A MULTI-OMICS APPROACH TO DESCRIBE CROSSTALK BETWEEN GUT MICROBIOTA AND HOST: EPIGENETICS AND VIROME ROLE

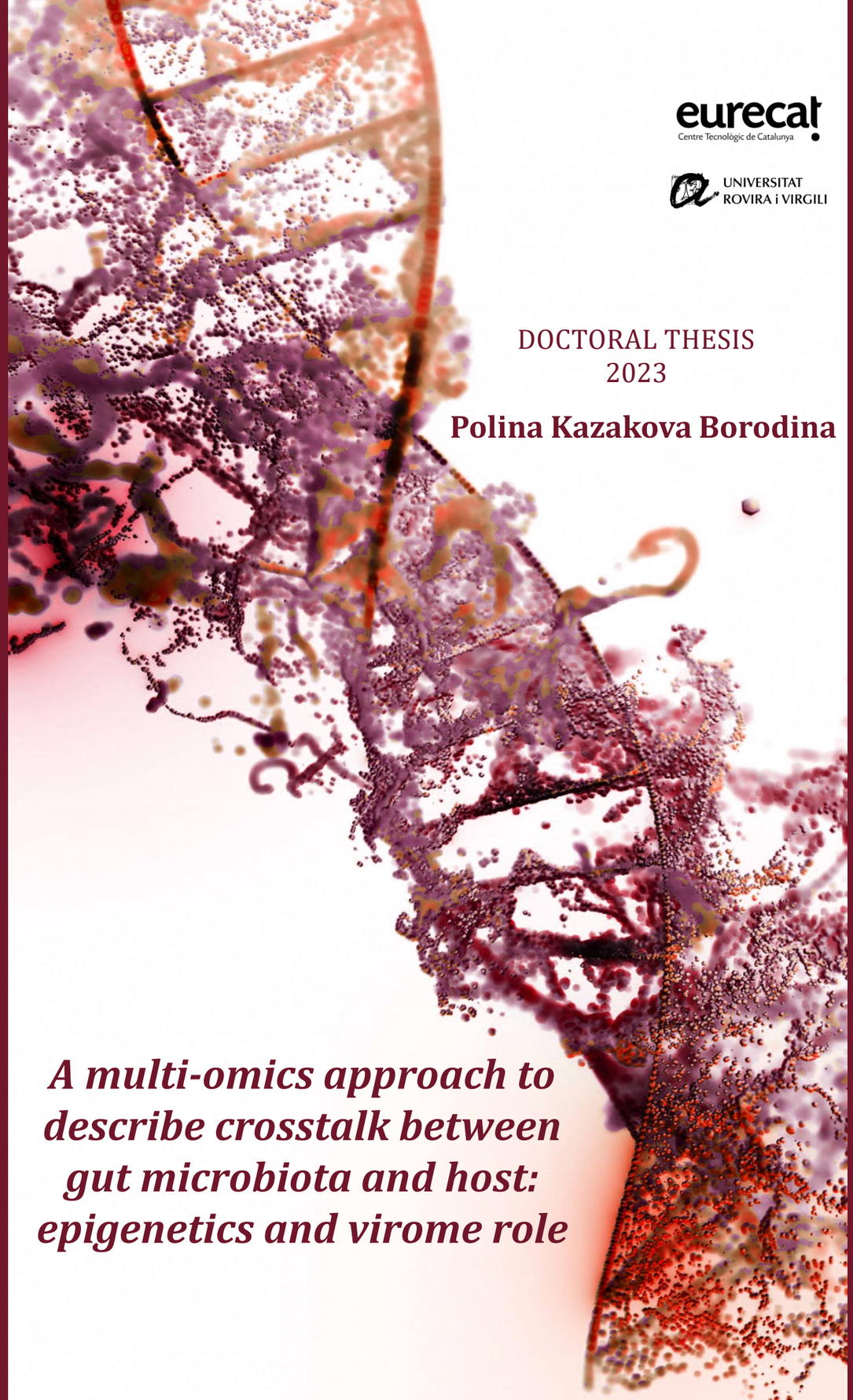
Polina Kazakova Borodina

UNIVERSITAT ROVIRA I VIRGILI
A MULTI-OMICS APPROACH TO DESCRIBE CROSSTALK BETWEEN GUT MICROBIOTA AND HOST:
EPIGENETICS AND VIROME ROLE
Polina Kazakova Borodina

Doctoral
Thesis
2023

A multi-omics approach to describe crosstalk between gut microbiota and host: epigenetics and virome role.

Polina Kazakova Borodina



eurecat
Centre Tecnològic de Catalunya

UNIVERSITAT
ROVIRA i VIRGILI

DOCTORAL THESIS
2023

Polina Kazakova Borodina

A multi-omics approach to describe crosstalk between gut microbiota and host: epigenetics and virome role

Polina Kazakova Borodina

A multi-omics approach to describe crosstalk between gut microbiota and host: epigenetic and virome role.

DOCTORAL THESIS

Supervised by Dr. Helena Torrell Galceran and Dr. Miquel Mulero Abellán

Eurecat, Centro de Ciencias Ómicas

Departament de Bioquímica i Biotecnologia

UNIVERSITAT ROVIRA i VIRGILI

eurecat
Centre Tecnològic de Catalunya

Reus 2023

UNIVERSITAT ROVIRA I VIRGILI
A MULTI-OMICS APPROACH TO DESCRIBE CROSSTALK BETWEEN GUT MICROBIOTA AND HOST: EPIGENETICS AND VIROME ROLE
Polina Kazakova Borodina

UNIVERSITAT ROVIRA i VIRGILI

DEPARTAMENT DE BIOQUÍMICA I BIOTECNOLOGIA
Campus Sescelades
C/ Marcel·lí Domingo, 1
43007 Tarragona
Tel. +34 977 55 95 21
Fax +34 977 55 82 32

FEM CONSTAR que aquest treball, titulat **A multi-omics approach to describe crosstalk between gut microbiota and host: epigenetic and virome role**, que presenta **Polina Kazakova Borodina** per a l'obtenció del títol de doctorat, ha estat realitzat sota la nostra direcció al Departament de Bioquímica i Biotecnologia de la Universitat Rovira i Virgili.

WE STATE that the present study, entitled **A multi-omics approach to describe crosstalk between gut microbiota and host: epigenetic and virome role**, presented by **Polina Kazakova Borodina** for the award of the degree of Doctor, has been carried out under our supervision at the Departament de Bioquímica i Biotecnologia from the Universitat Rovira.

Reus, 2 de Mayo de 2023

Els directors de la tesi doctoral

Doctoral thesis supervisors

Dr. Helena Torrell Galceran

Dr. Miquel Mulero Abellán

UNIVERSITAT ROVIRA I VIRGILI
A MULTI-OMICS APPROACH TO DESCRIBE CROSSTALK BETWEEN GUT MICROBIOTA AND HOST: EPIGENETICS AND VIROME ROLE
Polina Kazakova Borodina

Voyager, c'est bien utile, ça fait travailler l'imagination.

Tout le reste n'est que déceptions et fatigues.

Notre voyage à nous est entièrement imaginaire.

Voilà sa force.

Louis Ferdinand Céline

Science is always wrong.

It never solves a problem without creating ten more.

George Bernard Shaw

UNIVERSITAT ROVIRA I VIRGILI
A MULTI-OMICS APPROACH TO DESCRIBE CROSSTALK BETWEEN GUT MICROBIOTA AND HOST: EPIGENETICS AND VIROME ROLE
Polina Kazakova Borodina

This thesis has been carried out in the Center for omic Sciencies in collaboration with Universitat Rovira i Virgili (URV) thanks to the Vicente Lopez fellowship (Eurecat)

UNIVERSITAT ROVIRA I VIRGILI
A MULTI-OMICS APPROACH TO DESCRIBE CROSSTALK BETWEEN GUT MICROBIOTA AND HOST: EPIGENETICS AND VIROME ROLE
Polina Kazakova Borodina

SUMMARY 13

ABBREVIATIONS 19

INTRODUCTION 21

1 Intestinal microbiota 23

1.1 FUNCTIONS OF THE INTESTINAL MICROBIOTA. 23
1.1.1 MICROBIOTA INVOLVEMENT IN GENERAL METABOLISM. 23
1.1.2 MICROBIOTA AS A HOST DEFENSE MECHANISM AGAINST INFECTIOUS THREATS. 24
1.1.3 TROPHIC. 25
1.2 MODULATORY FACTORS OF MICROBIOTA AND HOST INTERACTIONS. 25
1.2.1 DIET. 25
1.2.2 GENETICS. 25
1.2.3 AGE. 26
1.2.4 ANTIBIOTICS. 26
1.2.5 GUT VIROME. 27
1.2.6 EPIGENETICS. 29
1.2.7 OTHERS. 30
1.3 ACQUISITION AND DEVELOPMENT OF INTESTINAL MICROBIOTA IN EARLY LIFE STAGES. 31
1.3.1 MAIN DRIVERS OF EARLY LIFE COLONIZATION. 32
1.4 THE HOLISTIC STUDY OF THE INTESTINAL MICROBIOTA: MULTIMETAOMIC APPROACHES. 38

2 Celiac disease 45

2.1 CED PATHOPHYSIOLOGY 48
2.2 RISK FACTORS 49
2.2.1 GENETIC AND IMMUNOGENETIC FACTORS 49
2.2.2 ENVIRONMENTAL FACTORS 50
2.3 ROLE OF THE INTESTINAL MICROBIOTA IN CED. 51
2.3.1 DYSBIOSIS IN CED. 51
2.3.2 EARLY INFECTIONS MIGHT INCREASE THE CED RISK. 52
2.3.3 EFFECTS OF INTESTINAL MICROBIOTA ON THE PATHOPHYSIOLOGY OF CED. 54

2.3.4 PROBIOTICS IN CED. 55
2.3.5 PREBIOTICS AND POSTBIOTICS IN CED. 57
2.4 A GLUTEN-FREE DIET (GFD) **57**

3 references **59**

HYPOTHESIS AND OBJECTIVES **77**

RESULTS **83**

CHAPTER 1 **85**

MANUSCRIPT 1 **87**
MANUSCRIPT 2 **123**

CHAPTER 2 **159**

MANUSCRIPT 3 **161**

GENERAL DISCUSSION **199**

REFERENCES **209**

CONCLUSIONS **213**

ACKNOWLEDGEMENTS **217**

LIST OF PUBLICATIONS **219**

LIST OF CONFERENCE PAPERS **221**

SUMMARY

The gut microbiota plays a crucial role in maintaining human health and homeostasis. Therefore, the development of gut microbiota during early life significantly impacts later health. Dysbiosis, or changes in the gut microbiota composition, has been associated with long-term health disorders such as celiac disease. However, while the bacterial microbiota has been extensively studied, the virome, which is much more diverse than all other kingdoms, is relatively less understood. Transkingdom interplay between bacteria and viruses impacts host health beyond their individual effects, emphasizing the importance of exploring this relationship.

To gain further insight into the relationship between the microbiota and its host in early life stages and in the context of childhood celiac disease, this doctoral thesis aims to describe the microbiota of meconium and stool samples from milk-fed infants and to compare the possible alterations of healthy children with that of children with celiac disease adhering to a gluten-free diet. Additionally, it assesses the effect of probiotic administration in an animal model of celiac disease.

By utilizing a combination of different omics sciences, this thesis has the potential to uncover valuable insights into the functionality and activity of the intestinal microbiota and its relationship with diseases.

The research in this thesis will contribute to a better understanding of the significance of transkingdom interplay and its roles in host homeostasis and health conditions.

UNIVERSITAT ROVIRA I VIRGILI
A MULTI-OMICS APPROACH TO DESCRIBE CROSSTALK BETWEEN GUT MICROBIOTA AND HOST: EPIGENETICS AND VIROME ROLE
Polina Kazakova Borodina

RESUM

La microbiota intestinal juga un paper crucial en el manteniment de la salut i l'homeòstasi humanes, per tant, el desenvolupament de la microbiota intestinal durant els primers anys de vida té un impacte significatiu a la salut posterior. La disbiosi, que fa referència als canvis en la composició de la microbiota intestinal, s'ha associat amb trastorns de salut a llarg termini, com la malaltia celíaca. Tot i això, encara que la microbiota bacteriana ha estat àmpliament estudiada, el viroma, que és molt més divers que tots els altres regnes, és relativament menys conegut. La interacció trans-regne entre bacteris i virus té un impacte en la salut de l'hoste més enllà dels seus efectes individuals, cosa que emfatitza la importància d'explorar aquesta relació.

Amb l'objectiu d'aprofundir en la relació entre la microbiota i el seu hoste en etapes primerenques de la vida i en el context de la malaltia celíaca infantil, aquesta tesi doctoral es proposa descriure la microbiota de meconi i mostres de femta de lactants alimentats amb llet, així com avaluar possibles alteracions en nens celíacs adherits a una dieta sense gluten en comparació amb nens sans. A més, s'avalua l'efecte de l'administració de probiòtics en un model animal de malaltia celíaca.

En utilitzar una combinació de diferents ciències òmiques, aquesta tesi té el potencial de descobrir informació valuosa sobre la funcionalitat i l'activitat de la microbiota intestinal i la seva relació amb les malalties. La investigació en aquesta tesi contribuirà a una millor comprensió de la importància de la interacció entre regnes i el seu paper a l'homeòstasi de l'hoste i les condicions de salut.

UNIVERSITAT ROVIRA I VIRGILI
A MULTI-OMICS APPROACH TO DESCRIBE CROSSTALK BETWEEN GUT MICROBIOTA AND HOST: EPIGENETICS AND VIROME ROLE
Polina Kazakova Borodina

RESUMEN

La microbiota intestinal juega un papel crucial en el mantenimiento de la salud y la homeostasis humanas, por lo tanto, el desarrollo de la microbiota intestinal durante los primeros años de vida tiene un impacto significativo en la salud posterior. La disbiosis, que se refiere a los cambios en la composición de la microbiota intestinal, se ha asociado con trastornos de salud a largo plazo, como la enfermedad celíaca. Sin embargo, aunque la microbiota bacteriana ha sido ampliamente estudiada, el viroma, que es mucho más diverso que todos los demás reinos, es relativamente menos conocido. La interacción trans-reino entre bacterias y virus tiene un impacto en la salud del huésped más allá de sus efectos individuales, lo que enfatiza la importancia de explorar esta relación.

Con el objetivo de profundizar en la relación entre la microbiota y su huésped en etapas tempranas de la vida y en el contexto de la enfermedad celíaca infantil, esta tesis doctoral se propone describir la microbiota de meconio y muestras de heces de lactantes alimentados con leche, así como evaluar posibles alteraciones en niños celíacos adheridos a una dieta sin gluten en comparación con niños sanos. Además, se evalúa el efecto de la administración de probióticos en un modelo animal de enfermedad celíaca.

Al utilizar una combinación de diferentes ciencias ómicas, esta tesis tiene el potencial de descubrir información valiosa sobre la funcionalidad y la actividad de la microbiota intestinal y su relación con las enfermedades. La investigación en esta tesis contribuirá a una mejor comprensión de la importancia de la interacción entre reinos y su papel en la homeostasis del huésped y las condiciones de salud.

UNIVERSITAT ROVIRA I VIRGILI
A MULTI-OMICS APPROACH TO DESCRIBE CROSSTALK BETWEEN GUT MICROBIOTA AND HOST: EPIGENETICS AND VIROME ROLE
Polina Kazakova Borodina

ABBREVIATIONS

CeD	Celiac disease
cDNA	Complementary DNA
EVs	Extracellular vesicles
FOXP3	Forkhead box P3
GABA	Gamma-aminobutyric acid
GWAS	Genome-Wide Association Studies
GFD	Gluten-free diet
HLA	Human leukyne antigen
HMO	Human milk oligosaccharides
IgA	Immunoglobulin A
IFN	Interferon
IL	Interleuquine
IM	Intestinal microbiota
MS	Mass spectrometry
r-RNA	Ribosomal RNA
SCFA	Short-chain fatty acids
SNPs	Single-nucleotide polymorphism
TLR	Toll-like receptor
t-RNA	Transfer RNA
TG2	Transglutaminase 2

TNF	Tumor necrosis factor
VLP	Virus-like particle
WHO	World Health Organization
3'UTRs	3' untranslated regions

INTRODUCTION

UNIVERSITAT ROVIRA I VIRGILI
A MULTI-OMICS APPROACH TO DESCRIBE CROSSTALK BETWEEN GUT MICROBIOTA AND HOST: EPIGENETICS AND VIROME ROLE
Polina Kazakova Borodina

INTRODUCTION

1 INTESTINAL MICROBIOTA

The human microbiome refers to the collection of all microorganisms that reside on or within human tissues and biofluids, as well as the corresponding anatomical sites in which they reside. The human intestinal tract is considered one of the densest ecosystems, and it is dominated principally by bacteria, making up the bulk of the gut microbiome. In addition to bacteria, the gut microbiome includes three additional domains of life: Archaea, Eukarya (fungi and protozoa), and viruses. The microbiota serves the host by interacting directly or indirectly with host cells and regulating numerous biological pathways involved in immunity and energy homeostasis. Indeed, many researchers consider the gut microbiota to be a virtual organ, and its interindividual variation affords opportunities for precision medicine and personalized nutrition. The healthy adult gut microbiota is generally characterized by rich species diversity, with the dominant phyla being Firmicutes and Bacteroidetes and, to a lesser extent, Actinobacteria, Proteobacteria, and Verrucomicrobia [1]. Although the gut microbiota tends to remain stable over long periods, it can be influenced by various factors.

1.1 Functions of the intestinal microbiota.

The functions of the intestinal microbiota fall into three categories: metabolic, defensive, and trophic [2].

1.1.1 Microbiota involvement in general metabolism.

The gut microbiota is crucial for general metabolism, delivering vital nutrients and bioactive compounds. It also produces essential vitamins, such as B vitamins (cobalamin, folic acid, biotin, thiamine, riboflavin, and pantothenic acid), as well as K vitamins.

Short-chain fatty acids (SCFAs) are the main metabolites produced in the colon by bacterial fermentation of dietary fibers and resistant starch. SCFAs, which are saturated aliphatic organic acids that consist of one to six carbons, include acetate (C2), propionate (C3), and butyrate (C4), of which acetate and propionate are produced mainly by Bacteroidetes and butyrate by Firmicutes. Bacteroidetes produce acetate, which is essential for cholesterol synthesis and stimulates leptin secretion, regulating energy balance and appetite [3]. They also produce propionate, which plays a crucial role in gluconeogenesis. Butyrate is the primary energy source for colonocytes [4] and promotes the proliferation of healthy colonocytes while inducing terminal differentiation and apoptosis in transformed cells. Additionally, SCFAs promote the generation of T cells [5] and the production of antimicrobial peptides, thus supporting the maturation and development of the immune system.

Growing evidence supports the idea that SCFAs also manifest crucial physiological functions in several organs, such as the brain, where they act as major inhibitors of the neurotransmitter gamma-aminobutyric acid (GABA) [6,7]. Another mechanism by which SCFAs regulate systemic functions is through the inhibition of histone deacetylase activity, thus promoting the acetylation of lysine residues present in nucleosomal histones throughout various cell populations [8].

In addition to SCFAs, the organic acids resulting from bacterial metabolism of dietary polyphenols or unassimilated amino acids are also crucial. For example, lactic acid participates in the regulation of intestinal peristalsis [9].

1.1.2 Microbiota as a Host Defense Mechanism Against Infectious Threats.

The microbiota plays a crucial role in maintaining intestinal homeostasis by performing various functions, such as anti-infectious, anti-inflammatory, and immune modulating roles. Nonpathogenic and pathogenic bacteria compete for the same intestinal niches and substrates.

Thus, nonpathogenic bacteria could prevent attachment and subsequent entry of pathogenic bacteria into epithelial cells [10]. Additionally, resident bacteria produce bacteriocins and change the environmental conditions to inhibit the growth of their competitors [11]. On the other hand, the immune response to bacteria relies on innate and adaptive components, such as immunoglobulin secretions.

1.1.3 Trophic.

Apart from metabolic and defensive functions, the gut microbiota also performs trophic functions. It promotes epithelial cell proliferation and differentiation (crypt formation and angiogenesis), stimulates intestinal motor activity and neuroendocrine pathways of gut origin, and has effects on distal organs (central nervous system and liver). The dialog between host and bacteria at the mucosa interface seems to play a part in the development of a competent immune system.

1.2 Modulatory factors of microbiota and host interactions.

As humans radiate across the planet, diverse lifestyles, diets, locations and genetic differences likely lead to variations in microbiota within populations. Several factors are known to affect human gut microbiota composition.

1.2.1 Diet.

Diet remains the most important determinant in shaping the microbiota composition, diversity, and richness even throughout adulthood. Different dietary patterns result in different microbiota communities. A fiber-rich diet is associated with greater richness and diversity. Conversely, a diet rich in proteins and fats, common in Europe, has been correlated with high abundance of Bacteroidetes and decreased levels of Firmicutes [12].

1.2.2 Genetics.

Several associations have been found between the microbiome and genes associated with the host's innate immunity. Pattern recognition receptors sense microorganisms in the intestines, thereby modulating

microbiome composition and microbiome-associated disease [13]. Evidence suggests that host genetics influences the acquisition and development of the infant gut microbiota [14].

1.2.3 Age.

The gut microbiota varies throughout different stages of life. The principal period of establishment and development is the first year of life, after which the microbiota evolves through three distinct phases in the dynamics of the most abundant phyla (Actinobacteria, Bacteroidetes, Firmicutes, Proteobacteria, and Verrucomicrobia). The first phase is the development phase (3-14 months), followed by the transitional phase (15-30 months) and the last phase, the stable phase (more than 31 months) [15]. In preadolescence (7-12 years), the gut microbiota is more diverse and contains significantly more Firmicutes and Actinobacteria than in healthy adults [16]. In adulthood, the gut microbiota remains stable over a long period and, in general, is characterized by rich species diversity. Aging has a significant impact on the gut microbiota, with severe changes in composition and functionality. After the age of 65, microbial diversity declines, opportunistic pathogens increase in abundance and a decrease in species associated with SCFA was observed [17]. Notably, some studies have shown that gut microbiota diversity inversely correlates with biological age but not chronological age. Therefore, what happens in the gut microbiota with advancing biological age can be very different from what happens with chronological age [18].

1.2.4 Antibiotics.

Antibiotics are known to significantly impact the gut microbiota, causing a rapid and long-lasting shift in its taxonomic composition. Broad-spectrum antibiotics often target nonpathogenic strains, resulting in overgrowth of certain taxa and elevated levels of resistant strains [19]. This leads to postantibiotic dysbiosis, which can have various functional effects, such as changes in host-microbiome mutualism and the interaction between species and metabolic activity and regulation. The most prominent changes

associated with antibiotic use include a loss of diversity of the phyla Firmicutes and Bacteroidetes and overgrowth of the family Enterobacteriaceae. However, reduced diversity does not necessarily mean a reduction in the overall number of bacteria, as susceptible bacteria are eliminated and resistant bacteria multiply and take their place [20].

Research has linked early life antibiotic therapy to several negative outcomes later in life, including the development of obesity, allergies, asthma, celiac disease, and inflammatory bowel disease [21,22]. Therefore, it is crucial to use antibiotics judiciously and to consider their potential impacts on the gut microbiota when prescribing them.

1.2.5 Gut Virome.

Viruses are incredibly diverse and abundant biological entities that exist in various environments, including the human body. They possess a genome consisting of either DNA or RNA, which can be either double-stranded or single-stranded, encapsulated in a protein coat called the capsid. Viruses are believed to be the most abundant and diverse biological entities on our planet, with an estimated population of 1031 viruses. They are the most numerous biological entities on Earth and inhabit diverse environments ranging from the oceans to hydrothermal vents to the human body [23]. In the human body, the viral fraction alone accounts for 1013 VLPs (virus-like particles), plus prophages in bacterial genomes [24]. There is much more diversity among viruses than we can find among all the other kingdoms, mainly due to their ability to parasitize many types of organisms. For this reason, the scientific community is investing great effort into understanding their interactions with their guests, and coevolution between parasites and hosts has been observed.

The gut virome is the collection of viruses that inhabit the gastrointestinal tract, particularly the large intestine. The gut virome is made up of many different types of viruses, with bacteriophages being the most abundant. They are viruses that infect bacteria and require bacterial hosts for

propagation. Phages exhibit two distinct life cycles in bacteria: a lytic cycle and a lysogenic cycle. During the lytic cycle, the phage redirects the host metabolism toward the production of new phages, which are released through lysis. In contrast, in the lysogenic cycle, phages integrate their genetic material into the bacterial chromosome and enter a dormant state until the detection of a suitable induction signal, after which the phage genome becomes excised and continues to direct lytic growth [25]. The gut virome is a complex ecosystem of viruses that interact with each other, the host's immune system, and the gut microbiome to maintain a healthy balance [26]. Although research on the gut virome is still in its early stages, recent studies have shown that changes in the gut virome may be associated with a variety of health conditions, including inflammatory bowel disease, obesity, and metabolic disorders [27–30].

Several factors, including diet, antibiotic administration or other medications, host immunity and genetics, geography, and age, can influence the human gut virome composition. However, individuals with the same diet have more similar viral compositions than those with different diets [31], and there is an age-dependent pattern in viral diversity [32].

Thus, the human virome is particularly extensive and, while undercharacterized, is recognized as an integral part of the healthy human ecosystem [28–30,33].

1.2.5.1 Transkingdom interplay and its impact on host homeostasis.

Interactions between viruses and bacteria can have a significant impact on the host’s health and disease, and bacteriophages can influence the structure of the bacterial microbiome through predator–prey relationships [34]. This model predicts that oscillations in prey (bacteria) precede oscillations in predators (bacteriophages). For example, the shift in virome occurs in parallel when the infant microbiota adopts an adult-like composition [35]. Bacteriophages can also indirectly benefit their host by

acting as a "biological weapon" and killing competitor bacteria (Figure 1), thus playing a protective role in the vacant ecological niche of the newborn gut. However, it is essential to consider that phages can also transfer genes to bacteria, modifying their phenotype (e.g., antibiotic resistance or virulence factors). This phenomenon highlights the potential risks associated with the indiscriminate use of antibiotics in early life stages and the need for the responsible use of antibiotics in clinical settings.

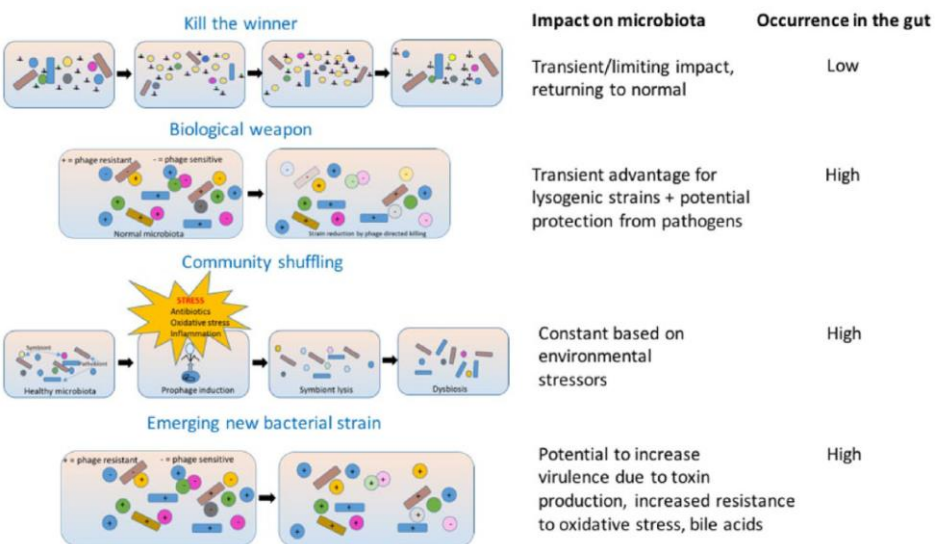


Figure 1: Proposed mechanisms underlying phage-mediated intestinal dysbiosis. Taken from Mukhopadhya et al. 2019.

1.2.6 Epigenetics.

The relationship between the intestinal microbiota, the immune system and the intestinal epithelium is complex and dynamic. The intestinal microbiota produces a wide range of metabolites that may have epigenetic functions either as regulators or as substrates and can induce epigenetic modifications in the intestine, such as DNA methylation. Another epigenetic mechanism that operates in the intestine functions through microRNAs (miRNAs), which are short (approximately 22 nt) single-stranded noncoding RNAs that control gene expression by binding to the 3'-untranslated regions (3'UTR) of other regulated transcripts. This binding can result in mRNA degradation or translation of its target gene. Previous

studies have demonstrated that miRNAs produced by intestinal epithelial cells can modulate gene expression posttranscriptionally and shape the gut microbiota via extracellular vesicles (EVs), which contain miRNAs [36]. In another study, the authors demonstrated that bacterial gene expression is directly altered when bacteria are cultured with human or mouse fecal miRNAs [37]. Analogous to how human cells produce EVs, bacteria also produce membrane vesicles or outer membrane vesicles (OMVs), which function similarly to EVs by mediating cell-to-cell exchange with small signaling molecules inside. The endogenous microbiota has been found to affect how miRNAs are expressed [38], although it is important to note that the secondary structure of miRNAs contained in OMVs does not resemble that of eukaryotic miRNAs. As a result, this miRNA-like molecule cannot directly regulate gene expression or interfere with human miRNA modulation [39]. Further studies are needed to validate these different hypotheses. Similar to protein regulators of gene expression, miRNAs play important roles in the control of developmental, differentiation and disease processes. Certain viruses have acquired and manipulated host miRNA genes to enhance infection, while others harness specific host miRNAs for viral function, such as genome replication [40]. Therefore, modulation of miRNAs by the intestinal microbiota may have the potential to affect the expression of many host genes, particularly in disorders where the composition of the microbiota is shifted toward less favorable species. This could represent a novel therapeutic approach.

1.2.7 Others.

Numerous factors have been found to impact the IM. Exercise and physical activity have been shown to enhance microbiota diversity and are positively correlated with protein intake and creatine kinase levels. In particular, athletes are exposed to external factors such as healthy lifestyles, as well as intrinsic adaptations to endurance training such as tissue hypoxia [41]. There are also discernible patterns of functional maturation of the gut communities in infants and children living in geographically and culturally

distinct settings. For example, a study of rural groups in Malawi and Venezuela found stark differences in the microbiota composition of nonmodern societies compared to Western groups [42]. Specifically, the Bacteroidetes to Firmicutes ratio was higher in these nonmodern societies. Therefore, when evaluated globally, the human gut microbiome can provide crucial information about how environmental and cultural variables affect the richness and makeup of the intestinal microbiota.

1.3 Acquisition and development of intestinal microbiota in early life stages.

The symbiotic relationship between infants and their gut microbiota is established at birth and evolves during the first few years of life. The establishment of a stable adult-like microbiota occurs between 2-4 years of age and is dominated by Firmicutes and Bacteroidetes [43]. As an infant's gut microbiota develops over time, there is a gradual increase in its alpha diversity, although its beta diversity is reduced by the age of 12 months (as it is higher at birth than in adults) [44]. These dynamic processes of development and maturation are influenced by various perinatal conditions, including external factors such as illness, antibiotic treatment, and changes in diet.

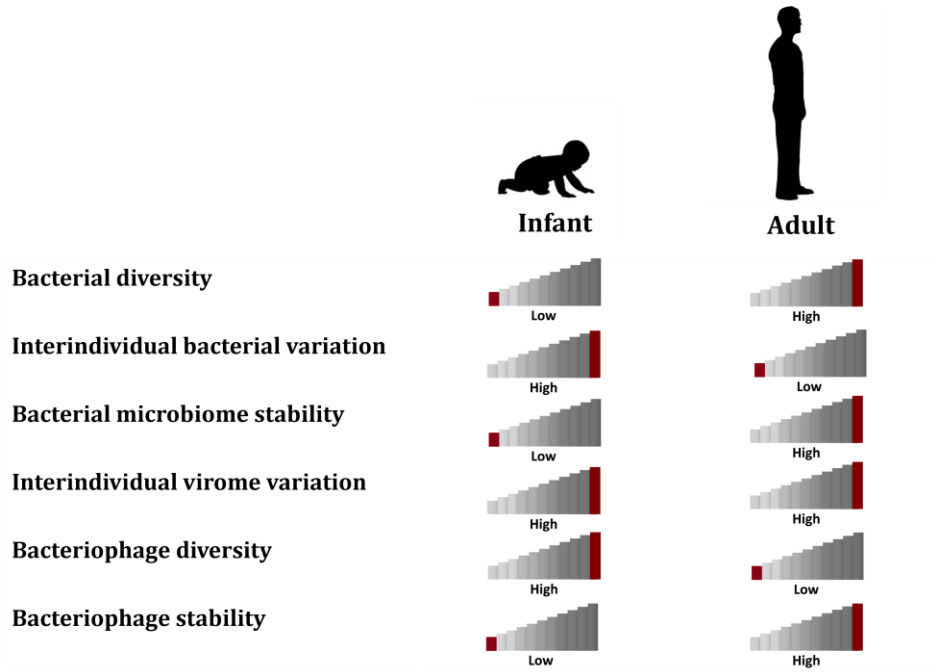


Figure 2: Composition of the human gut bacteriome and virome during infancy (under one year old) and adulthood (over eighteen years old). Adapted from Lim et al. 2016.

1.3.1 Main drivers of early life colonization.

During pregnancy, various factors can influence the composition of the maternal microbiota, including antibiotic treatment, diet, infections, stress, and host genetics. At birth, the infant's gut microbiota is primarily established through vertical transmission from the mother. The mode of delivery (vaginal delivery or cesarean section) and gestational age are important factors in this transmission process. Additionally, the type of feeding is a major determinant of the early microbial composition in infants.

1.3.1.1 Colonization of the fetus by the maternal microbiota.

In the past, it was believed that newborns were born completely sterile and that colonization of the newborn intestinal tract began during and after birth. However, several studies have suggested that fetal colonization may occur before birth. While there is no direct evidence to confirm this, and since the mechanisms by which bacteria pass from the mother to the fetus are still unknown, the results obtained thus far have

been attributed to possible contamination [45]. However, some studies have found bacteria or bacterial DNA in the placenta, amniotic fluid and fetal membranes, which are potential sources of the maternal microbiota that can colonize or infect the fetus [46,47] [48]. Additionally, meconium, the first stool of a newborn, has also been found to contain bacteria that are similar to those found in the mother's placenta [49–51]. These bacteria are typically nonpathogenic commensals belonging to the Firmicutes, Proteobacteria and Fusobacteria phyla [52]. However, compared with the adult fecal microbiota composition, meconium has lower species diversity, greater variation between subjects, and a higher proportion of Proteobacteria [53]. All of these findings support the idea of in utero colonization, although more investigation is required to fully understand the mechanisms involved.

1.3.1.1.1 Virome acquisition during early development.

For a long time, the inability to culture microorganisms from samples taken from healthy deliveries has supported the idea that neonates are born sterile. However, recent studies have challenged this hypothesis. It was first documented that pathogenic viruses such as human cytomegalovirus and herpes simplex can be vertically transmitted during pregnancy, but these viruses are associated with disease states, not health. However, several studies of the gut virome during infant development have shown that the colonization of viral populations is highly dynamic. Some studies detected VLPs in meconium samples taken shortly after birth and found that viral diversity increased within just a few hours after delivery [54]. These results suggest rapid colonization after rupture of membranes and delivery.

During the first few months of life, the gut phagosome community structure is composed primarily of a rich and diverse collection of phages, with the majority belonging to the Caudovirales order. However, as age increases, phage richness decreases, and by the age of two years, the composition

becomes dominated by Microviridae [35]. In contrast, eukaryotic viruses show low diversity during the first days of life, with an increase over two years before decreasing in childhood, suggesting that these viruses are primarily obtained from environmental sources, presumably due to an underdeveloped immune system [32]. Certain virus families, such as Picornaviridae, Adenoviridae, Reoviridae and Anelloviridae, exhibit sporadic distribution over time and are associated with seasonal infections such as influenza [55]. When comparing the abundance of viruses with respect to their lifecycle, it has been shown that lytic phages, which include Microviridae and crAssphages, are relatively rare in the first month of life but become more common later [54].

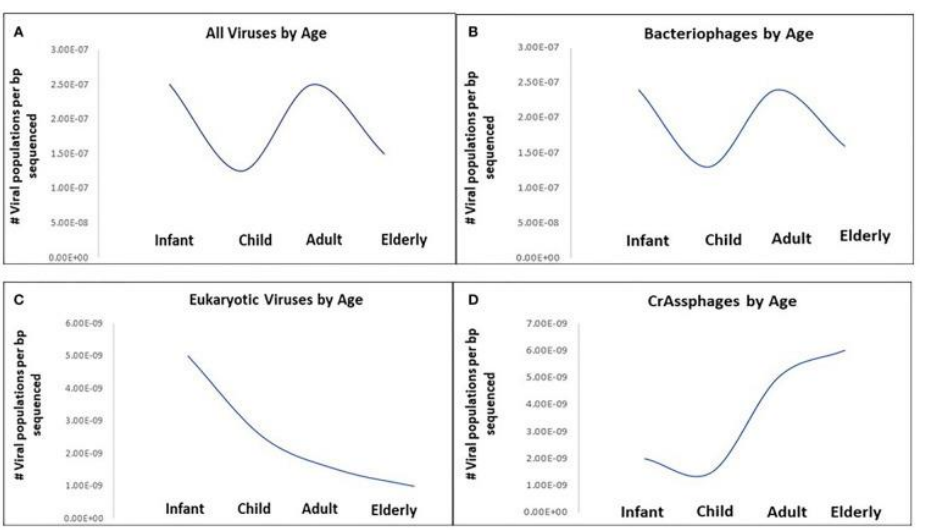


Figure 3: Variation in virome composition throughout a person's life. The virome diversity in healthy individuals changes with age across four age groups: infants (0–3 years), children (3–18 years), adults (18–65 years), and the elderly (65+). Taken from Spencer et al. 2022.

Previous research has investigated the impact of diet on the gut virome, and findings suggest that the infant virome may be partially transmitted from the mother through breast milk [56]. Additionally, breast milk can contain antiviral components, including maternal antibodies, which can provide protection against intestinal infections.

1.3.1.2 Mode of delivery.

Several studies have highlighted the impact of the mode of delivery on early life gut microbiota development, recognizing it as an important driver. Generally, vaginally delivered infants tend to harbor gut microbiota similar to their mother's vaginal and fecal microbiota, characterized by Prevotella, Sneathia and Lactobacillus genera [57]. Conversely, infants born via cesarean section are not directly exposed to maternal microbiota, and their fecal microbiota reflects the maternal skin and oral microbiota (or even the hospital environment), which are dominated by *Propionibacterium*, *Corynebacterium*, and *Streptococcus* genera [44]. In addition, these studies have demonstrated that cesarean-section delivered infants exhibit lower alpha diversity and are less frequently colonized by members of Bifidobacterium and Bacteroidetes, while being more often colonized by members of Clostridium [44].

1.3.1.3 Gestational age.

Gestational age at birth is a crucial determinant of infant microbiota colonization. Preterm infants, regardless of the degree of prematurity, are exposed to unique environmental conditions that can pose significant health challenges. The use of antibiotics and parenteral feeding are some of the factors that can disrupt the natural pattern of microbiota acquisition, leading to an aberrant establishment or deviating composition of the intestinal microbiota. The intestinal microbiota of preterm infants exhibits higher levels of facultative anaerobic bacteria, including *Enterobacteria*, *Enterococcus* and *Lactobacillus*, but lower levels of anaerobes, including *Bifidobacterium* and *Bacteroides*, compared to term infants [58]. The immature microbiota can pose a significant threat to preterm infants, as their immune and nervous systems are not yet fully developed, which can affect their short- and long-term health..

1.3.1.4 Feeding type.

The infant feeding mode, which can be breastfeeding, formula feeding or a combination of both, is a key factor influencing microbiota

colonization in early life. Extensive research has been conducted on the differences between these two types of feeding, which have been widely recognized and documented.

Exclusive breastfeeding is recommended by the WHO (World Health Organization) for the first 6 months of life, followed by supplemental breastfeeding for up to 2 years and beyond. The health benefits of breastfeeding have been known throughout history and are largely dependent on the duration of breastfeeding and the age at which complementary foods are introduced. Human milk contains a variety of nutrients and pro-microbial and antimicrobial agents, such as lysozyme and lactoferrin, as well as maternal IgAs, which play a protective role, and endo-cannabinoids, which promote a regulatory and more tolerogenic immune system [59]. Human milk oligosaccharides (HMOs), such as galactooligosaccharides, are the third most abundant solid component after lactose and lipids. They cannot be digested by infants, but some of the infant's gut bacteria, including Bifidobacterium, Lactobacillus and Bacteroidetes, can degrade these HMOs and obtain smaller molecules. These smaller molecules, mainly those fermented by Bifidobacterium, are used to produce short-chain fatty acids (SCFAs) [60], which play an important role in colon health.

Formula-fed infants are exposed to different carbohydrates and micronutrients and are not exposed to maternal skin microbiota. These differences result in distinct microbial colonization patterns of the gut. Typically, formula-fed infants experience an early shift toward an adult-like microbiota composition [44]. There are substantial differences in the gut microbiota compositions of breastfed vs. formula-fed infants. Breastfed infants contain higher levels of *Bifidobacteria* and *Lactobacilli* and lower levels of potential pathogens than their formula-fed counterparts. *Enterococci* and *Clostridia* primarily constitute the microbiota of formula-fed infants. Another study suggested that the microbiota of formula-fed infants is enriched by an anaerobic community, while the microbiota of

breastfed infants is typically composed of aerobic organisms [61]. Breastfed infants exhibit lower alpha diversity and different beta diversity measures compared to their formula-fed counterparts during the first year of life [62].

The role of the microbiota in the development of some illnesses, such as atopic disease, inflammatory bowel disease or even type 1 diabetes, has been well established [63,64]. Therefore, although the microbial alterations induced by formula feeding may be minor, they could potentially contribute to the development of these diseases later in life.

Taking all of this evidence into account, the supplementation of formula milk with prebiotics and probiotics has shown promise in altering the microbiota composition toward a more desirable breastfeeding pattern and stimulating an immune response [65]. There are studies in which milk formula has been improved by the inclusion of specific oligosaccharides, resulting in the establishment of a Bifidobacterium-dominant microbiota that is more similar to that of breastfed infants [61].

1.3.1.5 Introduction of solid food.

The timing of solid food introduction into an infant's diet is a crucial factor in microbiota development. The European Society for Pediatric Gastroenterology, Hepatology and Nutrition (ESPGHAN) recommends introducing solid food no earlier than 17 weeks of age and no later than 26 weeks. Changing an infant's diet can shift the types and prevalence of the microbial community in the gut, as certain phyla/species may be better adapted to utilize specific substrates. Gut microbiota produce degradative enzymes for fermentation and breakdown of nondigestible carbohydrates, such as plant cell wall polysaccharides, celluloses, and xylans, which are not encoded by the mammalian genome [66]. The introduction of solid food initiates the maturation of infant gut microbiota toward that of an adult, which presents an increase in functional genes associated with carbohydrate utilization and vitamin biosynthesis [67]. Along with the introduction of solid foods, there is a large shift in the gut microbiota from

Bifidobacterium dominant to Bacteroidetes and Firmicutes dominant [68]. Moreover, studies have validated higher levels of SCFA after the introduction of solid foods to infants [67].

1.3.1.6 Environment.

The influence of environmental factors on the colonization pattern of infant gut microbiota is an area of ongoing research, and the effects of family size, structure, and birth order remain to be established [69]. Additionally, geographical location may indirectly impact early-life microbiota by shaping dietary habits and lifestyle [70].

1.3.1.7 Antibiotic.

Antibiotic treatment during early life can have significant effects on the development of the microbiota. Shockingly, more than 50% of children aged 0-18 receive antibiotics [71], which can disrupt the actively developing infant gut microbiota, leading to potential long-term health problems. Even three months after amoxicillin treatment, disruption of the microbiota can be observed [72]. Antibiotics in infants can shift the composition of the microbiota, increasing the abundance of potentially pathogenic Enterobacteriaceae, reducing the percentages of microbial taxa associated with a healthy microbiota, such as Lactobacillales and Bifidobacteriaceae [73], and lowering overall microbial diversity. Epidemiological studies suggest that antibiotic use in early life is linked to an increased risk of developing allergic diseases such as asthma, atopic disease, and type 1 diabetes [74].

1.4 The holistic study of the intestinal microbiota: multimetaomic approaches.

The application of different omics offers tremendous potential in uncovering the components and relationships between the intestinal microbiome and diseases. By analyzing multiple types of data simultaneously, including genomics, transcriptomics, proteomics, and metabolomics, researchers can gain a comprehensive understanding of the

microbial community's functional capabilities and how it interacts with the host organism. This can lead to the identification of new microbial diagnostic markers and the development of personalized treatments for various diseases. These terms change to metagenomics, metatranscriptomics, metaproteomics, and metabolomics when omics are used to study the collective genetic material, gene expression, proteins, and small molecules of the complete microbial communities, rather than individual organisms.

Metagenomics, which involves the examination of all genetic material recovered directly from environmental or living samples, has provided crucial insights into the characterization of gut microbiota [75]. This can help to identify specific strains and their functional potential, such as the ability to produce beneficial metabolites. The characterization of gut microbiota requires a robust assortment of instrumental and analytical techniques and statistical analyses to assist researchers in preparing, analyzing, and interpreting results from the microbiome. Within the maturation of modern microbiology and the development of next-generation sequencing technologies, research has increasingly focused on the complex microbial communities that interact with the host to influence disease processes. Classical studies of the gut microbiota have been limited by cultivation techniques that can only capture 10-30% of the gut microbiota biodiversity. The introduction of Sanger sequencing technology in the late 1970s greatly impacted first-generation sequencing due to its simplicity and precision, although it presents some limitations because it is time-consuming and expensive to identify microbial biodiversity, especially for complex microbial communities. Alternative sequencing technologies such as next-generation sequencing (NGS) have been developed to overcome these limitations. NGS offers higher throughput sequencing platforms capable of generating hundreds of thousands of sequence reactions in a faster and more cost-effective way. However, third-generation sequencing has made the greatest advances in nanotechnology by utilizing

nanopore sequencing, a technique that involves the use of nanoscale pores to read single-stranded DNA as it passes through the pore. This technique enables real-time sequencing of DNA by measuring changes in electrical current as each nucleotide passes through the pore.

Currently, the gold standard technique for metagenomics studies of bacterial taxonomy is the analysis of 16S rRNA gene sequences by NGS. This gene is present in all microbes and, while highly conserved, has diverged over time to provide unique sequences that can be used to identify specific taxonomies and determine the frequency of each member within a community. Although partial sequencing of the 16S rRNA gene can typically identify the genus of a bacteria, it is less reliable for determining the species (with an accuracy ranging from 65% to 83%) [76]. Another approach commonly used for microbial analysis is shotgun metagenomics, which involves nontargeted sequencing to identify all microbial genomes present in a sample. This technique provides similar information to 16S rRNA gene amplicon sequencing (i.e., taxonomy and abundance) but also allows for the identification of gene coding sequences and functional annotation by examining the functions associated with identified genes [77]. However, there are several limitations to this technique, including the inability to distinguish between microbial expressions, the need for higher sequencing coverage than 16S rRNA gene sequence analysis, and the requirements for substantial quantity and high quality of DNA samples [78]. The typical workflow for microbial analysis using either of these techniques based on NGS involves five main steps: cell lysis, DNA isolation and extraction, amplification and sequencing of the amplicon libraries and bioinformatic analysis. To support these studies, it is important to continually update curated reference databases and use computational sequencing algorithms. Choosing the right database is therefore a crucial step in ensuring the accuracy and relevance of the results obtained.

As a result of these technical advances, metagenomics is an extremely powerful tool, but it cannot provide information on the regulation of gene

expression or reveal the activity of the microbial community in a defined environment. The metatranscriptomics approach uses a similar analytical concept to shotgun metagenomics, but it focuses on RNA transcribed from microbial cells and selective removal of the interfering nucleic acids (DNA, t-RNA, rRNA), allowing for evaluations of the expression activities of these organisms [79]. A standard RNAseq workflow includes the isolation of total RNA from microbiome samples, RNA enrichment, fragmentation, cDNA synthesis, and preparation of transcriptome libraries for sequencing [80]. One of the major limitations is separating mRNA from other abundant RNA, such as rRNA. Similar to metagenomics, bioinformatic data analysis is fundamental to the comprehensive functional characterization of RNA molecules.

Related to transcriptomics is the study of miRNA (microRNA): RNA-seq-based technology is also the most recent significant method for miRNA analysis, enabling simultaneous detection of all miRNAs as well as other small RNA species. Similar to the metatranscriptome workflow, small RNA-seq involves the isolation of total RNA, small RNA enrichment/isolation, cDNA synthesis, library construction and sequencing [81–83]. Similar to many other omics sciences, miRNA computational analysis is complex and involves numerous steps and analysis tools [83].

Metabolomics, defined as the systematic measurement of small molecules (metabolites), focuses on profiling the metabolites produced by the microbiota, as well as by the host itself, which can provide insights into the metabolic functions of the microbiota and its interactions with host metabolism. Identification of the metabolome in blood, urine, and feces enables a comprehensive elucidation of the ongoing physiological processes and the investigation of the cross-talk between host-microbiome and among microbial species [84].

In addition to providing insights into the basic biology of the microbiota, metabolomics can also be used to identify potential biomarkers of disease and to develop new diagnostic tools and treatments.

Metabolomics can be divided into two main strategies: targeted and nontargeted analysis. Targeted metabolomics is the analysis of a specific set of metabolites using quantitative methods. Nontargeted metabolomics, on the other hand, is an untargeted approach that aims to identify as many metabolites as possible, providing a rapid snapshot of the metabolic profile of samples. This approach is often used when the interest is to discover new metabolites or biomarkers associated with a particular disease or physiological condition. There are different detection methods and analytical platforms used in metabolomics, including mass spectrometry (MS) and nuclear magnetic resonance (NMR) technologies. MS-based metabolomics relies on the separation and detection of metabolites based on their mass-to-charge (m/z) ratio, has higher sensitivity and selectivity than NMR, and can be used to allow targeted and untargeted metabolomics [85]. In contrast, NMR-based metabolomics detects metabolites based on their unique nuclear magnetic resonance spectra and is mainly useful to determine metabolic fingerprints for the identification and quantification of compounds in nontargeted metabolomics. NMR is a nondestructive technique that requires minimal sample preparation and that can analyze samples directly, with high reproducibility, although it has less sensitivity than MS methods, detecting only metabolites present in high concentrations [86].

A major challenge in omics microbiome research stems from the considerable computational effort required to process the enormous amounts of data generated by the continuous improvement of next-generation sequencing and mass spectrometry technologies. On the other hand, such integration of omics science will provide more evidence regarding biological mechanisms, ultimately opening new prospects for the development of new therapeutic strategies and personalized medicine.

1.4.1.1 Methodology for virome study.

Recent investigations of the human virome, utilizing metagenomic sequencing and other techniques, have elucidated features of human virome diversity at different body sites, the relationship to disease states, and mechanisms of human virome establishment during early life. In the past, viruses were characterized by electron microscopy and culture-based methods [87]. Recent advances in NGS, combined with advanced bioinformatics tools, make it possible to characterize virome richness, gene functions, and associations with disease phenotypes. However, there are several methodological and conceptual challenges to studying the virome, including the potential for contamination, the difficulty of distinguishing between active and inactive viruses, and the complex interactions between viruses and the host immune system. Focusing on the methodologies, it is important to remark that viruses do not contain a conserved genomic region analogous to the 16S rRNA gene in bacteria, so viromics require whole genome sequencing (WGS) methodologies, which are more costly and time-consuming than other sequencing methods, such as 16S rRNA gene targeted sequencing. Moreover, in terms of analysis, WGS data can be more complex and challenging to analyze than other types of sequencing data [88]. Another challenge is the proportion of viral genomic material, which is small in comparison with total nucleic acids in microbial communities: a single contaminating host cell is the equivalent of half a million virions of some species in total nucleic acid length because viruses have a small genome (1.8 Kbp-2.5 Mbp), so enrichment methods are required to recover the most abundant community members. In VLP (virus-like particle) enrichment [89], a common approach for removing large particles such as host and bacterial cells from samples is to use centrifugation at a lower speed and subsequent filtration with 0.4-0.2 μm filters (which can retain large viruses and reduce the amount of recovered virus): the final step is concentration, which might be performed with precipitation or ultracentrifugation. Depending on how the method is performed, it can be biased toward isolating specific phage types, and some large viruses can also be retained,

reducing the final recovered virus. Finally, DNase and RNase treatment removes unencapsulated exogenous DNA/RNA. After VLP enrichment, it is necessary to use two different protocols to create libraries and analyze DNA and RNA viruses separately. The raw sequencing data were processed using bioinformatics tools to identify viral sequences, assemble viral genomes, and classify the viral taxa. Bioinformatic analysis poses another major challenge due to weak, inaccurate, or insufficient annotations of many viruses in public databases and limited homology between viral sequences and reference databases [90]. Generally, 40-90% of viral sequences cannot be taxonomically assigned by alignment to any reference nucleotide or amino acid sequence and are named “viral dark matter” [91]. To address these issues, major viromics research relies on de novo assembly to recover viral genomes from metagenomes. The latest problem is related to viral taxonomy because numerous viruses lack classification at all taxonomic levels. Viral taxonomy is especially complex, as, traditionally, viruses were classified by their morphology or host, and viral taxonomic distinctions may not necessarily rely on biological or phylogenetic characteristics [92]. In summary, there are different critical points in virome studies, starting in sample collection, where a considerable biomass of samples is needed, and maintaining the samples at -80 °C is necessary to preserve the integrity of the virions until sample analysis. Next are the requirements for an enrichment step and nucleic acid extraction, being extremely careful to avoid environmental contamination, and the situation that the step of preparing the genomic libraries can generate a bias depending on the protocol used [93]. Despite these serious methodological constraints, virome studies are growing and offer great potential to provide valuable insights into the diversity, ecology, and evolution of viral communities in gut microbiota: examples include the recent study which revealed that the gut virome is more diverse than previously thought, with an estimated 140,000 different viral species present in the human gut [94], and the discovery of crAssphage, which is the most abundant human-associated virus known (90% of sequences in gut virome) [95].

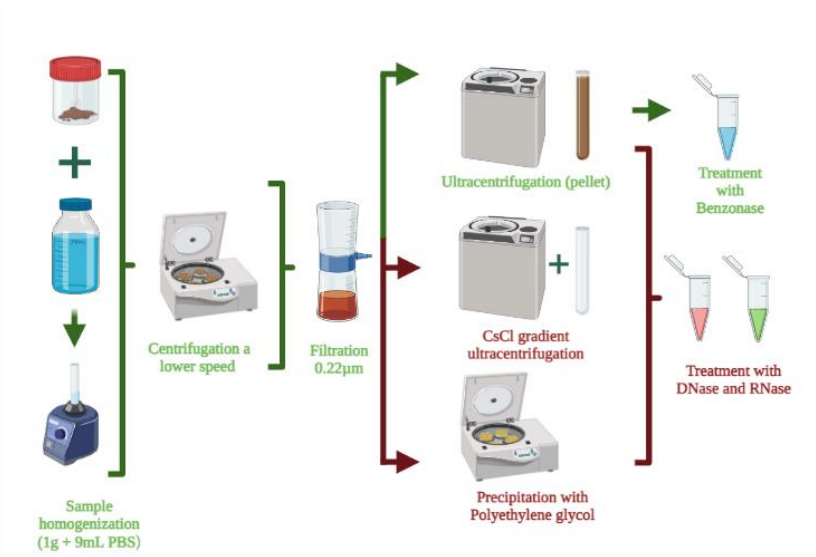


Figure 4: Schematic representation of the different steps in the VLP enrichment protocol that were compared in the study.

2 CELIAC DISEASE

Celiac disease (CeD) is an autoimmune disorder that occurs in genetically predisposed people who develop an immune response triggered by the ingestion of gluten proteins [96]. The clinical presentation of CeD is diverse and can range from severe to asymptomatic [97]. While the primary clinical manifestations (Figure 5) are related mainly to the small intestine, there are also some extraintestinal symptoms that can occur. CeD can be classified into several forms based on symptomatology, including classical (the most common), nonclassical, latent, potential/screen-detected and refractory [98]. The classical form is associated with intestinal manifestations, such as chronic diarrhea, weight loss and malabsorption, and is most prevalent in early childhood. The nonclassical form is characterized by extraintestinal symptoms, including fatigue, iron deficiency and osteoporosis, and it typically appears in adulthood [99]. The latent form is characterized by the absence of clinical manifestations but positive serological tests and genetic markers associated with CeD. The refractory form is the most difficult to treat, as patients are unresponsive to gluten-free diets.

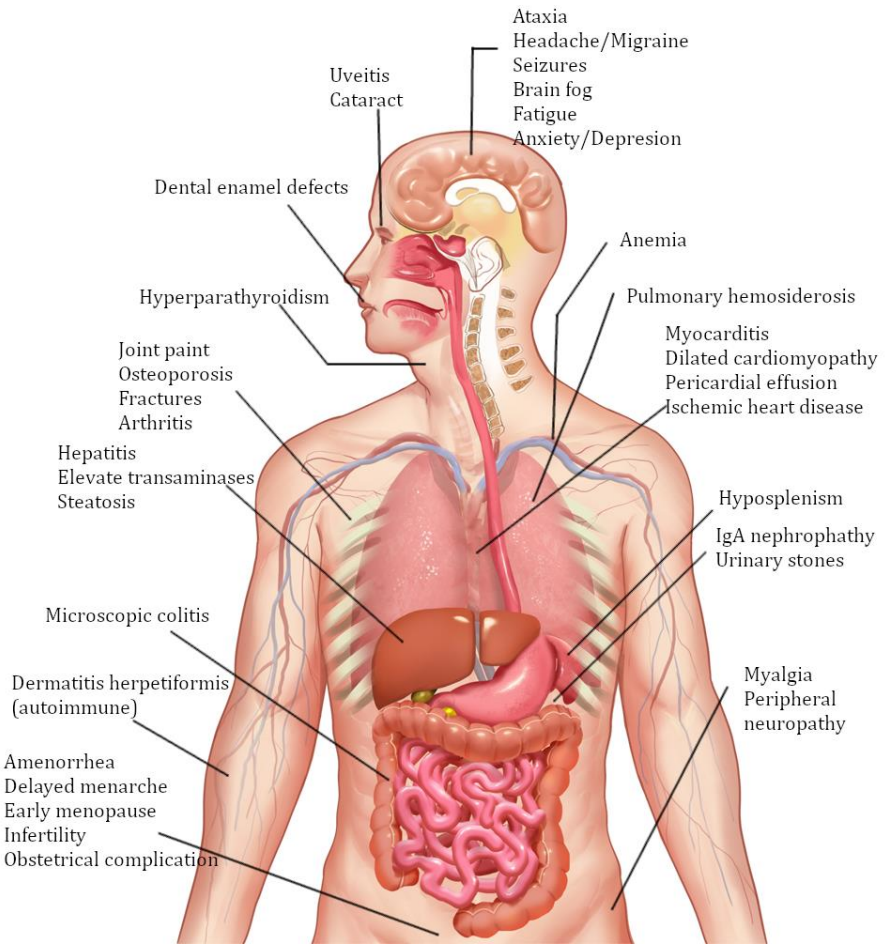


Figure 5: Extraintestinal manifestations and associated conditions linked to CeD. Adapted from Therrien et al. 2021.

Several screening studies of CeD in Europe, America, Asia, and Africa have been conducted to investigate the physiopathology and incidence of this disease [100,101]. While improvements in the diagnosis of CeD have contributed to some extent to the observed increase in disease prevalence and incidence worldwide, it is likely that a true rise in the incidence rate is also a factor [102]. The global prevalence of CeD is currently estimated to be approximately 1%, but it varies across geographic regions, with the highest incidence rates reported in the western Sahara region at 5.6% [103]. However, it is important to note that different diagnostic methods may impact the reported prevalence rates. For instance, if the diagnostic method relies on seroprevalence, the prevalence may be overestimated at 1.4%

because these tests can cross-react with other autoimmune conditions. Conversely, if the diagnostic method involves duodenal biopsy analysis, the prevalence may be underestimated at 0.7% because not all forms of CeD present with intestinal atrophy. This lack of a standardized diagnostic method can be attributed to what is commonly referred to as the "iceberg theory". In 1991, Richard F. Logan first used the iceberg metaphor to illustrate the diagnostic challenges associated with CeD. The visible part of the iceberg represents the diagnosed cases of CeD (which are a minority), while the submerged part of the iceberg represents asymptomatic and latent forms of the disease that are not immediately visible.

CeD can only be managed through strict adherence to a lifelong gluten-free diet (GFD). Within a few weeks of starting this diet, the primary symptoms of CeD typically improve, as the absence of dietary gluten allows for the normalization of serological markers and restoration of atrophy in the intestinal microvilli. However, the GFD has many disadvantages, the primary one being the difficulty of strictly following it, as many food products and medicines contain traces of gluten [104]. Nevertheless, in recent years, the gluten-free products market has expanded considerably (it was valued at USD 5.9 billion in 2021) [105], offering a wide range of gluten-free foods. Despite this, the GFD requires nutritional supplementation due to vitamin and mineral deficiencies. Consequently, alternative treatments are being investigated, including endopeptidases to digest gluten [106], transglutaminase 2 inhibitors [107], gluten vaccines [108], and others. These emerging therapeutic options have the potential to be tested in clinical trials.

People with CeD are at elevated risk of developing intestinal malignancies, with lymphoma and small bowel adenocarcinoma being the most common [109,110]. Even with treatment, CeD patients have a 2-fold greater risk of coronary artery disease and a 4-fold higher risk of small bowel cancers. Moreover, several studies have documented an association between CeD and increased mortality, likely due to chronic inflammation, energy, and

vitamin malnutrition. In fact, cardiovascular disease and cancer are the primary causes of death in individuals with celiac disease [111].

2.1 CeD pathophysiology

Celiac disease is primarily caused by an inappropriate adaptive immune response to gluten-derived peptides. Wheat gluten protein has been extensively studied [112], and it is composed of two subunits: the first is composed of monomeric gliadins, and the second is formed by polymeric glutenins. These subunits contain high levels of proline and glutamine amino acids, making gluten protein resistant to proteolytic degradation by gastrointestinal enzymes.

When gluten is consumed in individuals with CeD, it triggers an immune response in the small intestine, resulting in symptoms such as intestinal villus atrophy [113,114]. This immune response causes gluten peptides to pass through the intestinal epithelial barrier, reaching the lamina propria. Once in the lamina propria, tissue transglutaminase deamidates gluten peptides, which are then presented by DQ22+ or DQ8+ antigen-presenting cells to pathogenic CD4+ T cells, which drive a T helper cell type 1 (TH1) response. This leads to the development of celiac lesions, causing intraepithelial and lamina propria infiltration by inflammatory cells, crypt hyperplasia, and villus atrophy.

Thus, the overactivation of various immune processes occurs via different mechanisms (Figure 6). First, CD4+ T cells can stimulate B cells to produce autoantibodies such as transglutaminase 2 (TG2). These antibodies induce changes in the cytoskeleton of enterocytes through actin redistribution, resulting in cell damage. Second, the production of proinflammatory cytokines by CD4+ T cells induces TH1 cells to immediately produce interleukin-21 (IL-21) and interferon-γ (IFN-γ), exacerbating intestinal inflammation. This also leads to fibroblast or lamina propria mononuclear cell secretion of matrix metalloproteinases, which are responsible for tissue remodeling and result in villus atrophy and crypt hyperplasia. Another

contributing factor is the production of IL-15 by intestinal cells, which stimulates the production of IFN-γ and contributes to the selective expansion of the intraepithelial lymphocyte population independently of antigen presentation. This indicates the crucial role of IL-15 in the inflammatory process and the resulting lesions in the intestine.

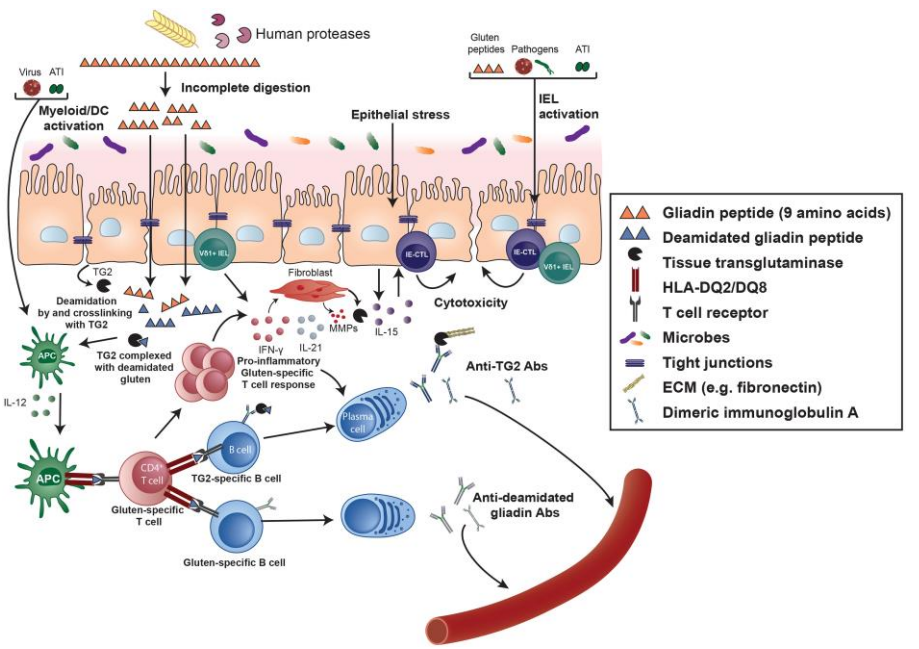


Figure 6: Pathogenesis in celiac disease. Taken from Verdu et al. 2021.

2.2 Risk Factors

2.2.1 Genetic and Immunogenetic Factors

There is much evidence supporting the role of genetics in the development of CeD. Over 90% of patients with CeD are DQ2 positive, and most of the others are DQ8 positive. It is important to note that reports on a patient's HLA status (HLA-DQ2 positive or HLA-DQ8 positive) should include the presence of both haplotypes and allelic components [115].

HLA contributes to an estimated 35% of the overall genetic predisposition for CeD. Genome-wide association studies have revealed that another 18% is attributed to non-DQ HLA genes [116]. The remaining 90% of genetic predisposition is associated with polymorphisms of non-protein-encoding

DNA regions [117]. Notably, there is significant overlap in single nucleotide polymorphisms (SNPs) between CeD and other autoimmune diseases, such as autoimmune thyroid disease. This is likely because these diseases are also tightly linked to HLA-DQ2 and HLA-DQ8 [118]. The human genome comprises approximately 30,000 genes and over 20 million SNPs. The first genome-wide association study (GWAS) for CeD was conducted in 2007 in England, analyzing more than 310,000 SNPs in 767 cases and 1,422 controls. Regions containing eight SNPs demonstrated the strongest association with CeD, with these SNPs found to be related to the genes encoding two cytokines (IL2 and IL21) involved in the proinflammatory immune response [119].

Genetic investigations of CeD have identified 42 non-HLA loci associated with the disease [117]. However, the discovery of additional biomarkers, such as differential expression in RNA sequencing analysis [120], is necessary. Moreover, in the most recent epigenetic analysis, aberrant DNA methylation was found in multiple genes that exert a role in immunity and inflammatory responses, including the proinflammatory TNFα and IL21 pathways [121].

2.2.2 Environmental Factors

2.2.2.1 Diet.

A previous study on Swedish children [122] examined the interplay of several factors in infant feeding, including the amount of gluten given, the timing of gluten introduction, and whether breastfeeding was ongoing when gluten was introduced. The study found that the timing of gluten introduction alone was not a significant factor in the development of CeD.

Based on these data, the European Society for Pediatric Gastroenterology, Hepatology and Nutrition [123] recommends introducing gluten between 4-7 months of age and avoiding gluten introduction while the infant is still being breastfed.

Although there is some debate about the protective effects of breastfeeding and its optimal duration, research suggests that human milk oligosaccharides play a role in preventing CeD by promoting the growth of beneficial bacteria, such as *Bifidobacterium spp*., and by preventing the growth of potential pathogens, such as *Clostridium difficile* [124]. In addition, human milk oligosaccharides enhance barrier integrity by making enterocytes less vulnerable to bacterial-induced inflammation.

2.2.2.2 Other risk factors

The results of a longitudinal study showed that children with a high frequency of rotavirus infections may be at increased risk for developing CeD if they are genetically predisposed [125]. In addition, another study found that infections with nonpathogenic reovirus might trigger CeD by disrupting intestinal immune homeostasis [126]. The role of intestinal infections in developing CeD is further described in section 2.3.2.

Another study linked antibiotic use with an increased risk of developing CeD, potentially due to intestinal dysbiosis [127]. However, there is currently no evidence linking in utero exposure to antibiotics with an increased risk of CeD [22]. Other drugs, such as interferon and ribavirin, may activate CeD [128].

2.3 Role of the intestinal microbiota in CeD.

Actual data suggest a multifactorial etiology of CeD in which the gut microbiota appears to be a critical player.

2.3.1 Dysbiosis in CeD.

The gut microbiota composition is established in the first two years of life and remains relatively stable in healthy adults [44]. It is crucial to maintain a diverse and well-balanced intestinal microbiota for good health. Pathogenic and symbiotic microbiota coexist in a healthy body; however, disturbance in this balance can lead to dysbiosis for various reasons, which has been identified as a risk factor for CeD [129]. Studies of fecal samples and duodenal mucosa biopsies showed differences in the microbiota

composition between children and adults with active CeD, patients with CeD treated following a gluten-free diet (GFD), and healthy controls [130] [131] [132,133]. These studies have demonstrated an imbalance in the Gram-negative to Gram-positive bacteria ratio, characterized by higher levels of Bacteroides (P*revotella*), Proteobacteria, and *Clostridium* and lower levels of *Bifidobacteria, Lactobacillus, Streptococcus,* and *Staphylococcus* in CeD individuals compared to healthy controls. Conversely, another study found that *Barnesiella* was higher in healthy controls than in individuals with CeD or their first-degree relatives, while *Dorea, Akkermansia* and *Prevotella* genera were more prevalent in fecal samples from controls than in those from CeD patients [133]. Interestingly, the diversity of stool microbiota was found to be significantly higher in CeD children than in healthy controls [134]. The concept of dysbiosis as a risk factor for CeD is further supported by an epidemic study that found a higher number of rod-shaped bacteria in CeD patients [129].

Prospective studies have investigated the relationship between the gut microbiota composition of children and the development of CeD. Infants who remained healthy showed an increase in bacterial diversity, specifically an increase in the Firmicutes family, which was not exhibited in infants who developed CeD [135].

To understand the biochemical mechanism underlying the effect of intestinal microbiota on CeD, germ-free mice were colonized with bacteria isolated from the small intestine of CeD patients and healthy controls. The results showed a protective effect of *Lactobacillus spp*. by reducing the immunogenicity of gluten-producing proteases. In contrast, *Pseudomonas aeruginosa* was associated with CeD development [136].

2.3.2 Early infections might increase the CeD risk.

It has been suggested that childhood infections may increase the risk of developing CeD later in life. One study found that children who experienced more than 10 episodes of respiratory or intestinal infections

were at higher risk for developing CeD than those who experienced fewer than 4 such events [137]. Additionally, prospective cohort studies on the microbiome in CeD have shown a correlation between a higher incidence of Clostridium difficile infections and CeD. C. difficile produces various toxins that increase intestinal permeability and induce a T-cell response, potentially contributing to the development of CeD [138]. However, the role of *Helicobacter pylori* infections in the pathogenesis of CeD is not well established. While some studies have found a low rate of gastric H. pylori infections in patients with CeD compared to those with normal duodenal mucosa [139], other studies have reported a higher incidence of H. pylori infection in patients with CeD compared to those with nonceliac peptic ulcers [140]. Further research is needed to fully understand the relationship between H. pylori and CeD.

It is important to note that CeD has been associated with not only bacterial infections but also viral infections. However, the role of the virome, which is an essential component of the microbiota, is often overlooked. Recent studies have emphasized the impacts of changes in the enteric virome composition on microbial diversity and functionality, particularly in the context of CeD. Research has suggested that early viral infections, such as those caused by adenovirus, enterovirus, rotavirus, and reovirus, may play a role in the development of CeD. Although most of these studies have been cross-sectional or epidemiological in nature, some have identified adenovirus, which is typically found in the intestine, as a potential trigger for CeD development [141].

Recent research from a birth cohort study has suggested that exposure to enterovirus in early childhood, specifically between the ages of one and two years old, may increase the risk for developing CeD [142].

Other studies have suggested that rotavirus infections may also trigger CeD in genetically susceptible children [143]. Children with CeD have been found to have a higher frequency of rotavirus gastroenteritis, and vaccination

against rotavirus has been shown to prevent the onset of CeD. In fact, one study found that the prevalence of CeD was lower in vaccinated children than in those who received a placebo [144].

Reovirus is another virus that has been suggested to play a role in the pathogenetic mechanisms of CeD development. A recent study showed that while reovirus infections may trigger the development of T-cell immunity to gluten as well as activation of TG2, additional events are necessary for the induction of anti-TG2 antibodies and intestinal atrophy [126].

2.3.3 Effects of intestinal microbiota on the pathophysiology of CeD.

The complex pathophysiology of CeD allows for the potential influence of the intestinal microbiota, which can enhance or attenuate immunopathology through different mechanisms. Broadly speaking, there are three main principles at play: first, the microbiota's proteolytic activity can impact the generation of immunogenic peptides from gluten; second, it can affect the integrity of the gut barrier; and third, it can modulate the immune response. Evidence supporting these mechanisms is described below.

The commensal microbiota comprises various species that are involved in processing gluten. One example is *Lactobacilli*, which release peptidases that break down peptides of gluten and alter their immunogenic potential. A study characterizing the adult intestinal microbiota's role in gluten hydrolysis identified 31 isolated strains that exhibit extracellular proteolytic activity against gluten protein, while 27 strains demonstrated peptidolytic activity toward the 33-mer peptide [145].

The physical barrier of the intestine includes mucous, tight junctions, and epithelial cells, which work together to regulate permeability. Zonulin is one of the proteins that regulates the permeability of tight junctions. Overproduction of zonulin can cause tight junctions to become more permeable, leading to increased intestinal permeability or a "leaky gut"

[146]. In celiac disease, gliadin peptides can bind to proinflammatory cytokine receptors and liberate zonulin, contributing to the disruption of tight junctions and increased intestinal permeability [147]. Interestingly, some bacterial species have been shown to modulate zonulin levels. For example, some bacteria, such as *Escherichia coli* and *Shigella*, can increase zonulin levels, while others, such as *Lactobacillus plantarum*, can decrease them [148,149]. These findings suggest that probiotics and other microbial interventions may offer therapeutic potential in regulating zonulin levels and improving gut barrier function.

The intestinal microbiota plays an important role in modulating the immune response. There are similarities between the activation of the innate and adaptive immune systems by gliadin peptides and alterations in the intestinal microbiota. Pathogenic bacteria can activate the immune system through activation of TLRs (Toll-like receptors) or T cells [150]. Studies have shown that certain strains of bacteria present in the intestinal microbiota of individuals with celiac disease can activate human dendritic cells, leading to an inflammatory phenotype and the development of an immune response [151]. Alterations in the intestinal microbiota can also affect their metabolites, such as by creating an imbalance in butyrate production, which increases the expression of a nonfunctional spliced form of FOXP3 (a transcription factor fundamental for the suppressive function and differentiation of T cells). This epigenetic change can increase the risk of developing autoimmunity [152].

2.3.4 Probiotics in CeD.

Therapeutic modulation of the gut microbiota composition has become a promising strategy for managing CeD. The goal of this strategy is to establish and maintain eubiosis, a balance between beneficial and harmful bacteria.

Various methods can be used to modulate the gut microbiota, including dietary regulation, prebiotics, and probiotics. Probiotics are live organisms

that, when ingested in adequate quantities, provide a health benefit to the host. By modulating the microbial population and increasing the number of beneficial colonizing microbes, probiotics can directly or indirectly influence the CeD forecast. Furthermore, probiotics have shown the ability to hydrolyze immunogenic gluten peptides, which reduces their immunogenicity [153,154]. This ability to modify gluten metabolism and modulate the intestinal immune response has opened new possibilities for treating CeD and its symptoms.

Some studies have demonstrated the potential benefits of probiotics in improving CeD-related symptoms and changing the intestinal microbiota. One study conducted in children showed that the Firmicutes/Bacteroidetes ratio was reduced in CeD patients compared to the control group, but probiotic treatment led to a restoration of the physiological Firmicutes/Bacteroidetes ratio [155]. Various candidates have been used for disease immunomodulation, such as different strains of *Bifidobacterium* or a combination of *Lactobacilli* and *Bifidobacterium. Bifidobacterium longum* and *Bifidobacterium bifidus* have been shown to downregulate T-cell pathways associated with CeD, decrease inflammatory cytokines, reduce CeD symptoms, and alleviate intestinal inflammation, leading to a decrease in IgA in fecal samples compared to the placebo group [156]. Moreover, Bifidobacterium lactis is capable of neutralizing the toxicity of gliadin [157], and various strains of *Lactobacillus* have been proposed to fully break down gluten through their endopeptidase activity and decrease its harmful effects on patients with celiac disease [158]. In contrast, other studies found no differences in intestinal microbiota or symptomatology after administration of a probiotic cocktail that included *Bifidobacterium* and *Lactobacillus* [159]. This suggests that it is plausible that the probiotic species did not survive the physiological environment of the upper gastrointestinal tract [160].

Using a murine model of CeD, different studies have demonstrated positive impacts of probiotics on the digestion of gliadin peptides, the intestinal barrier, and the immune system and mucosa. For example, *Bifidobacterium*

longum reduced proinflammatory cytokine synthesis and attenuated gliadin-induced intestinal damage [161–163]. On the other hand, *Lactobacillus casei* recovered basal TNF-α levels and improved villus blunting [164].

2.3.5 Prebiotics and Postbiotics in CeD.

Prebiotics are fibers that resist gastric acidity and mammalian enzyme hydrolysis and are fermented by intestinal microbiota. They selectively stimulate the growth and activity of potential health-promoting species, particularly *Bifidobacterium* and *Lactobacillus* [165].

On the other hand, postbiotics are functional bioactive compounds produced by fermentation that promote health. They comprise various metabolites, such as short-chain fatty acids (SCFAs). According to current research, postbiotics have direct immunomodulatory effects and clinically relevant benefits. Evidence has shown that they can alleviate symptoms in a range of diseases, such as infant colic, adult atopic dermatitis and different forms of diarrhea [166].

2.4 A gluten-free diet (GFD)

The gluten-free diet (GFD) is the only known effective treatment for those diagnosed with CeD. The GFD requires the complete exclusion of gluten-containing cereals (wheat, barley, rye, oats, spelt, and Kamut) and instead focuses on consuming naturally gluten-free foods such as fruits, vegetables, legumes, unprocessed meat, fish, and so on. However, this dietary change can be quite challenging for some individuals, especially those with low incomes, due to the significantly higher cost of gluten-free substitute foods compared to their gluten-containing counterparts. [167]. Additionally, social pressures, particularly during adolescence, can hinder strict adherence to the GFD. Notably, gluten is the most commonly consumed food protein in Western societies, with an average intake of approximately 10-20 g per person per day [168]. Therefore, excluding this protein from the diet can require a drastic change in eating habits. As a

result, patients with newly diagnosed CeD should be referred to a qualified dietitian who can help them navigate the complexities of the GFD. It is important not only to identify hidden sources of gluten, but also to incorporate healthy gluten-free substitute grains that provide adequate fiber and nutrients [169]. Deficiencies in iron, calcium, zinc, vitamin B12, vitamin D, and folate are common micronutrient inadequacies claimed for newly diagnosed CeD patients [170]. Therefore, it is crucial to ensure that the GFD is nutritionally balanced to avoid any adverse health consequences.

With the introduction of a GFD, patients experience substantial and rapid improvement in symptoms, including those beyond the typical symptoms of diarrhea, steatorrhea, and weight loss [171]. In the case of the intestinal microbiota, cross-sectional studies have shown that CeD patients still exhibit intestinal microbiota alterations that are not entirely normalized even after adhering to a GFD [172]. In contrast, another study found that the numbers of *Enterobacteria* and *Staphylococci* in untreated CeD patients were higher compared to controls, but these differences were nearly restored when CeD patients adhered to a long-term GFD [130]. Another study examined the effects of a GFD on the intestinal microbiota composition of ten healthy subjects and found that a GFD was associated with a reduction in beneficial gut bacteria (specifically *Bifidobacterium* and *Lactobacillus*) and higher proportions of Escherichia coli and *Enterobacteriaceae* [173].

3 REFERENCES

1. Falony, G.; Joossens, M.; Vieira-Silva, S.; Wang, J.; Darzi, Y.; Faust, K.; Kurilshikov, A.; Bonder, M.J.; Valles-Colomer, M.; Vandeputte, D.; et al. Population-Level Analysis of Gut Microbiome Variation. *Science (1979)* **2016**, *352*, 560–564, doi:10.1126/SCIENCE.AAD3503/SUPPL_FILE/TABLE_S9.XLSX.

2. Guarner, F.; Malagelada, J.R. Gut Flora in Health and Disease. *Lancet* **2003**, *361*, 512–519, doi:10.1016/S0140-6736(03)12489-0.

3. Halaas, J.L.; Gajiwala, K.S.; Maffei, M.; Cohen, S.L.; Chait, B.T.; Rabinowitz, D.; Lallone, R.L.; Burley, S.K.; Friedman, J.M. Weight-Reducing Effects of the Plasma Protein Encoded by the Obese Gene. *Science* **1995**, *269*, 543–546, doi:10.1126/SCIENCE.7624777.

4. Leblanc, J.G.; Chain, F.; Martín, R.; Bermúdez-Humarán, L.G.; Courau, S.; Langella, P. Beneficial Effects on Host Energy Metabolism of Short-Chain Fatty Acids and Vitamins Produced by Commensal and Probiotic Bacteria. *Microb Cell Fact* **2017**, *16*, 79, doi:10.1186/s12934-017-0691-z.

5. Furusawa, Y.; Obata, Y.; Fukuda, S.; Endo, T.A.; Nakato, G.; Takahashi, D.; Nakanishi, Y.; Uetake, C.; Kato, K.; Kato, T.; et al. Commensal Microbe-Derived Butyrate Induces the Differentiation of Colonic Regulatory T Cells. *Nature* **2013**, *504*, 446–450, doi:10.1038/nature12721.

6. Avoli, M.; Krnjević, K. The Long and Winding Road to Gamma-Amino-Butyric Acid as Neurotransmitter. *Canadian Journal of Neurological Sciences* **2016**, *43*, 219–226, doi:10.1017/CJN.2015.333.

7. Dalile, B.; Van Oudenhove, L.; Vervliet, B.; Verbeke, K. The Role of Short-Chain Fatty Acids in Microbiota-Gut-Brain Communication. *Nat Rev Gastroenterol Hepatol* **2019**, *16*, 461–478, doi:10.1038/S41575-019-0157-3.

8. Stilling, R.M.; van de Wouw, M.; Clarke, G.; Stanton, C.; Dinan, T.G.; Cryan, J.F. The Neuropharmacology of Butyrate: The Bread and Butter of the Microbiota-Gut-Brain Axis? *Neurochem Int* **2016**, *99*, 110–132, doi:10.1016/J.NEUINT.2016.06.011.

9. Sugawara, T.; Sawada, D.; Ishida, Y.; Aihara, K.; Aoki, Y.; Takehara, I.; Takano, K.; Fujiwara, S. Regulatory Effect of Paraprobiotic Lactobacillus Gasseri CP2305 on Gut Environment and Function. *Microb Ecol Health Dis* **2016**, *27*, doi:10.3402/MEHD.V27.30259.

10. Bernet, M.F.; Brassart, D.; Neeser, J.R.; Servin, A.L. Lactobacillus Acidophilus LA 1 Binds to Cultured Human Intestinal Cell Lines and Inhibits Cell Attachment and Cell Invasion by Enterovirulent Bacteria. *Gut* **1994**, *35*, 483, doi:10.1136/GUT.35.4.483.

11. Juturu, V.; Wu, J.C. Microbial Production of Bacteriocins: Latest Research Development and Applications. *Biotechnol Adv* **2018**, *36*, 2187–2200, doi:10.1016/J.BIOTECHADV.2018.10.007.

12. David, L.A.; Maurice, C.F.; Carmody, R.N.; Gootenberg, D.B.; Button, J.E.; Wolfe, B.E.; Ling, A. v.; Devlin, A.S.; Varma, Y.; Fischbach, M.A.; et al. Diet Rapidly and Reproducibly Alters the Human Gut Microbiome. *Nature* **2014**, *505*, 559–563, doi:10.1038/NATURE12820.

13. Gagliardi, A.; Totino, V.; Cacciotti, F.; Iebba, V.; Neroni, B.; Bonfiglio, G.; Trancassini, M.; Passariello, C.; Pantanella, F.; Schippa, S. Rebuilding the Gut Microbiota Ecosystem. *Int J Environ Res Public Health* **2018**, *15*, doi:10.3390/ijerph15081679.

14. Turpin, W.; Espin-Garcia, O.; Xu, W.; Silverberg, M.S.; Kevans, D.; Smith, M.I.; Guttman, D.S.; Griffiths, A.; Panaccione, R.; Otley, A.; et al. Association of Host Genome with Intestinal Microbial Composition in a Large Healthy Cohort. *Nat Genet* **2016**, *48*, 1413–1417, doi:10.1038/NG.3693.

15. Stewart, C.J.; Ajami, N.J.; O'Brien, J.L.; Hutchinson, D.S.; Smith, D.P.; Wong, M.C.; Ross, M.C.; Lloyd, R.E.; Doddapaneni, H.V.; Metcalf, G.A.; et al. Temporal Development of the Gut Microbiome in Early Childhood from the TEDDY Study. *Nature* **2018**, *562*, 583–588, doi:10.1038/S41586-018-0617-X.

16. Agans, R.; Rigsbee, L.; Kenche, H.; Michail, S.; Khamis, H.J.; Paliy, O. Distal Gut Microbiota of Adolescent Children Is Different from That of Adults. *FEMS Microbiol Ecol* **2011**, *77*, 404–412, doi:10.1111/J.1574-6941.2011.01120.X.

17. Biagi, E.; Rampelli, S.; Turroni, S.; Quercia, S.; Candela, M.; Brigidi, P. The Gut Microbiota of Centenarians: Signatures of Longevity in the Gut Microbiota Profile. *Mech Ageing Dev* **2017**, *165*, 180–184, doi:10.1016/J.MAD.2016.12.013.

18. Kim, S.; Jazwinski, S.M. The Gut Microbiota and Healthy Aging. *Gerontology* **2018**, *64*, 513, doi:10.1159/000490615.

19. Jakobsson, H.E.; Jernberg, C.; Andersson, A.F.; Sjölund-Karlsson, M.; Jansson, J.K.; Engstrand, L. Short-Term Antibiotic Treatment Has Differing Long-Term Impacts on the Human Throat and Gut Microbiome. *PLoS One* **2010**, *5*, doi:10.1371/JOURNAL.PONE.0009836.

20. Pallav, K.; Dowd, S.E.; Villafuerte, J.; Yang, X.; Kabbani, T.; Hansen, J.; Dennis, M.; Leffler, D.A.; Kelly, C.P. Effects of Polysaccharopeptide from Trametes Versicolor and Amoxicillin on the Gut Microbiome of Healthy Volunteers: A Randomized Clinical Trial. *Gut Microbes* **2014**, *5*, 458–467, doi:10.4161/GMIC.29558.

21. Turta, O.; Rautava, S. Antibiotics, Obesity and the Link to Microbes - What Are We Doing to Our Children? *BMC Med* **2016**, *14*, doi:10.1186/S12916-016-0605-7.

22. Mårild, K.; Ludvigsson, J.; Sanz, Y.; Ludvigsson, J.F. Antibiotic Exposure in Pregnancy and Risk of Coeliac Disease in Offspring: A Cohort Study. *BMC Gastroenterol* **2014**, *14*, doi:10.1186/1471-230X-14-75.

23. Suttle, C.A. Viruses in the Sea. *Nature* **2005**, *437*, 356–361, doi:10.1038/NATURE04160.

24. Liang, G.; Bushman, F.D. The Human Virome: Assembly, Composition and Host Interactions. *Nature Reviews Microbiology 2021 19:8* **2021**, *19*, 514–527, doi:10.1038/s41579-021-00536-5.

25. Weinbauer, M.G. Ecology of Prokaryotic Viruses. *FEMS Microbiol Rev* **2004**, *28*, 127–181, doi:10.1016/J.FEMSRE.2003.08.001.

26. Clokie, M.R.J.; Millard, A.D.; Letarov, A. V.; Heaphy, S. Phages in Nature. *Bacteriophage* **2011**, *1*, 31, doi:10.4161/BACT.1.1.14942.

27. Wang, W.; Jovel, J.; Halloran, B.; Wine, E.; Patterson, J.; Ford, G.; O'Keefe, S.; Meng, B.; Song, D.; Zhang, Y.; et al. Metagenomic Analysis of Microbiome in Colon Tissue from Subjects with Inflammatory Bowel Diseases Reveals Interplay of Viruses and Bacteria. *Inflamm Bowel Dis* **2015**, *21*, 1419–1427, doi:10.1097/MIB.0000000000000344.

28. Zhao, G.; Vatanen, T.; Droit, L.; Park, A.; Kostic, A.D.; Poon, T.W.; Vlamakis, H.; Siljander, H.; Härkönen, T.; Hämäläinen, A.M.; et al. Intestinal Virome Changes Precede Autoimmunity in Type I Diabetes-Susceptible Children. *Proc Natl Acad Sci U S A* **2017**, *114*, E6166–E6175, doi:10.1073/PNAS.1706359114.

29. Nakatsu, G.; Zhou, H.; Wu, W.K.K.; Wong, S.H.; Coker, O.O.; Dai, Z.; Li, X.; Szeto, C.H.; Sugimura, N.; Lam, T.Y.T.; et al. Alterations in Enteric Virome Are Associated With Colorectal Cancer and Survival Outcomes. *Gastroenterology* **2018**, *155*, 529-541.e5, doi:10.1053/J.GASTRO.2018.04.018.

30. Han, M.; Yang, P.; Zhong, C.; Ning, K. The Human Gut Virome in Hypertension. *Front Microbiol* **2018**, *9*, doi:10.3389/FMICB.2018.03150/FULL.

31. Minot, S.; Sinha, R.; Chen, J.; Li, H.; Keilbaugh, S.A.; Wu, G.D.; Lewis, J.D.; Bushman, F.D. The Human Gut Virome: Inter-Individual Variation and Dynamic Response to Diet. *Genome Res* **2011**, *21*, 1616–1625, doi:10.1101/GR.122705.111.

32. Gregory, A.C.; Zablocki, O.; Zayed, A.A.; Howell, A.; Bolduc, B.; Sullivan, M.B. The Gut Virome Database Reveals Age-Dependent Patterns of Virome Diversity in the Human Gut. *Cell Host Microbe* **2020**, *28*, 724-740.e8, doi:10.1016/J.CHOM.2020.08.003.

33. Wang, W.L.; Xu, S.Y.; Ren, Z.G.; Tao, L.; Jiang, J.W.; Zheng, S. Sen Application of Metagenomics in the Human Gut Microbiome. *World Journal of Gastroenterology : WJG* **2015**, *21*, 803, doi:10.3748/WJG.V21.I3.803.

34. Rodriguez-Valera, F.; Martin-Cuadrado, A.B.; Rodriguez-Brito, B.; Pašić, L.; Thingstad, T.F.; Rohwer, F.; Mira, A. Explaining Microbial Population Genomics through Phage Predation. *Nat Rev Microbiol* **2009**, *7*, 828–836, doi:10.1038/NRMICRO2235.

35. Lim, E.S.; Zhou, Y.; Zhao, G.; Bauer, I.K.; Droit, L.; Ndao, I.M.; Warner, B.B.; Tarr, P.I.; Wang, D.; Holtz, L.R. Early Life Dynamics of the Human Gut Virome and Bacterial Microbiome in Infants. *Nat Med* **2015**, *21*, 1228–1234, doi:10.1038/nm.3950.

36. Dalmasso, G.; Nguyen, H.T.T.; Yan, Y.; Laroui, H.; Charania, M.A.; Ayyadurai, S.; Sitaraman, S. V.; Merlin, D. Microbiota Modulate Host Gene Expression via MicroRNAs. *PLoS One* **2011**, *6*, doi:10.1371/JOURNAL.PONE.0019293.

37. Liu, S.; da Cunha, A.P.; Rezende, R.M.; Cialic, R.; Wei, Z.; Bry, L.; Comstock, L.E.; Gandhi, R.; Weiner, H.L. The Host Shapes the Gut Microbiota via Fecal MicroRNA. *Cell Host Microbe* **2016**, *19*, 32–43, doi:10.1016/J.CHOM.2015.12.005.

38. Singh, N.; Shirdel, E.A.; Waldron, L.; Zhang, R.H.; Jurisica, I.; Comelli, E.M. The Murine Caecal MicroRNA Signature Depends on the Presence of the Endogenous Microbiota. *Int J Biol Sci* **2012**, *8*, 171–186, doi:10.7150/IJBS.8.171.

39. Kang, S.M.; Choi, J.W.; Lee, Y.; Hong, S.H.; Lee, H.J. Identification of MicroRNA-Size, Small RNAs in Escherichia Coli. *Curr Microbiol* **2013**, *67*, 609–613, doi:10.1007/S00284-013-0411-9.

40. Jopling, C.L.; Yi, M.K.; Lancaster, A.M.; Lemon, S.M.; Sarnow, P. Modulation of Hepatitis C Virus RNA Abundance by a Liver-Specific MicroRNA. *Science* **2005**, *309*, 1577–1581, doi:10.1126/SCIENCE.1113329.

41. Mohr, A.E.; Jäger, R.; Carpenter, K.C.; Kerksick, C.M.; Purpura, M.; Townsend, J.R.; West, N.P.; Black, K.; Gleeson, M.; Pyne, D.B.; et al. The Athletic Gut Microbiota. *J Int Soc Sports Nutr* **2020**, *17*, doi:10.1186/S12970-020-00353-W.

42. 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. Human Gut Microbiome Viewed across Age and Geography. *Nature 2012 486:7402* **2012**, *486*, 222–227, doi:10.1038/nature11053.

43. Tanaka, M.; Nakayama, J. Development of the Gut Microbiota in Infancy and Its Impact on Health in Later Life. *Allergology International* **2017**, *66*, 515–522, doi:10.1016/J.ALIT.2017.07.010.

44. Bäckhed, F.; Roswall, J.; Peng, Y.; Feng, Q.; Jia, H.; Kovatcheva-Datchary, P.; Li, Y.; Xia, Y.; Xie, H.; Zhong, H.; et al. Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. *Cell Host Microbe* **2015**, *17*, 690–703, doi:10.1016/j.chom.2015.04.004.

45. Lauder, A.P.; Roche, A.M.; Sherrill-Mix, S.; Bailey, A.; Laughlin, A.L.; Bittinger, K.; Leite, R.; Elovitz, M.A.; Parry, S.; Bushman, F.D. Comparison of Placenta Samples with Contamination Controls Does Not Provide Evidence for a Distinct Placenta Microbiota. *Microbiome* **2016**, *4*, doi:10.1186/S40168-016-0172-3.

46. Neu, J. The Microbiome during Pregnancy and Early Postnatal Life. *Semin Fetal Neonatal Med* **2016**, *21*, 373–379, doi:10.1016/J.SINY.2016.05.001.

47. Sweeney, E.L.; Kallapur, S.G.; Gisslen, T.; Lambers, D.S.; Chougnet, C.A.; Stephenson, S.A.; Jobe, A.H.; Knox, C.L. Placental Infection With Ureaplasma Species Is Associated With Histologic Chorioamnionitis and Adverse Outcomes in Moderately Preterm and Late-Preterm Infants. *J Infect Dis* **2016**, *213*, 1340–1347, doi:10.1093/INFDIS/JIV587.

48. DiGiulio, D.B.; Romero, R.; Amogan, H.P.; Kusanovic, J.P.; Bik, E.M.; Gotsch, F.; Kim, C.J.; Erez, O.; Edwin, S.; Relman, D.A. Microbial Prevalence, Diversity and Abundance in Amniotic Fluid During Preterm Labor: A Molecular and Culture-Based Investigation. *PLoS One* **2008**, *3*, e3056, doi:10.1371/JOURNAL.PONE.0003056.

49. Hansen, R.; Scott, K.P.; Khan, S.; Martin, J.C.; Berry, S.H.; Stevenson, M.; Okpapi, A.; Munro, M.J.; Hold, G.L. First-Pass Meconium Samples from Healthy Term Vaginally-Delivered Neonates: An Analysis of the Microbiota. *PLoS One* **2015**, *10*, e0133320, doi:10.1371/JOURNAL.PONE.0133320.

50. Nagpal, R.; Tsuji, H.; Takahashi, T.; Kawashima, K.; Nagata, S.; Nomoto, K.; Yamashiro, Y. Sensitive Quantitative Analysis of the Meconium Bacterial Microbiota in Healthy Term Infants Born Vaginally or by Cesarean Section. *Front Microbiol* **2016**, *7*, 1997, doi:10.3389/FMICB.2016.01997/BIBTEX.

51. Hu, J.; Nomura, Y.; Bashir, A.; Fernandez-Hernandez, H.; Itzkowitz, S.; Pei, Z.; Stone, J.; Loudon, H.; Peter, I. Diversified Microbiota of Meconium Is Affected by Maternal Diabetes Status. *PLoS One* **2013**, *8*, e78257, doi:10.1371/JOURNAL.PONE.0078257.

52. Lauder, A.P.; Roche, A.M.; Sherrill-Mix, S.; Bailey, A.; Laughlin, A.L.; Bittinger, K.; Leite, R.; Elovitz, M.A.; Parry, S.; Bushman, F.D. Comparison of Placenta Samples with Contamination Controls Does Not Provide Evidence for a Distinct Placenta Microbiota. *Microbiome* **2016**, *4*, 1–11, doi:10.1186/S40168-016-0172-3/FIGURES/4.

53. Ardissone, A.N.; De La Cruz, D.M.; Davis-Richardson, A.G.; Rechcigl, K.T.; Li, N.; Drew, J.C.; Murgas-Torrazza, R.; Sharma, R.; Hudak, M.L.; Triplett, E.W.;

et al. Meconium Microbiome Analysis Identifies Bacteria Correlated with Premature Birth. *PLoS One* **2014**, *9*, doi:10.1371/JOURNAL.PONE.0090784.

54. Liang, G.; Zhao, C.; Zhang, H.; Mattei, L.; Sherrill-Mix, S.; Bittinger, K.; Kessler, L.R.; Wu, G.D.; Baldassano, R.N.; DeRusso, P.; et al. The Stepwise Assembly of the Neonatal Virome Is Modulated by Breastfeeding. *Nature* **2020**, *581*, 470–474, doi:10.1038/s41586-020-2192-1.

55. Lofgren, E.; Fefferman, N.H.; Naumov, Y.N.; Gorski, J.; Naumova, E.N. Influenza Seasonality: Underlying Causes and Modeling Theories. *J Virol* **2007**, *81*, 5429–5436, doi:10.1128/JVI.01680-06.

56. Pannaraj, P.S.; Ly, M.; Cerini, C.; Saavedra, M.; Aldrovandi, G.M.; Saboory, A.A.; Johnson, K.M.; Pride, D.T. Shared and Distinct Features of Human Milk and Infant Stool Viromes. *Front Microbiol* **2018**, *9*, 1162, doi:10.3389/fmicb.2018.01162.

57. Dominguez-Bello, M.G.; Costello, E.K.; Contreras, M.; Magris, M.; Hidalgo, G.; Fierer, N.; Knight, R. 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*, 11971–11975, doi:10.1073/PNAS.1002601107.

58. Arboleya, S.; Binetti, A.; Salazar, N.; Fernández, N.; Solís, G.; Hernández-Barranco, A.; Margolles, A.; de los Reyes-Gavilán, C.G.; Gueimonde, M. Establishment and Development of Intestinal Microbiota in Preterm Neonates. *FEMS Microbiol Ecol* **2012**, *79*, 763–772, doi:10.1111/J.1574-6941.2011.01261.X.

59. Maynard, C.L.; Elson, C.O.; Hatton, R.D.; Weaver, C.T. Reciprocal Interactions of the Intestinal Microbiota and Immune System. *Nature* **2012**, *489*, 231–241, doi:10.1038/NATURE11551.

60. le Hurou-Luron, I.; Blat, S.; Boudry, G. Breast- v. Formula-Feeding: Impacts on the Digestive Tract and Immediate and Long-Term Health Effects. *Nutr Res Rev* **2010**, *23*, 23–36, doi:10.1017/S0954422410000065.

61. Marcobal, A.; Barboza, M.; Sonnenburg, E.D.; Pudlo, N.; Martens, E.C.; Desai, P.; Lebrilla, C.B.; Weimer, B.C.; Mills, D.A.; German, J.B.; et al. Bacteroides in the Infant Gut Consume Milk Oligosaccharides via Mucus-Utilization Pathways. *Cell Host Microbe* **2011**, *10*, 507–514, doi:10.1016/J.CHOM.2011.10.007.

62. Chong, C.Y.L.; Vatanen, T.; Alexander, T.; Bloomfield, F.H.; O'Sullivan, J.M.; the DIAMOND study group Factors Associated With the Microbiome in Moderate-Late Preterm Babies: A Cohort Study From the DIAMOND Randomized Controlled Trial. *Front Cell Infect Microbiol* **2021**, *11*, doi:10.3389/FCIMB.2021.595323.

63. Navab-Moghadam, F.; Sedighi, M.; Khamseh, M.E.; Alaei-Shahmiri, F.; Talebi, M.; Razavi, S.; Amirmozafari, N. The Association of Type II Diabetes with Gut Microbiota Composition. *Microb Pathog* **2017**, *110*, 630–636, doi:10.1016/j.micpath.2017.07.034.

64. Ley, D.; Desseyn, J.L.; Mischke, M.; Knol, J.; Turck, D.; Gottrand, F. Early-Life Origin of Intestinal Inflammatory Disorders. *Nutr Rev* **2017**, *75*, 175–187, doi:10.1093/nutrit/nuw061.

65. Sherman, P.M.; Cabana, M.; Gibson, G.R.; Koletzko, B. v.; Neu, J.; Veereman-Wauters, G.; Ziegler, E.E.; Walker, W.A. Potential Roles and Clinical Utility of Prebiotics in Newborns, Infants, and Children: Proceedings from a Global Prebiotic Summit Meeting, New York City, June 27-28, 2008. *J Pediatr* **2009**, *155*, doi:10.1016/J.JPEDS.2009.08.022.

66. Flint, H.J.; Scott, K.P.; Duncan, S.H.; Louis, P.; Forano, E. Microbial Degradation of Complex Carbohydrates in the Gut. *Gut Microbes* **2012**, *3*, doi:10.4161/GMIC.19897.

67. Koenig, J.E.; Spor, A.; Scalfone, N.; Fricker, A.D.; Stombaugh, J.; Knight, R.; Angenent, L.T.; Ley, R.E. Succession of Microbial Consortia in the Developing Infant Gut Microbiome. *Proc Natl Acad Sci U S A* **2011**, *108 Suppl 1*, 4578–4585, doi:10.1073/PNAS.1000081107.

68. Palmer, C.; Bik, E.M.; DiGiulio, D.B.; Relman, D.A.; Brown, P.O. Development of the Human Infant Intestinal Microbiota. *PLoS Biol* **2007**, *5*, 1556–1573, doi:10.1371/journal.pbio.0050177.

69. Fouhy, F.; Ross, R.P.; Fitzgerald, G.; Stanton, C.; Cotter, P.D. Composition of the Early Intestinal Microbiota. *https://doi.org/10.4161/gmic.20169* **2012**, *3*, 203–220, doi:10.4161/GMIC.20169.

70. Arrieta, M.C.; Stiemsma, L.T.; Amenyogbe, N.; Brown, E.; Finlay, B. The Intestinal Microbiome in Early Life: Health and Disease. *Front Immunol* **2014**, *5*, 427, doi:10.3389/FIMMU.2014.00427/BIBTEX.

71. Clark, R.H.; Bloom, B.T.; Spitzer, A.R.; Gerstmann, D.R. Reported Medication Use in the Neonatal Intensive Care Unit: Data from a Large National Data Set. *Pediatrics* **2006**, *117*, 1979–1987, doi:10.1542/PEDS.2005-1707.

72. Ready, D.; Lancaster, H.; Qureshi, F.; Bedi, R.; Mullany, P.; Wilson, M. Effect of Amoxicillin Use on Oral Microbiota in Young Children. *Antimicrob Agents Chemother* **2004**, *48*, 2883, doi:10.1128/AAC.48.8.2883-2887.2004.

73. Arboleya, S.; Sánchez, B.; Milani, C.; Duranti, S.; Solís, G.; Fernández, N.; de Los Reyes-Gavilán, C.G.; Ventura, M.; Margolles, A.; Gueimonde, M. Intestinal Microbiota Development in Preterm Neonates and Effect of Perinatal Antibiotics. *J Pediatr* **2015**, *166*, 538–544, doi:10.1016/J.JPEDS.2014.09.041.

74. Langdon, A.; Crook, N.; Dantas, G. The Effects of Antibiotics on the Microbiome throughout Development and Alternative Approaches for Therapeutic Modulation. *Genome Med* **2016**, *8*, doi:10.1186/S13073-016-0294-Z.

75. Tringe, S.G.; Rubin, E.M. Metagenomics: DNA Sequencing of Environmental Samples. *Nat Rev Genet* **2005**, *6*, 805–814, doi:10.1038/NRG1709.

76. Petti, C.A. Detection and Identification of Microorganisms by Gene Amplification and Sequencing. *Clin Infect Dis* **2007**, *44*, 1108–1114, doi:10.1086/512818.

77. Quince, C.; Walker, A.W.; Simpson, J.T.; Loman, N.J.; Segata, N. Shotgun Metagenomics, from Sampling to Analysis. *Nat Biotechnol* **2017**, *35*, 833–844, doi:10.1038/nbt.3935.

78. Raes, J.; Korbel, J.O.; Lercher, M.J.; von Mering, C.; Bork, P. Prediction of Effective Genome Size in Metagenomic Samples. *Genome Biol* **2007**, *8*, doi:10.1186/GB-2007-8-1-R10.

79. Bashiardes, S.; Zilberman-Schapira, G.; Elinav, E. Use of Metatranscriptomics in Microbiome Research. *Bioinform Biol Insights* **2016**, *10*, 19–25, doi:10.4137/BBI.S34610.

80. Gosalbes, M.J.; Durbán, A.; Pignatelli, M.; Abellan, J.J.; Jiménez-Hernández, N.; Pérez-Cobas, A.E.; Latorre, A.; Moya, A. Metatranscriptomic Approach to Analyze the Functional Human Gut Microbiota. *PLoS One* **2011**, *6*, 17447, doi:10.1371/JOURNAL.PONE.0017447.

81. Hafner, M.; Landgraf, P.; Ludwig, J.; Rice, A.; Ojo, T.; Lin, C.; Holoch, D.; Lim, C.; Tuschl, T. Identification of MicroRNAs and Other Small Regulatory RNAs Using CDNA Library Sequencing. *Methods* **2008**, *44*, 3–12, doi:10.1016/J.YMETH.2007.09.009.

82. Raabe, C.A.; Tang, T.H.; Brosius, J.; Rozhdestvensky, T.S. Biases in Small RNA Deep Sequencing Data. *Nucleic Acids Res* **2014**, *42*, 1414, doi:10.1093/NAR/GKT1021.

83. Chen, L.; Heikkinen, L.; Wang, C.L.; Yang, Y.; Knott, K.E.; Wong, G. MiRToolsGallery: A Tag-Based and Rankable MicroRNA Bioinformatics Resources Database Portal. *Database* **2018**, *2018*, 4, doi:10.1093/DATABASE/BAY004.

84. Tang, J. Microbial Metabolomics. *Curr Genomics* **2011**, *12*, 391–403, doi:10.2174/138920211797248619.

85. Vernocchi, P.; Vannini, L.; Gottardi, D.; del Chierico, F.; Serrazanetti, D.I.; Ndagijimana, M.; Guerzoni, M.E. Integration of Datasets from Different Analytical Techniques to Assess the Impact of Nutrition on Human

Metabolome. *Front Cell Infect Microbiol* **2012**, *2*, 156, doi:10.3389/FCIMB.2012.00156.

86. Lamichhane, S.; Sen, P.; Dickens, A.M.; Orešič, M.; Bertram, H.C. Gut Metabolome Meets Microbiome: A Methodological Perspective to Understand the Relationship between Host and Microbe. *Methods* **2018**, *149*, 3–12, doi:10.1016/J.YMETH.2018.04.029.

87. Adams, M.Hancock. Bacteriophages. *Interscience publishers* **1959**.

88. Shkoporov, A.N.; Hill, C. Bacteriophages of the Human Gut: The "Known Unknown" of the Microbiome. *Cell Host Microbe* **2019**, *25*, 195–209, doi:10.1016/j.chom.2019.01.017.

89. Hall, R.J.; Wang, J.; Todd, A.K.; Bissielo, A.B.; Yen, S.; Strydom, H.; Moore, N.E.; Ren, X.; Huang, Q.S.; Carter, P.E.; et al. Evaluation of Rapid and Simple Techniques for the Enrichment of Viruses Prior to Metagenomic Virus Discovery. *J Virol Methods* **2014**, *195*, 194–204, doi:10.1016/j.jviromet.2013.08.035.

90. Sutton, T.D.S.; Clooney, A.G.; Ryan, F.J.; Ross, R.P.; Hill, C. Choice of Assembly Software Has a Critical Impact on Virome Characterisation. *Microbiome* **2019**, *7*, 1–15, doi:10.1186/S40168-019-0626-5/FIGURES/4.

91. Krishnamurthy, S.R.; Wang, D. Origins and Challenges of Viral Dark Matter. *Virus Res* **2017**, *239*, 136–142, doi:10.1016/J.VIRUSRES.2017.02.002.

92. Gibbs, A.J. Viral Taxonomy Needs a Spring Clean; Its Exploration Era Is Over. *Virol J* **2013**, *10*, doi:10.1186/1743-422X-10-254.

93. Chen, M.; Song, P.; Zou, D.; Hu, X.; Zhao, S.; Gao, S.; Ling, F. Comparison of Multiple Displacement Amplification (MDA) and Multiple Annealing and Looping-Based Amplification Cycles (MALBAC) in Single-Cell Sequencing. *PLoS One* **2014**, *9*, doi:10.1371/JOURNAL.PONE.0114520.

94. Gregory, A.C.; Zablocki, O.; Zayed, A.A.; Howell, A.; Bolduc, B.; Sullivan, M.B. The Gut Virome Database Reveals Age-Dependent Patterns of Virome Diversity in the Human Gut. *Cell Host Microbe* **2020**, *28*, 724-740.e8, doi:10.1016/J.CHOM.2020.08.003.

95. Yutin, N.; Makarova, K.S.; Gussow, A.B.; Krupovic, M.; Segall, A.; Edwards, R.A.; Koonin, E. v. Discovery of an Expansive Bacteriophage Family That Includes the Most Abundant Viruses from the Human Gut. *Nat Microbiol* **2018**, *3*, 38–46, doi:10.1038/S41564-017-0053-Y.

96. di Sabatino, A.; Corazza, G.R. Coeliac Disease. *Lancet* **2009**, *373*, 1480–1493, doi:10.1016/S0140-6736(09)60254-3.

97. Ludvigsson, J.F.; Leffler, D.A.; Bai, J.C.; Biagi, F.; Fasano, A.; Green, P.H.R.; Hadjivassiliou, M.; Kaukinen, K.; Kelly, C.P.; Leonard, J.N.; et al. The Oslo

Definitions for Coeliac Disease and Related Terms. *Gut* **2013**, *62*, 43–52, doi:10.1136/gutjnl-2011-301346.

98. Sood, A.; Midha, V.; Makharia, G.; Thelma, B.K.; Halli, S.S.; Mehta, V.; Mahajan, R.; Narang, V.; Sood, K.; Kaur, K. A Simple Phenotypic Classification for Celiac Disease. *Intest Res* **2018**, *16*, 288, doi:10.5217/IR.2018.16.2.288.

99. Leffler, D.A.; Green, P.H.R.; Fasano, A. Extraintestinal Manifestations of Coeliac Disease. *Nat Rev Gastroenterol Hepatol* **2015**, *12*, 561–571, doi:10.1038/NRGASTRO.2015.131.

100. Roberts, S.E.; Morrison-Rees, S.; Thapar, N.; Benninga, M.A.; Borrelli, O.; Broekaert, I.; Dolinsek, J.; Martin-de-Carpi, J.; Mas, E.; Miele, E.; et al. Systematic Review and Meta-Analysis: The Incidence and Prevalence of Paediatric Coeliac Disease across Europe. *Aliment Pharmacol Ther* **2021**, *54*, 109–128, doi:10.1111/APT.16337.

101. Ludvigsson, J.F.; Murray, J.A. Epidemiology of Celiac Disease. *Gastroenterol Clin North Am* **2019**, *48*, 1–18, doi:10.1016/J.GTC.2018.09.004.

102. Lebwohl, B.; Rubio-Tapia, A. Epidemiology, Presentation, and Diagnosis of Celiac Disease. *Gastroenterology* **2021**, *160*, 63–75, doi:10.1053/J.GASTRO.2020.06.098.

103. Catassi, C.; Rätsch, I.M.; Gandolfi, L.; Pratesi, R.; Fabiani, E.; el Asmar, R.; Frijia, M.; Bearzi, I.; Vizzoni, L. Why Is Coeliac Disease Endemic in the People of the Sahara? *Lancet* **1999**, *354*, 647–648, doi:10.1016/S0140-6736(99)02609-4.

104. Theethira, T.G.; Dennis, M. Celiac Disease and the Gluten-Free Diet: Consequences and Recommendations for Improvement. *Dig Dis* **2015**, *33*, 175–182, doi:10.1159/000369504.

105. Gluten-Free Products Market Size Report, 2022-2030 Available online: https://www.grandviewresearch.com/industry-analysis/gluten-free-products-market (accessed on 29 March 2023).

106. Ehren, J.; Móron, B.; Martin, E.; Bethune, M.T.; Gray, G.M.; Khosla, C. A Food-Grade Enzyme Preparation with Modest Gluten Detoxification Properties. *PLoS One* **2009**, *4*, e6313, doi:10.1371/JOURNAL.PONE.0006313.

107. Rauhavirta, T.; Oittinen, M.; Kivistö, R.; Männistö, P.T.; Garcia-Horsman, J.A.; Wang, Z.; Griffin, M.; Mäki, M.; Kaukinen, K.; Lindfors, K. Are Transglutaminase 2 Inhibitors Able to Reduce Gliadin-Induced Toxicity Related to Celiac Disease? A Proof-of-Concept Study. *Journal of Clinical Immunology 2012 33:1* **2012**, *33*, 134–142, doi:10.1007/S10875-012-9745-5.

108. Goel, G.; King, T.; Daveson, A.J.; Andrews, J.M.; Krishnarajah, J.; Krause, R.; Brown, G.J.E.; Fogel, R.; Barish, C.F.; Epstein, R.; et al. Epitope-Specific

Immunotherapy Targeting CD4-Positive T Cells in Coeliac Disease: Two Randomised, Double-Blind, Placebo-Controlled Phase 1 Studies. *Lancet Gastroenterol Hepatol* **2017**, *2*, 479–493, doi:10.1016/S2468-1253(17)30110-3.

109. Verbeek, W.H.M.; van de Water, J.M.W.; Al-Toma, A.; Oudejans, J.J.; Mulder, C.J.J.; Coupé, V.M.H. Incidence of Enteropathy--Associated T-Cell Lymphoma: A Nation-Wide Study of a Population-Based Registry in The Netherlands. *Scand J Gastroenterol* **2008**, *43*, 1322–1328, doi:10.1080/00365520802240222.

110. Brousse, N.; Meijer, J.W.R. Malignant Complications of Coeliac Disease. *Best Pract Res Clin Gastroenterol* **2005**, *19*, 401–412, doi:10.1016/J.BPG.2005.02.002.

111. Ludvigsson, J.F.; Montgomery, S.M.; Ekbom, A.; Brandt, L.; Granath, F. Small-Intestinal Histopathology and Mortality Risk in Celiac Disease. *JAMA* **2009**, *302*, 1171–1178, doi:10.1001/JAMA.2009.1320.

112. Shewry, P. What Is Gluten—Why Is It Special? *Front Nutr* **2019**, *6*, 101, doi:10.3389/FNUT.2019.00101/BIBTEX.

113. Kagnoff, M.F. Celiac Disease: Pathogenesis of a Model Immunogenetic Disease. *J Clin Invest* **2007**, *117*, 41–49, doi:10.1172/JCI30253.

114. Schuppan, D.; Junker, Y.; Barisani, D. Celiac Disease: From Pathogenesis to Novel Therapies. *Gastroenterology* **2009**, *137*, 1912–1933, doi:10.1053/J.GASTRO.2009.09.008.

115. Liu, E.; Dong, F.; Barón, A.E.; Taki, I.; Norris, J.M.; Frohnert, B.I.; Hoffenberg, E.J.; Rewers, M. High Incidence of Celiac Disease in a Long-Term Study of Adolescents With Susceptibility Genotypes. *Gastroenterology* **2017**, *152*, 1329-1336.e1, doi:10.1053/J.GASTRO.2017.02.002.

116. Gutierrez-Achury, J.; Zhernakova, A.; Pulit, S.L.; Trynka, G.; Hunt, K.A.; Romanos, J.; Raychaudhuri, S.; van Heel, D.A.; Wijmenga, C.; de Bakker, P.I.W. Fine Mapping in the MHC Region Accounts for 18% Additional Genetic Risk for Celiac Disease. *Nat Genet* **2015**, *47*, 577–578, doi:10.1038/NG.3268.

117. Withoff, S.; Li, Y.; Jonkers, I.; Wijmenga, C. Understanding Celiac Disease by Genomics. *Trends in Genetics* **2016**, *32*, 295–308, doi:10.1016/J.TIG.2016.02.003.

118. Kahaly, G.J.; Frommer, L.; Schuppan, D. Celiac Disease and Endocrine Autoimmunity – the Genetic Link. *Autoimmun Rev* **2018**, *17*, 1169–1175, doi:10.1016/J.AUTREV.2018.05.013.

119. van Heel, D.A.; Franke, L.; Hunt, K.A.; Gwilliam, R.; Zhernakova, A.; Inouye, M.; Wapenaar, M.C.; Barnardo, M.C.N.M.; Bethel, G.; Holmes, G.K.T.; et al. A Genome-Wide Association Study for Celiac Disease Identifies Risk Variants

in the Region Harboring IL2 and IL21. *Nat Genet* **2007**, *39*, 827, doi:10.1038/NG2058.

120. Bragde, H.; Jansson, U.; Fredrikson, M.; Grodzinsky, E.; Söderman, J. Celiac Disease Biomarkers Identified by Transcriptome Analysis of Small Intestinal Biopsies. *Cellular and Molecular Life Sciences* **2018**, *75*, 4385–4401, doi:10.1007/S00018-018-2898-5/FIGURES/4.

121. Cielo, D.; Galatola, M.; Fernandez-Jimenez, N.; de Leo, L.; Garcia-Etxebarria, K.; Loganes, C.; Tommasini, A.; Not, T.; Auricchio, R.; Greco, L.; et al. Combined Analysis of Methylation and Gene Expression Profiles in Separate Compartments of Small Bowel Mucosa Identified Celiac Disease Patients' Signatures. *Scientific Reports 2019 9:1* **2019**, *9*, 1–12, doi:10.1038/s41598-019-46468-2.

122. Ivarsson, A.; Persson, L.; Nyström, L.; Ascher, H.; Cavell, B.; Danielsson, L.; Dannaeus, A.; Lindberg, T.; Lindquist, B.; Stenhammar, L.; et al. Epidemic of Coeliac Disease in Swedish Children. *Acta Paediatr* **2000**, *89*, 165–171, doi:10.1080/080352500750028771.

123. Szajewska, H.; Shamir, R.; Mearin, L.; Ribes-Koninckx, C.; Catassi, C.; Domellof, M.; Fewtrell, M.S.; Husby, S.; Papadopoulou, A.; Vandenplas, Y.; et al. Gluten Introduction and the Risk of Coeliac Disease: A Position Paper by the European Society for Pediatric Gastroenterology, Hepatology, and Nutrition. *J Pediatr Gastroenterol Nutr* **2016**, *62*, 507–513, doi:10.1097/MPG.0000000000001105.

124. Fallani, M.; Young, D.; Scott, J.; Norin, E.; Amarri, S.; Adam, R.; Aguilera, M.; Khanna, S.; Gil, A.; Edwards, C.A.; et al. Intestinal Microbiota of 6-Week-Old Infants across Europe: Geographic Influence beyond Delivery Mode, Breast-Feeding, and Antibiotics. *J Pediatr Gastroenterol Nutr* **2010**, *51*, 77–84, doi:10.1097/MPG.0B013E3181D1B11E.

125. Stene, L.C.; Honeyman, M.C.; Hoffenberg, E.J.; Haas, J.E.; Sokol, R.J.; Emery, L.; Taki, I.; Norris, J.M.; Erlich, H.A.; Eisenbarth, G.S.; et al. Rotavirus Infection Frequency and Risk of Celiac Disease Autoimmunity in Early Childhood: A Longitudinal Study. *Am J Gastroenterol* **2006**, *101*, 2333–2340, doi:10.1111/J.1572-0241.2006.00741.X.

126. Bouziat, R.; Hinterleitner, R.; Brown, J.J.; Stencel-Baerenwald, J.E.; Ikizler, M.; Mayassi, T.; Meisel, M.; Kim, S.M.; Discepolo, V.; Pruijssers, A.J.; et al. Reovirus Infection Triggers Inflammatory Responses to Dietary Antigens and Development of Celiac Disease. *Science* **2017**, *356*, 44, doi:10.1126/SCIENCE.AAH5298.

127. Mårild, K.; Ye, W.; Lebwohl, B.; Green, P.H.R.; Blaser, M.J.; Card, T.; Ludvigsson, J.F. Antibiotic Exposure and the Development of Coeliac

Disease: A Nationwide Case-Control Study. *BMC Gastroenterol* **2013**, *13*, doi:10.1186/1471-230X-13-109.

128. Adinolfi, L.E.; Mangoni, E.D.; Andreana, A. Interferon and Ribavirin Treatment for Chronic Hepatitis C May Activate Celiac Disease. *Am J Gastroenterol* **2001**, *96*, 607–608, doi:10.1111/J.1572-0241.2001.03574.X.

129. Ou, G.; Hedberg, M.; Hörstedt, P.; Baranov, V.; Forsberg, G.; Drobni, M.; Sandström, O.; Wai, S.N.; Johansson, I.; Hammarström, M.L.; et al. Proximal Small Intestinal Microbiota and Identification of Rod-Shaped Bacteria Associated with Childhood Celiac Disease. *Am J Gastroenterol* **2009**, *104*, 3058–3067, doi:10.1038/AJG.2009.524.

130. Collado, M.C.; Donat, E.; Ribes-Koninckx, C.; Calabuig, M.; Sanz, Y. Specific Duodenal and Faecal Bacterial Groups Associated with Paediatric Coeliac Disease. *J Clin Pathol* **2009**, *62*, 264–269, doi:10.1136/JCP.2008.061366.

131. Sánchez, E.; Donat, E.; Ribes-Koninckx, C.; Calabuig, M.; Sanz, Y. Intestinal Bacteroides Species Associated with Coeliac Disease. *J Clin Pathol* **2010**, *63*, 1105–1111, doi:10.1136/JCP.2010.076950.

132. Nistal, E.; Caminero, A.; Herrán, A.R.; Arias, L.; Vivas, S.; Ruiz De Morales, J.M.; Calleja, S.; Sáenz De Miera, L.E.; Arroyo, P.; Casqueiro, J. Differences of Small Intestinal Bacteria Populations in Adults and Children with/without Celiac Disease: Effect of Age, Gluten Diet, and Disease. *Inflamm Bowel Dis* **2012**, *18*, 649–656, doi:10.1002/IBD.21830.

133. Bodkhe, R.; Shetty, S.A.; Dhotre, D.P.; Verma, A.K.; Bhatia, K.; Mishra, A.; Kaur, G.; Pande, P.; Bangarusamy, D.K.; Santosh, B.P.; et al. Comparison of Small Gut and Whole Gut Microbiota of First-Degree Relatives With Adult Celiac Disease Patients and Controls. *Front Microbiol* **2019**, *10*, 137–140, doi:10.3389/FMICB.2019.00164.

134. Sanz, Y.; Sánchez, E.; Marzotto, M.; Calabuig, M.; Torriani, S.; Dellaglio, F. Differences in Faecal Bacterial Communities in Coeliac and Healthy Children as Detected by PCR and Denaturing Gradient Gel Electrophoresis. *FEMS Immunol Med Microbiol* **2007**, *51*, 562–568, doi:10.1111/J.1574-695X.2007.00337.X.

135. Olivares, M.; Walker, A.W.; Capilla, A.; Benítez-Páez, A.; Palau, F.; Parkhill, J.; Castillejo, G.; Sanz, Y. Gut Microbiota Trajectory in Early Life May Predict Development of Celiac Disease. *Microbiome* **2018**, *6*, 1–11, doi:10.1186/s40168-018-0415-6.

136. Caminero, A.; Galipeau, H.J.; McCarville, J.L.; Johnston, C.W.; Bernier, S.P.; Russell, A.K.; Jury, J.; Herran, A.R.; Casqueiro, J.; Tye-Din, J.A.; et al. Duodenal Bacteria From Patients With Celiac Disease and Healthy Subjects Distinctly Affect Gluten Breakdown and Immunogenicity. *Gastroenterology* **2016**, *151*,

670–683, doi:10.1053/J.GASTRO.2016.06.041/ATTACHMENT/75C9F16D-7A02-411B-9391-A0095C437CC2/MMC1.PDF.

137. Mårild, K.; Kahrs, C.R.; Tapia, G.; Stene, L.C.; Størdal, K. Infections and Risk of Celiac Disease in Childhood: A Prospective Nationwide Cohort Study. *Am J Gastroenterol* **2015**, *110*, 1475–1484, doi:10.1038/AJG.2015.287.

138. Lebwohl, B.; Nobel, Y.R.; Green, P.H.R.; Blaser, M.J.; Ludvigsson, J.F. Risk of Clostridium Difficile Infection in Patients With Celiac Disease: A Population-Based Study. *Am J Gastroenterol* **2017**, *112*, 1878–1884, doi:10.1038/AJG.2017.400.

139. Lebwohl, B.; Blaser, M.J.; Ludvigsson, J.F.; Green, P.H.R.; Rundle, A.; Sonnenberg, A.; Genta, R.M. Decreased Risk of Celiac Disease in Patients With Helicobacter Pylori Colonization. *Am J Epidemiol* **2013**, *178*, 1721, doi:10.1093/AJE/KWT234.

140. Tumgor, G.; Agin, M.; Doran, F.; Cetiner, S. Frequency of Celiac Disease in Children with Peptic Ulcers. *Dig Dis Sci* **2018**, *63*, 2681–2686, doi:10.1007/S10620-018-5174-5.

141. Lähdeaho, M.L.; Lehtinen, M.; Rissa, H.R.; Hyöty, H.; Reunala, T.; Mäki, M. Antipeptide Antibodies to Adenovirus E1b Protein Indicate Enhanced Risk of Celiac Disease and Dermatitis Herpetiformis. *Int Arch Allergy Immunol* **1993**, *101*, 272–276, doi:10.1159/000236457.

142. Oikarinen, M.; Puustinen, L.; Lehtonen, J.; Hakola, L.; Simell, S.; Toppari, J.; Ilonen, J.; Veijola, R.; Virtanen, S.M.; Knip, M.; et al. Enterovirus Infections Are Associated With the Development of Celiac Disease in a Birth Cohort Study. *Front Immunol* **2020**, *11*, doi:10.3389/FIMMU.2020.604529.

143. Stene, L.C.; Honeyman, M.C.; Hoffenberg, E.J.; Haas, J.E.; Sokol, R.J.; Emery, L.; Taki, I.; Norris, J.M.; Erlich, H.A.; Eisenbarth, G.S.; et al. Rotavirus Infection Frequency and Risk of Celiac Disease Autoimmunity in Early Childhood: A Longitudinal Study. *Am J Gastroenterol* **2006**, *101*, 2333–2340, doi:10.1111/J.1572-0241.2006.00741.X.

144. Hemming-Harlo, M.; Lähdeaho, M.L.; Mäki, M.; Vesikari, T. Rotavirus Vaccination Does Not Increase Type 1 Diabetes and May Decrease Celiac Disease in Children and Adolescents. *Pediatr Infect Dis J* **2019**, *38*, 539–541, doi:10.1097/INF.0000000000002281.

145. Herrán, A.R.; Pérez-Andrés, J.; Caminero, A.; Nistal, E.; Vivas, S.; Ruiz de Morales, J.M.; Casqueiro, J. Gluten-Degrading Bacteria Are Present in the Human Small Intestine of Healthy Volunteers and Celiac Patients. *Res Microbiol* **2017**, *168*, 673–684, doi:10.1016/J.RESMIC.2017.04.008.

146. Serek, P.; Oleksy-Wawrzyniak, M. The Effect of Bacterial Infections, Probiotics and Zonulin on Intestinal Barrier Integrity. *Int J Mol Sci* **2021**, *22*, 11359, doi:10.3390/IJMS222111359.

147. Fasano, A. Zonulin, Regulation of Tight Junctions, and Autoimmune Diseases. *Ann N Y Acad Sci* **2012**, *1258*, 25, doi:10.1111/J.1749-6632.2012.06538.X.

148. Heyman, M.; Abed, J.; Lebreton, C.; Cerf-Bensussan, N. Intestinal Permeability in Coeliac Disease: Insight into Mechanisms and Relevance to Pathogenesis. *Gut* **2012**, *61*, 1355–1364, doi:10.1136/GUTJNL-2011-300327.

149. Liu, Y.; Alookaran, J.J.; Rhoads, J.M. Probiotics in Autoimmune and Inflammatory Disorders. *Nutrients* **2018**, *10*, doi:10.3390/NU10101537.

150. Sjöberg, V.; Sandström, O.; Hedberg, M.; Hammarström, S.; Hernell, O.; Hammarström, M.L. Intestinal T-Cell Responses in Celiac Disease – Impact of Celiac Disease Associated Bacteria. *PLoS One* **2013**, *8*, e53414, doi:10.1371/JOURNAL.PONE.0053414.

151. D'Argenio, V.; Casaburi, G.; Precone, V.; Pagliuca, C.; Colicchio, R.; Sarnataro, D.; Discepolo, V.; Kim, S.M.; Russo, I.; del Vecchio Blanco, G.; et al. Metagenomics Reveals Dysbiosis and a Potentially Pathogenic N. Flavescens Strain in Duodenum of Adult Celiac Patients. *Am J Gastroenterol* **2016**, *111*, 879–890, doi:10.1038/AJG.2016.95.

152. Serena, G.; Yan, S.; Camhi, S.; Patel, S.; Lima, R.S.; Sapone, A.; Leonard, M.M.; Mukherjee, R.; Nath, B.J.; Lammers, K.M.; et al. Proinflammatory Cytokine Interferon-γ and Microbiome-derived Metabolites Dictate Epigenetic Switch between Forkhead Box Protein 3 Isoforms in Coeliac Disease. *Clin Exp Immunol* **2017**, *187*, 490, doi:10.1111/CEI.12911.

153. Angelis, M. de; Rizzello, C.G.; Fasano, A.; Clemente, M.G.; Simone, C. de; Silano, M.; Vincenzi, M. de; Losito, I.; Gobbetti, M. VSL#3 Probiotic Preparation Has the Capacity to Hydrolyze Gliadin Polypeptides Responsible for Celiac Sprue. *Biochim Biophys Acta* **2006**, *1762*, 80–93, doi:10.1016/J.BBADIS.2005.09.008.

154. Herrán, A.R.; Pérez-Andrés, J.; Caminero, A.; Nistal, E.; Vivas, S.; Ruiz de Morales, J.M.; Casqueiro, J. Gluten-Degrading Bacteria Are Present in the Human Small Intestine of Healthy Volunteers and Celiac Patients. *Res Microbiol* **2017**, *168*, 673–684, doi:10.1016/J.RESMIC.2017.04.008.

155. Quagliariello, A.; Aloisio, I.; Bozzi cionci, N.; Luiselli, D.; D'Auria, G.; Martinez-Priego, L.; Pérez-Villarroya, D.; Langerholc, T.; Primec, M.; Mičetić-Turk, D.; et al. Effect of Bifidobacterium Breve on the Intestinal Microbiota of Coeliac Children on a Gluten Free Diet: A Pilot Study. *Nutrients* **2016**, *8*, doi:10.3390/NU8100660.

156. Olivares, M.; Castillejo, G.; Varea, V.; Sanz, Y. Double-Blind, Randomised, Placebo-Controlled Intervention Trial to Evaluate the Effects of Bifidobacterium Longum CECT 7347 in Children with Newly Diagnosed Coeliac Disease. *Br J Nutr* **2014**, *112*, 30–40, doi:10.1017/S0007114514000609.

157. Lindfors, K.; Blomqvist, T.; Juuti-Uusitalo, K.; Stenman, S.; Venäläinen, J.; Mäki, M.; Kaukinen, K. Live Probiotic Bifidobacterium Lactis Bacteria Inhibit the Toxic Effects Induced by Wheat Gliadin in Epithelial Cell Culture. *Clin Exp Immunol* **2008**, *152*, 552–558, doi:10.1111/J.1365-2249.2008.03635.X.

158. Francavilla, R.; de Angelis, M.; Rizzello, C.G.; Cavallo, N.; Dal Bello, F.; Gobbetti, M. Selected Probiotic Lactobacilli Have the Capacity To Hydrolyze Gluten Peptides during Simulated Gastrointestinal Digestion. *Appl Environ Microbiol* **2017**, *83*, doi:10.1128/AEM.00376-17.

159. Harnett, J.; Myers, S.P.; Rolfe, M. Probiotics and the Microbiome in Celiac Disease: A Randomised Controlled Trial. *Evidence-based Complementary and Alternative Medicine* **2016**, *2016*, doi:10.1155/2016/9048574.

160. Bezkorovainy, A. Probiotics: Determinants of Survival and Growth in the Gut. *Am J Clin Nutr* **2001**, *73*, 399s–405s, doi:10.1093/AJCN/73.2.399S.

161. Olivares, M.; Laparra, M.; Sanz, Y. Influence of Bifidobacterium Longum CECT 7347 and Gliadin Peptides on Intestinal Epithelial Cell Proteome. *J Agric Food Chem* **2011**, *59*, 7666–7671, doi:10.1021/JF201212M.

162. Laparra, J.M.; Sanz, Y. Bifidobacteria Inhibit the Inflammatory Response Induced by Gliadins in Intestinal Epithelial Cells via Modifications of Toxic Peptide Generation during Digestion. *J Cell Biochem* **2010**, *109*, 801–807, doi:10.1002/jcb.22459.

163. Laparra, J.M.; Olivares, M.; Gallina, O.; Sanz, Y. Bifidobacterium Longum CECT 7347 Modulates Immune Responses in a Gliadin-Induced Enteropathy Animal Model. *PLoS One* **2012**, *7*, doi:10.1371/journal.pone.0030744.

164. D'Arienzo, R.; Stefanile, R.; Maurano, F.; Mazzarella, G.; Ricca, E.; Troncone, R.; Auricchio, S.; Rossi, M. Immunomodulatory Effects of Lactobacillus Casei Administration in a Mouse Model of Gliadin-Sensitive Enteropathy. *Scand J Immunol* **2011**, *74*, 335–341, doi:10.1111/J.1365-3083.2011.02582.X.

165. Bindels, L.B.; Delzenne, N.M.; Cani, P.D.; Walter, J. Towards a More Comprehensive Concept for Prebiotics. *Nat Rev Gastroenterol Hepatol* **2015**, *12*, 303–310, doi:10.1038/NRGASTRO.2015.47.

166. Wegh, C.A.M.; Geerlings, S.Y.; Knol, J.; Roeselers, G.; Belzer, C. Postbiotics and Their Potential Applications in Early Life Nutrition and Beyond. *Int J Mol Sci* **2019**, *20*, doi:10.3390/IJMS20194673.

167. Shah, S.; Akbari, M.; Vanga, R.; Kelly, C.P.; Hansen, J.; Theethira, T.; Tariq, S.; Dennis, M.; Leffler, D.A. Patient Perception of Treatment Burden Is High in Celiac Disease Compared with Other Common Conditions. *Am J Gastroenterol* **2014**, *109*, 1304–1311, doi:10.1038/AJG.2014.29.

168. Cohen, I.S.; Day, A.S.; Shaoul, R. Gluten in Celiac Disease—More or Less? *Rambam Maimonides Med J* **2019**, *10*, doi:10.5041/RMMJ.10360.

169. Thompson, T.; Dennis, M.; Higgins, L.A.; Lee, A.R.; Sharrett, M.K. Gluten-Free Diet Survey: Are Americans with Coeliac Disease Consuming Recommended Amounts of Fibre, Iron, Calcium and Grain Foods? *J Hum Nutr Diet* **2005**, *18*, 163–169, doi:10.1111/J.1365-277X.2005.00607.X.

170. Welstead, L. The Gluten-Free Diet in the 3rd Millennium: Rules, Risks and Opportunities. *Diseases* **2015**, *3*, 136–149, doi:10.3390/DISEASES3030136.

171. Murray, J.A.; Watson, T.; Clearman, B.; Mitros, F. Effect of a Gluten-Free Diet on Gastrointestinal Symptoms in Celiac Disease. *Am J Clin Nutr* **2004**, *79*, 669–673, doi:10.1093/AJCN/79.4.669.

172. Bonder, M.J.; Tigchelaar, E.F.; Cai, X.; Trynka, G.; Cenit, M.C.; Hrdlickova, B.; Zhong, H.; Vatanen, T.; Gevers, D.; Wijmenga, C.; et al. The Influence of a Short-Term Gluten-Free Diet on the Human Gut Microbiome. *Genome Med* **2016**, *8*, 1–11, doi:10.1186/S13073-016-0295-Y/TABLES/4.

173. de Palma, G.; Nadal, I.; Collado, M.C.; Sanz, Y. Effects of a Gluten-Free Diet on Gut Microbiota and Immune Function in Healthy Adult Human Subjects. *Br J Nutr* **2009**, *102*, 1154–1160, doi:10.1017/S0007114509371767.

UNIVERSITAT ROVIRA I VIRGILI
A MULTI-OMICS APPROACH TO DESCRIBE CROSSTALK BETWEEN GUT MICROBIOTA AND HOST: EPIGENETICS AND VIROME ROLE
Polina Kazakova Borodina

HYPOTHESIS AND OBJECTIVES

UNIVERSITAT ROVIRA I VIRGILI
A MULTI-OMICS APPROACH TO DESCRIBE CROSSTALK BETWEEN GUT MICROBIOTA AND HOST: EPIGENETICS AND VIROME ROLE
Polina Kazakova Borodina

HYPOTHESIS AND OBJECTIVES

The infant gut microbiota is established at birth and is shaped during the first years of life. The development and maturation of the gut microbiota are very dynamic processes, and they are influenced by a variety of conditions, with infant feeding patterns (mainly breastfeeding or formula feeding) being one of the key factors determining the colonization of the intestinal microbiota. The gut microbiome plays an important role in human health and homeostasis; thus, the development of gut microbiota in early life impacts later health. Although bacteria are the most abundant microorganisms in the gut microbiota and have been studied the most, the virome is also part of the dynamic network of microorganisms that inhabit the gastrointestinal tract, and it is suggested to be highly involved in host immune system development. Since the gut microbiota has been confirmed to serve a key role in overall health, changes in its composition, known as dysbiosis, have been associated with long-term health disorders such as obesity, atopic diseases, and chronic inflammatory diseases. The gut bacteriome–host relationship is regulated by a complex network of interactions. The host can affect the microbial ecosystem through a variety of mechanisms, including small RNAs and other epigenetic elements. At the same time, the bacteriome is highly correlated with the virus population and metabolites.

There is also a strong relationship between intestinal dysbiosis and celiac disease. Celiac disease is an autoimmune disorder that is principally characterized by inflammation of the small intestine coupled with other extraintestinal manifestations. The global prevalence is approximately 1%, and currently, the only effective treatment against celiac disease is lifelong adherence to a strict gluten-free diet (GFD). Several studies have confirmed

the differences in terms of bacterial composition in healthy individuals and those with CeD, even if they adhered to a GFD. The intestinal microbiota could play an important role in mitigating the symptoms of CeD. Moreover, patients with persistent symptoms on a long-term GFD display an altered microbiota composition. This leaves the use of probiotics as a possible treatment to restore the bacterial population and improve CeD-related symptoms. The FAO/WHO definition of a probiotic is "live microorganisms which when administered in adequate amounts confer a health benefit on the host".

Considering previous information, we proposed two hypotheses. The first hypothesis, a descriptive hypothesis, suggested that epigenetic factors and the virome may be modulating factors for the bacterial population in the first stages of life and influenced by diet.

To verify this hypothesis, the main objective was to evaluate, following a multiomic approach, whether the viral population and the miRNA composition in stool samples from newborns and nursing babies were dependent on the type of feeding and on the bacterial population.

Different specific goals were proposed to reach this objective:

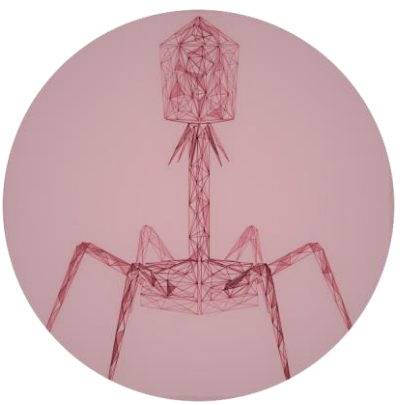
1. To describe the bacterial and viral communities and smallRNome in newborns (meconium samples) (Chapter 1).
2. To assess the effect of the type of feeding (breastfeeding/formula feeding of nursing babies) on the bacteriome, virome, smallRNome, and metabolome profile by NMR to prove the modulatory effects of the virome and miRNome on the bacterial community (Chapter 1).

The second hypothesis was that the metabolic intestinal profile added to bacterial and viral populations may be altered in CeD children adhering to a GFD in comparison with healthy children and that these alterations might be modulated by probiotic administration. To corroborate the confidence of this hypothesis, the objective was to determine the alterations in the intestinal phenotypes of children with celiac disease compared to healthy

children and evaluate the restorative effects on this intestinal phenotype enabled by the administration of a probiotic in an animal model of celiac disease.

3. To describe the differences in the intestinal phenotypes in celiac children adhering to a GFD compared with those of healthy children by following a multiomic approach (bacterial and viral community and metabolic profiling) (Chapter 2).
4. To verify the effect of a probiotic cocktail on the intestinal phenotypes (barrier integrity, bacteriome, metatranscriptome, and metabolome profile) in an animal model of celiac disease (Chapter 2).

UNIVERSITAT ROVIRA I VIRGILI
A MULTI-OMICS APPROACH TO DESCRIBE CROSSTALK BETWEEN GUT MICROBIOTA AND HOST: EPIGENETICS AND VIROME ROLE
Polina Kazakova Borodina



RESULTS

UNIVERSITAT ROVIRA I VIRGILI
A MULTI-OMICS APPROACH TO DESCRIBE CROSSTALK BETWEEN GUT MICROBIOTA AND HOST: EPIGENETICS AND VIROME ROLE
Polina Kazakova Borodina

Chapter 1

To corroborate that epigenetic factors and the virome may be modulating factors of the bacterial population in the first stages of life and influenced by diet.

UNIVERSITAT ROVIRA I VIRGILI
A MULTI-OMICS APPROACH TO DESCRIBE CROSSTALK BETWEEN GUT MICROBIOTA AND HOST: EPIGENETICS AND VIROME ROLE
Polina Kazakova Borodina

Manuscript 1

Multiomic Approach to Analyze Infant Gut Microbiota: Experimental and Analytical Method Optimization

Helena Torrell, Adrià Cereto-Massagué, Polina Kazakova, Lorena García, Héctor Palacios and Núria Canela.

Published in Biomolecules

Impact Factor: 6.064; Q2, *Biochemistry & Molecular Biology*

doi.org/10.3390/biom11070999

Multiomic Approach to Analyze Infant Gut Microbiota: Experimental and Analytical Method Optimization

Helena Torrell, Adrià Cereto-Massagué, Polina Kazakova, Lorena García, Héctor Palacios and Núria Canela

Eurecat, Centre Tecnològic de Catalunya, Centre for Omic Sciences (COS), Joint Unit Universitat Rovira i Virgili-EURECAT, Unique Scientific and Technical Infrastructures (ICTS), 43204 Reus, Spain

Correspondence: helena.torrell@eurecat.org

Keywords: next-generation sequencing; Ion Torrent; metagenomics; microbiome; early infancy microbiome; metabolomics; multiomics approach.

ABSTRACT

Background: The human intestinal microbiome plays a central role in overall health status, especially in early life stages. 16S rRNA amplicon sequencing is used to profile its taxonomic composition; however, multiomic approaches have been proposed as the most accurate methods for study of the complexity of the gut microbiota. In this study, we propose an optimized method for bacterial diversity analysis that we validated and complemented with metabolomics by analyzing fecal samples. Methods: Forty-eight different analytical combinations regarding (1) 16S rRNA variable region sequencing, (2) a feature selection approach, and (3) taxonomy assignment methods were tested. A total of 18 infant fecal samples grouped depending on the type of feeding were analyzed by the proposed 16S rRNA workflow and by metabolomic analysis. Results: The results showed that the sole use of V4 region sequencing with ASV identification and VSEARCH for taxonomy assignment produced the most accurate results. The application of this workflow showed clear differences between fecal samples according to the type of feeding, which correlated with changes in the fecal metabolic profile. Conclusion: A multiomic approach using real fecal samples from 18 infants with different types of feeding demonstrated the effectiveness of the proposed 16S rRNA-amplicon sequencing workflow.

1. INTRODUCTION

The gut microbiota is the population of microorganisms (mainly bacteria but also viruses, protozoa, and fungi) and their collective genetic material present in the gastroin- testinal tract. The intestinal microbiota plays an important role in human health, and the disruption of its composition, named dysbiosis, has been proposed as one of the most important factors involved in gastrointestinal diseases and other illnesses [1]. The gut

microbiota is highly malleable and can be altered throughout the lifespan by environmental factors, such as diet and medications. The development of the microbiota begins at birth, when very low microbial diversity is observed in the gastrointestinal tract of newborns, with the microbial population dependent on the mode of delivery (cesarean section or vaginal delivery). Many studies have shown a correlation between the acquired microbiota during the first stages of life and the development of different diseases throughout life [2,3]. Pediatric intestinal disorders cover a wide range of injuries and conditions that impact a child's intestines, being celiac disease (1%) [4], irritable bowel syndrome (9.3% to 35.5%) [5], and ulcerative colitis (0–3%) [6] the most relevant. Intestinal disorders can result from numerous conditions, but intestinal microbiota is thought to play an important role in illness development.

Complex transformations requiring microbial collaborations that are tightly regu-lated and coupled through microbial community interactions occur in the gut microbiota. There are several projects which aim to better our understanding of how the intestinal microbiota impacts human health and disease, one of the most important being The Human Microbiome Project (HMP) supported by the National Institute of Health (NIH).

The current trend in the application of omics pipelines in analyzing the actual status of diseases is evident [7–9]. Metagenomics is the study of a community of microorganisms by analyzing genomic sequences directly obtained from samples with no need to isolate and clone individual species [10]. The development of next-generation sequencing (NGS) techniques has allowed the production of high-quality and cost-effective genomic data, enough to identify and even relatively quantify microbial taxonomic units [11]. The common approach for identifying and classifying microbial species in a sample is to PCR amplify and further sequence a region (or several) of the 16S ribosomal RNA (16S rRNA) gene [12]. The use of different 16S rRNA gene regions means that the results of different studies are often not directly

comparable, diminishing the value of the inferences that can be drawn. Biases in the detection of bacteria for a given region are known to be caused by the choice of primers and the amplification protocol [11,13–20].

After improvement of sequencing methods and the appearance of NGS [21], more sequences of metagenomic data from different samples were obtained. Consequently, many 16S rRNA gene databases have been established. The most important 16S databases are the Ribosomal Database Project, RDP [22], GreenGenes [23,24] Silva [25], and Eztaxon-e [26]. To accurately identify the microbial composition of metagenomic data, traditional approaches cluster sequences into operational taxonomic units (OTUs) [27]. There are several tools for OTU selection. One of the first tools was DOTUR, [28] but UCLUST and USEARCH [29] became the most commonly used tools in the last decade, and VSEARCH [30] has emerged as an open-source alternative to USEARCH. A more recent approach for taxonomic analysis is the use of amplicon sequence variants (ASVs). Instead of clustering sequences by similarity, all unique sequences are retained after filtering low-quality and erroneous reads, which allows for far greater resolution for detecting different yet very similar sequences. ASV methods are thus able to resolve sequence differences by as little as a single nucleotide change, which allows this method to avoid similarity-based operational clustering units all together.

There are several bioinformatic tools for the identification of taxonomic classifications of metagenomes, each of which works with algorithms and pipelines. MG-RAST [31] is an online tool for annotation and taxonomic classification of metagenomes. ESPRIT [32] is an algorithm that removes low-quality reads, computes pairwise distances, groups OTUs and calculates statistical inferences to estimate the species richness [33]. EMIRGE [34] is a program for sequencing short reads of Illumina, and it is an iterative method to assemble small subunit ribosomal genes (SSUs) of metagenomic data and estimate relative taxa abundance [35]. Two of the most commonly used

programs are MOTHUR [36] and Quantitative Insights Into Microbial Ecology (QIIME) [37,38]. The pipelines of these programs are similar, with their main difference being in scope; one is a single program reimplementing everything (MOTHUR), and the other is a collection of scripts leveraging already existing tools, where one feeds to the next in a chain of inputs and outputs, which gives it access to a wider range of algorithms.

Due to the wide range of analytical possibilities, we aimed to select an optimized and validated workflow to assess bacterial diversity by comparing the results with a known bacterial community comprised of eleven species. The relative performance characteristics to be compared included employing two distinct methods of 16S rRNA library preparation (custom fusion-tag primers and a commercial kit named Ion 16S Metagenomics kit from Life Technologies) and several informatic configurations regarding feature selection and taxonomy assignment when using QIIME2 [39]. We selected QIIME2 due to its customiz-able input and output options and to enable increased transparency, reproducibility, and the use of open-source methods.

Once the optimized workflow was identified, it was validated through fecal samples from 18 infants with different types of feeding (breastfed and formula-fed) in the first month of life. Additionally, as recent multiomic approaches have been proposed as the most accurate methods for the study of the complexity of gut microbiota functions [9], 16S rRNA metagenome analysis was complemented with fecal metabolome profiling by nuclear magnetic resonance (NMR).

2. MATERIALS AND METHODS

2.1. Method Optimization

2.1.1. Bacterial Strains and DNA Preparation

Eleven representative microorganisms were used in this study. These strains include Lactobacillus brevis 216, Pediococcus parvulus 3911, Lactobacillus plantarum 220, Pediococcus pentosaceus 4208, Lactobacillus hilgardii 4786, Bacteroides coprophilus, Prevotella copri, Glu-conobacter oxydans, Acetobacter malorum 14377, Lactobacillus buchnerii 4111T, and Escherichia coli DBH10.

The genomic bacterial DNA from pure cultures was obtained by extraction with the QIAamp DNA stool kit (Qiagen, Hilden, Germany) following the protocol for Isolation of DNA from Stool for Pathogen detection (with slight modifications). The bacterial pellet or the stool samples were homogenized with 2 mL of lysis buffer ASL. Incubation at 96 C was performed for 10 min unaided by any combination of digestion enzymes or detergents to break the membranes of both Gram+ and Gram- bacteria. Then, the manufacturer's protocol was followed for the rest of the steps. The final elution was performed with 200 L of nuclease-free water. DNA quantification was performed with a Nanodrop2000 spectrophotometer (ThermoFisher Scientific, MA, USA).

2.1.2. Library Preparation

Two different methods were used for amplifying the selected 16S rRNA gene regions.

Tag-Fusion Primers

Partial 16S rRNA gene sequences were amplified from the extracted DNA using primer pair 341F-532R, which targets the V3 region of the 16S rRNA gene sequence, primer pair 515F-806R, which targets the V4 region, and primer pair 967F-1046R, which targets the V6 region (Table 1). The coordinates are based on the 16S rRNA gene of Escherichia coli strain K-12 substr. MG1655.

Table 1. Fusion-tag primer sequences used to amplify the V3, V4, and V6 16S rRNA gene regions.

Name and direction	ADAPTOR	KEY	BARCODE	SPACER	TARGET 16S PRIMER
V3_forward	CCATCTCATCCCT GCGTGTCTCCGAC	TCAG	CTAAGGTAAC	GAT	CCTACGGGRSGCAGCAG
V3_reverse	CCTCTCTATGGGC AGTCGGTGAT			CC	ATTACCGCGGCTGCT
V4_forward	CCATCTCATCCCTG CGTGTCTCCGAC	TCAG	CTAAGGTAAC	GAT	GTGCCAGCMGCCGCGGTAA
V4_reverse	CCTCTCTATGGGCA GTCGGTGAT			CC	GGACTACHVGGGTWTCTAAT
V6_forward	CCATCTCATCCCTG CGTGTCTCCGAC	TCAG	CTAAGGTAAC	GAT	CAACGCGAAGAACCTTACC
V6_reverse	CCTCTCTATGGGCA GTCGGTGAT			CC	CGACAGCCATGCANCACCT

Primer selection (Table1) was performed by searching the scientific literature [40–42], assessing their specificity using BLAST tools, and ensuring that they matched conserved regions of the 16S rRNA gene [43].

These primers were designed to include at their 5’ end one of the two adaptor se- quences used in the Ion Torrent sequencing library preparation protocol, linking a unique tag barcode of 10 bases to identify different samples (Table1).

The PCR cycle parameters used for the V4 and V6 regions were 3 min at 94 °C, 25 cycles of 30 s at 94 °C, 45 s at 57 °C and 60 s at 72 °C, followed by 2 min at 72 °C. For the V3 region, the thermal cycle was 5 min at 94 °C, 25 cycles of 30 s at 94 °C, 45 s at 55 °C and 60 s at 72 °C followed by 10 min at 72 °C. The amplification conditions employed were 18.6 µL HiFi Platinum

(Life Technologies, Carlsbad, CA, USA), 1.5 µL of each primer at 2.5 µM, and 50 ng of genomic DNA in a final volume of 25 µL. Reactions were carried out by using a Verity Thermocycler (Applied Biosystems, Foster City, CA, USA). PCR products were confirmed by a 2% agarose gel, and specific bands were excised and then purified using a Nucleospin Gel and PCR clean up kit (Macherey-Nagel, Berlin, Germany), dissolving the gel slices at room temperature by vortexing (without heat). The concentration of the PCR amplicons was analyzed by electrophoresis on a Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Equimolar pools of each fragment were combined, and 26 pM equimolar pools of all strains were also formed.

Ion 16s Metagenomic Kit from Life Technologies

Seven regions of the 16S rRNA gene were amplified with an Ion 16s Metagenomics kit (Life Technologies, Carlsbad, CA, USA), purified using AgentCourt AMPure beads (Beckman Coulter, Brea, CA, USA) and quantified using a Universal Library Quantitation kit (Life Technologies, Carlsbad, CA, USA). All procedures were performed following the manufacturer's instructions. These procedures included amplification of regions V2, V3, V4, V6, V7, V8, and V9 by two primer pools, ligation of the specific Ion Torrent adaptors and barcodes, nick repair, and final DNA purification.

2.1.3. Ion Torrent PGM Sequencing

Once the libraries were created, they were diluted to 26 pM DNA prior to clonal amplification. Emulsion PCR and ion sphere particle enrichment were carried out using the Ion PGM Template OT2 Hi-Q Kit (Life Technologies, Carlsbad, CA, USA) according to the manufacturer's instructions. Afterward, the samples were prepared for sequencing by employing the Ion PGM Hi-Q Sequencing Kit (Life Technologies, Carlsbad, CA, USA). Prepared samples were loaded on a 318 chip and sequenced using the Ion Torrent PGM applying 840 flows.

After sequencing, the individual sequence reads were filtered by PGM software to remove low-quality and polyclonal sequences. Those reads were then converted to FASTQ files and analyzed using QIIME2.

2.1.4. QIIME2 Analysis

The sequences were divided according to the library preparation method: kit (Ion 16s Metagenomic kit from Life Technologies, Carlsbad, CA, USA) and tag-fusion primers, which were further divided according to the amplified region: V3, V4, and V6. Combinations of the data from the latter were also analyzed (V3 + V4, V3 + V6, V4 + V6, and V3 + V4 + V6). All of the regions included in the commercial 16S metagenomics kit were analyzed together.

From this point, several configurations were tried, with every possible combination of the following parameters:

Feature Selection approach: Recent amplicon sequence variant (ASV) approach using dada2 [44] or the traditional operational taxonomic unit (OTU) picking method using VSEARCH [29] with 99% homology. Quality control retained sequences for which all bases had a quality score of 20 or higher.

Taxonomy assigning method: The default consensus alignment was conducted with VSEARCH, BLAST [45], or a recently proposed newer approach through a sklearn machine learning method (Naïve Bayes). The database used for both the consensus alignment reference and to train the naïve Bayes model was the Silva database [46] at 99% homology (which has been shown to outperform the 97% that is widely used [47].

All these possibilities yielded a total of 48 different combinations (Figure1).

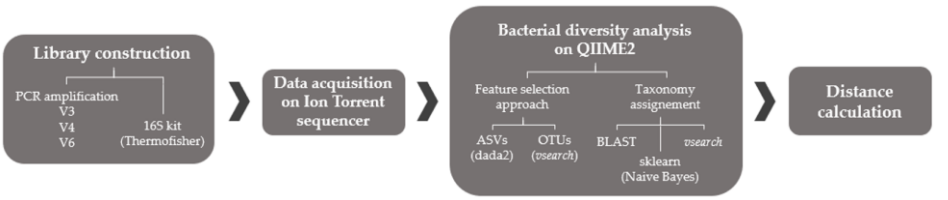


Figure 1. Scheme of all of the possible analytical combinations.

2.1.5. Performance Comparison

To compare the performance of each method in terms of accuracy with respect to the actual sample composition, there is the need for a metric that can evaluate each method's performance. This metric should be able to compare the difference between the expected results and the obtained results in terms of percentage abundance. We chose to use Bray–Curtis dissimilarity to compare the conditions. A perfect analysis method would yield 9.09% of reads identified for each of the 11 species and 0% of the remaining reads, which would give a Bray–Curtis dissimilarity of 0 for the sample composition. A method absolutely incapable of identifying any read as one of the expected species would yield a dissimilarity of 1. Thus, the closer the dissimilarity is to 0, the closer the result's analysis is to the actual composition of the sample. The further its absolute value increases, the further the analysis's results deviate from reality.

The performance of all methods was compared at three different taxonomic levels: family, genus, and species. The best workflow was selected according to its lower dissim- ilarity value. Additionally, the performance of ASV identification versus OTU picking regardless of library construction method was statistically analyzed using the Holm–Šídák post hoc adjustment of the Kruskal–Wallis test.

2.2. Method Validation

To confirm the workflow capacity to describe bacterial diversity in gut microbiota, the optimized procedure was applied to 18 infant fecal samples.

The infants were under six months old and exclusively fed milk. The infants were classified depending on the type of feeding (breast feeding, n = 9 or formula feeding, n = 9). Additionally, the fecal metabolic profile was also analyzed through NMR in 15 of the previous 18 fecal samples (breast feeding, n = 9, and formula feeding, n = 6). A multiomic correlation was performed to relate the metabolite changes with gut microbiota differences in composition (Figure2). Samples were stored at −80 ◦C until metagenomic and metabolic analysis.

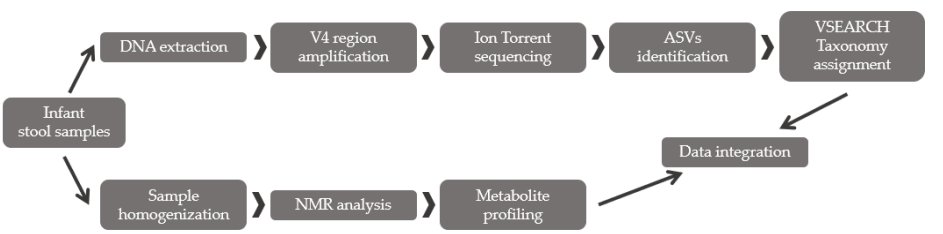


Figure 2. Scheme of the applied multiomic approach: bacterial diversity analysis combined with fecal metabolite profiling.

2.2.1. Fecal Bacterial Diversity Analysis through Partial 16S rRNA Sequencing

DNA was extracted from approximately 200 mg of fecal sample using a QIAamp® DNA Stool Mini Kit (Qiagen Inc., Hilden, Germany) according to the manufacturer's instructions. Library preparation, sequencing, and taxonomy assignment were performed as described in Section3.1. The analysis included amplicon sequence variant (ASV) identi- fication, taxonomy assignment, and alpha-diversity analysis (Shannon index). Taxonomic abundances were compared between experimental groups using the Holm–Šídák post hoc adjustment of the Kruskal–Wallis test.

2.2.2. Fecal Metabolite Profiling by NMR

Fecal samples (50 mg) were homogenized with phosphate-buffered saline

(PBS). The homogenates were centrifuged at 15,000 g for 15 min at 4 ◦C. Two-hundred microliters of the supernatants were separated from the pellet and mixed with 400 µL of D2O phosphate buffer (PBS 0.05 mM, pH 7.4, 99.5% D2O). The mixture was transferred into 5 mm NMR glass tubes for NMR measurement. The NMR spectra were measured at a 600.20 MHz frequency while using an Avance III-600 Bruker spectrometer equipped with a 5 mm PABBO BB-1H/D Z-GRD probe. A standard 1H NOESY presaturation pulse sequence (RD-90◦-t1-90◦-tm-90◦-acquire, noesypr1d) was used with water suppression. A recycle delay (RD) of 5.0 s, a mixing time (tm) of 100 ms, an acquisition time of 3.4 s, and a 90◦ pulse of 21.16 µs were used for all of the samples. NMR spectra were processed using TopSpin 3.5pl4 software (Bruker Biospin, Conventry, UK). An exponential line broadening of 0.3 Hz was applied before Fourier transformation. Metabolite identification was carried out using information from the literature, public databases, and proprietary software (Chenomx NMR Suite 8.5®, Human Metabolite DataBase, Biological Magnetic Resonance Data Bank), and normalization was performed using the PULCON (PULse length-based CONcentration determination) methodology (ERETIC®), which is based on the principle of reciprocity. Orthogonal partial least squares-discriminant analysis (OPLS-DA) was performed with Metaboanalyst software 5.0 (https://www.metaboanalyst.ca/; accessed on 23 October 2020). The statistical calculation of the significance of the correlated OPLS-DA metabolites ($p < 0.05$) was obtained using the website for statistical computation VassarStats (http://www.vassarstats.net/; accessed on 27 October 2020).

2.2.3. Integration between Metabolites and Gut Bacterial Composition

Metabolite levels and microbial diversity were analyzed through neural networks with MMVECs (Metabolite-Microbe VECtor) [48] to test whether there was co-occurrence between the type of feeding, the obtained metabolites through NMR and the detected bacterial taxonomic categories.

This was achieved by training the neural network to predict metabolite compositions based solely on the presence of each individual microbe and then inferring co-occurrence probabilities from the trained model and the taxonomic profile of each sample.

3. RESULTS

3.1. Method Optimization

The 16S rRNA gene NGS run produced 5,141,956 reads from the two library prepara- tion protocols. These reads were transformed to FASTQ format, and two FASTQ files were used as input in the QIIME2 analysis workflows. All results were split depending on the 16S rRNA gene region used to assign taxonomy prior to the QIIME2 analysis. Analyzing the dissimilarity data in detail, they clearly showed that use of the V4 region applying the ASV approach for feature selection and VSEARCH for taxonomy assignment produces the most accurate results at the three studied taxonomic levels (Table2). However, the differ- ences between this combination and the combination of ASV identification with BLAST or sklearn for taxonomy assignment is minimal at the genus and family levels (Figure3). Thus, ASV identification performed better than OTU picking at the family and genus levels regardless of the taxonomy assignment method ($p = 0.001$). However, at the species level, OTU picking + VSEARCH showed significantly lower dissimilarity values than the other combinations involving ASV identification (corrected $p < 0.05$, Figure3), although these results were conditioned by the 16S rRNA gene region considered (Table2). On the other hand, the incorporation of the V6 region into the analysis causes high dissimilarity values due to its impossibility of assigning the Bacteroidaceae family (data not shown). Thus, the exclusive use of the V3 and V4 regions seems more suitable as the dissimilarity values are smaller. However, the combination of V3 + V4 regions did not increase the performance of the analysis, showing that the amplification of only the V4 region was sufficient to correctly assess the taxonomy, at least at the genus and family levels.

Table 2. This table shows the dissimilarity values obtained at three taxonomic levels (family, genus and species), depending on the library construction method, the feature selection approach and the taxonomy assignment method.

Library	Feature	Taxonomy	Bray–Curtis Dissimilarity		
Construction	Selection		Species	Genus	Family
V4	ASV	VSEARCH	0.5576	0.1182	0.1182
V3_V4	ASV	VSEARCH	0.5577	0.1182	0.1182
V3	ASV	VSEARCH	0.5930	0.1186	0.1186
V3_V4_V6	ASV	VSEARCH	0.6221	0.2084	0.1756
V4_V6	ASV	VSEARCH	0.6222	0.2087	0.1758
V3	OTU	VSEARCH	0.6966	0.2342	0.1405
V3_V6	OTU	VSEARCH	0.7268	0.2617	0.1903
V3_V6	OTU	BLAST	0.7627	0.2601	0.1916
V3	OTU	sklearn	0.7695	0.2387	0.1445
kit	OTU	VSEARCH	0.7712	0.2506	0.1532
V6	OTU	BLAST	0.7770	0.3955	0.3580
V3_V4	OTU	VSEARCH	0.7791	0.4680	0.4374
kit	ASV	BLAST	0.7792	0.3138	0.3137
V3_V4_V6	OTU	VSEARCH	0.7825	0.4392	0.4086
V3	OTU	BLAST	0.7872	0.2338	0.1405
V6	OTU	VSEARCH	0.7916	0.3961	0.3562
kit	ASV	VSEARCH	0.7978	0.3138	0.3137
kit	OTU	BLAST	0.8077	0.2477	0.1529
V3_V4_V6	OTU	BLAST	0.8088	0.4386	0.4096
V3_V4	OTU	BLAST	0.8208	0.4682	0.4378
V4_V6	OTU	VSEARCH	0.8262	0.5376	0.5264
V4	ASV	BLAST	0.8305	0.1182	0.1182
V3_V4	ASV	BLAST	0.8307	0.1182	0.1182

V4_V6	OTU	BLAST	0.8312	0.5371	0.5278
V4_V6	ASV	BLAST	0.8351	0.2088	0.1759
V3_V4_V6	ASV	BLAST	0.8351	0.2085	0.1757
V3_V6	ASV	VSEARCH	0.8413	0.4719	0.4536
V6	ASV	VSEARCH	0.8438	0.4761	0.4577
V6	ASV	BLAST	0.8443	0.5015	0.4584
V3_V6	ASV	BLAST	0.8445	0.4970	0.4543
V4	OTU	VSEARCH	0.8454	0.6350	0.6346
kit	OTU	sklearn	0.8568	0.3560	0.2626
V4	OTU	BLAST	0.8614	0.6355	0.6353
V3	ASV	BLAST	0.8631	0.1186	0.1186
kit	ASV	sklearn	0.8724	0.5185	0.5181
V3_V6	OTU	sklearn	0.8756	0.3614	0.2275
V3	ASV	sklearn	0.8914	0.1186	0.1186
V3_V4	ASV	sklearn	0.8972	0.1182	0.1182
V4	ASV	sklearn	0.8973	0.1182	0.1182
V3_V4	OTU	sklearn	0.9144	0.6126	0.5830
V3_V4_V6	ASV	sklearn	0.9282	0.2321	0.1864
V4_V6	ASV	sklearn	0.9284	0.2325	0.1869
V3_V4_V6	OTU	sklearn	0.9367	0.5926	0.5302
V3_V6	ASV	sklearn	0.9958	0.6680	0.5621
V6	OTU	sklearn	0.9958	0.5720	0.4090
V6	ASV	sklearn	0.9968	0.6740	0.5672
V4_V6	OTU	sklearn	0.9985	0.7347	0.6812
V4	OTU	sklearn	10.000	0.8434	0.8434

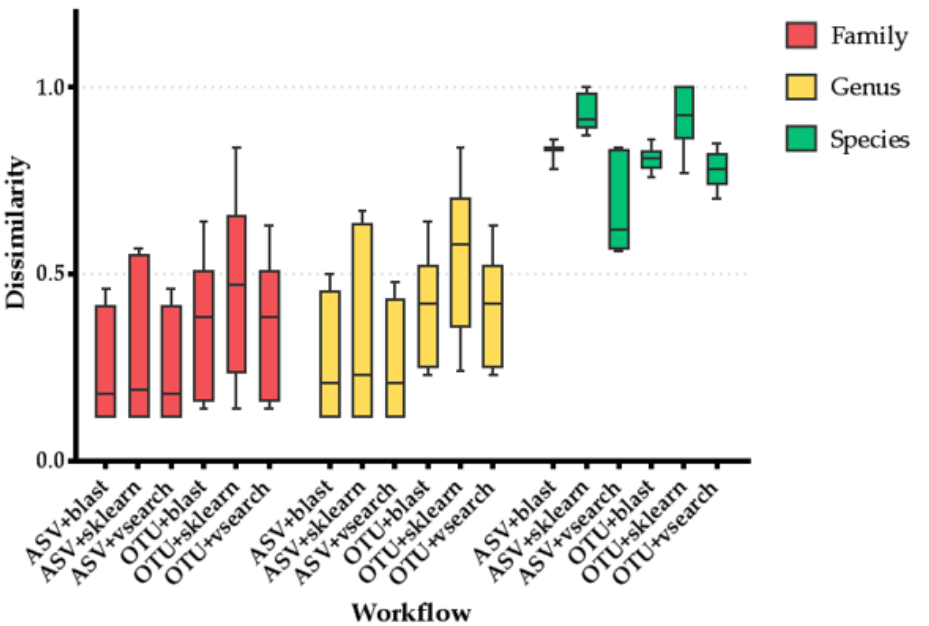


Figure 3. This graph shows the dissimilarity to the actual composition for the combinations of the QI- IME feature selection approach and taxonomy assignment method at three different taxonomic levels. The following comparisons showed statistically significant differences at the species level (corrected $p < 0.05$): ASV + sklearn vs. ASV + VSEARCH; ASV + sklearn vs. OTU + blast; ASV + sklearn vs. OTU + VSEARCH; ASV + VSEARCH vs. OTU + VSEARCH; OTU + sklearn vs. OTU + VSEARCH.

Notably, when comparing the expected and obtained results (Table3) per taxonomic level, it is clear that they are almost the same at the family and genus levels but not at the species level with six of the species not identified. These data also confirm that the addition of V3 region sequences does not improve the taxonomy assignment accuracy.

Table 3. This table shows the expected relative taxonomic abundance from the synthetic sample (Expected) and the relative abundance reported (Obtained) by the ASV + VSEARCH analysis with the V3, V4 and combination of both regions at different taxonomic resolutions.

Species	Expected	Obtained V3+V4	Obtained V3	Obtained V4
Lactobacillus brevis	9.09%	8.80%	6.62%	8.80%
Pediococcus pentosaceus	9.09%	10.05%	8.83%	10.06%
Lactobacillus plantarum	9.09%	11.89%	11.48%	11.89%
Gluconobacter oxydans	9.09%	8.14%	7.06%	8.14%
Bacteroides coprophilus	9.09%	0.00%	0.00%	0.00%
Lactobacillus hilgardii	9.09%	0.02%	0.00%	0.02%
Escherichia coli	9.09%	9.99%	9.93%	9.99%
Prevotella copri	9.09%	0.00%	0.00%	0.00%
Pediococcus parvulus	9.09%	0.00%	0.00%	0.00%
Acetobacter malorum	9.09%	0.00%	0.00%	0.00%
Lactobacillus buchneri	9.09%	0.00%	0.00%	0.00%
Other	0	51.11%	56.07%	51.09%
Genus	Expected	Obtained V3+V4	Obtained V3	Obtained V4
Bacteroides	9.09%	9.33%	10.30%	9.32%
Prevotella	9.09%	7.63%	8.68%	7.62%
Lactobacillus	36.36%	37.08%	36.72%	37.08%

Pediococcus	18.18%	18.96%	18.54%	18.96%
Acetobacter	9.09%	8.76%	8.76%	8.76%
Gluconobacter	9.09%	8.14%	7.06%	8.15%
Escherichia	9.09%	9.99%	9.93%	9.99%
Other	0.00%	0.10%	0.00%	0.10%
Family	Expected	Obtained V3+V4	Obtained V3	Obtained V4
Bacteroidaceae	9.09%	9.33%	10.30%	9.32%
Prevotellaceae	9.09%	7.63%	8.68%	7.62%
Lactobacillaceae	54.55%	56.04%	55.26%	56.04%
Acetobacteraceae	18.18%	16.90%	15.82%	16.91%
Enterobacteriaceae	9.09%	9.99%	9.93%	9.99%
Other	0	0.10%	0.00%	0.10%

3.2. Method Validation

3.2.1. Differences in Bacterial Diversity between the Experimental Groups

The sequencing run generated 5,423,214 reads that were used to determine ASVs from the V4 region of the 16S rRNA gene, which were then used to summarize the relative abundance of the microbial clades at different taxonomic levels. These differences are mainly due to changes in the Firmicutes and Bacteroidetes phyla (predominant in the breast-fed group (44% and 30%) and Proteobacteria (predominant in the formula-fed group (57%)). These differences were also transferred to lower taxonomic levels (Figure4a). A significant decrease in the *Enterobacteriaceae* family ($p = 0.0064$) was observed in the breast-fed group compared to the formula-fed group, paired with an increase in the *Staphylococcaceae* ($p = 0.016$), *Porticoccaceae* ($p = 0.022$), and *Immundisolibacteraceae* ($p = 0.022$) families. Additionally, greater species richness was observed in the samples belonging to the breastfed group, although the differences were not

significant (Figure4b).

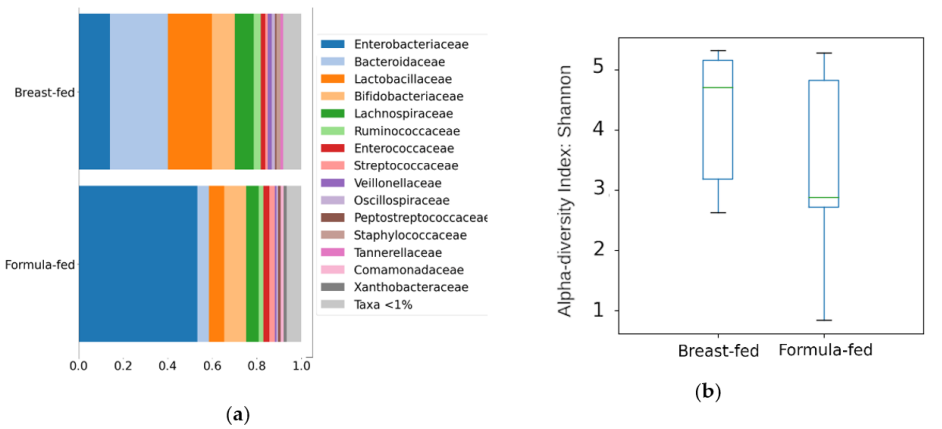


Figure 4. (a) Taxonomic composition of the community at the Family level using a Stacked Bar plot. (b) Alpha diversity measure using Shannon, Y axis represent Shannon score values.

3.2.2. Multivariate Analysis Showed Differences in the Fecal Metabolome Depending on the Type of Feeding

A total of 27 metabolites were identified and quantified after the alignment and normal- ization of the spectra. Table4shows the mean concentration per each metabolite per group, and additionally raw data has been upload to MetaboLights (ebi.ac.uk/metabolights; ac- cessed on: 6 June 2021) with the accession number MTBLS2942. To analyze the differences between the two groups, a pairwise OPLS-DA model was performed to compare the ef- fect of lactation on the fecal metabolome. A model with predictive ability [R2X = 0.218; Q2 = 0.307] was obtained comparing the fecal metabolites of both groups. The metabo- lites that were significantly altered ($p < 0.05$) in this model and found to be higher in the formula-fed group were valine, phenylalanine, lactate, isoleucine, glycine, citrate, choline, aspartate, alanine, and 2-hydroxy-3-methylbutyric. We also observed tendencies in differ- ent metabolites, such as an increase in butyrate concentration ($p = 0.078$) and isoleucine ($p = 0.077$) in the formula-fed group. No significant

metabolites were found to be increased in breastfed infants compared with the formula-fed group (Figure5) [49].

Table 4. Mean concentration per group in mmol/mg of representative fecal metabolites analysed by Nuclear Magnetic Resonance. Desvest means standard deviation and SEM stands for standard error of the mean.

Metabolite	Mean	Desvest	SEM	Mean	Desvest	SEM
Valine	1.25	0.76	0.25	3.51	2.43	0.99
Uracil	0.05	0.06	0.02	0.14	0.08	0.04
Tyrosine	0.12	0.08	0.03	0.16	0.09	0.04
Tryprophan	0.06	0.05	0.02	0.11	0.06	0.02
Succinate	2.12	2.12	0.71	6.17	10.47	4.28
Propionate	1.73	1.24	0.41	2.33	1.23	0.50
Phenylalanine	0.12	0.07	0.02	0.30	0.17	0.07
Methionine sulfoxide	0.03	0.03	0.01	0.04	0.04	0.02
Methionine	0.14	0.08	0.03	0.33	0.20	0.08
Leucine	0.69	0.39	0.13	1.58	1.05	0.43
Lactate	0.48	0.37	0.12	7.82	8.12	3.32
Isoleucine	0.79	0.51	0.17	1.96	1.17	0.48
Hypoxanthine	0.06	0.03	0.01	0.10	0.07	0.03
Glycine	0.41	0.18	0.06	2.61	2.17	0.89
Glutamate	0.83	0.71	0.24	1.48	1.08	0.44
Fumarate	0.03	0.02	0.01	0.02	0.01	0.00
Fucose	0.21	0.30	0.10	0.25	0.17	0.07
Formate	0.02	0.01	0.00	0.07	0.07	0.03
Citrate	0.18	0.15	0.05	1.18	1.15	0.47
Choline	0.07	0.04	0.01	0.17	0.06	0.02
Butyrate	1.21	1.24	0.41	3.99	2.78	1.13
Aspartate	0.21	0.12	0.04	0.67	0.70	0.29
Alanine	1.31	0.86	0.29	3.24	2.06	0.84
Acetate	6.72	5.71	1.90	16.64	12.19	4.98
5-Aminopentanoate	0.82	1.70	0.57	1.22	1.41	0.58
2-Hydroxy-3-methylbutyric	0.04	0.03	0.01	0.11	0.07	0.03
2-3-Butanediol	0.23	0.16	0.05	0.44	0.60	0.25

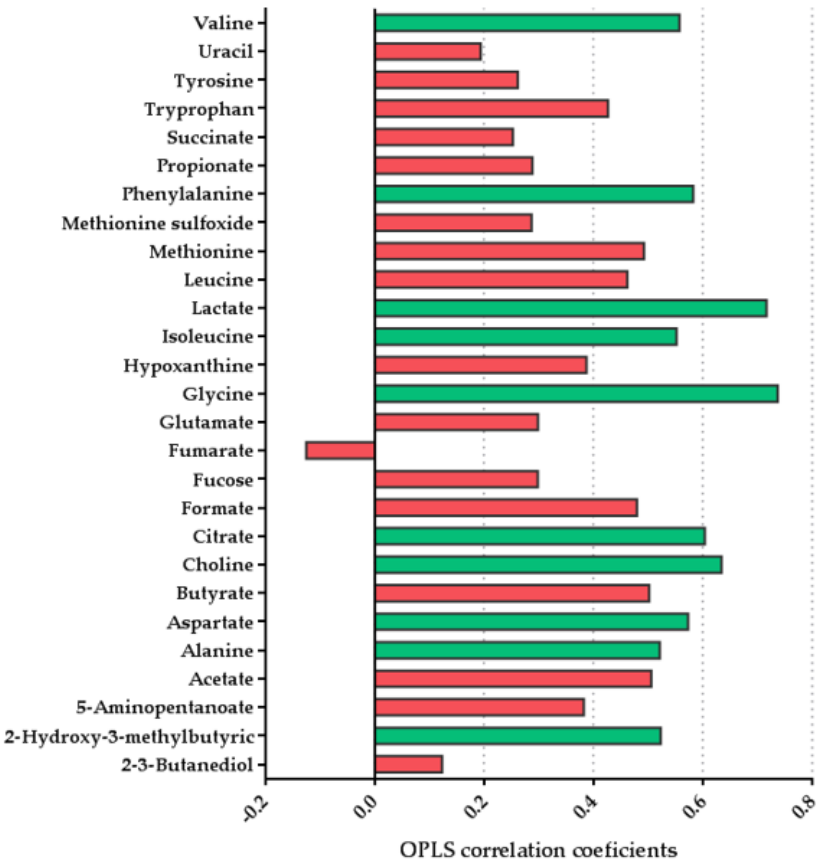


Figure 5. OPLS-DA line plot indicating metabolite differentiation between the groups (n = 9 breast- fed group and n = 6 formula-fed group). Positive correlation coefficients indicate higher levels in the formula-fed group. The significant metabolite bars are represented in green ($p < 0.05$).

3.2.3. Integration between Metabolites and Gut Microbiome

Some of the analyzed metabolites co-occurred with the bacterial population in the infant fecal samples. The most relevant co-occurrences were observed in aspartate with a positive correlation with the Staphylococcales, Clostridiales, Lachnospirales, and Pasteurellales orders; lactate with a negative correlation in the Clostridiales order; and glycine with a positive

correlation in the Bifidobacteriales order and a negative correlation with the Lachnospirales order (Figure6).

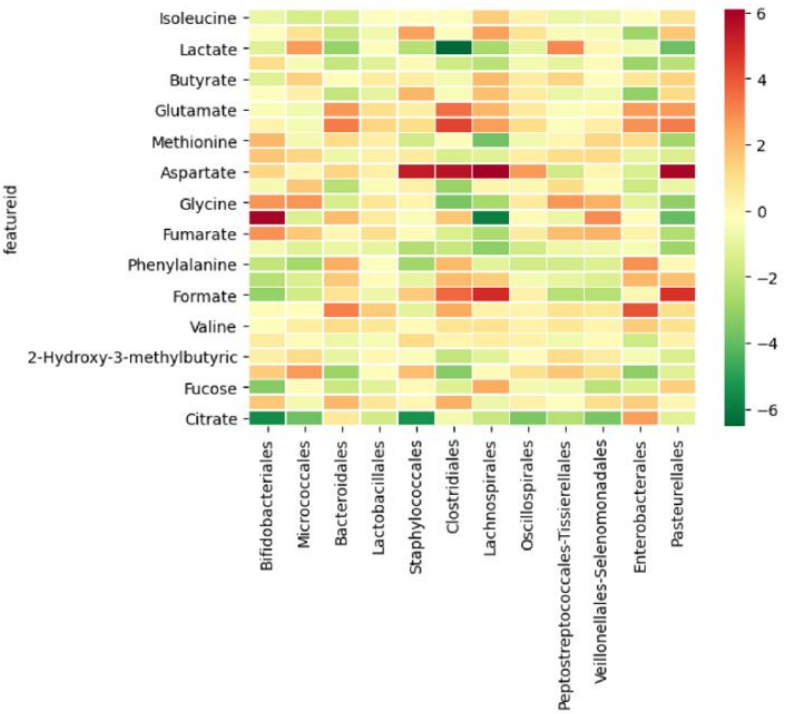


Figure 6. Heatmap with correlation co-occurrence between metabolites and the bacterial community at the order level. Darker red shows a stronger positive co-ocurrence (both occur together); dark green shows a stronger negative co-ocurrence (one occurs in the absence of the other one); pale yellow shows no significant co-ocurrence.

4. DISCUSSION

4.1. Method Optimization

These results showed that the analysis of the V4 region using ASV identification and VSEARCH for taxonomy assignment produced the most accurate results at the three studied taxonomic levels.

These results demonstrated that the precise assessment of the composition of a given microbiota population depends on several items, either

methodological or analytical. Different choices applied to the same sample may lead to different outcomes in describing the microbial population. Here, we used an Ion Torrent Machine to generate genomic data. It has been documented that Ion Torrent exhibits a higher rate of sequencing errors than data from the Illumina platform [50]; however, in partial 16S rRNA sequencing approaches, both platforms are generally comparable [51].

Regardless of the sequencing platform, the first aspect to consider is the selection of 16S rRNA gene regions to be sequenced. We examined the V3, V4, and V6 regions and analyzed the complete 16S rRNA gene amplification mediated by a 16S metagenomics kit from Life Technologies to compare the accuracy of all library preparation approaches. This commercial kit allows the sequencing of several regions of the 16S rRNA gene, but in our study, it did not lead to the expected results. One explanation for this discordance can be erroneous primer design, varying amplification efficiencies and coverage rates of the primers contained in this kit (unknown sequences), or high error and chimera rates [52]. These results do not agree with a previous study that also analyzed the performance of the Thermo Fisher 16S kit compared with a commercial service based on V4 sequencing and concluded that the Ion 16S kit generally allowed the greatest number of taxonomic identifi- cations across the domain Bacteria. However, it used different bioinformatic pipelines to assess taxonomy [20]. Nevertheless, we insist on considering cost and convenience factors. For instance, tag-fusion primers are easy to design and cheap to acquire when working on the Ion Torrent platform (depending on how many samples are multiplexed in a single run and how many variable regions are chosen). Furthermore, library construction through tag-fusion primers requires less hands-on time and fewer steps than commercial kits. In addition, the amplification of solely one variable region per sample permits more samples per run, as more reads can be generated per sample without losing resolution.

For variable region selection, V3 and V4 are currently the most popular

regions to be examined, although V6 has also traditionally been used [40,53,54]. We observed that the V6 region led to inaccurate results, suggesting insufficient specificity for that region or for the designed primers, perhaps due to polymorphisms accumulated in the conserved regions or perhaps due to its short length. Some studies have suggested that the full sequencing of the V3–V4 regions provides the closest analog to sequencing the entire 16S rRNA gene [55], while others have proposed that the single use of the V4 region is sufficient [56,57]. However, all of them applied OTU picking for feature selection instead of ASV identification. In our study, although the combination of the V3 and V4 regions generated more reads, the sequencing information obtained from the V4 region alone was able to assign the correct taxonomic rank to all of the identified ASVs at the genus level. Regardless of the informatic workflow applied, our approach failed in assigning taxonomy at the species level, probably due to the low taxonomic and phylogenetic resolution of the 16S rRNA gene analyzed regions. The reasons behind these peculiarities may be related to the functional variable regions: the V4, V5, and V6 regions directly take part in translation and are responsible for binding tRNAs and interacting with the 23S rRNA; therefore, these regions should be the most conserved. It is logical to think that more conserved regions should be sufficiently distinct only at higher taxa levels, while less conserved regions could distinguish among the lower levels [55]. To solve this issue, in experiments where it is highly important to determine the bacterial population at the species level, it has been proposed to additionally sequence the rRNA internal transcribed spacer (ITS) [58]. This taxonomic profiling can be performed using tag-fusion primers and can be analyzed on QIIME because specific ITS databases are available.

Regarding feature selection approaches [21], we applied QIIME's open-reference OTU picking method using VSEARCH, which also used a quality filter and checked for chimeras in the input reads before OTU picking. We

also performed ASV identification, for which QIIME offers two possible backends: DADA2 and deblur [59]. Per QIIME2 documentation, deblur should be used only on Illumina reads, and as our reads were single- end reads generated with a different technology, we selected DADA2 instead. Notably, OTU approaches have served the microbiome community for many years and will likely still find use in the future in specific circumstances; however, there is an ever-increasing amount of evidence that ASV approaches are the superior choice for most future analyses [60]. Thus, in this study, ASVs were expected to outperform OTU-based approaches, and they did so at the family and genus levels. The unexpected result at the species level, with no consistent winner between OTUs and ASVs across different pipelines, may be due to the previous inability to obtain species-level resolution with this dataset [61]. A recent study also showed that DADA2 offered the best sensitivity as compared to other bioinformatic pipelines for the analysis of amplicon sequence data and concluded that ASV-level workflows offer superior resolution compared to OTU-level workflows [61].

The next step involves assigning taxonomies to the features (OTUs and ASVs), which again can be done with several different methods. QIIME's available consensus alignment methods, using VSEARCH and BLAST, respectively, were tested, alongside the machine learning taxonomy assignment method based on a naïve Bayes model trained on the same reference database (Silva) [48]. Considering the results, the VSEARCH consensus alignment provides more reliable and consistent results, topping the results for all pipeline combinations. However, the difference from the other 2 methods is very small in most cases, while the sklearn-based method had the widest dispersion in performance. Similar results were shown in a previous study in which a sklearn naïve Bayes machine-learning classifier, alignment-based taxonomy consensus methods based on VSEARCH, and BLAST met or exceeded the species-level accuracy of other

commonly used methods [62].

4.2. Method Validation through a Multiomics Approach

With the optimized workflow, and only selecting the V4 region of the 16S rRNA gene, it was possible to characterize the infant gut microbiota in fecal samples and separate them into two groups depending on the type of feeding, replicating already published studies. The small sample size of this study (n = 18 for microbiota analysis and n = 15 for metabolomic profile) must be noted. Although it is the main limitation; this sample size is enough for an exploratory study. A significant decrease in the Enterobacteriaceae family (Mann-Whitney, $p = 0.0064$) was observed in the breast-fed group compared to the formula-fed group, paired with an increase in the *Staphylococcaceae, Porticoccaceae*, and *Immundisolibacteraceae* families. These results agree with previous studies comparing the influence of the type of feeding on gut microbiota in infants and when using a similar workflow with only the V4 region of the 16S rRNA gene with the QIIME 1.8 and Greengenes reference databases [63].

Additionally, the multiomic approach, with the inclusion of metabolic profiling from feces, contributes to separating the two groups and is also correlated with changes in the microbiota. We observed tendencies in different metabolites, such as an increase in butyrate concentration and isoleucine in the formula-fed group, as reported previously [49]. Levels of choline were higher in feces of the formula-fed group in contrast with a previous study where the choline concentration was higher in breast-fed infants, but they were using serum samples [64]. These results might be due to high levels of these metabolites in the formula milk administered to infants. The observed positive correlation between bacteria from the Firmicutes phylum and butyrate may be explained by its ability to produce this short-chain fatty acid (SCFA) [65,66]. The negative correlation between Bifidobacteriales and citrate and other SCFAs that we observed was also

described previously [67]. It is important to note that intrinsic factors such as genetics and other environmental factors such as lifestyle may be influencing both microbiota and metabolic phenotypes, and thus contributing to differentiate both experimental groups.

5. CONCLUTION

It has been shown that microbiota community description profiling can be affected by the specific target region used, library preparation method and analysis workflow because all of them will give different results in terms of error rates and biases. We concluded that the sequencing of the V4 region only, using ASV identification for feature selection and VSEARCH for taxonomy assignment, produces the most accurate results when working on the Ion Torrent NGS platform. However, all combinations struggled to correctly identify the samples at the species level, but the results up to the genus level were satisfactory. Notably, this setup outperformed the use of the commercial primer kit.

The optimized workflow applied to fecal samples in an infant nutrition study demon- strated its ability to assign taxonomy and to discern between samples and groups when working with a fecal matrix. There is a need for workflow standardization when analyzing bacterial diversity through 16S rRNA gene partial sequencing, and here we provide a validated protocol which can be applied either in research, clinical or commercial field to accurate determine bacterial communities from a variety of samples. Furthermore, its complementation with other omic data, such as metabolite profiling, increases the potential use of metagenomics to enlarge the knowledge of microbiota functions and relationships with host physiopathology.

REFERENCES

1. Yang, T.; Santisteban, M.; Rodriguez, V.; Vermali, R.; Ahmari, N.; Carvajal, J.M.; Zadeh, M.; Gong, M.; Qi, Y.; Zubcevic, J.; et al. Gut Dysbiosis Is Linked to

Hypertension. Hypertension 2015, 65, 1331–1340. [CrossRef] [PubMed]

2. Clemente, J.C.; Ursell, L.K.; Parfrey, L.W.; Knight, R. The Impact of the Gut Microbiota on Human Health: An Integrative View. Cell 2012, 148, 1258–1270. [CrossRef]

3. Chang, C.; Lin, H. Dysbiosis in gastrointestinal disorders. Best Pract. Res. Clin. Gastroenterol. 2016, 30, 3–15. [CrossRef]

4. Caio, G.; Volta, U.; Sapone, A.; Leffler, D.A.; de Giorgio, R.; Catassi, C.; Fasano, A. Celiac disease: A comprehensive current review. BMC Med. 2019, 17, 142. [CrossRef]

5. Ibrahim, N.K. A systematic review of the prevalence and risk factors of irritable bowel syndrome among medical students. Turk. J. Gastroenterol. 2016, 27, 10–16. [CrossRef]

6. Ng, S.C.; Shi, H.Y.; Hamidi, N.; Underwood, F.E.; Tang, W.; Benchimol, E.I.; Panaccione, R.; Ghosh, S.; Wu, J.C.Y.; Chan, F.K.L.; et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: A systematic review of population-based studies. Lancet 2017, 390, 2769–2778. [CrossRef]

7. Tilocca, B.; Pieroni, L.; Soggiu, A.; Britti, D.; Bonizzi, L.; Roncada, P.; Greco, V. Gut–Brain Axis and Neurodegeneration: State-of-the-Art of Meta-Omics Sciences for Microbiota Characterization. Int. J. Mol. Sci. 2020, 21, 4045. [CrossRef]

8. Zhang, X.; Li, L.; Butcher, J.; Stintzi, A.; Figeys, D. Advancing functional and translational microbiome research using meta-omics approaches. Microbiome 2019, 7, 154. [CrossRef]

9. Guirro, M.; Costa, A.; Gual-Grau, A.; Mayneris-Perxachs, J.; Torrell, H.; Herrero, P.; Canela, N.; Arola, L. Multi-omics approach to elucidate the gut microbiota activity: Metaproteomics and metagenomics connection. Electrophoresis 2018, 39, 1692–1701. [CrossRef]

10. Handelsman, J. Magnetic resonance in Pb x Nb y O z -ceramics as a system containing. Microbiol. Mol. Biol. Rev. 2004, 68, 1477–1481. [CrossRef]

11. Salipante, S.J.; Kawashima, T.; Rosenthal, C.; Hoogestraat, D.R.; Cummings, L.A.; Sengupta, D.J.; Harkins, T.T.; Cookson, B.T.; Hoffman, N.G. Performance Comparison of Illumina and Ion Torrent Next-Generation Sequencing Platforms for 16S rRNA-Based Bacterial Community Profiling. Appl. Environ. Microbiol. 2014, 80, 7583–7591. [CrossRef] [PubMed]

12. Lane, D.J.; Pace, B.; Olsen, G.J.; Stahl, D.A.; Sogin, M.L.; Pace, N.R. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses (reverse transcriptase/dideoxynudeotide). Evolution 1985, 82, 6955–6959.

13. Shaufi, M.A.M.; Sieo, C.C.; Chong, C.W.; Gan, H.M.; Ho, Y.W. Deciphering chicken gut microbial dynamics based on high- throughput 16S rRNA metagenomics analyses. Gut Pathog. 2015, 7, 4–12. [CrossRef] [PubMed]

14. Decuypere, S.; Meehan, C.J.; Van Puyvelde, S.; De Block, T.; Maltha, J.; Palpouguini, L.; Tahita, M.; Tinto, H.; Jacobs, J.; Deborggraeve, S. Diagnosis of Bacterial Bloodstream Infections: A 16S Metagenomics Approach. PLoS Negl. Trop. Dis. 2016, 10, e0004470. [CrossRef] [PubMed]

15. Claesson, M.J.; Wang, Q.; O'Sullivan, O.; Greene-Diniz, R.; Cole, J.R.; Ross, R.P.; O'Toole, P.W. Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. Nucleic Acids Res. 2010, 38, e200. [CrossRef]

16. Lagier, J.-C.; Armougom, F.; Million, M.; Hugon, P.; Pagnier, I.; Robert, C.; Bittar, F.; Fournous, G.; Gimenez, G.; Maraninchi, M.; et al. Microbial culturomics: Paradigm shift in the human gut microbiome study. Clin. Microbiol. Infect. 2012, 18, 1185–1193. [CrossRef]

17. Pfleiderer, A.; Lagier, J.-C.; Armougom, F.; Robert, C.; Vialettes, B.; Raoult, D. Culturomics identified 11 new bacterial species from a single anorexia nervosa stool sample. Eur. J. Clin. Microbiol. Infect. Dis. 2013, 32, 1471–1481. [CrossRef]

18. Ku, H.-J.; Lee, J.-H. Development of a novel long-range 16S rRNA universal primer set for metagenomic analysis of gastrointestinal microbiota in newborn infants. J. Microbiol. Biotechnol. 2014, 24, 812–822. [CrossRef]

19. Kim, M.; Morrison, M.; Yu, Z. Evaluation of different partial 16S rRNA gene sequence regions for phylogenetic analysis of microbiomes. J. Microbiol. Methods 2011, 84, 81–87. [CrossRef]

20. Sperling, J.L.; Silva-Brandão, K.; Brandão, M.; Lloyd, V.; Dang, S.; Davis, C.; Sperling, F.; Magor, K. Comparison of bacterial 16S rRNA variable regions for microbiome surveys of ticks. Ticks Tick-Borne Dis. 2017, 8, 453–461. [CrossRef]

21. Callahan, B.J.; McMurdie, P.J.; Holmes, S.P. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J. 2017, 11, 2639–2643. [CrossRef]

22. Larsen, N.; Olsen, G.J.; Maidak, B.L.; McCaughey, M.J.; Overbeek, R.A.; Macke, T.J.; Marsh, T.L.; Woese, C.R. The ribosomal database project. Nucleic Acids Res. 1993, 21, 3021–3023. [CrossRef]

23. DeSantis, T.Z.; Hugenholtz, P.; Larsen, N.; Rojas, M.; Brodie, E.L.; Keller, K.; Huber, T.; Dalevi, D.; Hu, P.; Andersen, G.L. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl. Environ. Microbiol. 2006, 72, 5069–5072. [CrossRef]

24. McDonald, D.; Price, M.N.; Goodrich, J.; Nawrocki, E.P.; DeSantis, T.Z.; Probst, A.; Andersen, G.L.; Knight, R.; Hugenholtz, P. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J. 2011, 6, 610–618. [CrossRef] [PubMed]

25. Pruesse, E.; Quast, C.; Knittel, K.; Fuchs, B.; Ludwig, W.; Peplies, J.; Glöckner, F.O. SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. 2007, 35, 7188–7196. [CrossRef] [PubMed]

26. Kim, O.-S.; Cho, Y.-J.; Lee, K.; Yoon, S.-H.; Kim, M.; Na, H.; Park, S.-C.; Jeon, Y.S.; Lee, J.-H.; Yi, H.; et al. Introducing EzTaxon-e: A prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. Int. J. Syst. Evol. Microbiol. 2012, 62, 716–721. [CrossRef] [PubMed]

27. Chan, E.R.; Hester, J.; Kalady, M.; Xiao, H.; Li, X.; Serre, D. A novel method for determining microflora composition using dynamic phylogenetic analysis of 16S ribosomal RNA deep sequencing data. Genomics 2011, 98, 253–259. [CrossRef]

28. Schloss, P.D.; Handelsman, J.; Schloss, P.D.; Handelsman, J. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl. Environ. Microbiol. 2005, 71, 1501–1506. [CrossRef]

29. Edgar, R.C. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 2010, 26, 2460–2461. [CrossRef]

30. Rognes, T.; Flouri, T.; Nichols, B.; Quince, C.; Mahé, F. VSEARCH: A versatile open source tool for metagenomics. PeerJ 2016, 4, e2584. [CrossRef] [PubMed]

31. Meyer, F.; Paarmann, D.; Souza, M.D.; Olson, R.; Glass, E.M.; Kubal, M.; Paczian, T.; Rodriguez, A.; Stevens, R.; Wilke, A.; et al. The metagenomics RAST server—A public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinform. 2008, 9, 386. [CrossRef] [PubMed]

32. Sun, Y.; Cai, Y.; Liu, L.; Yu, F.; Farrell, M.L.; Mckendree, W.; Farmerie, W. ESPRIT: Estimating species richness using large collections of 16S rRNA pyrosequences. Nucleic Acids Res. 2009, 37, e76. [CrossRef]

33. Sun, Y.; Cai, Y.; Mai, V.; Farmerie, W.; Yu, F.; Li, J.; Goodison, S. Advanced computational algorithms for microbial community analysis using massive 16S rRNA sequence data. Nucleic Acids Res. 2010, 38, e205. [CrossRef]

34. Miller, C.S.; Baker, B.J.; Thomas, B.C.; Singer, S.W.; Banfield, J.F. EMIRGE: Reconstruction of full-length ribosomal genes from microbial community short read sequencing data. Genome Biol. 2011, 12, R44. [CrossRef]

35. Ong, S.H.; Kukkillaya, V.U.; Wilm, A.; Lay, C.; Ho, E.X.P.; Low, L.; Hibberd, M.L.; Nagarajan, N. Species Identification and Profiling of Complex Microbial Communities Using Shotgun Illumina Sequencing of 16S rRNA Amplicon Sequences. PLoS ONE 2013, 8, e60811. [CrossRef]

36. Schloss, P.D.; Westcott, S.L.; Ryabin, T.; Hall, J.R.; Hartmann, M.; Hollister, E.B.; Lesniewski, R.A.; Oakley, B.B.; Parks, D.H.; Robinson, C.J.; et al. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. Appl. Environ. Microbiol. 2009, 75, 7537–7541. [CrossRef]

37. Caporaso, J.G.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F.D.; Costello, E.K.; Fierer, N.; Peña, A.G.; Goodrich, J.K.; Gordon, J.I.; et al. QIIME Allows Analysis of High-Throughput Community Sequencing data. Nat. Methods 2010, 7, 335–336. [CrossRef]

38. Kuczynski, J.; Lauber, C.L.; Walters, W.A.; Parfrey, L.W.; Clemente, J.C.; Gevers, D.; Knight, R. Experimental and analytical tools for studying the human microbiome. Nat. Rev. Genet. 2011, 13, 47–58. [CrossRef]

39. Bolyen, E.; Rideout, J.R.; Dillon, M.R.; Bokulich, N.A.; Abnet, C.C.; Al-Ghalith, G.A.; Alexander, H.; Alm, E.J.; Arumugam, M.; Asnicar, F.; et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat. Biotechnol. 2019, 37, 852–857. [CrossRef] [PubMed]

40. Milani, C.; Hevia-Gonzalez, A.; Foroni, E.; Duranti, S.; Turroni, F.; Lugli, G.A.; Sanchez, B.; Martin, R.; Gueimonde, M.; Van Sinderen, D.; et al. Assessing the Fecal Microbiota: An Optimized Ion Torrent 16S rRNA Gene-Based Analysis Protocol. PLoS ONE 2013, 8, e68739. [CrossRef] [PubMed]

41. Tridico, S.R.; Murray, D.C.; Addison, J.; Kirkbride, K.P.; Bunce, M. Metagenomic

analyses of bacteria on human hairs: A qualitative assessment for applications in forensic science. Investig. Genet. 2014, 5, 16. [CrossRef] [PubMed]

42. Whiteley, A.; Jenkins, S.; Waite, I.; Kresoje, N.; Payne, H.; Mullan, B.; Allcock, R.; O'Donnell, A. Microbial 16S rRNA Ion Tag and community metagenome sequencing using the Ion Torrent (PGM) Platform. J. Microbiol. Methods 2012, 91, 80–88. [CrossRef]

43. Wang, Y.; Qian, P.Y. Conservative Fragments in Bacterial 16S rRNA Genes and Primer Design for 16S Ribosomal DNA Amplicons in Metagenomic Studies. PLoS ONE 2009, 4, e7401. [CrossRef] [PubMed]

44. Callahan, B.J.; Mcmurdie, P.J.; Rosen, M.J.; Han, A.W.; Johnson, A.J.A.; Holmes, S.P. DADA2: High-resolution sample inference from Illumina amplicon data. Nat. Methods 2016, 13, 581–583. [CrossRef] [PubMed]

45. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. J. Mol. Biol. 1990, 215, 403–410. [CrossRef]

46. Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glöckner, F.O. The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools. Nucleic Acids Res. 2012, 41, D590–D596. [CrossRef]

47. Edgar, R.C. Updating the 97% identity threshold for 16S ribosomal RNA OTUs. Bioinformatics 2018, 34, 2371–2375. [CrossRef]

48. Morton, J.T.; Aksenov, A.A.; Nothias, L.F.; Foulds, J.R.; Quinn, R.A.; Badri, M.H.; Swenson, T.L.; Van Goethem, M.; Northen, T.R.; Vazquez-Baeza, Y.; et al. Learning representations of microbe–metabolite interactions. Nat. Methods 2019, 16, 1306–1314. [CrossRef] [PubMed]

49. He, X.; Parenti, M.; Grip, T.; Lönnerdal, B.; Timby, N.; Domellöf, M.; Hernell, O.; Slupsky, C.M. Fecal microbiome and metabolome of infants fed bovine MFGM supplemented formula or standard formula with breast-fed infants as reference: A randomized controlled trial. Sci. Rep. 2019, 9, 11589. [CrossRef]

50. Loman, N.J.; Misra, R.V.; Dallman, T.J.; Constantinidou, C.; Gharbia, S.E.; Wain, J.; Pallen, M. Performance comparison of benchtop high-throughput sequencing platforms. Nat. Biotechnol. 2012, 30, 434–439. [CrossRef]

51. Onywera, H.; Meiring, T.L. Comparative analyses of Ion Torrent V4 and Illumina V3-V4 16S rRNA gene metabarcoding methods for characterization of cervical microbiota: Taxonomic and functional profiling. Sci. Afr. 2020, 7, e00278.

[CrossRef]

52. D'Amore, R.; Ijaz, U.Z.; Schirmer, M.; Kenny, J.G.; Gregory, R.; Darby, A.; Shakya, M.; Podar, M.; Quince, C.; Hall, N. A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling. BMC Genom. 2016, 17, 1–20. [CrossRef] [PubMed]

53. Huse, S.M.; Dethlefsen, L.; Huber, J.; Welch, D.M.; Relman, D.; Sogin, M.L. Exploring Microbial Diversity and Taxonomy Using SSU rRNA Hypervariable Tag Sequencing. PLoS Genet. 2008, 4, e1000255. [CrossRef]

54. Liu, Z.; Lozupone, C.; Hamady, M.; Bushman, F.; Knight, R. Short pyrosequencing reads suffice for accurate microbial community analysis. Nucleic Acids Res. 2007, 35, e120. [CrossRef] [PubMed]

55. Bukin, Y.S.; Galachyants, Y.P.; Morozov, I.V.; Bukin, S.V.; Zakharenko, A.S.; Zemskaya, T. The effect of 16S rRNA region choice on bacterial community metabarcoding results. Sci. Data 2019, 6, 190007. [CrossRef] [PubMed]

56. Soergel, D.A.W.; Dey, N.; Knight, R.; Brenner, S.E. Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences. ISME J. 2012, 6, 1440–1444. [CrossRef] [PubMed]

57. Di Segni, A.; Braun, T.; Benshoshan, M.; Barhom, S.F.; Saar, E.G.; Cesarkas, K.; Squires, J.E.; Keller, N.; Haberman, Y. Guided Protocol for Fecal Microbial Characterization by 16S rRNA-Amplicon Sequencing. J. Vis. Exp. 2018, 2018, e56845. [CrossRef]

58. Milani, C.; Lugli, G.A.; Turroni, F.; Mancabelli, L.; Duranti, S.; Viappiani, A.; Mangifesta, M.; Segata, N.; Van Sinderen, D.; Ventura, M. Evaluation of bifidobacterial community composition in the human gut by means of a targeted amplicon sequencing (ITS) protocol. FEMS Microbiol. Ecol. 2014, 90, 493–503. [CrossRef] [PubMed]

59. Amir, A.; McDonald, D.; Navas-Molina, J.A.; Kopylova, E.; Morton, J.T.; Xu, Z.Z.; Kightley, E.P.; Thompson, L.R.; Hyde, E.R.; Gonzalez, A.; et al. Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. mSystems 2017, 2, e00191-16. [CrossRef]

60. Caruso, V.; Song, X.; Asquith, M.; Karstens, L. Performance of Microbiome Sequence Inference Methods in Environments with Varying Biomass. mSystems 2019, 4, e00163-18. [CrossRef]

61. Prodan, A.; Tremaroli, V.; Brolin, H.; Zwinderman, A.H.; Nieuwdorp, M.; Levin,

E. Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing. PLoS ONE 2020, 15, e0227434. [CrossRef]

62. Bokulich, N.A.; Kaehler, B.D.; Rideout, J.R.; Dillon, M.; Bolyen, E.; Knight, R.; Huttley, G.A.; Caporaso, J.G. Gregory Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2Js q2-feature-classifier plugin. Microbiome 2018, 6, 90. [CrossRef] [PubMed]

63. Wang, Z.; Neupane, A.; Vo, R.; White, J.; Wang, X.; Marzano, S.-Y.L. Comparing Gut Microbiome in Mothers' Own Breast Milk- and Formula-Fed Moderate-Late Preterm Infants. Front. Microbiol. 2020, 11, 891. [CrossRef]

64. Ilcol, Y.O.; Ozbek, R.; Hamurtekin, E.; Ulus, I.H. Choline status in newborns, infants, children, breast-feeding women, breast-fed infants and human breast milk. J. Nutr. Biochem. 2005, 16, 489–499. [CrossRef] [PubMed]

65. Louis, P.; Flint, H.J. Formation of propionate and butyrate by the human colonic microbiota. Environ. Microbiol. 2017, 19, 29–41. [CrossRef] [PubMed]

66. Palomo-Buitrago, M.E.; Sabater-Masdeu, M.; Moreno-Navarrete, J.M.; Infantes, E.C.; Rodríguez, M.A.; Coll, C.; Ramió, L.; Palomino-Schätzlein, M.; Gutiérrez-Carcedo, P.; Pérez-Brocal, V.; et al. Glutamate interactions with obesity, insulin resistance, cognition and gut microbiota composition. Acta Diabetol. 2019, 56, 569–579. [CrossRef]

67. Salazar, N.; Dewulf, E.M.; Neyrinck, A.; Bindels, L.B.; Cani, P.D.; Mahillon, J.; de Vos, W.M.; Thissen, J.-P.; Gueimonde, M.; Reyes-Gavilan, C.D.L.; et al. Inulin-type fructans modulate intestinal Bifidobacterium species populations and decrease fecal short-chain fatty acids in obese women. Clin. Nutr. 2015, 34, 501–507. [CrossRef]

Manuscript 2

Gut microbiome and small RNA integrative-omic perspective of meconium and milk-fed infant stool samples

Polina Kazakova, Nerea Abasolo, Sara Martínez de Cripán, Emili Marquès Adrià Cereto-Massagué, Lorena García, Núria Canela, Ramon Tormo, Helena Torrell.

Submited to: International Journal of Molecular Sciences

Impact factor: 6.208; Q1, Biochemistry & Molecular Biology

Gut microbiome and small RNA integrative-omic perspective of meconium and milk-FED infant stool samples.

Polina Kazakova[a], Nerea Abasolo[a], Sara Martínez de Cripán[a], Emili Marquès[b] Adrià Cereto-Massagué[a], Lorena García[a], Núria Canela[a], Ramon Tormo[c], Helena Torrell[a]

[a]Eurecat, Centre Tecnològic de Catalunya, Centre for Omic Sciences (COS), Joint Unit Universitat Rovira i Virgili-EURECAT, Unique Scientific and Technical Infrastructures (ICTS), Reus, Spain
[b]Clínica Terres de l'Ebre, Tortosa, Spain.
[c]ESPGHAN, European Society for Paedriatic Gastroenterology, Hepatology and Nutrition, Switzerland; and Gastroenterology and Nutrition Pediatric Center, Barcelona, Spain.

Correspondence: helena.torrell@eurecat.org

Keywords: bacteriome; virome; smallRNome; gut microbiota; meconium; holobiont; multiomics; metagenomics; breast-fed; formula-fed

ABSTRACT The human gut microbiome plays an important role in health, and its initial development is condi-tioned by many factors, such as feeding. It has also been claimed that this colonization is guided by bacterial populations, the dynamic virome, and transkingdom interactions between host and mi-crobial cells, partially mediated by epigenetic signaling. In this article, we characterized the bacte-riome, virome, and smallRNome and their interaction in the meconium and stool samples from infants. Bacterial and viral DNA and RNA were extracted from the meconium and stool samples of 2- to 4-month-old milk-fed infants. The bacteriome, DNA and RNA virome, and smallRNome were assessed using 16S rRNA V4 sequencing, viral enrichment sequencing, and small RNA sequencing protocols, respectively. Data pathway analysis and integration were performed using the R package mixOmics. Our findings showed that the bacteriome differed among the three groups, while the virome and smallRNome presented significant differences, mainly between the meconium and stool of milk-fed infants. The gut environment is rapidly acquired after birth, and it is highly adaptable due to the interaction of environmental factors. Additionally, transkingdom interactions between vi-ruses and bacteria can influence host and smallRNome profiles. However, virome characterization has several protocol limitations that must be considered.

1.INTRODUCTION.

The human gut harbors more than 100 trillion different microorganisms (mainly bacteria but also viruses, protozoa, archaea, and fungi) that are in a symbiotic relationship with the host. The gut microbiota plays a crucial role in human health, and there is crosstalk between the intestinal microbiota and immune development, metabolism, neurogenesis, gastrointestinal integrity, and many other systems across the lifespan, beginning during fetal development [1] [2]. Thus, the role of the microbiota in numerous intestinal

and extraintestinal diseases has become steadily apparent [3,4], and more research is now being performed on how alterations in the early gut microbiota could influence child and adult health [5].

The acquisition and development of the gut microbiota in infancy is generally believed to begin at birth, with the mode of delivery being the first contact with microbia (vaginal population for vaginal delivery and skin population for cesarean delivery) [6]. Although the womb has long been assumed to be sterile, some evidence of nonpathogenic bacteria in the placenta, amniotic fluid, and fetal gut has questioned this concept [7,8]. Regardless of this debate's outcome, microbes from maternal and environmental sources rapidly and densely colonize the neonate at birth. The development and maturation of the gut microbiota are dynamic and nonrandom processes where positive and negative interactions between significant bacterial taxa start within a few hours after delivery [9–11]. Following birth, another factor that determines early colonization is the type of milk feeding: breast-fed or formula-fed. It has been suggested that breastfeeding can have a protective effect on illness development due to its repercussions on gut microbiota composition, for example, in celiac disease [12]. The mother's age, length of gestation (i.e., full-term versus preterm), smoking habits and body mass index are also important factors to be considered. After the first 6 months of life, the gastrointestinal tract slowly acquires a more complex bacterial community, which substantially increases its diversity when solid food is introduced. Microbiota diversity expands during the first two years of life. At this moment, the community converges toward an adult-like state, and it remains stable throughout the years until old age, the moment when microbiota diversity starts decreasing [13].

Although the microbiota is mainly composed of bacteria and viruses, most published microbiome studies have focused on the bacterial community, as

they clearly dominate the microbiome, and techniques used for bacterial community studies are more developed and standardized than those for the viral population (namely, the virome) [14]. However, the gut virome is highly correlated with the intestinal bacterial population, and its main composition corresponds to the Caudovirales order [15,16], formed mostly by bacteriophages. Bacteriophages are viruses that infect bacteria and thus have various mechanisms that could control the density, diversity, and network interactions inside gut-associated bacterial communities, such as lysogeny and gene transfer [17,18]. The gut virome is also acquired from an early age, and similar to the bacteriome, many factors influence its shaping (e.g., type of feeding) [19].

Although the external environment plays an important role in shaping the gut microbiome community, the host itself can modulate the microbial ecosystem through different mechanisms, such as epigenetic factors, including small RNAs.

Small RNAs consist of different types of regulatory small transcripts, with microRNAs (miRNAs) being the best known and most studied class, and they have been described as key regulators in multiple cellular functions. miRNAs are a family of very stable small noncoding RNAs containing approximately 20 nucleotides that regulate gene expression [20], and changes in their expression and function can be associated with numerous diseases. Recent studies have suggested bidirectional interactions between host cells and gut microbiota via miRNAs that participate in shaping the gut microbiota after they are secreted from intestinal epithelial cells. Likewise, host miRNA expression can be influenced by the microbiota through microbe-derived metabolites that might potentially influence the host physiology [21–24].

To gain a better understanding of how the gut microbial composition is shaped in early life stages, we conducted a fecal holo-omic study in newborn infants (meconium samples were obtained) and in exclusively milk-fed

infants either with human milk (breast-fed group) or formula milk (formula-fed group). This study included the analysis of fecal bacterial and viral populations as well as the identification of host small RNA signaling in feces.

2. RESULTS.

2.1. Bacteriome analysis

The microbiota population at the phylum level showed differences among the three groups. Proteobacteria was the predominant phylum in meconium samples (41.6%) and in the formula-fed group (46.4%), mainly due to the Enterobacteriaceae family, which was less prevalent in the breast-fed group (27.8%). The distribution of the Firmicutes phylum was similar between the three experimental groups (30-33%), and the fundamental difference in the meconium group compared to the two other groups was the higher presence of the Bacteroidetes and Actinobacteria phyla. However, none of these differences were statistically significant. In contrast, the Verrucomicrobiota phylum showed that the Akkermansia genus was significantly different in the formula-fed group, whose abundance was lower in comparison to the other two groups ($p<0.05$).

Nevertheless, α diversity measured using the Shannon index remained unchanged between groups (Figure 1B). However, when β diversity (between-sample diversity) was measured by unweighted UniFrac and compared using ANOSIM and PERMANOVA analysis, both tests showed significant segregation of the bacterial composition across different categories (Figure 1A), with the breast-fed group showing the most differentiated cluster (ANOSIM $R = 0.07$, $p = 0.03$; PERMANOVA pseudo-$F = 2.11$, $p = 0.02$).

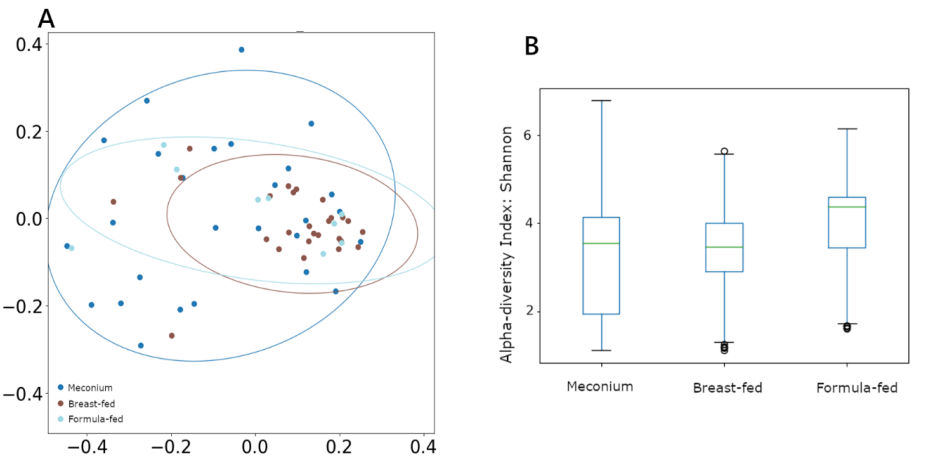


Figure 1: A) β-diversity measured using unweighted UniFrac distances. B) α-diversity measured using Shannon, with the Y-axis representing the Shannon score values.

Afterwards, split statistical analysis was performed to determine the differences between meconium and stool samples from milk-fed infants to assess the effect of the feeding type.

2.1.1. Taxonomic Differences in the Gut Bacteriome of Meconium, Formula-fed, and Breast-fed Infants

The statistical comparisons of the taxa among the meconium group versus the other two groups showed that the *Prevotellaceae* family ($p<0.001$), specifically *Prevotella* ($p<0.001$), *Paraprevotella* ($p<0.001$), and *Alloprevotella* ($p<0.01$) genera, were higher in the meconium group. There was also a higher presence of *Treponema* ($p<0.01$) and Rikenellaceae ($p<0.001$) genera in this group. On the other hand, *Enterococcus* ($p<0.01$), *Epulopiscium* ($p<0.01$) and *Lactobacillus* ($p<0.05$) genera were lower in the meconium group compared to the other two.

Separately, the *Enterobacteri*aceae family, represented by *Citrobacter* ($p<0.01$), *Enterobacter* ($p<0.01$) and *Klebsiella* ($p<0.05$) genera, was less

abundant in the meconium samples compared to the formula-fed group, in contrast to *Akkermansia* ($p<0.05$) and *Muribacullaceae* ($p<0.05$) genera, which are more abundant in the meconium samples.

When comparing the meconium and breast-fed groups, the *Veillonella* genera ($p<0.05$) had a lower presence in meconium samples.

2.1.2. Influence of fed type on microbiota.

When comparing the two types of feeding, the formula-fed group presented higher levels of *Prevotellaceae* and *Enterobacteriaceae* families and a lower abundance of *Bacteriodaceae* (Figure 2), and these differences were maintained at the genus level, as already mentioned above. Additionally, at the genus level, the formula-fed group showed higher levels of *Micrococcus*, *Dorea*, and *Lactococcus* ($p<0.05$). In contrast, the *Lactobacillae* family was more abundant in the breast-fed group ($p<0.05$), while the genera *Parasutterella, Butyricimonas*, *Desulfovibrio* and *Acinetobacter* ($p<0.05$) were more abundant.

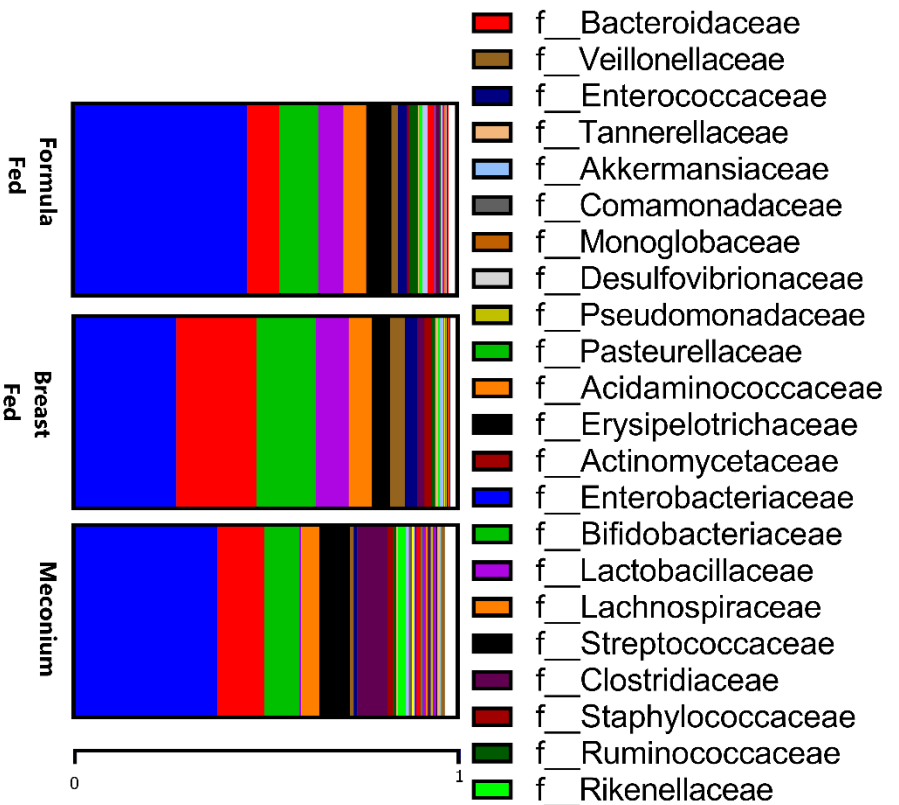


Figure 2: Relative abundances of taxa at the family level in the three studied groups

2.2. Virome analysis

Since viral enrichment was performed from stool samples, DNA and RNA virus analyses were performed separately and interrogated at the species level. After sequencing, only a small percentage of reads could be assigned to viral taxonomy (19.14% for DNA viruses and 1.86% for RNA viruses).

2.2.1. Differences in DNA virome, but unchanged RNA virome.

The sequences obtained from the DNA virome were assigned to 495 different species, predominantly of the order Caudovirales (phages). For RNA viruses, 54 different species were identified: 53.7% of species belonged to bacteriophages, 31.5% to animal viruses, 7.4% to cloning vectors, 3.7%

to phytoviruses, and the remaining 3.7% were crAssphages or undetermined.

There was no difference between α-diversity (measured using the Shannon index) values among groups in either DNA viruses or RNA viruses. However, when β-diversity was measured by Bray–Curtis and compared using ANOSIM and PERMANOVA analysis, both tests showed significant segregation of the DNA viral composition across different categories (Figure 3A), with the breast-fed group showing the least differentiated cluster (ANOSIM R = 0.29, p = 0.01; PERMANOVA pseudo-F = 2.47, p = 0.01). Any difference was observed when comparing RNA viral populations (Figure 3B).

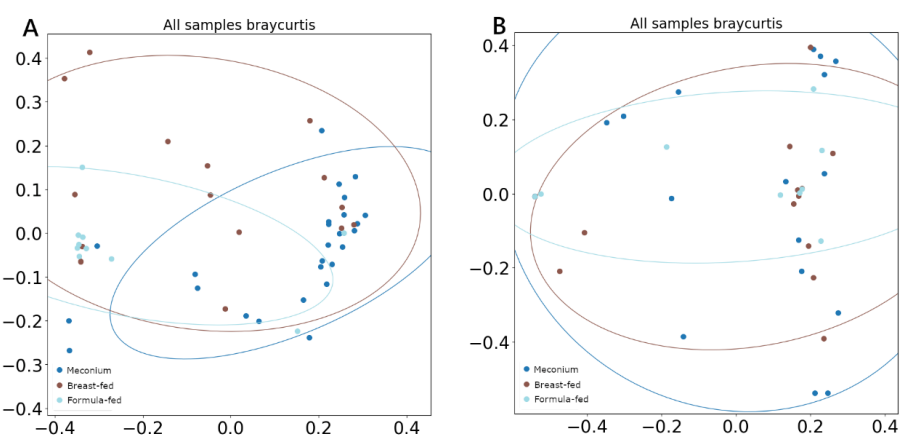


Figure 3: PCoA A) β diversity of DNA viruses measured with Bray–Curtis, B) β diversity RNA viruses measured with Bray–Curtis

2.2.2. Presence of exclusive bacteriophages in milk-fed groups.

Acinetobacter phages, *Bifidobacterium* phages, *Lactococcus* phages, *Leuconostoc* phages, *Mycobacterium* phages, *Stenotrophomonas* phages and *Yersinia* phages were detected only in samples from milk-fed infants but not in meconium samples ($p<0.05$). Additionally, *Aeromonas* phages and *Bacteroidetes* phages were detected only in the breast-fed group, while

Klebsiella and *Lactobacillus* phages were found only in the formula-fed group.

2.2.3. Virome differences among the experimental groups.

The abundance of more than half of *Enterococcus* phages, *Pseudomonas* phages and *Streptococcus* phages was significantly higher in breast-fed samples ($p<0.05$) and detected only in a very low proportion of the meconium samples. Additionally, the breast-fed group showed a higher abundance ($p<0.05$) of *Escherichia* phage than the formula-fed group. In contrast, eleven of thirteen species of CrAss phages and *Staphylococcus* phages SauM Remus were predominant in meconium samples ($p<0.05$).

Finally, meconium samples presented a higher abundance ($p<0.05$) of two RNA viruses (*Shamonda* virus and *Oxbow* virus) than the other two groups.

2.3. Transkingdom correlation

Several bacterial families correlate with phages and DNA/RNA viral species. *Pseudomonadaceae* and *Veillonellaceae* families strongly correlated with *Burkholderia* phage phiE094 (0.56 and 0.62, respectively). This makes sense in *Pseudomonadaceae* since phage phiE094 is a lytic phage mainly hosting species from the Proteobacteria phylum but not in *Veillonellaceae* (Firmicutes phylum). As expected, another positive correlation was observed between a group of *Bacteroidetes* phages and the Bacteroidetes phylum (0.39) (Figure 4). In contrast, the *Serratia* (-0.27), *Stenotrophomonas* (-0.32), *Aeromonas* (-0.31) and *Citrobacter* (-0.30) phages were negatively correlated with the Bacteroidetes phylum. Additionally, as shown in Figure 5, the Bacteroidetes phylum was positively correlated with the *Shamonda* virus (0.47) and *Oxbow* virus (0.50) and negatively correlated with *Enterobacteria* phage (-0.35). Another association found was the positive correlation between *Gammaherpesvirus* (0.3, human virus) and the Deinococcota phylum.

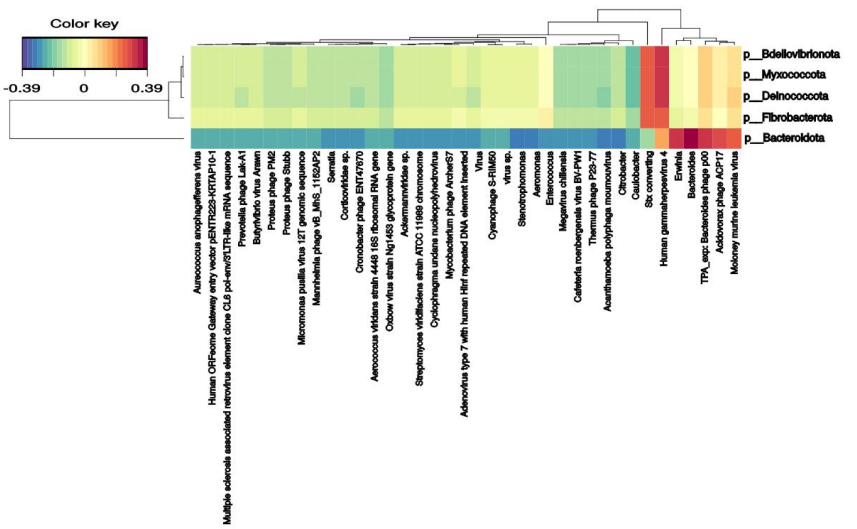


Figure 4: Clustered Image Map from the sparse Partial Least Squares (sPLS) on the DNA virus and phages (on the X-axes) and Bacterial phyla (on the Y-axis). It shows the relationship between the variables of each omic dataset. A ±0.25 threshold was used.

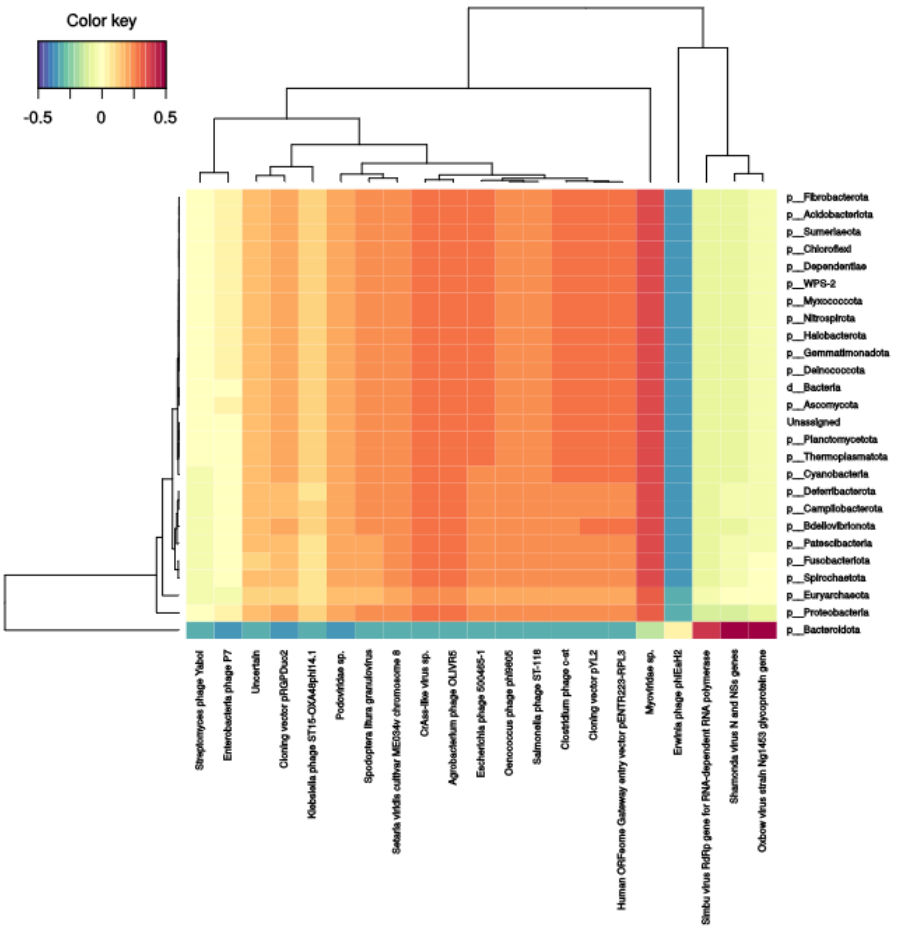


Figure 5: Clustered Image Map from the sparse Partial Least Squares (sPLS) on the RNA virus (on the X-axes) and Bacterial phyla (on the Y-axis). It shows the relationship between the variables of each omic dataset. A ±0.3 threshold was used.

2.4. SmallRNome analysis

Starting from an average of 20.4 million single-end reads per sample, an average of 0.26% of the reads were assigned to human small RNA annotations. Collectively, 1918 hsa-miRNAs and 1514 hsa-sncRNAs were assigned with at least one read. For the differential expression analysis, both

sets were combined, and only genes with 3 counts in at least 8 samples were considered, resulting in 227 human small RNAs.

In human annotation, in the meconium group, 13% of the assigned reads were for miRNAs, and the remaining 87% were assigned to sncRNA. In the breast-fed group, the percentages were 28% for miRNAs and 72% for sncRNA, and in the formula-fed group, the percentages were 17% for miRNAs and 83% sncRNA. The remaining reads not aligned to hsa-miRNAs and hsa-sncRNAs were further mapped against the human genome to identify those derived from human RNAs. However, 91.12% of the input reads remained unaligned. Then, these reads were mapped to bacterial, archaea and virus genomes, receiving the highest percentage of reads with bacteria (on average, 99% of the assigned reads), followed by virus with 0.73% and less than 0.01% of the aligned reads associated with archaea.

2.4.1. SmallRNA expression differences between meconium and milk-fed sample.

Several significant differences were found in small RNA expression when comparing meconium samples with the other two experimental groups (Figure 6). Specifically, ten small RNAs were upregulated, and eight were downregulated in meconium samples compared with formula-fed samples. Similarly, upon comparing meconium with breast-fed samples, twenty small RNAs were upregulated, and ten were downregulated in the meconium group (considering a threshold of 1.5). See supplementary online material for further details (Tables S1, S2 and S3).

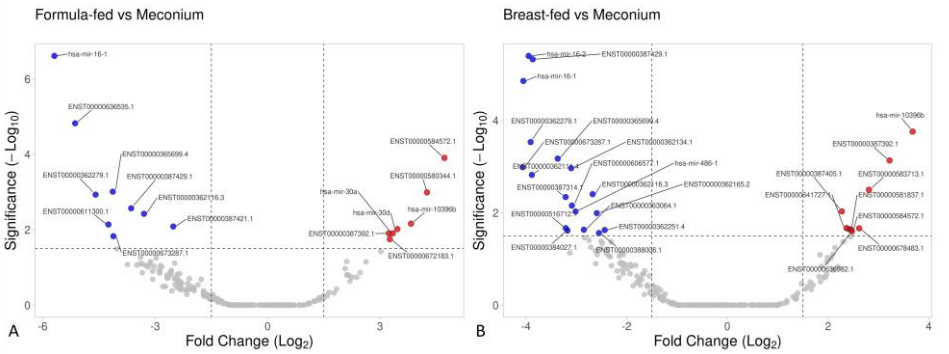


Figure 6: Volcano plot based on small RNA data. A) The meconium group versus the breast-fed group and B) the meconium group versus the formula-fed group

2.4.2. Interference of meconium miRNA in metabolic pathways.

For a deeper functional analysis, the target genes for 17 differentially expressed miRNAs were predicted using several online algorithms (described in the methods section) and mapped to the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways using the KEGG Mapper. False discovery rate (FDR) correction was calculated. As presented in Table 1, six possible pathways were listed with a $p < 0.001$, including the signaling pathways of transforming growth factor β (TGF-β), fatty acid metabolism and adherens junction, among others.

Table 1. Significantly targeted pathways

KEGG pathway	FDR adjustment	targeted genes	number of involved miRNAs
Fatty acid biosynthesis	< 0.001	7	6
Cell cycle	< 0.001	87	6
Fatty acid metabolism	< 0.001	31	7
Adherens junction	< 0.001	55	9
Lysine degradation	< 0.001	33	9

TGF-beta signaling pathway	< 0.001	55	7

2.4.3. Main difference in mascRNA among milk-fed infants.

A single significant difference was found between the breast-fed and formula-fed groups: MALAT1-associated small cytoplasmic RNA (ENST00000611300.1) was downregulated in the formula-fed group compared to that in the breast-fed group (considering a threshold of 1.5).

2.4.4. Correlation between smallRNA profile and bacterial population.

To investigate the relationships between small RNAs and the intestinal microbiome, all small RNA expression levels and microbial family abundance were analyzed together. As shown in Figure 7, the expression of several small RNAs correlated (positively and negatively) with bacterial families. For instance, there were strong positive correlations between MT-IC-201 and five bacterial families, Bacteroidia (0.73) and *Peptococcaceae* (0.73), *Spirochaetaceae* (0.60), *Erwiniaceae* (0.62) and *Muribaculaceae* (0.64). mir 4792-201 had a positive correlation with five different families: *Clostridiales* (0.65), *Clostridia* (0.63), *Spirochaetaceae* (0.68), Bacteroidia (0.66) and *Peptococcaceae* (0.60). On the other hand, the Bacteroidales order had the largest number of correlations; in addition to the two previously mentioned correlations, it had six more positive correlations (one mitochondrial RNA (0.62), one SnoRNA (0.60) and four miRNAs (0.6-0.66)) and five negative correlations (mir-103b-1 (-0.58), mir-378a (-0.56), mir-101-1 (-0.56), has-let-7a-2 (-0.58) and mir-103a-2 (-0.57)). The other family that had many correlations was the *Peptococcaceae* family, which presented five positive (two mitochondrial RNAs (0.60) and three miRNAs (0.6-0.64)) and five negative (mir-103b-1 (-0.56), mir-378i (-0.55), mir-101-1 (-0.56), hsa-let-7a-2 (-0.58) and mir-103a-2 (-0.57)) correlations.

Finally, the *Muribaculaceae* family had one positive correlation with mitochondrial RNA (0.63).

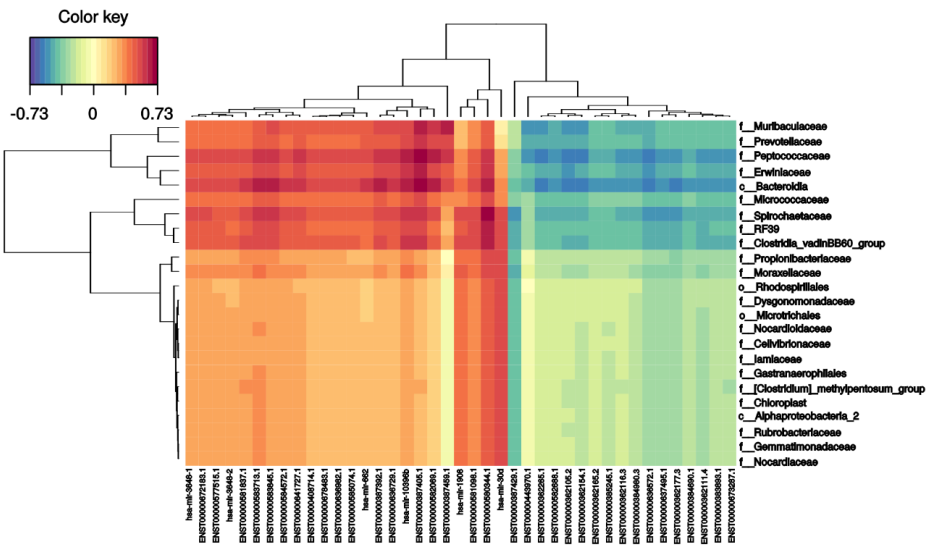


Figure 7: Sparse partial least squares (sPLS). The following graph indicates the correlation between the variables of each dataset: X-axis with sncRNAs, and the Y-axis with the bacteriome highest taxonomic level that can be classified (p_phylum, c_class, o_order, f_family). The ±0.5 threshold was used for generation of the figure.

3. DISCUSSION.

The gut microbiota, which includes bacteria, archaea, fungi and viruses, plays an important role in human health. The development of an adult intestinal microbiota begins with the primary colonization of the infant gut, the composition of which may be affected by several early-life factors, such as birth mode or feeding type [25]. In the present study, we investigated the bacteriome and virome composition of neonates at the moment of delivery (meconium), and, we characterized the bacterial and viral populations

depending on the feeding type (breast/formula-milk) in 2- to 4-month-old infants. As in many other studies, bacteria were detected in all meconium samples [26–28], contrasting the past hypothesis that considered meconium sterile [29]. Metagenomic analysis of meconium samples showed large interindividual differences and low species diversity, with Proteobacteria in the highest proportion (42%), followed by Firmicutes (30%), Bacteroidetes (18%), Actinobacteria (10%), Verrucomicrobia (0.4%) and Desulfobacterota (0.2%). These results are consistent with previous studies showing a high abundance of Proteobacteria and a lower abundance of Bacteroidetes [30,31]. The Proteobacteria predomination in meconium samples is explained by its similarity with the bacterial communities found in the mother's placenta, regardless of the method of delivery, and different from those found in the maternal vagina, according to previous studies [32]. In our study, the most abundant genera in meconium samples were *Escherichia-Shigella*, *Bacteroides*, *Bifidobacterium, Streptococcus, Clostridia, Staphylococcus, and Enterococcus*, with the last two previously reported as highly abundant [31,33]. As we expected, meconium samples presented statistically significant differences in the microbial population compared to the 2-4-month-old infant groups. *Muribaculaceae* ($p<0.001$), *Prevotella* ($p<0.01$) and the *Propionibacteriaceae* family ($p<0.05$) were higher in the meconium group than in the infants' stool from later developmental stages. Despite these differences, the bacterial composition of the meconium group was more similar to that of the formula-fed group than to that of the breast-fed group, suggesting that maternal milk has a higher modulating effect than formula, as published previously [34]. It is worth noting that the virome has not been previously fully explored. In our study, the meconium group presented a high abundance of crAssphages compared to the other two lactating groups, agreeing with the published evidence that crAssphage abundance increases with age [35] and supporting the hypothesis of vertical transmission during delivery from the

mother. In contrast, the eukaryotic viruses in the meconium group were in low abundance, and many species were not even present in comparison with the two milk-fed groups. Our results showed that viral species richness presented low values at month 0 (meconium) but higher richness after four months and reported that the vast majority of taxonomic classifications were phage families, according to a previous study [36].

Regarding the type of feeding, the infant intestinal microbiota was predominantly represented by microorganisms from the Proteobacteria phylum (46% in formula-fed and 28% in breast-fed) and Firmicutes phylum (33% in both groups), followed by those from the Bacteroidetes phylum (11% in the formula-fed group and 23% in the breast-fed group) and Actinobacteria phylum (11% in the formula-fed group and 16% in the breast-fed group). Alpha diversity was higher in the formula-fed group, and beta diversity clearly differentiated the groups into two clusters according to previous studies [33]. In fact, the formula-fed group presented higher levels of *Prevotellaceae* and *Enterobacteriaceae* families and a lower abundance of *Bacteriodaceae*, partially in agreement with a previous study that showed higher levels of *Enterobacteriaceae,* but no changes in *Bacteriodaceae* [37]. Focusing on the virome analysis, there were few significant differences in DNA viruses according to the type of feeding: only *Shigella* phage SfIV, *Burkholderia* phages, and *Streptomyces* phages presented greater relative abundances in the formula-fed group. Additionally, *Escherichia* phage JLK-2012 and *Microviridae* sp., both RNA viruses, were only present in the breast-fed group. These results contrasted with previous studies [18] where viruses infecting human cells were found only in formula-fed infants. Additionally, a previous study [35] showed that after four months of life, human cell viruses were more prominent, including *Adenoviridae*, *Anelloviridae*, *Caliciviridae*, and *Picornaviridae*, but we did not

find any of these species; this result can be explained by the size of these viruses that might have remained in the filter during the enrichment steps. Nevertheless, human pathogens (eukaryotic viruses) are comparatively well documented but are outside of infections, and their abundance is low in a healthy human gut. In contrast, phages are a natural component of every environment's intricate microbiome, depending on if it is a place where they can live freely, such as the human gut. There is evidence that phages and bacteria coexist and evolve together, but their interaction in the gut environment is poorly described [38]. To analyze this behavior, we correlated both profiles (bacteriome and virome) for the whole cohort. Notably, most of the positive correlations occurred between a bacteriophage and its host, i.e., the greater the presence of a bacteriophage, the greater the presence of its bacterial host. This agrees with the hypothesis that bacteria–phage interactions work as a network in which cross-infective phages invade other bacteria, in addition to their putative bacterial hosts, to obtain a dynamic equilibrium with all microbial communities of the gut microbiome to regulate gut homeostasis [39].

It is important to note that the current knowledge of the virome is very limited, and most previous virome profiling studies have focused on the DNA virome [40]. Here, we carried out two separate protocols for the DNA and RNA viromes, and as expected, we mainly identified both DNA and RNA bacteriophages [40]. However, as in previous studies, several protocol limitations can be identified [41]. First, the experimental enrichment method introduces some bias; for example, large viruses, such as herpesviruses, may be retained in microfiltration. Second, viruses have a small genome, and their proportion in comparison with bacteria and the host genome is very small. Third, two different protocols were needed for the study of RNA and DNA viruses. Virus taxonomy classification is an additional challenge. Whereas viruses are traditionally classified according to their morphology, their classification based on genomic sequences is

more complex. Global viral diversity has not yet been characterized, and many viruses do not have reference sequences in databases. Due to this shortcoming, 90% of the virome sequenced reads that we obtained did not share homology with any reference database, a situation also reflected in other published studies [42,43].

While many studies have focused on how crucial a balanced microbiota is for homeostasis, epigenetic signaling controlling microbiota evolution has received less attention. miRNAs have recently been described to interact with the gut microbiota in a reciprocal manner and affect the host's health status. In this study, we characterized the fecal and, importantly, meconium small RNAs for the first time to determine differences between infant milk feeding and the impact on the intestinal holobiont. To add more evidence to this statement, we integrated bacteriome data with smallRNome to gain a better understanding of the overall regulation. We observed several miRNAs that were expressed significantly differently in meconium compared with both milk-fed group samples. For example, mir-30d and mir-30a were upregulated in the meconium group in comparison with the formula-fed group. These two miRNAs belong to the miR-30 family and play a crucial regulatory role in the development of tissues and organs, and in the pathogenesis of clinical diseases [44]. These miRNAs are, for the first time, linked with infant microbiota or type of feeding. Because these miRNAs come from the mother, it is not surprising to find that newborns may be more protected against developmental problems [45] than milk-fed infants. On the other hand, MALAT1-associated small cytoplasmic RNA was found to be upregulated in breast-fed samples compared to formula-fed samples. This small RNA has been described to regulate TLR-induced proinflammatory and antiviral responses, suggesting its participation in the immune response in the early stages of life [46]. Thus, breast-fed infants show an improved and well-prepared response to inflammation and viral

infection. Moreover, mir-24 and mir-29 were downregulated in the meconium group in comparison to in the breast-fed group, and high expression of both miRNAs in maternal plasma was associated with a high risk of preeclampsia [47]. Moreover, mir-21 was downregulated in the meconium group compared to the breast-fed group. Low expression of this miRNA in the placenta was associated with intrauterine growth restriction [48] or with maternal cigarette smoking during pregnancy [59]. Additionally, some metabolic pathways were predicted to be affected by some of these miRNAs that were expressed significantly differently in meconium samples. One of the detected pathways was TGF-β signaling, which regulates many aspects of physiological embryogenesis and adult tissue homeostasis [50], and was associated with infant birth weight [51]. Another metabolic pathway was related to adherens junctions, which may play a crucial role in regulating the intestinal barrier. In addition, we report strong correlations between the *Clostridia*, *Spirochaetaceae*, *Erwiniaceae*, *Peptococcaceae*, RF39, and *Bacteroidia* families and certain miRNAs, such as mir4792-201. This miRNA has been previously described to target the FOXC1 gene (Forkhead Box C1), which plays a role in the oxidative stress response, suggesting that these bacterial taxa are involved in this metabolic process [52]. In addition, the Bacteroidia order had a strong correlation with two mitochondrial genes (ENST00000636729.1 and ENST00000387392.1) and two miRNAs (mir4472-2, mir-10396b), and *Peptoccocaceae* had a negative correlation with mir101-1, which has been described to suppress different virus replications by targeting different genes [53,54]. These findings suggest that the smallRNome mediates host regulation of the intestinal microbiota from early development, which is an effective strategy for establishing a structured and dynamic holobiont. However, for complete understanding, the complex network of interactions between miRNAs and their targets that also depend on the cell type, location, and tissue condition must be considered [55].

Finally, we can surmise that our analysis of the intestinal virome uncovered substantial variation and associations with the corresponding bacteriome and several factors, such as the smallRNome. These results provide the basis for a better understanding of microbial ecology and its relationship with the host.

4. MATERIAL AND METHODS.

4.1. Study design and sample collection

The study included a total of 64 vaginally delivered healthy full-term infants; 27 of them were newborns, and meconium samples were obtained within 0-48 h after birth; the remaining 37 were infants under four months old (3.2 ± 0.7) and were divided into 25 breast-fed and 12 formula-fed infants. To ensure the accuracy of our results, we excluded infants whose mothers had illnesses or were on medication during pregnancy, as well as those who had undergone restricted diets. Newborns with pathologies were also excluded, as were infants whose birth weight fell below the 25th or above the 75th Spanish percentile. The breast-fed group was formed by infants fed breast milk exclusively for 2-4 months after birth. The formula-fed group was formed by infants fed formula exclusively for 2-4 months, whose mothers voluntarily chose to feed their babies formula. The formula milk used in this study was compliant with the Commission Delegated Regulation (EU) 2016/127 regarding composition. While different commercially available formula milks were used, they all met the same nutritional requirements. For all cases, stool samples were collected and immediately frozen at -20 °C. Ethical approval was granted on June 28, 2018. Informed consent was obtained from the parents of the eligible infants. The study is in accordance with the ethical standards of the Declaration of Helsinki.

4.2. Bacteriome analysis

Bacterial DNA was extracted from approximately 200 mg of meconium or fecal sample using a QIAamp® DNA Stool Mini Kit (Qiagen Inc., Hilden, Germany) according to the manufacturer's instructions. DNA quantity and purity were assessed using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA).

The V4 variable region of the 16S rRNA gene was amplified by PCR as described previously [56].

The amplicon libraries were pooled and diluted to 35 pM before clonal amplification. The Ion 510 and 520 and 530 Ion Chef Kit (Life Technologies, Carlsbad, CA, USA) was employed for template preparation. Next-generation sequencing of the clonally amplified 16S rRNA libraries was performed on an Ion GeneStudio S5 system (Life Technologies, Carlsbad, CA, USA) following the manufacturer's instructions. The generated reads were quality filtered, analyzed with QIIME2 (2022.2), and passed for classification into amplicon sequence variants (ASVs) to DADA2, using only reads of at least 200 bp and truncating at that length. These ASVs were taxonomically classified with VSEARCH against the Silva database at 99% homology [56]. Alpha diversity was calculated as Shannon, Chao1, and Faith indices, and beta diversity was calculated with Bray-Curtis, Aitchison, Jaccard, and UniFrac (weighted and unweighted) distances. PERMANOVA and ANOSIM analyses were performed on the beta diversity data. The taxonomic abundance for each taxon at every level was compared using a Kruskal-Wallis test and corrected by Holm-Sidak correction. Statistical power analysis for the comparisons was calculated according to Equations D and E from Ferdous et al. [57]. The results are shown in Supplementary Table S4.

4.3. Virome analysis

Each sample was weighed and resuspended to a final concentration of 10% (w/v) in autoclaved phosphate-buffer saline (PBS) buffer (Thermo Fisher

Scientific, Waltham, MA, USA) and vigorously vortexed until reaching a completely homogeneous suspension. The suspension was centrifuged at 4800 × *g* for 10 min at 4 °C to remove/clarify large particles that may be present in the samples, such as organic matter or host cells. The supernatant was collected and filtered through a 0.22 µm filter Steritop (Millipore Sigma, Hayward, CA, USA) to retain the bacterial cells. The filtrate was ultracentrifuged (14 mL, Polypropylene Tube, 14 × 95 mm—50 Pk, and SW 40 Ti Swinging-Bucket Rotor Package) at 900,000× *g* for 90 min at 4 °C to reduce the liquid volume and concentrate VLPs. The pellet was resuspended in 199 µL of enzyme buffer and treated with 25 U of Benzonase® Nuclease (Millipore Sigma, Hayward, CA, USA) at 37 °C for 90 min to digest non-particle-protected nucleic acids. Subsequently, viral nucleic acids were extracted by a Quick-DNA/RNA Viral Kit (Zymo Research, Irvine, CA, USA) according to the manufacturer's instructions and eluted into 36 µL of RNase-free water. The DNA and RNA purity were evaluated by the A260/A280 ratio using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). Nucleic acids were divided into two aliquots, one of which was treated with 1 U of DNase I (Invitrogen™, Waltham, MA, USA) at 37 °C for 90 min to obtain pure RNA, and the other aliquot was treated with 10 U of RNase ONE™ ribonuclease (Promega, Madison, WI, USA) at 37 °C for 30 min to obtain pure DNA.

Afterwards, DNA libraries were generated by the Ion Xpress™ Plus Fragment Library Kit for the AB Library Builder™ System (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's instructions, and libraries were reamplified and purified manually after their creation. RNA libraries were created by a Total RNA-Seq Kit v2 (Thermo Fisher Scientific, Waltham, MA, USA) for whole transcriptome libraries according to the manufacturer's instructions. DNA and RNA libraries were quantified by electrophoresis at TapeStation using High-sensitivity DNA ScreenTape Analysis (Agilent Technologies, Santa Clara, CA, USA). DNA and RNA

template libraries were performed using the Ion Chef System (Thermo Fisher Scientific, Waltham, MA, USA) and sequenced using the Ion GeneStudio S5 System (Thermo Fisher Scientific, Waltham, MA, USA). All steps in the Ion GeneStudio S5 System (Life Technologies, Carlsbad, CA, USA), including amplification through sequencing, were performed according to the manufacturer's recommendations.

For both DNA and RNA, each sample's reads were assembled using SPAdes genome assembler 3.15.5. The resulting contigs were classified as viral, potentially viral, or nonviral using viralVerify 1.1 (https://github.com/ablab/viralVerify, accessed on 9 September 2022).

With the Pfam-A HMM database and a BLAST against the NCBI "nt" database, contigs marked as nonviral or that matched against cellular organisms or plasmids were discarded. The abundance of each contig within each sample was derived from its coverage. The remaining contigs were assigned a taxonomy based on their best match in the BLAST results. The contigs from all samples were clustered at 90% homology to form OTUs (operational taxonomic units) comparable across samples. From these OTUs, alpha and beta diversity analyses and statistical analyses, such as those for the bacteriome, were performed, with the only difference being avoiding metrics that rely on phylogenetic distances. The results are shown in Supplementary Table S4.

4.4. Analysis

Total RNA was extracted from 250 mg of meconium or fecal samples using Direct-zol™ RNA Miniprep Plus (Zymo Research, Irvine, CA, USA) according to the manufacturer's instructions, and the RNA concentration was quantified using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). Libraries were created by a TruSeq Small RNA Library kit (Illumina, San Diego, CA, USA) according to the manufacturer's instructions, and were sequenced by Illumina NextSeq2000.

sRNA-seq pipeline analysis was performed following a previously described approach [58–60].

FastQC software v0.11.9 ((http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used for quality control (QC) of FASTQ files. Samples were preprocessed with Cutadapt (version 2.9), discarding reads shorter than 14 nt and imposing a maximum error rate equal to 0.15 for mismatches, insertions, and deletions. Trimmed reads were mapped against hsa-miRNAs from miRBase using the BWA Algorithm v. 0.7.17-r1188. Unaligned reads were aligned against hsa-sncRNA sequences shorter than 80 bp from Ensembl with BWA default parameters. The quantification of miRNA and sncRNA was performed with SAMtools and merged into a unique smallRNA count matrix. From this count matrix, differential expression analysis was performed with DESeq2.

The reads that were left unmapped were aligned with BWA against the Hg38 genome from Ensembl. The reads that were still unmapped on the human genome were then analyzed by Kraken for metatranscriptomic analysis. The statistical power of the comparisons for each miRNA was calculated according to Equations D and E from Ferdous et al. [57]. The results are shown in Supplementary Table S4.

4.5. Metabolic pathway analysis

Target genes for 14 miRNAs that were significant when comparing meconium with the other two groups were predicted using at least four of these public database algorithms online: Diana MicroT, miRanda, miRDB, PicTar and miRNAMap. Target genes predicted by at least two different tools were mapped to Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways using KEGG Mapper and enriched by Fisher's exact test (confidence interval 95%) with FDR correction using R Software version 4.2.2 (R Development Core Team, 2013, Vienna, Austria).

4.6. Omics data integration

Integration of the bacteriome, RNA and DNA virome and sRNA transcriptome was performed in R version 4.2.2 with the mixOmics package [63]. We evaluated each pair of omics data because of sample limitations. We calculated the sPLS model to identify the most discriminative features between each pair of omics data to evaluate their associations. sPLS in regression mode was applied to rlog-normalized sncRNA counts and log-ratio-transformed relative abundance for the bacteriome and virome, respectively. The models were tuned based on 10-fold cross-validation, and optimal parameters were chosen according to the highest mean correlation measure for each pair of omics.

5. CONCLUSIONS.

Despite the limited number of samples available from the formula-fed group, our results indicate that the gut environment of newborns, as assessed through three examined omic levels, is more similar to that of the formula-fed group than to the breast-fed group. This supports the idea that diet has a significant impact on gut microbiota and confirms the modulating effects of breast-feeding, likely mainly triggered by the intake of colostrum. Furthermore, our study highlights a clear difference in the virome between newborns and four-month-old infants, underscoring the dynamic nature of viral populations and their role in shaping community assembly and host health. Notably, we found trans-kingdom correlations between virome components and bacteria, suggesting additional layers of complexity in host–microbial homeostasis. While we acknowledge certain protocol limitations and analysis shortcomings, our findings point to the need for further investigation into these areas.

In addition, our study is the first to identify miRNAs in meconium samples, indicating the existence of epigenetic mechanisms before birth that putatively interact with the host intestinal system, modulating and

controlling its homeostasis. We also found that the miRNome profile, similar to the virome, rapidly changes in the first months of life. However, additional research is needed to fully understand the precise mechanisms underlying the interaction between the microbiota and miRNA.

In conclusion, our findings suggest that the gut environment is rapidly acquired after birth and highly malleable, with environmental factors and genetic responses creating complex molecular interactions at the host–microbiota interface. These interactions may play a significant role in regulating predisposition to future diseases.

REFERENCES

1. Leblanc, J.G.; Chain, F.; Martín, R.; Bermúdez-Humarán, L.G.; Courau, S.; Langella, P. Beneficial effects on host energy metabolism of short-chain fatty acids and vitamins produced by commensal and probiotic bacteria. *Microb. Cell Fact.* 2017, *16*, 79. https://doi.org/10.1186/s12934-017-0691-z.

2. Silva, Y.P.; Bernardi, A.; Frozza, R.L. The Role of Short-Chain Fatty Acids From Gut Microbiota in Gut-Brain Communication. *Front. Endocrinol. (Lausanne)* 2020, *11*, 25. https://doi.org/10.3389/fendo.2020.00025.

3. Navab-Moghadam, F.; Sedighi, M.; Khamseh, M.E.; Alaei-Shahmiri, F.; Talebi, M.; Razavi, S.; Amirmozafari, N. The association of type II diabetes with gut microbiota composition. *Microb. Pathog.* 2017, *110*, 630–636. https://doi.org/10.1016/j.micpath.2017.07.034.

4. Yang, T.; Santisteban, M.M.; Rodriguez, V.; Li, E.; Ahmari, N.; Carvajal, J.M.; Zadeh, M.; Gong, M.; Qi, Y.; Zubcevic, J.; et al. Gut Dysbiosis Is Linked to Hypertension. *Hypertension* 2015, *65*, 1331–1340. https://doi.org/10.1161/hypertensionaha.115.05315.

5. Sarkar, A.; Yoo, J.Y.; Dutra, S.V.O.; Morgan, K.H.; Groer, M. The Association between Early-Life Gut Microbiota and Long-Term Health and Diseases. *J. Clin. Med.* 2021, *10*, 459. https://doi.org/10.3390/jcm10030459.

6. Akagawa, S.; Tsuji, S.; Onuma, C.; Akagawa, Y.; Yamaguchi, T.; Yamagishi, M.; Yamanouchi, S.; Kimata, T.; Sekiya, S.-I.; Ohashi, A.; et al. Effect of Delivery Mode and Nutrition on Gut Microbiota in Neonates. *Ann. Nutr. Metab.* 2019, *74*, 132–139. https://doi.org/10.1159/000496427.

7. Rautava, S.; Kainonen, E.; Salminen, S.; Isolauri, E. Maternal probiotic supplementation during pregnancy and breast-feeding reduces the risk of eczema in the infant. *J. Allergy Clin. Immunol.* 2012, *130*, 1355–1360. https://doi.org/10.1016/j.jaci.2012.09.003.

8. Rackaityte, E.; Halkias, J.; Fukui, E.M.; Mendoza, V.F.; Hayzelden, C.; Crawford, E.D.; Fujimura, K.E.; Burt, T.D.; Lynch, S.V. Viable bacterial colonization is highly limited in the human intestine in utero. *Nat. Med.* 2020, *26*, 599–607. https://doi.org/10.1038/s41591-020-0761-3.

9. Tanaka, M.; Nakayama, J. Development of the gut microbiota in infancy and its impact on health in later life. *Allergol. Int.* 2017, *66*, 515–522. https://doi.org/10.1016/j.alit.2017.07.010.

10. Turroni, F.; Milani, C.; Duranti, S.; Lugli, G.A.; Bernasconi, S.; Margolles, A.; Di Pierro, F.; Van Sinderen, D.; Ventura, M. The infant gut microbiome as a microbial organ influencing host well-being. *Ital. J. Pediatr.* 2020, *46*, 16. https://doi.org/10.1186/s13052-020-0781-0.

11. Gars, A.; Ronczkowski, N.M.; Chassaing, B.; Castillo-Ruiz, A.; Forger, N.G. First Encounters: Effects of the Microbiota on Neonatal Brain Development. *Front. Cell. Neurosci.* 2021, *15*, 212. https://doi.org/10.3389/fncel.2021.682505.

12. Ivarsson, A.; Hernell, O.; Stenlund, H.; Persson, L.. Breast-feeding protects against celiac disease. *Am. J. Clin. Nutr.* 2002, *75*, 914–921. https://doi.org/10.1093/ajcn/75.5.914.

13. Xiao, L.; Zhao, F. Microbial transmission, colonisation and succession: From pregnancy to infancy. *Gut* 2023, *72*, 772–786. https://doi.org/10.1136/gutjnl-2022-328970.

14. Fulci, V.; Stronati, L.; Cucchiara, S.; Laudadio, I.; Carissimi, C. Emerging Roles of Gut Virome in Pediatric Diseases. Int. J. Mol. Sci. 2021, 22, 4127. https://doi.org/10.3390/ijms22084127.

15. Norman, J.M.; Handley, S.A.; Baldridge, M.T.; Droit, L.; Liu, C.Y.; Keller, B.C.; Kambal, A.; Monaco, C.L.; Zhao, G.; Fleshner, P.; et al. Disease-Specific Alterations in the Enteric Virome in Inflammatory Bowel Disease. Cell 2015, 160, 447–460. https://doi.org/10.1016/j.cell.2015.01.002.

16. Minot, S.; Bryson, A.; Chehoud, C.; Wu, G.D.; Lewis, J.D.; Bushman, F.D. Rapid evolution of the human gut virome. Proc. Natl. Acad. Sci. USA 2013, 110, 12450–12455. https://doi.org/10.1073/pnas.1300833110.

17. Carding, S.R.; Davis, N.; Hoyles, L. Review article: The human intestinal virome in health and disease. Aliment. Pharmacol. Ther. 2017, 46, 800–815. https://doi.org/10.1111/apt.14280.

18. Nishijima, S.; Nagata , N.; Kiguchi, Y.; Kojima, Y.; Miyoshi-Akiyama, T.; Kimura, M.; Ohsugi, M.; Ueki, K.; Oka, S.; Mizokami, M.; et al. Extensive gut virome variation and its associations with host and environmental factors in a population-level cohort. Nat. Commun. 2022, 13, 5252. https://doi.org/10.1038/s41467-022-32832-w.

19. Pannaraj, P.S.; Ly, M.; Cerini, C.; Saavedra, M.; Aldrovandi, G.M.; Saboory, A.A.; Johnson, K.M.; Pride, D.T. Shared and Distinct Features of Human Milk and Infant Stool Viromes. Front. Microbiol. 2018, 9, 1162. https://doi.org/10.3389/fmicb.2018.01162.

20. Esteller, M. Non-coding RNAs in human disease. Nat. Rev. Genet. 2011, 12, 861–874. https://doi.org/10.1038/nrg3074.

21. Dalmasso, G.; Nguyen, H.T.T.; Yan, Y.; Laroui, H.; Charania, M.A.; Ayyadurai, S.; Sitaraman, S.V.; Merlin, D. Microbiota Modulate Host Gene Expression via MicroRNAs. PLoS ONE 2011, 6, e19293. https://doi.org/10.1371/journal.pone.0019293.

22. Liu, S.; Weiner, H.L. Control of the Gut Microbiome by Fecal MicroRNA. Microbial Cell 2016, 3, 176–177.

23. Liu, S.; Liu, X.; Guo, J.; Yang, R.; Wang, H.; Sun, Y.; Chen, B.; Dong, R. The Association Between Microplastics and Microbiota in Placentas and Meconium: The First Evidence in Humans. Environ. Sci. Technol. 2022 Oct 21 https://doi.org/10.1021/acs.est.2c04706.

24. Liu, S.; da Cunha, A.P.; Rezende, R.M.; Cialic, R.; Wei, Z.; Bry, L.; Comstock, L.E.; Gandhi, R.; Weiner, H.L. The Host Shapes the Gut Microbiota via Fecal MicroRNA. Cell Host Microbe 2016, 19, 32–43. https://doi.org/10.1016/j.chom.2015.12.005.

25. Ma, J.; Li, Z.; Zhang, W.; Zhang, C.; Zhang, Y.; Mei, H.; Zhuo, N.; Wang, H.; Wang, L.; Wu, D. Comparison of gut microbiota in exclusively breast-fed and formula-fed babies: A study of 91 term infants. Sci. Rep. 2020, 10, 15792. https://doi.org/10.1038/s41598-020-72635-x.

26. Dominguez-Bello, M.G.; Costello, E.K.; Contreras, M.; Magris, M.; Hidalgo, G.; Fierer, N.; Knight, R. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. Proc. Natl. Acad. Sci. USA 2010, 107, 11971–11975. https://doi.org/10.1073/pnas.1002601107.

27. Gosalbes, M.J.; Durbán, A.; Pignatelli, M.; Abellan, J.J.; Jiménez-Hernández, N.; Pérez-Cobas, A.E.; Latorre, A.; Moya, A. Metatranscriptomic Approach to Analyze the Functional Human Gut Microbiota. PLoS ONE 2011, 6, e17447. https://doi.org/10.1371/journal.pone.0017447.

28. Hansen, R.; Scott, K.P.; Khan, S.; Martin, J.C.; Berry, S.H.; Stevenson, M.; Okpapi, A.; Munro, M.J.; Hold, G.L. First-Pass Meconium Samples from Healthy Term Vaginally-Delivered Neonates: An Analysis of the Microbiota. PLoS ONE 2015, 10, e0133320. https://doi.org/10.1371/journal.pone.0133320.

29. Koleva, P.T.; Bridgman, S.L.; Kozyrskyj, A.L. The Infant Gut Microbiome: Evidence for Obesity Risk and Dietary Intervention. Nutrients 2015, 7, 2237–2260. https://doi.org/10.3390/nu7042237.

30. Hu, J.; Nomura, Y.; Bashir, A.; Fernandez-Hernandez, H.; Itzkowitz, S.; Pei, Z.; Stone, J.; Loudon, H.; Peter, I. Diversified Microbiota of Meconium Is Affected by Maternal Diabetes Status. PLoS ONE 2013, 8, e78257. https://doi.org/10.1371/journal.pone.0078257.

31. Ardissone, A.N.; De La Cruz, D.M.; Davis-Richardson, A.G.; Rechcigl, K.T.; Li, N.; Drew, J.C.; Murgas-Torrazza, R.; Sharma, R.; Hudak, M.L.; Triplett, E.W.; et al. Meconium Microbiome Analysis Identifies Bacteria Correlated with Premature Birth. PLoS ONE 2014, 9, e90784. https://doi.org/10.1371/journal.pone.0090784.

32. Lim, E.S.; Zhou, Y.; Zhao, G.; Bauer, I.K.; Droit, L.; Ndao, I.M.; Warner, B.B.; Tarr, P.I.; Wang, D.; Holtz, L.R. Early life dynamics of the human gut virome and bacterial microbiome in infants. Nat. Med. 2015, 21, 1228–1234. https://doi.org/10.1038/nm.3950.

33. Martin, R.; Makino, H.; Yavuz, A.C.; Ben-Amor, K.; Roelofs, M.; Ishikawa, E.; Kubota, H.; Swinkels, S.; Sakai, T.; Oishi, K.; et al. Early-Life Events, Including Mode of Delivery and Type of Feeding, Siblings and Gender, Shape the Developing Gut Microbiota. PLoS ONE 2016, 11, e0158498. https://doi.org/10.1371/journal.pone.0158498.

34. Laursen, M.F.; Bahl, M.I.; Michaelsen, K.F.; Licht, T.R. First Foods and Gut Microbes. Front. Microbiol. 2017, 8, 356. https://doi.org/10.3389/fmicb.2017.00356.

35. Gregory, A.C.; Zablocki, O.; Zayed, A.A.; Howell, A.; Bolduc, B.; Sullivan, M.B. The Gut Virome Database Reveals Age-Dependent Patterns of Virome Diversity in the Human Gut. Cell Host Microbe 2020, 28, 724–740.e8. https://doi.org/10.1016/j.chom.2020.08.003.

36. Liang, G.; Zhao , C.; Zhang, H.; Mattei, L.; Sherrill-Mix, S.; Bittinger, K.; Kessler, L.R.; Wu, G.D.; Baldassano, R.N.; DeRusso, P.; et al. The stepwise assembly of the neonatal virome is modulated by breastfeeding. Nature 2020, 581, 470–474. https://doi.org/10.1038/s41586-020-2192-1.

37. Forbes, J.D.; Azad, M.B.; Vehling, L.; Tun, H.M.; Konya, T.B.; Guttman, D.S.; Field, C.J.; Lefebvre, D.; Sears, M.R.; Becker, A.B.; et al. Association of Exposure to Formula in the Hospital and Subsequent Infant Feeding Practices with Gut Microbiota and Risk of Overweight in the First Year of Life. JAMA Pediatr. 2018, 172, e181161. https://doi.org/10.1001/jamapediatrics.2018.1161.

38. Clokie, M.R.J.; Millard, A.D.; Letarov, A.V.; Heaphy, S. Phages in nature. Bacteriophage 2011, 1, 31–45. https://doi.org/10.4161/bact.1.1.14942.

39. Iorio, A.; Biazzo, M.; Gardini, S.; Muda, A.O.; Perno, C.F.; Dallapiccola, B.; Putignani, L. Cross-correlation of virome–bacteriome–host–metabolome to study respiratory health. Trends Microbiol. 2022, 30, 34–46. https://doi.org/10.1016/j.tim.2021.04.011.

40. Bikel, S.; Valdez-Lara, A.; Cornejo-Granados, F.; Rico, K.; Canizales-Quinteros, S.; Soberón, X.; Del Pozo-Yauner, L.; Ochoa-Leyva, A. Combining metagenomics, metatranscriptomics and viromics to explore novel microbial interactions: Towards a systems-level understanding of human microbiome. Comput. Struct. Biotechnol. J. 2015, 13, 390–401. https://doi.org/10.1016/j.csbj.2015.06.001.

41. Garmaeva, S.; Sinha, T.; Kurilshikov, A.; Fu, J.; Wijmenga, C.; Zhernakova, A. Studying the gut virome in the metagenomic era: Challenges and perspectives. BMC Biol. 2019, 17, 84. https://doi.org/10.1186/s12915-019-0704-y.

42. Dutilh, B.E.; Cassman, N.; McNair, K.; Sanchez, S.E.; Silva, G.G.Z.; Boling, L.; Barr, J.J.; Speth, D.R.; Seguritan, V.; Aziz, R.K.; et al. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. Nat. Commun. 2014, 5, 4498. https://doi.org/10.1038/ncomms5498.

43. Mirzaei, M.K.; Maurice, C.F. Ménage à trois in the human gut: Interactions between host, bacteria and phages. Nat. Rev. Microbiol. 2017, 15, 397–408. https://doi.org/10.1038/nrmicro.2017.30.

44. Zhang, X.; Dong, S.; Jia, Q.; Zhang, A.; Li, Y.; Zhu, Y.; Lv, S.; Zhang, J. The microRNA in ventricular remodeling: The miR-30 family. Biosci. Rep. 2019, 39, BSR20190788. https://doi.org/10.1042/bsr20190788.

45. Raymond, F.; Lefebvre, G.; Texari, L.; Pruvost, S.; Metairon, S.; Cottenet, G.; Zollinger, A.; Mateescu, B.; Billeaud, C.; Picaud, J.-C.; et al. Longitudinal Human Milk miRNA Composition over the First 3 mo of Lactation in a Cohort of Healthy Mothers Delivering Term Infants. J. Nutr. 2022, 152, 94–106. https://doi.org/10.1093/jn/nxab282.

46. Sun, T.; Wei, C.; Wang, D.; Wang, X.; Wang, J.; Hu, Y.; Mao, X. The small RNA mascRNA differentially regulates TLR-induced proinflammatory and antiviral responses. JCI Insight 2021, 6, e150833. https://doi.org/10.1172/jci.insight.150833.

47. Akgör, U.; Ayaz, L.; Çayan, F. Expression levels of maternal plasma microRNAs in preeclamptic pregnancies. J. Obstet. Gynaecol. 2021, 41, 910–914. https://doi.org/10.1080/01443615.2020.1820465.

48. Maccani, M.A.; Padbury, J.F.; Marsit, C.J. miR-16 and miR-21 Expression in the Placenta Is Associated with Fetal Growth. PLoS ONE 2011, 6, e21210. https://doi.org/10.1371/journal.pone.0021210.

49. Maccani, M.A.; Avissar-Whiting, M.; Banister, C.E.; McGonnigal, B.; Padbury, J.F.; Marsit, C.J. Maternal cigarette smoking during pregnancy is associated with downregulation ofmiR-16,miR-21, andmiR-146ain the placenta. Epigenetics 2010, 5, 583–589. https://doi.org/10.4161/epi.5.7.12762.

50. Tzavlaki, K.; Moustakas, A. TGF-β Signaling. Biomolecules 2020, 10, 487. https://doi.org/10.3390/biom10030487.

51. Rodil-Garcia, P.; Arellanes-Licea, E.D.C.; Montoya-Contreras, A.; Salazar-Olivo, L.A. Analysis of MicroRNA Expression in Newborns with Differential Birth Weight Using Newborn Screening Cards. Int. J. Mol. Sci. 2017, 18, 2552. https://doi.org/10.3390/ijms18122552.

52. Liu, P.; Pu, J.; Zhang, J.; Chen, Z.; Wei, K.; Shi, L. Bioinformatic analysis of miR-4792 regulates Radix Tetrastigma hemsleyani flavone to inhibit proliferation, invasion, and induce apoptosis of A549 cells. OncoTargets Ther. 2019, 12, 1401–1412. https://doi.org/10.2147/ott.s182525.

53. Wang, Y.; Tian, H. miR-101 suppresses HBV replication and expression by targeting FOXO1 in hepatoma carcinoma cell lines. Biochem. Biophys. Res. Commun. 2017, 487, 167–172. https://doi.org/10.1016/j.bbrc.2017.03.171.

54. Sharma, S.; Chatterjee, A.; Kumar, P.; Lal, S.; Kondabagil, K. Upregulation of miR-101 during Influenza A Virus Infection Abrogates Viral Life Cycle by Targeting mTOR Pathway. Viruses 2020, 12, 444. https://doi.org/10.3390/v12040444.

55. Yarani, R.; Shojaeian, A.; Palasca, O.; Doncheva, N.T.; Jensen, L.J.; Gorodkin, J.; Pociot, F. Differentially Expressed miRNAs in Ulcerative Colitis and Crohn's Disease. Front. Immunol. 2022, 13, 865777. https://doi.org/10.3389/fimmu.2022.865777.

56. Torrell, H.; Cereto-Massagué, A.; Kazakova, P.; García, L.; Palacios, H.; Canela, N. Multiomic Approach to Analyze Infant Gut Microbiota:

Experimental and Analytical Method Optimization. Biomolecules 2021, 11, 999. https://doi.org/10.3390/biom11070999.

57. Ferdous, T.; Jiang, L.; Dinu, I.; Groizeleau, J.; Kozyrskyj, A.L.; Greenwood, C.M.T.; Arrieta, M.-C. The rise to power of the microbiome: Power and sample size calculation for microbiome studies. Mucosal Immunol. 2022, 15, 1060–1070. https://doi.org/10.1038/s41385-022-00548-1.

58. Kulkarni, N.; Alessandrì, L.; Panero, R.; Arigoni, M.; Olivero, M.; Ferrero, G.; Cordero, F.; Beccuti, M.; Calogero, R.A. Reproducible bioinformatics project: A community for reproducible bioinformatics analysis pipelines. BMC Bioinform. 2018, 19, 5–13. https://doi.org/10.1186/s12859-018-2296-x.

59. Ferrero, G.; Cordero, F.; Tarallo, S.; Arigoni, M.; Riccardo, F.; Gallo, G.; Ronco, G.; Allasia, M.; Kulkarni, N.; Matullo, G.; et al. Small non-coding RNA profiling in human biofluids and surrogate tissues from healthy individuals: Description of the diverse and most represented species. Oncotarget 2017, 9, 3097–3111. https://doi.org/10.18632/oncotarget.23203.

60. Tarallo, S.; Ferrero, G.; Gallo, G.; Francavilla, A.; Clerico, G.; Realis Luc, A.; Manghi, P.; Thomas, A.M.; Vineis, P.; Segata, N.; et al. Altered Fecal Small RNA Profiles in Colorectal Cancer Reflect Gut Microbiome Composition in Stool Samples. Msystems 2019, 4, e00289-19. https://doi.org/10.1128/msystems.00289-19.

61. Rohart, F.; Gautier, B.; Singh, A.; Lê Cao, K.-A. mixOmics: An R package for 'omics feature selection and multiple data integration. PLoS Comput. Biol. 2017, 13, e1005752. https://doi.org/10.1371/journal.pcbi.1005752.

Supplementary Material

Chapter 2

To determine the alterations in the intestinal phenotype of children with celiac disease compared to healthy children and evaluate the restoring effect on this intestinal phenotype of the administration of a probiotic in an animal model of celiac disease.

UNIVERSITAT ROVIRA I VIRGILI
A MULTI-OMICS APPROACH TO DESCRIBE CROSSTALK BETWEEN GUT MICROBIOTA AND HOST: EPIGENETICS AND VIROME ROLE
Polina Kazakova Borodina

Manuscript 3

Multiomic Assessment of Gut Microbiome and Probiotic Intervention in Celiac Disease: Implications for Early Prevention and Treatment Strategies

Polina Kazakova, Héctor Palacios, Sara Martínez de Cripán, Adrià Cereto-Massagué, Nerea Abasolo, Ana Antolín, Juan Maria Alcaide-Hidalgo. Xavier Escoté, Ramón Tormo, Nuria Canela, Helena Torrell.

Under preparation

Multiomic Assessment of Gut Microbiome and Probiotic Intervention in Celiac Disease: Implications for Early Prevention and Treatment Strategies

Polina Kazakova[a], Héctor Palacios[a], Sara Martínez de Cripán[a], Adrià Cereto-Massagué[a], Nerea Abasolo[a], Ana Antolín[b], Juan Maria Alcaide-Hidalgo[b]. Xavier Escoté[b], Ramón Tormo[c], Nuria Canela[a], Helena Torrell[a].

[a] Eurecat, Centre Tecnològic de Catalunya, Centre for Omic Sciences (COS), Joint Unit Universitat Rovira i Virgili-EURECAT, Unique Scientific and Technical Infrastructures (ICTS), Reus, Spain

[b] Eurecat, Centre Tecnològic de Catalunya, Unitat de Nutrició i Salut, Reus, Spain

[c] ESPGHAN, European Society for Paedriatic Gastroenterology, Hepatology and Nutrition, Switzerland; and Gastroenterology and Nutrition Pediatric Center, Barcelona, Spain

Correspondence: helena.torrell@eurecat.org

Keywords: microbiota, multiomics, holobiont, celiac disease, probiotic treatment, dysbiosis, metabolomic

Abstract:

Background: Coeliac disease (CeD) is an autoimmune multifactorial enteropathy induced by the ingestion of gluten that affects the integrity of small intestine and results in numerous symptoms. The only effective treatment against CeD is lifelong adherence to a strict gluten-free diet (GFD), but, in some cases, this treatment does not totally restore the gut environment. Probiotics could potentially be used to restore gut dysbiosis in CeD and aid in gliadin digestion. The goal of the study is to provide the necessary knowledge to implement early preventive interventions that can re-establish gluten tolerance and reduce the risk of developing CeD.

Methods: The study compared the bacteriome, virome, and metabolome of celiac children who follow GFD with those of healthy controls to identify potential biomarkers of CeD symptoms. Additionally, the study evaluates the effect of a probiotic cocktail in an animal model of induced CeD using a multiomic approach.

Results: Children with CeD who adhere to GFD still exhibit differences in bacterial and viral microbiota communities, as well as differences in metabolites compared to non-celiac control children. Additionally, in the study of the gliadin-induced enteropathy in rats as a CeD model, and the evaluation of the potential of VSL#3 as probiotic treatment to reverse some alterations, reveals that VSL#3 partially restores phenotypical alterations.

Conclusion: The study found that GFD in CeD patients did not fully restore the normal microbiota. Rat models showed potential for probiotic testing, requiring further research.

1. INTRODUCTION

Celiac disease (CeD) is an autoimmune enteropathy that results from a complex interaction between genetic and environmental factors. It is primarily induced by the ingestion of gluten, the main protein component found in wheat, barley, and other grains [1]. The gluten peptides, specifically gliadin, trigger and induce an autoimmune response when gliadin interacts

with antigens in cells of the small intestine mucosa. This response is associated with damage to the small intestine and the most histological characteristics are abnormal morphology, such as irregular epithelial cells, hyperplasia of crypts, disorganized villi, and degree of villous atrophy [2]. These morphological alterations result in numerous symptoms such as malabsorption, intestinal cancer and extraintestinal disorders (anemia and Dermatitis herpetiformis). The genetical predisposition to CeD is strongly associated with HLA class II, the HLA-DQ2/DQ8 haplotypes (it has been seen that 90-95% of CeD patients carry these haplotypes [3]). However, it is recognized that expression of these alleles is necessary for CeD development, but not sufficient for the disease occurrence. This fact indicates that CeD is a multifactorial disorder where numberless genes interact with environmental factors to cause CeD [4]. Notably, environmental factors such as lifestyle, type of diet (first exposition to gluten in infancy), antibiotic treatment and intestinal microbiota composition during childhood might induce and influence the development or progression of CeD [5,6].

Currently, the only effective treatment against celiac disease is lifelong adherence to a strict gluten-free diet (GFD) [7] by permanently eliminating cereals of wheat and barley, although medicines such as anti-tissue-transglutaminase IgA antibodies are yet under study [8]. However, in some cases, gastrointestinal symptoms persist despite a GFD [9], and long-term effects of GFD are associated with vitamin and mineral deficit [10].

The interplay between microbiota and the host is of paramount importance for the maintenance of intestinal integrity. Microbial population, including bacteria, archaea, yeast, and viruses is stablished early in life and research is now performed on how alterations in the early gut microbiota could influence child and adult health. The main bacterial phyla composing the gut microbiota are Firmicutes, Bacteroidetes, and Actinobacteria [11] but viruses also colonized the intestinal tract in a large proportion and highly correlate with bacterial population [12]. The homeostatic composition of

the microbiota can be disrupted by environmental agents and endogenous signals, and increase the risk of an immune condition, such as CeD, among others [13]. This disruption is called dysbiosis and is translated into reduction in microbial diversity and loss of beneficial bacteria [14]. Dysbiosis in the gut microbiota community has been shown in previous studies when comparing CeD patients and healthy people. What is more, some of these alterations are persisting in CeD individuals with GFD, mainly due to the reduced intake of polysaccharides. In fact, it has been reported that CeD individuals following a GFD presented a lower concentration of butyric, valeric and propionic acid (short-chain fatty acids known to protect intestinal mucosa) and a higher concentration of L-lactic acid in the stool [15]. This restricted diet results in pathogenic bacterial species overgrowing (*Enterobacteriaceae* and *Escherichia coli*) and other beneficial bacteria like *Bifidobacterium, Lactobacillus, Bifidobacterium longum, Clostridium lituseburense* and *Fecalibacterium prausnitzii* decreasing [16].

Other studies have suggested that gut microbiota may affect gluten digestion [17]. It is known that *Lactobacilli* and *Bifidobacterium* have a protective role because they release peptidases that may be involved in the breakdown of peptides in gluten [18]. This digestion decreases immunogenicity and reduces the toxicity of gluten. For this reason, among others, these strains have the potential to be used as probiotics in CeD patients in order to restore gut dysbiosis and aid in gliadin digestion reducing its toxicity and inflammatory effects. This, in turn, could help to improve the functioning of intestinal epithelial cells [19]. Hypothetically, the replacement of the missing commensal microbes with specific strains or a combination of defined strains could aid in the prevention and treatment of such disorders.

Altered microbiota in patients with treated CeD has been described, although it has not been provided a comprehensive understanding of the effects of this disbiosys. In this study, we compare the bacteriome, virome and metabolome of celiac children following GFD with those of healthy

controls to identify potential biomarkers of CeD symptomatology. Additionally, we evaluate in a rat model of induced CeD [20,21] the effect of the probiotic cocktail VSL#3, known to hydrolyze gluten proteins [20], using a multiomic approach that combines metagenomics, metatranscriptomics, metabolomics and the assessment of intestinal damage. This study aims to provide the foundational necessary knowledge to implement early preventive interventions that can re-establish tolerance and reduce the risk of developing CeD.

2. RESULTS

2.1 Multiomic comparison of fecal microbiota between children with CeD and healthy children

The study involved 27 children, who were divided into two groups: celiac patients adhering to a GFD (referred to as CeD group) consisting of 17 children (10 male and 7 female) with an average age of 8.2; and the control group consisting of 10 children (7 male and 3 female) with an average age of 6.2. Next-generation sequencing (NGS) was applied to identify and relatively quantify the bacteria and virus diversity present in the stool samples. Nuclear magnetic resonance (NMR) was performed to assess their metabolic activities.

2.1.1 Significant bacterial taxonomic differences between CeD and control groups

At the phylum level, both groups were found to be dominated by the same three phyla: Bacteroidetes, Firmicutes and Proteobacteria. The ratio of Firmicutes/Bacteroidetes were 0.36 in the control group and 0.65 in CeD group. Our analysis revealed significant differences in the abundance of 16 taxa at different taxonomic levels between the two groups ($p<0.05$). Seven genera were found exclusively in the control group, namely *Treponema, Gluconobacter, Peptostreptococcus*, *Lachnospiraceae* group NK3A20, *Prevotellacea* UCG004, F082 (pertaining to Bacteroidales order) and WCHB1 (pertaining to Kiritimatiellae class), indicating their potential as probiotics for preventing CeD if confirmed in future studies. Families *Campylobacteraceae, Butyricicocaceae* and *Micrococcaceae* were

significantly more abundant in the CeD group, while *Acholeplasmataceae* and *Barnesiellaceae* families and *Negativicoccus, Lachnospiraceae* and *Erysipelotrichaceae* genera were more abundant in the control group. The richness of the gut bacteriomes did not differ significantly between the two groups based on the Mann–Whitney U test of the Chao1 and Shannon indexes. Additionally, the fecal bacteriome communities of control and CeD groups were not clearly distinguishable based on β diversity based on Bray–Curtis distance (PERMANOVA, $p > 0.05$).

2.1.2 Children with CeD exhibit a different virome population compared to the control group.

The stool virome composition was analyzed at the genus level and DNA and RNA viromes were analyzed separately. The DNA virome primarily consists of sequences assigned to 365 different species, with the order Caudovirales (predominant bacterial virus) being the most prevalent in both groups. Notably, the presence of *Lactococcal* phage was restricted to CeD patients adhering to GFD, along with *Arthrobacter* phage, *Pseudomonas* phage pf16, *Thermus* phage P23-77 and *Staphylococcus* phage phiSA012. In contrast, the control group exhibited lower levels of three species of *Enterobacteria* phage and two species of *Escherichia* phage, while *Fecalibacterium* phage was more abundant. Although Shannon alpha diversity did not significantly different between CeD adhered to GFD and control groups, significant differences were observed in β diversity measured using Bray-Curtis dissimilarity. Specifically, control group samples were found to be significantly different from CeD group samples at the species level, indicating that the former were more clustered than the latter (PERMANOVA pseudo-F = 1.14, $p = 0.020$) (Figure 1A).

Analysis of RNA virome identified 65 different species, with 17% belonging to plant viruses, 23% to animal viruses and 33% to bacterial viruses. Of these, only 32 were detected in the control group. Salmonella phage 118970 was highly represented in both groups, accounting for 12% and 6% of relative abundance in the control and CeD groups, respectively,

but the differences are not significant between the two groups. Indeed, α diversity, measured using the Shannon index showed statistically significant differences between the two studied groups ($p<0.05$), being higher in CeD adhered to GFD. On the other hand, β diversity did not change, as shown in figure 1B (PERMANOVA pseudo-F =1.04, p = 0.277).

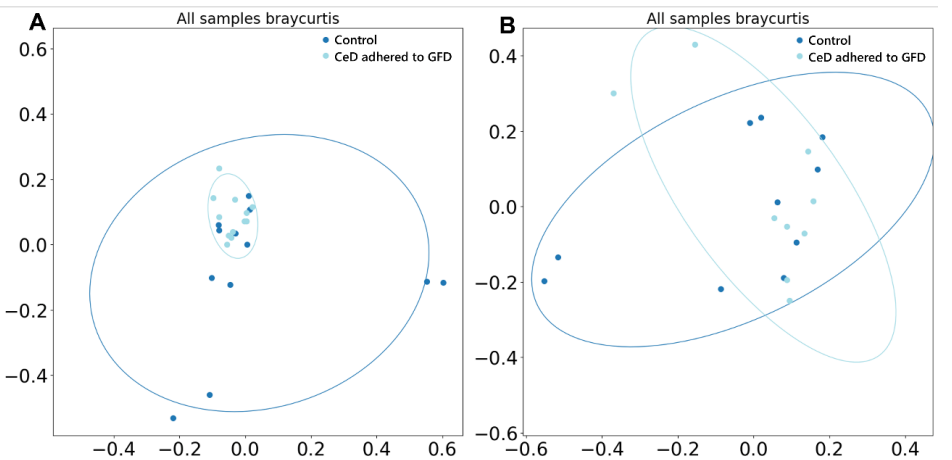


Figure 1: A) DNA virome β-diversity measured using bray Curtis. B) RNA virome β-diversity measured using bray Curtis.

2.1.3 Fumarate and glycerol changed significantly among groups.

31 metabolites were identified in feces after data filtering and normalization of the spectra which were used for subsequent statistical analysis (supplementary table 1). In this constructed OPLS-DA model, the R and Q2 values were 0.547 and 0.942 respectively. To validate the model, a permutation test was performed allowing 1000 permutations. The significance of the model was $p<0.05$, verifying the validity of the model.

Results showed that fumarate was significantly higher, and glycerol was found significantly lower ($p<0.05$) in the control group compared with CeD group (Figure 2).

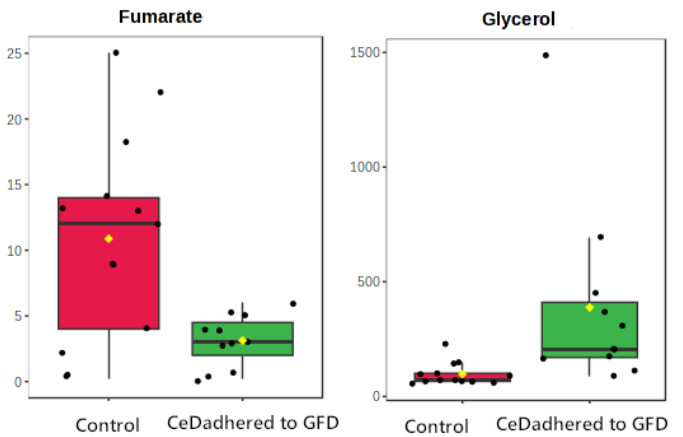


Figure 2: Boxplots from Fumarate and Glycerol concentration in two studied groups.

Metabolic pathway analysis was conducted using *Homo sapiens* (KEGG) as the pathway library to compare the two studied groups. The analysis revealed that two metabolic pathways were differentially regulated. Particularly, the Glycerolipid Metabolism and the Galactose Metabolism, both involving glycerol, were upregulated in the CeD group (Figure 3).

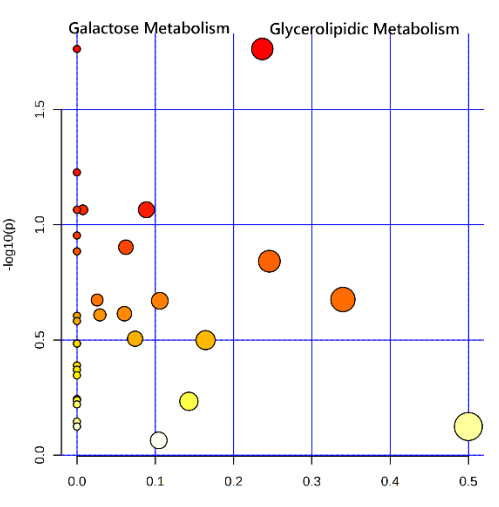


Figure 3: Graph of pathway impact and relative p-value for the comparison of metabolic pathway between CeD and control groups. Pathways were matched based on their p values from the pathway enrichment analysis and the pathway impact values obtained from the pathway topology analysis. The color and size of each circle are based on p-values and pathway impact values, respectively. Small p-values and large pathway impact circles indicate that the pathway is significantly perturbed.

2.2 Intestinal integrity assessment and multiomic approach for microbiota description in a rat model of celiac disease

A total of 30 offspring (n=10 per group) were obtained by crossing adult Wistar rats. One group served as the control health group (CON group), while the second group was sensitized with IFN-γ and fed with gliadin to mimic CeD (CD group). The third group was induced to elicit CeD and treated with the probiotic cocktail VSL#3 (CD+PB group).

2.2.1 Alterations in crypt depth and villi height were detected in rats with induced CeD.

Microscopic examination of CON animals (crypt depth, villi height and villus/crypt ratio) revealed normal ileum morphology while examination of the CeD model showed typical lesions of celiac disease such as tissue fragility (Figure 4).

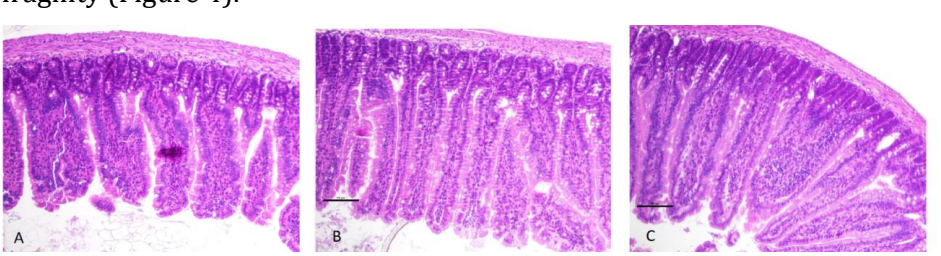


Figure 4: A photomicrograph of the ileum, A: CON animal, B: CD animal, C: CD+PB animal.

Statistically significant differences ($p<0.05$) were found among the groups. Crypt depth was lower in the CON group than in the CD group, resulting in

a higher villus/crypt ratio in the controls. In contrast, villi height and villus/crypt ratio were higher in the CD+PB group compared to the other groups. Finally, the CD and CD+PB groups differed in all three parameters tested, with the CD group having deeper crypts, lower villus height and lower villus/crypt ratio compared to the CD+PB (Figure 5).

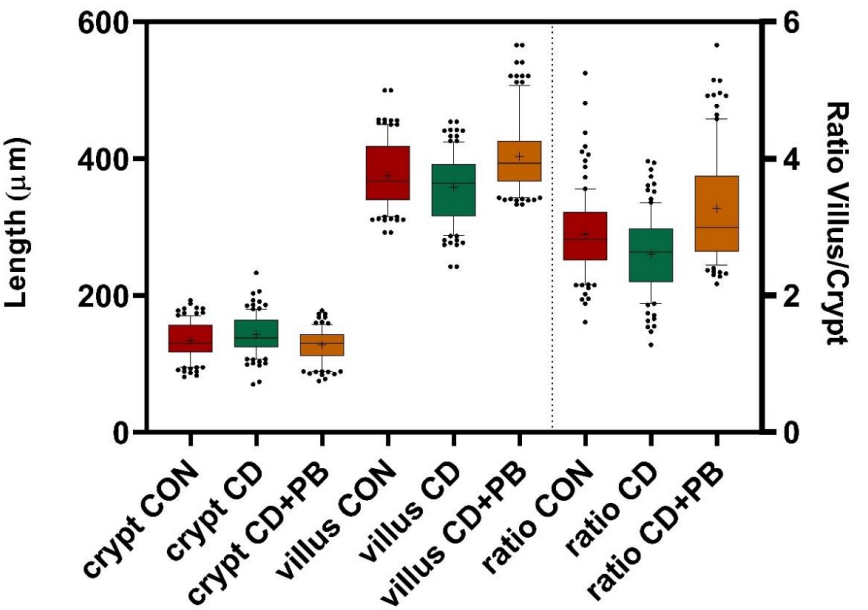


Figure 5: Values of the three parameters measured in the three studied groups using 10th-90th percentile range. Depth and villi height (in µm) are shown on the left, while villus/crypt ratio is shown in the right.

2.2.2 Bacterial and viral communities changed due to induced CeD but were not restored after probiotic administration.

There are several differences in the microbial population among the three studied groups.

Regarding richness, the CD group showed the highest alpha diversity score (Shannon index), while the CON group had the lowest score. However, no statistically significant differences were found between groups.

Focusing on bacterial taxonomy, we observed a trend with a p-value close to 0.05 for three families of the Cyanobacteria phylum: *Scytonemataceae, Synechococcales* and *Aphanothecaceae*. The CON group had higher abundance of the first two families while the CD+PB group had higher abundance of the third family. There were no statistically significant differences in bacterial distribution between the CD and CD+PB groups, except for a decreasing trend in the Firmicutes phylum in the CD+PB group ($p = 0.057$). In terms of β diversity, the CON group formed a distinctive cluster, while there was no differentiation between CD and CD+PB groups, indicating that the probiotic did not alter the bacterial community (Figure 6A).

Additionally, we found no statistically significant difference in the abundance of the specific eight bacterial strains forming the probiotic cocktail VSL#3, which was administered to the CD+PB group. This result suggests that this bacterial mixture may not have effectively established in the intestinal environment during probiotic treatment.

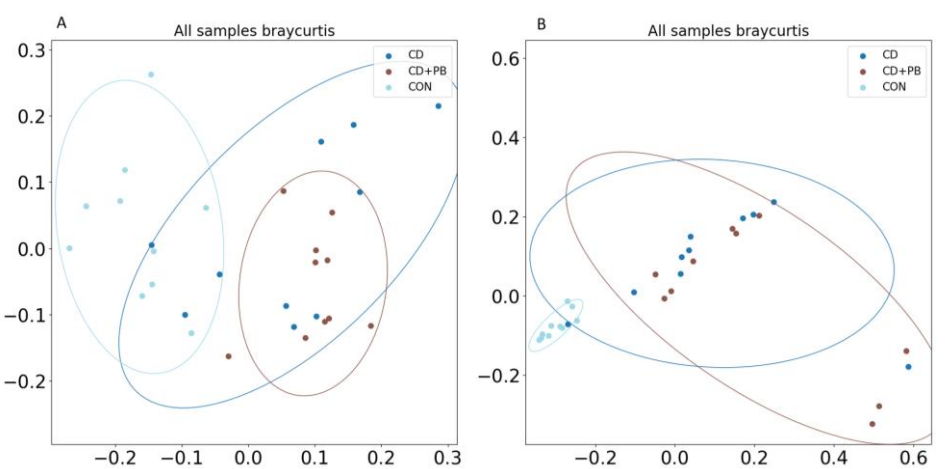


Figure 6: PCoA of bacterial diversity (A) and viral diversity (B), Bray-Curtis PERMANOVA ANOSIM $p<0.01$ in both cases.

In terms of viral taxonomic composition, all groups were found to be dominated by Caudovirales order, which accounted for 80-88% of all identified viruses. However, while the CON group consisted solely of the

Myoviridae family, the CD and CD+PB groups exhibited this abundance divided between two families: *Myoviridae* and *Siphoviridae* (48% and 32% respectively). Although this dissimilarity was not found to be statistically significant, a significant difference was observed in the Herpesvirales order ($p = 0.011$) between the three studied groups. Additionally, β diversity analysis, as depicted in figure 6B, revealed that the CON group formed a distinct cluster separated from the other two groups, indicating a significant shift in the virome population upon induction of CeD.

To comprehensively explore the genomic data obtained from the metagenomic sequencing, a functional analysis was conducted based on gene presence and content. Gene identification was assigned to metabolic pathways and compared between experimental groups. As a result, the lactose and galactose degradation pathways were found to be downregulated ($p = 0.04$) in CD and CD+PB groups compared to the CON group.

2.2.3 Only control group significantly differs at metatranscriptome analysis.

Fifteen samples (5 per group) were analyzed, and no significant differences were found in functional analysis. However, it is worth noting that the CON group metratranscriptomes appeared to differ from those of the other two groups (Figure 7). Bray-Curtis β diversity analysis revealed significant differences between the CON group and the other two groups ($p < 0.05$, PERMANOVA). In all three groups, a high percentage of the reads were assigned to hypothetical proteins, with over 80% in CD and CD+PB groups, and 34% in CON group which difficulted the functional assignment.

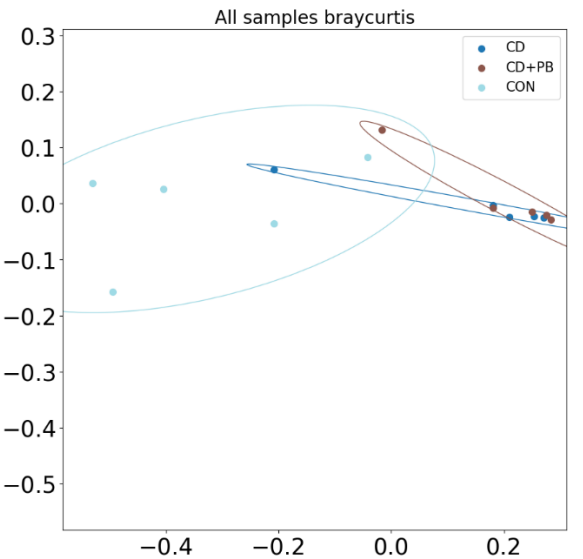


Figure 7: PCoA of functional analysis using bray Curtis.

However, when analyzing the expression of the transcripts coming from the host, we observed that *Muc2* gene expression was 5-fold higher in CD compared to CD+PB group (p = 0.052). This gene is known to be involved in mucus production. In contrast, we also noted that *NF- κB* inhibitor alpha (ia) gene expression was 5-fold lower in CD group than in the CD+PB (p = 0.061). This gene is involved in inhibiting NF-κB signaling pathway.

2.2.4 The metabolic profile is different between the three experimental groups.

After undergoing data filtering and normalization, 43 metabolites were identified in fecal samples, and their concentrations were found to be differently distributed among the three groups (Figure 8). In fact, when comparing the CD group to the CON group using a t-test, the concentration of nine metabolites were found to be significantly different. Specifically, methionine, tyrosine, asparagine, threonine, valine, leucine, alanine, and hypoxanthine concentrations were higher in the CD group, while succinate concentration was lower. Moreover, the PB+CD group showed higher levels of 2-oxoglutarate and lower levels of uracil, compared to the CON group.

Interestingly, the multivariate analysis 2D OPLS-DA demonstrated that the three groups could be distinguished based on their metabolic profile. The VIP scores generated by the OPLS-DA model with higher score than 1 were identified as significant contributors to the clustering of the three different groups in the OPLS-DA (Supplementary figure 1).

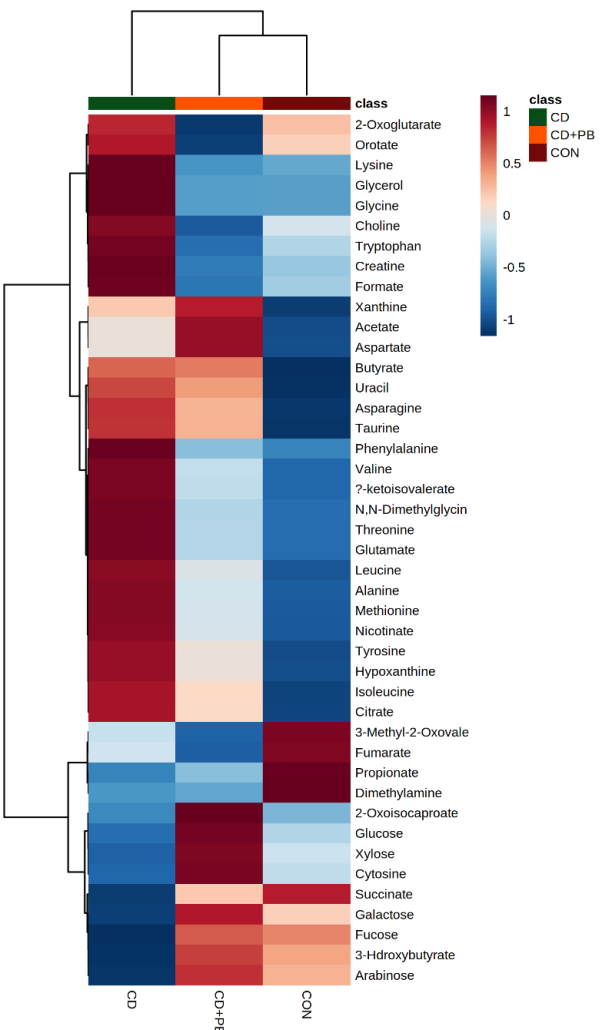


Figure 8: Heatmap of metabolites concentration in each experimental group.

The metabolic pathways were analyzed using *Rattus norvegicus* (KEGG) as the pathway model library and a comparison was made between CD and CD+PB groups. The results showed that three metabolic pathways were

differentially regulated. The pentose and glucuronate interconversion pathway (involving D-xylose metabolite) and the glycolysis pathway (involving beta-D-glucose and Acetate metabolites) were upregulated in the CD+PB group. Conversely, the cysteine and methionine metabolism pathway (involving L-methionine metabolite) was downregulated in the CD+PB group (Figure 9).

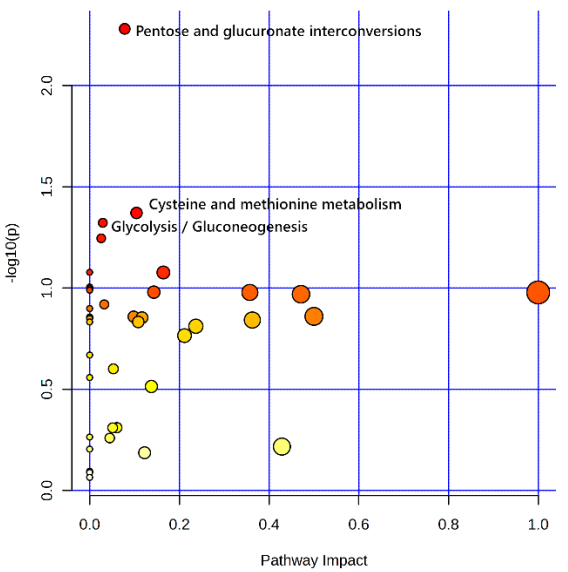


Figure 9: Graph of pathway impact and relative p-values for the comparison of metabolic pathways between CD and CD+PB groups. The pathways were matched based on their p values from the pathway enrichment analysis and pathway impact values from the pathway topology analysis. The color and size of each circle are based on the p-values and pathway impact values, respectively. Small p-values and large pathway impact circles indicate that the pathway is significantly perturbed.

3. DISCUSSION

3.1 Gut dysbiosis and metabolic alterations in CeD patients adhered to GFD

There is evidence that gut dysbiosis may contribute to the development of inflammatory disorders, such as inflammatory bowel diseases, Crohn's disease, and Celiac disease [22]. However, it is not yet clear to what extent the dysbiosis observed in previous research is involved in CeD pathogenesis or if it is a secondary effect of disease pathology [23]. To shed light on this issue, the first part of our study aimed to investigate the differences in the human fecal microbiota of enfants with CeD who were adherent to a GFD, compared to control subjects without the disease, offering the opportunity to identify host characteristics and associated changes to the CeD microbiome in an integrated multi-omic framework. For this reason, we assessed the profiles of fecal bacteriome, virome and metabolome. Nonetheless, these results must be interpreted with caution as sample size limitation should be considered.

As a main finding, it can be stated that the GFD treatment did not fully restore the normal intestinal microbiota in the studied CeD patients, as evidenced by ongoing differences compared to healthy individuals. as reported in previous studies [24][25] Moreover, certain pathogenic bacterial strains associated with CeD such as the *Lachnospiracea* and *Prevotellaceae*, were not detected in CeD patients in our study but were detected in healthy individuals, which contrasts with previous studies [26]. The relative abundance of the *Butyricicocaceae* family was increased in the CeD group, and this family is known to produce butyrate and has probiotic potential, as already reported [27]. Similarly, the *Campylobacteraceae* family was increased in the CeD group, and this has been found to correlate with metabolic markers [28].

In the present study, we conducted a DNA and RNA viral profile analysis of CeD children who adhered to GFD and compared their profile with those of healthy children. Most previous viromics profiling studies have focused

solely on the DNA virome, as they only require a single methodological protocol [29]. However, in our case, we carried out two separate protocols one for the DNA and the other for RNA virome. Upon comparative analysis of the two groups, we found that, at the level of individual phages, the *Lactococcal* phage was only detected in CeD children, which contrasts with previous studies that suggested a protective role of this phage because of its lower presence in CeD samples [30]. On the other hand, the lower presence of *Enterobacteria* phages in healthy children was consistent with a previous study and we also found three *Escherichia* phages that were less abundant in healthy controls. In terms of the RNA virome, we identified 65 species, eight of which had a relative abundance of between 80-90% (93% CeD group and 83% control group). CeD patients presented lower viral DNA and RNA β-diversity than healthy controls. The causality of lower diversity to diseases remains to be identified. However, healthy individuals have normally high bacterial diversity, unlike individuals with some pathology where bacterial diversity decreases [31]. This result suggests that bacterial and viral diversity are closely correlated.

Our study found that adhering to a GFD may lead to a lower level of beneficial metabolites and an increase in toxic metabolites in patients with CeD compared to a control group. These differences could be due to bacterial dysbiosis or/and the composition of GFD, which differs from a non-restricted diet. De Angelis et al. reported alterations in fecal levels of six metabolites in CeD patients treated with GFD compared to healthy children: tryptophan, proline, asparagine, histidine, methionine, trimethylamine-N-ox and tyramine were lower in CeD patients. However, the most studied microbial metabolites considered to affect the system homeostasis, acetate, butyrate, and propionate [32] (three are SCFA), did not show significant changes between the two studied groups in our study which contrasts with previous reports [15]. Nevertheless, we observed significant differences in fumarate and glycerol levels, with the former being higher in the control group and the latter being lower. These findings differ from those reported

previously by Upadhyay et al [33] where glycerol was lower in intestinal mucosa but higher in blood plasma in samples of CeD patients compared to healthy controls. It's worth to note, however, that their study analyzed mucosa directly, while we analyzed stool samples. The metabolic pathways analysis also showed an alteration in glycerolipidic metabolism pathway agreeing with a previous study [33] and in Galactose metabolism, which may be related to malabsorption caused by CeD [34].

In summary, the results suggest a close relationship between CeD and gut microbiome and metabolome. Bacteriome, virome and metabolome were altered, indicating that either changes in the bacteriome perturbed the virome and metabolome, or inversely, an altered metabolome perturbed the bacteriome and virome. These differences could be caused by several factors, including disease activity in host tissues, an altered microbiome, and by external factors such as variations in subject diet and medication use. Moreover, when examining the metabolomic outcome, we can hypothesize that GFD resulted in a significant decrease in the intake of certain natural ingredients which may partially account for the reduction in beneficial gut bacteria populations.

3.2 Impact of probiotic supplementation on intestinal mucosa, microbiota, and metabolite composition in Gliadin-Induced Enteropathy rat model

The second part of the study aimed to investigate the effects of gliadin-induced enteropathy on the intestinal mucosa, altered microbiota population and activity, and also metabolite composition, using rats as a model for CeD. Additionally, the study explored the potential of probiotic treatment to reverse some of these alterations, given that probiotics are considered a promising approach for modulating the microbiome and promote an anti-inflammatory state.

Analysis of the histopathological sections of the ileum revealed significant differences between the control group and the group with induced CeD (CD). Notably, the depth of the crypt, was found greater in the induced CeD.

Additionally, villus/crypt ratio was higher in control (CON) group, indicating that the depth of the crypt relative to the villi, was greater in the CD group. When comparing the CD group with the probiotic treated group (CD+PB), significant differences were observed in the three parameters (crypt depth, villi height and villus/crypt ratio). The CON group and CD+PB group presented significant differences in villi and ratio; suggesting that the administration of VSL#3 could partially restore the phenotypical alterations observed in the ileum at the histopathological analysis. These findings are consistent with previous research [21] which demonstrated that gliadin administration and IFN-gamma sensitization are necessary for inducting intestinal damage and causing alterations at different levels. Moreover, *B. longum* and *Lactobacillus casei* have been found to be able to restore intestinal damage.

Numerous human clinical trials have investigated the effects of probiotic supplementation on celiac disease. However, the results have been inconsistent, and further research is needed. For instance, Harnett et al. conducted a 12-week study in which 5g of VSL#3, containing probiotic strains capable of hydrolyzing gluten, was administered to human subjects. Despite this promising feature, the study found no significant changes in fecal microbiota counts, blood safety parameters, or clinical symptoms. These findings suggest that additional research is necessary to determine the potential of VSL#3 as a therapy for celiac disease [35].

We used a rat model to evaluate the effects of administering eight bacterial strains as a probiotic (VSL#3 solution). Our results showed that the CD+PB group did not exhibit a significant increase in bacterial counts compared to the control group. Interestingly, we also found that α-diversity of the microbiota was lower in the control group than in both CD and CD+PB groups. This finding contrasts with the common view that richness tends to decrease in some diseases.

Cyanobacteria and Verrucomicrobia are typically underrepresented in human healthy microbiome [36], and conveniently, our study found

differences in Cyanobacteria phyla (including *Scytonemataceae, Synechococcales* and *Aphanothcaceae* families), among the three groups we examined. Although prior studies have shed some light on the connection between Cyanobacteria and specific illnesses, the current evidence is primarily based on correlations [37]. Additionally, we observed a significant difference in the Herpesvirales, being higher in the CD+PB group, and lower in the CON group. Herpesviruses are known for their ability to establish lifelong infections within hosts and may therefore be considered a component of the microbiome [38]. In terms of metabolic pathways, our analysis of the whole metagenome revealed downregulation of the Lactose and Galactose degradation pathways in CD and CD+PB groups compared to the control group. This finding is consistent with prior research indicating that lactose malabsorption can occur as a primary or secondary disorder due to intestinal diseases, including celiac disease resulting from the loss of lactase enzyme in the damaged small intestinal villi [39]. Notably, we did not find any significant difference between the CD and CD+PB groups, suggesting that the VSL#3 supplementation was not effective in restoring the gut microbiota.

In the metatranscriptomic analysis, in terms of clustering and functional analysis, no differences were found between the CD and the CD+PB groups, while CON group significantly differed from the other two groups. However, given the high percentage of hypothetical proteins combing from bacterial transcripts, a definite conclusion cannot be drawn. Conversely, the *Muc2* and *NF-κB ia* genes showed nearly significant statistically differences in our study. In particular, the *Muc2* gene was five times higher in CD group than in the CD+PB group. Previous research has suggested that *Muc2* is overexpressed in response to intestinal damage in certain intestinal diseases [40], and our results could indicate that the probiotic treatment may help to improve intestinal permeability and inhibit the overexpression of genes related to mucus production. Moreover, we observed that the expression of *NF-κB ia*, an inhibitor of NF-κB pathway, was five times lower in the CD group compared with CD+PB group. It is worth noting that the

activation of *NF-κB* has been shown to induce the expression of *Muc2* gene in the intestine [41]. Therefore, the histopathological effects observed in the CD+PB group may be related to the modulation of NF-kb pathway and repression of the *Muc2* gene expression.

The metabolic profile of the fecal content revealed 43 metabolites which is twelve more than the number of metabolites identified in the human feces. Elevated levels of different amino acids were observed in the induced CeD group when compared to the control group. Specifically, methionine, tyrosine, asparagine, threonine, valine, leucine, and alanine were found to be at higher levels. Previous research has suggested that the differences in amino acids levels between patients with CeD and healthy individuals may be linked to inflammatory processes [42]. Additionally, the probiotic group exhibited higher levels of 2-oxoglutarate (2-OG) than the control group. 2-OG serves as a necessary substrate in a variety of oxidative reactions that are catalyzed by 2-OG-dependent dioxygenase, according to prior studies [43].

Further analysis of the metabolic pathways showed a significant alteration in the metabolism of cysteine and methionine in the CD group. This finding is consistent with the earlier observation [42] that the concentration of L-methionine was higher in CeD. In addition, the analysis revealed that the energy generation pathways, specifically glycolysis and gluconeogenesis, were altered in the CD+PB group. Specifically, Beta-D-glucose and acetate were found to be at higher levels in this group than in the other two groups. Finally, the pentose and glucuronate interconversion pathways were also found to be altered in the probiotic group, as evidenced by the increase of D-xylose which is involved in this pathway.

In summary, the effectiveness of VSL#3 in restoring histopathological alterations was confirmed. However, microbial dysbiosis was not restored. On the contrary, changes were observed at the functional level, making the CD+PB group more similar to the control group, particularly at the metabolic level. There are very few studies on the use of probiotics for treating CeD in humans. Previously, Olivares et al. demonstrated the

potential of *Bifidobacterium longum* to improve the health of CeD patients with altered gut microbiota, and a skewed immune response even on a gluten-free diet, to have better health. Based on the results of Olivares et al. and our own study, further research is needed to determine the optimal probiotic strain and dosage.

4. CONCLUSION

The study found that GFD treatment did not fully restore the normal intestinal microbiota in the studied CeD patients, as there were ongoing differences compared to healthy individuals. This suggests that apart from the nutritional deficiencies resulting from a GFD, reduced intake of polysaccharides (prebiotics) can lead to changes in the intestinal microbiota. Additionally, the similarities between the virome and bacteriome behavior of alpha and beta diversity in both groups suggest a correlation between them.

Rat models can be used as preclinical models for testing interventions before human trials, in order to study specific intestinal diseases such as CeD In our study, probiotic treatment in rats resulted in noticeable histopathological changes, which could potentially improve symptoms in humans. Further research is required to determine the effectiveness of this treatment and identify the optimal probiotic strain and dosage.

5. MATERIAL AND METHODS

All fecal samples in this study were immediately frozen after collection and stored at -80ºC until analysis. Ethical approval was granted on June 28, 2018. Informed consent was obtained from the parents of the eligible children. The study is in accordance with the ethical standards of the Declaration of Helsinki.

5.1 Multiomic fecal analysis from CeD and control children

Bacteriome analysis

The fecal Infant Bacteriome, including DNA extraction, PCR amplification, library preparation, and sequencing have been previously described [44]. Briefly, DNA was extracted using QIAamp DNA Stool Mini Kit (Qiagen Inc., Hilden, Germany), ag-Fusion Primers of the V4 region of the 16S rRNA gene were used to create libraries. Quality assessment and quantification of each library was performed with TapeStation (Agilent Technologies, Santa Clara, CA, USA). An equimolar mixture of the libraries was clonally amplified with Ion Chef System and sequenced with Ion Torrent GenStudio S5 (Thermo Fisher Scientific, Waltham, MA, USA), achieving up to 100 thousand reads per sample. The generated reads were quality filtered and analyzed with QIIME2 (2022.2) and passed for classification into Amplicon Sequence Variants (ASV) to DADA2, using only reads of at least 200 bp and truncating at that length. These ASVs were taxonomically classified with VSEARCH against the Silva database at a 99% homology [45]. Alpha diversity was calculated as Shannon, Chao1 and Faith indices, and beta diversity was calculated with Bray-Curtis, Aitchison, Jaccard and Unifrac (weighted and unweighted) distances.

Virome analysis

The fecal Infant virome analysis was performed as we described previously [46]. Briefly, the virus-like particles (VLP) were enriched, and fecal samples were resuspended in phosphate buffer saline (PBS) and centrifugated at low speed. The supernatant was filtered and later ultracentrifuged. The pellet was resuspended and treated with Bensonase Nuclease (Merck, Germany). Nucleic acids were extracted using Quick-DNA/RNA Viral Kit (Zymo Research, Irvine, CA) and divided into two aliquots, with one of them treated with DNase and the other aliquot treated with RNase. DNA and RNA libraries were prepared separately; with DNA libraries created by Ion Xpress™ Plus Fragment Library Kit for AB Library Builder™ System (Thermo Fisher Scientific, Waltham, MA, USA) and RNA libraries created by Total RNA-Seq Kit v2 for Whole Transcriptome Libraries. Libraries were clonally amplified on the Ion Chef System and sequencing was on the Ion

Torrent S5 (Thermo Fisher Scientific, Waltham, MA, USA), achieving up to 2 million reads per sample. For both DNA and RNA, each sample's reads were assembled using SPAdes genome assembler 3.15.5. The resulting contigs were classified as viral, potentially viral or non-viral using viralVerify 1.1 (https://github.com/ablab/viralVerify).

With the Pfam-A HMM database and a BLAST against the NCBI "nt" database. Contigs marked as non-viral or that matched against cellular organisms or plasmids were discarded. The abundance of each contig within each sample was derived from its coverage. The remaining contigs were assigned a taxonomy based on their best match in the BLAST results. The contigs from all the samples were clustered at 90% homology to form OTUs (Operational Taxonomic Units) comparable across samples. From these OTUs, alpha and beta diversity analysis like those of the bacteriome were performed.

Metabolome analysis

Fecal Metabolite Profiling by 1H-NMR protocol has been previously described [44] Fecal samples were homogenized with PBS and subsequently centrifuged to separate both phases. The upper phase was mixed with deuterated PBS to be analyzed by NMR. The 1H NMR spectra were measured at 600.20 MHz frequency at 300K on an Avance III-600 Bruker spectrometer using a 5 mm PABBO BB-1H/D Z-GRD probe. Metabolite identification was carried out using information from the literature, public databases, and proprietary software (Chenomx NMR Suite 8.5®, Human Metabolite DataBase, Biological Magnetic Resonance Data Bank), and normalization was performed using the PULCON (PULse length-based CONcentration determination) methodology (ERETIC®), which is based on the principle of reciprocity. Orthogonal partial least squares-discriminant analysis (OPLS-DA) was performed with Metaboanalyst software 5.0 (https://www.metaboanalyst.ca/; accessed on 23 October 2020). The statistical calculation of the significance of the correlated OPLS-DA metabolites ($p < 0.05$) was obtained using the website for statistical

computation VassarStats (http://www.vassarstats.net/; accessed on 27 October 2020.

5.2 Intestinal damage assessment and multiomic analysis in celiac disease rat model

Rat model development

To investigate the effects of probiotic supplementation on celiac disease (CeD) in rats, we followed an experimental design similar to that described by Laparra and Selim [21] . Experimental animal procedures were approved by the Animal Ethics Committee of the Technological Unit of Nutrition and Health of Eurecat (Reus, Spain) and the Generalitat de Catalunya (10771). The experimental protocols followed the "Principles of Laboratory Care" guidelines and was carried out in accordance with the European Communities Council Directive (86/609/EEC). Adult Wistar rats were crossed to obtain the offspring, and cross-fostering was performed to reduce potential genetic effects. Thirty offspring were used in the study, (Table S2). and were distributed according to whether they were induced to CeD, and whether they were supplemented with a probiotic cocktail called VSL#3 or a placebo protein from day 7 after birth:

- CON group: formed by 10 animals fed with breast milk and supplemented with placebo protein.
- CD group: formed by 10 animals induced to CeD and supplemented with placebo protein.
- CD+PB: formed by 10 animals induced to CeD and supplemented with VSL#3 cocktail.

To induce CeD, we administered 1000 U of IFN-γ intraperitoneally to the offspring on their birthday, following established protocols previously described by other authors. The CON group received intraperitoneal saline (IFN-γ vehicle) and hypoallergenic formula milk as a placebo for gliadin treatment. Prior to administration, the gliadins were digested with acid

digestion and pepsin (0.2N HCl) followed by trypsin digestion and enzyme inactivation by boiling. The gliadins were then dissolved in hypoallergenic formula milk and administered to the offspring, with the amount increasing over time (see Supplemental Table 2). The day of birth was considered day 0 of the study.

Probiotic supplementation included eight dried bacterial strains (*Streptococcus thermophilus, Bifidobacterium breve, Bifidobacterium longum, Bifidobacterium infantis, Lactobacillus acidophilus, Lactobacillus plantarum, Lactobacillus paracasei* and *Lactobacillus bulgaricus*) prepared in the commercial solution VSL#3. Dosage was set at kg 2580 billion bacteria/kg and prepared my mixing it with protein surimi.

Histopathologic assessment:

A piece of the intestine from each animal was fixed in 4% diluted formaldehyde for a posterior histopathological analysis to evaluate the crypts and villus length using Harris haematoxylin and eosin (H&E) staining (n=10 per group). After 24 hours of fixation, successive dehydration (alcohol/ethanol 70%, 96% and 100% plus xylol/dimethyl benzene) and paraffin infiltration-immersion at 52ºC (Citadel 200, Thermo Scientifica) were performed. Training paraffin block (Thermo Scientific Histostar) and subsequent successive 2µm thickness sections (Microm HM 355S, Thermo Scientific) were performed. Sections were deposited above slides (JP Selecta Paraffin Bath) and followed to automated H&E staining (Varistain Gemini, Thermo Scientific). The histological sections of the ileum were observed with a Nikon Eclipse Ti microscope using a 10X objective. The intestinal integrity was assessed by histopathological sections of the ileum and was focused on the following parameters: crypt depth, villus height, and rate of villus/crypt. Per each sample, five intestinal villi and 10 Lieberkühn crypts were measured, which were chosen at random considering that the villi were complete and in continuity with the mucosa and that crypts presented well-defined light. Statistical analysis was performed using SPSS version

26.0 (SPSS Inc., Chicago, IL, USA). Levene's test was performed to determine the homogeneity of variances. Mann-Whitney U-test was used and $p< 0.05$ was considered statistically significant.

Metagenome analysis:

Total fecal DNA was extracted using the DNeasy PowerSoil Pro Kit (Qiagen Inc., Hilden, Germany) according to the manufacturer's instructions. The purity of extracted DNA were assessed using Nanodrop2000 spectrophotometer (ThermoFisher Scientific, MA, USA). and were quantified using Qubit dsDNA BR Assay kit (ThermoFisher Scientific, MA, USA) was used for DNA quantification. For posterior libraries preparation at least 2ug of DNA was used using Illumina DNA Prep Kit (Illumina, USA). Prepared libraries were quantified with Qubit 1X ds DNA HS Assay kit and their quality and length were performed with high Sensitivity DNA ScreenTape Analysis (Agilent Technologies, USA). An equimolar pool was prepared at 2nM and diluted to 750pM with 1% of PhyX control. The indexed libraries were sequenced in NextSeq2000, achieving up to 7 million 150pbx2 PE reads per sample.

The generated reads were classified using Kraken 2.1.2 with the "Standard plus Protozoa and Fungi" Kraken database [47]. α and β diversity analyses were further derived from the results. Using the Holm-Sidak post hoc adjustment of the Kruskal-Wallis test, taxonomic abundances were compared between the experimental groups, and PERMANOVA and ANOSIM analyses were performed on the beta diversity difference between groups.

Metatranscriptome analysis:

Total fecal RNA was extracted using RNeasy PowerMicrobiome Kit (Qiagen Inc., Hilden, Germany) according to the manufacturer's instructions. Quality of extracted RNA was performed with RNA ScreenTape Analysis based on

RNA Integrity Number (RIN). Indexed libraries were created using 100ng of RNA with the Illumina Stranded Total RNA Prep with Rib-Zero Kit (Illumina, USA). The quality and concentration of the prepared libraries were checked with high Sensitivity DNA ScreenTape Analysis (Agilent Technologies, USA) and Qubit 1X ds DNA HS Assay kit, respectively. Finally, the libraries were pooled equimolarly to 750pM final concentration with 1% of PhyX control. Prepared libraries were sequenced in NextSeq2000 achieving up to 20 million 76pbx2 PE reads per sample.

The metatranscriptome was analyzed using SAMSA2 [48] and α and β diversity analyses of the found taxa were performed with QIIME2. Alpha diversity was calculated as Shannon and Chao1 indices, and beta diversity was calculated with Bray-Curtis, Aitchison, and Jaccard distances. PERMANOVA and ANOSIM analyses were performed on the beta diversity distances among groups. Taxonomic abundance for each taxon at every level has been compared using Kruskal-Wallis tests and corrected by Holm-Sidak post hoc correction.

Metabolome analysis

The fecal Metabolite Profiling by NMR was performed as described in section 5.1

REFERENCES

1. di Sabatino, A.; Corazza, G.R. Coeliac Disease. *Lancet* **2009**, *373*, 1480–1493, doi:10.1016/S0140-6736(09)60254-3.

2. Dickson, B.C.; Streutker, C.J.; Chetty, R. Coeliac Disease: An Update for Pathologists. *J Clin Pathol* **2006**, *59*, 1008, doi:10.1136/JCP.2005.035345.

3. Kneepkens, C.M.F.; von Blomberg, B.M.E. Clinical Practice: Coeliac Disease. *Eur J Pediatr* 2012, *171*, 1011–1021.

4. Anderson, R.P.; Henry, M.J.; Taylor, R.; Duncan, E.L.; Danoy, P.; Costa, M.J.; Addison, K.; Tye-Din, J.A.; Kotowicz, M.A.; Knight, R.E.; et al. A Novel Serogenetic Approach Determines the Community Prevalence of Celiac

Disease and Informs Improved Diagnostic Pathways. *BMC Med* **2013**, *11*, doi:10.1186/1741-7015-11-188.

5. Leonard, M.M.; Karathia, H.; Pujolassos, M.; Troisi, J.; Valitutti, F.; Subramanian, P.; Camhi, S.; Kenyon, V.; Colucci, A.; Serena, G.; et al. Multi-Omics Analysis Reveals the Influence of Genetic and Environmental Risk Factors on Developing Gut Microbiota in Infants at Risk of Celiac Disease. *Microbiome* **2020**, *8*, doi:10.1186/S40168-020-00906-W.

6. Roos, S.; Liedberg, G.M.; Hellström, I.; Wilhelmsson, S. Persistent Symptoms in People With Celiac Disease Despite Gluten-Free Diet: A Concern? *Gastroenterol Nurs* **2019**, *42*, 496–503, doi:10.1097/SGA.0000000000000377.

7. Green, P.H.R.; Jabri, B. Coeliac Disease. *Lancet* **2003**, *362*, 383–391, doi:10.1016/S0140-6736(03)14027-5.

8. di Sabatino, A.; Biagi, F.; Lenzi, M.; Frulloni, L.; Lenti, M.V.; Giuffrida, P.; Corazza, G.R. Clinical Usefulness of Serum Antibodies as Biomarkers of Gastrointestinal and Liver Diseases. *Digestive and Liver Disease* 2017, *49*, 947–956.

9. Caio, G.; Volta, U.; Sapone, A.; Leffler, D.A.; De Giorgio, R.; Catassi, C.; Fasano, A. Celiac Disease: A Comprehensive Current Review. *BMC Medicine 2019 17:1* **2019**, *17*, 1–20, doi:10.1186/S12916-019-1380-Z.

10. Lerner, A.; O'Bryan, T.; Matthias, T. Navigating the Gluten-Free Boom: The Dark Side of Gluten Free Diet. *Front Pediatr* **2019**, *7*, 414, doi:10.3389/FPED.2019.00414/BIBTEX.

11. Tap, J.; Mondot, S.; Levenez, F.; Pelletier, E.; Caron, C.; Furet, J.P.; Ugarte, E.; Muñoz-Tamayo, R.; Paslier, D.L.E.; Nalin, R.; et al. Towards the Human Intestinal Microbiota Phylogenetic Core. *Environ Microbiol* **2009**, *11*, 2574–2584, doi:10.1111/J.1462-2920.2009.01982.X.

12. Liang, G.; Bushman, F.D. The Human Virome: Assembly, Composition and Host Interactions. *Nature Reviews Microbiology 2021 19:8* **2021**, *19*, 514–527, doi:10.1038/s41579-021-00536-5.

13. Marchesi, J.R.; Adams, D.H.; Fava, F.; Hermes, G.D.A.; Hirschfield, G.M.; Hold, G.; Quraishi, M.N.; Kinross, J.; Smidt, H.; Tuohy, K.M.; et al. The Gut Microbiota and Host Health: A New Clinical Frontier. *Gut* **2016**, *65*, 330–339, doi:10.1136/gutjnl-2015-309990.

14. Dydensborg Sander, S.; Nybo Andersen, A.M.; Murray, J.A.; Karlstad, Ø.; Husby, S.; Størdal, K. Association Between Antibiotics in the First Year of Life and Celiac Disease. *Gastroenterology* **2019**, *156*, 2217–2229, doi:10.1053/J.GASTRO.2019.02.039.

15. Zafeiropoulou, K.; Nichols, B.; Mackinder, M.; Biskou, O.; Rizou, E.; Karanikolou, A.; Clark, C.; Buchanan, E.; Cardigan, T.; Duncan, H.; et al. Alterations in Intestinal Microbiota of Children With Celiac Disease at the Time of Diagnosis and on a Gluten-Free Diet. *Gastroenterology* **2020**, *159*, 2039-2051.e20, doi:10.1053/J.GASTRO.2020.08.007.

16. de Palma, G.; Nadal, I.; Collado, M.C.; Sanz, Y. Effects of a Gluten-Free Diet on Gut Microbiota and Immune Function in Healthy Adult Human Subjects. **2021**, doi:10.1017/S0007114509371767.

17. Bascuñán, K.A.; Araya, M.; Roncoroni, L.; Doneda, L.; Elli, L. Dietary Gluten as a Conditioning Factor of the Gut Microbiota in Celiac Disease. *Advances in Nutrition* **2020**, *11*, 160–174, doi:10.1093/ADVANCES/NMZ080.

18. Lindfors, K.; Blomqvist, T.; Juuti-Uusitalo, K.; Stenman, S.; Venäläinen, J.; Mäki, M.; Kaukinen, K. Live Probiotic Bifidobacterium Lactis Bacteria Inhibit the Toxic Effects Induced by Wheat Gliadin in Epithelial Cell Culture. *Clin Exp Immunol* **2008**, *152*, 552–558, doi:10.1111/J.1365-2249.2008.03635.X.

19. Olivares, M.; Laparra, M.; Sanz, Y. Influence of Bifidobacterium Longum CECT 7347 and Gliadin Peptides on Intestinal Epithelial Cell Proteome. *J Agric Food Chem* **2011**, *59*, 7666–7671, doi:10.1021/JF201212M.

20. Angelis, M. De; Rizzello, C.G.; Fasano, A.; Clemente, M.G.; Simone, C. De; Silano, M.; Vincenzi, M. De; Losito, I.; Gobbetti, M. VSL#3 Probiotic Preparation Has the Capacity to Hydrolyze Gliadin Polypeptides Responsible for Celiac Sprue. *Biochim Biophys Acta* **2006**, *1762*, 80–93, doi:10.1016/J.BBADIS.2005.09.008.

21. Laparra, J.M.; Sanz, Y. Bifidobacteria Inhibit the Inflammatory Response Induced by Gliadins in Intestinal Epithelial Cells via Modifications of Toxic Peptide Generation during Digestion. *J Cell Biochem* **2010**, *109*, 801–807, doi:10.1002/jcb.22459.

22. Laparra, J.M.; Olivares, M.; Gallina, O.; Sanz, Y. Bifidobacterium Longum CECT 7347 Modulates Immune Responses in a Gliadin-Induced Enteropathy Animal Model. *PLoS One* **2012**, *7*, doi:10.1371/journal.pone.0030744.

23. Clemente, J.C.; Ursell, L.K.; Parfrey, L.W.; Knight, R. The Impact of the Gut Microbiota on Human Health: An Integrative View. *Cell* **2012**, *148*, 1258–1270, doi:10.1016/j.cell.2012.01.035.

24. Sanz, Y.; Sánchez, E.; Marzotto, M.; Calabuig, M.; Torriani, S.; Dellaglio, F. Differences in Faecal Bacterial Communities in Coeliac and Healthy Children as Detected by PCR and Denaturing Gradient Gel Electrophoresis. *FEMS Immunol Med Microbiol* **2007**, *51*, 562–568, doi:10.1111/J.1574-695X.2007.00337.X.

25. Nistal, E.; Caminero, A.; Vivas, S.; Ruiz De Morales, J.M.; Sáenz De Miera, L.E.; Rodríguez-Aparicio, L.B.; Casqueiro, J. Differences in Faecal Bacteria Populations and Faecal Bacteria Metabolism in Healthy Adults and Celiac Disease Patients. *Biochimie* **2012**, *94*, 1724–1729, doi:10.1016/J.BIOCHI.2012.03.025.

26. Sanz, Y. Effects of a Gluten-Free Diet on Gut Microbiota and Immune Function in Healthy Adult Humans. *Gut Microbes* **2010**, *1*, 135–137, doi:10.4161/GMIC.1.3.11868.

27. de Angelis, M.; Vannini, L.; di Cagno, R.; Cavallo, N.; Minervini, F.; Francavilla, R.; Ercolini, D.; Gobbetti, M. Salivary and Fecal Microbiota and Metabolome of Celiac Children under Gluten-Free Diet. *Int J Food Microbiol* **2016**, *239*, 125–132, doi:10.1016/J.IJFOODMICRO.2016.07.025.

28. Geirnaert, A.; Steyaert, A.; Eeckhaut, V.; Debruyne, B.; Arends, J.B.A.; van Immerseel, F.; Boon, N.; van de Wiele, T. Butyricicoccus Pullicaecorum, a Butyrate Producer with Probiotic Potential, Is Intrinsically Tolerant to Stomach and Small Intestine Conditions. *Anaerobe* **2014**, *30*, 70–74, doi:10.1016/J.ANAEROBE.2014.08.010.

29. le Chatelier, E.; Nielsen, T.; Qin, J.; Prifti, E.; Hildebrand, F.; Falony, G.; Almeida, M.; Arumugam, M.; Batto, J.M.; Kennedy, S.; et al. Richness of Human Gut Microbiome Correlates with Metabolic Markers. *Nature* **2013**, *500*, 541–546, doi:10.1038/NATURE12506.

30. Nishijima, S.; Nagata, N.; Kiguchi, Y.; Kojima, Y.; Miyoshi-Akiyama, T.; Kimura, M.; Ohsugi, M.; Ueki, K.; Oka, S.; Mizokami, M.; et al. Extensive Gut Virome Variation and Its Associations with Host and Environmental Factors in a Population-Level Cohort. *Nat Commun* **2022**, *13*, doi:10.1038/S41467-022-32832-W.

31. El Mouzan, M.; Assiri, A.; Al Sarkhy, A.; Alasmi, M.; Saeed, A.; Al-Hussaini, A.; AlSaleem, B.; Al Mofarreh, M. Viral Dysbiosis in Children with New-Onset Celiac Disease. *PLoS One* **2022**, *17*, doi:10.1371/JOURNAL.PONE.0262108.

32. Mosca, A.; Leclerc, M.; Hugot, J.P. Gut Microbiota Diversity and Human Diseases: Should We Reintroduce Key Predators in Our Ecosystem? *Front Microbiol* **2016**, *7*, 455, doi:10.3389/FMICB.2016.00455/BIBTEX.

33. Gonçalves, P.; Araújo, J.R.; di Santo, J.P. A Cross-Talk Between Microbiota-Derived Short-Chain Fatty Acids and the Host Mucosal Immune System Regulates Intestinal Homeostasis and Inflammatory Bowel Disease. *Inflamm Bowel Dis* **2018**, *24*, 558–572, doi:10.1093/IBD/IZX029.

34. Upadhyay, D.; Singh, A.; Das, P.; Mehtab, J.; Dattagupta, S.; Ahuja, V.; Makharia, G.K.; Jagannathan, N.R.; Sharma, U. Abnormalities in Metabolic Pathways in Celiac Disease Investigated by the Metabolic Profiling of Small

Intestinal Mucosa, Blood Plasma and Urine by NMR Spectroscopy. *NMR Biomed* **2020**, *33*, doi:10.1002/NBM.4305.

35. Siddiqui, Z.; Osayande, A.S. Selected Disorders of Malabsorption. *Prim Care* **2011**, *38*, 395–414, doi:10.1016/J.POP.2011.05.002.

36. Angelis, M. De; Rizzello, C.G.; Fasano, A.; Clemente, M.G.; Simone, C. De; Silano, M.; Vincenzi, M. De; Losito, I.; Gobbetti, M. VSL#3 Probiotic Preparation Has the Capacity to Hydrolyze Gliadin Polypeptides Responsible for Celiac Sprue Probiotics and Gluten Intolerance. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **2006**, *1762*, 80–93, doi:10.1016/J.BBADIS.2005.09.008.

37. Huttenhower, C.; Gevers, D.; Knight, R.; Abubucker, S.; Badger, J.H.; Chinwalla, A.T.; Creasy, H.H.; Earl, A.M.; Fitzgerald, M.G.; Fulton, R.S.; et al. Structure, Function and Diversity of the Healthy Human Microbiome. *Nature* **2012**, *486*, 207, doi:10.1038/NATURE11234.

38. Hu, C.; Rzymski, P. Non-Photosynthetic Melainabacteria (Cyanobacteria) in Human Gut: Characteristics and Association with Health. *Life* **2022**, *12*, doi:10.3390/LIFE12040476.

39. Dreyfus, D.H. Herpesviruses and the Microbiome. *J Allergy Clin Immunol* **2013**, *132*, 1278–1286, doi:10.1016/J.JACI.2013.02.039.

40. Jankowiak, C.; Ludwig, D. [Frequent Causes of Diarrhea: Celiac Disease and Lactose Intolerance]. *Med Klin (Munich)* **2008**, *103*, 413–424, doi:10.1007/S00063-008-1061-8.

41. Johansson, M.E.V.; Sjövall, H.; Hansson, G.C. The Gastrointestinal Mucus System in Health and Disease. *Nat Rev Gastroenterol Hepatol* **2013**, *10*, 352–361, doi:10.1038/NRGASTRO.2013.35.

42. Chen, W.; Zhang, Z.; Zhang, S.; Zhu, P.; Ko, J.K.S.; Yung, K.K.L. MUC1: Structure, Function, and Clinic Application in Epithelial Cancers. *Int J Mol Sci* **2021**, *22*, doi:10.3390/IJMS22126567.

43. Sevinc, E.; Haluk Akar, H.; Sevinc, N.; Arslan, D.; Sezgin, G.C.; Kendirci, M. Amino Acid Levels in Children with Celiac Disease. *Nutr Hosp* **2015**, *32*, 139–143, doi:10.3305/nh.2015.32.1.9066.

44. Araújo, W.L.; Martins, A.O.; Fernie, A.R.; Tohge, T. 2-Oxoglutarate: Linking TCA Cycle Function with Amino Acid, Glucosinolate, Flavonoid, Alkaloid, and Gibberellin Biosynthesis. *Front Plant Sci* **2014**, *5*, 552, doi:10.3389/FPLS.2014.00552/BIBTEX.

45. Torrell, H.; Canela, N.; Cereto-Massagué, A.; Kazakova, P.; García, L.; Palacios, H. Multiomic Approach to Analyze Infant Gut Microbiota: Experimental and Analytical Method Optimization. *Biomolecules 2021, Vol. 11, Page 999* **2021**, *11*, 999, doi:10.3390/BIOM11070999.

46. Bokulich, N.A.; Kaehler, B.D.; Rideout, J.R.; Dillon, M.; Bolyen, E.; Knight, R.; Huttley, G.A.; Gregory Caporaso, J. Optimizing Taxonomic Classification of Marker-Gene Amplicon Sequences with QIIME 2's Q2-Feature-Classifier Plugin. *Microbiome* **2018**, *6*, 90, doi:10.1186/s40168-018-0470-z.

47. Lu, J.; Rincon, N.; Wood, D.E.; Breitwieser, F.P.; Pockrandt, C.; Langmead, B.; Salzberg, S.L.; Steinegger, M. Metagenome Analysis Using the Kraken Software Suite. *Nat Protoc* **2022**, *17*, 2815–2839, doi:10.1038/S41596-022-00738-Y.

48. Westreich, S.T.; Treiber, M.L.; Mills, D.A.; Korf, I.; Lemay, D.G. SAMSA2: A Standalone Metatranscriptome Analysis Pipeline. *BMC Bioinformatics* **2018**, *19*, 1–11, doi:10.1186/S12859-018-2189-Z/TABLES/2.

SUPPLEMENTARY INFORMATION

S1: Identified metabolites with concentration and statistical analis.

metabolite	Mean	Desvest	SEM	Mean	Desvest	SEM	T test
Valerate	0,472	0,178	0,178	0,415	0,165	0,048	0,441
Isoleucine	3,465	1,232	1,232	3,664	1,522	0,439	0,742
Propionate	5,518	5,179	5,179	4,229	2,048	0,591	0,438
Lactate	4,291	9,526	9,526	1,373	0,453	0,131	0,302
Alanine	1,346	0,492	0,492	1,133	0,350	0,101	0,248
Butyrate	4,956	2,797	2,797	5,694	2,009	0,580	0,480
Acetate	15,073	10,529	10,529	10,750	4,840	1,397	0,218
Glutamine	2,499	0,892	0,892	2,921	1,060	0,306	0,328
Succinate	0,548	0,792	0,792	0,187	0,084	0,024	0,133
Methylamine	0,268	0,106	0,106	0,240	0,093	0,027	0,568
Methionine	0,618	0,194	0,194	0,602	0,224	0,065	0,863
Methionine sulfoxide	0,109	0,062	0,062	0,084	0,029	0,008	0,209
Aspartate	0,937	0,556	0,556	0,616	0,400	0,115	0,130
Trimethylamine	0,082	0,043	0,043	0,061	0,028	0,008	0,155
Choline	0,055	0,021	0,021	0,042	0,025	0,007	0,212
Glycine	1,788	1,430	1,430	1,011	0,407	0,118	0,086
Glucose	1,658	1,065	1,065	1,621	1,157	0,334	0,938
Uracil	0,295	0,232	0,232	0,445	0,228	0,066	0,145
Fumarate	0,031	0,021	0,021	0,109	0,076	0,022	0,005
Tyrosine	0,434	0,152	0,152	0,459	0,193	0,056	0,752
Tryptophan	0,130	0,066	0,066	0,149	0,086	0,025	0,584
Formate	0,053	0,015	0,015	0,057	0,027	0,008	0,608
Valine	4,006	1,449	1,449	4,302	1,854	0,535	0,686
Leucine	1,624	0,576	0,576	1,706	0,704	0,203	0,769
Dimethyl Sulfone	0,172	0,076	0,076	0,158	0,076	0,022	0,651
Butyrate	4,720	3,106	3,106	5,173	2,256	0,651	0,697
Glycerol	3,880	3,859	3,859	0,974	0,477	0,138	0,017
Dimethylamine	0,038	0,019	0,019	0,026	0,012	0,004	0,087
Lysine	1,736	0,635	0,635	1,908	0,769	0,222	0,571
Malonate	0,596	0,351	0,351	0,371	0,155	0,045	0,059
Threonine	1,700	1,702	1,702	1,068	0,373	0,108	0,225

S2: diet administered to rats as a function of time

	Day 1 to 4	Day 5 to 6	Day 7 to 11	Day 12 to 18	Day 19 to 21	Day 22 to 28
Volume per offspring (ul)	12,5	30	50	200	400	600
The stock of gliadins (ul)	366,67	880	1067	4267	8533	12800
Condensed milk (ul)	183,33	440	533	2133	4266	6400

UNIVERSITAT ROVIRA I VIRGILI
A MULTI-OMICS APPROACH TO DESCRIBE CROSSTALK BETWEEN GUT MICROBIOTA AND HOST: EPIGENETICS AND VIROME ROLE
Polina Kazakova Borodina

GENERAL DISCUSSION

UNIVERSITAT ROVIRA I VIRGILI
A MULTI-OMICS APPROACH TO DESCRIBE CROSSTALK BETWEEN GUT MICROBIOTA AND HOST: EPIGENETICS AND VIROME ROLE
Polina Kazakova Borodina

GENERAL DISCUSSION

The human gut microbiome's diversity and richness play a crucial role in maintaining overall health. Numerous studies have established a correlation between childhood factors that affect the gut microbiome, immunological development, and disease outcomes in adulthood. While it is well known that microbiome members and host communities engage in transkingdom interactions [1], the majority of published studies have primarily focused on bacterial communities. However, other cofactors, such as the virome [2–4], miRNAs [5], and metabolome, are also highly correlated with the bacterial population and host health [6]. Therefore, these aspects must be taken into account when examining how the gut microbiome affects human health.

The present doctoral thesis aims to provide a comprehensive understanding of the interaction between the microbiota and its host during early life stages and in the context of childhood CeD. Specifically, the thesis aims to achieve the following objectives: 1) to describe and correlate the microbiome (including bacteriome and virome) and epigenetic factors (such as small RNAs) of meconium, 2) to assess the effects of type of feeding in 4-month-old infants and compare them to the meconium microbiome, 3) to investigate the differences in the intestinal ecosystem of celiac children on a GFD versus healthy children, and 4) to evaluate the impacts of a probiotic cocktail on the intestinal phenotypes in an animal model that mimics CeD. By attaining these goals, this thesis aims to advance our knowledge of the intricate interactions between the host and microbiota in both healthy and pathological stages.

To analyze infant gut microbiota, we first optimized an experimental and analytical method using a multiomic approach (Manuscript 1). For more accurate comparison between studies, it is important to standardize the

analytical and bioinformatics protocols. While there are few studies that thoroughly analyze meconium samples, our validated analysis using 16S rRNA sequencing of the bacteriome confirmed the presence of bacteria in meconium. This finding challenges the widely accepted idea of a sterile womb paradigm and suggests that factors in the early stages of development can impact the microbiota, with potential long-lasting effects on health and well-being [7–9]. Our study showed that the meconium microbiome was mainly represented by the phyla Proteobacteria and Firmicutes, followed by the phyla Bacteroidetes and Actinobacteria, consistent with previous results [10,11].

In addition to the bacteriome, virus-like particles (VLPs) were detected and taxonomically identified. Like the bacteriome, there was a low interindividual similarity and a low richness of species in the VLPs. Interestingly, we observed a high abundance of crAssphages in the meconium group compared to milk-fed infants, consistent with previous evidence suggesting that crAssphage abundance increases with age and may be vertically transmitted during delivery from the mother [12]. Additionally, we identified several bacteriophages exclusive to milk-fed infants. However, it is worth noting that the current understanding of the virome is still limited, and several limitations in our protocol, as well as in previous studies, should be considered [13].

Therefore, we tested different methodologies to determine the optimum protocol for VLP enrichment. During this process, we identified two critical points in the workflow. First, it is crucial to eliminate any host and bacterial cells, as they contain large amounts of genetic material that can contaminate virome libraries. However, it is important to minimize bias as much as possible during this process [14,15]. Second, we had to carry out separate protocols for DNA and RNA virome analysis to include all viral populations. Once the genomic data were created, we faced a major limitation in the form of “viral dark matter” sequences, which did not share homology with any sequences in databases and accounted for approximately 90% of the

sequenced virome. This is a common issue in virome studies and has been reported previously [16]. Our findings suggest that the virome is incredibly diverse, with the majority of its sequences remaining unexplored.

To gain a better understanding of meconium composition, we conducted a smallRNome analysis, which, to our knowledge, is the first study to describe the small RNA profile in meconium samples (Manuscript 2). Our analysis revealed significant differences in small RNA composition between meconium samples and those from milk-fed infants. We found that many of these miRNAs are associated with different signaling pathways, such as fatty acid biosynthesis and metabolism, cell cycle, and TGF-beta signaling pathways, all of which have been linked to infant birth weight [17]. Additionally, two miRNAs belonging to the miR-30 family, which are known to regulate development and disease [18], were found to be upregulated in meconium samples compared with milk-fed infants.

Our results from the three omics studies support the vertical transmission of the microbiome to the infant during birth [19].

To continue studying the effects of the gut microbiome on the host and to explore the impacts of epigenetic factors and the virome on the bacteriome in early stages of life, we proposed the second aim of this thesis: to examine the influences of different types of diet in nursing infants using a multiomic approach (Manuscript 1 and Manuscript 2). Our bacteriome findings showed that the microbiota changes rapidly during the first months of life. Although the same four phyla (Proteobacteria, Firmicutes, Bacteroidetes, and Actinobacteria) predominated in milk-fed infants, there was higher species richness in formula-fed infants. We also observed a significant decrease in the Enterobacteriaceae family in breastfed infants, accompanied by an increase in the Bacteriodaceae family, which is partially consistent with previous studies that used a similar workflow [20].

In contrast with previous studies on the virome, which identified numerous human viruses, our study resulted in low identification of this type of virus.

We suspect that this may be due to the enrichment protocol and the bias that occurs during filtration [4,21]. Most of the identified sequences were assigned to bacteriophages of the Caudovirales order, which is consistent with previous characterizations of the gut virome [22,23]. To provide evidence of the transkingdom interplay between the virome and bacteriome and to demonstrate that viruses modulate their bacterial hosts through different mechanisms, we correlated these two omics tools. Our findings supported the idea that populations of phages and hosts oscillate with time and that bacteriophages have been shown to drive host diversification both theoretically and experimentally [24–26]. Specifically, positive correlations were observed between a bacteriophage and its host in our study.

We did not find any significant differences in small RNA expression when comparing breastfed infants to formula-fed infants.

The fecal metabolome provides a functional readout of microbial activity and can be interpreted as an intermediate phenotype mediating host–microbiome interactions [27]. Using nontargeted NMR, we identified ten metabolites that differed between the two groups of milk-fed infants. Breastfed infants showed lower levels of butyrate, isoleucine and choline, agreeing with the previous report of production of choline by the gut microbiome [28]. We also found a positive correlation between butyrate and the Firmicutes phylum, which is consistent with the fact that SCFA-producing species belong to this phylum [29]. Additionally, aspartate was positively correlated with four bacterial orders (Staphylococcales, Clostridiales, Lachnospirales, and Pasteurellales), some of which contain species that produce amino acids, such as Clostridium [30]. Taken together, these results suggest that SCFAs positively impact the gut microbiome and play an important role in maintaining host homeostasis.

Our study continuously investigated the potential link between early fecal microbiota composition and disease risk in CeD. By examining this relationship, we hoped to gain a better understanding of the role that early

microbial colonization may play in the pathogenesis of CeD. Thus, we proposed the second hypothesis, in which children with CeD who adhere to a GFD may exhibit different bacterial and viral populations and gut metabolic profiles than healthy children. We further suggested that probiotic supplementation may help to restore these alterations. Finally, we aimed to compare the fecal bacteriome, virome, and metabolome of CeD children adhering to a GFD with those of healthy children to identify any significant differences (Manuscript 3).

In general, there were few discernible differences in the bacteriome between the two groups, but it is important to note that the sample size was small, which could limit the interpretation of the results. Despite this, both groups of children had similar levels of richness and similarity, and there were no major differences at different taxonomic levels. The most prevalent phyla in both groups were Bacteroidetes and Firmicutes, which together accounted for over 90% of the total abundance. However, the ratio of Firmicutes/Bacteroidetes was lower in healthy children, which is consistent with previous studies that suggested that a lower ratio is indicative of healthier individuals, since a higher ratio is associated with obesity or aging [31]. In the healthy group, only seven genera were present, with three of them (Prevotella, Lachnospira, and Spirochaeta) typically associated with pathogenicity and commonly found in the microbiota of individuals with illness [32]. Similarly, the virome in CeD when adherent to a GFD presented lower evenness and a higher abundance of Enterobacteria phages, which is consistent with a previous study [33]. This indicates that a GFD may not restore the virome to a healthy state. Notably, the DNA virome analysis identified a large percentage of reads (over 30%) belonging to Siphoviridae, which includes 1,166 species [34]. The RNA virome analysis also showed a high abundance of Salmonella phage in both groups. These results suggest that there may be a tight correlation between the bacteriome and virome, particularly with respect to richness and evenness in both omics.

At a more functional level, the gut metabolome in CeD when adhered to a GFD was not fully restored, and some metabolic differences remained when compared to healthy controls. For instance, the concentration of fumarate was higher, and the concentration of glycerol was lower in the healthy group. Paying attention to the three main SCFAs involved in bacterial metabolism (acetate, butyrate, and propionate), previous studies showed alterations in propionate concentration in CeD patients [35]. However, we did not observe any differences between the two groups. Furthermore, those with CeD who adhered to a GFD showed alterations in glycerolipid and galactose metabolism, with higher concentrations of glycerol. Altered glycerolipid metabolism is related to certain disease conditions, such as human obesity [36]. Based on these results, it appears that children with CeD who adhere to a GFD may continue to experience malabsorption due to their condition. This finding is consistent with previous research [37].

The last objective of this thesis was to investigate the effects of probiotics on a CeD animal model by using a multiomics approach to evaluate the restoration of the animal phenotypes (Manuscript 3). First, our results confirm that the rats were able to mimic CeD, as evidenced by the histologic lesions observed in the ileum histopathological sections, which are consistent with previously reported characteristics of this enteropathy [38]. We measured three parameters (crypt depth, villus height and villus/crypt ratio) from histopathological sections of the ileum and observed improvement in these parameters in CeD animals treated with probiotics, which made them similar to the control group and not the CeD group. Second, metagenomic analysis results disagreed with previous reports indicating that gut microbiome richness decreases in some pathologies [39]. In our study, the control group had the lowest richness among the three studied groups. The most significant difference was observed in virus composition, as the samples from the control group were highly clustered, indicating a distinct viral diversity pattern. In the probiotic group, there were no differences observed in the eight bacterial strains that make up the

administered probiotic compared to the other two groups. Thus, the probiotic cocktail did not appear to colonize the intestine or modulate the overall diversity of the intestinal microbiota. Bacterial and viral richness and evenness between the CeD treated with probiotic group and CeD-induced group did not differ.

To address discrepancies between phenotype and metagenomics, we conducted a functional study using metatranscriptomics and metabolomics. However, the metatranscriptome results did not distinguish between the induced CeD group and CeD group treated with probiotics, while the control group could be clearly differentiated. This was complicated by a high percentage of reads assigned to hypothetical proteins (80% in the CeD and CeD treated with probiotic groups and 30% in the control group), hindering the analysis and interpretation of the results. In contrast, our results for cecal metabolome profiling indicated an inflammatory process in induced celiac animals, as evidenced by an elevated concentration of seven amino acids as previously reported [40]. This finding supports the pathophysiological results of CeD mimicry in rats. Interestingly, the induced CeD samples were less clustered than the two other groups, indicating that the probiotic treatment returned the samples to a similar grouping. Several metabolites differentiated between the induced CeD group and group of CeD treated with probiotics were involved in energy generation pathways such as glycolysis and gluconeogenesis. This could be a symptom of malabsorption.

In general, probiotic supplementation was found to be insufficient in completely restoring the alterations in rats that mimic CeD. While some level of restoration was observed at the histopathological level, there was no restoration of dysbiosis in the microbiome. Additionally, the restoration at the functional level was only partial, particularly with regard to metabolic function.

Considering all the results of this thesis, they confirm the rapid evolution of the microbiome during the early stages of life and demonstrate the important roles of various components in maintaining host homeostasis. Consistent with previous studies, we found that the virome is strongly correlated with the bacteriome and may also be linked to different metabolites [41]. Nonetheless, it is worth noting that the protective role of breastfeeding against CeD may not be a direct consequence of gut microbiota but rather an indirect consequence. One proposed theory suggests that intestinal infections can disrupt intestinal homeostasis and increase intestinal permeability, potentially contributing to the development of celiac disease [42]. However, in this context it is also crucial to consider the potential roles of the microbiota in inhibiting viral infections and stimulating the host immune system [1].

It should be noted, however, that further standardization of virome studies is needed to allow for better comparison and reproducibility of the results. Currently, differences in VLP enrichment and bioinformatic analysis can limit the ability to compare and reproduce study findings.

Finally, the multiomics approach provides a unique opportunity to investigate the complex network of interactions between microbiota and host. The study of multiomics offers a compelling method for gaining a deeper understanding of the mechanisms underlying alterations in host-microbiome interactions and identifying ways to restore them to a healthy state. Recent advancements in multiomics have opened a window into the functionality and activity of the intestinal microbiota. However, it is essential to acknowledge the fundamental challenges, such as the volume and complexity of data, that require examination. Despite these challenges, the multiomics approach continues to be a useful method for improving our comprehension of the microbiome and how it affects human health.

REFERENCES

1. Belkaid, Y.; Hand, T.W. Role of the Microbiota in Immunity and Inflammation. *Cell* **2014**, *157*, 121, doi:10.1016/J.CELL.2014.03.011.

2. Bikel, S.; Valdez-Lara, A.; Cornejo-Granados, F.; Rico, K.; Canizales-Quinteros, S.; Soberón, X.; Del Pozo-Yauner, L.; Ochoa-Leyva, A. Combining Metagenomics, Metatranscriptomics and Viromics to Explore Novel Microbial Interactions: Towards a Systems-Level Understanding of Human Microbiome. *Comput Struct Biotechnol J* **2015**, *13*, 390–401, doi:10.1016/J.CSBJ.2015.06.001.

3. Iorio, A.; Biazzo, M.; Gardini, S.; Muda, A.O.; Perno, C.F.; Dallapiccola, B.; Putignani, L. Cross-Correlation of Virome-Bacteriome-Host-Metabolome to Study Respiratory Health. *Trends Microbiol* **2022**, *30*, 34–46, doi:10.1016/J.TIM.2021.04.011.

4. Liang, G.; Bushman, F.D. The Human Virome: Assembly, Composition and Host Interactions. *Nature Reviews Microbiology 2021 19:8* **2021**, *19*, 514–527, doi:10.1038/s41579-021-00536-5.

5. Louwies, T.; Johnson, A.C.; Orock, A.; Yuan, T.; Greenwood-Van Meerveld, B. The Microbiota-Gut-Brain Axis: An Emerging Role for the Epigenome. *Exp Biol Med* 2020, *245*, 138–145.

6. Nyholm, L.; Koziol, A.; Marcos, S.; Botnen, A.B.; Aizpurua, O.; Gopalakrishnan, S.; Limborg, M.T.; Gilbert, M.T.P.; Alberdi, A. Holo-Omics: Integrated Host-Microbiota Multi-Omics for Basic and Applied Biological Research. *iScience* **2020**, *23*, 101414, doi:10.1016/J.ISCI.2020.101414.

7. Yassour, M.; Jason, E.; Hogstrom, L.J.; Arthur, T.D.; Tripathi, S.; Siljander, H.; Selvenius, J.; Oikarinen, S.; Hyöty, H.; Virtanen, S.M.; et al. Strain-Level Analysis of Mother-to-Child Bacterial Transmission during the First Few Months of Life. *Cell Host Microbe* **2018**, *24*, 146-154.e4, doi:10.1016/J.CHOM.2018.06.007.

8. Shi, Y.C.; Guo, H.; Chen, J.; Sun, G.; Ren, R.R.; Guo, M.Z.; Peng, L.H.; Yang, Y.S. Initial Meconium Microbiome in Chinese Neonates Delivered Naturally or by Cesarean Section. *Sci Rep* **2018**, *8*, doi:10.1038/S41598-018-21657-7.

9. Hansen, R.; Scott, K.P.; Khan, S.; Martin, J.C.; Berry, S.H.; Stevenson, M.; Okpapi, A.; Munro, M.J.; Hold, G.L. First-Pass Meconium Samples from Healthy Term Vaginally-Delivered Neonates: An Analysis of the Microbiota. *PLoS One* **2015**, *10*, doi:10.1371/JOURNAL.PONE.0133320.

10. Del Chierico, F.; Vernocchi, P.; Petrucca, A.; Paci, P.; Fuentes, S.; Praticò, G.; Capuani, G.; Masotti, A.; Reddel, S.; Russo, A.; et al. Phylogenetic and Metabolic Tracking of Gut Microbiota during Perinatal Development. *PLoS One* **2015**, *10*, doi:10.1371/JOURNAL.PONE.0137347.

11. Hu, J.; Nomura, Y.; Bashir, A.; Fernandez-Hernandez, H.; Itzkowitz, S.; Pei, Z.; Stone, J.; Loudon, H.; Peter, I. Diversified Microbiota of Meconium Is Affected by Maternal Diabetes Status. *PLoS One* **2013**, *8*, e78257, doi:10.1371/JOURNAL.PONE.0078257.

12. Gregory, A.C.; Zablocki, O.; Zayed, A.A.; Howell, A.; Bolduc, B.; Sullivan, M.B. The Gut Virome Database Reveals Age-Dependent Patterns of Virome Diversity in the Human Gut. *Cell Host Microbe* **2020**, *28*, 724-740.e8, doi:10.1016/J.CHOM.2020.08.003.

13. Garmaeva, S.; Sinha, T.; Kurilshikov, A.; Fu, J.; Wijmenga, C.; Zhernakova, A. Studying the Gut Virome in the Metagenomic Era: Challenges and Perspectives. *BMC Biol* **2019**, *17*, doi:10.1186/S12915-019-0704-Y.

14. Santiago-Rodriguez, T.M.; Hollister, E.B. Potential Applications of Human Viral Metagenomics and Reference Materials: Considerations for Current and Future Viruses. *Appl Environ Microbiol* **2020**, *86*, 1–12, doi:10.1128/AEM.01794-20/ASSET/CEC893D2-8C34-40D1-88E8-E1AC74389FEA/ASSETS/GRAPHIC/AEM.01794-20-F0005.JPEG.

15. Jansen, D.; Matthijnssens, J. The Emerging Role of the Gut Virome in Health and Inflammatory Bowel Disease: Challenges, Covariates and a Viral Imbalance. *Viruses* **2023**, *15*, doi:10.3390/V15010173.

16. Santiago-Rodriguez, T.M.; Hollister, E.B. Unraveling the Viral Dark Matter through Viral Metagenomics. *Front Immunol* **2022**, *13*, 5508, doi:10.3389/FIMMU.2022.1005107/BIBTEX.

17. Rodil-Garcia, P.; Arellanes-Licea, E. del C.; Montoya-Contreras, A.; Salazar-Olivo, L.A. Analysis of MicroRNA Expression in Newborns with Differential Birth Weight Using Newborn Screening Cards. *Int J Mol Sci* **2017**, *18*, doi:10.3390/IJMS18122552.

18. Mao, L.; Liu, S.; Hu, L.; Jia, L.; Wang, H.; Guo, M.; Chen, C.; Liu, Y.; Xu, L. MiR-30 Family: A Promising Regulator in Development and Disease. *Biomed Res Int* **2018**, *2018*, doi:10.1155/2018/9623412.

19. Avershina, E.; Lundgård, K.; Sekelja, M.; Dotterud, C.; Storrø, O.; Øien, T.; Johnsen, R.; Rudi, K. Transition from Infant- to Adult-like Gut Microbiota. *Environ Microbiol* **2016**, *18*, 2226–2236, doi:10.1111/1462-2920.13248.

20. Wang, Z.; Neupane, A.; Vo, R.; White, J.; Wang, X.; Marzano, S.Y.L. Comparing Gut Microbiome in Mothers' Own Breast Milk- and Formula-Fed Moderate-Late Preterm Infants. *Front Microbiol* **2020**, *11*, doi:10.3389/FMICB.2020.00891.

21. Aggarwala, V.; Liang, G.; Bushman, F.D. Viral Communities of the Human Gut: Metagenomic Analysis of Composition and Dynamics. *Mob DNA* **2017**, *8*, 1–10, doi:10.1186/s13100-017-0095-y.

22. Zuo, T.; Lu, X.J.; Zhang, Y.; Cheung, C.P.; Lam, S.; Zhang, F.; Tang, W.; Ching, J.Y.L.; Zhao, R.; Chan, P.K.S.; et al. Gut Mucosal Virome Alterations in Ulcerative Colitis. *Gut* **2019**, *68*, 1169–1179, doi:10.1136/gutjnl-2018-318131.

23. Shkoporov, A.N.; Hill, C. Bacteriophages of the Human Gut: The "Known Unknown" of the Microbiome. *Cell Host Microbe* **2019**, *25*, 195–209, doi:10.1016/j.chom.2019.01.017.

24. Carding, S.R.; Davis, N.; Hoyles, L. Review Article: The Human Intestinal Virome in Health and Disease. *Aliment Pharmacol Ther* **2017**, *46*, 800–815, doi:10.1111/apt.14280.

25. Rodriguez-Valera, F.; Martin-Cuadrado, A.B.; Rodriguez-Brito, B.; Pašić, L.; Thingstad, T.F.; Rohwer, F.; Mira, A. Explaining Microbial Population Genomics through Phage Predation. *Nat Rev Microbiol* **2009**, *7*, 828–836, doi:10.1038/NRMICRO2235.

26. Clokie, M.R.J.; Millard, A.D.; Letarov, A. V.; Heaphy, S. Phages in Nature. *Bacteriophage* **2011**, *1*, 31, doi:10.4161/BACT.1.1.14942.

27. Deng, K.; Xu, J. jian; Shen, L.; Zhao, H.; Gou, W.; Xu, F.; Fu, Y.; Jiang, Z.; Shuai, M.; Li, B. yan; et al. Comparison of Fecal and Blood Metabolome Reveals Inconsistent Associations of the Gut Microbiota with Cardiometabolic Diseases. *Nature Communications 2023 14:1* **2023**, *14*, 1–12, doi:10.1038/s41467-023-36256-y.

28. Chávez-Carbajal, A.; Nirmalkar, K.; Pérez-Lizaur, A.; Hernández-Quiroz, F.; Ramírez-Del-Alto, S.; García-Mena, J.; Hernández-Guerrero, C. Gut Microbiota and Predicted Metabolic Pathways in a Sample of Mexican Women Affected by Obesity and Obesity Plus Metabolic Syndrome. *Int J Mol Sci* **2019**, *20*, doi:10.3390/IJMS20020438.

29. Louis, P.; Flint, H.J. Formation of Propionate and Butyrate by the Human Colonic Microbiota. *Environ Microbiol* 2017, *19*, 29–41.

30. Manjasetty, B.A.; Chance, M.R.; Burley, S.K.; Panjikar, S.; Almo, S.C. Crystal Structure of Clostridium Acetobutylicum Aspartate Kinase (Ca AK): An Important Allosteric Enzyme for Amino Acids Production. *Biotechnol Rep (Amst)* **2014**, *3*, 73–85, doi:10.1016/J.BTRE.2014.06.009.

31. Gagliardi, A.; Totino, V.; Cacciotti, F.; Iebba, V.; Neroni, B.; Bonfiglio, G.; Trancassini, M.; Passariello, C.; Pantanella, F.; Schippa, S. Rebuilding the Gut Microbiota Ecosystem. *Int J Environ Res Public Health* **2018**, *15*, doi:10.3390/ijerph15081679.

32. de Angelis, M.; Vannini, L.; di Cagno, R.; Cavallo, N.; Minervini, F.; Francavilla, R.; Ercolini, D.; Gobbetti, M. Salivary and Fecal Microbiota and Metabolome

of Celiac Children under Gluten-Free Diet. *Int J Food Microbiol* **2016**, *239*, 125–132, doi:10.1016/J.IJFOODMICRO.2016.07.025.

33. el Mouzan, M.; Assiri, A.; al Sarkhy, A.; Alasmi, M.; Saeed, A.; Al-Hussaini, A.; AlSaleem, B.; al Mofarreh, M. Viral Dysbiosis in Children with New-Onset Celiac Disease. *PLoS One* **2022**, *17*, doi:10.1371/JOURNAL.PONE.0262108.

34. Mayneris-Perxachs, J.; Castells-Nobau, A.; Arnoriaga-Rodríguez, M.; Garre-Olmo, J.; Puig, J.; Ramos, R.; Martínez-Hernández, F.; Burokas, A.; Coll, C.; Moreno-Navarrete, J.M.; et al. Caudovirales Bacteriophages Are Associated with Improved Executive Function and Memory in Flies, Mice, and Humans. *Cell Host Microbe* **2022**, *30*, 340-356.e8, doi:10.1016/J.CHOM.2022.01.013.

35. Zafeiropoulou, K.; Nichols, B.; Mackinder, M.; Biskou, O.; Rizou, E.; Karanikolou, A.; Clark, C.; Buchanan, E.; Cardigan, T.; Duncan, H.; et al. Alterations in Intestinal Microbiota of Children With Celiac Disease at the Time of Diagnosis and on a Gluten-Free Diet. *Gastroenterology* **2020**, *159*, 2039-2051.e20, doi:10.1053/J.GASTRO.2020.08.007.

36. Prentki, M.; Madiraju, S.R.M. Glycerolipid Metabolism and Signaling in Health and Disease. *Endocr Rev* **2008**, *29*, 647–676, doi:10.1210/ER.2008-0007.

37. Siddiqui, Z.; Osayande, A.S. Selected Disorders of Malabsorption. *Prim Care* **2011**, *38*, 395–414, doi:10.1016/J.POP.2011.05.002.

38. Laparra, J.M.; Olivares, M.; Gallina, O.; Sanz, Y. Bifidobacterium Longum CECT 7347 Modulates Immune Responses in a Gliadin-Induced Enteropathy Animal Model. *PLoS One* **2012**, *7*, doi:10.1371/journal.pone.0030744.

39. Mosca, A.; Leclerc, M.; Hugot, J.P. Gut Microbiota Diversity and Human Diseases: Should We Reintroduce Key Predators in Our Ecosystem? *Front Microbiol* **2016**, *7*, 455, doi:10.3389/FMICB.2016.00455/BIBTEX.

40. Sevinc, E.; Haluk Akar, H.; Sevinc, N.; Arslan, D.; Sezgin, G.C.; Kendirci, M. AMINO ACID LEVELS IN CHILDREN WITH CELIAC DISEASE. *Nutr Hosp* **2015**, *32*, 139–143, doi:10.3305/NH.2015.32.1.9066.

41. Duan, J.; Wang, W.; Jiang, T.; Bai, X.; Liu, C. Viral Metagenomics Combined with Metabolomics Reveals the Role of Gut Viruses in Mouse Model of Depression. *Front Microbiol* **2022**, *13*, doi:10.3389/FMICB.2022.1046894.

42. Marasco, G.; Di Biase, A.R.; Schiumerini, R.; Eusebi, L.H.; Iughetti, L.; Ravaioli, F.; Scaioli, E.; Colecchia, A.; Festi, D. Gut Microbiota and Celiac Disease. *Dig Dis Sci* **2016**, *61*, 1461–1472, doi:10.1007/S10620-015-4020-2.

CONCLUSIONS

UNIVERSITAT ROVIRA I VIRGILI
A MULTI-OMICS APPROACH TO DESCRIBE CROSSTALK BETWEEN GUT MICROBIOTA AND HOST: EPIGENETICS AND VIROME ROLE
Polina Kazakova Borodina

CONCLUSIONS

1. The potential utility of metagenomics to advance the understanding of microbiota functions and connections to host homeostasis is increased by the complementarity of the metagenome with other omics data, such as metabolite profiling.

2. We confirmed the existence of miRNA in meconium samples, pointing to the possibility of prenatal epigenetic processes that interact with the host intestinal system to modulate and regulate its homeostasis.

3. Our results supported the hypothesis of vertical transmission during delivery, especially considering the high abundance of crAssphage in meconium samples.

4. Our study also indicates that the miRNome profile, similar to the virome, rapidly changes in the first months of life, reflecting the dynamic and malleable nature of the developing gut microbiota. This highlights the importance of early life interventions to promote health and prevent future illnesses.

5. Transkingdom connections between bacteria and virome components illustrate the complex interactions within the microbiota and the need to consider multiple levels of microbial community interactions when studying host-microbial balance.

6. The GFD does not completely restore the gut microbiome of CeD children, but the differences may also be due to other dietary factors.

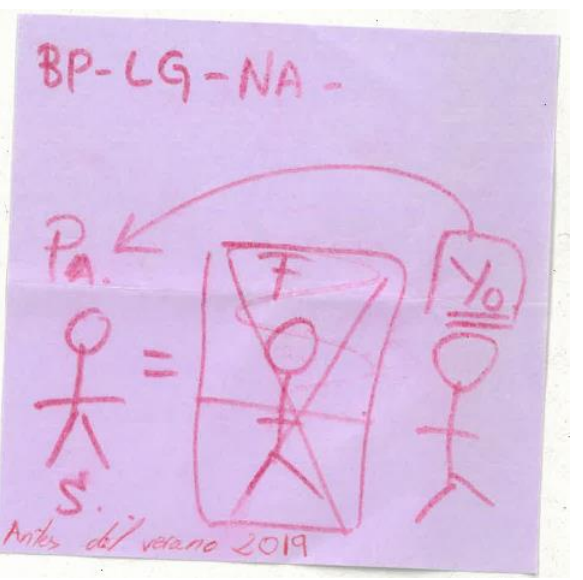
However, this suggests that a GFD alone may not be enough to fully restore gut homeostasis in CeD patients.

7. The administration of VSL#3 in CeD-induced rats does not fully restore gut homeostasis, but it does improve damage to the ilium at the pathophysiological level.

8. A multiomic approach is a powerful tool for studying the relationship between taxonomic and functional alterations in the gut holobiont. This approach allows for a more comprehensive understanding of the complex interactions within the microbiota and their impacts on host health.

9. The major limitation of many emerging omics technologies, including viromics, is the necessity for improved bioinformatic analysis and databases. Additionally, standardizing both analytical and bioinformatics protocols is imperative to enable reliable comparisons of results across various studies.

AKNOWLEDGMENTS



LIST OF PUBLICATIONS

Papers included in the thesis:

A) Published papers

Torrell, H.; Cereto-Massagué, A.; Kazakova, P.; García, L.; Palacios, H.; Canela, N. Multiomic Approach to Analyze Infant Gut Microbiota: Experimental and Analytical Method Optimization. Biomolecules 2021, 11, 999. https://doi.org/10.3390/biom11070999

Kazakova P, Abasolo N, de Cripan SM, Marquès E, Cereto-Massagué A, Garcia L, Canela N, Tormo R, Torrell H. Gut Microbiome and Small RNA Integrative-Omic Perspective of Meconium and Milk-FED Infant Stool Samples. International Journal of Molecular Sciences. 2023; 24(9):8069. https://doi.org/10.3390/ijms24098069C)

UNIVERSITAT ROVIRA I VIRGILI
A MULTI-OMICS APPROACH TO DESCRIBE CROSSTALK BETWEEN GUT MICROBIOTA AND HOST: EPIGENETICS AND VIROME ROLE
Polina Kazakova Borodina

LIST OF CONFERENCE PAPERS

World of Microbiome: Pregnancy, Birth and Infancy & Digestive and Metabolic. 2020. GUT VIROME AND BACTERIOME IN EARLY INFANCY P. Kazakova, L. Garcia, A. Cereto, N. Canela, H. Torrell; Poster Presentation.

UNIVERSITAT ROVIRA I VIRGILI
A MULTI-OMICS APPROACH TO DESCRIBE CROSSTALK BETWEEN GUT MICROBIOTA AND HOST: EPIGENETICS AND VIROME ROLE
Polina Kazakova Borodina