Studying vertical microbiome transmission from mothers to infants by strain-level metagenomic profiling

Francesco Asnicar*,[1], Serena Manara*,[1], Moreno Zolfo[1], Duy Tin Truong[1], Matthias Scholz[1], Federica Armanini[1], Pamela Ferretti[1], Valentina Gorfer[2], Anna Pedrotti[2], Adrian Tett[1,#], Nicola Segata[1,#]

* Equal contribution
# Corresponding authors: adrianjames.tett@unitn.it, nicola.segata@unitn.it

[1] Centre for Integrative Biology, University of Trento, Italy
[2] Azienda Provinciale per i Servizi Sanitari, Trento, Italy

Abstract

The gut microbiome starts to be shaped in the first days of life and continues to increase its diversity during the first months. Several investigations are assessing the link between the configuration of the infant gut microbiome and infant health, but a comprehensive strain-level assessment of vertically transmitted microbes from mother to infant is still missing. We longitudinally collected fecal and breast milk samples from multiple mother-infant pairs during the first year of life, and applied shotgun metagenomic sequencing followed by strain-level profiling. We observed several specific strains including those from *Bifidobacterium bifidum*, *Coprococcus comes*, and *Ruminococcus bromii*, that were present in samples from the same mother-infant pair, while being clearly distinct from those carried by other pairs, which is indicative of vertical transmission. We further applied metatranscriptomics to study the *in vivo* expression of vertically transmitted microbes, for example *Bacteroides vulgatus* and *Bifidobacterium* spp., thus suggesting that transmitted strains are functionally active in the two rather different environments of the adult and infant guts. By combining longitudinal microbiome sampling and newly developed computational tools for strain-level microbiome analysis, we showed that it is possible to track vertical transmission of members of the microbiome from mother to infants and characterize their transcriptional activity. Our work poses the basis for surveying at larger scale the sources of microbial diversity in the infants and starts associating these transmissions with the subsequent longer-term development of a healthy or dysbiotic microbiome.

Importance

Early infant exposure is important in the acquisition and ultimate development of a healthy infant microbiome. There is increasing support that the maternal microbial reservoir is a key route of microbial transmission, yet much is inferred from the observation of shared species in mother and infant. Common species, *per se*, does not necessarily equate vertical transmission as species exhibit considerable strain heterogeneity and it is therefore imperative to identify shared strains. We demonstrate here the potential of shotgun metagenomics and strain-level resolution to identify vertical transmission events via the maternal source. Combined with a metatranscriptomic

approach, we show the potential not only to identify and track the fate of microbes in the early infant microbiome but also identify the metabolically active members. These approaches will ultimately provide important insights into the acquisition, development and community dynamics of the infant microbiome.

Introduction

The community of microorganisms that dwell in the human gut has been shown to play an integral role in human health (1-3), for instance by harvesting nutrients that would be inaccessible otherwise (4), modulating host metabolism and immune system (5) and preventing infections by occupying the ecological niches that could otherwise be exploited by pathogens (6). The essential role of the intestinal microbiome is probably best exemplified by the successful treatment of dysbiotic states such as chronic life threatening *Clostridium difficile* infections using microbiome transplantation therapies (7-9).

The gut microbiome is a dynamic ecological community which is shaped by multiple factors throughout an individual's life. The early development of the infant microbiome has been shown to be particularly crucial for future health (10-12) and a few studies have investigated the factors that are important in defining its first structure (13-16). In particular, gestational age at birth (16), mode of delivery (13, 14), early antibiotic treatments (17), have all been shown to influence the gut microbial composition in the short term, and the pace of its development in the longer term.

Vertical transmission of bacteria from the body and breast milk of the mother to her infant has gained attention as an important source of microbial colonization (13, 18-20). The extent to which the mother directly contributes to the infant's gut microbial diversity is likely more crucial than microbial transmission from exposure to the wider environment (21, 22), including the delivery room (23). Early cultivation-based and cultivation-free methods (16S rRNA community profiling and a single metagenomic study) have suggested that the mother could transfer microbes to the infant by breastfeeding (24), and that the route of delivery has the potential of seeding the infant gut with members of the mother's vaginal community (10, 13, 25, 26).

Studies of the vertical transmission of microbes from mothers to infants have however hitherto focused on the cultivable fraction of the community (27) or lacked strain-level resolution (10). At a species level resolution, where the same microbial species have been found to be shared between mothers and their infants (12, 28), it still remains inconclusive if vertical transmission has actually occurred as the same species are also observed in unrelated mothers and infants. Because different individuals will often carry different strains of common species (29, 30), it is crucial to profile bacteria at a finer strain-level to ascertain the route of transfer. This has been performed only for specific microbes by cultivation methods (15, 27), but many vertically transmitted microorganisms remain hard to cultivate (15), thus the true extent of microbial transmission remains unknown. A further crucial aspect still largely unexplored is the fate of vertically acquired strains i.e. are they merely transient or functionally active and therefore more likely to be early colonizers of the infant intestine. Although studies have reported the

transcriptional activity of intestinal microbes under different conditions (31-34), no studies have applied metatranscriptomics to characterize the activity of vertically transmitted microbes *in vivo*.

In this work, we present and validate a shotgun metagenomic pipeline to track mother-to-infant vertical transmission of microbes by applying the strain-level computational profiling to all the members of the infant microbiome. Moreover, we assessed the transcriptional activity of vertically transmitted microbes to elucidate if transferred strains are not only present but also metabolically active in the infant gut.

Results and Discussion

We analyzed the vertical transmission of microbes from mother-to-infant by enrolling 5 mother-infant pairs and collecting fecal samples and breast milk (see **Methods**) at three months of age of the infant (Timepoint 1). Two mother-infant pairs were additionally sampled at 10 months (Timepoint 2) and one pair at 16 months post birth (Timepoint 3). We applied shotgun metagenomic sequencing to 24 microbiome samples (8 mother fecal samples, 8 infant fecal samples, and 8 milk samples) generating 1.2G reads (39.6M s.d. 28.7M reads/sample, see **Supplementary Table S1**). Metatranscriptomics (90.55M s.d. 46.86M reads/sample) was also applied on fecal samples of two pairs to investigate the differential expression profiles of the bacterial strains in the gut of mothers and their infants.

Shared mother-infant microbial species

In our cohort, the infant intestinal microbiome is dominated by *Escherichia coli* and *Bifidobacterium* spp., such as *B. longum*, *B. breve* and *B. bifidum* (**Fig. 1A**), which is consistent with what has been described previously (11, 35, 36). As expected, mothers' intestines have a more diverse microbiome with high abundances of *Prevotella copri*, *Clostridiales* (e.g. *Coprococcus* spp. and *Faecalibacterium prausnitzii*), and *Bacteroidales* (e.g. *Parabacteroides merdae* and *Alistipes putredinis*). Interestingly, the post-weaning infant microbiome of Pair 5 (Timepoint 3, 16 months post birth), had already shifted towards a more "mother-like" composition (**Fig 1B**), with an increase in diversity and the appearance of *Parabacteroides merdae*, *Coprococcus* spp., and *Faecalibacterium prausnitzii* (12, 36). Nevertheless, this 16-months old infant still retains some infant microbiome signatures, such as a high abundance of *Bifidobacteria* that are present only at low levels in mothers' samples (**Fig. 1A** and **1C**).

We extracted and successfully sequenced microbial DNA from 7 out of 8 milk samples. Profiling of the microbial composition of milk is hindered by the high abundance of interfering molecules (proteins, fats, proteases - e.g. plasmin - and calcium ions) (37-39) that affect the efficiency of extraction and amplification steps. Even so, we obtained an average of 3.08 s.d. 1.5 Gb (100 bp reads) per sample. In each sample we extracted 26 s.d. 56 Mb of non-human reads (higher than the only other metagenomic study (40), see **Supplementary Table S1**).

Milk samples had a limited microbial diversity at the first sampling time (Timepoint 1, three months post birth), and included skin-associated bacteria like *Corynebacterium kroppenstedtii* and *Staphylococcus epidermidis*. Cutaneous taxa, however, were only observed in low abundancies

in the gut microbiome of infants, confirming that skin microbes are not colonizers of the human gut (**Fig. 1A**). At later timepoints the milk samples were enriched in *B. breve* and bacteria usually found in the oral cavity, such as *Streptococcus* and *Veillonella* spp. The presence of oral taxa in milk has already been observed by previous studies based on 16S rRNA sequencing (13, 19, 20, 24) and shotgun metagenomics (40). This could be caused by the retrograde flux into the mammary gland during breastfeeding (41) whereby cutaneous microbes of the breast and from the infant oral cavity are transmitted to the breast glands (42). These observations are summarized in the ordination analysis (**Fig. 1B**), in which the different microbiome samples (infant feces, mother feces, and milk) clustered by type, with weaning representing a key factor in the shift from an infant to an adult-like microbiome structure (12, 36, 43).

Comparing the species present both in the mother and infant pairs (**Fig. 1C**), we observed that many shared species occur at much higher abundance in the infant than in the mother (e.g.
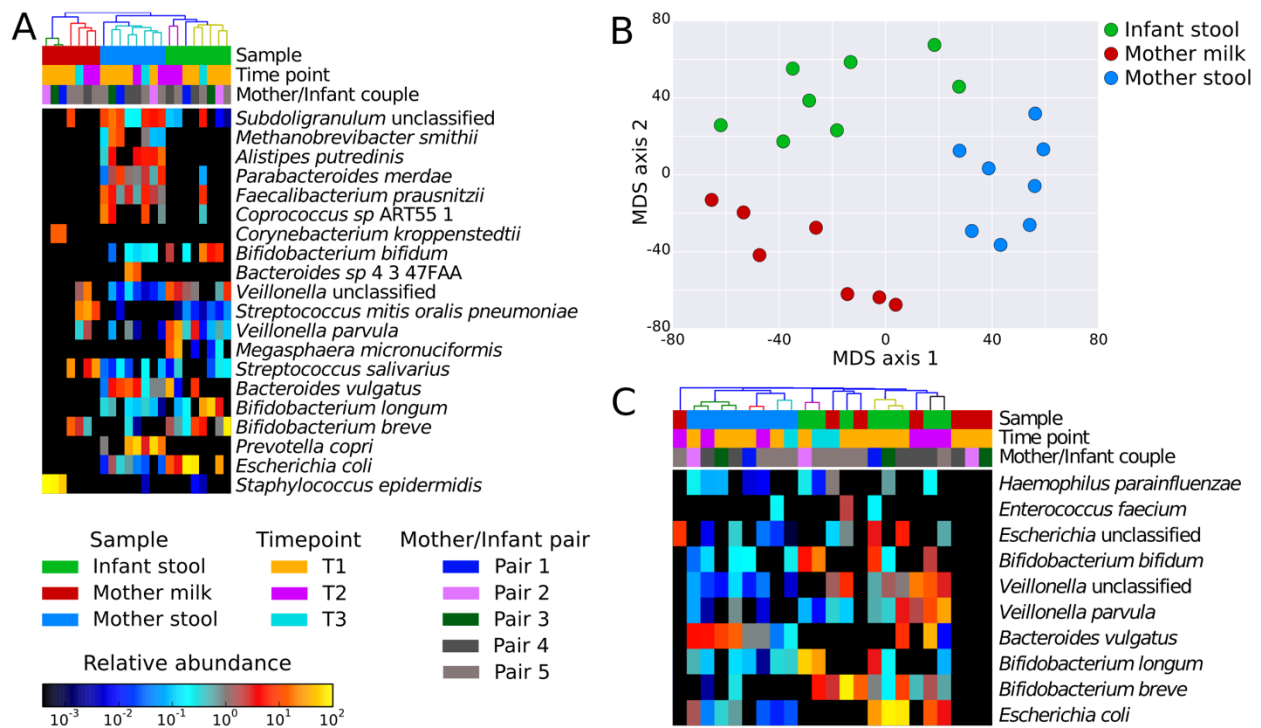

Figure 1. Microbial composition of mother and infant samples, and shared bacteria within mother-infant pairs. (A) Quantitative microbial taxonomic composition of the metagenomic samples from milk and fecal samples of mothers and infants as estimated by MetaPhlAn2 analysis (45) (only the 20 most abundant species are indicated). As can be seen milk samples present a low microbial richness compared to fecal samples. **(B)** Ordination plot of microbiome composition showing clustering of the three different sample types: mother feces, infant feces, and breast milk samples. The two infant samples close to the cluster of mother feces and in between the clusters of mothers and infants, are from later timepoints, denoting the convergence of the infant microbiome towards an adult-like one. **(C)** The abundances of the 10 microbial species detected (>0.1% abundance) in at least one infant and their respective mother (shared species have been identified based on samples from Timepoint 1 only).

Escherichia, *Bifidobacterium* and *Veillonella* spp.) possibly due to the lower level of species diversity and therefore competition in the gut. *Bacteroides vulgatus* was instead found at relatively high abundances in both the infant and the mother of Pair 4 at both Timepoint 1 and Timepoint 2. The presence of shared species in infant-mother pairs seen here and elsewhere (13, 15, 16, 24, 44) confirms that mothers are a potential reservoir of microbes vertically transmissible to infants, but it remains unproven whether the same strain is transmitted to the infant from the mother or if a more complex transmission route is involved.

Strains shared between mother and infants are indicative of vertical transmission

At a species level resolution, it has been shown that different individuals have a core of microbial species, but these common species consist of distinct strains (29, 30). To analyze microbial transmission from mother to infant it is therefore crucial to assess whether they carry the same microbial strain for that species. To this end, we further analyzed the metagenomic samples at a finer strain level resolution. This was achieved by applying a recent strain-specific pangenome-based method called PanPhlAn (29) as well as a novel method, called StrainPhlAn (see **Methods**), that identifies single-nucleotide variants (SNVs) in species-specific marker genes.

Using this SNVs-based analysis, we first confirmed considerable strain-level heterogeneity in the species present in multiple mothers also with respect to available reference genomes (**Fig. 2** and **Fig. S2**). This heterogeneity is not observed between mother-infant pairing, as in the case of *Bifidobacterium* spp., *Ruminococcus bromii*, and *Coprococcus comes*. The infant of Pair 4 at Timepoint 2 harbors, for example, a strain of *B. bifidum* which matches the one in his mother at 99.96% sequence identity and yet at the same time is clearly distinct from the *B. bifidum* strains harbored by other infants that differ by at least 0.64% nucleotides (the average nucleotide diversity with the other strains in other pairs is 0.74%, **Fig. 2A**). The same is true for the strains of *C. comes* (99.87% intra-pair similarity, 0.87% and 1.63% divergence compared to the closest strain and average, respectively, **Fig. 2B**) and *R. bromii* (99.93% similar and 1.53% and 2.63% diverse – same as above, **Fig. 2C**) that are shared by Pair 5. The mother-infant sharing of the same strain was confirmed by strain-level pangenome analysis (29) that showed that the strains from the same pair carry the same unique gene repertoire (**Supplementary Fig. S3**). While it is not possible to exclude the independent acquisition of strains from a shared environmental source, the finding that mother-infant pairs harbor some shared strains represents strong evidence of vertical microbiome transmission.

The mother-infant Pair 5 at Timepoint 3 harbors the highest number of shared species with 35.2% of the species present in the infant microbiome at a relative abundance >0.1% that are also present in the mother according to MetaPhlAn2 profiles. Among the 35.2% of shared species, the mother and infant of Pair 5 shared several strains (**Fig. 2** and **Supplementary Fig. S2**), according to the results of both PanPhlAn and StrainPhlAn (see for example *B. adolescentis* and *C. comes*, **Fig. 2** and **Supplementary Fig. S2** and **S3**). However, some strains that were shared within mother-infant Pair 5 at earlier timepoints were replaced by different ones at Timepoint 3. The *R. bromii* strain harbored by the infant at Timepoint 3 is in fact different from the one found at Timepoint 2 and in the mother at both timepoints (**Fig. 2C**) and a similar event is observed for the

latter infant timepoint for *B. breve* (**Supplementary Fig. S2B**) and *B. longum* (**Supplementary Fig. S2C**). Although it is not possible to generalize these results because of the small sample size, these replacement events suggest that the originally acquired maternal strains can later be substituted with different ones from the same species (46, 47).

We then extended our analysis to the viral organisms detectable from metagenomes and metatranscriptomes, as viruses also can be potentially vertical transmitted. The DNA viruses identified from our metagenome samples largely consisted of bacteriophages from the *Caudovirales* order, a common order found in the intestine as previously reported (3, 48), but the low breadth of coverage for many of them made it difficult to identify pair-specific phage variants (**Supplementary Table ST3**). When looking at RNA viruses from metatranscriptomics, we
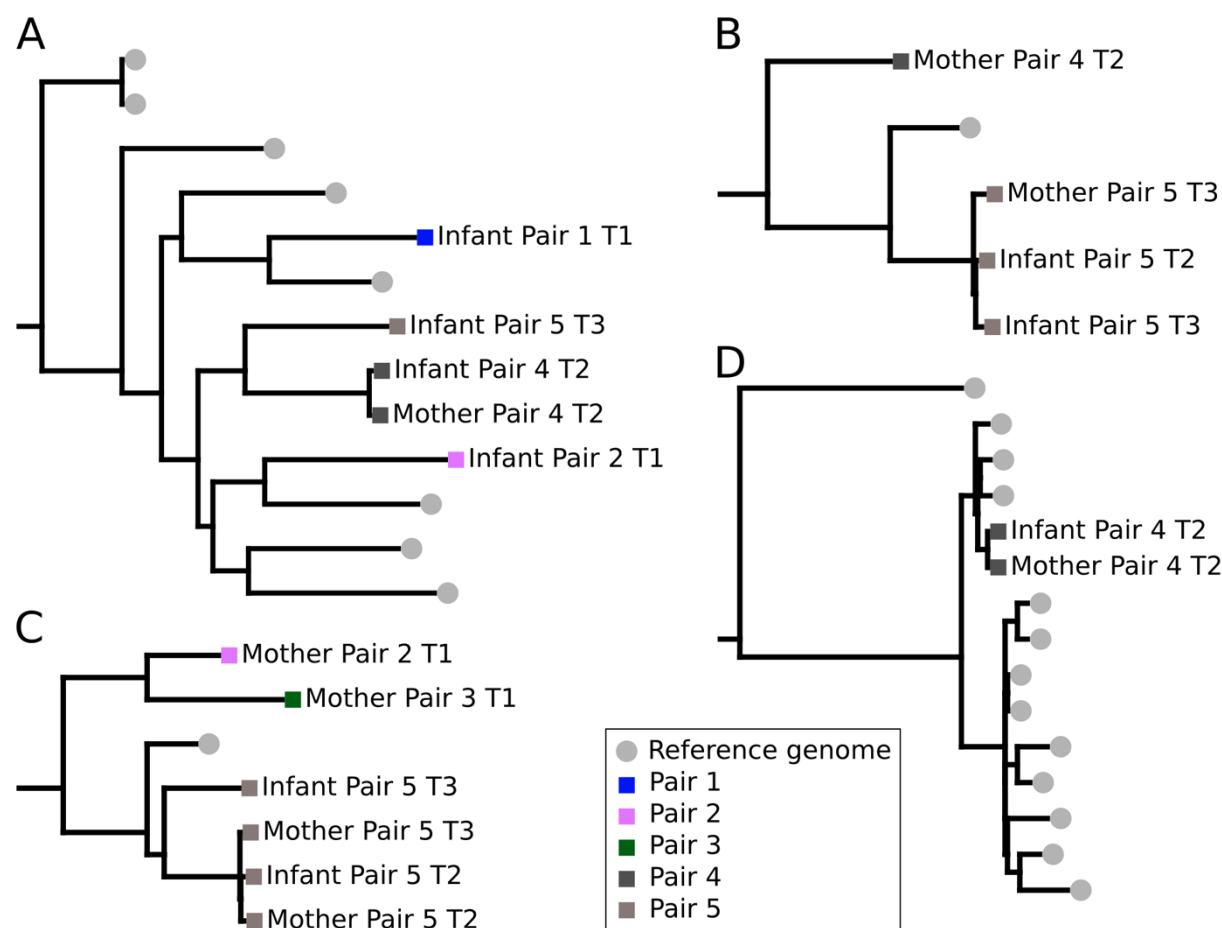

Figure 2. Strain-level phylogenetic trees for microbes present in both the mother and infant. Phylogenetic trees built by using species-specific markers (StrainPhlAn, see **Methods**) confirming the presence of the same strain in the mother and infant intestinal microbiome, thus suggesting vertical transmission. Available reference genomes were included in the phylogenetic trees. Here we report three bacterial species, namely **(A)** *Bifidobacterium bifidum*, **(B)** *Coprococcus comes*, and **(C)** *Ruminococcus bromii*, and the most abundant viral species found in Pair 4 **(D)** Pepper mild mottle virus. Other species-specific phylogenetic trees (*B. adolescentis*, *B. breve*, and *B. longum*) are reported in the **Supplementary Figure S2**.

detected the Pepper Mild Mottle Virus (PMMoV), a single-stranded positive-sense RNA virus of the genus *Tobamovirus*, in all of the four metatranscriptomes from pairs 4 and 5. Surprisingly, transcripts from the PMMoV where in greater abundance than all the other microbial transcripts found for the mother of Pair 4. PMMoV has already been reported in the gut microbiome (49-51), and other viruses of the same family have been shown to be able to enter and persist in eukaryotic cells (52, 53). The high abundance of PMMoV in mother-infant Pair 4 allowed us to reconstruct its full genome (99.9%) and to perform a phylogenetic analysis showing that both the mother and the infant share a PMMoV with the very same genomic sequence, which is highly similar but clearly distinct from the PMMoV reference genomes (27 SNVs in total, **Fig. 2D**). Although the coverage was lower, the same evidence of a shared PMMoV was observed also within Pair 5. The analysis of PMMoV polymorphisms within each sample also suggests the coexistence of different PMMoV haplotypes (**Supplementary Figure S6**) in the same host. Although vertical transmission of RNA viruses and PMMoV specifically would be intriguing, because of the age and dietary habits of the infants (**Supplementary Table ST1**) this finding could be related to the exposure to a common food source (54).

Differences in the overall functional potential and expression in mothers and infant

The physiology of the mammary gland (milk) as well as the adult and infant intestine is reflected by niche specific microbial communities. We complemented the taxonomic analysis above by employing HUMAnN2 (see **Methods**) to characterize the overall functional potential of these inhabiting communities. As expected, there was considerably overlap in the functionality of the mothers' and infants' gut microbiomes when looking at the relative abundance of microbial pathways (**Fig. 3A**). Nevertheless, there were notable differences, for instance the infants' microbiome showed a higher potential for utilization of intestinal mucin as a carbon source and for folate biosynthesis, (11, 55, 56) while displaying a lower potential for starch degradation, in line with what had been observed previously (11, 55-58). Mucin utilization specifically by infant-gut microbial communities is reflective of the higher abundance of mucin degrading *Bifidobacteria* observed from the taxonomic analyses (11, 55-57), whereas increased folate biosynthesis (11, 55, 56, 58) and decreased starch degradation (4) have been suggested to be a response to the limited dietary intake in infants with respect to adults. Interestingly, the post-weaning infant of Pair 5 (16 months post birth) clustered together with the adults' intestinal samples (**Fig. 3B**), suggesting that the shift towards an adult-like microbiome observed in the taxonomic profiling (**Fig. 1B**) is also reflected by or is as a consequence of a change in community functioning. Among the most prevalent pathways in the milk microbiomes we observed those involved in galactose and lactose degradation (59) as well as in aromatic compounds biosynthesis (**Supplementary Fig. S4**). This is specifically true for chorismate production, a key intermediate for the biosynthesis of essential amino acids and vitamins found in milk (58) (**Fig. 3C** and **Supplementary Fig. S4**).

To further evaluate the functional capacity of the gut associated microbiomes and analyze the *in vivo* transcription, we performed metatranscriptomics on the feces of two mother-infant pairs (see **Methods**). HUMAnN2 was used to identify differences in the transcriptional levels of pathways in mothers and infants gut. The most notable global difference was that fermentation pathways were

highly transcribed in the mother metatranscriptome compared to infant. This reflects the transition of the gut from aerobic to anaerobic and the associated shift from facultative anaerobes to obligate anaerobes over the first few months of life (60, 61). The same is true for pathways involved in starch degradation, which are not only poorly represented but also negligibly expressed in the infants' microbiomes (59). What is evident is that the transcriptional patterns for different members vary considerably, as illustrated for Pair 4 and 5 (**Fig. 4A**, and **Supplementary Fig. S5A** respectively). For example, in Pair 4 we observe in the infant that *B. vulgatus* is more transcriptionally active (average of 2.7 s.d. 2.5 normalized transcript abundance - NTA, see **Methods**) than both *E. coli* (average 0.4 s.d. 0.6 NTA) and *Bifidobacteria* spp. (average 0.01 s.d. 0.01 NTA), although whether these differences are physiologically significant remains less clear.
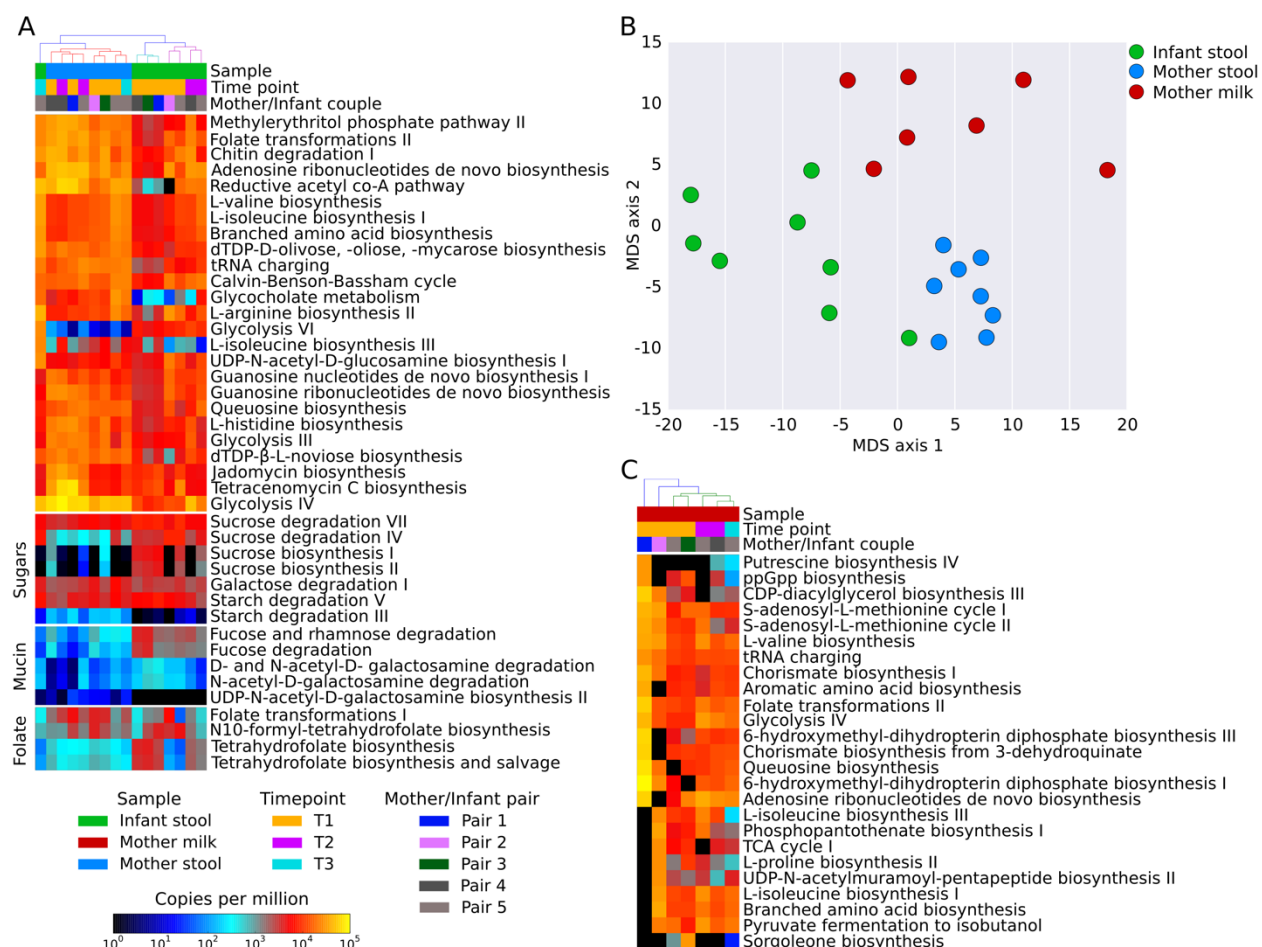

Figure 3. Functional potential analyses. (A) HUMAnN2 heatmap reporting the 25 most abundant pathways in the fecal samples of mothers and infants. Added at the bottom, specific pathways of interest: starch, mucin, and folate metabolism. **(B)** MDS result on functional potential, showing the differences between fecal samples of mothers and infants, and milk. In particular, the infant-feces point in the mother-feces cluster corresponds to the Timepoint 3 of the Pair 5, showing a shift from the infant microbiome towards and adult-like microbiome. **(C)** HUMAnN2 results on the 25 most abundant pathways found only in the milk samples.

Strain-specific transcriptional differences in mothers and infants

To further explore the transcriptional activity of the intestinal microbiomes and more specifically to ascertain which individual microbial members are metabolically active in the gut, we employed the strain-specific metatranscriptomic approach implemented in PanPhlAn (29) (see **Methods**). Of particular interest is the transcriptional activity of the shared mother-infant strains that based on our strain-level analyses are likely to have been vertically acquired by the infant from a maternal route. Such transcriptional analyses can clarify whether these transmitted strains are not only present but also functioning in the infant gut suggesting that the transmitted strains are likely colonizing. For three transmitted species in Pair 4 (*B. vulgatus, E. coli* and *B. bifidum*) we show that they are both active in the mother and infant intestine (**Fig. 4B**). Of note is *B. bifidum* that is more active in the infant with respect to the mother (**Fig. 4B**), which would be expected as
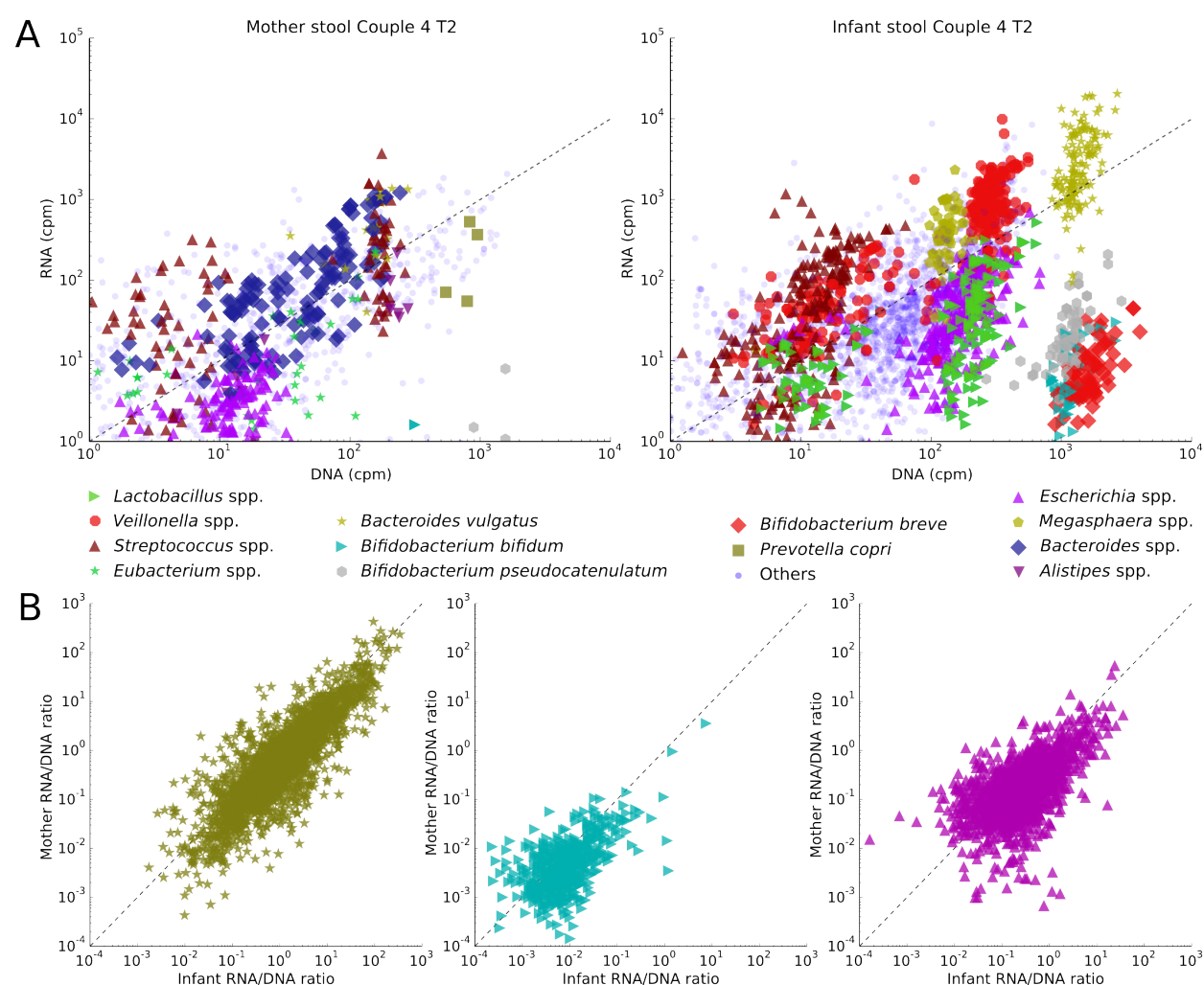

Figure 4. Metabolic pathways expression in mother and infant of Pair 4 at Timepoint 2. (A) Scatterplots showing the transcription rate of metabolic pathways of different species and genera of interest for both the mother and infant of Pair 4 at Timepoint 2. **(B)** Comparison between gene families' transcription rate in mother and infant gut microbiomes.

this species is a known early colonizer of the infant gut (11, 35, 36). Interestingly, the *B. bifidum* strain of Pair 5 shows the opposite behavior (Supplementary Fig. S5B). We postulate that this is because the infant of Pair 5 was post weaning (10% breast milk diet) compared to the pair 4 (90% breast milk diet) and reflects the change in substrate availability from breast milk to solid food, which might have a detrimental effect on the bifidobacterial population (36, 62, 63). Moreover, in support of our metagenomics analyses that the microbiome of infant of Pair 5 is shifting towards a more adult like structure (**Fig. 1B**), we observe high transcriptional activity for *R. bromii,* a commonly adult associated species, which could be seen as hallmark of this transition (64, 65).

It is well established that metatranscriptomic profiling provides a more accurate account of the actual community functioning than metagenomics alone. Here we show that the combination of both approaches affords the exploration of which members are not only transmitted but are also actively participating in the community, and therefore a more detailed account of the microbial community dynamics.

Conclusions

Human associated microbiomes are complex and dynamic communities that are continuously interacting with each other, their host, as well as under influence from environmental sources of microbial diversity. Identifying and understanding the transmission from these external sources is crucial to understand, in particular, how the infant gut microbiome is colonized and ultimately develops into an adult like composition. However, detecting direct transmission is not a trivial task: many species are ubiquitous in host-associated and in the wider environment alike, yet they actually comprise a myriad of different strains and phenotypic capabilities. Therefore, detection of microbial transmission events requires the ability to characterize microbes at the strain level. The epidemiological tracking of pathogens from cultivation-based isolate sequencing has proven successful (66, 67) but it relies on time-consuming protocols and can focus only on a limited number of species. In contrast, while there have been some examples of strain level tracking from metagenomic data (67, 68), this still remains challenging. In this study, we developed methods for identifying the vertical flow of microorganisms from mothers to their infants and show that maternal microbial sources are important in the development of the infant gut microbiome.

We demonstrate that high-resolution computational methods applied to shotgun metagenomic and metatranscriptomic data enable the tracking of strains and the following of strain-specific transcriptional patterns across mother-infant pairs. In our cohort of five mother-infant pairs, we detected several species with substantial genetic diversity between different pairs but identical genetic profiles in the mother and her infant, indicative of vertical transmission. These include some *Bifidobacteria* typical of the infant gut (i.e. *B. longum*, *B. breve*, *B. bifidum*, and *B. adolescentis*) but also *Clostridiales* usually found in the adult intestine (i.e. *R. bromii* and *C. comes*) and viral organisms. These results thus confirm that the infant receives a maternal microbial imprinting that is likely crucial for the development of the gut microbiome in the first years of life.

The strain-level investigation of vertical transmitted microbes was followed by transcriptional activity characterization of the transmitted strains in the mother and infant environments. We found that the transcriptional activity of shared strains within mother-infant pairs proved to be different in the two hosts, suggesting that successful colonization of maternally transmitted microbes.

Altogether, our work provides the preliminary results and methodology to expand our knowledge on how microbial strains are transmitted across microbiomes. Expanding the cohort size and considering other potential microbial sources of transmission, such as additional mother and infant body sites, as well as other family members (i.e. fathers and siblings) and environment (hospital and house surfaces), will likely shed light on the key determinants in early infant exposure and the seeding and development of the infant gut microbiome.

Materials and methods

Sample collection and storage

In total five mother-infant pairs were enrolled. Fecal samples and breast milk were collected for all pairs at three months (Timepoint 1), additional samples were collected for Pair 4 and 5 at 10 months (Timepoint 2) and for Pair 5 only at 16 months (Timepoint 3) (**Supplementary Table SX**). All aspects of recruitment, sample and data processing was approved by the local ethics committee. Fecal samples were collected in sterile Faeces tubes (Sarstedt, Nümbrecht, Germany) and immediately stored at -20°C. In those cases, where metatranscriptomics was applied, a fecal aliquot was removed prior to freezing the remaining feces. This aliquot was stored at 4°C and RNA extracted within two hours of sampling to preserve RNA integrity. Milk was expressed and collected mid flow by mothers into 15 ml centrifuge tubes (VWR, Milan, Italy) and immediately stored at -20°C. Within 48 hours of collection all milk/fecal samples were moved -80°C storage until processed.

Nucleic acids extraction for metagenomic analysis

DNA was extracted from feces using the QIAamp DNA Stool Mini Kit (Qiagen, Netherlands), Milk DNA was extracted using the PowerFood Microbial DNA Isolation Kit (MO BIO, California, USA), both according to manufacturer's specifications. Extracted DNA was purified using Agencourt AMPure XP kit (Beckman Coulter, California, USA). Metagenomic libraries where constructed using the Nextera XT DNA Library Preparation Kit (Illumina, California, USA) according to manufacturer instructions, and sequenced on the HiSeq 2500 platform (Illumina, California, USA) at an expected sequencing depth of 6 Gb/library.

Nucleic acids extraction for metatranscriptomic analysis

Fecal samples for metatranscriptomic profiling were pretreated as described previously (69). Briefly, 110 µl of lysis buffer (30 mM Tris·Cl, 1 mM EDTA, pH 8.0, 1.5 mg/ml of proteinase K and 15 mg/ml of lysozyme) was added to 100 mg of feces and incubated at room temperature for 10

minutes. After pretreatment, samples were treated with 1200 µl of QIAGEN RLT Plus buffer (from AllPrep DNA/RNA Mini Kit, QIAGEN, Netherlands) containing 1% volume beta-mercaptoethanol, and transferred into 2 ml sterile screw-cap tubes (Starstedt, Germany) filled with 1 ml of zirconia-silica beads <0.1 mm (BioSpec Products, Oklahoma, USA). Tubes were placed on a Vortex-Genie 2 with Vortex Adapter 13000-V1-24 (MO BIO, California, USA) and shaken at maximum speed for 15 minutes. Lysed fecal samples were homogenized using QIAshredder spin columns (QIAGEN, Netherlands) and homogenized sample lysates were then extracted with the AllPrep DNA/RNA Mini Kit (QIAGEN, Netherlands) according to manufacturer's specifications. Extracted RNA and DNA were purified using Agencourt RNAClean XP and Agencourt AMPure XP (Beckman Coulter, California, USA) kits respectively. Total RNA samples were subjected to rRNA depletion and metatranscriptomic libraries were prepared using the ScriptSeq Complete Gold Kit (Epidemiology) – Low Input (Illumina, California, USA). Metagenomic libraries were prepared with Nextera XT DNA Library Preparation Kit (Illumina, California, USA). All libraries were sequenced on the HiSeq 2500 platform (Illumina, California, USA) at an expected depth of 6 Gb/library.

Sequencing data pre-processing

The metagenomes and metatranscriptomes were pre-processed by removing low quality reads (mean quality less than 25), trimming low quality positions (quality less than 15), and removing reads shorter than 90 nucleotides using FastqMcf (70). Further quality control steps involved the removal human reads and the reads from the Illumina spike-in (bacteriophage phiX174), by mapping the reads against the corresponding genomes with Bowtie2 (71). Metatranscriptomes were additionally processed to remove ribosomal RNA, by mapping the reads against 16S and 23S rRNA gene databases (SILVA_119.1_SSURef_Nr99_tax_silva and SILVA_119_LSURef_tax_silva (72)), and to remove contaminant adapters using trim_galore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) with parameters: -q 0, --nextera, and --stringency 5. The milk sample of mother-infant Pair 4 at Timepoint 1 was discarded from further analyses because of the low number of microbial reads (less than 400,000 bp) obtained after the quality control (**Supplementary Table S1**). All metagenomes and metatranscriptomes have been deposited and are available at the NCBI Sequence Read Archive under the BioProject accession number PRJNA339914.

Taxonomic and strain-level analysis

Taxonomic profiling was performed with MetaPhlAn2 (45) (with default parameters) on all the 24 metagenomic samples sequenced. MetaPhlAn2 uses clade-specific markers for taxonomically profiling shotgun metagenomic data and quantify the clades present in the microbiome with specie-level resolution.

Strain-level profiling was performed with PanPhlAn (29) and a novel strain-level method called StrainPhlAn. PanPhlAn is a pangenome-based approach that profiles the presence/absence pattern of species-specific genes in the metagenomes. The genes presence/absence profiles are then used to characterize the strain-specific gene repertoire of the members of the microbiome. PanPhlAn has been executed using the following parameters: --min_coverage 1, --left_max 1.70, and --right_min 0.30. StrainPhlAn is a novel complementary method based on SNVs analysis that

reconstructs the genomic sequence of species-specific markers. StrainPhlAn builds the strain-level phylogeny of microbial species by reconstructing the consensus marker sequences of the dominant strain for each detected species. The extracted consensus sequences are multiple-aligned using MUSCLE version v3.8.1551 (73) (default parameters) and the phylogeny is reconstructed using RAxML version 8.1.15 (74) (parameters, -m GTRCAT and -p 1234). StrainPhlAn is available with supporting documentation at http://segatalab.cibio.unitn.it/tools/strainphlan.

Functional profiling from metagenomes and metatranscriptomes

The functional potential and transcriptomic analysis was performed with both HUMAnN2 (75) and PanPhlAn (29). HUMAnN2 selects the most representative species from a metagenome and then builds a custom database of pathways and genes that is used as a mapping reference for the coupled metatranscriptomic sample to quantify transcript abundances. We computed the normalized transcript abundance (NTA), that we define as the average coverage of a genomic region in the metatranscriptomic versus the corresponding metagenomic sample normalized by the total number of reads in each sample. PanPhlAn infers the expression of the strain-specific gene families by extracting them from the metagenome and matching them in the metatranscriptome. PanPhlAn has been executed using the following parameters: --rna_norm_percentile 90 and --rna_max_zeros 90.

Profiling of DNA and RNA viruses

We investigated the presence of viral and phage genomes by mapping the reads present in the metagenomes and metatranscriptomes against 7,194 viral genomes available in RefSeq (release 77). The average coverage and sequencing depth were computed with samtools (76) and bedtools (77).

The presence of the Pepper Mild Mottle Virus (PMMoV) was confirmed by mapping the reference genome (NC_003630) against the metatranscriptomic samples from mother-infant of Pair 4 and Pair 5. In the mother and infant of Pair 4, 424,510 and 119 reads were mapped respectively, while in the mother and infant of Pair 5, 1,444 and 61 of the reads were mapped respectively. In the two mothers (Pair 4 and 5) the breadth of coverage is 0.99 and 0.98, and the average coverage is 6,562 and 22, respectively. In the two infants (Pairs 4 and 5), the breadth of coverage is of 0.6 and 0.5, and the average coverage is of 1.81 and 0.95, respectively. Additionally, we extracted the shared fractions of the PMMoV genome present in both mother and infant of Pair 4, together with the same regions of all the available reference genomes (n=13 and specifically accessions LC082100.1, KJ631123.1, AB550911.1, AY859497.1, KU312319.1, KP345899.1, NC_003630.1, M81413.1, KR108207.1, KR108206.1, AB276030.1, AB254821.1, LC082099.1).The resulting sequences were aligned using MUSCLE version v3.8.1551 (default parameters) and the resulting alignment was used to build a phylogenetic tree with RAxML v.8.1.15 (parameters, -m GTRCAT and -p 1234).

Statistical analyses and data visualization

The taxonomic and functional heatmaps were generated using hclust2 (parameters: --f_dist_f euclidean, --s_dist_f braycurtis, and -l) available at https://bitbucket.org/nsegata/hclust2. The Multi-Dimensional Scaling plots were computed with the sklearn Python package (78).

Biomarker discovery (**Supplementary Fig. S4)** was performed by applying LEfSe (79) (with parameter: -l 3.0) on HUMAnN2 profiles. The two functional trees (**Supplementary Fig. S4**) have been automatically annotated with export2graphlan.py (GraPhlAn package) and displayed with GraPhlAn (80) using default parameters.

Acknowledgements

Bibliography

1. **Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende DR, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan M, Batto JM, Hansen T, Le Paslier D, Linneberg A, Nielsen HB, Pelletier E, Renault P, Sicheritz-Ponten T, Turner K, Zhu H, Yu C, Li S, Jian M, Zhou Y, Li Y, Zhang X, Li S, Qin N, Yang H, Wang J, Brunak S, Dore J, Guarner F, Kristiansen K, Pedersen O, Parkhill J, Weissenbach J, et al.** 2010. A human gut microbial gene catalogue established by metagenomic sequencing. Nature **464:**59-65.

2. **Clemente JC, Ursell LK, Parfrey LW, Knight R.** 2012. The impact of the gut microbiota on human health: an integrative view. Cell **148:**1258-1270.

3. **Tamburini S, Shen N, Wu HC, Clemente JC.** 2016. The microbiome in early life: implications for health outcomes. Nat Med **22:**713-722.

4. **Backhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI.** 2005. Host-bacterial mutualism in the human intestine. Science **307:**1915-1920.

5. **Palm NW, de Zoete MR, Flavell RA.** 2015. Immune-microbiota interactions in health and disease. Clin Immunol **159:**122-127.

6. **Stecher B, Hardt WD.** 2011. Mechanisms controlling pathogen colonization of the gut. Curr Opin Microbiol **14:**82-91.

7. **Fuentes S, van Nood E, Tims S, Heikamp-de Jong I, ter Braak CJ, Keller JJ, Zoetendal EG, de Vos WM.** 2014. Reset of a critically disturbed microbial ecosystem: faecal transplant in recurrent Clostridium difficile infection. ISME J **8:**1621-1633.

8. **Khoruts A, Dicksved J, Jansson JK, Sadowsky MJ.** 2010. Changes in the composition of the human fecal microbiome after bacteriotherapy for recurrent Clostridium difficile-associated diarrhea. J Clin Gastroenterol **44:**354-360.

9.   **Britton RA, Young VB.** 2014. Role of the intestinal microbiota in resistance to colonization by Clostridium difficile. Gastroenterology **146:**1547-1553.

10.   **Backhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, Li Y, Xia Y, Xie H, Zhong H, Khan MT, Zhang J, Li J, Xiao L, Al-Aama J, Zhang D, Lee YS, Kotowska D, Colding C, Tremaroli V, Yin Y, Bergman S, Xu X, Madsen L, Kristiansen K, Dahlgren J, Wang J.** 2015. Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. Cell Host Microbe **17:**690-703.

11.   **Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP, Heath AC, Warner B, Reeder J, Kuczynski J, Caporaso JG, Lozupone CA, Lauber C, Clemente JC, Knights D, Knight R, Gordon JI.** 2012. Human gut microbiome viewed across age and geography. Nature **486:**222-227.

12.   **Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO.** 2007. Development of the human infant intestinal microbiota. PLoS Biol **5:**e177.

13.   **Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N, Knight R.** 2010. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. Proc Natl Acad Sci U S A **107:**11971-11975.

14.   **Azad MB, Konya T, Maughan H, Guttman DS, Field CJ, Chari RS, Sears MR, Becker AB, Scott JA, Kozyrskyj AL, Investigators CS.** 2013. Gut microbiota of healthy Canadian infants: profiles by mode of delivery and infant diet at 4 months. CMAJ **185:**385-394.

15.   **Milani C, Mancabelli L, Lugli GA, Duranti S, Turroni F, Ferrario C, Mangifesta M, Viappiani A, Ferretti P, Gorfer V, Tett A, Segata N, van Sinderen D, Ventura M.** 2015. Exploring Vertical Transmission of Bifidobacteria from Mother to Child. Appl Environ Microbiol **81:**7078-7087.

16.   **La Rosa PS, Warner BB, Zhou Y, Weinstock GM, Sodergren E, Hall-Moore CM, Stevens HJ, Bennett WE, Jr., Shaikh N, Linneman LA, Hoffmann JA, Hamvas A, Deych E, Shands BA, Shannon WD, Tarr PI.** 2014. Patterned progression of bacterial populations in the premature infant gut. Proc Natl Acad Sci U S A **111:**12522-12527.

17.   **Greenwood C, Morrow AL, Lagomarcino AJ, Altaye M, Taft DH, Yu Z, Newburg DS, Ward DV, Schibler KR.** 2014. Early empiric antibiotic use in preterm infants is associated with lower bacterial diversity and higher relative abundance of Enterobacter. J Pediatr **165:**23-29.

18.   **Aagaard K, Riehle K, Ma J, Segata N, Mistretta TA, Coarfa C, Raza S, Rosenbaum S, Van den Veyver I, Milosavljevic A, Gevers D, Huttenhower C, Petrosino J, Versalovic J.** 2012. A metagenomic approach to characterization of the vaginal microbiome signature in pregnancy. PLoS One **7:**e36466.

19.   **Hunt KM, Foster JA, Forney LJ, Schutte UM, Beck DL, Abdo Z, Fox LK, Williams JE, McGuire MK, McGuire MA.** 2011. Characterization of the diversity and temporal stability of bacterial communities in human milk. PLoS One **6:**e21313.

20.   **Cabrera-Rubio R, Collado MC, Laitinen K, Salminen S, Isolauri E, 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.

21.   **Flores GE, Caporaso JG, Henley JB, Rideout JR, Domogala D, Chase J, Leff JW, Vazquez-Baeza Y, Gonzalez A, Knight R, Dunn RR, Fierer N.** 2014. Temporal variability is a personalized feature of the human microbiome. Genome Biol **15:**531.

22.   **Song SJ, Lauber C, Costello EK, Lozupone CA, Humphrey G, Berg-Lyons D, Caporaso JG, Knights D, Clemente JC, Nakielny S, Gordon JI, Fierer N, Knight R.** 2013. Cohabiting family members share microbiota with one another and with their dogs. Elife **2:**e00458.

23. **Shin H, Pei Z, Martinez KA, 2nd, Rivera-Vinas JI, Mendez K, Cavallin H, Dominguez-Bello MG.** 2015. The first microbial environment of infants born by C-section: the operating room microbes. Microbiome **3:**59.

24. **Jost T, Lacroix C, Braegger CP, Rochat F, Chassard C.** 2014. Vertical mother-neonate transfer of maternal gut bacteria via breastfeeding. Environ Microbiol **16:**2891-2904.

25. **Biasucci G, Rubini M, Riboni S, Morelli L, Bessi E, Retetangos C.** 2010. Mode of delivery affects the bacterial community in the newborn gut. Early Hum Dev **86 Suppl 1:**13-15.

26. **Dominguez-Bello MG, De Jesus-Laboy KM, Shen N, Cox LM, Amir A, Gonzalez A, Bokulich NA, Song SJ, Hoashi M, Rivera-Vinas JI, Mendez K, Knight R, Clemente JC.** 2016. Partial restoration of the microbiota of cesarean-born infants via vaginal microbial transfer. Nat Med **22:**250-253.

27. **Makino H, Kushiro A, Ishikawa E, Muylaert D, Kubota H, Sakai T, Oishi K, Martin R, Ben Amor K, Oozeer R, Knol J, Tanaka R.** 2011. Transmission of intestinal Bifidobacterium longum subsp. longum strains from mother to infant, determined by multilocus sequencing typing and amplified fragment length polymorphism. Appl Environ Microbiol **77:**6788-6793.

28. **Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, Egholm M, Henrissat B, Heath AC, Knight R, Gordon JI.** 2009. A core gut microbiome in obese and lean twins. Nature **457:**480-484.

29. **Scholz M, Ward DV, Pasolli E, Tolio T, Zolfo M, Asnicar F, Truong DT, Tett A, Morrow AL, Segata N.** 2016. Strain-level microbial epidemiology and population genomics from shotgun metagenomics. Nat Methods **13:**435-438.

30. **Schloissnig S, Arumugam M, Sunagawa S, Mitreva M, Tap J, Zhu A, Waller A, Mende DR, Kultima JR, Martin J, Kota K, Sunyaev SR, Weinstock GM, Bork P.** 2013. Genomic variation landscape of the human gut microbiome. Nature **493:**45-50.

31. **Turnbaugh PJ, Quince C, Faith JJ, McHardy AC, Yatsunenko T, Niazi F, Affourtit J, Egholm M, Henrissat B, Knight R, Gordon JI.** 2010. Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins. Proc Natl Acad Sci U S A **107:**7503-7508.

32. **Maurice CF, Haiser HJ, Turnbaugh PJ.** 2013. Xenobiotics shape the physiology and gene expression of the active human gut microbiome. Cell **152:**39-50.

33. **Gosalbes MJ, Abellan JJ, Durban A, Perez-Cobas AE, Latorre A, Moya A.** 2012. Metagenomics of human microbiome: beyond 16s rDNA. Clin Microbiol Infect **18 Suppl 4:**47-49.

34. **Bao G, Wang M, Doak TG, Ye Y.** 2015. Strand-specific community RNA-seq reveals prevalent and dynamic antisense transcription in human gut microbiota. Front Microbiol **6:**896.

35. **Kurokawa K, Itoh T, Kuwahara T, Oshima K, Toh H, Toyoda A, Takami H, Morita H, Sharma VK, Srivastava TP, Taylor TD, Noguchi H, Mori H, Ogura Y, Ehrlich DS, Itoh K, Takagi T, Sakaki Y, Hayashi T, Hattori M.** 2007. Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. DNA Res **14:**169-181.

36. **Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R, Angenent LT, Ley RE.** 2011. Succession of microbial consortia in the developing infant gut microbiome. Proc Natl Acad Sci U S A **108 Suppl 1:**4578-4585.

37. **Bickley J, Short JK, McDowell DG, Parkes HC.** 1996. Polymerase chain reaction (PCR) detection of Listeria monocytogenes in diluted milk and reversal of PCR inhibition caused by calcium ions. Lett Appl Microbiol **22:**153-158.

38.   **Cremonesi P, Castiglioni B, Malferrari G, Biunno I, Vimercati C, Moroni P, Morandi S, Luzzana M.** 2006. Technical note: Improved method for rapid DNA extraction of mastitis pathogens directly from milk. J Dairy Sci **89:**163-169.

39.   **Schrader C, Schielke A, Ellerbroek L, Johne R.** 2012. PCR inhibitors - occurrence, properties and removal. J Appl Microbiol **113:**1014-1026.

40.   **Ward TL, Hosid S, Ioshikhes I, Altosaar I.** 2013. Human milk metagenome: a functional capacity analysis. BMC Microbiol **13:**116.

41.   **Ramsay DT, Kent JC, Owens RA, Hartmann PE.** 2004. Ultrasound imaging of milk ejection in the breast of lactating women. Pediatrics **113:**361-367.

42.   **Jeurink PV, van Bergenhenegouwen J, Jimenez E, Knippels LM, Fernandez L, Garssen J, Knol J, Rodriguez JM, Martin R.** 2013. Human milk: a source of more life than we imagine. Benef Microbes **4:**17-30.

43.   **Costello EK, Stagaman K, Dethlefsen L, Bohannan BJ, Relman DA.** 2012. The application of ecological theory toward an understanding of the human microbiome. Science **336:**1255-1262.

44.   **Faith JJ, Guruge JL, Charbonneau M, Subramanian S, Seedorf H, Goodman AL, Clemente JC, Knight R, Heath AC, Leibel RL, Rosenbaum M, Gordon JI.** 2013. The long-term stability of the human gut microbiota. Science **341:**1237439.

45.   **Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, Tett A, Huttenhower C, Segata N.** 2015. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nat Methods **12:**902-903.

46.   **Morowitz MJ, Denef VJ, Costello EK, Thomas BC, Poroyko V, Relman DA, Banfield JF.** 2011. Strain-resolved community genomic analysis of gut microbial colonization in a premature infant. Proc Natl Acad Sci U S A **108:**1128-1133.

47.   **Sharon I, Morowitz MJ, Thomas BC, Costello EK, Relman DA, Banfield JF.** 2013. Time series community genomics analysis reveals rapid shifts in bacterial species, strains, and phage during infant gut colonization. Genome Res **23:**111-120.

48.   **Ogilvie LA, Jones BV.** 2015. The human gut virome: a multifaceted majority. Front Microbiol **6:**918.

49.   **Victoria JG, Kapoor A, Li L, Blinkova O, Slikas B, Wang C, Naeem A, Zaidi S, Delwart E.** 2009. Metagenomic analyses of viruses in stool samples from children with acute flaccid paralysis. J Virol **83:**4642-4651.

50.   **Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, Rohwer F, Gordon JI.** 2010. Viruses in the faecal microbiota of monozygotic twins and their mothers. Nature **466:**334-338.

51.   **Zhang T, Breitbart M, Lee WH, Run JQ, Wei CL, Soh SW, Hibberd ML, Liu ET, Rohwer F, Ruan Y.** 2006. RNA viral community in human feces: prevalence of plant pathogenic viruses. PLoS Biol **4:**e3.

52.   **de Medeiros RB, Figueiredo J, Resende Rde O, De Avila AC.** 2005. Expression of a viral polymerase-bound host factor turns human cell lines permissive to a plant- and insect-infecting virus. Proc Natl Acad Sci U S A **102:**1175-1180.

53.   **Balique F, Colson P, Barry AO, Nappez C, Ferretti A, Moussawi KA, Ngounga T, Lepidi H, Ghigo E, Mege JL, Lecoq H, Raoult D.** 2013. Tobacco mosaic virus in the lungs of mice following intra-tracheal inoculation. PLoS One **8:**e54993.

54.   **Colson P, Richet H, Desnues C, Balique F, Moal V, Grob JJ, Berbis P, Lecoq H, Harle JR, Berland Y, Raoult D.** 2010. Pepper mild mottle virus, a plant virus associated with specific immune responses, Fever, abdominal pains, and pruritus in humans. PLoS One **5:**e10041.

55.   **Marcobal A, Barboza M, Sonnenburg ED, Pudlo N, Martens EC, Desai P, Lebrilla CB, Weimer BC, Mills DA, German JB, Sonnenburg JL.** 2011. Bacteroides in the

infant gut consume milk oligosaccharides via mucus-utilization pathways. Cell Host Microbe **10:**507-514.

56. **Tailford LE, Crost EH, Kavanaugh D, Juge N.** 2015. Mucin glycan foraging in the human gut microbiome. Front Genet **6:**81.

57. **Turroni F, Milani C, van Sinderen D, Ventura M.** 2011. Genetic strategies for mucin metabolism in Bifidobacterium bifidum PRL2010: an example of possible human-microbe co-evolution. Gut Microbes **2:**183-189.

58. **LeBlanc JG, Milani C, de Giori GS, Sesma F, van Sinderen D, Ventura M.** 2013. Bacteria as vitamin suppliers to their host: a gut microbiota perspective. Curr Opin Biotechnol **24:**160-168.

59. **Flint HJ, Scott KP, Duncan SH, Louis P, Forano E.** 2012. Microbial degradation of complex carbohydrates in the gut. Gut Microbes **3:**289-306.

60. **Brown CT, Sharon I, Thomas BC, Castelle CJ, Morowitz MJ, Banfield JF.** 2013. Genome resolved analysis of a premature infant gut microbial community reveals a Varibaculum cambriense genome and a shift towards fermentation-based metabolism during the third week of life. Microbiome **1:**30.

61. **Houghteling PD, Walker WA.** 2015. Why is initial bacterial colonization of the intestine important to infants' and children's health? J Pediatr Gastroenterol Nutr **60:**294-307.

62. **Turroni F, Peano C, Pass DA, Foroni E, Severgnini M, Claesson MJ, Kerr C, Hourihane J, Murray D, Fuligni F, Gueimonde M, Margolles A, De Bellis G, O'Toole PW, van Sinderen D, Marchesi JR, Ventura M.** 2012. Diversity of bifidobacteria within the infant gut microbiota. PLoS One **7:**e36957.

63. **Turroni F, Foroni E, Serafini F, Viappiani A, Montanini B, Bottacini F, Ferrarini A, Bacchini PL, Rota C, Delledonne M, Ottonello S, van Sinderen D, Ventura M.** 2011. Ability of Bifidobacterium breve To Grow on Different Types of Milk: Exploring the Metabolism of Milk through Genome Analysis. Applied and Environmental Microbiology **77:**7408-7417.

64. **Walker AW, Ince J, Duncan SH, Webster LM, Holtrop G, Ze XL, Brown D, Stares MD, Scott P, Bergerat A, Louis P, McIntosh F, Johnstone AM, Lobley GE, Parkhill J, Flint HJ.** 2011. Dominant and diet-responsive groups of bacteria within the human colonic microbiota. Isme Journal **5:**220-230.

65. **Scott KP, Antoine JM, Midtvedt T, van Hemert S.** 2015. Manipulating the gut microbiota to maintain health and treat disease. Microb Ecol Health Dis **26:**25877.

66. **Gardy JL, Johnston JC, Ho Sui SJ, Cook VJ, Shah L, Brodkin E, Rempel S, Moore R, Zhao Y, Holt R, Varhol R, Birol I, Lem M, Sharma MK, Elwood K, Jones SJ, Brinkman FS, Brunham RC, Tang P.** 2011. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. N Engl J Med **364:**730-739.

67. **Loman NJ, Constantinidou C, Christner M, Rohde H, Chan JZ, Quick J, Weir JC, Quince C, Smith GP, Betley JR, Aepfelbacher M, Pallen MJ.** 2013. A culture-independent sequence-based metagenomics approach to the investigation of an outbreak of Shiga-toxigenic Escherichia coli O104:H4. JAMA **309:**1502-1510.

68. **Li SS, Zhu A, Benes V, Costea PI, Hercog R, Hildebrand F, Huerta-Cepas J, Nieuwdorp M, Salojarvi J, Voigt AY, Zeller G, Sunagawa S, de Vos WM, Bork P.** 2016. Durable coexistence of donor and recipient strains after fecal microbiota transplantation. Science **352:**586-589.

69. **Giannoukos G, Ciulla DM, Huang K, Haas BJ, Izard J, Levin JZ, Livny J, Earl AM, Gevers D, Ward DV, Nusbaum C, Birren BW, Gnirke A.** 2012. Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes. Genome Biol **13:**R23.

70. **Aronesty E.** 2013. Comparison of sequencing utility programs. The Open Bioinformatics Journal **7**.

71.   **Langmead B, Salzberg SL.** 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods **9:**357-359.

72.   **Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO.** 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res **41:**D590-596.

73.   **Edgar RC.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res **32:**1792-1797.

74.   **Stamatakis A.** 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics **30:**1312-1313.

75.   **Franzosa EA, Morgan XC, Segata N, Waldron L, Reyes J, Earl AM, Giannoukos G, Boylan MR, Ciulla D, Gevers D, Izard J, Garrett WS, Chan AT, Huttenhower C.** 2014. Relating the metatranscriptome and metagenome of the human gut. Proc Natl Acad Sci U S A **111:**E2329-2338.

76.   **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S.** 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics **25:**2078-2079.

77.   **Quinlan AR, Hall IM.** 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics **26:**841-842.

78.   **Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E.** 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research **12:**2825-2830.

79.   **Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C.** 2011. Metagenomic biomarker discovery and explanation. Genome Biol **12:**R60.

80.   **Asnicar F, Weingart G, Tickle TL, Huttenhower C, Segata N.** 2015. Compact graphical representation of phylogenetic data and metadata with GraPhlAn. PeerJ **3:**e1029.