UNIVERSITAT POLITÈCNICA DE VALÈNCIA

ESCOLA TÈCNICA SUPERIOR D´ENGINYERIA AGRONÒMICA I DEL MEDI NATURAL

METAGENOMIC STUDY OF THE HUMAN GUT VIROME DURING THE FIRST YEAR OF LIFE

Final Degree Project

Bachelor's Degree in Biotechnology

Academic Year: 2013-2014

Valencia, September 2014

Author: David Mayo Muñoz
Tutor: Santiago Vilanova Navarro
Co-tutor: María Pilar Francino Puget

Metagenomic study of the human gut virome during the first year of life

Author: David Mayo Muñoz
Tutor: Dr. Santiago Vilanova Navarro
Co-tutor: Dra. Maria Pilar Francino Puget

Valencia, September 2014

ABSTRACT

Despite the important role of the microbiota in the human gastrointestinal tract (GIT) and its impact on life-long health, the successional process through which this microbial community develops during infancy is still poorly understood. Specially, little is known about how the amount and type of viruses present in the GIT, i. e. the virome, varies throughout this period and about the role this collection of viruses may play in the assembly of the GIT microbiota. In this study, the patterns of taxonomic change of the GIT viral community were analyzed in a birth cohort of healthy infants during the first year of life.

Fecal samples were collected at different timepoints for the infants and their mothers. Viral DNA was isolated, randomly amplified using phi29 polymerase and sequenced through high-throughput 454 pyrosequencing. Then, the obtained viral sequences were taxonomically assigned through bioinformatic analyses to establish the composition of the virome present in each sample. Finally, statistical analyses were performed in order to assess the process of viral dynamics through time.

Through the analysis of taxonomic richness and diversity, as well as clustering based on viral composition and abundance, it has been observed that individual instances of viral dynamics in the gut are distinct, and clear patterns have not been identified. However, general trends have been observed in virome development, such as the decrease of viral taxonomic richness and diversity during the first year of life, contrary to the behaviour of the bacterial microbiota, and a directionality of change through time towards the taxonomic composition of the maternal virome.

Key words:

Virome, GIT microbiota, taxonomic richness, taxonomic diversity.

Estudio metagenómico del viroma instestinal humano en el primer año de vida

Autor: David Mayo Muñoz
Tutor: Dr. Santiago Vilanova Navarro
Cotutor: Dra. Maria Pilar Francino Puget

Valencia, Septiembre 2014

RESUMEN

A pesar de la importante función de la microbiota en el tracto gastrointestinal y su impacto en la salud durante toda la vida, el proceso de sucesión a través del cual se desarrolla esta comunidad microbiana durante la infancia es aún poco conocido. Especialmente, se sabe poco sobre como la cantidad y tipo de virus, es decir, el viroma, varía a través de este periodo y sobre la función que este conjunto de virus puede desarrollar en el ensamblaje de la microbiota del tracto gastrointestinal. En este estudio se analizaron los patrones de cambio taxonómico en la comunidad viral del tracto gastrointestinal en una cohorte de niños sanos durante el primer año de vida.

Para ello, se utilizaron muestras fecales que fueron recogidas a diferentes tiempos para los niños y sus madres. El ADN viral fue aislado, amplificado aleatoriamente utilizando la polimerasa phi29 y secuenciado por pirosecuenciación 454. A continuación, las secuencias virales obtenidas fueron taxonómicamente asignadas por medio de análisis bioinformáticos para establecer la composición del viroma presente en cada muestra. Finalmente, se realizaron análisis estadísticos para evaluar el proceso de la dinámica viral a través del tiempo.

A través del análisis tanto de la riqueza y diversidad taxonómicas como del clustering de muestras en base a la composición y la abundancia viral, se ha observado que la dinámica viral en el intestino es distinta en cada individuo y no se han podido identificar patrones claros. Sin embargo, se han observado tendencias generales en el desarrollo del viroma, como el decrecimiento de la riqueza y la diversidad taxonómicas durante el primer año de vida, contrariamente al comportamiento de la microbiota bacteriana, y un cambio direccional a través del tiempo hacia la composición taxonómica del viroma materno.

Palabras clave:

Viroma, microbiota del tracto gastrointestinal, riqueza taxonómica, diversidad taxonómica.

ACKNOWLEDGEMENTS / AGRADECIMIENTOS:

Este Trabajo Final de Grado no habría sido podido ser realizado sin la ayuda de ciertas personas a las que quiero hacer mención y expresar mi agradecimiento.

En primer lugar, a mi tutora en el CSISP, Maria Pilar Francino, por haberme ofrecido este proyecto en el cual he aprendido tanto y que ha sido mi primera experiencia en el mundo de la investigación. Estoy muy agradecido por toda tu ayuda y atención, has sido muy buena tutora conmigo.

Mi gran gratitud y aprecio a Yvonne Vallés, quien me dio las mejores sugerencias de qué hacer en cada momento del presente estudio y de la que tanto he aprendido. Echaré en falta tu optimismo y entusiasmo en el trabajo. Te deseo lo mejor en esta nueva etapa que empiezas.

También quiero expresar mi agradecimiento a Vicente Pérez, quien me enseñó cómo manejarme en el laboratorio y que no ha dudado en resolverme toda cuestión que me ha surgido. Da gusto trabajar con personas como tú. También quiero recordar a todos aquellos en el laboratorio que me han echado una mano siempre que lo he necesitado.

A Santiago Vilanova, por haber accedido a ser mi tutor y recibirme en cualquier momento que se lo pidiera.

Quiero recordar a mi gran amigo Rafa Pinilla, quien me ayudó a entrar en el CSISP y me ha acompañado en algunos ratos de laboratorio. Gracias por estar a mi lado durante este último curso. Te voy a echar mucho de menos, no sabes cuánto.

A todas aquellas personas que se han acordado de mí en sus oraciones y las que me han alentado con sus ánimos durante este difícil momento. Todo ello me ha reconfortado de verdad.

Y en especial, a mis padres por su amor incondicional y por su apoyo en todo momento, por haberme educado para ser hoy quien soy, y a mis hermanos por su infinita paciencia conmigo. Este Trabajo Final de Grado va dedicado a ellos, y también a mis abuelos.

INDEX

FIGURE INDEX

INDEX OF TABLES

ABBREVIATIONS:

Bp: Base Pairs

BLAST: Basic Local Alignment Search Tool

CCA: Canonical Correspondence Analysis

DNA: Deoxyribonucleic acid

EDTA: Ethylenediaminetetraacetic acid

GIT: Gastro Intestinal Tract

HGT: Horizontal Genetic Transfer

MDA: Multiple Displacement Amplification

MIP: Mother-infant pairs

PBS: Phosphate Buffer Saline

PCR: Polymetase Chain Reaction

RNA: Deoxyribonucleic acid

rpm: Revolutions per minute

TBE: Tris-Borate-EDTA

TE: Tris-EDTA

TRIS: Tris(hydroxymethyl)aminomethane

UV: Ultraviolet

1. INTRODUCTION

1.1. The human gut microbiota

1.1.1. General traits

The gut microbiota harbors a great number of microorganisms such as bacteria, archaea, viruses and eukaryotic communities coexisting and maintaining homeostasis in such a complex ecosystem. It plays essential roles in gut maturation, digestion, nutrition, angiogenesis, immune system modulation, protection from pathogens and metabolism, contributing to the complementation of metabolic features not coded by the human genome (Sekirov et al., 2010; Collado et al., 2013).

The microbiota colonizes every surface of the body: the skin and the genitourinary, gastrointestinal and respiratory tracts. The gastrointestinal tract (GIT) is the most heavily colonized mainly because it is a large organ, so it has a bigger surface for microbial colonization and it is rich in nutrients for the microbes (Sekirov et al., 2010). The intestinal microbiota is composed by many different species and it is not homogeneous, but rather varies along the different parts of the GIT (Sekirov et al., 2010). The organisms of the microbiota greatly exceed the number of human body cells (Whitman et al., 1998), however, only a small fraction of the microbial biodiversity is known.

1.1.2. Factors influencing the GIT microbial community

The gastrointestinal tract of a normal fetus is not sterile, which has been proven by analyzing the meconium, the newborn's first intestinal discharge (Gosalbes et al., 2013). In a study by Breitbart et al. (2008), it was observed that the most abundant fecal viral sequences did not originate from breast-milk or formula, suggesting a non-dietary initial source of viruses. At birth, the composition of the human gut immediately changes, when infants are exposed to a complex microbial population from the mother, particularly vaginal and fecal microbiomes, and surrounding environment (Mackie et al., 1999). It makes sense that similarities have been observed between the intestinal microbiota of infants and the vaginal microbiota of their mothers (Mandar and Mikelsaar, 1996). Moreover, infants delivered through Cesarean section show less similarities with the vaginal microbiota of their mothers compared with vaginally delivered infants (Domínguez-Bello et al., 2010).

However, although the maternal microbiota and mode of delivery seem to be a major factor in shaping our gut microbial community, there are many factors, such as type of feeding during the first months of life (breast milk or formula), disease, antibiotic treatment, human physiology, genetic endowment and characteristics of the individual's gut environment (pH, immune response) that are closely linked to the maintenance of or to compositional changes in the gut microbiota (Turnbaugh et al., 2007; Thompson-Chagoyan et al., 2007; Dicksved et al., 2007; Penders et al., 2006). All of these incomes and their relations influence in shaping the infant's gut microbiota

and immune system, determining the individual's health and propensity to immune diseases (Gosalbes et al., 2013).

1.1.3. Influence of diet

The type of feeding may interfere greatly in the regulation of the intestinal microbiota. The study of Bezirtzoglou et al. (2011) was developed to determine the bacterial composition in fecal samples of newborn infants under different types of feeding and found differences in fecal microbiota between breast-fed infants and formula-fed infants. Moreover, a shift toward bacterial consortia characteristic of the adult microbiota is observed with the introduction of solid foods (Palmer et al., 2007). On the other hand, the microbiome in the first months is enriched in genes that facilitate lactate utilization, and, before the introduction of solid food, functional genes involved in plant polysaccharide metabolism are present, preparing the infant gut for an adult diet (Koenig, 2011).

In the Minot et al. (2011) study, it was observed that in the viral population, the significant changes among individuals were detectably associated with switching to a defined diet. Although individuals eating the same food did not show identical viromes, the gut virome changed significantly during the change in diet by alteration of the proportions of pre-existing populations, so that subjects on the same diet showed more similar virome composition.

1.1.4. Variability among individuals

The microbial composition of the mammalian intestine exhibits large inter-individual variability, depending on internal and external factors, such as genetic factors, age, diet and health (Collado et al., 2013).

In spite of the variability among individuals, there are a number of physical, chemical and mechanical properties that are, in general, similar across individuals, such as temperature, pH and surface tension values bound within limited ranges (McConnell et al., 2008). In this way, a considerable degree of inter-individual convergence of GIT communities is expected as a response to common selective pressures. Hence, many studies have been developed for the detection of a taxonomic core that would be shared by all individuals (Turnbaugh et al., 2007; Qin et al., 2010; Jalanka-Tuovinen et al., 2011) and some of them have indicated the predominance of a core gut microbiota (Figure 1).

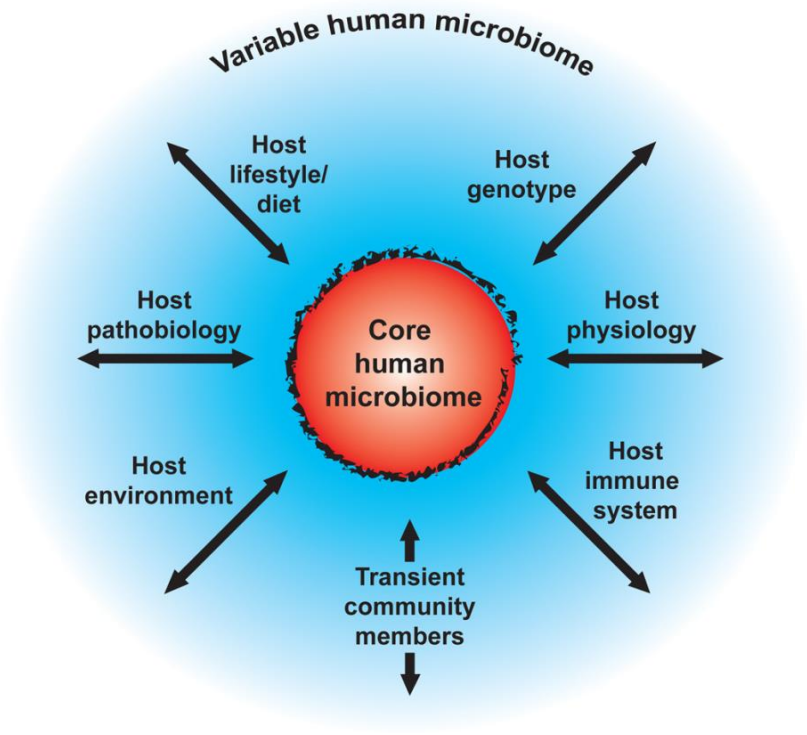


Figure 1. Core gut microbiota (Turnbaugh et al., 2007).

1.1.5. Variability with time. Successional process

This variability is also reflected with time, predominantly during the first year of life. Gut microbiota varies in temporal dynamics and there are parallelisms between functional and taxonomic changes. The taxonomic composition of the microbiota shapes the functional capacities of the community, and, consequently, successional variability with time may affect host physiology, metabolism and immunity.

Different studies have shown that the first stages of infant gut microbiota development are characterized by high levels of variability among individuals and a largely uneven distribution of taxa. Nevertheless, as infant development progresses, the intestinal microbiota begins to stabilize and to resemble that of a young adult with a more even taxa distribution (Palmer et al., 2007; Avershina et al., 2014; Mackie et al., 1999) (Figure 2).

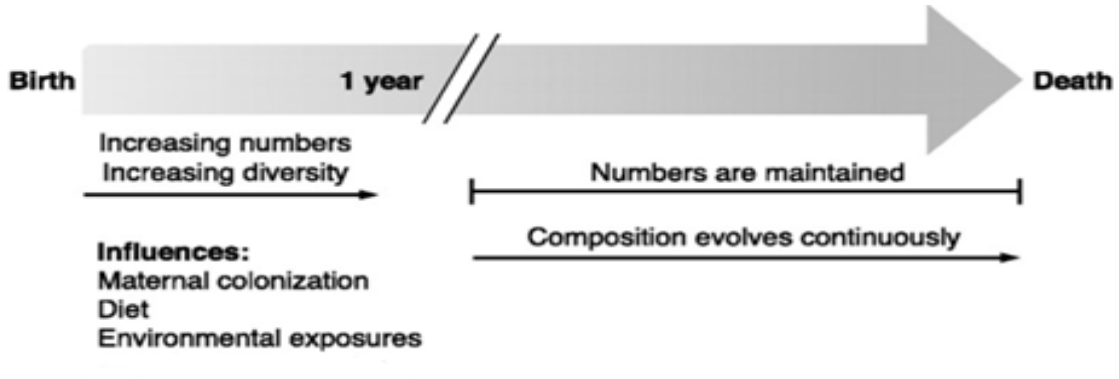


Figure 2. Succesional process of the gut microbiota towards the adult composition. (Sekirov et al., 2010).

1.2. The GIT viral community

1.2.1. General traits and variability

The viruses of the human gut remain largely unknown and poorly understood and, although the role of viruses in human health is very important, many studies have been rather focused in other environments, such as seawater or sediments (Breitbart et al., 2002; Breitbart et al., 2004; Venter et al., 2004; Angly et al., 2006; Bench et al., 2007; Schoenfeld et al., 2008; López-Bueno et al., 2009; Tamaki et al., 2011), and few studies about viral diversity in the human gut have been carried out (Kim et al., 2011; Reyes et al., 2010). Moreover, many of those studies are based on samples from a few individuals and use shotgun library construction with limited resolution (Breitbart et al., 2008; Breitbart et al., 2003; Zhang et al., 2006).

Analyzing and characterizing viral community dynamics in the environment is complicated due to the lack of adequate cultured-based methods, because viruses depend on the culturing of the host cells, and less than 1% of microbial hosts can be cultured in vitro (Pérez-Brocal et al., 2013). In addition, viruses lack conserved genetic elements that can be used for a taxonomic classification, like the 16S rRNA gene, commonly used for bacteria and archaea (Rohwer and Edwards, 2002). Therefore, alternative approaches are necessary to characterize the viral diversity and dynamics. Metagenomic approaches, such as high-throughput sequencing, are opening new ways of assessing the viral diversity in a wider range of environments, including the human GIT. In fact, our knowledge of viral diversity has been widely revolutionized in the last years thanks to these approaches. Metagenomic analyses, in combination with tangential flow filtration for viral particle isolation and concentration from large volume samples, have been employed to survey the DNA viruses in seawater (Breitbart et al., 2002) and from human feces (Breitbart et al., 2003).

Adults contain a large and diverse community of both DNA and RNA viruses (Breitbart et al., 2003; Zhang et al., 2006). In adult fecal samples, the majority of the DNA viruses are phages and the majority of RNA viruses are plant viruses. Animal viruses are uncommon in the human healthy GIT. The concentration of phages varies

among individuals and no correlations have been detected with age or sex (Breitbart et al., 2003; Breitbart et al., 2008).

While bacteria are enriched in genes for synthesis of amino acid and carbohydrate precursors, metabolism, ion transport, translational machinery and cell wall/membrane biogenesis, viruses are enriched in genes for replication, recombination and repair. Therefore, this confirms in the wild the long-held view based on experimental work in the laboratory that viruses recruit host cell machinery for translation, energy production, and synthesis of macromolecular precursors while using their own coding capacity to encode functions for replication (Minot et al., 2011).

Although viromes are exclusive to each individual independently of the degree of genetic relatedness, among other studies, Reyes et al. (2010) found a significantly higher degree of similarity in the fecal viral communities between twins and their mothers than in unrelated individuals, and stability of viral communities over time.

1.2.2. Ecological role of phages and potential for antibacterial therapy

Phages, the most abundant viruses found in the human gut, are thought to play important roles in the GIT and are also likely to do so in the gastrointestinal ecology of the developing infant. They contribute to human health, acting in concert with the immune system, defending the organism by controlling invading pathogens or helping control local inflammatory and autoimmune reaction and demonstrate anticancer activity (Gorski and Weber-Dabrowska, 2005). The frequent occurrence of horizontal genetic transfer (HGT) among gut microbes (Tamames and Moya, 2008) and the presence of a high number of phage-related genes from the human gut microbial metagenome implies a likely role of viruses in gut homeostasis (Kim et al., 2011). Moreover, they may influence food web processes because they have the capacity to act in nutrient cycling and carbon flow processes, regulating microbial communities through lytic and lysogenic replication through the release of cell lysis products (Weinbauer, 2004).

Phages have the potential to show a strong influence on the diversity, abundance and structure of bacterial communities, as they contribute to bacterial host mortality and genetic diversity. In the study of Reyes et al. (2010) characterizing fecal viromes, it was found that phage and bacterial abundance do not show cyclic changes of the dominant taxa in the human gut. This is different from the expected Lotka-Volterra predator-prey relationships (Bohannan and Lenski, 1997), known as “kill the winner” dynamics (Rodriguez-Valera et al., 2009), where there are episodes of outgrowth of particular bacterial species followed by blooms of their phage, observed in many other environments (Rodriguez-Brito et al., 2010).

Despite interpersonal variations in viromes, intrapersonal diversity is very low with viromes dominated by a few temperate phages that exhibit notable genetic stability (Reyes et al., 2010). So, a different viral life cycle is observed, where temperate (not lytic) viruses are contributors to microbial phenotypes through provision of adaptive genes. In fact, viruses, particularly bacteriophages, are one of the main drivers of the

evolutionary change of microorganisms through HGT (Jagadish et al., 1993; Hwang et al., 1994). This dynamic can change the metabolic features of bacteria and the lifestyles of pathogens and many of the differences between closely related microbial strains arise from prophage insertions. Further research is needed to clarify if the changes in phage abundance are merely a result of changes in abundance of their hosts, or if additional mechanisms are involved, although initial data indicates a possible contribution of lysogenic induction (Minot et al., 2011).

Phages have been used with a high efficacy in preventing and treating bacterial infections in animals and animal models of infection (Smith and Huggins, 1983; Hermoso et al., 2007). Phage therapy is an alternative to the antibiotic-based antimicrobial therapies, potentially useful against antibiotic-resistant pathogens. The use of phage for the control of infectious agents has advantages such as their ability to attack specifically target cells by hijacking the DNA replication machinery, replicate inside the cell and killing the host; the multiplying effect and their inability to attack mammalian host cells or tissue (Rea et al., 2013).

But, in contrast, they can affect health negatively by eliminating probiotic strains (Ventura et al., 2011) or introducing new phenotypic traits to the bacterial hosts, such as antibiotic resistance and the ability to produce potent toxins to make a pathogen more aggressively virulent (Davis and Waldor, 2002).

1.3. The importance of studies in infants

The alteration of functional diversity within the microbiota community can affect health status. Accordingly, many metagenomic studies have focused their efforts in the assessment of the functional diversity present in the GIT microbiota of healthy individuals, to detect potential deviations in individuals affected by different diseases (Collado et al., 2013; Turnbaugh et al., 2007). But, most of these studies have been developed on adult individuals, while the assessment on infants has remained underexplored, even though infancy is the critical period for gut microbiota assembly with strong repercussions for immune and metabolic development. Different studies highlight that the microbe-host interactions occurring during infancy represent a main determinant of life-long health or disease (Rautava et al., 2004; Penders et al., 2007).

In spite of its importance, the process of microbial succession in the gut of infants is still poorly understood. In addition, the gut microbiota in infants has been mostly explored in cross-sectional studies (Kim et al., 2011) and in a few longitudinal studies limited to a few infants (Reyes et al., 2010; Vallès et al., 2012). It is important to know that cross-sectional studies reflect the microbiota of different individuals in a single point and do not follow individuals through time, and, therefore, cannot evaluate the microbiota succession and dynamics. So, to understand the succession of the microbiome community in the human gut during infancy it is necessary to develop longitudinal studies that follow infants through a period of time. Through understanding the gut colonization patterns in healthy infants with time, we will be able to produce better definitions of the deviations in this process that result in microbiota imbalances and disease. Therefore, the development of strategies for the formation of

health-promoting microbiotas that could be maintained throughout the life of the host will be possible.

1.4. Our study

Taking into account all of the above, the present study explores the patterns of taxonomic change along time during GIT virome development in a birth cohort of 7 infants, to complement the study of Vallès et al. (2014), where the corresponding analysis with the bacterial gut community was developed. With this aim, a set of fecal samples from healthy infants throughout the first year of life has been used, and virus sequences have been obtained to characterize the phylogenetic composition and genetic repertoire of the virome present in each sample. Fecal samples were collected at five different timepoints of the infant: at one week, one month, three months (last breastfeeding timepoint), seven months (after introduction of solid foods) and one year after birth.

Moreover, the microbiota present in the mother before and one-year after childbirth was also collected and sequenced, in order to assess the progression of the infant's microbiota towards an adult-like state and to see the changes that occur within the mother during pregnancy.

We have also analyzed the diversity and structure of these viral communities in relation to those of the bacterial communities from the same samples, which were determined by direct metagenomic sequencing of total DNA in the study of Vallès et al. (2014).

2. OBJECTIVES

The main objective of this study is the analysis of the human gut virome dynamics during the first year of life by characterizing the patterns of taxonomic change along time during GIT virome development in a birth cohort of 7 infants. For this purpose, a series of partial objectives have been achieved:

1. The isolation and pyrosequencing of DNA viruses of the intestinal microbiota of the infants and their mothers.

2. Description of the gut viral community along development and of how these communities vary or not between individuals.

3. Comparison of the dynamics of the infant gut viral communities to that of the bacterial communities in the same samples.

3. MATERIAL AND METHODS

3.1. Infant cohort, sampling and sample preparation

The study has been developed from the samples used in the study of Vallès et al. (2014). In that study, 21 women were recruited, all residents of the city of Valencia, who were contacted during midwife visits, and 13 were retained throughout the length of the study. At the moment of delivery, these women were between 29 and 42 years of age and had not taken antibiotics in at least three months before the onset of labor. Infants' samples were collected at one week (I1), one month (I2), three months (I3, before introduction of solid foods), seven months (I4, after introduction of solid foods) and one year after birth (I5), and maternal samples were collected within one week prior to delivery (MA) and one year after (MB).

In addition to fecal samples, throughout the 12-months sampling period information was obtained regarding the infants' diet, general health and intake of antibiotics and other drugs, by means of specifically designed questionnaires that were given to the infants' parents. This information allowed us to establish that all infants remained healthy throughout most of the sampling period and that solid foods were introduced into their diets between the 3- and 7- months samplings, following patterns typical of Spanish Mediterranean infant diets (Capdevila et al., 1998).

The fecal samples were collected by the mothers and stored in home freezers until brought to the laboratory, where they were stored at -80ºC until processing. Samples were homogenized in a 50% RNA later/phosphate saline buffer solution and centrifuged for two minutes at 2000 rpm. Only the supernatant resulting from the latter spin was used for further processing.

The protocol used for the sample preparation and viral DNA-RNA extraction has been the one followed in the study of Pérez-Brocal et al. (2013) with slight modifications.

In order to optimize viral load, samples were centrifuged at least twice but not more than three times depending on the amount of solids in suspension, at 12000 rpm for 7.5 min at 4ºC. After each centrifugation, the supernatant was transferred to a new tube. The supernatants were then serially filtered through 0.45-µm and 0.20-µm pore-size cellulose acetate filters (Sartorius Stedim Biotech, Goettingen, Germany).

To remove unprotected nucleic acids of bacterial or eukaryotic origin, the filtered samples were further treated with a cocktail of DNAses/RNAses composed of 14 U of Turbo DNAse (Ambion, Austin, TX, USA), 20 U of Benzonase (Novagen, Inc., Madison, WI, USA), and 20 U of RNAse A (Invitrogen, Carlsbad, CA, USA) in DNAse buffer (Ambion), for 120 min at 37°C. Then, the cocktail was inactivated by adding EDTA (0.5 M, pH 8) and incubating for 10 min at 75ºC.

3.2. Viral DNA-RNA extractions

Intact nucleic acids contained in the viral capsids were extracted using the QiAamp viral RNA extraction kit (Qiagen), following the manufacturer's instructions for viral DNA/RNA extraction. During the extraction with this kit, elutions of DNA/RNA were obtained and stored at -80ºC to posterior DNA amplification. Finally, a portion of them was taken and incubated at 37 °C for 30 min, with RQ1 DNAse (Promega), to remove the DNA coextracted with the RNA. These elutions were also stored at -80ºC to a posterior retrotranscription to cDNA and sequencing. But in this study, we have focused on DNA viruses.

3.3. DNA amplification

The amplification of viral DNA for sequencing was carried out with a whole genome amplification strategy using the Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare, Amersham, UK), incubating for 3 h at 30 °C followed by phi29 DNA polymerase inactivation for 10 min at 65 °C.

This DNA amplification is non-PCR based and uses the Φ29 DNA polymerase, and random hexamer primers for unbiased whole genome amplification (WGA). This is a multiple displacement amplification (MDA) which is used for linear DNA molecules, such as genomic DNA (Silander and Saarela, 2008). The MDA method is based on isothermal amplification, and does not require changing the reaction temperature.

MDA provides many improvements in respect to the other amplification techniques, such as PCR. The Φ29 DNA polymerase has high processivity and fidelity, i.e. a lower error frequency (Paez JG et al, 2004), and this allows highly uniform amplification across the genome. Hence, it is possible to carry out the robust amplification of entire genomes from little amounts of DNA sample (Silander and Saarela, 2008). So this method has been used in WGA and holds the potential to allow large-scale studies, such as sequencing-based genetic studies, realised directly from primary sample sources, like genomes isolated from feces (Kim et al., 2011) (Minot et al., 2011), tissue cells (Pérez-Brocal et al., 2013) or blood (Hosono et al., 2003).

Following the GenomiPhi V2 protocol, DNA is briefly heat denatured, then cooled in a sample buffer containing random hexamers that anneal to the DNA. A master mix containing Phi 29 DNA polymerase, additional random hexamers, nucleotides, salts, and buffers is added. Isothermal amplification proceeds at 30°C, followed by heat inactivation of the polymerase at 65°C (Figure 3).

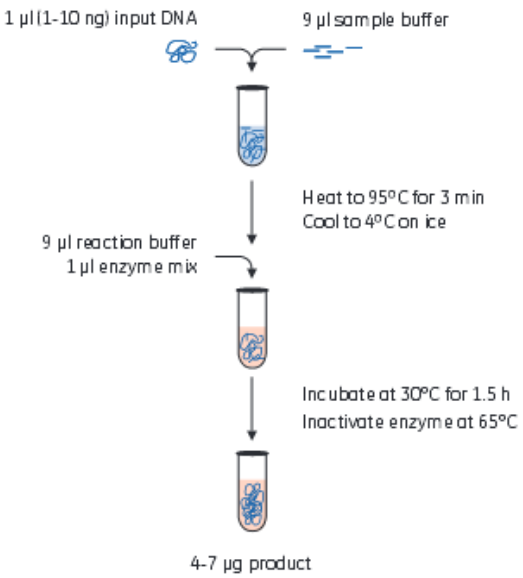


Figure 3. GenomiPhi V2 DNA Amplification Kit protocol (Rech et al., 2007).

3.4. DNA quantification

The resulting DNA amplification was confirmed and quantified by fluorometric measurement using Picogreen®, taking 1 µL of DNA per sample.

On the one hand, a 1:200 Picogreen dilution with TE 1X is prepared and, on the other hand, 1 µL of DNA is taken and mixed with 24 µL of TE 1X. Then, 25 µL of the Picogreen dilution are added to the DNA dilution and, after mixing, incubated in darkness for 3 min. The 50 µL are transferred to the measuring cuvette and the quantification is performed in the fluorimeter.

3.5. Gel electrophoresis

The amplified products were also confirmed by gel electrophoresis, performed with 0.8% agarose gel and using TBE (Tris/Borate/EDTA) 1X buffer solution (89 mM Tris (pH 7.6), 89 mM boric acid, 2 mM EDTA). The molecular weight marker used was “GeneRuler, ready-to-use, 100-10000 bp DNA ladder Mix; 0.1µg/µL” (Thermo Scientific). 0.5 µL of amplified sample was taken and mixed with 1 µL of Loading MIX (50 µL loading buffer 3X, 0.2 µL Gel Red) and 3.5 µL of mQ H20, before loading them on the gel wells. All gels were run during a period of 30 min and a 110 V electric field was applied. Gel results were visualized under UV light, taking the proper precautions, and photographs of the results were taken using a polaroid photo documentation camera.

3.6. Sequencing

Randomly amplified viral genomes were used for high-throughput shotgun sequencing. Samples were prepared for sequencing by adding a barcode and pooling them in groups of 4 samples per run. Then, samples were sequenced simultaneously using the 454 pyrosequencing Genome Sequencer FLX titanium plus, on an eighth of a PicoTiterPlate device (Roche Diagnostics, Mannheim, Germany).

3.7. Bioinformatic analysis

3.7.1. Processing

To analyze the viral metagenomes, raw sequence reads from each sample were filtered for low quality signal or ambiguous characters (<30 out of 40 quality units assigned by the 454) with the Genome Sequencer FLX System Software Package 2.3 (Roche).

Reads were assigned to their corresponding samples according to their barcode-tagged primer sequences, and primers, linkers, and adaptors were trimmed from the sequences.

Removal of rRNA, 16S, 23S and 5S sequences, was performed using the database search tool hmmer, based on profile hidden markov models (profile HMM) (Eddy, 1998), to search the databases of ribosomal sequences.

Elimination of human sequences was performed following a BLAST search against human subset database from the Nucleotide Collection Nr/Nt (NCBI) considering an e-value cutoff of 0.001.

3.7.2. Viral taxonomic identification and nomenclature

Filtered reads were compared against a custom made nonredundant viral and prophage nucleotide database (Virusdb_11_11_2013) using tBLASTx considering an e-value cutoff of 0.001, and a set of Perl programs was used to automate the process of taxonomic binning at class, order, family, genus, and species level. This viral database is a collection composed of the complete viral genomes from the NCBI (http://www.ncbi.nlm.nih.gov/, from NCBI Refseq and Genbank), the viral genomes from the EBI (European Bioinformatics Institute, http://www.ebi.ac.uk/), the complete and partial viral sequences from the database DDBJ (DNA Data Bank of Japan, http://www.ddbj.nig.ac.jp/), and viruses and prophages from ACLAME (A CLAssification of Mobile genetic Elements, http://aclame.ulb.ac.be/).

The nomenclature we used for viruses and prophages was initially generated using the ‘‘Fetch taxonomic representation’’ tool implemented on the Galaxy platform. Next, we used in house Perl scripts to convert the Galaxy output into a standardized abundance table containing four of the taxonomic levels accepted for viruses (order,

family, genus, and species) by inheriting the higher or, if not possible, the lower adjacent taxonomic-level tag to fill in the missing taxonomic levels of each bin. This way a nonredundant taxonomy for all entries was generated. In addition, prophages adopted their bacterial–host taxonomy with addition of the tag ‘‘phage’’.

3.7.3. Viral community structure and diversity estimation

The resulting output files from our tBLASTx-based method used for taxonomic identification were converted into operational taxonomic units (OTUs) tables by customized Perl scripts and can thus be used for the analysis of the diversity and structure of the viral communities.

We assessed the taxonomic richness and diversity of the viral communities by means of several estimators. The richness estimators N (Observed species) and Chao1 (Chao, 1984) and the Shannon diversity index (Shannon, 1948) were calculated using the library ‘vegan’ from the R package (R Development Core Team, 2013). The Chao1 estimator was chosen because it is shown to be one of the most reliable non-parametric estimators of species richness in species-rich samples (Basualdo, 2011). The Shannon index was preferred for species diversity because of its use of natural logarithms of relative species abundances, which reduces the weight of the more abundant species and renders it sensitive to the changes in rare species, which are common in infant gut microbiota samples.

3.7.4. Virome composition, clustering, dynamics and directionality of change

The R package (version 2.15) was employed for comparative analyses of the virome taxonomic composition. Heatmaps and clustering analyses were based on the Bray-Curtis distance as a measure of dissimilarity (Bray and Curtis, 1957). To explore further the pattern of similarities among timepoints we performed Canonical Correspondence Analysis (CCA) using Gower distances (Gower, 1966), for taxonomic data sets.

Microsoft Excel from Microsoft Office Professional Plus 2010 was employed to obtain graphics illustrating dynamics of taxonomic richness and diversity through gut microbiota development of each infant, for both the viral and bacterial community.

IBM SPSS Statistics (Version 20) was employ to obtain boxplots presenting the range of viral and bacterial taxon richness and diversity for all infants and maternal samples analyzed in this study. In addition, the significance (p value) of pairwise t-tests were obtained, comparing richness and diversity between different timepoints and between all infant and maternal samples.

4. RESULTS AND DISCUSSION

4.1. Samples and sequencing

The aim of this study was to explore the patterns of taxonomic change along time during GIT virome development to complement the Vallès et al. (2014) study, where the corresponding analysis with the bacterial gut community was developed. For this purpose, the study has been developed from the samples used in the Vallès et al. (2014) study, where women having healthy pregnancies and stating their intention to exclusively breastfeed their infants during at least three months were recruited to the study.

Vallès et al. (2014) analyzed fecal samples of 13 Mother-Infant Pairs (MIPs) at two different timepoints for the mother, one week prior to delivery (MA) and one year after (MB), and at five different timepoints for the infant, one week (I1), one month (I2), three months (I3, before introduction of solid foods), seven months (I4, after introduction of solid foods) and one year after birth (I5). However, not enough material for virome analysis remained for many of the collected samples, so some of the mother-infant pairs and the I2 timepoint could not be analyzed and we finally had available to analyze the fecal samples of 7 mother-infant pairs at six of the seven timepoints, an overall total of 42 samples. But, moreover, due to technical difficulties, we finally obtained the viral sequences of 25 samples.

All 7 infants in the analyzed MIPs were born at term (>37 weeks of gestation), five of them by vaginal delivery and two by C-section. Six infants were exclusively breastfed during at least three months and one was partially breastfed during the first month and formula-fed thereafter (Table 1).

Table 1. Information regarding mothers and infants obtained from questionnaires answered by the infants' parents. Table modified from Vallès et al. (2014).
MIP: Mother Infant Pair.
a. For MA samples we report whether antibiotics were given during childbirth and the specific antibiotic given. In the case of C-sections, we report administration of amoxicillin, which is the standard practice in Spanish hospitals. None of the mothers had taken antibiotics before childbirth for at least three months.
b. Oftalmowell is a combination of gramicidin, neomycin and polymyxin B.

Sample	Age	Sex	Delivery	Antibiotics Mother[a]	Antibiotics Infant	Diet	Sequenced
MIP01.MA	29	-	-	No	-	-	Yes
MIP01.MB	30	-	-	-	-	-	No
MIP01.I1	1 Week	Male	Vaginal	-	-	Breast milk	No
MIP01.I3	3 Months	-	-	-	-	Breast milk	Yes
MIP01.I4	7 Months	-	-	-	-	Solid foods	Yes
MIP01.I5	1 Year	-	-	-	-	Solid foods	No
MIP03.MA	30	-	-	No	-	-	No
MIP03.MB	31	-	-	No	-	-	Yes
MIP03.I1	1 Week	Female	Vaginal	Amoxicillin	Oftalmowell[b]	Breast milk	No
MIP03.I3	3 Months	-	-	Cefuroxime	-	Breast milk	Yes

MIP03.I4	7 Months	-	-	-	-	Solid foods	Yes
MIP03.I5	1 Year	-	-	Amoxicillin	Cefuroxime	Solid foods	Yes
MIP06.MA	42	-	-	Amoxicillin	-	-	Yes
MIP06.MB	43	-	-	-	-	-	Yes
MIP06.I1	1 Week	Female	C-section	-	-	Breast milk	No
MIP06.I3	3 Months	-	-	-	-	Breast milk	Yes
MIP06.I4	7 Months	-	-	-	-	Solid foods	Yes
MIP06.I5	1 Year	-	-	-	Amoxicillin	Solid foods	Yes
MIP07.MA	31	-	-	Amoxicillin	-	-	No
MIP07.MB	32	-	-	-	-	-	Yes
MIP07.I1	1 Week	Male	C-section	-	-	Breast milk	No
MIP07.I3	3 Months	-	-	-	-	Breast milk	No
MIP07.I4	7 Months	-	-	-	-	Solid foods	Yes
MIP07.I5	1 Year	-	-	-	-	Solid foods	Yes
MIP08.MA	30	-	-	No	-	-	Yes
MIP08.MB	31	-	-	-	-	-	Yes
MIP08.I1	1 Week	Female	Vaginal	-	-	Breast milk	No
MIP08.I3	3 Months	-	-	-	-	Breast milk	No
MIP08.I4	7 Months	-	-	-	-	Solid foods	Yes
MIP08.I5	1 Year	-	-	-	-	Solid foods	Yes
MIP09.MA	30	-	-	No	-	-	Yes
MIP09.MB	31	-	-	-	-	-	No
MIP09.I1	1 Week	Male	Vaginal	Amoxicillin	-	Mixed	Yes
MIP09.I3	3 Months	-	-	-	-	Formula	Yes
MIP09.I4	7 Months	-	-	-	-	Solid foods	Yes
MIP09.I5	1 Year	-	-	-	-	Solid foods	Yes
MIP16.MA	39	-	-	Amoxicillin	-	-	Yes
MIP16.MB	40	-	-	-	-	-	No
MIP16.I1	1 Week	Male	Vaginal	-	-	Breast milk	No
MIP16.I3	3 Months	-	-	-	-	Breast milk	No
MIP16.I4	7 Months	-	-	-	-	Solid foods	No
MIP16.I5	1 Year	-	-	-	-	Solid foods	No

After quality filtering, we obtained a total of 480835 reads with an average of 19233 reads per sample and an average length of 566 bp (Table 2).

Very few ribosomal RNA-like sequences were found in the fecal samples, an average of 1.56, indicating that ribosomal contamination was essentially eliminated during the isolation procedure.

Regarding human DNA sequences, most of the samples showed little amounts of them, less than 3% of the sequence reads after removal of rRNA sequences. But some of the samples (MIP01.I3, MIP01.I4 and MIP09.I1) showed high amounts of human sequences, over 94% of the sequence reads. This may reflect the copurification

of human sequences with the virions from the fecal samples during tangential flow filtration and the incorrect removal of unprotected nucleic acids of human origin during the purification step.

Based on the analyses of viral sequences, an average of 68.5% of the sequence reads after removal of human sequences were not identified in tBLASTx searches against the viral database (classified as unknown), and an average of 31.5% of the sequences showed a significant degree of similarity to virus (classified as known). Comparing these results with those obtained in the Pérez-Brocal et al. (2013) study, where just 12.01% of the reads showed homology to viral hits in tBLASTx analyses, we have obtained a high percentage of viral reads.

Table 2. Summary of the number of reads generated by 454 pyrosequencing before and after processing.

Samples	Raw reads	rRNA sequences	Reads after removal of rRNA sequences	Human sequences	Reads after removal of human sequences	Unknown sequences	Viral reads
MIP01.MA	30527	0	30527	4	30523	12014	18509
MIP01.I3	19664	1	19663	19099	564	256	308
MIP01.I4	25582	1	25581	24171	1410	866	544
MIP03.MB	11692	0	11692	25	11667	10536	1131
MIP03.I3	14853	0	14853	48	14805	12570	2235
MIP03.I4	11371	0	11371	305	11066	8517	2549
MIP03.I5	6402	0	6402	31	6371	2706	3665
MIP06.MA	27509	2	27507	301	26983	18920	8286
MIP06.MB	8670	0	8670	40	8630	8409	221
MIP06.I3	6904	0	6904	12	6892	5250	1642
MIP06.I4	10429	0	10429	24	10405	5002	5403
MIP06.I5	1796	0	1796	9	1787	676	1111
MIP07.MB	11939	0	11939	163	11776	4431	7345
MIP07.I4	17798	0	17798	6	17792	6859	10933
MIP07.I5	9723	0	9723	98	9625	7749	1876
MIP08.MA	26309	1	26308	98	26210	24487	1723
MIP08.MB	9985	5	9980	65	9915	3814	6101
MIP08.I4	34142	0	34142	127	34015	24439	9576
MIP08.I5	25544	0	25544	69	25475	14539	10936
MIP09.MA	20063	0	20063	35	20028	17914	2114
MIP09.I1	35716	1	35715	35395	320	241	79
MIP09.I3	41836	27	41809	4784	37025	21785	15240
MIP09.I4	27074	0	27074	260	26814	20563	6251
MIP09.I5	26074	0	26074	301	25773	21500	4273
MIP16.MA	19233	1	19232	143	19089	16671	2418
Total	480835	39	480796	85613	394960	270714	124469
Average	19233,4	1,56	19231,84	3424,52	15798,4	10828,56	4978,76

4.2. Viral and bacterial richness and diversity within and between samples

The dynamics of richness and diversity in the infant microbiota from the first week to the one-year mark were assessed by calculating the Chao1 estimator (Chao, 1984) and the Shannon index (Shannon, 1948). Richness, i. e. the number of taxa present in a community, is estimated with the Chao1 estimator, while the Shannon index of diversity takes into account both richness and evenness, i.e. how similar the abundances of the different taxa are.

The dynamics of taxonomic richness and diversity are complicated to evaluate because the MIPs are not complete (not all the samples were sequenced) and it is not possible to observe the dynamics along the entire evolution of the infants. In addition, there is a lot of variability in both the taxonomic richness and diversity changes undergone by the different infants during the year. Nevertheless, some trends can also be discerned. The estimators of taxonomic richness and diversity for the different samples are presented in Table 3, their dynamics can be observed in the graphics of Figure 4 and their ranges in the boxplots of Figure 5.

Regarding the maternal virome, changes between the perinatal period (MA) and one year after childbirth (MB) are suggested in Figure 5. The viral taxonomic richness median value of MA samples is lower than that of MB samples, as is also the case for bacterial taxonomic richness when only the samples included in the viral analyses are considered. In contrast, the median value for viral taxonomic diversity is higher in MA samples, again similarly to the case for bacterial diversity. Richness decreases accompanied by a diversity increase can ensue from the loss of rare taxa, which results in a more homogeneous distribution of taxa in the community. However, the differences between MA and MB samples in richness and diversity are not significant in pairwise t-tests (Table 4), probably due to the low number of samples compared.

With respect to the infant virome, although it is impossible to perform a comparison of individual temporal profiles due to the incompleteness of the data, the overall trends depicted in Figure 4A, 4C, 5A and 5C seem to indicate that viral richness and diversity tend to decrease through infant development. In nearly all infants, both viral richness and diversity are lower at the last timepoint sampled than at the first, including in MIP09, the only one for which viral sequences were obtained for all infant timepoints. These results are in contrast to those detected for the bacterial microbiota, both for the subset of samples analyzed here (Figure 4B, 4D, 5B and 5D) and for the entire infant cohort (Vallès et al. 2014), where richness and diversity increase significantly between I1 and I5. In particular, for all infants for which the comparison can be made, the viral richness and diversity values at I5 are lower than those that had been attained by I3, before the introduction of solid foods. This is reflected in a significant ($p=0,039$) diversity change from I3 to I5 when all samples are considered (Table 4). The contrasting results between the temporal trends of the infant gut viral and bacterial communities might indicate that bacterial communities turn more resistant to viral infection as they become more rich and complex.

Table 3. Taxonomic richness (N and Chao1) and diversity (Shanon Index).

	Virus			Bacteria		
	N	**Chao1**	**Shannon**	**N**	**Chao1**	**Shannon**
MIP01.MA	32	102	0,46	326	387,96	2,03
MIP01.I3	81	435	3,23	246	350,34	1,58
MIP01.I4	101	207,94	3,27	178	234,4	2,25
MIP03.MB	27	93	1,45	349	446,02	1,41
MIP03.I3	229	406,16	3,23	77	92,83	0,81
MIP03.I4	252	391,79	3,85	134	164,27	1,24
MIP03.I5	197	329,1	2,08	140	189,14	2,14
MIP06.MA	302	458,89	3,51	341	405,77	3,11
MIP06.MB	56	105,58	3,05	302	408,05	1,59
MIP06.I3	196	331,37	3,84	279	347,47	1,6
MIP06.I4	397	586,85	4,05	187	321,17	1,83
MIP06.I5	64	115	2,26	367	449,91	2,31
MIP07.MB	164	410,57	2,24	389	467,91	2,32
MIP07.I4	91	213,77	0,36	249	298,29	1,96
MIP07.I5	209	314,39	3,45	240	345,84	1,56
MIP08.MA	175	311,5	4,05	286	357,2	1,61
MIP08.MB	405	677,6	2,4	278	379,86	2,29
MIP08.I4	421	609,75	4,1	187	269,5	1,65
MIP08.I5	79	286	2,46	354	473,11	1,5
MIP09.MA	35	61,25	1,96	347	446,38	2,17
MIP09.I1	62	251,43	4,01	147	231,91	0,42
MIP09.I3	245	286,13	2,96	158	211,2	1,82
MIP09.I4	14	24	1,64	137	183,87	1,5
MIP09.I5	24	37,75	1,46	320	393,15	2,06
MIP16.MA	117	185,08	2,76	229	287,41	2,22

Table 4. Significance (p value) of pairwise t-tests comparing richness and diversity between different timepoints and between all infant and maternal samples. Significant values (p<0,05) are colored red.

Pairwise timpepoints	**Viral Richness**	**Viral Diversity**	**Bacterial Richness**	**Bacterial Diversity**
I3-I4	0,807	0,531	0,942	0,325
I4-I5	0,304	0,471	0,069	0,441
I3-I5	0,071	**0,039**	0,172	0,145
I5-MB	0,525	0,905	0,353	0,969
MA-MB	0,563	0,724	0,187	0,370
Infants-Mothers	0,682	0,324	**0,002**	0,052

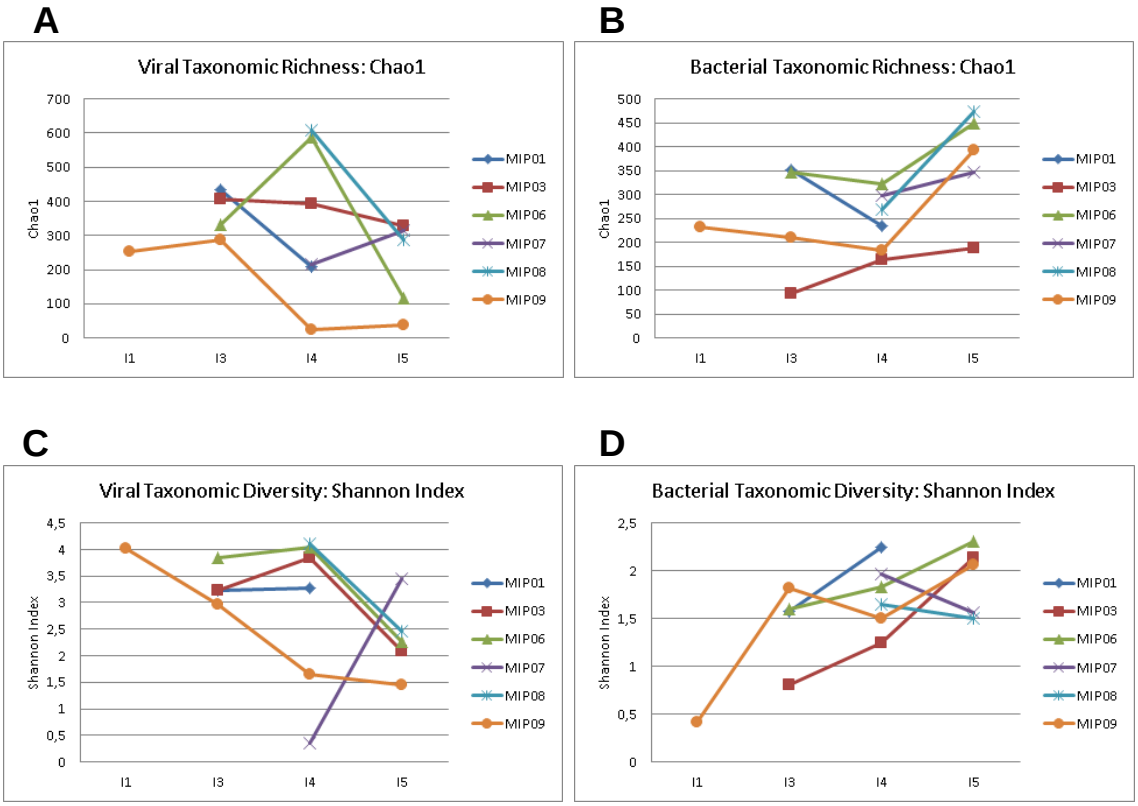
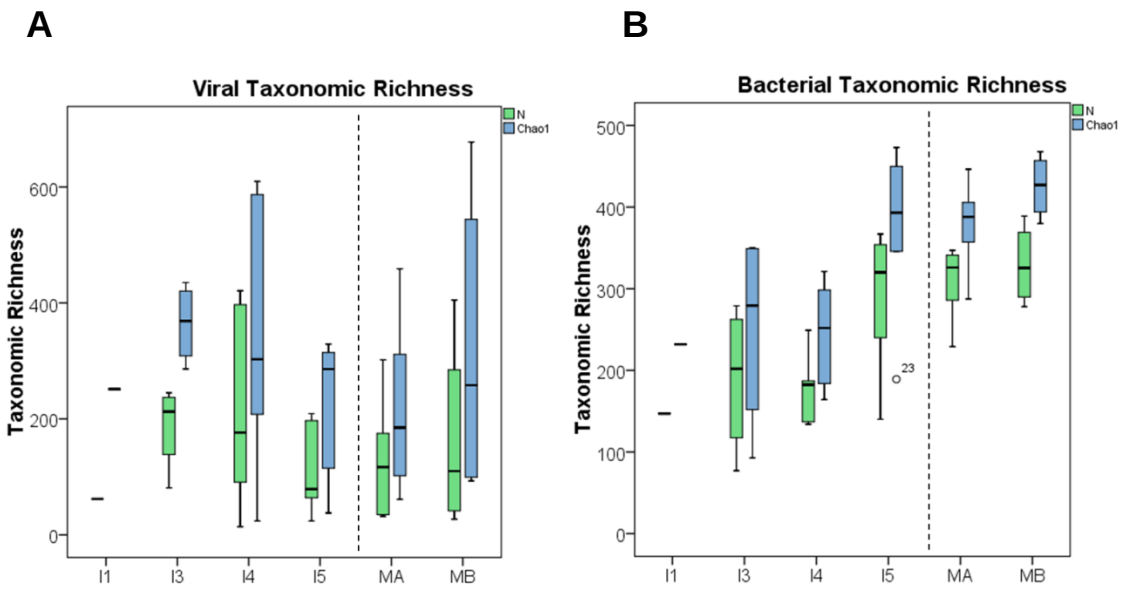


Figure 4. Graphics illustrating dynamics of taxonomic richness and diversity through infant gut microbiota development. The evolution of the viral (A) and bacterial (B) taxonomic richness (Chao1 estimator), as well as the viral (C) and bacterial (D) taxonomic diversity (Shannon index) for each infant is depicted in each graphic.



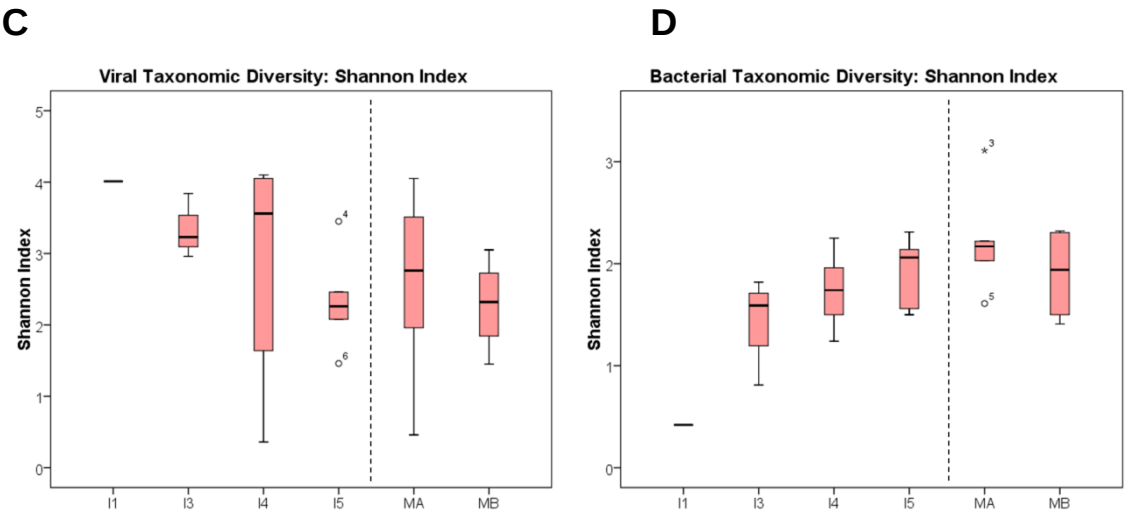


Figure 5. Boxplots presenting the range of viral (A) and bacterial (B) taxon richness and viral (C) and bacterial (D) taxon diversity for all infant and maternal samples analyzed in this study.

4.3. Comparison of sample clustering based on viral composition and abundance

Clustering analysis of samples using the Bray-Curtis distance (Bray and Curtis, 1957) provides a global comparison of all infant and maternal samples according to the taxonomic composition of their virome. Figure 6 illustrates the uneven frequencies of viral types within each sample and the high variability across samples.

A priori, one could expect that the samples within one MIP would be more similar among each other than to any other sample, due to permanence of viral communities through time in the infants and to the fact that mother and infant are genetically and environmentally linked; but, on the other hand, infants at the same timepoint could rather have similar viral communities due to similarities in their bacterial microbiotas, potentially brought about by age-related physiological requirements. However, sample clustering shows no clear pattern of clustering, neither by MIP nor by timepoint.

Nevertheless, a trend of convergence towards the taxonomic composition of the maternal virome by the end of the year is suggested. A different distribution of samples according to age can be detected in the clustering analysis. The taxonomic heatmap in Figure 6 shows a cluster (a) that contains nine infant samples but only two maternal samples, and on the other hand, another cluster (b) that contains all the other maternal samples, most of the I5 samples and only a few samples from other infant timepoints. The clustering of most maternal and I5 samples may reflect that the adult GIT harbors a virome that is distinct in taxonomic composition from those present in infants throughout most of their first year of life, but that by one-year of age the viral community in the infant gut has converged towards the adult taxonomic composition. A very similar trend in clustering patterns is observed at the level of bacterial microbiota composition for the entire infant cohort (Vallès et al. 2014), suggesting that the

convergence in bacterial community composition may drive a concomitant convergence in the infant gut viral communities.

On the other hand, MA and MB samples of the same mother do not group together in the clustering analysis, as they do not in terms of bacterial microbiota composition (Vallès et al. 2014), reflecting that the maternal gut microbiota is altered during pregnancy.

Regarding the specific composition of the samples, shown on the heatmap (Figure 6A), a prevalence of bacteriophages over other types of viruses is observed, in accordance with previous analyses of the human gut virome (Breitbart et al., 2003). Particularly, the maternal and I5 cluster (cluster b) shows a high percentage of Chlamydia phages that are not detected in younger infants. Interestingly, chlamydias are not present at a significant abundance (>1%) in any of the bacterial communities in these samples. On the other hand, the younger infants cluster (cluster a) cannot be characterized by the general presence of any one type of phage, but a small cluster of three infant samples shows an abundance of enteric phages, including Escherichia coli, Salmonella and Cronobacter phages.

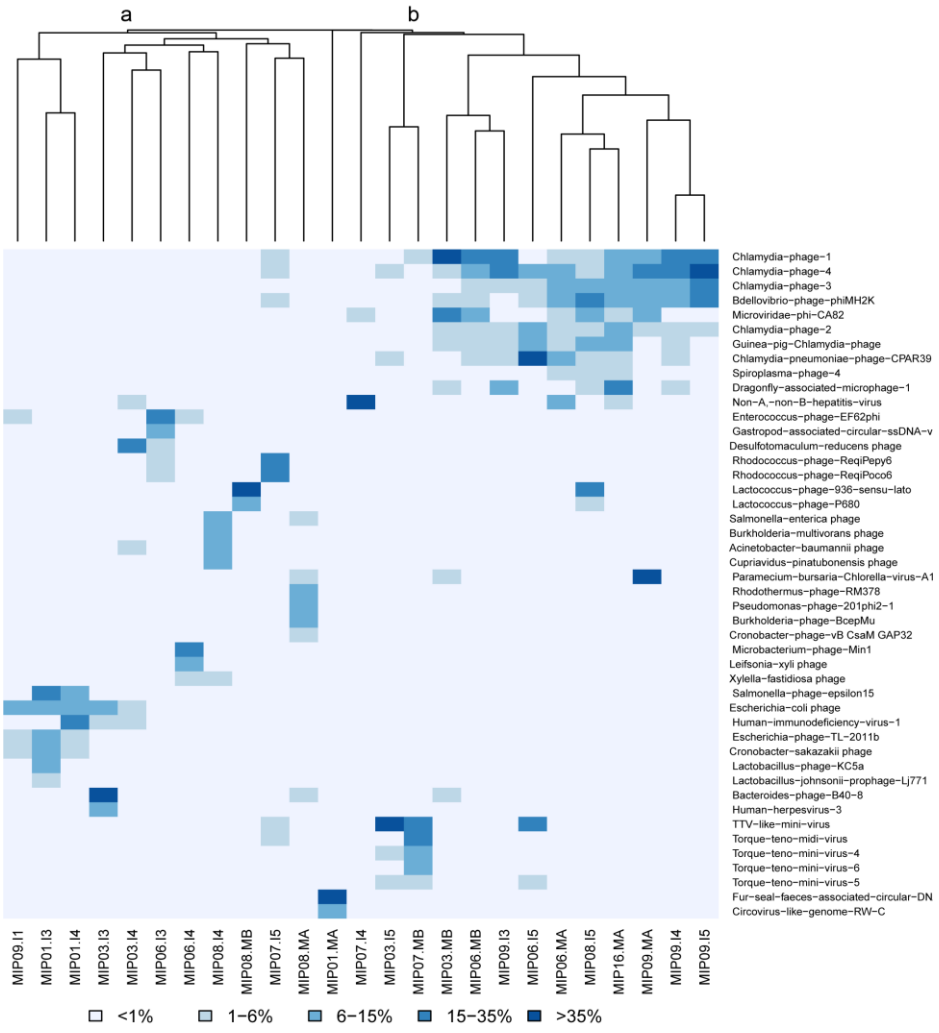


Figure 6. Heatmap and clustering of individual gut virome samples for taxonomic composition. Clustering was based on Bray-Curtis distances. Only taxa above 1% abundance in at least one sample are depicted. Each sample is identified at the bottom of the heatmaps.

4.4. Directionality in taxonomic change along development

In order to investigate whether an overall pattern of change towards the taxonomic composition of the maternal virome can be discerned in the infant samples, a Canonical Correspondence Analysis (CCA) was performed. The CCA enables a visualization of how the variation among samples is distributed.

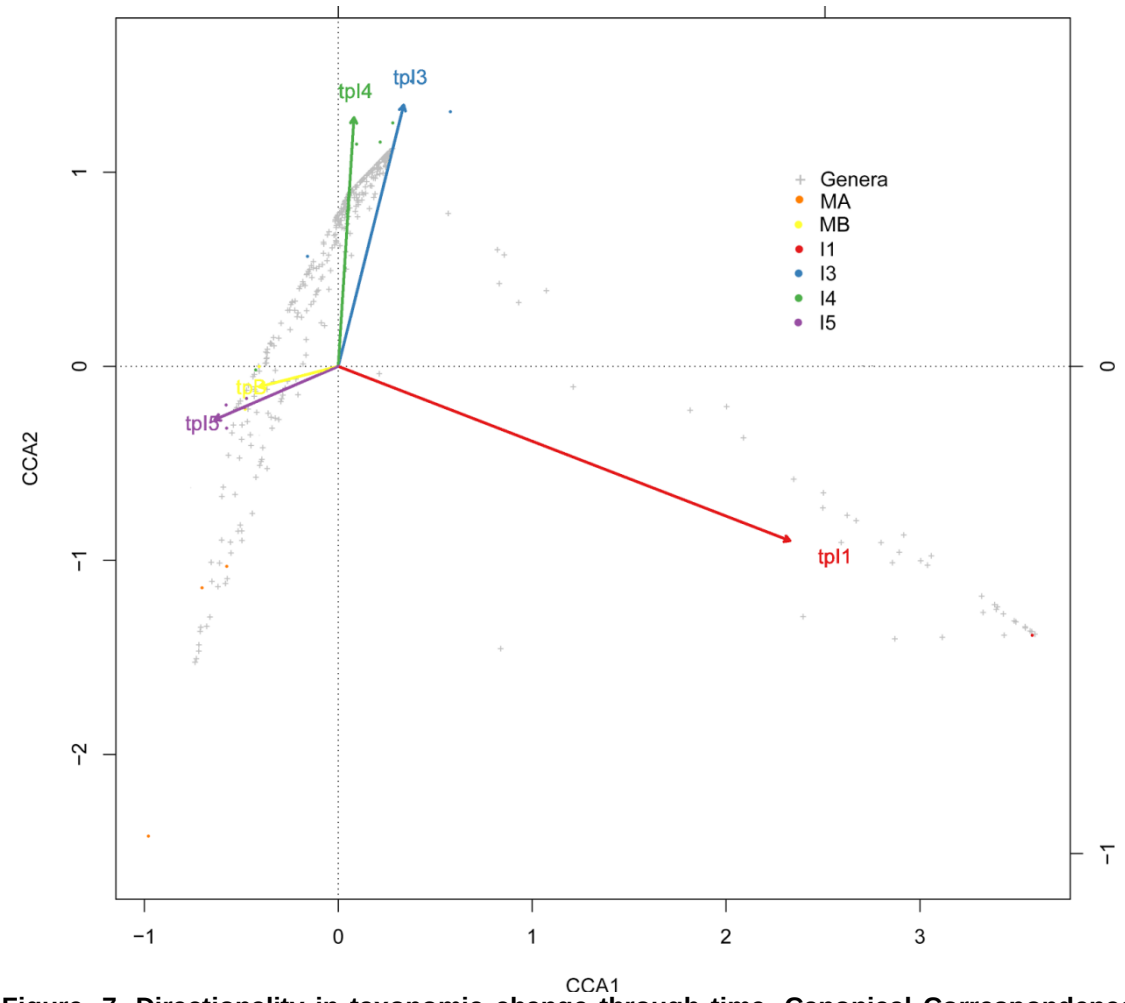


Figure 7. Directionality in taxonomic change through time. Canonical Correspondence Analysis (CCA) of taxonomic data, showing that the main axis (CCA1) separates infant timepoints I1, I3 and I4 from I5, MA and MB. The direction of the timepoint arrows indicates the main axis of deviation from the reference maternal timepoint (MA).

The CCA (Figure 7) reveals a pattern of change with a clear directionality towards the taxonomic composition of the maternal virome, confirming the findings obtained in the clustering analysis showing that I5 clustered with the maternal samples. The first axis of the CCA separates the majority of infant samples (I1, I3 and I4) from the I5 and maternal samples (MA and MB), reflecting that progressive change has resulted in a virome more similar to that of the mothers. The I1 timepoint is represented by a single sample in the analysis, but it can be seen that this sample is the most distant in composition from the maternal samples, whereas I3 an I4 are similar to each other and closer to the I5 and maternal samples. This pattern is analogous to that seen in Vallès et al. (2014) for the taxonomic composition of the bacterial microbiota.

5. CONCLUSIONS

The lack of sufficient material for virome analysis and other technical factors have not allowed us to obtain the viral sequences of the complete MIPs. This has hindered the analyses because we could not do a complete tracking of the virome evolution throughout the first year of the infants. Nevertheless, several interesting preliminary results have been obtained on the basis of the subset of samples for which sequences could be generated.

No common successional pattern across individuals in terms of magnitude and direction of taxon richness and diversity change through time has been found, as the partial temporal profiles detected for each individual are distinct. However, the overall trends seem to indicate that viral richness and diversity tend to decrease during the first year of life, in contrast to the results detected in the bacterial microbiota.

No clear clustering pattern across samples according to virome taxonomic composition has been observed. However, a trend of convergence towards the maternal taxonomic composition by the end of the year has been found and confirmed by the CCA.

Completion of sequencing for all available samples from the cohort and further bioinformatic and statistical analyses will be required to better describe and interpret the changes in the infant gut virome during the first year of life, and to unravel the relation among such changes and the process of bacterial succession in this environment.

6. BIBLIOGRAPHY

ANGLY, F.E.; FELTS, B.; BREITBART, M. ET AL. (2006). The marine viromes of four oceanic regions. PLoS Biol 2006; 4: e368.

AVERSHINA, E.; STORRO, O.; OIEN, T.; JOHNSEN, R.: POPE, P. ET AL. (2014). Major faecal microbiota shifts in composition and diversity with age in a geographically restricted cohort of mothers and their children. FEMS Microbiol Ecol 87: 280–290.

BASUALDO, C. (2011). Choosing the best non-parametric richness estimator for benthic macroinvertebrates databases. Rev Soc Entomol Argent 70: 27–38.

BENCH, SR.; HANSON, T.E.; WILLIAMSON, K.E. ET AL. (2007). Metagenomic characterization of Chesapeake Bay virioplankton. Appl Environ Microbiol 2007; 73: 7629–7641.

BEZIRTZOGLOU, E.; TSIOTSIAS, A.; WELLING, G.W. (2011). Microbiota profile in feces of breast- and formula-fed newborns by using fluorescence in situ hybridization (FISH). Anaerobe 2011; 17:478–82.

BOHANNAN, B.J.M.; LENSKI, R.E. (1997). Effect of resource enrichment on a chemostat community of bacteria and bacteriophage. Ecology 78: 2303–2315.

BRAY, J.R.; CURTIS, J.T. (1957). An ordination of upland forest communities of southern Wisconsin. Ecological Monographs 27: 325–349.

BREITBART, M.; FELTS, B.; KELLEY, S. ET AL. (2004). Diversity and population structure of a nearshore marine sediment viral community. Proc Biol Sci 2004; 271: 565–574.

BREITBART, M.; HAYNES, M.; KELLEY, S. ET AL. (2008). Viral diversity and dynamics in an infant gut. Res Microbiol 2008; 159: 367–373.

BREITBART, M.; HEWSON, I.; FELTS, B. ET AL. (2003). Metagenomic analyses of an uncultured viral community from human feces. J Bacteriol 2003; 185: 6220–6223.

BREITBART, M.; SALAMON, P.; ANDRESEN, B. ET AL. (2002). Genomic analysis of uncultured marine viral communities. Proc Natl Acad Sci USA 2002; 99: 14250–14255.

CAPDEVILA, F.; VIZMANOS, B.; MARTI-HENNEBERG, C. (1998). Implications of the weaning pattern on macronutrient intake, food volume and energy density in non-breastfed infants during the first year of life. J Am Coll Nutr 17: 256–262.

CHAO, A. (1984). Nonparametric estimation of the number of classes in a population. Scandinavian Journal of Statistics 11: 256–270.

COLLADO, M.C.; D'AURIA, G.; MIRA, A.; FRANCINO, M.P. (2013). Human Microbiome and Diseases: A Metagenomic Approach. In: Watson RR and Preedy VR,

editors. Bioactive Food as Dietary Interventions for Liver and Gastrointestinal Disease. San Diego: Academic Press. pp. 235–249.

DAVIS, B.M.; WALDOR, M.K. (2002). Mobile genetic elements and bacterial pathogenesis. In Craig NL, Gragie R, Gellert M, Lambowitz AM (eds) Mobile DNA II. ASM Press: Washington, DC, USA, 2002. pp. 1040–1055.

DICKSVED, J.; FLOISTRUP, H.; BERGSTROM, A.; ROSENQUIST, M.; PERSHAGEN, G.; SCHEYNIUS, A.; ROOS, S.; ALM, J.S.; ENGSTRAND, L.; BRAUN-FAHRLANDER, C.; VON MUTIUS, E.; JANSSON, J.K. (2007). Molecular fingerprinting of the fecal microbiota of children raised according to different lifestyles. Appl. Environ. Microbiol. 73, 2284e2289.

DOMÍNGUEZ-BELLO, M.G.; COSTELLO, E.K.; CONTRERAS, M. ET AL. (2010). Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. Proc Natl Acad Sci U S A 2010; 107: 11971–5.

EDDY, S.R. (1998). Profile hidden markov models. Vol. 9 no. 9 1998. Pages 755-763

GÓRSKI, A.; WEBER-DABROWSKA, B. (2005). The potential role of endogenous bacteriophages in controlling invading pathogens. *Cell Mol Life Sci* 2005;**62**: 511–519.

GOSALBES, M.J.; LLOP, S.; VALLÈS, Y.; MOYA, A.; BALLESTER, F. ET AL. (2013). Meconium microbiota types dominated by lactic acid or enteric bacteria are differentially associated with maternal eczema and respiratory problems in infants. Clin Exp Allergy 43: 198–211.

GOWER, J.C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53: 325–338.

HERMOSO, J.A.; GARCÍA, J.L.; GARCÍA, P. (2007). Taking aim on bacterial pathogens: from phage therapy to enzybiotics. Current Opinion in Microbiology; Volume 10, Issue 5, October 2007, Pages 461–472

HOSONO, S.; FARUQI, A.F.; DEAN, F.B.; DU, Y.; SUN, Z.; WU, X.; DU, J.; KINGSMORE, S.F.; EGHOLM, M.; LASKEN, R.S. (2003). Unbiased whole-genome amplification directly from clinical samples. Genome Res 13:954–964

HWANG, D.J.; ROBERTS, I.M.; WILSON, T.M. (1994). Expression of tobacco mosaic virus coat protein and assembly of pseudovirus particles in Escherichia coli. Proc Natl Acad Sci U S A 91: 9067–9071.

JAGADISH, M.N.; HUANG, D.; WARD, C.W. (1993). Site-directed mutagenesis of a potyvirus coat protein and its assembly in Escherichia coli. J Gen Virol 74: 893–896.

JALANKA-TUOVINEN, J.; SALONEN, A; NIKKILA, J.; IMMONEN, O.; KEKKONEN, R. ET AL. (2011). Intestinal microbiota in healthy adults: temporal analysis reveals individual and common core and relation to intestinal symptoms. PLoS One 6: e23035.

KIM, M.S.; PARK, E.J.; ROH, S.W. ET AL. (2011). Diversity and abundance of single-stranded DNA viruses in human feces. Appl Environ Microbiol 2011; 77: 8062–8070.

KOENIG, J.E.; SPOR, A.; SCALFONE, N.; FRICKER, A.D.; STOMBAUGH, J. ET AL. (2011). Succession of microbial consortia in the developing infant gut microbiome. Proc Natl Acad Sci U S A 108 Suppl 1: 4578–4585.

KOREN, O.; GOODRICH, J.K.; CULLENDER, T.C.; SPOR, A.; LAITINEN, K. ET AL. (2012). Host remodeling of the gut microbiome and metabolic changes during pregnancy.Cell 150: 470–480.

KUROKAWA, K.; ITOH, T.; KUWAHARA. T.; OSHIMA, K.; TOH, H.; ET AL. (2007). Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. DNA Res 14: 169–181.

LÓPEZ-BUENO, A.; TAMAMES, J.; VELÁZQUEZ, D. ET AL. (2009). High diversity of the viral community from an Antarctic lake. Science 2009; 326: 858–861.

MACKIE, R.I.; SGHIR, A.; GASKINS, H.R. (1999). Developmental microbial ecology of the neonatal gastrointestinal tract. *Am J Clin Nutr* 69: 1035S–1045S, 1999.

MANDAR, R.; MIKELSAAR, M. (1996). Transmission of mother's microflora to the newborn at birth. *Biol Neonate*69: 30–35, 1996.

REA M.C.; DEBEBE ALEMAYEHU, R.; ROSS, P.; HILL, C. (2013). Gut solutions to a gut problem: bacteriocins, probiotics and bacteriophage for control of Clostridium difficile infection. Journal of Medical Microbiology (2013), 62, 1369–1378

MCCONNELL, E.L.; BASIT, A.W.; MURDAN, S. (2008). Measurements of rat and mouse gastrointestinal pH, fluid and lymphoid tissue, and implications for in-vivo experiments. J Pharm Pharmacol 60: 63–70.

MINOT, S.; SINHA, R.; CHEN, J. ET AL. (2011). The human gut virome: inter-individual variation and dynamic response to diet. Genome Res 2011; 21: 1616–1625.

PAEZ, J.G.; LIN, M.; BEROUKHIM, R.; LEE, J.C.; ZHAO, X.; RICHTER, D.J.; GABRIEL, S.; HERMAN, P.; SASAKI, H.; ALTSHULER, D. ET AL. (2004). Genome coverage and sequence fidelity of phi29 polymerase-based multiple strand displacement whole genome amplification. Nucleic Acids Res 32:e71

PALMER, C; BIK, E.M.; DIGIULIO, D.B.; RELMAN, D.A.; BROWN, P.O. (2007). Development of the human infant intestinal microbiota. PLoS Biol 5: e177.

PENDERS, J.; STOBBERINGH, E.E.; VAN DEN BRANDT, P.A.; THIJS, C. (2007). The role of the intestinal microbiota in the development of atopic disorders. Allergy 62: 1223–1236.

PENDERS, J.; THIJS, C.; VINK, C.; STELMA, F.F.; SNIJDERS, B.; KUMMELING, I.; VAN DEN BRANDT, P.A.; STOBBERINGH, E.E. (2006). Factors influencing the composition of the intestinal microbiota in early infancy. Pediatrics 118, 511e521.

PÉREZ-BROCAL, V.; GARCÍA-LÓPEZ, R.; VÁZQUEZ-CASTELLANOS, J. F.; NOS, P.; BELTRÁN, B.; LATORRE, A.; MOYA, A. (2013). Study of the Viral and Microbial Communities Associated With Crohn's Disease: A Metagenomic Approach. *Clinical and Translational Gastroenterology* 4, e36; doi:10.1038/ctg.2013.9

QIN, J. ET AL. 2010. A human gut microbial gene catalogue established by metagenomic sequencing. Nature 464:59–65.

R DEVELOPMENT CORE TEAM. (2010). R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

RAUTAVA, S.; RUUSKANEN, O.; OUWEHAND, A.; SALMINEN, S.; ISOLAURI, E. (2004). The hygiene hypothesis of atopic disease–an extended version. J Pediatr Gastroenterol Nutr 38: 378–388.

RECH, R.; KAPOLKA, K.; LAVRENOV K.; GARNOVA, E.; LAVASINI, S.; DEADMAN, R.; HAMILTON, S. (2007). Genomic DNA preparation using illustra GenomiPhi V2 and HY DNA Amplification Kits. G. Kumar, Gene Analysis Department, GE Healthcare, 800 Centennial Avenue, Piscataway, NJ 08854

REYES, A.; HAYNES, M.; HANSON, N. ET AL. (2010). Viruses in the faecal microbiota of monozygotic twins and their mothers. Nature 2010; 466: 334–338.

RODRIGUEZ-BRITO, B.; LI, L.;WEGLEY, L.; FURLAN, M.; ANGLY, F.; BREITBART, M.; BUCHANAN, J.; DESNUES, C.; DINSDALE, E.; EDWARDS, R.; ET AL. (2010). Viral and microbial community dynamics in four aquatic environments. ISME J 4: 739–751.

RODRIGUEZ-VALERA, F.; MARTIN-CUADRADO, A.B.; RODRIGUEZ-BRITO, B.; PASIC, L.; THINGSTAD, T.F.; ROHWER, F.; MIRA, A. (2009). Explaining microbial population genomics through phage predation. Nat Rev Microbiol 7: 828–836.

ROHWER, F.; EDWARDS, R. (2002). The phage proteomic tree: A genome-based taxonomy for phage. J Bacteriol 184: 4529–4535.

SCHOENFELD, T.; PATTERSON, M.; RICHARDSON, P.M. ET AL. (2008). Assembly of viral metagenomes from Yellowstone hot springs. Appl Environ Microbiol 2008; 74: 4164–4174.

SEKIROV, I.; RUSSELL, S.L.; ANTUNES, L.C. ET AL. (2010). Gut microbiota in health and disease. *Physiol Rev* 2010; **90**: 859–904.

SHANNON, C.E. (1948). A mathematical theory of communication. Bell System Technical Journal 27: 379–656.

SILANDER, K.; SAARELA, J. (2008). Whole Genome Amplification with Phi29 DNA Polymerase to Enable Genetic or Genomic Analysis of Samples of Low DNA Yield. *Genomics Protocols Methods in Molecular Biology*, 439: pp 1-18

SMITH, H.W.; HUGGINS, M. B. (1983). Effectiveness of Phages in Treating Experimental Escherichia coli Diarrhoea in Calves, Piglets and Lambs. Journal of General Microbiology (1983), 129, 2659-2675.

TAMAKI, H.; ZHANG, R.; ANGLY, F.E. ET AL. (2011). Metagenomic analysis of DNA viruses in a wastewater treatment plant in tropical climate. Environ Microbiol 2011; 14: 441–452.

TAMAMES, J.; MOYA, A. (2008). Estimating the extent of horizontal gene transfer in metagenomic sequences. BMC Genomics 9:136.

THOMPSON-CHAGOYAN, O.C.; MALDONADO, J.; GIL, A. (2007). Colonization and impact of disease and other factors on intestinal microbiota. Digest Dis. Sci. 52, 2069e2077.

TURNBAUGH, P.J.; LEY, R.E.; HAMADY, M.; FRASER-LIGGETT, C.M.; KNIGHT, R. ET AL. (2007). The human microbiome project. Nature 449: 804–810.

VALLÈS, Y.; ARTACHO, A.; PASCUAL-GARCÍA, A.; FERRÚS, M.L.; GOSALBES, M.J.; ABELLÁN, J.J.; FRANCINO, M.P. (2014). Microbial succession in the gut: directional trends of taxonomic and functional change in a birth cohort of spanish infants. PLoS Genet. 2014 Jun 5;10(6):e1004406. doi: 10.1371/journal.pgen.1004406. eCollection 2014.

VALLÈS, Y.; GOSALBES, M.J.; DE VRIES, L.E.; ABELLAN, J.J.; FRANCINO, M.P. (2012). Metagenomics and development of the gut microbiota in infants. Clin Microbiol Infect 18 Suppl 4: 21–26.

VENTER, J.C.; REMINGTON, K.; HEIDELBERG, J.F. ET AL. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. Science 2004; 304: 66–74.

VENTURA, M.; SOZZI, T.; TURRONI, F. ET AL. (2011). The impact of bacteriophages on probiotic bacteria and gut microbiota diversity. Genes Nutr 2011; 6: 205–207.

WEGRZYN, G.; THOMAS, M.S. (2002). Modulation of the susceptibility of intestinal bacteria to bacteriophages in response to Ag43 phase variation—a hypothesis. *Med Sci Monit* 2002; 8: HY15–HY18.

WEINBAUER, M.G. (2004). Ecology of prokaryotic viruses. FEMS Microbiol Rev 28: 127–181.

WHITMAN, W.B.; COLEMAN, D.C.; WIEBE, W.J. (1998). Prokaryotes: the unseen majority. *Proc Natl Acad Sci USA* 1998; **95**: 6578–6583.

ZHANG, T.; BREITBART, M.; LEE, W.H. ET AL. (2006). RNA viral community in human feces: prevalence of plant pathogenic viruses. PLoS Biol 2006; 4: 0108–0118.