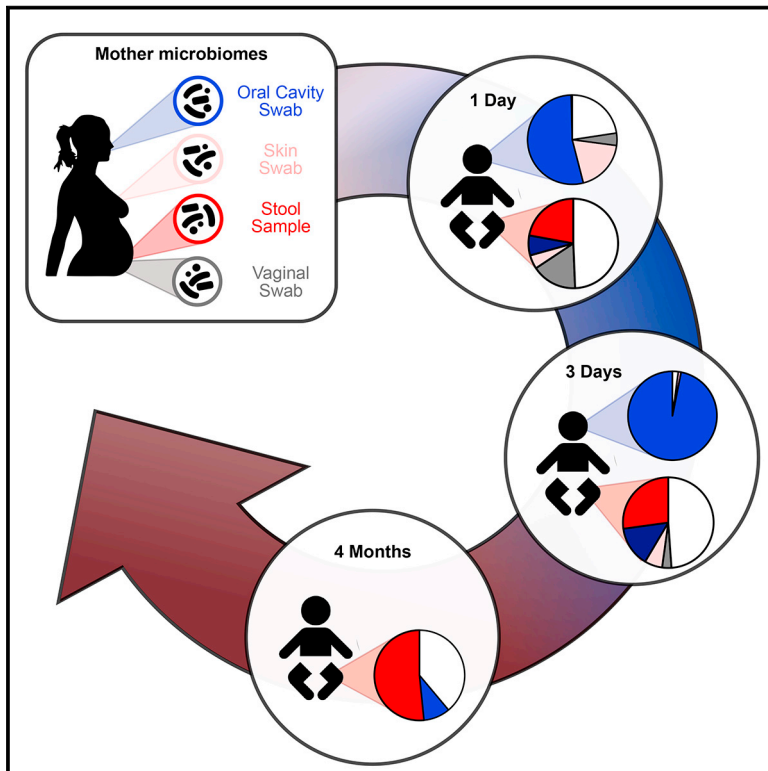# Cell Host & Microbe

Mother-to-Infant Microbial Transmission from Different Body Sites Shapes the Developing Infant Gut Microbiome

Graphical Abstract

Authors

Pamela Ferretti, Edoardo Pasolli, Adrian Tett, ..., Curtis Huttenhower, Peer Bork, Nicola Segata

Correspondence

nicola.segata@unitn.it

In Brief

Ferretti et al. use metagenomics with strain-resolved computational profiling to characterize the transfer of microbes from mothers to their infants during their first 4 months of life. Multiple maternal body sites contribute to the developing infant microbiome, with maternal gut strains providing the largest contribution of colonizing microorganisms.

Highlights

- Strain-resolved metagenomics was used to track mother-to-infant microbiome transfer

- Microbial strains from multiple maternal body sites transfer to the infant microbiome

- The early microbial diversity in the infant gut is rapidly shaped by niche selection

- The maternal gut microbiome is the source of the majority of transmitted strains

CellPress

Mother-to-Infant Microbial Transmission from Different Body Sites Shapes the Developing Infant Gut Microbiome

Pamela Ferretti,[1,2] Edoardo Pasolli,[1,11] Adrian Tett,[1,11] Francesco Asnicar,[1,11] Valentina Gorfer,[3] Sabina Fedi,[3] Federica Armanini,[1] Duy Tin Truong,[1] Serena Manara,[1] Moreno Zolfo,[1] Francesco Beghini,[1] Roberto Bertorelli,[4] Veronica De Sanctis,[4] Ilaria Bariletti,[3] Rosarita Canto,[3] Rosanna Clementi,[3] Marina Cologna,[3] Tiziana Crifò,[3] Giuseppina Cusumano,[3] Stefania Gottardi,[3] Claudia Innamorati,[3] Caterina Masè,[3] Daniela Postai,[3] Daniela Savoi,[3] Sabrina Duranti,[5] Gabriele Andrea Lugli,[5] Leonardo Mancabelli,[5] Francesca Turroni,[5] Chiara Ferrario,[5] Christian Milani,[5] Marta Mangifesta,[5,6] Rosaria Anzalone,[5] Alice Viappiani,[6] Moran Yassour,[7] Hera Vlamakis,[7] Ramnik Xavier,[7] Carmen Maria Collado,[8] Omry Koren,[9] Saverio Tateo,[3] Massimo Soffiati,[3] Anna Pedrotti,[3] Marco Ventura,[5] Curtis Huttenhower,[7,10] Peer Bork,[2] and Nicola Segata[1,12,*]

[1]Centre for Integrative Biology, University of Trento, 38123 Trento, Italy
[2]European Molecular Biology Laboratory, Structural and Computational Biology Unit, 69117 Heidelberg, Germany
[3]Azienda Provinciale per i Servizi Sanitari, 38123 Trento, Italy
[4]NGS Facility, Laboratory of Biomolecular Sequence and Structure Analysis for Health, Centre for Integrative Biology, University of Trento, 38123 Trento, Italy
[5]Laboratory of Probiogenomics, Department of Chemistry, Life Sciences and Environmental Sustainability, University of Parma, 43124 Parma, Italy
[6]GenProbio srl, 43124 Parma, Italy
[7]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[8]Institute of Agrochemistry and Food Technology, National Research Council, Paterna, 46980 Valencia, Spain
[9]Faculty of Medicine, Bar Ilan University, Safed 1311502, Israel
[10]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA
[11]These authors contributed equally
[12]Lead Contact
*Correspondence: nicola.segata@unitn.it
https://doi.org/10.1016/j.chom.2018.06.005

SUMMARY

The acquisition and development of the infant microbiome are key to establishing a healthy host-microbiome symbiosis. The maternal microbial reservoir is thought to play a crucial role in this process. However, the source and transmission routes of the infant pioneering microbes are poorly understood. To address this, we longitudinally sampled the microbiome of 25 mother-infant pairs across multiple body sites from birth up to 4 months postpartum. Strain-level metagenomic profiling showed a rapid influx of microbes at birth followed by strong selection during the first few days of life. Maternal skin and vaginal strains colonize only transiently, and the infant continues to acquire microbes from distinct maternal sources after birth. Maternal gut strains proved more persistent in the infant gut and ecologically better adapted than those acquired from other sources. Together, these data describe the mother-to-infant microbiome transmission routes that are integral in the development of the infant microbiome.

INTRODUCTION

The complex microbial community that inhabits humans, the microbiome, is an integral aspect of human health. In what is undoubtedly a complex interplay between host genetics and environmental conditions, these resident microbes support many functions in the human body, including the facilitation of nutrient absorption that would otherwise be inaccessible to the host (Flint et al., 2012; Wasielewski et al., 2016), the training and modulation of the immune system (Kau et al., 2011; Thaiss et al., 2016), and protection against pathogenic organisms (Buffie and Pamer, 2013). Dysbiosis of this harmonious relationship has reported to be linked to many diseases in adults, including inflammatory bowel diseases (IBDs) (Morgan et al., 2012; Parekh et al., 2015), type 2 diabetes (Qin et al., 2012; Upadhyaya and Banerjee, 2015), and colorectal cancer (Kostic et al., 2012; Vogtmann et al., 2016; Zeller et al., 2014). Similarly, in infants it is associated with IBD (Kolho et al., 2015), Crohn's disease (Gevers et al., 2014; Wang et al., 2016), type 1 diabetes (Kostic et al., 2015), necrotizing enterocolitis (Elgin et al., 2016; Ward et al., 2016), and asthma (Attar, 2015). While the importance of the host-microbiome interplay is not in question, the mechanisms by which an infant acquires these microbes, and from what source, remain largely unexplored.

The long-held belief that an infant is sterile at birth (Escherich, 1988) has been challenged by an increasing number of reports

offering evidence of the occurrence of intra-uterine seeding (Aagaard et al., 2014; Perez-Munoz et al., 2017; Rodriguez et al., 2015), but the role and importance of prenatal microbial colonization are still open to debate (Jimenez et al., 2008; Perez-Munoz et al., 2017; Walker et al., 2017). What is clearer is that extensive microbial colonization begins postpartum. Several crucial factors have been linked to the early development of the infant microbiome, including the mode of delivery (Dominguez-Bello et al., 2010, 2016) and gestational age at birth (La Rosa et al., 2014), as well as other influencing factors including maternal and infant antibiotic usage (Lemas et al., 2016; Yassour et al., 2016) and feeding method (formula or breastfeeding) (Backhed et al., 2015). Wider environmental exposure (Shin et al., 2015) and early intimate relations, particularly with the mother (Asnicar et al., 2017; Backhed et al., 2015; Korpela et al., 2018; Nayfach et al., 2016), also play a pivotal role in the early microbial acquisition and community succession in the infant.

Much of what is known about the acquisition of the early infant microbiome has been obtained using cultivation-based (Makino et al., 2011) or taxonomic profiling limited to a species-level resolution (Backhed et al., 2015). As individuals frequently share common species, inferring transmission at lower taxonomic resolutions, even at the species level, is not sufficient as a species can comprise multiple subspecies strain variants, which can be specific to different individuals (Backhed et al., 2015; Truong et al., 2017). It is essential, therefore, to use strain-level profiling to identify and quantify the instances of transmission from external sources to the infant. This has been shown for a limited number of cultivable species (Makino et al., 2011; Milani et al., 2015), and we previously demonstrated that the maternal microbial reservoir is an important source in the early acquisition of microbial species and strains in the infant gut (Asnicar et al., 2017; Korpela et al., 2018). Yet there has been no comprehensive assessment of the multiple potential maternal sources of microbial transmission, and how ultimately they contribute to the acquisition of the infant microbiome within hours of birth and over the first few months of life.

To this end, we investigated what is arguably the most important intimate relationship in the development of the early infant microbiome, the mother (Mueller et al., 2015). Focusing on five potential maternal sources of microbial transmission (skin, breast milk, fecal, vaginal, and oral), 25 mother and infant pairs were recruited and the infants longitudinally sampled from birth up to the first 4 months of life. Using high-resolution shotgun metagenomics (Quince et al., 2017) with improved strain-level computational profiling of known and poorly characterized microbiome members (Segata, 2018), we followed the transmission and assessed the impact of the maternal microbiomes on the development of infant oral and fecal microbial communities from birth to 4 months of life.

RESULTS

A Metagenomic Framework to Study Vertical Microbial Transmission

We enrolled 25 healthy pregnant women who vaginally delivered healthy newborns at full term. For each mother, we sampled the stool as a proxy of the lower intestine, and four additional body sites: the oral cavity (tongue dorsum swabs), the skin (intermammary cleft swabs), the vagina (vaginal introitus swabs), and the breast milk. Each newborn was sampled at two sites, the gut and the oral cavity, from birth up to 4 months postpartum (Figure 1A). All infants were exclusively breastfed at 3 days, 96% at 1 month, and 56% at 4 months. Of the 44% of non-exclusively breastfed infants at 4 months, 16% of the infants were exclusively formula-fed (Table S1A). All samples were shotgun sequenced, yielding a total of 216 high-quality metagenomes (STAR Methods), with an average of 5.73 (±7.26) Gbases per sample after quality control. DNA extraction from breast milk was not feasible in the majority of the cases, and when enough DNA was recovered a large fraction was found to be of human origin; thus, no high-quality milk microbiome samples were retained for further analysis.

Quantitative taxonomic profiling performed with MetaPhlAn2 (Truong et al., 2015) revealed that the diversity (Figure 1B), structure (Figure 1C), and composition (Figure S1) of the microbial communities are, as previously described (Human Microbiome Project Consortium, 2012a), distinct at each body site. The diversity of the infant microbiome was significantly lower than all the maternal microbiomes except for the *Lactobacillus*-dominated vaginal microbiome (Figure 1B), which is known to have a low diversity (Gajer et al., 2012; Human Microbiome Project Consortium, 2012b; Ravel et al., 2011), and was confirmed by rarefaction analysis to account for variance in sequence depths (Figure S2A). The gut and tongue microbiomes in infants were instead found to have a very high inter-subject variability, particularly at early time points (Figure 1D), compared with that of the mothers (Figure S2B). Interestingly, although the infants were all vaginally delivered, clustering based on initial sampling revealed that the early infant microbiomes did not consistently resemble one specific maternal body site (Figure 1C). For example, in some cases the early infant fecal samples (at 1 and 3 days) clustered with the maternal vaginal samples, while in other instances they clustered with the mothers' fecal samples. The high inter-subject diversity and lack of uniformity in the composition of the infant microbiomes at 1 day of life suggest that the initial exposure and seeding of the microbiome is largely stochastic, with each infant being influenced to a varying degree by the different maternal microbiomes (i.e., vagina, skin, oral, fecal), by potential prebirth *in utero* microbial acquisition, or by contact with other environmental sources.

Early Acquisition of Microbial Diversity in the Infant Is Subject to Subsequent Rapid Niche Selection

We observed a high species diversity in the infant fecal microbiomes at the first time point (T1, within 24 hr of birth; Figure 1B). This diversity decreased over the first week postpartum (p < 0.05 for both T2 and T3) before recovering over time (Figure 1B). Interestingly, the gut microbiome of exclusively breastfed infants at T4 (4 months) showed a lower diversity than the ones that gradually switched to formula feeding (Figure S2C). The high diversity at day 1 (T1) reflects the rapid influx of microbes, the pioneering microbiome, from maternal as well as other environmental sources after birth, and is consistent with a previous report using 16S rRNA taxonomic profiling (Wampach et al., 2017). Some of these microbes are present only transiently in some infants (Table S2A) as they are probably poorly adapted or unsuited to colonize the
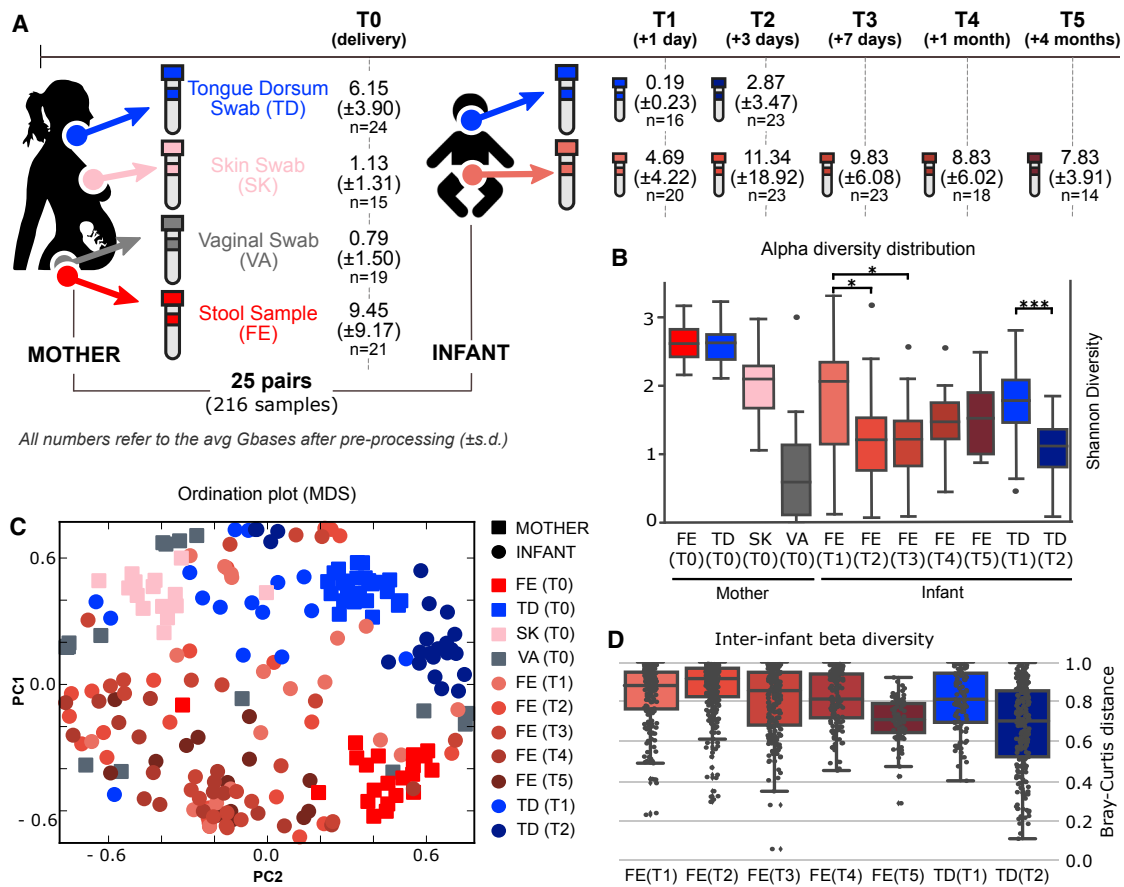
Figure 1. Longitudinal Metagenomic Sequencing of the Microbiome of Mother-Infant Pairs

(A) Samples were collected from 25 mother-infant pairs and metagenomically sequenced. Samples were taken from the stool (FE), skin (SK), vagina (VA), and tongue dorsum (TD) of the mothers and from the stool and tongue dorsum of the infants. Sampling of the infant started within 24 hr from delivery and continued for up to 4 months (STAR Methods). All samples were shotgun sequenced and the average depth (in Gbases) of the quality-controlled and human DNA-free samples are reported.

(B) Alpha-diversity distributions for each sample type and time point (*p < 0.05, ***p < 0.001).

(C) Ordination plot (MDS) of all the samples that passed preprocessing based on the Bray-Curtis distance between samples highlights the spatial clustering of samples with respect to both different body sites and longitudinal time points.

(D) Beta diversities (Bray-Curtis on log-scaled relative abundances) between samples within each infant body site (gut and tongue dorsum) across time points.

infant lower gastrointestinal tract and, as such, are easily lost or replaced. Others are instead more likely to be true early colonizers. In particular, we observed that some species (e.g., *Alistipes putredinis*, *Clostridium innocuum*, *Haemophilus parainfluenzae*, *Prevotella melaninogenica*, and *Streptococcus parasanguinis*) found in the infant stool at day 1 (T1) were not present at subsequent time points (see Table S2A for the full list). This was expected as *P. melaninogenica*, *C. innocuum*, and *Lactobacillus crispatus* are not typically associated with the infant fecal microbiome, never being found to constitute more than 1% abundance in any of the three metagenomic datasets (Backhed et al., 2015; Kostic et al., 2015; Vatanen et al., 2016) available for infants in *curatedMetagenomicData* (Pasolli et al., 2017) (Table S2B). Among the species that were lost in the infants between day 1 and the subsequent time points, 80% were shared with at least one body site of the respective mother (Table S2A). Many of these shared but transient microbes most likely originated from maternal body sites other than the maternal stool

(11 times from tongue dorsum, 1 from vagina, and 5 from skin), which could suggest their unsuitability to colonize the infant gut. In contrast, other more typical fecal species (e.g., *Bacteroides vulgatus*, *Bifidobacterium longum*, and *Bifidobacterium breve*) persisted from birth up to at least 4 months of age, when sample collection ended, suggesting that they are indeed colonizing the infant gut (Duranti et al., 2017; Milani et al., 2015).

Prepartum, the infant is subjected to largely anaerobic conditions *in utero*, but at birth the infant gut is mostly facultative aerobic, i.e., oxygen permissive (Houghteling and Walker, 2015; Rodriguez et al., 2015). At day 1 (T1), we found more strict anaerobic species (on average 4.7 species per sample) than aerobic (0.4) or facultative aerobic species (4.4), albeit present at rather low abundances (16.2% in relative abundance; Figure S3). Of these strict anaerobes, 40% of the species were in common with the corresponding mother. On the third day (T2), these strict anaerobes had drastically decreased in number by 26% in the stool and 75% in the tongue dorsum. Over the same period,

the abundance of facultative anaerobes increased in the stool (+9.1%) and slightly decreased in the tongue dorsum (−6.3%). These facultative anaerobes are known early colonizers of the infant gut, and these pioneering species mediate the shift from aerobic to anaerobic conditions typically associated with the adult state (Backhed et al., 2015; Houghteling and Walker, 2015; Rodriguez et al., 2015). Probably as a result of this transition toward anaerobic conditions, starting from the first week of age (T3), the number and the relative abundance of strictly anaerobic species increased over time (Figure S3).

This process of early and rapid acquisition of microbial species followed by selection and succession is reflected by inter-infant microbiome distances, which, after the third day, show a decrease as the infant microbiomes converge toward a more defined composition (Figure 1D). This is explained by the decreasing effect of the original direct seeding of the infant microbiome from different maternal body sites in different mother-infant pairs. Nevertheless, the infant gut microbiome at 4 months is still markedly different from that of the mother (Permanova on Bray-Curtis dissimilarity, p < 0.0001), confirming that a longer time window is required to fully appreciate the maturation process from infant to an adult-like state (Koren et al., 2012; Yatsunenko et al., 2012).

Infants Are Enriched with Microbial Species Present in the Microbiomes of Their Mothers

The infant gut microbiomes displayed a large proportion of species in common with their mothers for all mother-infant pairs. At day 1 (T1), about 50% of the microbial population in the infant gut belonged to species also present in at least one of the sampled maternal body sites (Figure 2A), and this fraction was relatively stable over time (50.7% at day 1 [T1], 48.3% at day 3 [T2], 52.2% at 1 week [T3], 37% at 1 month [T4], and 61.2% at 4 months [T5]). Shared species were notably in lower relative abundance in the mothers compared with their infants (between 2.7% and 5.5% in the mothers, depending on the reference infant time point). This suggests that it is the fitness of the microbial organisms reaching the infant gut that plays a greater role than the quantitative contribution of microbial seeding occurring for each species. The species observed in the infants but not shared with their respective mothers are likely to have been acquired from environmental sources, including other individuals that have had contact with the infant (Korpela et al., 2018). In the oral cavity, the relative abundance of shared species between infants and any maternal microbiomes was even more pronounced with 77.6% (day 1, T1) and 95.4% (day 3, T2) of the infant microbiome being shared with the mother (Figure 2B). Also in this case, the shared species were present at low relative abundances in the maternal tongue dorsum compared with the infants (5.7% at day 1 [T1] and 6.6% at day 3 [T2]).

We then considered the species shared between the infant and their mothers, but not with the other mothers in the cohort. Infants at day 1 (T1) shared significantly more microbial species with their mothercompared with other mothers (p < 0.05; Figure 2D). This was confirmed when looking at the fraction of shared species over the total number of species (p < 0.01) and at the cumulative relative abundance of the shared species (p < 0.05; Figure 2D). At 3 days of age (T2), infants still significantly (p < 0.01) shared a larger fraction of species with

their mother compared with others, before gradually losing their species-level remembrance of the maternal microbiome (Figure 2D).

To identify the source of the infant microbiome, we then analyzed separately the four sequenced maternal microbiomes as potential reservoirs of microbial transmission. A reanalysis of the Human Microbiome Project data (Human Microbiome Project Consortium, 2012b) confirmed that the compositions of the adult microbiomes belonging to the four body sites considered here are very distinct, with no species shared at comparable abundances across body sites, and this allowed us to compute the lists of body site-specific microbes reported in Table S3. All the maternal body sites contributed to the common mother-infant species with the mothers' stool microbiome accounting for 22.1% of the overall microbial abundance in the infant gut followed by the vagina (16.3%), the oral cavity (7.2%), and the skin (5%). Over time, the abundance of typical vaginal, oral, and cutaneous species decreased, suggesting that these species are likely transient inhabitants of the lower gastrointestinal tract. For instance, most of the vaginal species, which constituted up to 16.3% of the total abundance in the infant stool at day 1 (T1), were either lost or at undetectable levels by 1 week of age (T3). Interestingly, while abundant in the vaginal community, lactobacilli were rarely identified in the infant guts (Figure 2C). This is perhaps a consequence of lactobacilli being acidophilic, preferring a pH of approximately 3.5 (Ma et al., 2012), and the infant gut being near neutral (Evans et al., 1988). This might also explain the persistence of the more pH-tolerant (pH 5.5) vaginal species, *Gardnerella vaginalis* and *Atopobium vaginae*, at least over the first days of life (Figure 2C), although other explanations involving for example nutrient requirements are also plausible. The infant oral microbiome mirrors the trend observed in the stool, with the initial presence of species common to multiple maternal body sites, rapidly followed by the predominance of species more typically associated with oral microbial communities, at 3 days of life (T2; Figure 2B).

Mothers Transmit a Substantial Fraction of the Strains from Shared Species

To add support to vertical mother-to-infant transmission events, it is necessary to identify the same strain variants within the mother-infant pairs. Because considerable individualized strain-level heterogeneity has been observed in the human microbiome (Franzosa et al., 2015; Schloissnig et al., 2013; Truong et al., 2017), finding the same strain in the mother and in the infant would give strong evidence of intra-pair transmission. We thus implemented a novel combination of metagenomic strain-profiling tools (STAR Methods), expanding on methods validated previously (Asnicar et al., 2017). By coupling single-nucleotide variant (SNV) profiling (Truong et al., 2017) with gene-content-based profiling (Scholz et al., 2016), we characterized all the strains with sufficient coverage and applied a conservative threshold (0.1 in the normalized phylogenetic distance; STAR Methods) to call strain identity across paired mother-infant samples and infer vertical transmission (Figure 3A). Although there is no current consensus on the definition of a microbial strain (Segata, 2018), here we adopt the operational definition outlined elsewhere (Truong et al., 2017).
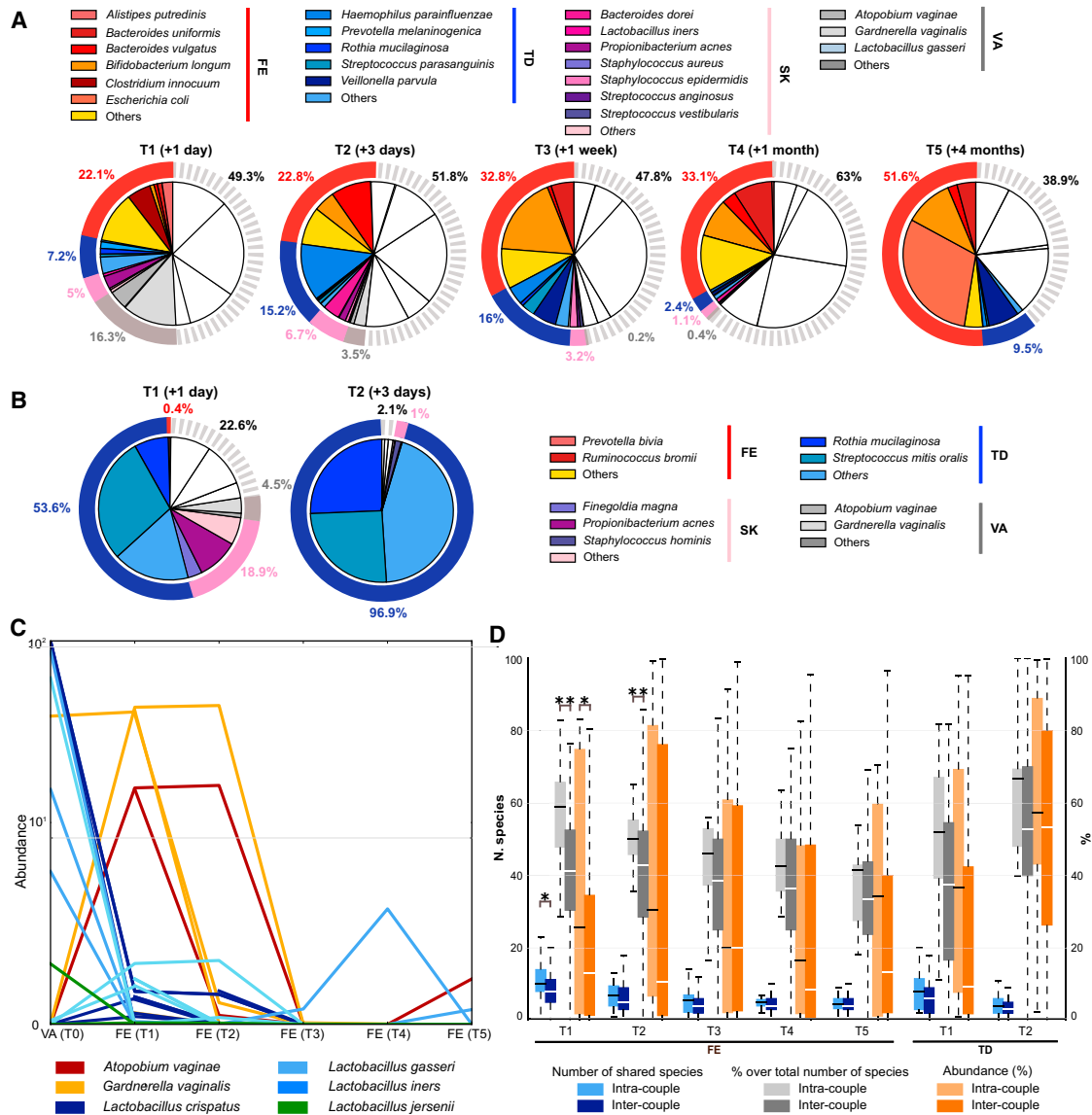
Figure 2. Microbial Species Common to the Mothers and Their Infants

(A and B) Average taxonomic composition of the infant stool (A) and tongue dorsum (B) microbiomes over time. The colored sectors indicate species that are found in the infant and his/her mother. White portions refer to species not found in the maternal body sites. External rings show the cumulative abundance of bacterial species per maternal body site.

(C) Relative abundances of the most abundant vaginal bacteria in the mothers and in the gut of their infants. Each line represents a mother-infant pair.

(D) Number, percentage over total number, and cumulative abundance of identified microbial species that are shared between each mother and her own infant (intra-pair), and between each mother and unrelated infants (inter-pair) (*p < 0.05 and **p < 0.01, t test).

This defines strain identity based on genomic identity using a similarity threshold on the SNV rate or gene content, which is consistent with short-term intra-subject strain variation (after excluding strain replacement), while also accounting for sequencing noise and non-adaptive low-frequency variants (Truong et al., 2017).

With this approach, we detected a total of 52 strains shared within mother and infant pairs (out of 317 cases of shared mother-infant species with typable strains), i.e., a 16.4% strain transmission rate. By comparison, we found 46 strains in infants that were in common with unrelated mothers out of a total of 6,319 mother-infant inter-pair shared species. Strain sharing between infants and unrelated mothers is thus a rare event, with a sharing rate of 0.73%. This rate is even lower than we previously observed for strain sharing among an adult worldwide population (Truong et al., 2017), which was based solely on SNV-based profiling, indicating that the combined approach of gene content and SNV strain profiling used here is more conservative. Overall intra-pair strain sharing was >22-fold higher than inter-pair sharing (Fisher test, $p < 1 \times 10^{-15}$), demonstrating the influence of the maternal source in shaping the infant microbiome.
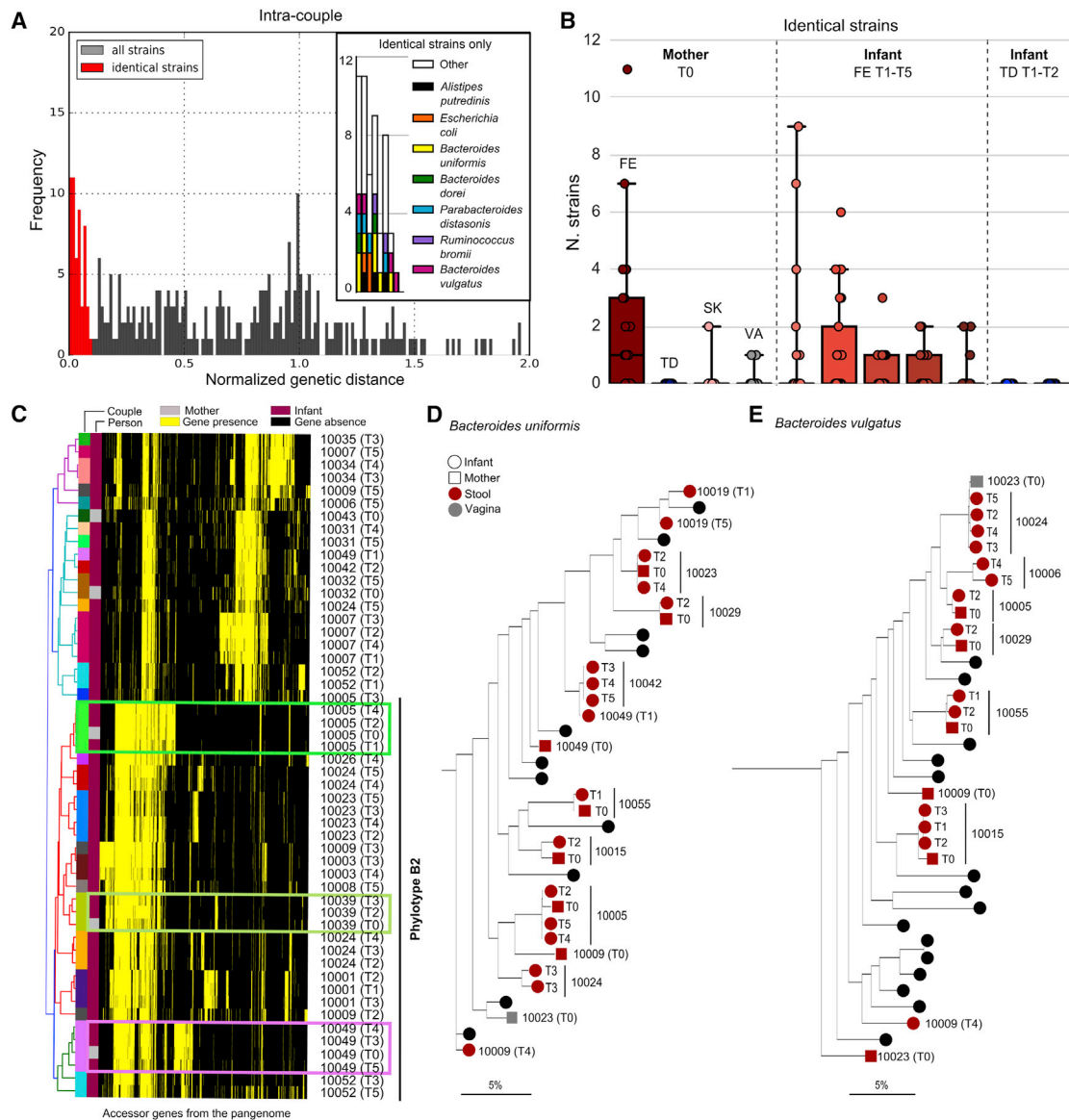
Figure 3. Strain Transmission between Mothers and Their Infants

(A and B) The distribution of the normalized strain intra-pair distances (STAR Methods) (A) and the number of vertical transmitted strains for each maternal source and each infant recipient body site and time point (B).

(C) *Escherichia coli* strain-specific gene content as identified by PanPhlAn. Strains with clear evidence of vertical transmission are indicated with boxes.

(D and E) Mother-infant phylogenies for *Bacteroides uniformis* and *Bacteroides vulgatus* as inferred by StrainPhlAn. Maternal body sites are represented by squares and infant body sites by circles. Mother-infant pairs with at least two samples are labeled with the pair ID in the trees (black circles otherwise).

The Maternal Gut Microbiome Is the Source of the Majority of Transmitted Strains

Considering the maternal sources of transmission, the gut microbiome was the largest donor of the infant-acquired strains (Figure 3B). Common strains were also found in the maternal skin and vaginal microbiomes, but to a lesser extent. The least important route of transmission appears to be the oral cavity, with little evidence of strain sharing between mother and infant. The number of strains in common between mother and infant gradually decreases over time, with 23 at T1 (1 day) and 28 at T2 (3 days), compared with 10 at T3 (1 week), 10 at T4 (1 month),

and 6 at T5 (4 months). This supports the hypothesis of selection for niche-specific bacteria from the pool of maternal strains seeding the infant gut.

Among the species for which we observed mother-infant transmission, *Escherichia coli* can be tracked very effectively using pangenome analysis (Scholz et al., 2016) because of its genomic plasticity and large set of accessory genes. Transmission and persistent colonization by *E. coli* strains were evident in three cases (Figure 3C) all belonging to the *E. coli* B2 phylotype, known for enhanced persistence in the infant gut (Nowrouzian et al., 2006). While other *E. coli* phylotypes were present in
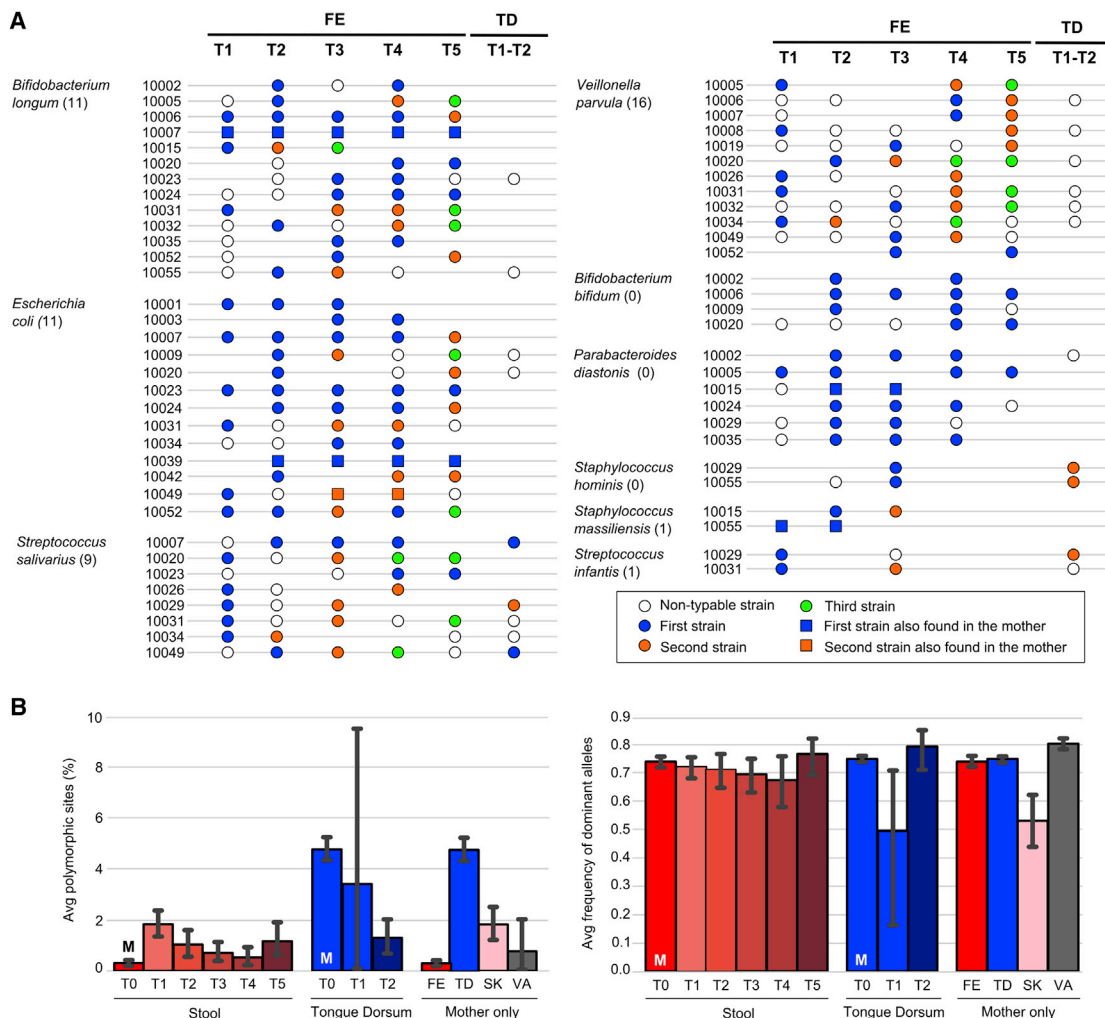
Figure 4. Strain Persistence, Strain Replacement Events, and Strain Heterogeneity

(A) Map of the strain dynamics in longitudinal infant stool (FE) samples for selected species (for full map, see Figure S6). The tongue dorsum (TD) column shows the species for which at least one of the strains found in stool was also present on the tongue dorsum. Blue circles represent the first strain of the species identified in the infant, whereas orange and green circles denote the second and third longitudinally identified strain, respectively. Empty circles refer to species for which strain profiling was not possible in the specific sample. Missing samples and samples lacking the species are not reported. The total number of infant replacement events observed in each species is shown in parentheses.

(B) Mean percentages of polymorphic sites and average frequency of the dominant alleles in polymorphic sites for each body site and time point ("M" indicates maternal samples). Color coding is as per Figure 1. p values are reported in Table S4. Error bars refer to 95% confidence intervals.

infants, none of these were shared with their mothers (Figure 3C), which is suggestive of a non-maternal source for these strains. Other typically gut-associated species belonging to bifidobacteria (Figure S4) and *Bacteroides* (Figures 3D and 3E) displayed clear maternal routes of transmission confirmed by SNV identity patterns. As a validation for inferring strains from metagenomes, a comparison was made with strains cultivated from sample aliquots for a subset of the mother-infant pairs. Metagenomically inferred strains were near identical (>99% nucleotide identity) to those identified via single isolate sequencing (Figure S5). Cultivation also confirmed that breast milk is yet another maternal source of bacterial strains colonizing the infant, with two genomes of *B. longum* and *Bifidobacterium bifidum* strains isolated from the same milk sample

(STAR Methods) recovered in the corresponding infant stool (Figure S5).

Vertically Transmitted Microbes Are More Likely to Be Stable Colonizers

Vertical microbial transmission from the mother to the infant can either be transient or lead to longer-lasting colonization of the infant gut (Korpela et al., 2018). Of the vertically transferred strains, 17 were identified at more than one time point in the infant (Figures 4A and S6). In 12 of these 17 cases, after the first occurrence of the maternally acquired strain we found no subsequent replacement by another conspecific strain, i.e., 70.5% of the strains were retained and 29.5% replaced. In contrast, the 163 strains present in the infant at more than one time point, but

without evidence that the mother was the source, were found to be replaced in 119 cases (73% replacement) and retained in 44 (27% retention). Vertically transmitted strains therefore seem to have a better fitness for colonization than strains without evidence of acquisition from the mother (70.5% versus 27% stable colonization, Fisher test, p < 0.001). This supports the intriguing hypothesis that maternal strains are likely to be more ecologically adaptable in the infant compared with non-maternal strains.

Conspecific Strain Diversity within Fecal Species Is Higher in the Infant Than in the Mother

We next investigated the total strain heterogeneity for each species in the microbiome of the infants compared with that of the mothers. To estimate conspecies strain heterogeneity and dominance, we analyzed the number of polymorphic nucleotide positions in the single-copy marker genes of each detected strain, as well as the average frequency of the dominant allelic variant in polymorphic positions. The analysis of maternal gut samples confirmed that the adult human gut tends to harbor only one strain of a given species (Truong et al., 2017), with an average fraction of polymorphic sites of 0.31% (Figure 4B). The infant gut microbiome at day 1 (T1) instead has a very high conspecific strain heterogeneity with 6.1-fold more polymorphisms than the mother (p = 1 × 10^{-7}). As observed, the early infant microbiome at day 1 postpartum is characterized by a high species diversity (Figure 1B), which is thus also accompanied by a high strain diversity, further suggesting that the pioneering microbiome is a complex community of microbes shaped by the process of ecological selection over time. Correspondingly, at later time points there is a decrease in the intra-species polymorphic rates up to 1 month (T4), to levels comparable with those of the mothers (no significant difference at 1 month compared with the mother). Simultaneously, a higher relative frequency of the dominant strain is observed (Figure 4B). Samples collected from the infant at 4 months of age (T5) then suggest that the strain diversity is increasing and remains significantly higher than the diversity in the mothers (p = 0.0014), potentially as a consequence of the increased exposure of the infant to other possible sources of microbial seeding from the environment. Comparing the conspecific strain diversity of the infant over time with the other maternal body sites (Figure 4B), we identified markedly different levels of heterogeneity (Table S4), with the maternal tongue dorsum significantly more strain diverse than the infant gut (p < 1 × 10^{-10} for all time points). The maternal skin and vaginal microbiomes have instead a strain diversity in line with that of the infant stool (Table S4). Interestingly, and in contrast to the stool, the maternal oral strain diversity compared with infants is significantly higher (p = 2 × 10^{-9} at T2, t test). Nevertheless, in the infant oral cavity, we identified the same pattern observed in the gut, namely a high species and conspecific strain diversity (Figures 1B and 4B) followed by a rapid decline in species and strain heterogeneity due to selection, which is observed to start after a few days postpartum.

Oral Bacteria Seed the Gut Microbiome in Infants to a Greater Extent Than in the Mothers

The oral cavity is the gateway to the gastrointestinal tract, but the analysis of the oral and gut microbial communities in adults has shown them to be distinct with minimal overlap (Human Micro-

biome Project Consortium, 2012b; Lloyd-Price et al., 2017). To our knowledge, the extent of species and strain sharing in the oral and gut of infants has not been previously characterized. In our cohort, we found an average of 9.8 species in common between the stool and the tongue dorsum samples in the infant at day 1 (T1) and 7.2 species at 3 days of life (T2). As a baseline of comparison, the average in the mothers was significantly lower with only 5.5 shared species (p < 0.001 for the two time points T1 and T2; Figure 5A). In both infants and mothers, the common species were more abundant on the tongue dorsum compared with the stool (Figure 5A). However, while for the mothers almost all oral-gut shared species were at very low abundance in the stool compared with the tongue dorsum (average of 0.96%), in contrast in the infant they represented about a quarter of the total abundance of the stool microbiome (average of 24.85% and 27.09% for T1 and T2, respectively). This suggests that the oral-gut axis in infants is a rather permissive interface, and that continuous seeding of the infant gut via the oral cavity is more relevant in infants than in adults, perhaps as a consequence of reduced acid production in infants at birth as the gastrointestinal tract develops. Interestingly, the number of shared oral-gut species already decreased in the infant by day 3; however, the abundance of the shared species increased in the oral cavity, suggesting that the orally acquired species at least transiently colonize the gut.

In the infant, the most commonly shared oral-gut species at day 1 (T1) were *Gardnerella vaginalis*, *Propionibacterium acnes*, *Prevotella bivia*, *Atopobium vaginae*, and *Prevotella melaninogenica*, while at day 3 (T2) they were mainly *Rothia mucilaginosa*, *Streptococcus parasanguinis*, and *Streptococcus salivarius* (Tables S5A–S5D). When looking at strain identity within the shared oral-gut species, we found that *S. salivarius* (Figure 5C) and *R. mucilaginosa* (Figure 5B) were the species with the highest number of shared strains (Table S5E). These results suggest that *S. salivarius* and *R. mucilaginosa* might have an increased capacity to survive in both the oral cavity and the gut, at least for a limited time.

Strains Belonging to as yet Uncharacterized Species Are Also Vertically Transmitted

To perform strain profiling for microbes belonging to poorly characterized species without available genomes, we expanded our analysis by performing metagenomic assembly (Nurk et al., 2017) on each sample followed by contig binning (Kang et al., 2015), phylogenetic profiling (Segata et al., 2013), and whole-genome strain identity inference (STAR Methods). Overall, we reconstructed 1,132 metagenome-assembled genomes (on average five per sample; Table S6A) with sufficient quality (≥50% completeness, ≤5% contamination) to be amenable for strain tracking. Of these, 763 genomes could be assigned to a known species by applying a 95% percent identity threshold on the whole sequence (STAR Methods) and were therefore not further processed because these strains were captured by the reference-based methods already considered above. However, the remaining 369 genomes (Figure 6; Table S6A) did not belong to any of the 13,575 microbial species for which at least one reference genome is available (STAR Methods), including 36 that could not even be assigned below the level of family. The genera containing most of the unknown
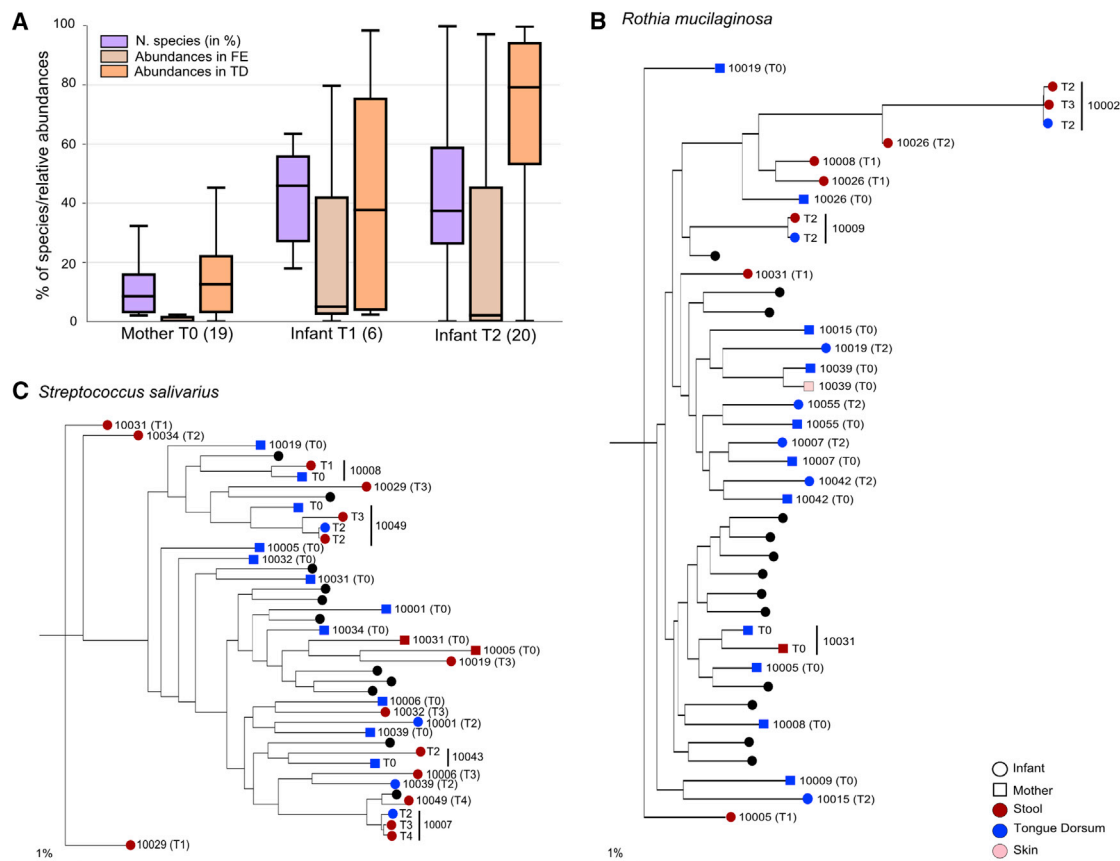
Figure 5. Infants Have More Shared Species and Strains between the Oral and Gut Microbiome Than Their Mothers
(A) Number of shared species normalized by the total number of species present in the gut and the tongue dorsum of each subject, and the cumulative abundances of shared species in the two body sites. The number of samples considered for the analysis is reported in parentheses.
(B and C) Transmission trees for *Streptococcus salivarius* (B) and *Rothia mucilaginosa* (C). Only pairs with more than two samples present in the tree are shown (black circles otherwise).

species were *Streptococcus* (32 genomes), *Clostridium* (31), and *Prevotella* (31).

Next, we compared the 369 taxonomically uncharacterized genomes against each other to identify the presence of the same strain in different metagenomic samples. Using a strict threshold of 99.5% identity over the full length of the genomes, we identified eight vertical transmission events (Figure 6; Table S6B). In six cases the strain sharing was between the mother and infant gut microbiomes. Two of these strains belonged to uncharacterized species in the *Akkermansia* and *Bacteroides* genera (less than 89% identity with the closest available genomes over less than 75% of the length), while for the other four strains classification was even more challenging and we could only infer that they belonged to four different phyla (Verrucomicrobia, Proteobacteria, Bacteroidetes, and Firmicutes; Table S6B). In addition to the fecal transmission routes, two vertical transmission events were also observed from other body sites with an uncharacterized Clostridiales strain (99.9% of similarity) shared by the maternal vaginal community and the stool of the infant, and an unknown *Leptospira* strain (99.9% of similarity) shared by the skin microbiome of the mother and the saliva microbiome of the infant. There was only one case of a strain from

an unknown species with 99.9% similarity within an unrelated mother-infant pair, strongly confirming the occurrence of vertical transmission for the eight genomes above (Fisher test, $p < 1 \times 10^{-9}$) and confirming that uncharacterized species have a role in the mother-to-infant microbial seeding.

DISCUSSION

We investigated here the early acquisition and development of the infant gut and oral microbiomes, and in particular the role of different maternal sources in this process, by means of a longitudinal multiple body site metagenomic approach. Among our main findings is the very high microbial diversity and strain heterogeneity in the pioneering infant gut microbiome even in the first day of life, which dramatically decreases within the first week before recovering and gradually increasing over the next 4 months. While we cannot discount the possibility of intra-uterine microbial acquisition, this suggests early seeding with an overall species diversity and strain heterogeneity far higher than previously appreciated, followed by steep selection forces that maintain only part of this early biodiversity. The selection process is corroborated by the decreasing strain heterogeneity
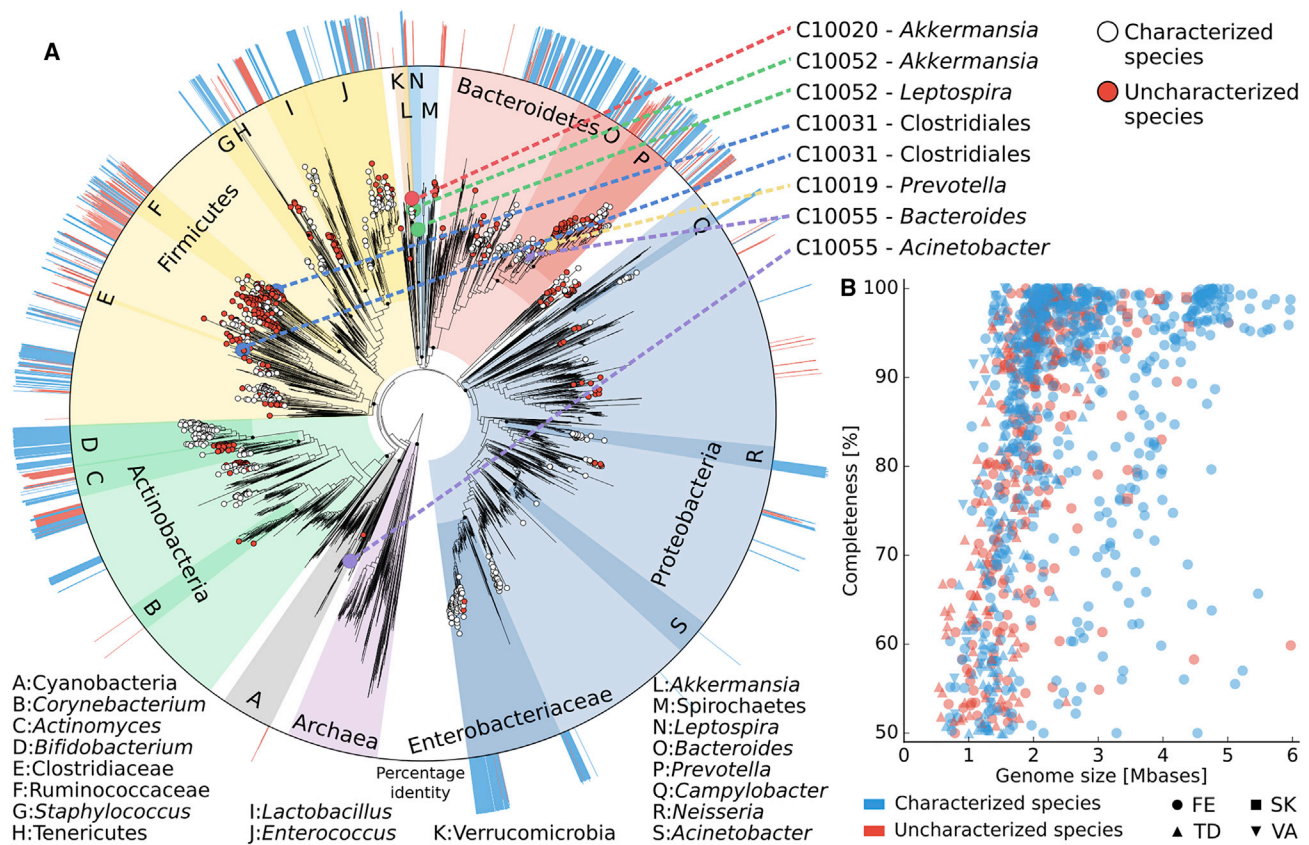
Figure 6. Phylogenetic Placement of 1,132 Metagenomically Reconstructed Genomes and Mother-to-Infant Transmission of Taxonomically Uncharacterized Strains

(A) We used PhyloPhlAn2 (Segata et al., 2013) to place the genomes reconstructed with metaSPAdes (Nurk et al., 2017) and binned with MetaBAT2 (Kang et al., 2015) (STAR Methods) on the microbial "Tree of Life" (Ciccarelli et al., 2006; Segata et al., 2013), which encompasses 4,000 species with available reference genomes. Leaf nodes without circles refer to reference genomes from known species, white circles indicate reconstructed genomes that are close (>95% identity) to a known species, and red circles show reconstructed genomes that cannot be assigned (<95% identity) to known species. The eight events of mother-to-infant transmission of strains from species yet to be described are called out on the top right, and the external ring of the phylogeny reports the percent identity of each leaf node against the closest genomes from known species (values below 95% are shown in red).

(B) The reconstructed genomes with completeness >50% from each body site are plotted with the corresponding completeness and genome size.

in the developing infant microbiome and the initial increase of facultative anaerobes, which are subsequently replaced by strict anaerobes consistently with the biochemical changes of the infant gut environment. Far from being a static process, microbial seeding from maternal sources is continuous, with some species and strains appearing in the infant at later time points. We thus describe how the microbial colonization process in the infant reflects a balance between influx of microbial strains and a process of niche selection. This balance is likely the key for the physiological development of the infant microbiome and should be further studied to unravel potential links with pathologies in childhood.

Another key finding of our study is that we identify microbial strains present in the infants for which there is strong evidence of transmission from their mothers, and that these strains are more likely to adapt to and persist in the infant gut than non-maternally acquired strains. This reinforces the importance of this vertical mother-to-infant microbial transmission from multiple sources, because even though the maternal microbiomes cannot explain many microbial strains present in the infant, the transmitted strains appear crucial in the developing microbiome

(Korpela et al., 2018). The mechanisms of this phenomenon should be further investigated and may be related to a combination of prebirth transmission, shared environmental factors, and common genetic factors that could partially explain the mother-infant strain specificity. Nevertheless, if we consider vertical microbial transmission as a physiological process under evolutionary pressure in recent human history, the study of vertically transmitted strains can provide the basis for better understanding the impact of non-vaginal birth (C-section) and non-exclusive breastfeeding.

The methodological approaches in this work are a novel combination of reference-based and assembly-based computational profiling that enables us to comprehensively describe mother-to-infant strain transmission and strain-level dynamics in the infant microbiome. The adoption of computational profiling tools with a resolution at the level of individual strains is key because no direct evidence of microbial transmission can be inferred when bacteria are categorized solely at the species level. In this work, we exploit and model the very high inter-subject variability of strains in the general population, and suggest

transmission only for the cases in which a mother and her infant have a strain similarity substantially higher than that found between unrelated subjects. Overall, we find 62 cases with strong evidence of strain transmission, eight of which are from species that are currently uncharacterized. We also find evidence of strains originating from multiple maternal sources, with the vaginal, skin, oral, and gut communities all contributing to the early infant microbiome. However, even after a few days postpartum, the contribution of the vaginal and skin microbiome already decreases. Here we focus on the mother, perhaps the most important familiar relationship in the development of the infant microbiome. However, the same techniques can be applied to describe other routes of transmission, including family members other than the mother (Korpela et al., 2018), and the hospital rooms (Brooks et al., 2017). These sets of microbial sources have not been studied together, and their integrated analysis will further contribute to understanding the mechanisms of early microbiome acquisition and subsequent development.

STAR★METHODS

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

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - ○ Experimental Design and Cohort Recruitment
- METHOD DETAILS
  - ○ Sample Collection
  - ○ DNA Extraction and Sequencing
  - ○ Metagenome Quality Control and Preprocessing
  - ○ Species- and Strain-Level Profiling
  - ○ Profiling by Metagenomic Assembly
  - ○ Integrated Cultivation and Metagenomic Analysis of Bifidobacteria
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and six tables and can be found with this article online at https://doi.org/10.1016/j.chom.2018.06.005.

AUTHOR CONTRIBUTIONS

V.G., A.P., and N.S. conceived the study; P.F., F. Armanini, I.B., R. Canto, R. Clementi, M.C., T.C., G.C., S.F., S.G., C.I., C. Masè, D.P., D.S., M.S., S.T., and A.P. designed, coordinated, and performed the sample collection; F. Armanini, R.B., V.D.S., S.D., G.A.L., L.M., F.T., C.F., C. Milani, M.V., M.M., R.A., A.V., and M.Y. performed the experiments; P.F., E.P., A.T., F. Asnicar, D.T.T., S.M., M.Z., F.B., H.V., R.X., C.M.C., O.K., C.H., P.B., and N.S. analyzed the data; and P.F., E.P., A.T., F. Asnicar, and N.S. wrote and revised the manuscript. All authors read and approved the final manuscript.

REFERENCES

Aagaard, K., Ma, J., Antony, K.M., Ganu, R., Petrosino, J., and Versalovic, J. (2014). The placenta harbors a unique microbiome. Sci. Transl. Med. 6, 237ra265.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Aronesty, E. (2011). ea-utils: Command-Line Tools for Processing Biological Sequencing Data (Expression Analysis).

Asnicar, F., Manara, S., Zolfo, M., Truong, D.T., Scholz, M., Armanini, F., Ferretti, P., Gorfer, V., Pedrotti, A., Tett, A., et al. (2017). Studying vertical microbiome transmission from mothers to infants by strain-level metagenomic profiling. mSystems 2, https://doi.org/10.1128/mSystems.00164-16.

Asnicar, F., Weingart, G., Tickle, T.L., Huttenhower, C., and Segata, N. (2015). Compact graphical representation of phylogenetic data and metadata with GraPhlAn. PeerJ 3, e1029.

Attar, N. (2015). Microbiome: early life dysbiosis increases asthma risk. Nat. Rev. Microbiol. 13, 660.

Backhed, F., Roswall, J., Peng, Y., Feng, Q., Jia, H., Kovatcheva-Datchary, P., Li, Y., Xia, Y., Xie, H., Zhong, H., et al. (2015). Dynamics and stabilization of the human gut microbiome during the first year of life. Cell Host Microbe 17, 690–703.

Brooks, B., Olm, M.R., Firek, B.A., Baker, R., Thomas, B.C., Morowitz, M.J., and Banfield, J.F. (2017). Strain-resolved analysis of hospital rooms and infants reveals overlap between the human and room microbiome. Nat. Commun. 8, 1814.

Buffie, C.G., and Pamer, E.G. (2013). Microbiota-mediated colonization resistance against intestinal pathogens. Nat. Rev. Immunol. 13, 790–801.

Ciccarelli, F.D., Doerks, T., von Mering, C., Creevey, C.J., Snel, B., and Bork, P. (2006). Toward automatic reconstruction of a highly resolved tree of life. Science 311, 1283–1287.

Dominguez-Bello, M.G., Costello, E.K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N., and Knight, R. (2010). Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. Proc. Natl. Acad. Sci. USA 107, 11971–11975.

Dominguez-Bello, M.G., De Jesus-Laboy, K.M., Shen, N., Cox, L.M., Amir, A., Gonzalez, A., Bokulich, N.A., Song, S.J., and Hoashi, M. (2016). Partial restoration of the microbiota of cesarean-born infants via vaginal microbial transfer. Nat. Med. 22, 250–253.

Duranti, S., Lugli, G.A., Mancabelli, L., Armanini, F., Turroni, F., James, K., Ferretti, P., Gorfer, V., Ferrario, C., Milani, C., et al. (2017). Maternal inheritance of bifidobacterial communities and bifidophages in infants through vertical transmission. Microbiome 5, 66.

Elgin, T.G., Kern, S.L., and McElroy, S.J. (2016). Development of the neonatal intestinal microbiome and its association with necrotizing enterocolitis. Clin. Ther. 38, 706–715.

Escherich, T. (1988). The intestinal bacteria of the neonate and breast-fed infant. 1884. Rev. Infect. Dis. 10, 1220–1225.

Evans, D.F., Pye, G., Bramley, R., Clark, A.G., Dyson, T.J., and Hardcastle, J.D. (1988). Measurement of gastrointestinal pH profiles in normal ambulant human subjects. Gut 29, 1035–1041.

Flint, H.J., Scott, K.P., Louis, P., and Duncan, S.H. (2012). The role of the gut microbiota in nutrition and health. Nat. Rev. Gastroenterol. Hepatol. *9*, 577–589.

Franzosa, E.A., Huang, K., Meadow, J.F., Gevers, D., Lemon, K.P., Bohannan, B.J.M., and Huttenhower, C. (2015). Identifying personal microbiomes using metagenomic codes. Proc. Natl. Acad. Sci. USA *112*, E2930–E2938.

Gajer, P., Brotman, R.M., Bai, G., Sakamoto, J., Schutte, U.M., Zhong, X., Koenig, S.S., Fu, L., Ma, Z.S., Zhou, X., et al. (2012). Temporal dynamics of the human vaginal microbiota. Sci. Transl. Med. *4*, 132ra152.

Gevers, D., Kugathasan, S., Denson, L.A., Vazquez-Baeza, Y., Van Treuren, W., Ren, B., Schwager, E., Knights, D., Song, S.J., Yassour, M., et al. (2014). The treatment-naive microbiome in new-onset Crohn's disease. Cell Host Microbe *15*, 382–392.

Houghteling, P.D., and Walker, W.A. (2015). Why is initial bacterial colonization of the intestine important to infants' and children's health? J. Pediatr. Gastroenterol. Nutr. *60*, 294–307.

Human Microbiome Project Consortium (2012a). A framework for human microbiome research. Nature *486*, 215–221.

Human Microbiome Project Consortium (2012b). Structure, function and diversity of the healthy human microbiome. Nature *486*, 207–214.

Jimenez, E., Marin, M.L., Martin, R., Odriozola, J.M., Olivares, M., Xaus, J., Fernandez, L., and Rodriguez, J.M. (2008). Is meconium from healthy newborns actually sterile? Res. Microbiol. *159*, 187–193.

Kang, D.D., Froula, J., Egan, R., and Wang, Z. (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ *3*, e1165.

Kau, A.L., Ahern, P.P., Griffin, N.W., Goodman, A.L., and Gordon, J.I. (2011). Human nutrition, the gut microbiome and the immune system. Nature *474*, 327–336.

Kolho, K.L., Korpela, K., Jaakkola, T., Pichai, M.V., Zoetendal, E.G., Salonen, A., and de Vos, W.M. (2015). Fecal microbiota in pediatric inflammatory bowel disease and its relation to inflammation. Am. J. Gastroenterol. *110*, 921–930.

Koren, O., Goodrich, J.K., Cullender, T.C., Spor, A., Laitinen, K., Backhed, H.K., Gonzalez, A., Werner, J.J., Angenent, L.T., Knight, R., et al. (2012). Host remodeling of the gut microbiome and metabolic changes during pregnancy. Cell *150*, 470–480.

Korpela, K., Costea, P., Coelho, L.P., Kandels-Lewis, S., Willemsen, G., Boomsma, D.I., Segata, N., and Bork, P. (2018). Selective maternal seeding and environment shape the human gut microbiome. Genome Res. *28*, 561–568.

Kostic, A.D., Gevers, D., Pedamallu, C.S., Michaud, M., Duke, F., Earl, A.M., Ojesina, A.I., Jung, J., Bass, A.J., Tabernero, J., et al. (2012). Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. Genome Res. *22*, 292–298.

Kostic, A.D., Gevers, D., Siljander, H., Vatanen, T., Hyotylainen, T., Hamalainen, A.M., Peet, A., Tillmann, V., Poho, P., Mattila, I., et al. (2015). The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. Cell Host Microbe *17*, 260–273.

La Rosa, P.S., Warner, B.B., Zhou, Y., Weinstock, G.M., Sodergren, E., Hall-Moore, C.M., Stevens, H.J., Bennett, W.E., Jr., Shaikh, N., Linneman, L.A., et al. (2014). Patterned progression of bacterial populations in the premature infant gut. Proc. Natl. Acad. Sci. USA *111*, 12522–12527.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Lemas, D.J., Yee, S., Cacho, N., Miller, D., Cardel, M., Gurka, M., Janicke, D., and Shenkman, E. (2016). Exploring the contribution of maternal antibiotics and breastfeeding to development of the infant microbiome and pediatric obesity. Semin. Fetal Neonatal Med. *21*, 406–409.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics *27*, 2987–2993.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009).

The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079.

Lloyd-Price, J., Mahurkar, A., Rahnavard, G., Crabtree, J., Orvis, J., Hall, A.B., Brady, A., Creasy, H.H., McCracken, C., Giglio, M.G., et al. (2017). Strains, functions and dynamics in the expanded human microbiome project. Nature *550*, 61.

Ma, B., Forney, L.J., and Ravel, J. (2012). Vaginal microbiome: rethinking health and disease. Annu. Rev. Microbiol. *66*, 371–389.

Makino, H., Kushiro, A., Ishikawa, E., Muylaert, D., Kubota, H., Sakai, T., Oishi, K., Martin, R., Ben Amor, K., Oozeer, R., et al. (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.

Milani, C., Mancabelli, L., Lugli, G.A., Duranti, S., Turroni, F., Ferrario, C., Mangifesta, M., Viappiani, A., Ferretti, P., Gorfer, V., et al. (2015). Exploring vertical transmission of bifidobacteria from mother to child. Appl. Environ. Microbiol. *81*, 7078–7087.

Morgan, X.C., Tickle, T.L., Sokol, H., Gevers, D., Devaney, K.L., Ward, D.V., Reyes, J.A., Shah, S.A., LeLeiko, N., Snapper, S.B., et al. (2012). Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol. *13*, R79.

Mueller, N.T., Bakacs, E., Combellick, J., Grigoryan, Z., and Dominguez-Bello, M.G. (2015). The infant microbiome development: mom matters. Trends Mol. Med. *21*, 109–117.

Nayfach, S., Rodriguez-Mueller, B., Garud, N., and Pollard, K.S. (2016). An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography. Genome Res. *26*, 1612–1625.

Nowrouzian, F.L., Adlerberth, I., and Wold, A.E. (2006). Enhanced persistence in the colonic microbiota of Escherichia coli strains belonging to phylogenetic group B2: role of virulence factors and adherence to colonic cells. Microbes Infect. *8*, 834–840.

Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P.A. (2017). metaSPAdes: a new versatile metagenomic assembler. Genome Res. *27*, 824–834.

Parekh, P.J., Balart, L.A., and Johnson, D.A. (2015). The influence of the gut microbiome on obesity, metabolic syndrome and gastrointestinal disease. Clin. Transl. Gastroenterol. *6*, e91.

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. *25*, 1043–1055.

Pasolli, E., Schiffer, L., Manghi, P., Renson, A., Obenchain, V., Truong, D.T., Beghini, F., Malik, F., Ramos, M., Dowd, J.B., et al. (2017). Accessible, curated metagenomic data through ExperimentHub. Nat. Methods *14*, 1023.

Perez-Munoz, M.E., Arrieta, M.C., Ramer-Tait, A.E., and Walter, J. (2017). A critical assessment of the "sterile womb" and "in utero colonization" hypotheses: implications for research on the pioneer infant microbiome. Microbiome *5*, 48.

Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., Liang, S., Zhang, W., Guan, Y., Shen, D., et al. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature *490*, 55–60.

Quince, C., Walker, A.W., Simpson, J.T., Loman, N.J., and Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. Nat. Biotechnol. *35*, 833.

Ravel, J., Gajer, P., Abdo, Z., Schneider, G.M., Koenig, S.S., McCulle, S.L., Karlebach, S., Gorle, R., Russell, J., Tacket, C.O., et al. (2011). Vaginal microbiome of reproductive-age women. Proc. Natl. Acad. Sci. USA *108* (*Suppl 1*), 4680–4687.

Rodriguez, J.M., Murphy, K., Stanton, C., Ross, R.P., Kober, O.I., Juge, N., Avershina, E., Rudi, K., Narbad, A., Jenmalm, M.C., et al. (2015). The composition of the gut microbiota throughout life, with an emphasis on early life. Microb. Ecol. Health Dis. *26*, 26050.

Schloissnig, S., Arumugam, M., Sunagawa, S., Mitreva, M., Tap, J., Zhu, A., Waller, A., Mende, D.R., Kultima, J.R., Martin, J., et al. (2013). Genomic variation landscape of the human gut microbiome. Nature *493*, 45–50.

Scholz, M., Ward, D.V., Pasolli, E., Tolio, T., Zolfo, M., Asnicar, F., Truong, D.T., Tett, A., Morrow, A.L., and Segata, N. (2016). Strain-level microbial epidemiology and population genomics from shotgun metagenomics. Nat. Methods *13*, 435–438.

Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., and Huttenhower, C. (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. Nat. Methods *9*, 811–814.

Segata, N., Bornigen, D., Morgan, X.C., and Huttenhower, C. (2013). PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. Nat. Commun. *4*, 2304.

Segata, N. (2018). On the road to strain-resolved comparative metagenomics. mSystems *3*, https://doi.org/10.1128/mSystems.00190-17.

Shin, H., Pei, Z., Martinez, K.A., 2nd, Rivera-Vinas, J.I., Mendez, K., Cavallin, H., and Dominguez-Bello, M.G. (2015). The first microbial environment of infants born by C-section: the operating room microbes. Nat. Med. *3*, 59.

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

Thaiss, C.A., Zmora, N., Levy, M., and Elinav, E. (2016). The microbiome and innate immunity. Nature *535*, 65–74.

Truong, D.T., Franzosa, E.A., Tickle, T.L., Scholz, M., Weingart, G., Pasolli, E., Tett, A., Huttenhower, C., and Segata, N. (2015). MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nat. Methods *12*, 902–903.

Truong, D.T., Tett, A., Pasolli, E., Huttenhower, C., and Segata, N. (2017). Microbial strain-level population structure and genetic diversity from metagenomes. Genome Res. *27*, 626–638.

Upadhyaya, S., and Banerjee, G. (2015). Type 2 diabetes and gut microbiome: at the intersection of known and unknown. Gut Microbes *6*, 85–92.

Vatanen, T., Kostic, A.D., d'Hennezel, E., Siljander, H., Franzosa, E.A., Yassour, M., Kolde, R., Vlamakis, H., Arthur, T.D., Hamalainen, A.M., et al. (2016). Variation in microbiome LPS immunogenicity contributes to autoimmunity in humans. Cell. *165*, 1551.

Ventura, M., Reniero, R., and Zink, R. (2001). Specific identification and targeted characterization of *Bifidobacterium lactis* from different environmental isolates by a combined multiplex-PCR approach. Appl. Environ. Microbiol. *67*, 2760–2765.

Vogtmann, E., Hua, X., Zeller, G., Sunagawa, S., Voigt, A.Y., Hercog, R., Goedert, J.J., Shi, J., Bork, P., and Sinha, R. (2016). Colorectal cancer and the human gut microbiome: reproducibility with whole-genome shotgun sequencing. PLoS One *11*, e0155362.

Walker, R.W., Clemente, J.C., Peter, I., and Loos, R.J.F. (2017). The prenatal gut microbiome: are we colonized with bacteria in utero? Pediatr. Obes. *12* (*Suppl 1*), 3–17.

Wampach, L., Heintz-Buschart, A., Hogan, A., Muller, E.E.L., Narayanasamy, S., Laczny, C.C., Hugerth, L.W., Bindl, L., Bottu, J., Andersson, A.F., et al. (2017). Colonization and Succession within the human gut microbiome by archaea, bacteria, and microeukaryotes during the first year of life. Front. Microbiol. *8*, 738.

Wang, F., Kaplan, J.L., Gold, B.D., Bhasin, M.K., Ward, N.L., Kellermayer, R., Kirschner, B.S., Heyman, M.B., Dowd, S.E., Cox, S.B., et al. (2016). Detecting microbial dysbiosis associated with pediatric Crohn disease despite the high variability of the gut microbiota. Cell Rep. *14*, 945–955.

Ward, D.V., Scholz, M., Zolfo, M., Taft, D.H., Schibler, K.R., Tett, A., Segata, N., and Morrow, A.L. (2016). Metagenomic sequencing with strain-level resolution implicates uropathogenic *E. coli* in necrotizing enterocolitis and mortality in preterm infants. Cell Rep. *14*, 2912–2924.

Wasielewski, H., Alcock, J., and Aktipis, A. (2016). Resource conflict and cooperation between human host and gut microbiota: implications for nutrition and health. Ann. N. Y. Acad. Sci. *1372*, 20–28.

Yassour, M., Vatanen, T., Siljander, H., Hamalainen, A.M., Harkonen, T., Ryhanen, S.J., Franzosa, E.A., Vlamakis, H., Huttenhower, C., Gevers, D., et al. (2016). Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. Sci. Transl. Med. *8*, 343ra381.

Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R.N., Anokhin, A.P., et al. (2012). Human gut microbiome viewed across age and geography. Nature *486*, 222–227.

Zeller, G., Tap, J., Voigt, A.Y., Sunagawa, S., Kultima, J.R., Costea, P.I., Amiot, A., Bohm, J., Brunetti, F., Habermann, N., et al. (2014). Potential of fecal microbiota for early-stage detection of colorectal cancer. Mol. Syst. Biol. *10*, 766.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
Man-Rogosa-Sharp (MRS)	Scharlau Chemie, Barcelona, Spain	N/A
Critical Commercial Assays		
PowerSoil DNA Isolation Kit	MoBio Laboratories Carlsbad, USA	N/A
NexteraXT DNA Library Preparation Kit	Illumina, California, USA	N/A
Deposited Data		
Raw sequencing data	This paper	NCBI-SRA BioProject number PRJNA352475
Assembled contigs	This paper	https://www.dropbox.com/s/7pygii8khtrriw0/genome_bins.tar.bz2?dl=0
Software and Algorithms		
FastqMcf	Aronesty, 2011	https://github.com/ExpressionAnalysis/ea-utils/blob/wiki/FastqMcf.md
BowTie2	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
trim_galore	N/A	http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
MetaPhlAn2 (version 2.6)	Segata et al., 2012; Truong et al., 2015	https://bitbucket.org/biobakery/metaphlan2
Seqtk	N/A	https://github.com/lh3/seqtk
PanPhlAn	Scholz et al., 2016	https://bitbucket.org/CibioCM/panphlan/wiki/Home
StrainPhlAn	Truong et al., 2017	https://bitbucket.org/biobakery/biobakery/wiki/strainphlan
RAxML	Stamatakis, 2014	https://sco.h-its.org/exelixis/web/software/raxml/
metaSPAdes (version 3.10.1)	Nurk et al., 2017	http://bioinf.spbau.ru/en/spades3.7
MetaBAT2 (version 2.12.1)	Kang et al., 2015	https://bitbucket.org/berkeleylab/metabat/wiki/Home
CheckM (version 1.0.7)	Parks et al., 2015	https://github.com/Ecogenomics/CheckM/wiki
BLASTn (version 2.6.0)	Altschul et al., 1990	https://blast.ncbi.nlm.nih.gov
pyani (version 0.2.6)	N/A	https://github.com/widdowquinn/pyani
PhyloPhlAn2	Segata et al., 2013	https://bitbucket.org/nsegata/phylophlan/wiki/phylophlan2
GraPhlAn	Asnicar et al., 2015	https://bitbucket.org/nsegata/graphlan/wiki/Home
Samtools	Li, 2011; Li et al., 2009	http://samtools.sourceforge.net/

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources, reagents, and software should be directed to and will be fulfilled by the Lead Contact, Nicola Segata (nicola.segata@unitn.it).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Experimental Design and Cohort Recruitment

A total of 25 pregnant women (26-43 years of age, BMI before pregnancy between 16.5 and 29.4) were recruited by Santa Chiara Hospital in Trento, Italy, between April 2015 and July 2016. The protocol of this study was approved by the Ethics Committee of Santa Chiara Hospital (Trento, Italy) and the Ethics Committee of the University of Trento. Parents were asked to provide informed consent and complete questionnaires regarding pregnancy data, maternal diet, medical records, and lifestyle before and during the pregnancy (Table S1A).

Exclusion criteria included delivery by Caesarean section, pre-term birth (<37 weeks), body temperature >38°C, birth weight <2.5 Kg, antibiotic treatments during pregnancy, intention to avoid formula feeding at least for the first six months. The infant feeding practices as well as antibiotic usage and complete medical records across the sampling period were also collected (Table S1A). At 4 months of age, 56% of infants were exclusively breastfed (14 out of 25), while 12% (3 out of 25) of mothers reported mixed feeding and 16% exclusive formula feeding (4 out of 25). For four pairs, the sampling was interrupted before the age of 4 months, and therefore no metadata in terms of feeding are available.

METHOD DETAILS

Sample Collection
The samples collection procedure was based on HMP sampling guidelines (Human Microbiome Project Consortium, 2012b). Skin samples were collected using Catch-All-Swabs (Epicentre Technologies, Wisconsin, USA) shortly after birth but before the skin-to-skin contact with the infant, by swabbing the upper area of the maternal breast (intermammary cleft). After pre-moistening with 2 ml SCF-1 buffer (50 mM Tris buffer, pH 7.6, 1mM EDTA, pH 8.0, and 0.5% Tween-20) (Human Microbiome Project Consortium, 2012b) contained in a 15 ml sterile screw top collection tube (Sarstedt, Nümbrecht, Germany), the swab head was rubbed back and forth for approximately 30 seconds over the area (repeating twice) before the swab was returned to the buffered solution. Vaginal swabs were collected before delivery, as soon as possible upon arrival at the delivery ward to reduce potential blood contamination. The swab was rubbed 5 times, with a circular motion, in the vaginal introitus and then the swab head was placed in a 15 ml sterile screw top collection tube containing 2 ml SCF-1 buffer. Maternal tongue dorsum swabs were collected shortly before or during delivery by rubbing a swab on the central area of the back of the tongue for approximately 5 seconds. The swab head was then placed in a 15 ml collection tube containing 2 ml SCF-1 buffer. The same procedure was used for sampling the infant tongue dorsum, at time of the collection of the first stool sample. Maternal breast milk was self-collected by the mothers starting at one day after delivery, using gloves to avoid skin contamination. Stool samples from the mother were collected during or shortly after the delivery by the hospital staff, using collection tubes specific for faecal material (Sarstedt, Nümbrecht, Germany). All infant stool samples were self-collected by the mother, following a detailed protocol. The samples collected directly at the hospital were frozen at -20°C immediately after the collection and moved to a -80°C facility within a week, where they remained stored until further analysis. Additional aliquots were collected and stored in 20% glycerol and kept at -20°C for the cultivation experiments. After leaving the hospital, normally three days after delivery, the mothers performed the collection of the infants' stool samples at home and put the samples immediately at -20°C, which were then delivered to the hospital staff within 12 hours.

DNA Extraction and Sequencing
DNA was extracted using the PowerSoil DNA Isolation Kit (MoBio Laboratories Carlsbad, USA), as described in the HMP protocol (Human Microbiome Project Consortium, 2012b). For the stool samples, a preliminary heating step (65°C for 10 minutes, 95°C for 10 minutes) was performed before extraction. For samples collected via swabbing (skin, oral, and vaginal) the head of the Catch-All Sample Collection Swab was removed and put in the PowerBead Tube. The 2 ml SCF-1 specimen-containing buffer was centrifuged at 1000g for 5 min and added to the PowerBead Tube for cell lysing performed using the MOBIO Vortex Adapter (MO BIO Catalog No. 13000-V1). DNA was recovered in 10 mM Tris pH 7.4 and quantified using Qubit 2.0 (Thermo Fisher Scientific, Massachusetts, USA) fluorometer as per manufacturer's instructions. Sequencing libraries were prepared using the NexteraXT DNA Library Preparation Kit (Illumina, California, USA), following the manufacturer's guidelines. The sequencing was performed on the HiSeq2500 (Illumina, California, USA).

Metagenome Quality Control and Preprocessing
Out of the 225 collected samples subjected to sequencing, 216 provided more than 5.1×10^9 total reads (average 23.96×10^6 reads per sample) and were therefore used for the downstream analysis. Of these samples, 119 (55%) were stool samples (21 from the mothers and 98 from the infants), 15 (6.9%) were skin swabs from the mother, 63 (29.2%) were oral cavity swabs (24 from the mothers and 39 from the infants), and 19 were derived from vaginal swabs (8.8%).

The generated raw metagenomes were processed with FastqMcf (Aronesty, 2011) by trimming positions with quality <15, removing low-quality reads (mean quality <25), and discarding reads shorter than 90 nt. Human and bacteriophage phiX174 (Illumina spike-ins) DNA were then removed using BowTie2 (Langmead and Salzberg, 2012) to map the reads against the reference genomes. The adapters were also discarded by trim_galore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) with parameters "-q 0 –nextera –stringency 5". All samples providing less than 50,000 reads were excluded from the downstream analysis. The number of reads for each sample after preprocessing is reported in Table S1B.

Species- and Strain-Level Profiling
Species-level quantitative taxonomic profiling was performed using MetaPhlAn2 (version 2.6) (Segata et al., 2012; Truong et al., 2015) on the post-processed reads. MetaPhlAn2 estimates the relative abundances of each known microbial species. These relative abundances are proportional to the underlying absolute species concentrations that cannot be quantified from metagenomic data alone. Taxonomic profiles included bacteria, archaea, microbial Eukaryotes, and viruses, and were inferred by MetaPhlAn2 using the ~1 M unique clade-specific marker genes identified from ~17,000 reference genomes (~13,500 bacterial and archaeal, ~3,500 viral, and

~110 eukaryotic). Rarefaction analysis was performed using Seqtk (https://github.com/lh3/seqtk) at 5 million reads per sample; 57 samples (28 of which from infants) were excluded because of insufficient sequence depth (less than 5 million reads per sample after sub-sampling).

Strain-level analysis was performed using a combination of gene-content-based profiling using PanPhlAn (Scholz et al., 2016), and single-nucleotide variant profiling using StrainPhlAn (Truong et al., 2017). PanPhlAn was applied on the preprocessed metagenomes using default parameters generating a presence/absence gene-family profile independently for each sample and each species present in the reference database. A species-specific gene-family matrix was obtained by combining profiles from different samples and was further processed with RAxML (Stamatakis, 2014) to generate a species-specific phylogenetic tree. Comparison among different trees was performed by normalizing distances in each tree by its median value. Finally, strain distance for any pair of samples was defined as the normalized phylogenetic distance on the corresponding tree. Similarly, StrainPhlAn was run on the preprocessed reads with default parameters and adding the options "-alignment_program mafft" and "–relaxed_parameters3″ for each reference species and by considering pair-specific markers. This generated a maximum of twenty-five (i.e., the number mother-infant pairs) phylogenetic trees for each species. A single species-specific value for each pair of samples was obtained by averaging their distances along the different trees. Strain distance was defined as for PanPhlAn including normalization of each tree by its median value. The final strain distance for any sample pair was obtained by taking the minimum value between the two strain distance values generated by PanPhlAn and StrainPhlAn. We considered a threshold of 0.1 on the strain distance for defining a pair of strains as the same strain. This threshold was chosen as the value dividing the two peaks in the bimodal distribution of all-versus-all normalized species-specific strains distances representing identical and thus possible transmitted strains (values smaller than 0.1) and clearly distinct strains (values bigger than 0.1). The same approach was adopted for all species under the assumption that this threshold is very stringent and that almost identical genomes represent almost identical strains irrespective of the species they belong to. Results reported in Figures 3A and 3B were obtained by considering the minimum strain distance value for each pair of subjects to avoid skewing of the statistics due to effects of multiple timepoints and multiple body sites effects. Strain-level phylogenies presented in Figures 3D and 3E and in Figures 5B and 5C were obtained from the StrainPhlAn analysis using RAxML (parameters: -m GTRCAT –p 1234) (Stamatakis, 2014). For the four phylogenies in the two figures we also reported the proportion of phylogenetic distance in percentage computed with respect to the total branch length of the respective phylogeny.

Results reported in Figure 1C were computed using the Bray-Curtis dissimilarity and the MDS algorithm, both implemented in the scikit-learn Python package (scikit-learn).

Profiling by Metagenomic Assembly

All samples were processed independently for *de novo* metagenomic assembly through metaSPAdes (Nurk et al., 2017) (version 3.10.1 using default parameters), discarding contigs shorter than 1000 nt. Reads were mapped to contigs using Bowtie2 (Langmead and Salzberg, 2012) (version 2.2.9; option "–very-sensitive-local"), and the mapping output was used for contig binning through MetaBAT2 (Kang et al., 2015) (version 2.12.1; option "-m 1500"). CheckM (Parks et al., 2015) (version 1.0.7; lineage specific workflow) was applied to the resulting bins, and only those with sufficient quality ($\geq$50% completeness, $\leq$5% contamination) were considered for further analysis. We used BLASTn (Altschul et al., 1990) (version 2.6.0; default parameters) to map the contigs of each reconstructed bin against all the microbial reference genomes (including bacteria, archaea, viruses, and microeukaryotes) available in the NCBI repository as of September 2016 for a total of 13,575 unique and named species. Each position of a given contig was assigned to the hit with the highest *bitscore*, and then the median percent identity across the mapping positions was computed. The average percentage identity for a genome was computed by averaging the percent identity values amongst its contigs. For detecting potential mother-to-infant transmissions, we considered only bins with an average percent identity below 95%, as genomes with available close references were already considered by the assembly-free approach described above. The 369 genomes without a close reference were then compared against each other to identify the presence of the same strain in different samples. We computed the average nucleotide identity (ANI) using the pyani tool (v.0.2.6; option "-m ANIb") and considered a threshold of 99.5% identity over the full length of the genomes to define strain identity. Because of the use of the whole genome instead of species-specific markers and pangenomes, the strain identity thresholds for metagenomically assembled strains and StrainPhlAn/PanPhlAn profiled strains are not directly comparable. Phylogenetic placement of the reconstructed genomes was obtained by running PhyloPhlAn2 (Segata et al., 2013), and visualized through GraPhlAn (Asnicar et al., 2015).

Integrated Cultivation and Metagenomic Analysis of Bifidobacteria

Bifidobacterium strains were inoculated in Man-Rogosa-Sharp (MRS) (Scharlau Chemie) supplemented with 0.05% (wt/vol) L-cysteine hydrochloride and incubated in an anaerobic atmosphere [2.99% (vol/vol) H2, 17.01% (vol/vol) CO2, and 80% (vol/vol) N2] in a chamber (Concept 400, Ruskin) at 37°C for 16 h. Chromosomal DNA was extracted as described previously (Ventura et al., 2001). Library preparation and Illumina sequencing were performed at the GenProbio srl (Parma, Italy) and previously described (Duranti et al., 2017).

We validated the microbial transmission results for bifidobacteria by searching for the presence of the 12 assembled genomes isolated in the metagenomes. We first mapped the metagenomics reads against the 12 isolates genomes using Bowtie2 (Langmead and Salzberg, 2012) (parameters: –very-sensitive-local –k 100000 –no-unal). Then, we performed a SNV-rate by analyzing the ".sam" output from Bowtie2 through Samtools (Li, 2011; Li et al., 2009) and custom Python scripts. For each isolate genome mapped against

a single metagenome, we considered the sites with at least two high-quality reads covering the single position in either forward or reverse orientation, as defined in the "DP4" Variant Call Format field.

QUANTIFICATION AND STATISTICAL ANALYSIS

The permanova analysis has been performed using the scikit-bio Python package (scikit-bio). When not specified otherwise, t-test was used to calculate p-values. All the other statistical analyses have been performed with open source software mentioned and referenced in the description of the analyses. The number and description of the samples are reported in Table S1B, and the statistical analyses always refer to the whole set of samples in the specific condition of interest.

DATA AND SOFTWARE AVAILABILITY

All metagenomes were deposited and are available at the NCBI Sequence Read Archive (SRA) under BioProject number PRJNA352475 and SRA accession number SRP100409. Taxonomic profiles along with metadata are available in the curated MetagenomicData package (Pasolli et al., 2017). Assembled contigs are available at https://www.dropbox.com/s/7pygii8khtrriw0/genome_bins.tar.bz2?dl=0. All the software packages used in this study are open source and publicly available and the new software developed for this study is contained in the new releases of PanPhlAn (Scholz et al., 2016) and StrainPhlAn (Truong et al., 2017) available at https://bitbucket.org/CibioCM/panphlan/wiki/Home and https://bitbucket.org/biobakery/biobakery/wiki/strainphlan respectively.

Supplemental Information

Mother-to-Infant Microbial Transmission

from Different Body Sites Shapes

the Developing Infant Gut Microbiome

Pamela Ferretti, Edoardo Pasolli, Adrian Tett, Francesco Asnicar, Valentina Gorfer, Sabina Fedi, Federica Armanini, Duy Tin Truong, Serena Manara, Moreno Zolfo, Francesco Beghini, Roberto Bertorelli, Veronica De Sanctis, Ilaria Bariletti, Rosarita Canto, Rosanna Clementi, Marina Cologna, Tiziana Crifò, Giuseppina Cusumano, Stefania Gottardi, Claudia Innamorati, Caterina Masè, Daniela Postai, Daniela Savoi, Sabrina Duranti, Gabriele Andrea Lugli, Leonardo Mancabelli, Francesca Turroni, Chiara Ferrario, Christian Milani, Marta Mangifesta, Rosaria Anzalone, Alice Viappiani, Moran Yassour, Hera Vlamakis, Ramnik Xavier, Carmen Maria Collado, Omry Koren, Saverio Tateo, Massimo Soffiati, Anna Pedrotti, Marco Ventura, Curtis Huttenhower, Peer Bork, and Nicola Segata
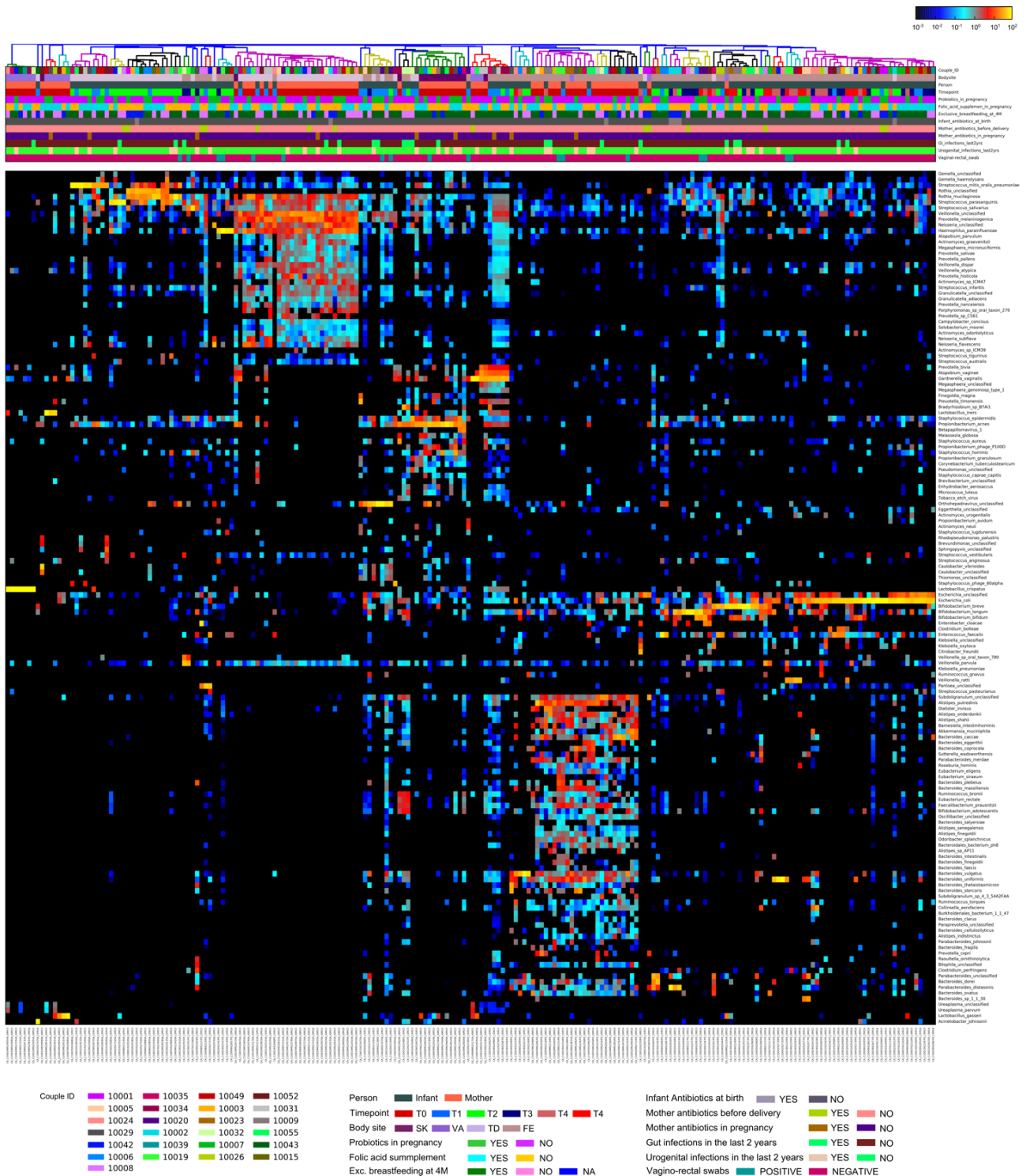
Figure S1. Related to Figure 1. Taxonomic composition and abundance at the species level of all collected samples. Collected metadata include probiotics, folic acid and antibiotics consumption during pregnancy, antibiotics administered to the mother immediately before the delivery, antibiotics administered to the infant shortly after birth, gut and urogenital infections in the mother in the past 2 years, result of the vaginal-rectal swab (standard hospital practice aimed at identifying the presence of *Streptococcus* species belonging to the group B), and breastfeeding practice at 4 months.
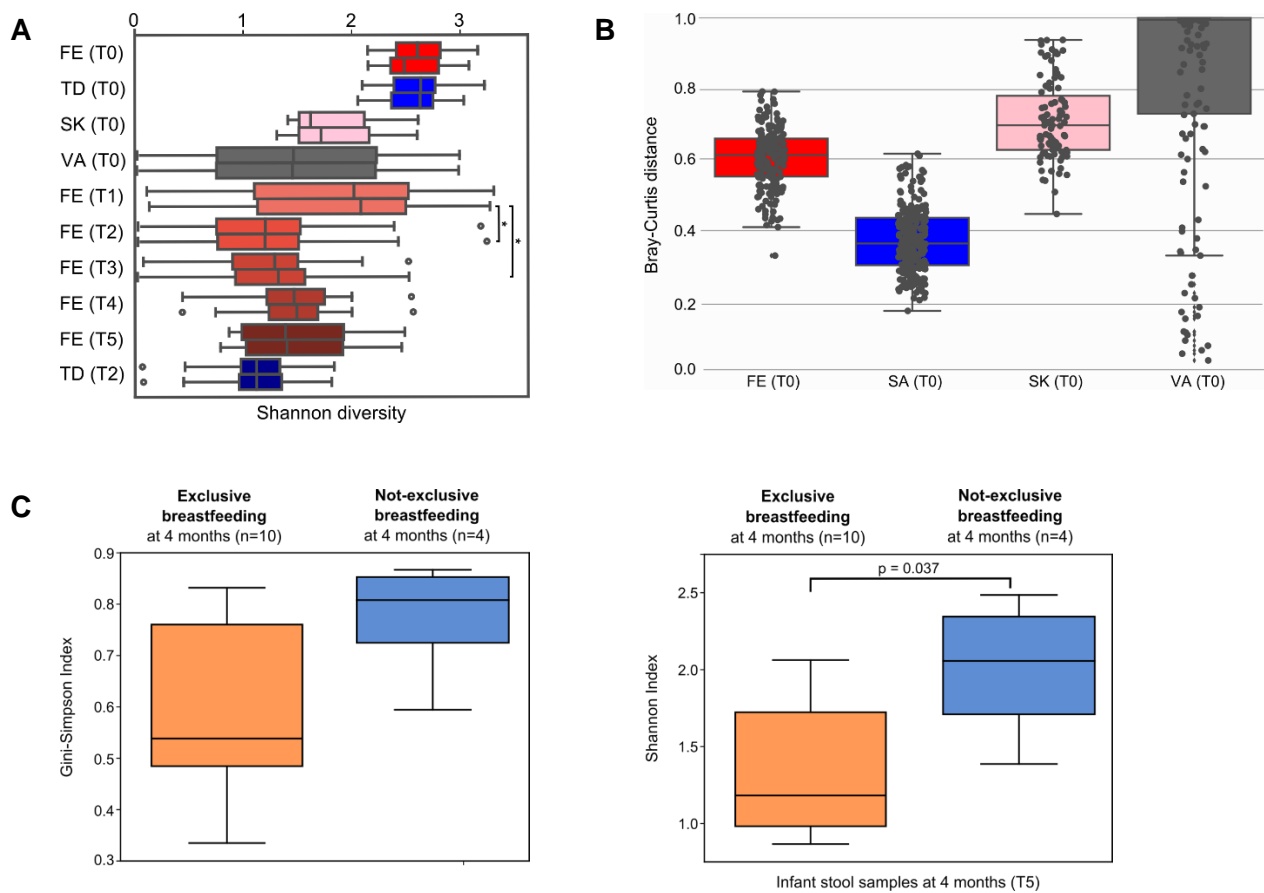
Figure S2. Related to Figure 1. Diversity measures of the sequenced metagenomes. (A) Comparison between standard (upper boxplot) and rarefied (lower boxplot) Shannon alpha diversity for each timepoint and body site. Out of 216 samples, 159 were above the rarefied threshold of 5 million bases. All the tongue dorsum samples at day one (T1) and 5 at day three (T2), 5 stool samples at day one (T1), 1 at one week (T3), 1 at 4 months (T5) and some vaginal, skin and tongue dorsum swabs from the mother (16, 12 and 1 respectively) were discarded in the process. We calculated the Shannon alpha diversity on the original data without rarefaction (upper boxplot) and the rarefied Shannon alpha diversity (lower boxplot) for each timepoint and body site. The * indicates a $p < 0.05$ (t-test). (**B). Inter-mother beta-diversity (Bray-Curtis on log-scaled relative abundances) for each maternal body site (stool, tongue dorsum, skin and vagina) at delivery.** The vaginal microbiome shows a higher inter-subject variability compared to the other maternal body sites, due to the presence of the different vaginal community types in the sampled population. **(C). Gini-Simpson and Shannon alpha diversity for faecal samples from exclusively breastfed and (partially or completely) formula-fed infants at 4 months.** Exclusive formula feeding refers to infants that switched from exclusive breastfeeding to a formula-only based diet between the first and the fourth month of age. Numbers of infants for each category are written in brackets (for 11 couples we did not have faecal samples at T5). Two-sided t-test was used to calculate p-values. Feeding mode data for every infant available in the **Table S1A.**
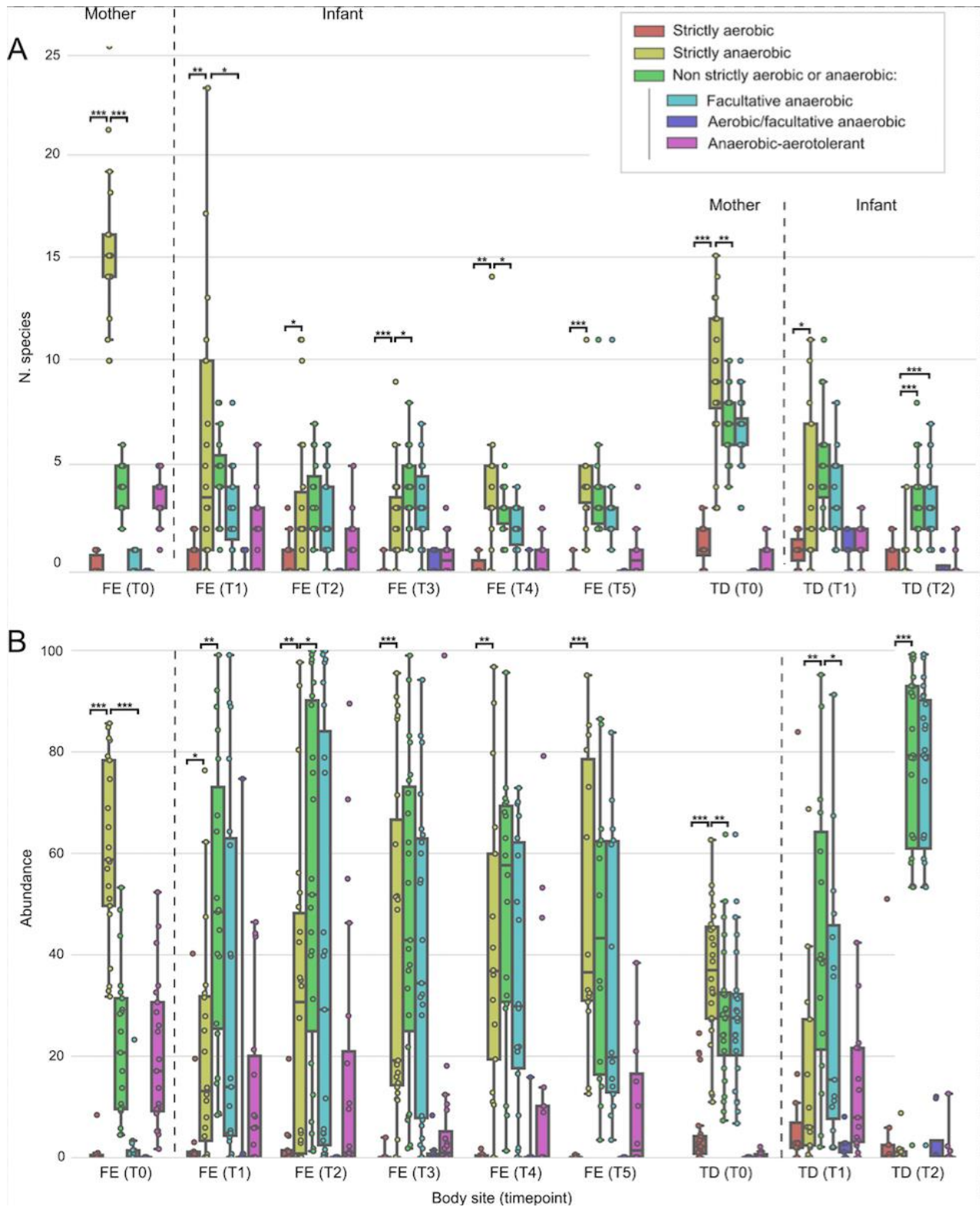
Figure S3. Related to Figure 1 and Figure 2. (A) Number of species and (B) their cumulative abundances present in the stool and oral samples, divided for their ability to utilize oxygen. Species not defined as "Unclassified" by MetaPhlAn2 and present with at least 0.5% of relative abundance were considered. The increase in strict anaerobes abundances, starting from T2 in the faecal samples, is mainly associated to *Bifidobacterium bifidum*, *Bifidobacterium breve*, *Bifidobacterium longum*, *Parabacteroides distasonis,* and *Veillonella parvula*. P-values referred to the aerotolerance classes within the same body site and timepoint are shown in the plot (*for P-values <0.05, ** for P-values <0.01 and *** for P-values <0.001, t-test).
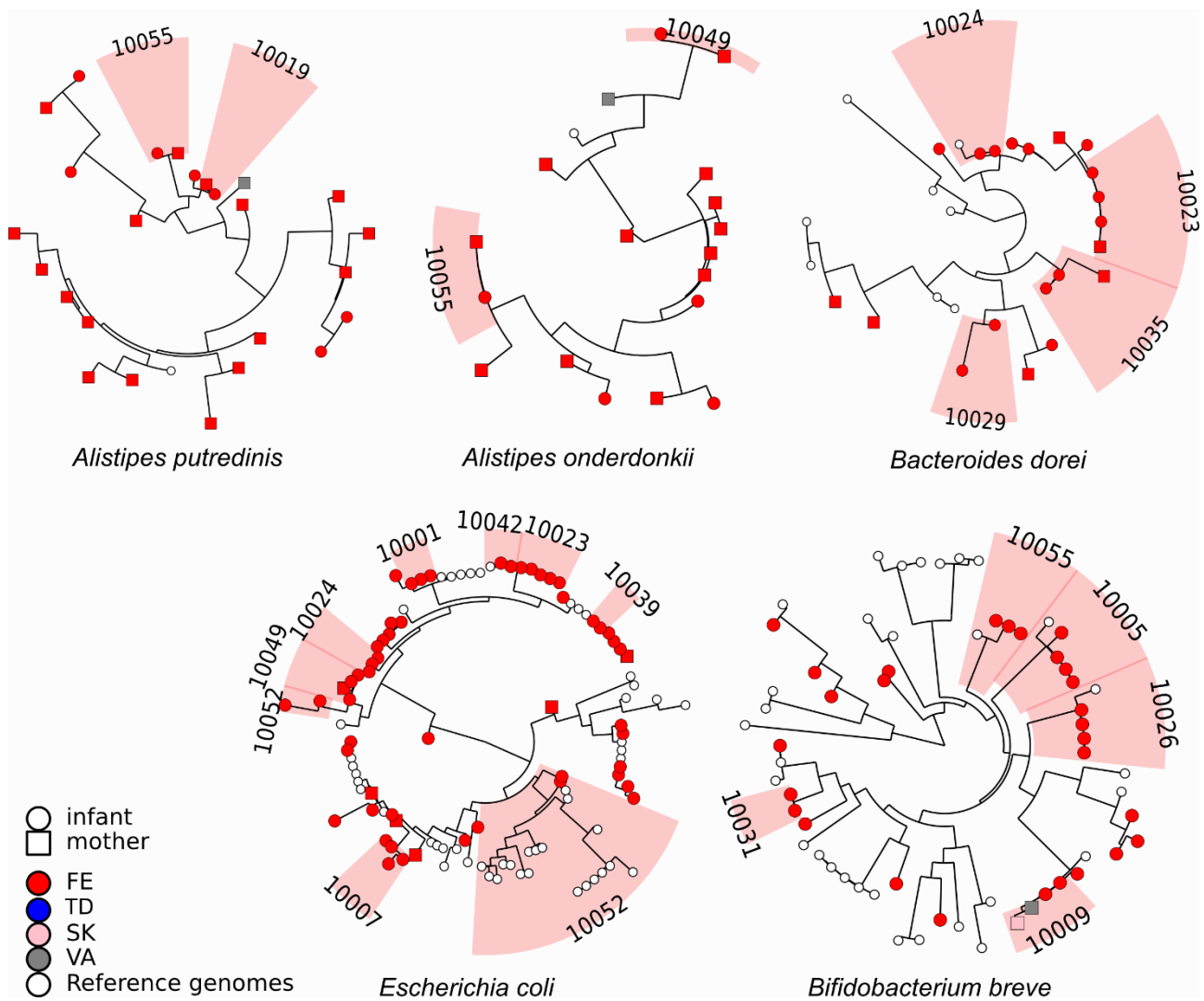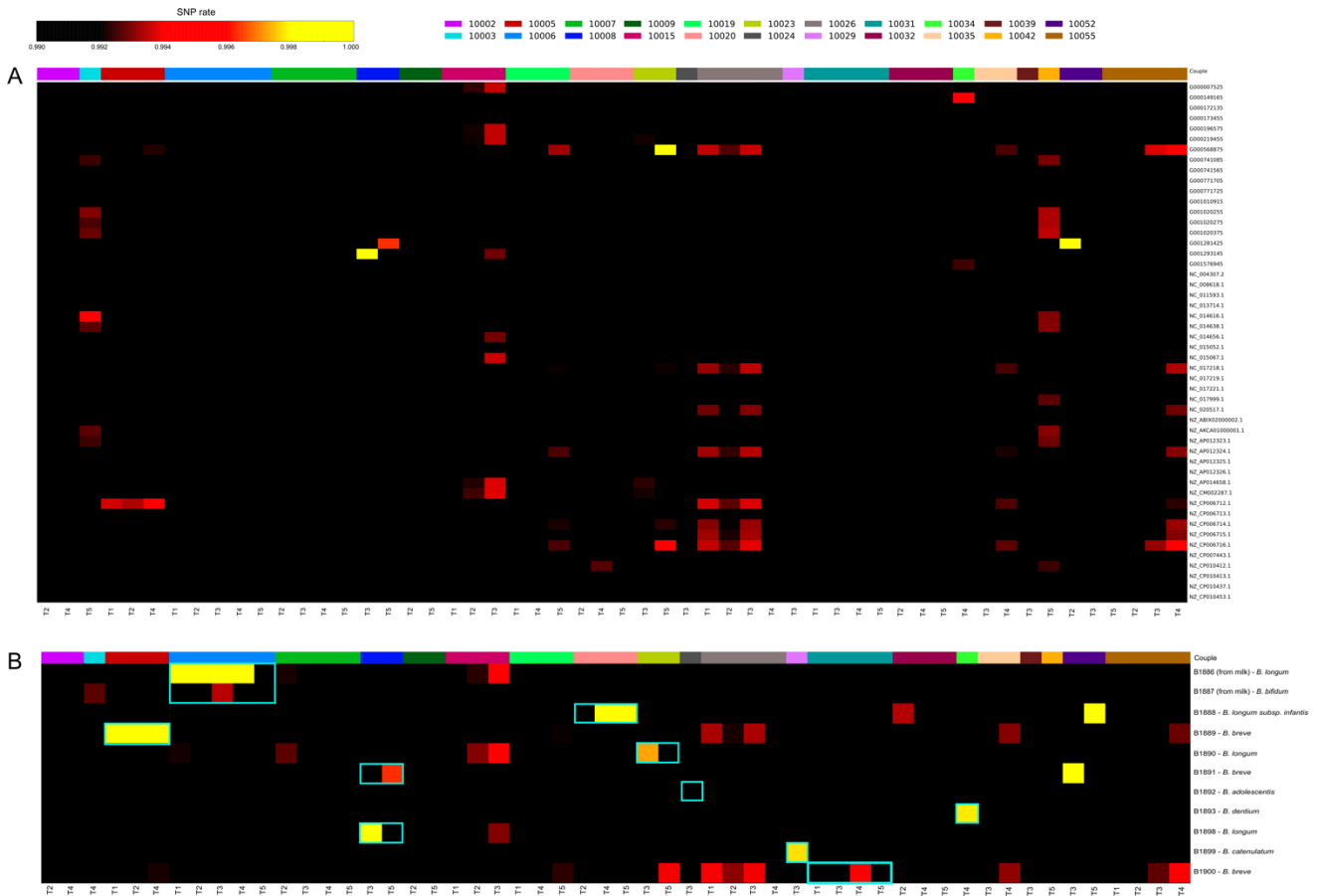
Figure S4. Related to Figure 3. Phylogenetic trees of vertically transmitted species in stool samples. Transmission between paired maternal- and infant- stool samples (represented by squares and circles, respectively) are highlighted with red shaded areas (pair ID is also shown).
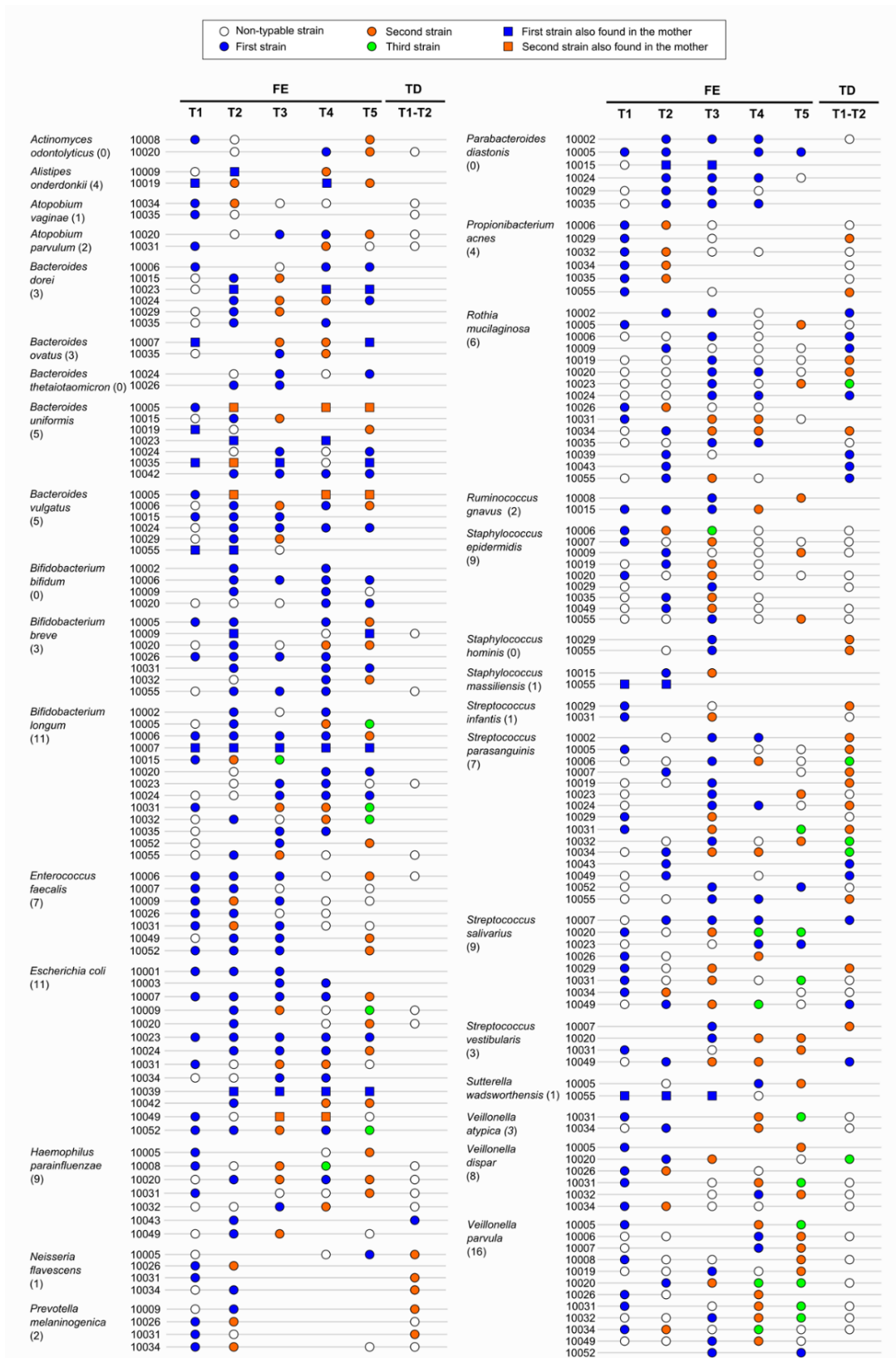
Figure S5. Related to Figure 3. Comparison of metagenomically inferred *Bifidobacterium* strains and Bifidobacteria cultured for a subset of the mother-infant pairs. Panel (**A**) shows the genetic distance between publically available reference genomes and the infant metagenomes. Panel (**B**) shows the genetic distance of the *Bifidobacterium* strains isolated from some of the milk and stool samples of this cohort identified in all the infant stool metagenomes. Cyan squares highlight the infant-mother pairs where the strain was isolated from. Two strains, *B. longum* (B1886) and *B. bifidum* (B1887), were isolated from a single milk sample (pair 10006). The *B. longum* (B1886) isolate was identified in the corresponding infant at four timepoints (>99.99% identity) and the *B. bifidum* (B1887) at one timepoint (>99.5% identity), validating the strain identification directly from metagenomes and the breast milk as a source of infant acquired microbes. Further, nine *Bifidobacterium* strains (belonging to *B. longum*, *B. breve*, *B. adolescentis*, *B. bifidum*, *B. catenulatum,* and *B. dentium* species) were isolated from faecal samples of eight infants. The isolates were identified in infants from eight different pairs, at >99.0% identity, validating again the possibility of identifying strains directly from metagenomes.

Figure S6. Related to Figure 4. Full list of strain replacement events in the infant body sites. Species with at least two pairs are shown. Empty circle present when only the identification at the species-level was available, i.e. the strain reconstruction was not possible (non-typable strain). Missing circle when the species is not present in the sample. In brackets, the number of replacement events per species (only infants stool samples are considered). In total, we identified 136 replacement events (on average 4.7 events per species and 5.4 per pair).