Metagenomics and development of the gut microbiota in infants

Yvonne Vallès¹, María José Gosalbes¹,², Lisbeth E. de Vries³,⁴, Juan José Abellán¹,² and M. Pilar Francino¹,⁵*

¹ Unitat Mixta d’Investigació en Genòmica i Salut, Centre Superior d’Investigació en Salut Pública/UVEG-Institut Cavanilles, Avda. de Catalunya 21, Valencia 46020, Spain
² CIBER en Epidemiología y Salud Pública (CIBEResp), Spain
³ Department of Veterinary Disease Biology, University of Copenhagen, Frederiksberg, Denmark
⁴ National Food Institute, Technical University of Denmark, Lyngby, Denmark
⁵ School of Natural Sciences, University of California Merced, CA, USA

* Corresponding author: francino_pil@gva.es

Abstract

The establishment of a balanced intestinal microbiota is essential for numerous aspects of human health, yet the microbial colonization of the gastrointestinal tract (GIT) of infants is both complex and highly variable among individuals. In addition, the GIT microbiota is often exposed to antibiotics, and may be an important reservoir of resistant strains and of transferable resistance genes since early infancy. We are investigating by means of diverse metagenomic approaches several areas of microbiota development in infants, including the deployment of functional capabilities at the community level, the presence of antibiotic resistances and the population dynamics of the most abundant genera.

Keywords: intestinal microbiota, antibiotics, metagenomics, feces, meconium
Although essential for health, the development of the gastrointestinal microbiota is poorly understood. 16S rDNA analyses have established that, throughout the first year of life, the GIT microbiota changes continuously until reaching an adult-like composition at roughly 12 months, but the process is highly variable among individuals [1]. Vaishampayan et al. [2] performed a case study where fecal samples from a mother and her infant at 1 and 11 months after delivery were utilized to construct 4 large-insert metagenomic libraries, one for each individual at each time point. This approach enabled the taxonomic and functional characterization of the microbiota in the samples. Although all samples shared the same phyla, there were large differences in composition at the genus level between mother and infant as well as between time points within individuals. Regardless, by one month of age the infant had already acquired a functional gene repertoire broadly analogous to that of the mother. From a population dynamics perspective, maternally inherited populations detected in the infant at one month had been lost by 11 months, suggesting that early colonizers can be easily replaced by externally acquired species. Such dynamics limit the potential for development of long-term coadaptations between specific bacterial and host genotypes. Rather, an intermittent pattern of interactions between different strains and human genotypes is likely to result in a diffuse process of coevolution among all interacting partners [2].

**Potential for maternal transmission of the GIT microbiota**

Nevertheless, the presence of maternal GIT phylotypes in the young infant raises the possibility of direct vertical transmission of GIT bacteria. How may such transmission take place? Infants acquire vaginal bacteria during their passage through the birth canal [3], but the vagina carries a limited diversity of bacteria with dominant genera distinct from those of the GIT [4]. However, during vaginal birth, infants are also exposed to maternal fecal microbes. In addition, although amniotic fluid and
meconium have traditionally been considered sterile under normal conditions, with bacterial colonization of the GIT starting at birth, recent studies have demonstrated the presence of bacteria in both environments even without rupture of membranes [5-8]. Furthermore, experiments with pregnant mice showed that a labelled bacterial strain administered orally could be recovered from meconium collected surgically from unborn fetuses [5]. Therefore, an internal route may exist enabling the transport of bacteria from the maternal to the fetal GIT. Such a route might invoke transport of maternal GIT bacteria through the bloodstream to the placenta from where they could reach the amniotic fluid and be swallowed by the fetus [5, 6]. The transport of GIT bacteria could be further facilitated by cells of the immune system, such as dendritic cells, that are capable of penetrating the gut epithelium and taking up bacteria from the lumen [9].

Few studies have employed culture-independent molecular analyses to investigate the potential bacterial diversity in meconium [3, 10]. We are performing 16S rDNA analyses of meconium samples from healthy babies, using high-throughput pyrosequencing, capable of detecting and identifying low frequencies of most taxa. In an attempt to separate the organisms collected during meconium formation from those added as the meconium was being passed, we are dividing our samples into internal and external portions. Our preliminary results with cloned 16S rDNA show similar bacterial compositions for both portions, although the internal one seems to contain a larger number of taxa. The most abundant bacterial families detected, such as the Enterobacteriaceae and the Streptococcaceae (Fig. 1), are common in infant feces, but not among the bacteria found on the skin of newborns [3]. These observations argue against external contamination dominating the bacterial composition of our meconium samples and rather suggest an internal origin of the detected bacteria.
**Antibiotic resistance in the GIT microbiota of infants**

From the functional genomics perspective, the availability of large-insert libraries facilitates the characterization of the genes and pathways present in the GIT microbiota. We have started this task by analyzing the distribution of genes related to antibiotic resistance in the libraries described in [2]. We have tested the libraries on ten different antibiotics and obtained clones resistant to every one, including many with multiple resistances, confirming that the GIT microbiota is an important reservoir of resistance genes. We are now studying the diversity of genes encoding the detected antibiotic resistances, identifying the bacteria that carry these genes, investigating potential horizontal gene transfers among them, and evaluating the occurrence of resistance transmission from mother to infant. To these aims, we are employing PCR screening and fosmid end-sequencing of resistant clones, and obtaining complete sequences for a number of resistance-carrying fosmids. These analyses are uncovering large numbers of different resistance genes, which may or not be shared between mother and infant, indicating that, by one month of age, the infant has probably acquired resistances both through maternal transmission and from the immediate environment. On the other hand, identical sequences of resistance genes are detected in different bacterial genera, families, and even phyla, suggesting that transfers of resistance genes can indeed occur among distantly related bacteria coinhabiting the GIT of an individual [11].

**Compositional and functional development of the GIT microbiota in cohorts of infants**

Although fosmid libraries are generating substantial insight into the dynamics of GIT bacterial populations, this methodology is prohibitive in labor and cost for the monitoring of microbiota development in numerous individuals. However, given the large inter-individual variability in this process [1], unveiling its general patterns will
require analysis of prospective cohorts with large numbers of infants. Such studies will enable us to address basic ecological questions regarding successional processes. In fact, the GIT microbial dynamics highlight an omnipresent ecological conundrum: both stochasticity and niche-driven succession are likely to play important roles during the onset and development of a community, but what is their relative contribution? And how does this relation change over time? In GIT microbiota development, does the microbiota of the mother play any role in setting up the stage for the infant’s? To address these questions we are collecting fecal samples from healthy Mother-Infant Pairs (MIPs) throughout the first year of the infant’s life and taking a metagenomics approach by direct pyrosequencing of bacterial DNA, cDNA, and amplicons from 16S and single copy protein-coding genes. We present here the metagenomic DNA analyses of the mother’s sample and the first 3 time points from the infant (at 1 week, 1 month and 3 months) for 4 different MIPs. The infants, except for one, were born vaginally and all have been almost exclusively breast-fed, although two of them ingested a bottle of formula shortly after birth.

Table 1 shows the numbers of phyla and genera identified in each sample, as well as the estimated taxon richness and the Shannon diversity index. Chao’s estimator shows that saturation has been reached in most of the samples, and the Shannon index indicates a large degree of unevenness in taxa abundance within samples at both taxonomic levels. Fig. 2A illustrates the uneven frequencies of genera within each sample, and the large variability existing across samples. At the level of phyla, although all samples share the same main groups, their relative frequencies also vary widely. Despite these diverse distributions, can any pattern of similarity be recognized? Because mother and infant are genetically and environmentally linked, one could expect that the samples within one MIP would be more alike; on the other hand, infants of the same age could share similar patterns of microbiota defined by similar physiological requirements. However, sample clustering
reveals no systematic pattern. Interestingly, in contrast to all other MIPs, all four MIP03 samples do group in the same cluster. Also, excepting M03, all other mother samples group separately from infants in a single cluster, but infant samples from MIPs 1, 2 and 6 do not group together, neither do they form clusters corresponding to the different age groups (Fig. 2A).

Contrary to the large differences in taxonomic composition, Fig. 2B shows substantial concordance across samples at the level of function. However, although taxonomy and function give strikingly different results in terms of sample variability, the dendrograms clustering samples according to each of these sequence classifications share several features (Fig. 2). According to function, samples from MIP03 appear again in the same cluster (although this cluster also contains M02-I1); mothers M01, M02 and M06 group separately from infants in a single cluster; and no clusters are recovered corresponding to individual infants or to infant age groups (Fig. 2B). The clustering of most maternal samples in both taxonomy and function suggests that the adult GIT contains a microbiota that is distinct in composition and gene repertoire from those present in infants throughout their first trimester of life, in accordance with a previous study [12]. Interestingly, the mother and infant in MIP03 were the only ones to have taken antibiotics, and this is the only MIP for which samples cluster together at the taxonomic and functional level. This pattern indicates that antibiotic treatment alters microbial colonization in the infant and that it may preferentially affect organisms encoding specific types of functions. This leads us to a final set of questions that we will address as we gather data for more time points and larger numbers of MIPs. How are gene functions partitioned among taxa in the developing microbiota of the GIT? Can we detect functionally equivalent taxa based on the functions assigned to their sequence reads and on their patterns of appearance, abundance and dependence on other taxa or on environmental variables during microbiota development? And, if equivalent taxa can be defined, are
their dynamics purely stochastic during colonization of the GIT, as predicted by the neutral theory of biodiversity [13]? Given the pattern of functional conservation in the face of taxonomical variability already observed in several analyses [2, 14-16], we anticipate that neutral dynamics will be apparent within functional equivalence classes. This would imply that groups of bacterial species share the same functional niches in the GIT, rather than each species occupying a small portion of a finely differentiated habitat. In that case, the number of specific ecological niches within the GIT that effectively guide microbial succession during colonization could be far below the number of species that inhabit the GIT of a single individual.

Acknowledgements

We thank Ferran Ballester and the INMA project (Infancia y Medio Ambiente) for providing meconium samples, Loreto Ferrús for her laboratory work, Nuria Jiménez, Llúcia Martínez and Alejandro Artacho for pyrosequencing and bioinformatic processing of sequence data at CSISP, and Feng Chen and his Sequencing Technology group for pyrosequencing and data processing at the DOE Joint Genome Institute. Financial support was received from the US National Institutes of Health (project R01DK066288) and the Spanish MICINN (project SAF2009-13032-C02-02 and project CSD2009-00006 of the CONSOLIDER program).

Transparency Declaration

The authors declare no conflicts of interest.
References


Table 1. Information concerning age at which the sample was taken, antibiotic use and infant's diet, and taxon richness and diversity per collected sample reported at phylum/genus level (p/g). N: number of taxa identified in each sample.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Age</th>
<th>Antibiotic (Mother)</th>
<th>Antibiotic (Infant)</th>
<th>Diet</th>
<th>N (p/g)</th>
<th>Chao1 taxon richness (p/g)</th>
<th>Shannon diversity (p/g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M01</td>
<td>30 years</td>
<td>-</td>
<td>-</td>
<td>Breast Milk</td>
<td>16</td>
<td>237</td>
<td>16.0</td>
</tr>
<tr>
<td>M01I1</td>
<td>1 week</td>
<td>-</td>
<td>-</td>
<td>Breast Milk</td>
<td>10</td>
<td>119</td>
<td>11.5</td>
</tr>
<tr>
<td>M01I2</td>
<td>1 month</td>
<td>-</td>
<td>-</td>
<td>Breast Milk</td>
<td>13</td>
<td>143</td>
<td>13.0</td>
</tr>
<tr>
<td>M01I3</td>
<td>3 months</td>
<td>-</td>
<td>-</td>
<td>Breast Milk</td>
<td>16</td>
<td>199</td>
<td>16.0</td>
</tr>
<tr>
<td>M02</td>
<td>36 years</td>
<td>-</td>
<td>-</td>
<td>Breast Milk</td>
<td>16</td>
<td>239</td>
<td>16.0</td>
</tr>
<tr>
<td>M02I1</td>
<td>1 week</td>
<td>-</td>
<td>-</td>
<td>Mixed</td>
<td>14</td>
<td>149</td>
<td>11.5</td>
</tr>
<tr>
<td>M02I2</td>
<td>1 month</td>
<td>-</td>
<td>-</td>
<td>Breast Milk</td>
<td>15</td>
<td>139</td>
<td>15.0</td>
</tr>
<tr>
<td>M02I3</td>
<td>3 months</td>
<td>-</td>
<td>-</td>
<td>Breast Milk</td>
<td>13</td>
<td>159</td>
<td>13.0</td>
</tr>
<tr>
<td>M03</td>
<td>31 years</td>
<td>-</td>
<td>-</td>
<td>Breast Milk</td>
<td>16</td>
<td>228</td>
<td>16.0</td>
</tr>
<tr>
<td>M03I1</td>
<td>1 week</td>
<td>Amoxicillin</td>
<td>Oftalmowell</td>
<td>Mixed</td>
<td>14</td>
<td>164</td>
<td>14.0</td>
</tr>
<tr>
<td>M03I2</td>
<td>1 month</td>
<td>Amoxicillin</td>
<td>-</td>
<td>Breast Milk</td>
<td>15</td>
<td>139</td>
<td>15.0</td>
</tr>
<tr>
<td>M03I3</td>
<td>3 months</td>
<td>Cefuroxima</td>
<td>-</td>
<td>Breast Milk</td>
<td>13</td>
<td>141</td>
<td>13.0</td>
</tr>
<tr>
<td>M06</td>
<td>42 years</td>
<td>-</td>
<td>-</td>
<td>Breast Milk</td>
<td>15</td>
<td>149</td>
<td>16.0</td>
</tr>
<tr>
<td>M06I1</td>
<td>1 week</td>
<td>-</td>
<td>-</td>
<td>Breast Milk</td>
<td>15</td>
<td>172</td>
<td>15.0</td>
</tr>
<tr>
<td>M06I2</td>
<td>1 month</td>
<td>-</td>
<td>-</td>
<td>Breast Milk</td>
<td>15</td>
<td>159</td>
<td>18.0</td>
</tr>
</tbody>
</table>
Figure legends.

Fig. 1. Histogram depicting the distribution of bacterial taxa of two meconium samples that were each separated into two parts: the external layers of the meconium (ME) and the internal part of the meconium (MI).

Fig. 2. Heatmaps and sample clustering based on taxon (A) and function (B) frequencies detected in DNA metagenomic analyses. (A) Taxonomic assignments at genus level; only genera that have a proportion greater than 1% in at least one sample are depicted. (B) Functional assignments based on the SEED functional classification [17]. Both taxonomic and functional assignments were obtained in the analysis platform MG-RAST (Meta Genome Rapid Annotation using Subsystem Technology; http://metagenomics.nmpdr.org/). Heatmaps and clustering were obtained in R [18]. Clustering was based on the Bray-Curtis distance measure as implemented in the R package VEGAN. Colors in the figure depict the percentage range of sequences assigned to each genus or functional category.
Fig. 1
Fig. 2

(A) Thermotoga, Thermomicrobium, Streptomyces, Staphylococcus, Shigella, Salmonella, Pseudomonas, Parabacteroides, Mycobacterium, Listeria, Lactobacillus, Kineococcus, Janibacter, Geobacillus, Flavobacterium, Escherichia, Enterococcus, Enterobacter, Desulfovibrio, Desulfobacter, Corynebacterium, Clostridium, Clostridium, Caldolobacter, Caldocellum, Rhodococcus, Bacillus, Actinobacteria.


Legend: <1%, 1–5%, 5–10%, 10–25%, >25%