contig contaminant database and any hit to this database resulted in the removal of such contig from sample contigs. After contaminated contig removal, for each family, the contigs were concatenated into one file and then merged to create longer contigs using minimus2 [49]. These familial contigs were queried against viral NR database using blastx and the contigs with a hit against this database were used to create the familial viral contig database. We also corroborated these contig analyses by identifying bacteriophage contigs with VirSorter (parameters—virome, -db 1) (Additional file 4: Figure S4) [50].

##### ***Contig transmission***

To identify viral contigs transmitted within each family group, BWA-MEM was used to map the QC reads of each sample to their corresponding familial viral contig database (version 0.7.17) [51]. The parameters -L 97,97 and -M were used to mark short reads as secondary.



The mapped reads were used to measure shared contigs between mother and infant and used to determine potential transmission of virome from mother to infant. If a mother and one of her related infants had a read mapped to the same contig, then it was assumed to be shared between mother and infant; otherwise, it was exclusive to mothers or infants. Only families that had a sample from mother and at least one of her infants were used to determine potential mother to infant transmission. VirSorter-identified contigs were also analyzed in an identical manner, by generating a VirSorter contig database and mapping reads to the database to determine potential mother to infant transmission (Additional file 4: Figure S4).

### Transkingdom analysis

To characterize transkingdom interactions in mother-to-infant transmission, Pearson's correlations were performed between bacterial ASVs and bacteriophage contigs that were shared between mothers and at least one of her infants. Bacterial ASVs and bacteriophage contigs that were present in only one family were excluded. Correlations were performed separately by delivery route and clustered on the vaginal delivery dataset using the R package gplots. To contrast differences in the C-section profile, the same clustering order (vaginal delivery) was used to visualize correlations in the C-section dataset. Pearson's correlation was also performed between the ASVs and VirSorter identified bacteriophages.

### Statistics

Mann-Whitney  $U$  test was used to test significance when there were two independent groups, for example, related twins versus unrelated twins. Kruskal-Wallis test and Dunn's multiple comparison were used to test significance between three or more groups, for example the feeding type of breastmilk, formula, or mixed. Pearson's correlation was used obtain correlation between transmitted viruses and bacteria. All  $p$  values were two-tailed.

### Code availability

A fully reproducible workflow of the analysis presented in this manuscript can be found at [https://github.com/RachelRodgers/Holtz-Lim\\_MotherTwinInfant\\_Virome\\_BacterialMicrobiome](https://github.com/RachelRodgers/Holtz-Lim_MotherTwinInfant_Virome_BacterialMicrobiome)

### Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s40168-019-0766-7>.

**Additional file 1: Figure S1.** Metadata analysis of bacterial ASVs. (A) Multidimensional scaling (MDS) plot of the unweighted UniFrac distance with normal confidence ellipses. Expanded version shows eighteen samples encapsulated by the confidence ellipse of the buffer control

samples (B) Output of MaAslin and random forest analysis. (C) Important ASVs and their statistical significance as assessed by MaAslin and random forest. (D) Heatmap of MaAslin identified ASVs for mothers and infants. (E) ROC curves and AUC measures for delivery route and mother/infant classification using random forest and pROC packages in R. Pseudo-probabilities are plotted on graph using Out-Of-Bag (OOB) sample tree classification votes. Only infant data used in vaginal vs. C-section classification while both mother and infant data used for mother vs. infant classification. (F) Weighted UniFrac pairwise comparisons. Statistical significance assessed by Mann-Whitney, and Kruskal-Wallis with Dunn's multiple correction (feeding type: breastmilk vs. formula; breastmilk vs. mix; formula vs. mix). (G) Weighted UniFrac pairwise comparisons for related mother infant pairs. Statistical significance assessed by Kruskal-Wallis and Mann-Whitney.

**Additional file 2: Figure S2.** Relative abundance of the 10 most frequently shared (mother-infant) ASVs. Taxonomy shown at the genus level.

**Additional file 3: Figure S3.** Metadata analysis of viral species. (A) Number of sequencing reads in infant and maternal samples. Statistical significance assessed by Mann-Whitney. (B) Output of MaAslin and random forest analysis. (C) Important viral species and their statistical significance as assessed by MaAslin and random forest. (D) Heatmap of MaAslin identified viral species for mothers and infants. (E) ROC curves and AUC measures for delivery route and mother/infant classification using random forest and pROC packages in R. Pseudo-probabilities are plotted on graph using Out-Of-Bag (OOB) sample tree classification votes. Only infant data used in vaginal vs. C-section classification while both mother and infant data used for mother vs. infant classification. (F) Bray-Curtis Dissimilarity pairwise comparisons. Statistical significance assessed by Mann-Whitney. Kruskal-Wallis with Dunn's multiple correction was used to assess feeding type (breastmilk vs. formula; breastmilk vs. mix; formula vs. mix) and Pre-pregnancy BMI (obese vs. overweight; obese vs. normal; overweight vs. normal). (G) Bray-Curtis Dissimilarity pairwise comparisons for related mother infant pairs. Statistical significance assessed by Mann-Whitney. Kruskal-Wallis with Dunn's multiple correction was used to assess feeding type (breastmilk vs. formula; breastmilk vs. mix; formula vs. mix) and Pre-pregnancy BMI (obese vs. overweight; obese vs. normal; overweight vs. normal).

**Additional file 4: Figure S4.** VirSorter viral contig transmission analysis. Analyses of VirSorter contigs were performed as in Fig. 4. (A) The number of infant VirSorter viral contigs shared with their mother (top), and their proportion out of the total infant virome (VirSorter identified) is shown (bottom). Percent of infant VirSorter contigs that are shared with their mother is shown in blue and infant only VirSorter contigs are shown in green. (B) Average VirSorter contig proportion of mothers' virome that is shared with infant or unique to the mother. (C) Average VirSorter contig proportion of infant virome that is shared with mother by delivery route. Statistical significance assessed by Mann-Whitney test. (D) Average VirSorter contig proportion of maternal virome that is shared with infant by delivery route. Statistical significance assessed by Mann-Whitney test. (E) Plot of the number of infant shared VirSorter viral contigs vs. the total number of maternal VirSorter contigs. Linear regression line fit to data is shown.

**Additional file 5: Figure S5.** Transkingdom interaction between bacteria and bacteriophage. Correlation of shared bacterial ASVs and viral contigs. (A) Heatmap shows the Pearson correlation between bacterial ASVs and bacteriophage contigs in infants from C-section delivery (left) clustered by hierarchical clustering. Second heatmap shows correlations from infants delivered vaginally clustered in the same order as correlations from infants delivered by C-section. (B) Heatmap shows the Pearson correlation between bacterial ASVs and VirSorter identified bacteriophage contigs in infants delivered vaginally (left) clustered by hierarchical clustering. Second heatmap shows correlations from infants delivered by C-section clustered in the same order as correlations from infants delivered vaginally. These VirSorter analyses corroborate the contig analyses shown in Fig. 5.

**Additional file 6: Table S1.** Cohort metadata and sample inclusion status.

**Additional file 7: Table S2.** Read counts and taxonomy of ASVs shared by both twins.

**Additional file 8: Table S3.** Taxonomy of the 10 most frequently transmitted ASVs.

**Acknowledgements**

We thank the families and their physicians for their participation in, and cooperation with, the study.

**Authors' contributions**

ESL and LRH conceived and designed the experiments. CR prepared samples for sequencing. CR performed PCR experiments. RR, RM, ESL, and SH processed and analyzed the sequencing data. PIT, BBW, and MN recruited the study participants and managed the metadata. ESL and LRH wrote and edited the manuscript. All authors read and approved the final manuscript.

**Funding**

This work was supported in part by the Children's Discovery Institute (MD-FR-2013-292), Doris Duke Charitable Foundation (2017076), March of Dimes (BOC 388999), NIH grant R00 DK107923, Arizona State University startup funding (ESL), the National Institutes of Health (5P30 DK052574 (Biobank, DDRCC) to P.I.T. and UH3AI083265 to P.I.T. and B.B.W.). P.I.T. and B.B.W. received funding from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, and the Foundation for the National Institutes of Health (made possible by support from the Gerber Foundation).

**Availability of data and materials**

Sequence data has been deposited to the European Nucleotide Archive under accession number PRJEB33578. Reads mapping to human have been removed from the submitted metagenomic sequence data. A fully reproducible workflow of the analysis presented in this manuscript can be found at [https://github.com/RachelRodgers/Holtz-Lim\\_MotherTwinInfant\\_Virome\\_BacterialMicrobiome](https://github.com/RachelRodgers/Holtz-Lim_MotherTwinInfant_Virome_BacterialMicrobiome)

**Ethics approval and consent to participate**

The institutional review board at Washington University School of Medicine and Missouri-Baptist Medical Center approved the study. Subjects provided informed consent to be included.

**Consent for publication**

Not applicable

**Competing interests**

P.I.T. is a consultant to Takeda Pharmaceuticals on childhood gastrointestinal diseases. The other authors declare that they have no competing interests.

**Author details**

<sup>1</sup>School of Life Sciences, Arizona State University, Tempe, AZ 85287, USA. <sup>2</sup>Center for Fundamental and Applied Microbiomics, The Biodesign Institute, Tempe, AZ 85287, USA. <sup>3</sup>Department of Pediatrics, Washington University School of Medicine, St. Louis, MO 63110, USA. <sup>4</sup>Department of Pathology & Immunology, Washington University School of Medicine, St. Louis, MO 63110, USA. <sup>5</sup>Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, MO 63110, USA.

Received: 4 September 2019 Accepted: 8 November 2019

Published online: 10 December 2019

**References**

- White RA, Bjornholt JV, Baird DD, Midtvedt T, Harris JR, Pagano M, Hide W, Rudi K, Moen B, Iszatt N, et al. Novel developmental analyses identify longitudinal patterns of early gut microbiota that affect infant growth. *PLoS Comput Biol*. 2013;9:e1003042.
- Britton RA, Young VB. Interaction between the intestinal microbiota and host in *Clostridium difficile* colonization resistance. *Trends Microbiol*. 2012;20:313–9.
- Frank DN, Robertson CE, Hamm CM, Kpadeh Z, Zhang T, Chen H, Zhu W, Sartor RB, Boedeker EC, Harpaz N, et al. Disease phenotype and genotype are associated with shifts in intestinal-associated microbiota in inflammatory bowel diseases. *Inflamm Bowel Dis*. 2011;17:179–84.
- Olszak T, An D, Zeissig S, Vera MP, Richter J, Franke A, Glickman JN, Siebert R, Baron RM, Kasper DL, Blumberg RS. Microbial exposure during early life has persistent effects on natural killer T cell function. *Science*. 2012;336:489–93.
- Stout MJ, Conlon B, Landeau M, Lee I, Bower C, Zhao Q, Roehl KA, Nelson DM, Macones GA, Mysorekar IU. Identification of intracellular bacteria in the basal plate of the human placenta in term and preterm gestations. *Am J Obstet Gynecol*. 2013;208:226 e221–7.
- Aagaard K, Ma J, Antony KM, Ganu R, Petrosino J, Versalovic J. The placenta harbors a unique microbiome. *Sci Transl Med*. 2014;6:237ra265.
- Collado MC, Rautava S, Aakko J, Isolauri E, Salminen S. Human gut colonisation may be initiated in utero by distinct microbial communities in the placenta and amniotic fluid. *Sci Rep*. 2016;6:23129.
- Lauder AP, Roche AM, Sherrill-Mix S, Bailey A, Laughlin AL, Bittinger K, Leite R, Elovitz MA, Parry S, Bushman FD. Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota. *Microbiome*. 2016;4:29.
- Perez-Munoz ME, Arrieta MC, Ramer-Tait AE, Walter J. A critical assessment of the "sterile womb" and "in utero colonization" hypotheses: implications for research on the pioneer infant microbiome. *Microbiome*. 2017;5:48.
- Lim ES, Rodriguez C, Holtz LR. Amniotic fluid from healthy term pregnancies does not harbor a detectable microbial community. *Microbiome*. 2018;6:87.
- de Goffau MC, Lager S, Sovio U, Gaccioli F, Cook E, Peacock SJ, Parkhill J, Charnock-Jones DS, Smith GCS. Human placenta has no microbiome but can contain potential pathogens. *Nature*. 2019;572:329.
- Yassour M, Jason E, Hogstrom LJ, Arthur TD, Tripathi S, Siljander H, Selvenius J, Oikarinen S, Hyoty H, Virtanen SM, et al. Strain-Level Analysis of Mother-to-Child Bacterial Transmission during the First Few Months of Life. *Cell Host Microbe*. 2018;24:146–154 e144.
- Asnicar F, Manara S, Zolfo M, Truong DT, Scholz M, Armanini F, Ferretti P, Gorfer V, Pedrotti A, Tett A, Segata N. Studying Vertical Microbiome Transmission from Mothers to Infants by Strain-Level Metagenomic Profiling. *mSystems*. 2017;2:e00164.
- Nayfach S, Rodriguez-Mueller B, Garud N, Pollard KS. An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography. *Genome Res*. 2016;26:1612–25.
- Milani C, Mancabelli L, Lugli GA, Duranti S, Turroni F, Ferrario C, Mangifesta M, Viappiani A, Ferretti P, Gorfer V, et al. Exploring Vertical Transmission of Bifidobacteria from Mother to Child. *Appl Environ Microbiol*. 2015;81:7078–87.
- Ferretti P, Pasolli E, Tett A, Asnicar F, Gorfer V, Fedi S, Armanini F, Truong DT, Manara S, Zolfo M, et al. Mother-to-Infant Microbial Transmission from Different Body Sites Shapes the Developing Infant Gut Microbiome. *Cell Host Microbe*. 2018;24:133–145 e135.
- Chu DM, Ma J, Prince AL, Antony KM, Seferovic MD, Aagaard KM. Maturation of the infant microbiome community structure and function across multiple body sites and in relation to mode of delivery. *Nat Med*. 2017;23:314–26.
- Backhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, Li Y, Xia Y, Xie H, Zhong H, et al. Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. *Cell Host Microbe*. 2015;17:690–703.
- Lim ES, Wang D, Holtz LR. The Bacterial Microbiome and Virome Milestones of Infant Development. *Trends Microbiol*. 2016;24:801–10.
- Lim ES, Zhou Y, Zhao G, Bauer IK, Droit L, Ndao IM, Warner BB, Tarr PI, Wang D, Holtz LR. Early life dynamics of the human gut virome and bacterial microbiome in infants. *Nat Med*. 2015;21:1228–34.
- Duranti S, Lugli GA, Mancabelli L, Armanini F, Turroni F, James K, Ferretti P, Gorfer V, Ferrario C, Milani C, et al. Maternal inheritance of bifidobacterial communities and bifidophages in infants through vertical transmission. *Microbiome*. 2017;5:66.
- McCann A, Ryan FJ, Stockdale SR, Dalmasso M, Blake T, Ryan CA, Stanton C, Mills S, Ross PR, Hill C. Viromes of one year old infants reveal the impact of birth mode on microbiome diversity. *PeerJ*. 2018;6:e4694.
- Yatsunenkov T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP, et al. Human gut microbiome viewed across age and geography. *Nature*. 2012;486:222–7.
- Zhao G, Vatanen T, Droit L, Park A, Kostic AD, Poon TW, Vlamakis H, Siljander H, Harkonen T, Hamalainen AM, et al. Intestinal virome changes precede autoimmunity in type I diabetes-susceptible children. *Proc Natl Acad Sci U S A*. 2017;114:E6166–75.
- Drell T, Stsepetova J, Simm J, Rull K, Alekseeva A, Antson A, Tillmann V, Metsis M, Sepp E, Salumets A, Mandar R. The Influence of Different Maternal Microbial Communities on the Development of Infant Gut and Oral Microbiota. *Sci Rep*. 2017;7:9940.
- Avershina E, Lundgard K, Sekelja M, Dotterud C, Storro O, Oien T, Johnsen R, Rudi K. Transition from infant- to adult-like gut microbiota. *Environ Microbiol*. 2016;18:2226–36.
- Moreno-Gallego JL, Chou SP, Di Rienzi SC, Goodrich JK, Spector TD, Bell JT, Youngblut ND, Hewson I, Reyes A, Ley RE. Virome Diversity Correlates with Intestinal Microbiome Diversity in Adult Monozygotic Twins. *Cell Host Microbe*. 2019;25:261–272 e265.



28. Baumann-Dudenhoeffer AM, D'Souza AW, Tarr PI, Warner BB, Dantas G. Infant diet and maternal gestational weight gain predict early metabolic maturation of gut microbiomes. *Nat Med*. 2018;24:1822–9.
29. Gibson MK, Wang B, Ahmadi S, Burnham CA, Tarr PI, Warner BB, Dantas G. Developmental dynamics of the preterm infant gut microbiota and antibiotic resistome. *Nat Microbiol*. 2016;1:16024.
30. Planer JD, Peng Y, Kau AL, Blanton LV, Ndao IM, Tarr PI, Warner BB, Gordon JI. Development of the gut microbiota and mucosal IgA responses in twins and gnotobiotic mice. *Nature*. 2016;534:263–6.
31. Moore AM, Ahmadi S, Patel S, Gibson MK, Wang B, Ndao MI, Deych E, Shannon W, Tarr PI, Warner BB, Dantas G. Gut resistome development in healthy twin pairs in the first year of life. *Microbiome*. 2015;3:27.
32. Gurnee EA, Ndao IM, Johnson JR, Johnston BD, Gonzalez MD, Burnham CA, Hall-Moore CM, McGhee JE, Mellmann A, Warner BB, Tarr PI. Gut Colonization of Healthy Children and Their Mothers With Pathogenic Ciprofloxacin-Resistant *Escherichia coli*. *J Infect Dis*. 2015;212:1862–8.
33. Gurnee EA, Ndao IM, McGhee JE, Warner BB, Tarr PI, Fritz SA, Burnham CA. Fecal carriage of methicillin-resistant *Staphylococcus aureus* and vancomycin-resistant *Enterococcus* in healthy children. *Antimicrob Agents Chemother*. 2014;58:1261–2.
34. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010;7:335–6.
35. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods*. 2016;13:581–3.
36. Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, Brown CT, Porras-Alfaro A, Kuske CR, Tiedje JM. Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res*. 2014;42:D633–42.
37. Bodenhofer U, Bonatesta E, Horejs-Kainrath C, Hochreiter S: msa: an R package for multiple sequence alignment. *Bioinformatics*. 2015;31:3997–9.
38. Schliep KP. phangorn: phylogenetic analysis in R. *Bioinformatics*. 2011;27:592–3.
39. McMurdie PJ, Holmes S: phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*. 2013;8:e61217.
40. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome*. 2018;6(1):226.
41. Kircher M, Heyn P, Kelso J. Addressing challenges in the production and analysis of illumina sequencing data. *BMC Genomics*. 2011;12:382.
42. Felix MA, Ashe A, Piffaretti J, Wu G, Nuez I, Belicard T, Jiang Y, Zhao G, Franz CJ, Goldstein LD, et al. Natural and experimental infection of *Caenorhabditis nematodes* by novel viruses related to nodaviruses. *PLoS Biol*. 2011;9:e1000586.
43. Zhao G, Wu G, Lim ES, Droit L, Krishnamurthy S, Barouch DH, Virgin HW, Wang D. VirusSeeker, a computational pipeline for virus discovery and virome composition analysis. *Virology*. 2017;503:21–30.
44. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlenn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H. vegan: Community Ecology Package. R package version 2.5-2; 2018.
45. Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV, Reyes JA, Shah SA, LeLeiko N, Snapper SB, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol*. 2012;13:R79.
46. Liaw A, Wiener M. Classification and Regression by randomForest. *R News*. 2002;2:18–22.
47. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Muller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12:77.
48. Peng Y, Leung HC, Yiu SM, Chin FY. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*. 2012;28:1420–8.
49. Sommer DD, Delcher AL, Salzberg SL, Pop M. Minimus: a fast, lightweight genome assembler. *BMC Bioinformatics*. 2007;8:64.
50. Roux S, Enault F, Hurwitz BL, Sullivan MB. VirSorter: mining viral signal from microbial genomic data. *PeerJ*. 2015;3:e985.
51. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*. 2013;1303:3397.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://www.biomedcentral.com/submissions)

