UNIVERSITÀ DEGLI STUDI DI SASSARI
INTERNATIONAL PhD SCHOOL IN BIOMOLECULAR
AND BIOTECHNOLOGICAL SCIENCES

Curriculum: Microbiology and immunology

XXVIII CICLO

Director: Prof. Leonardo A. Sechi

Development of pipeline to create and validate metagenomic datasets enabling microbiota associations with host traits

Tutor: Prof. Sergio Uzzau

PhD student: Valeria Manghina

La presente tesi è stata prodotta nell'ambito della scuola di dottorato in Biomolecular and Biotechnological Sciences dell'Università degli Studi di Sassari, a.a. 2014/2015 – XXVIII ciclo, con il supporto di una borsa di studio finanziata con le risorse del P.O.R. SARDEGNA F.S.E. 2007-2013 - Obiettivo competitività regionale e occupazione, Asse IV Capitale umano, Linea di Attività l.3.1.

Table of contents

List of tables

List of figures

Paragraph 5.2.2

Paragraph 5.3.1

Paragraph 5.4

Paragraph 5.5

Paragraph 5.6

Paragraph 5.7

Abstract

The microbiota can be depicted as a measurable organ consisting of microbial cells, and creating a unique ecosystem together with the host eukaryotic cells. Therefore, to understand the normal physiology and pathology of animal ecosystems, it is mandatory to tackle a comprehensive analysis of the host, the microbiota, and their interactions. Since over 99.8% of the microbes cannot be cultured, metagenomics offers a path to the study of their community structures and metabolic potential. These goals are achievable thanks to the recent advances in sequencing complex assortments of small genomes, through 16S and WGS approach. In spite of the great number of latest studies, there is not a defined and standardized pipeline to measure the microbial community; rather, significant efforts are needed to optimize sample preparation and data analysis workflows for metagenomics analysis of microbiome.

In the present PhD Thesis, a major task is aimed at developing rapid and efficient workflows for metagenomics analysis. These include the choice of the sample, the extraction methods more efficient according to sample features, choice of sequencing approaches and the statistical methods.

A developed workflow was successfully applied to investigate to the first time the sheep gut microbiome, to perform association studies linking gut microbiota to T1D host traits in mice and to assess dietary and immunogenetic background impact over gut microbiota in a translational murine model of NAFLD.

Chapter 1

Introduction

1.1 Microbiome: definitions

The concept of the human microbiome was first suggested by Joshua Lederberg, who coined the term ‘‘microbiome, to signify the ecological community of commensal, symbiotic, and pathogenic microorganisms that literally share our body space’’ (Lederberg and McCray 2001). The human body, both outside and inside, plays host to innumerable microbes. The microbial communities that colonize the human body help to protect us from pathogens, to digest food, and provide nutrients, signaling molecules affecting host functions, and a large array of antigens continuously challenging the immune system. It has been estimated that the human microbiome consists of up to 100 trillion cells, 10-fold the number of human cells. The microbes in the human body encode over 100 times the number of genes compared to genes contained within the human genome (Bäckhed *et al.*, 2005; Ley *et al.*, 2006a). The human microbiota may include bacteria, fungi, archaea, protist, and viruses, distributed throughout the human body (i.e. oral cavity, urogenital tract, skin, and gut). Given their associated functions and wide distribution, human microbiota have a significant influence on the physiology, nutrition, and immunity of the human body. Thus, a detailed understanding of the genetic basis of system biology in human requires knowledge of both the human genome and the associated microbial metagenome. Disruptions of these human-associated microbial communities or the imbalance of the relationships between these microbes and human host can therefore lead to severe health problems. To this extent, increasing attention is paid to the connection between the dynamics of the human microbiome and human health, from infectious disease to metabolic homeostasis.

1.2 The Human Microbiome Project

The HMP was launched by the National Institutes of Health (NIH) in the fall of 2007. One of the main aims of HMP was to use sequencing to examine the microbes associated with the human body. Its main purpose is to create resources for the research community, with a focus on building a "healthy cohort" reference database of human microbiome genome sequences (known as metagenomic sequences), computational tools to analyze complex metagenomic sequences, and clinical protocols for sampling the human microbiome. The "healthy cohort" project is a sequencing study of the microbiome based on sampling from 5 major body sites (18 subsites): nasal passages, oral cavities, skin, gastrointestinal (GI) tract, and urogenital tract. The body sites were selected by a panel of experts in human microbiology. The study recruited 300 adults (of whom half were women and half were men) who were clinically verified to be free of overt disease. About 20 percent of the study participants self-identified as a racial minority and 10 percent as Hispanic. Each participant was sampled up to three times over a 2-year period. Two kinds of sequencing data were collected: microbial taxonomic characterization using the 16S ribosomal ribonucleic acid (rRNA) marker gene and sequence data from entire microbial communities (i.e., metagenomic sequences). When the project started, the goal was to sequence 1,000 reference genomes. Today, the goal is to sequence 3,000 reference genomes. Results from the first 178 genomes and 550,000 genes sequenced were published in 2010. These sequences, representing two kingdoms (Bacteria and Archaea), nine phyla, 18 classes, and 24 orders, were distributed among the gastrointestinal tract, the urogenital/vaginal tract, the skin, the oral cavity, and the respiratory tract (Nelson *et al.*, 2010). In addition to the healthy cohort project, the HMP is managing a series of demonstration projects to evaluate associations between the microbiome and disease in the major body sites (i.e. eczema, Crohn's disease,

necrotizing enterocolitis, inflammatory bowel disease (IBD), ulcerative colitis and bacterial vaginosis). Additionally, the project is accumulating clinical and phenotype data associated with the healthy cohort sequencing data.

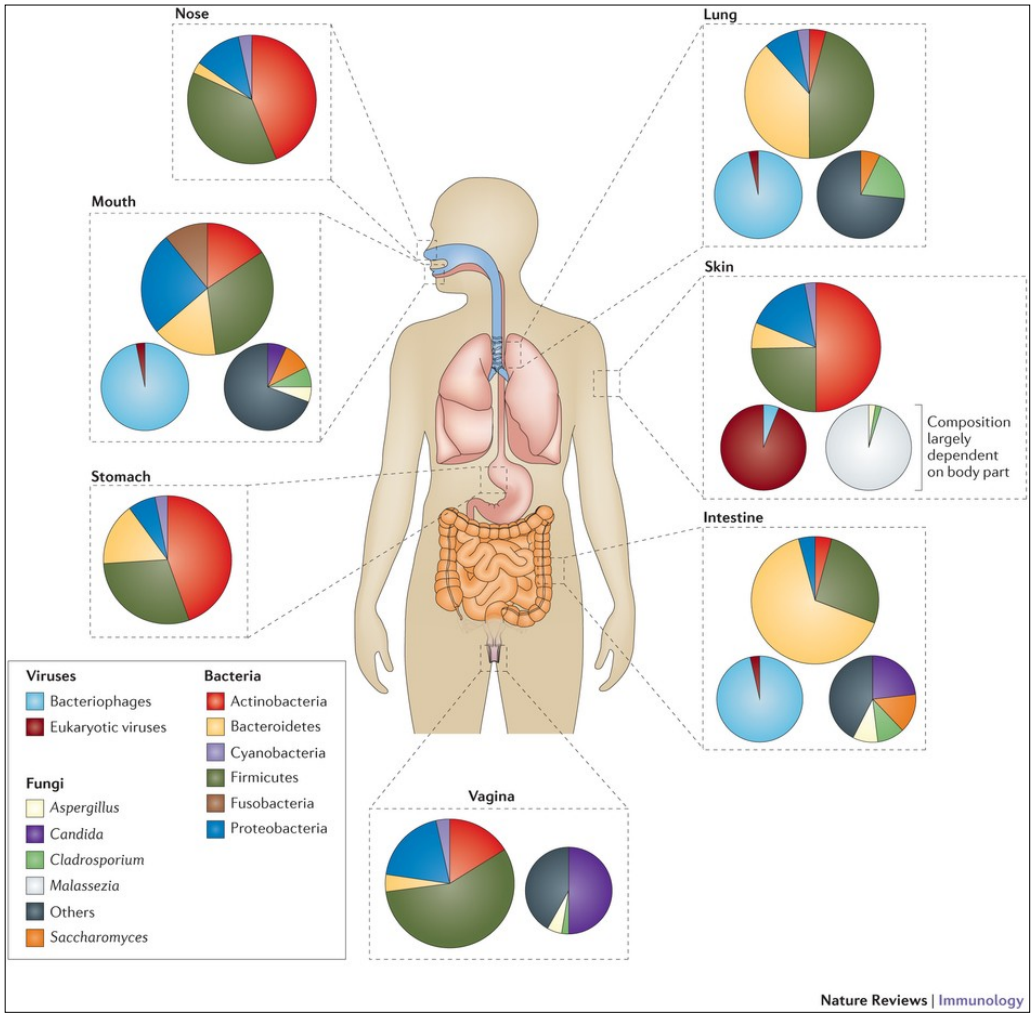

Figure 1. The figure shows the relative abundance of bacterial, fungal and viral communities at different body sites

1.3 Gut microbiome

The human gut is the habitat for a diverse and dynamic microbial ecosystem. Each individual has a unique signature gut flora, composed of some 100–1000 microbial species, composed predominantly of bacterial species (Tap *et al*., 2009, Qin *et al*., 2009). Gut microbiota and host have a symbiotic relationship due to their co-existence and co-evolution. On one hand, gut microbiota depend on the host for their growth and survival. On the other hand, most of the gut microbiota are non-pathogenic and can benefit the host in many ways: extraction of nutrients and energy from diet intake (Sonnenburg *et al*., 2005; Yatsunenko *et al*., 2012; Kau AL *et al*, 2011), protection from enteropathogen invasion (Fukuda *et al*., 2011), contribution to the development of a normal immune system or function (Olszak T *et al*., 2012; Ahern PP *et al*., 2014). In contrast, the imbalance between the gut microbiota and the host has been associated with many diseases. Colonization of the gastrointestinal (GI) tract starts at birth and evolves and changes over a lifetime, such that the adult human GI tract is home to a unique ecosystem of several billion bacteria. The infant's gut is first colonized by maternal and environmental bacteria during birth and continues to be populated through feeding and other contacts (Sekirov *et al*., 2010). Factors known to influence colonization include gestational age, mode of delivery (vaginal birth vs c-section), diet (breast milk vs formula) and exposure to antibiotics (Marques *et al*., 2010; Fouhy *et al*., 2012). The intestinal microbiota of newborns is characterized by low diversity and a relative dominance of the phyla Proteobacteria and Actinobacteria; thereafter, the microbiota becomes more diverse with the dominance of Firmicutes and Bacteroidetes, which characterizes the adult microbiota (Qin *et al*.,2010). In particular, in the human adult host, the proximal small intestine contains relatively small numbers of bacteria in healthy subjects (O'Hara *et al*., 2006). The microbiology of the terminal ileum represents a transition zone between the jejunum, containing

predominantly aerobic species, and the dense population of anaerobes found in the colon. Bacterial colony counts may be as high as 10^9 colony-forming units (CFU)/mL in the terminal ileum immediately proximal to the ileocecal valve, with a predominance of gram-negative organisms and anaerobes. On crossing into the colon, the bacterial load and variety of the enteric flora change dramatically. Loads of 10^{12} CFU/mL or greater may be found and are comprised mainly of anaerobes such as *Bacteroides*, *Porphyromonas*, *Bifidobacterium*, *Lactobacillus* and *Clostridium*, with anaerobic bacteria outnumbering aerobic bacteria by a factor of 100 to 1000:1. The predominance of anaerobes in the colon reflects the fact that oxygen concentrations in the colon are very low; the flora has simply adapted to survive in this "hostile" environment. While the gut microbiota evolves with age and varies in composition along the length of the GI tract, the microbiota composition of adult host is, in general, constant in time (Vanhoutte *et al*., 2004, Seksik *et al*., 2003). However, the gut microbiota composition is influenced by diet, socioeconomic conditions and health/disease status; indeed, imbalance of the gut microbiota, aka dysbiosis, can predispose individuals to a variety of disease states ranging from gut intrinsic disorders such as inflammatory bowel diseases (Nell *et al*., 2010), Crohn's disease and ulcerative colitis, and colonic cancer (Scanlan *et al*., 2008, Arthur *et al*., 2012) to systemic diseases such as allergic diseases (Kuitunen *et al*., 2009, McLoughlin *et al*., 2011) and metabolic syndromes such as obesity (Turnbaugh *et al*., 2009, Ley *et al*., 2010), diabetes (Wen *et al*., 2008, Musso *et al*., 2011, Qin *et al*., 2012), arteriosclerotic diseases (Koeth *et al*., 2013, Wang *et al*., 2011), and nonalcoholic steatohepatitis (NASH) (Henao-Mejia *et al*., 2012, Abu-Shanab *et al*., 2010). For these reasons, a current goal is to characterize the human microbiota, enabling the study of its variation according to factors such as population, genotype, disease status and profile, age, nutrition, as well as exposure to various medications, and dietary factors. Worldwide, scientific and commercial interest in the cross-talk between microbes and their human hosts

has also fueled by the recognition that the intestinal microbiota plays a pivotal role in many aspects of human health and disease.

1.4 Host-Microbiota interactions and diseases

In recent years, concerted efforts to identify, describe, and quantify the bacterial communities of the mammalian gastrointestinal tract have begun to shed the first lights. In this contest, the role of the microbiota in regulation of host energy balance and metabolism can be studied at various levels. The composition of the bacterial flora can then be monitored simultaneously with changes in the animal's body weight. One very powerful tool to address this question is germ-free mice raised in an environment completely devoid of bacteria. By simply comparing the physiology of germ-free mice with that of conventionally raised animals, one can obtain useful information about how bacteria can shape host metabolism. Experiments have shown that germ-free animals seem to be protected from diet-induced obesity (Backhed F. *et al*., 2007). However, this protective effect was later shown to be strongly dependent on the sugar compositions, that is type of sugar of these diets, not just the amount of fat-derived calories (Fleissner *et al*., 2010). This may well be because germ-free animals lack the bacterial enzymes needed to digest polysaccharides, leading, therefore, to a lower calorie intake. Complex polysaccharides are processed in the gut and fermenting microbes produce short-chain fatty acids (SCFAs). Butyrate, propionate, and acetate are SCFAs that can be directly used by colonocytes as an energy source or be further transported to the liver where they can be used as substrates for lipid synthesis (Wolever *et al*., 1991; Scheppach *et al.,*1994). Given these observations, a series of experiments were performed to show that the ability of carbohydrate fermentation together with the production of SCFAs was linked to induction of obesity. When germ-free mice were

colonized either by whole bacterial flora (Backhed F. *et al*., 2004) or by saccharolitic fermenting bacteria such as *Bacteroides thetaiotamicron* together with *Methanobrevibacter smithii*, which facilitates fermentation (Samuel BS *et al*.,2008), an increase in body weight and adiposity was observed. In other studies, introduction of HFD in conventionally raised animals was accompanied by a shift toward fermenters in gut flora. Animals on HFD had a microbial community that was characterized by a general decrease in microbial diversity and a phylogenetic shift from Bacteroidetes toward Firmicutes (Turnbaugh *et al*., 2004). This could further be attributed to an extensive bloom of one of the families within Firmicutes, namely *Erysipelotrichaceae* (Turnbaugh *et al*., 2008; Fleissner *et al*., 2010). Interestingly, Turnbaugh and collaborators have demonstrated that the phenotype was somewhat transferable: wild-type germfree mice colonized with flora from *ob/ob* mice (Turnbaugh *et al*., 2004) as well as mice that received HFD (Turnbaugh *et al*., 2008) were better at storing fat compared to those colonized with a normal flora. Of note, gut microbiota play an important role in the regulation of autoimmunity (Wu HJ et al., 2012; Longman et al., 2013) and its involvement has been suggested in the development of T1D as early as 1987 (Suzuki T *et al*., 1987). Following on from this, experiments using NOD mice that were transferred from specific pathogen-free (SPF) conditions to germ free (GF) conditions showed a marked change in insulitis and the incidence confirmed the role of gut microbiota as a regulator of islet-specific autoimmunity (Alam C *et al*., 2011; Wen L *et al*., 2008). The first gut microbiota study in humans for T1D compared the microbiome between four Finnish children with T1D and four age and HLA-DQ-matched healthy children (Giongo A *et al*., 2011; Brown CT *et al.*, 2011); subsequently, in the follow-up study, those Finnish children who developed T1D had a decreased ratio of Firmicutes vs Bacteroidetes, supporting a cross-sectional study showing that Bacteroidetes were more abundant in islet-specific autoantibody-positive children than in autoantibody negative children (de Goffau *et al.*, 2013; de

Goffau *et al*., 2014). Furthermore, heightened gut permeability has been demonstrated to be one of the phenomena that precede the clinical onset of T1D in animal models of autoimmune diabetes, as well as in patients with T1D and prediabetic individuals (Neu J *et al*., 2005; Vaarala *et al*., 2008). Evidence from animal studies has been largely derived from two rodent models: NOD mice and the BioBreeding diabetes-prone (BBDP) rat. It has been suggested that the imbalance of bacteria, such as Bacteroidetes, which ferment short-chain fatty acid (SCFA), can affect the gut permeability. Indeed, in parallel to the changed gut permeability, BBDP rats, before clinic onset, have a different gut bacterial composition from that of diabetes-resistant (BBDR) rats, with relatively higher abundance of Bacteroides sp. in diabetic rat (Schwartz RF *et al*., 2007; Brugman S et al., 2006). At disease onset, the gut bacterial profile was also different between BBDP and BBDR rats (Roesch LF *et al*., 2009). Specifically, the BBDP rats had a lower proportion of the probiotic-like bacteria, such as Bifidobacterium and Lactobacillus, but had higher numbers of Bacteroides, Ruminococcus and Eubacterium (Roesch LF *et al*., 2009). At the cellular level, there were also structural changes in the intestinal morphology accompanying the increased permeability in BBDP rats (Neu J *et al*., 2005; Graham S *et al*., 2004; Watts T *et al*., 2005) and at the molecular level, the expression of multiple tight junction proteins was down- or up-regulated, in both BBDP rats and T1D patients, thus affecting the gut permeability, including occludin, members of claudin family and zonulin (Sapone A *et al.*, 2006; Watts T *et al*., 2005; Vaarala *et al*., 2008).

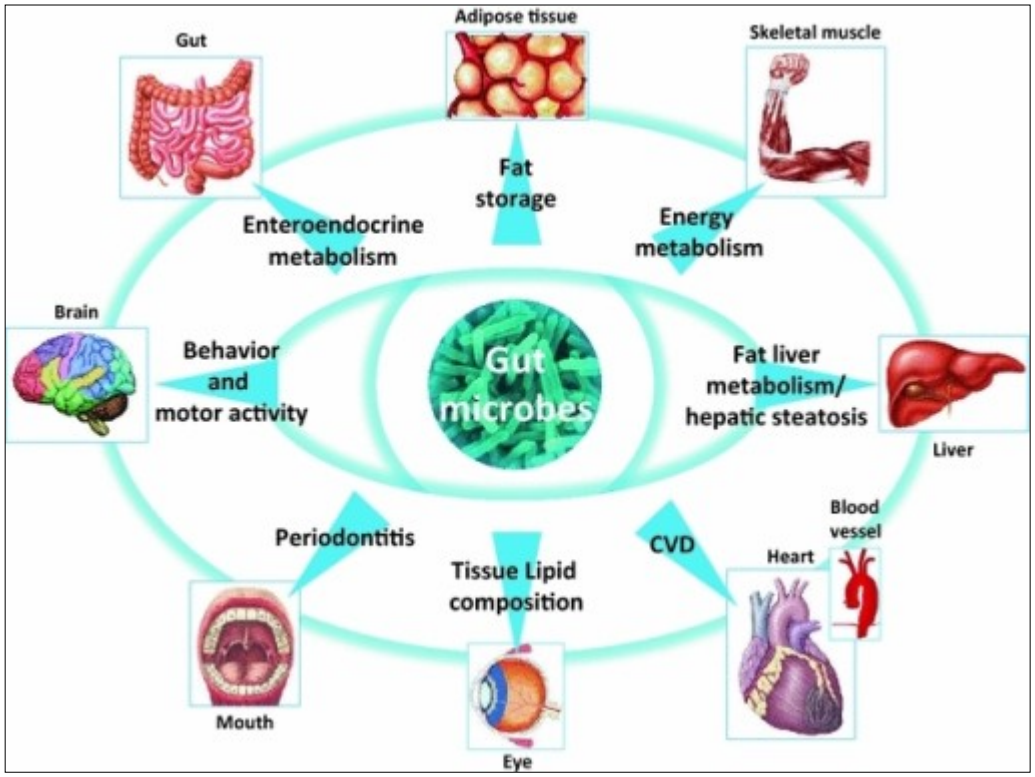

Figure 2. Gut microbiome and host interactions

Another aspect is related to the importance of the gut microbiota in the healthy development of the mammalian immune system (Brestoff *et al*., 2013). Direct evidence comes from germ-free mice, in which multiple defects in the gut immune system have been noted, including impaired development of gut-associated lymphoid tissue (GALT) (Bouskra *et al*., 2008), generation of colonic regulatory T cells (Atarashi K *et al*., 2011) and production of IgA (Moreau MC *et al*., 1978). Importantly, the profound effects that commensal microbiota have on immunity is not limited to the gut immune system, but extends to the systemic immune response (Mazmanian *et al*., 2005). Germ-free mice have an elevated IgE level and overall are skewed toward Th2 immune responses, which can be normalized by exposure to a diverse microbiota during early life (Cahenzli J *et al*., 2013). Although the various mechanisms by which gut microbiota regulate host immunity are, as yet poorly elucidated, a couple of mechanisms have been proposed: directly activating the innate immune response

through Toll-like receptors (TLR) by molecular patterns from gut bacteria (Kingma SD *et al*., 2011; Oh JZ *et al*., 2014; Round JL *et al*., 2011) or modulating immune responses via G-protein-coupled receptor (GPCR) by bacterially derived metabolites (Maslowski KM *et al*., 2009; Thangaraju M *et al*., 2009). The activation of innate immune responses by gut microbiota-derived molecular patterns mostly depend on bacteria cell wall components such as flagellin (Oh JZ *et al*., 2014; Vijay-Kumar M *et al*., 2010), LPS and polysaccharide A (PSA) (Round JL *et al*., 2011). This suggests that malfunction of the innate immune system may promote the development of metabolic syndrome through modification of the gut bacterial profile. More surprisingly, gut dysbiosis has been implicated in chronic metabolic disorders such as the non-alcoholic fatty liver disease (NAFLD) (Dumas ME *et al*., 2006; Wang *et al*., 2011). NAFLD encompasses a spectrum of hepatic pathologies. Accumulation of triglycerides in hepatocytes (hepatic steatosis) is the most common liver phenotype in NAFLD. Some individuals with hepatic steatosis develop nonalcoholic steatohepatitis (NASH), a more severe type of liver damage characterized by hepatic inflammation and liver cell death. Urinary metabolites of NAFLD-susceptible mice, when fed HFD, were enriched in microbiota-derived methylamines: dimethylamine, trimethylamine (TMA), and trimethylamine-N-oxide (TMAO). TMA is synthesized exclusively by symbiotic bacteria (Al-Waiz *et al*., 2007) and can be further transformed into TMAO by the microbiota themselves or at least in humans by the liver enzyme FMO3. The production of methylamines by microbiota results in decreased bioavailability of choline for the host and seems to trigger NAFLD in mice. This idea was further explored by Wang *et al*., who could link high levels of TMAO in plasma with increased risk for cardiovascular disease in humans (Wang *et al*., 2011). Furthermore, as accumulating literatures underpin the importance of the gut microbiome to intestinal functions, a novel concept of microbiome–gut–brain axis has been evolved. The gut receives regulatory signals from the Central

Nervous System (CNS) and vice versa. The term gut–brain-axis thus describes an integrative physiology concept that incorporates all, including afferent and efferent neural, endocrine, nutrient, and immunological signals between the CNS and the gastrointestinal system (Romijn *et al*., 2008). CNS can influence gut microbiome through neural and endocrine pathways in both direct and indirect manners. The autonomic nervous system (ANS) and hypothalamus–pituitary–adrenal (HPA) axis that liaise the CNS and viscera can modulate gut physiology such as motility, secretion and epithelial permeability as well as systemic hormones, which in turn affects the niche environment for microbiota and also host-microbiome interaction at the mucosae (Cryan and Dinan, 2012). The influence of microbiome on CNS functions is manifested in both normal and disease conditions. There is a crucial link between gut microbiome and CNS maturation under physiological state. External cues derived from indigenous commensal microbiota affect prenatal and postnatal developmental programming of the brain (Al-Asmakh *et al*., 2012; Douglas-Escobar *et al*., 2013). As multiple mechanisms guide the impact of microbiome on the CNS, it is therefore of particular interest to explore the role of microbiome in the regulation of CNS disorders. While there is still a lack of epidemiological evidence to connect microbiome with CNS pathologies, accumulating studies have underscored the importance of microbiome in a range of CNS disorders, as multiple sclerosis, anxiety, stress, depression and autism (Ochoa-Reparaz *et al*., 2011).

Chapter 2

Methods to study the microbiome

2.1 Metagenomics

Uncultured microorganisms comprise the majority of the planet's biological diversity. Microorganisms represent two of the three domains of life and in many environments, as many as 99% of the microorganisms cannot be cultured by standard techniques. Therefore, culture-independent methods are essential to understand the genetic diversity, population structure, and ecological roles of the majority of microorganisms. In this contest, the advent of high-throughput next-generation sequencing (NGS) has revolutionized the field of microbial ecology and brought classical environmental studies to another level. In fact, this type of technology has led to the establishment of the field of "metagenomics", defined as the direct genetic analysis of genomes contained within an environmental sample without the prior need for cultivating clonal cultures. Currently the term is also widely applied to studies performing amplification of certain genes of interest (marker gene amplification metagenomics), but initially, was only used for functional and sequence-based analysis of the collective microbial genomes contained in an environmental sample (full shotgun metagenomics) (Xia *et al*., 2011; Handelsman *et al*., 2009; Riesenfeld *et al.*, 2004).

	16S rRNA (amplicon sequencing)	Shotgun sequencing
Type of information produced	The taxonomic composition and phylogenetic structure of a microbial community expressed as Operational taxonomic units (OTUs).	Functional and process-level characterization of microbial communities as a whole, and the reconstruction of draft genome sequences for individual community members.
Application	Monitor bacterial populations.	Detect new members, new genes, and resolve complex taxonomies.
Ability to detect rare members of the community (sensitivity)	Highly sensitive. rRNA makes up 80% of total bacterial RNA .	Requires much deeper sequencing to achieve the same level of sensitivity.

Table 1. Main features of amplicon and shotgun sequencing

Such methodologies allow a much faster and elaborative genomic/genetic profile generation of an environmental sample at a very acceptable cost.

Full shotgun metagenomics has the capacity to fully sequence the majority of available genomes within an environmental sample.This creates a community biodiversity profile that can be further associated with functional composition analysis of known and unknown organism lineages, i.e., genera or taxa (Tringe

et al., 2005). This approach has evolved to address the questions of *who* is present in an environmental community, *what they are doing*, and *how* these microorganisms interact to sustain a balanced ecological niche. It further provides unlimited access to functional gene composition information derived from microbial communities inhabiting practical ecosystems.

Marker gene metagenomics is a fast and gritty way to obtain a community/taxonomic distribution profile using PCR amplification and sequencing of evolutionarily conserved marker genes, such as the 16S rRNA gene (Tringe *et al.*, 2008). This taxonomic distribution can subsequently be associated with environmental metadata derived from the sampling site under investigation.

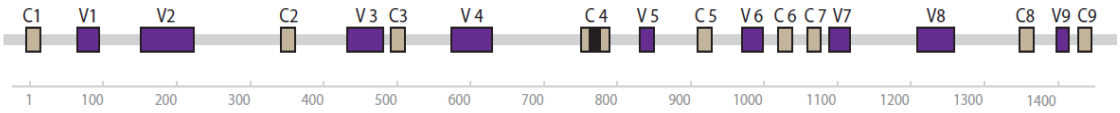

Figure 3. Schematic representation of the 16S rRNA gene. Location of variable (purple) and conserved (brown) regions in a canonical bacterial 16S rRNA. The black region is invariable in all bacteria

The use of 16S rRNA gene sequences to study bacterial phylogeny and taxonomy has been by far the most common housekeeping genetic marker used for a number of reasons. One of these include its presence in almost all bacteria, often existing as a multigene family, or operons; the function of the 16S rRNA gene over time has not changed, suggesting that random sequence changes are a more accurate measure of time (evolution); and the 16S rRNA gene (1,500 bp) is large enough for informatics purposes (Patel J B, 2001). The 16S rRNA gene includes interspersed conserved and nine hypervariable regions, which makes it well suited for PCR amplification and sequencing; in this process, probes are designed to hybridize to the conserved regions, allowing for amplification and sequencing of the variable regions. Focusing on a small part of the microbial genome lowers sequencing costs. When 16S rDNA sequences are available, it is

possible to cluster them in Operational Taxonomic Unit (OTU), according to a percent of similarity threshold (> 97%) in order to classify bacteria within the same, or different, OTUs. This approach has been particularly effective for monitoring fluctuations in microbial populations. Actually, for all these reasons, it has reached an extremely high level of reliability, becoming the most popular technique to perform taxonomic classification (Han *et al*., 2013; Nava and Stappenbeck, 2011; Santamaria *et al*., 2012).

2.2 Sequencing Technologies

454 Life Science and Illumina platforms for NGS

Two commonly used NGS technologies utilized to date are the 454 Life Sciences and the Illumina systems, with the ratio of usage shifting in favor of the latter recently.

The 454 pyrosequencer was the first next-generation sequencer to achieve commercial introduction in 2004 (Mardis *et al*., 2008). This DNA sequencing technique is based on the detection of released pyrophosphate (PPi) during DNA synthesis. In a cascade of enzymatic reactions, visible light is generated that is proportional to the number of incorporated nucleotides. The cascade starts with a nucleic acid polymerization reaction in which inorganic PPi is released as a result of nucleotide incorporation by polymerase. The release of pyrophosphate is conveyed into light using enzyme reactions, which is then converted into actual sequence information and because the added nucleotide is known, the sequence of the template can be determined. Its main difference from the classic Sanger sequencing is that pyrosequencing relies on the detection of pyrophosphate release on nucleotide incorporation rather than chain termination with dideoxynucleotides (Ronaghi *et al*., 2001). The 454 pyrosequencing technology can generate reads up to 1,000 bp in length and

~1,000,000 reads per run and this relatively long read length allows a significantly less error-prone assembly in shotgun metagenomics and permits greater annotation accuracy (Thomas *et al*., 2012; Wommack *et al*., 2008). The estimated cost of sequencing using 454 pyrosequencing technology is higher than other platforms such as Illumina (US $20 per Mb), and it has a relatively low coverage of 0.7 GB per sequencing run. Furthermore, noise generated by this 454 pyrosequencing technology affected different aspects of metagenomic data analysis and led to biased results.

In the last years, another sequencing service has become more popular, namely Illumina. Illumina sequencing generates many millions of highly accurate reads making it much faster and cheaper than other available sequencing methods. Illumina sequencing by synthesis is similar to Sanger sequencing, but it uses modified dNTPs containing a terminator which blocks further polymerization-so only a single base can be added by a polymerase enzyme to each growing DNA copy strand. The sequencing reaction is conducted simultaneously on a very large number (many millions in fact) of different template molecules spread out on a solid surface. The terminator also contains a fluorescent label, which can be detected by a camera. Only a single fluorescent color is used, so each of the four bases must be added in a separate cycle of DNA synthesis and imaging. Following the addition of the four dNTPs to the templates, the images are recorded and the terminators are removed. This chemistry is called "reversible terminators". Finally, another four cycles of dNTP additions are initiated. Since single bases are added to all templates in a uniform fashion, the sequencing process produces a set of DNA sequence reads of uniform length (Bentley DR *et al*., 2008; Mardis *et al*., 2008). Illumina is now the dominant vendor of high-throughput DNA sequencing machines. It has a variety of sequencing instruments dedicated to different applications. MiSeq, for example, has an output of 15 GB and 25 million sequencing reads of 300 bp in length;

clustered fragments can be sequenced from both ends (paired-end sequencing), which can be merged so that 600 bp reads can be obtained. Currently, Illumina sequencers has been developed in order to run smaller jobs at a much faster rate with relatively high throughput and they allow sample preparation sizes of < 20 ng DNA. When analyzing 16S metagenomics data, this technology obviates the need for time-consuming noise removal algorithms required for pyrosequencing and makes analysis less error prone (Handelsman., 2009). The greater coverage/yield generally offered by Illumina allows significant decrease of systematic errors. This advantage and the low cost (~US $0.50 per Mb) are the delineating factors that have turned Illumina into the preferred high-throughput sequencing technology for metagenomics studies.

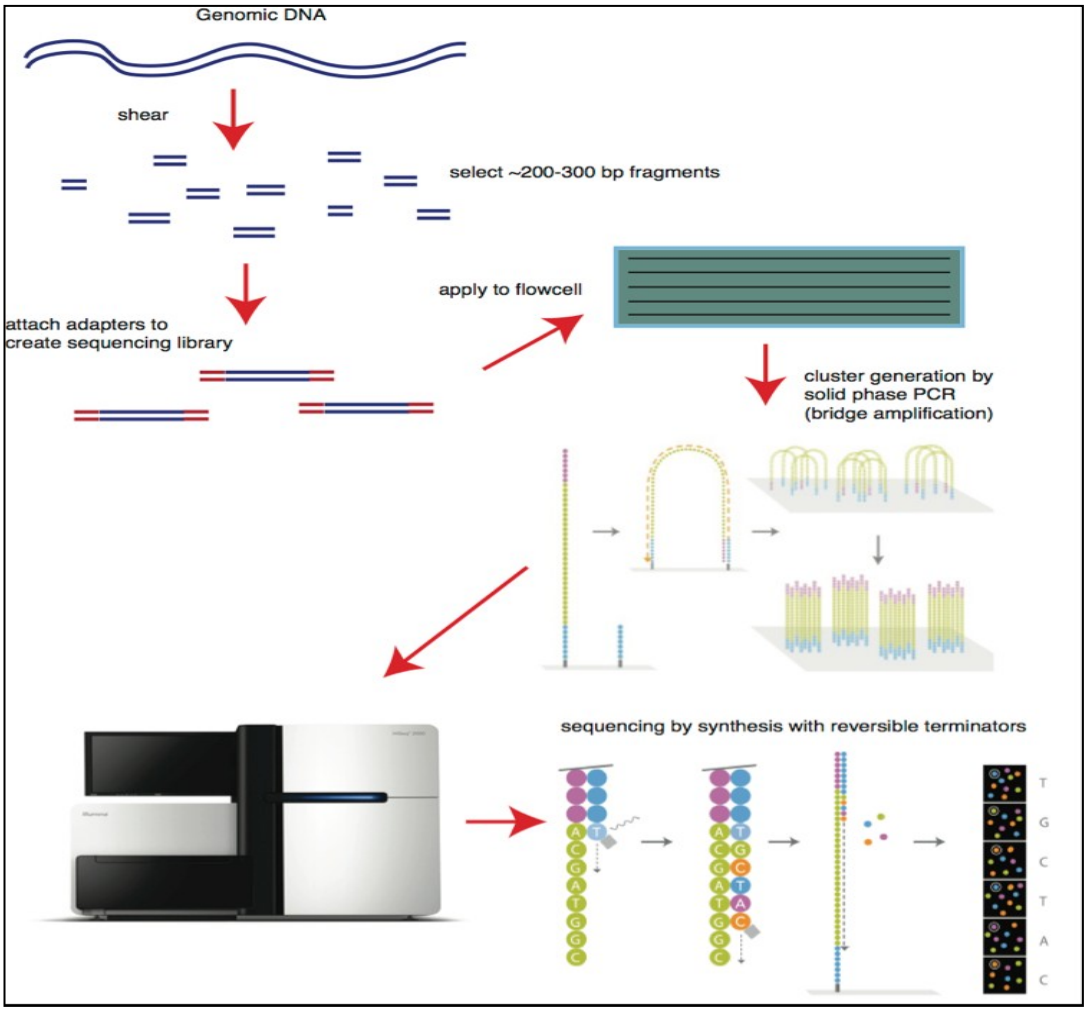

Figure 4. Illumina sequencing by synthesis

2.3 Other emerging sequencing technologies

Additional sequencing technologies are available and can potentially be used for metagenomic studies. These include the Applied Biosystems SOLiD, Pacific Biosciences, Ion Torrent and Oxford Nanopore technologies.

The Applied Biosystems SOLiD 5500 W Series sequencer offers higher coverage than 454 pyrosequencing but lower than Illumina (~120 GB per run); however, it can only guarantee a low error rate for sequencing reads of maximum 50 bp in length (Metzker *et al*., 2010). This reduces the possibility of generating a reliable and usable de novo assembly for shotgun metagenomics. An emerging sequencing technology that may have high impact on the fields of genomics and metagenomics was recently developed by Pacific Biosciences, PacBio (Metzker *et al*., 2010). This technology uses single-molecule real-time (SMRT) sequencing, which is a parallelized single-molecule DNA sequencing by synthesis. PacBio provides much longer read lengths (~10,000 bp) compared to the aforementioned technologies, thus having obvious advantages when addressing issues of annotation and assembly for shotgun metagenomics, but despite the high read length of PacBio, this technology is limited by high error rates and low coverage. In addition to the aforementioned technologies, which are based on optics, technologies such as Ion Torrent's semiconductor sequencing benchtop sequencer and Ion Proton are now coming into play. These technologies are based on the use of proton emission during polymerization of DNA in order to detect nucleotide incorporation. This system promises read lengths of > 200 bp and relatively high throughput, on the order of magnitude achieved by 454 Life Sciences systems. Additionally, it offers higher quality than 454, especially when sequencing homopolymers, but at a similar cost (about US $23 per Mb for the Ion Torrent PGM –314 Chip). An even more cutting-edge technology is currently under development by Oxford Nanopore technologies, which is developing "strand sequencing", a method of DNA

analysis that could potentially sequence completely intact DNA strands/polymers passed through a protein nanopore. This obviates the need for shotgun sequencing and aims to revolutionize the sequencing industry in the future. For metagenomics, this technology can have obvious advantages, as it will eliminate erroneous sequencing caused by shotgun metagenomics and exclude the need for the error-prone assembly step during data analysis; however, nanopore sequencing is at the moment non commercialized.

2.4 Metatranscriptomics

The metatranscriptome is the identity and quantity of a complete set of transcripts in a population of cells. While metagenomics tells us who is there and what they are capable of, based on their gene complement, metatranscriptomics tells us what they are doing at that moment. Neither primers nor probes are needed, so there is no need to anticipate important genes beforehand and transcripts from microbial assemblages are sequenced with little bias. Further, paralagous sequences which might crosshybridize on a microarray can be distinguished. The approach is particularly amenable to an experimental framework in which gene expression is monitored while a biotic or abiotic parameter is manipulated. Experimental metatranscriptomics is one of the most powerful tools for understanding the timing and regulation of complex microbial processes within communities and consortia, as well as microbial dexterity in response to changing conditions. Though metatranscriptomics is a small conceptual leap from metagenomics, practical considerations have slowed its development. One difficulty is that bacterial and archaeal mRNAs typically are not polyA tailed, so methods for specific capture of eukaryotic cDNAs are not applicable. This results in coextraction of the more abundant and stable rRNAs, which can lead to a disappointingly low yield of expressed gene sequences in a

large-scale sequencing run, potentially as low as 10% (Lamendella *et al*., 2012). Selectively removing rRNA from the total RNA pool or embracing the rRNA sequences for their insight into community structure (if amplification and other steps that could bias the rRNA pool are avoided) help mitigate this issue. Another technical challenge of working with RNA is a half-life on the order of minutes even under optimal conditions. On a conceptual level, there is not always a predictable relationship between mRNA abundance and protein activity, since genes can be constitutively expressed and enzyme activities can be regulated post-transcriptionally. Metaproteomics, a promising complementary technique, offers a better link to metabolic function but a less-resolved view of instantaneous regulatory responses.

2.5 Metaproteomics

Metaproteomics is the study of all protein samples recovered directly from environmental sources. This approach is able to provide details on the pathways that are actively functioning in a community, and on how the synthesis of specific proteins can change according to time, location, or environmental stimuli (Ottman *et al*., 2012). In particular, when analyzing a particular microbiome, it can be more important to know which functions are carried out by the microbial components present in a biological district, than which specific microbial species are present within. Different microorganisms can in fact perform the same function, thus a divergence in microbial composition between two samples is not always correlated to an equivalent altered microbial functionality. Although recent advances in DNA sequencing and proteomics technologies have opened the door to investigation of the structure and function of the gut microbiota without the necessity for cultivation, there have been very few efforts to date that have used a multi-‘‘omics’’ approach to study the

complex ecosystem in the gut. The ability to combine information about the identities of microbial community members (obtained from 16S rRNA gene-based measurements), metabolic potential (obtained from metagenome sequence data) and synthesis (obtained from metaproteome data) should enable exploration of the gut microbiota at multiple molecular levels simultaneously.

2.6 Critical issues in NGS technology

2.6.1 DNA extraction

Determining the bacterial community structure in fecal samples through amplification and sequence analysis of extracted DNA has revolutionized gastrointestinal microbiology research over recent years. These culture-independent techniques for assessing diversity have largely replaced traditional culture based approaches as they are considered to be less biased in terms of defining true diversity and considerably less labor-intensive (Amann *et al*., 1995). Due to the recent rapid increase in DNA-based phylogenetics of bacterial communities many different DNA extraction procedures are used, each with its own potential biases. In particular, the first step of DNA extraction - disruption and/or lysis of the bacterial membranes can be expected to be biased for specific bacterial taxa due to differences in cell wall structure and integrity of Gram-positive and Gram-negative bacteria. This step often involves bead-beating, thus most studies comparing methods of DNA extraction find that the major impact on the resulting measured community structure is caused by the use of bead-beating. Therefore, all methods rely on chemical or mechanical disruption, lysis using detergents, or a combination of these approaches. Previous studies have evaluated differences between DNA extraction methods from fecal samples, exploring detection with conventional PCR (Persson *et al*., 2011; McOrist *et al*., 2002), quantitative PCR (Nechvatal *et al*., 2008), bands on denaturing gradient

gel electrophoresis (DGGE) (Maukonen *et al*., 2012) and phylogenetic microarray (Salonen A. *et al*., 2010) Significant differences in relative abundance have been demonstrated when DNA was extracted using different methods from mock communities of bacteria and assessed by 16S rRNA sequencing (Milani *et al*., 2013). Wu *et al*. described the effect of different fecal extraction methods on 16S rRNA pyrosequencing, comparing QIAamp DNA Stool Minikit, MoBio PowerSoil DNA Isolation Kit and Stratec PSP Spin Stool DNA Kit (Wu *et al*., 2010). Therefore, the extraction of fecal DNA is a challenge, since feces not only contains bacterial and host cells but also many different substances derived from, for example, food, medicine, secondary cell metabolites etc. that can inhibit downstream PCR (Abu Al-Soud *et al*., 1998; Monteiro *et al*., 1997; Wilson *et al*., 1997). There is, therefore, the need to testing different protocols for DNA extraction for samples of different origin in order to obtain optimal results and avoid false interpretations on the bacterial compositions and bacterial diversity.

2.6.2 16S rRNA amplification

It is widely accepted that sequencing of the 16S rRNA gene reflects eubacterial evolution. Since the introduction of SSU rDNA-based molecular techniques, the study of microbial diversity in natural environments has advanced significantly. In addition, sequencing of the 16S rRNA gene has been widely applied in the field of microbial ecology and has resulted in a great number of sequences deposited in relevant databases, thus enhancing the value of 16S as the "gold standard" in microbial ecology. While the 16S rRNA gene fragment, containing one or more variable regions, is the preferred target marker gene for bacteria and archaea, this is not the case for fungi and eukaryotes where the preferred marker genes are the internal transcribed spacer (ITS) and 18S rRNA gene, respectively.

Taxonomic analysis for prokaryotes (i.e., bacteria and archaea) is regularly performed using 16S data derived from varying sequencing technologies (i.e., 454 pyrosequencing as well as Illumina, Solid and Ion Torrent). The protocols and methods used, however, vary considerably with regard to amplification primers, sequencing primers and sequencing technologies. Several studies have demonstrated that no single hypervariable region of 16S rRNA gene can differentiate among all bacteria. Clarridge showed that the initial 500-1500 bp of the 16S rRNA gene sequence was sufficient to discriminate among 100 bacteria (Clarridge, 2004); Chakravorty and colleagues have demonstrated that the hypervariable regions V2 (nucleotides 137-242), V3 (nucleotides 433-497) and V6 (nucleotides 986-1043) contain the maximum nucleotide heterogeneity and the maximum discriminatory power for the 110 bacterial species analyzed. Also, the hypervariable region V6 (986-1043) is the shortest hypervariable region with the maximum degree of sequence heterogeneity (Chakravorty *et al.,* 2007). Four hypervariable regions (V4, V5, V7 and V8) particularly V5, were less suitable for species identification due to a higher degree of sequence conservation compared to the other hypervariable regions. Furthermore, the results of Vasileiadis indicate that overall the most prominent hypervariable region for bacterial diversity studies was V3, even though it was outperformed in some of the tests. Despite its high performance during most tests, V4 had a reduced conservation of flanking sequence sites of the V region, V5 performed well in the non-redundant RDP database based analysis; however V5 did not resemble the full-length 16S rRNA gene sequence results as well as V3 and V4 did when the natural sequence frequency and occurrence approximation was considered in the virtual experiment. Although, the highly conserved flanking sequence regions of V6 provide the ability to amplify partial 16S rRNA gene sequences from very diverse owners, it was demonstrated that V6 was the least informative compared to the rest examined hypervariable regions (Vasileiadis *et al*., 2012).

In light of this background, it is important to study the most suitable approach during the pipeline definition.

2.6.3 16S Data analysis

There are different tools for 16S data analysis and denoising include QIIME (Caporaso *et al*., 2010), Mothur (Schloss *et al*., 2009), SILVAngs (Quast *et al*.,2013), MEGAN (Huson *et al*., 2007). After the demultiplexing of the dataset (the assignment of reads to samples using barcode information), the next step is OTU picking. For bacteria/archaea, it is accepted that OTUs of similarity greater than 97% correspond to the same species, but also other dissimilarity cut offs can be employed, if needed for the downstream analyses. There are numerous OTU picking strategies:

1) *De novo* is used if amplicons overlap and if a reference sequence collection is not available. It clusters all reads without using a reference and is quite expensive computationally, hence not very suitable for very large datasets.

2) Closed-reference is used if amplicons do not overlap and if a reference sequence collection is available. This approach discards reads that do not hit a reference sequence.

3) Open-reference is used if amplicons overlap and a reference dataset is available. This method clusters reads against a reference dataset, but if the reads do not match the reference, they are consequently clustered *de novo*. All the aforementioned are incorporated into QIIME. the most appropriate choice for the downstream analysis will depend on the type of data and the user. Taxonomic assignment of OTUs can be performed using a variety of algorithms. Currently QIIME supports numerous algorithms, such as BLAST, the RDP classifier, RTAX, Mothur classifier, and uclust, to search for the closest match

to an OTU from which a taxonomic lineage is inferred. This requires reference databases of marker genes. Some commonly utilized databases include: Greengenes (16S) (De Santis *et al*., 2006), Ribosomal Database Project (16S) (Cole JR *et al*., 2007), Silva (16S + 18S) (Quast *et al*., 2013; Pruesse *et al*., 2007).

2.6.4 QIIME

A very popular software for the analysis of microbial communities is the above mentioned QIIME (stands for Quantitative Insights Into Microbial Ecology). Initially QIIME was implemented for use of 454 pyrosequencing datasets only, i.e., using sff (Standard Flowgram Format) files, but currently QIIME has been modified to accept the fastq file format, thereby making the analysis of Illumina datasets possible. The QIIME developers provide users with extensive online tutorials for several workflows, and, moreover, QIIME is available as an open-source software package mostly implemented using the programming language PYTHON (Caporaso *et al*., 2010). It supports a wide range of microbial community analyses and visualizations that have been central to several recent high-profile studies, including network analysis, histograms of within or between sample diversity and analysis of whether 'core' sets of organisms are consistently represented in certain habitats. QIIME also provides graphical displays that allow users to interact with the data. This modularity allows alternative components for functionalities such as choosing operational taxonomic units (OTUs), sequence alignment, inferring phylogenetic trees and phylogenetic and taxon-based analysis of diversity within and between samples to be easily integrated and benchmarked against one another. QIIME output includes a representation of a taxonomic tree in Newick format, which can be visualized in applications such as FigTree, and a file in Biom (Biological Observation Matrix) format (McDonald D. *et al*., 2012) representing OTU tables. This file can be imported into MEGAN for visualization or into any other

statistical software requiring matrix-type data. In addition, alpha-diversity analysis (diversity within a sample, eg, Phylogenetic Diversity, Chao, etc.) and beta-diversity analysis (diversity across samples, eg, UniFrac, PCoA) (Lozupone *et al*.,2011), as well as taxonomic composition and phylogenetic analyses, are supported through QIIME (Kuczynski *et al*., 2011).

2.7 Final considerations on major challenges in the microbiome study

The past decade has clearly been a "golden age" for microbiome research. Extensive efforts to characterize the human microbiome, coupled with exciting advances in sequencing technologies and in computational techniques, have tremendously increased our knowledge about the diversity of the microbiome and about its composition in health and in disease, but numerous challenges are open again:

1. Modeling microbiome community dynamics. How do microbiomes change over time? What are the drivers of those changes? The environment? Health status? How does the prevalence of certain species affect other species in the community?

2. Linking microbiome function to community composition. How can the two different types of analyses (phylogenetic analysis and metabolic reconstruction) be linked so that more nuanced questions can be addressed? In other words, which organisms are responsible for which functions?

3. Integrating different types of -omics datasets. How can genomic data be integrated with transcriptomic, proteomic data integrated into a systems biology–level approach to studying these communities?

4. Correlating microbiome shifts with host phenotype. It can be very difficult to associate shifts in community composition (or functional state) with host phenotype when the phenotype in question is not well defined and when the impact of environmental change on that phenotype is unknown.

5. Definition of an univocal and efficiently analysis pipeline.

6. Meeting data volume and computational requirements. Development of an infrastructure for people to access available data, increase algorithm efficiency and reduce data redundancy.

Chapter 3

Aim of research proposal

In keeping with the considerations outlined in the introduction, the main objective of this project was the development of a workflow for gut microbiota characterization through the application of metagenomics.

Accordingly, the following aims were established:

- Development of robust methods of genomic DNA preparation from complex samples (i.e. stools) derived from different hosts;
- Comparative assessment of NGS technology approaches for the characterization of the gut microbiota (i.e., metagenomics vs 16S analyses);
- Assembling of bioinformatic pipelines for metagenomic data analysis;
- Pipeline validation in association studies linking microbiota with host traits.

Chapter 4

Materials and methods

4.1 Samples

4.1.1 Ovine fecal samples

The ovine fecal samples, kindly provided by Dr. Gavino Marogna (Istituto Zooprofilattico Sperimentale della Sardegna), were collected from five Sarda sheep belonging to the same flock. The animals were lactating females, free-grazing, and without evident clinical symptoms. Fecal samples were collected and stored at -80°C until use.

Two different approaches were used in this work to treat fecal samples. According to the first approach, different extraction methods were used to extract DNA directly from fecal sample. According to the second approach, the fecal sample was pretreated with a so-called "differential centrifugation" before being subjected to the various methods DNA extraction; briefly, after thawing at 4°C, fecal samples (approximately, 100 mg each) were resuspended in 10 ml of PBS, vortexed, shaken in a tube rotator for 45 minutes, and subjected to low-speed centrifugation at 500 x g for 5 minutes to eliminate gross particulate material; the supernatants were carefully transferred to clean polycarbonate centrifuge bottles (Beckman Coulter, Brea, CA, USA) and kept at 4°C, whereas the pellets were suspended again in PBS. The entire procedure was repeated for a total of three rounds. Then, the three supernatants obtained from each sample were centrifuged at 20,000 x g for 15 minutes, and the three derivative pellets were pooled after resuspension with the extraction buffer described below, and subjected to DNA extraction as detailed in the "DNA extraction" section in this chapter.

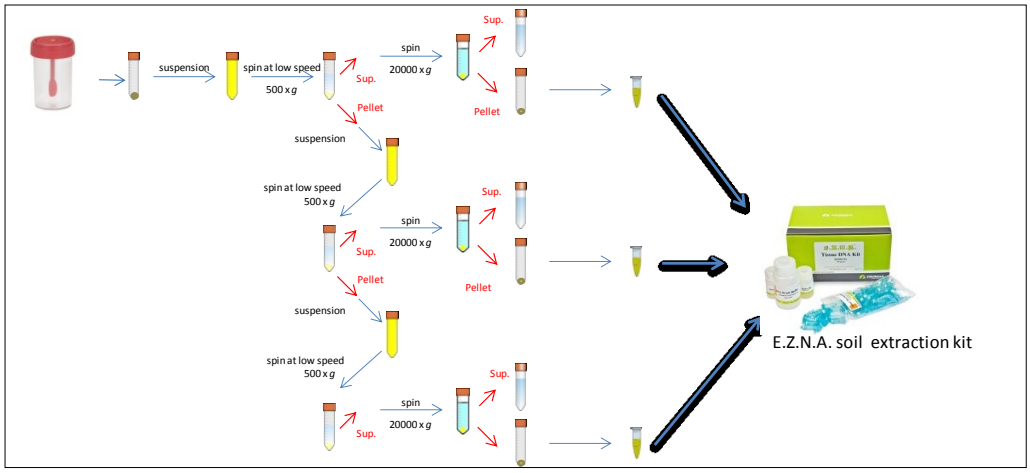

Figure 5. Differential centrifugation

4.1.2 Murine fecal samples and cecal content

The murine fecal samples and cecal contents (analyzed in Chapter 5) were kindly provided by Dr. Michael Silverman (Department of Microbiology and Immunobiology, Harvard Medical School, Boston, USA). Briefly, cecum was removed in hood under aseptic conditions and placed in 10 ml of cold PBS; cecum was opened longitudinally and vortex for 10-20 seconds to release content. The content was centrifuged 10 minutes at 10,000 x g at 4°C and the surnatante was aspirate off.

The murine intestinal contents analyzed in paragraph 5.7 were kindly provided by prof. Svegliati Baroni (Università politecnica delle Marche, Ancona, Italy). The samples were collected from mice raised under standard condition, and stored at -80°C until use.

4.1.3 Human stool samples

The human stool samples analyzed in paragraph 5.2.2 were kindly provided by prof. Francesco Cucca (Institute of Biomedical and Genetic Research, Centro Nazionale Ricerche, Cagliari, Italy). The samples were collected from Ogliastra population, during a recall for immune traits measurements in fully sequenced and/or genotyped individuals and stored at -80°C until use.

4.2 DNA extraction

Different DNA extraction methods were tested on different animal models to extract microbial genetic materials from different substrates. For this purpose we used two commercial kits that use the enzymatic and mechanical lysis for the microbial wall and membrane disruption, the QIAamp Fast DNA Stool (Qiagen, Hilden, Germany) and the E.Z.N.A. Soil DNA Kit (Omega Bio-Tek, Norcross, GA) respectively.

DNA was extracted according to the instruction of the QIAamp Fast DNA Stool QIAGEN; briefly, fecal sample was mixed with 1 ml InhibitEX buffer and vortexed until the sample was thoroughly homogenized. After homogenization, the suspension was heated for 5 minutes at 95°C for cells that are difficult to lyse (such as Gram-positive bacteria). Subsequently, the sample was centrifuged at full speed for 1 minute to pellet stool particles. Then, 200 ul of supernatant was added to 15 ul proteinase K. After adding AL buffer, ethanol was added to lysate. Finally, DNA bound to the QIAamp membrane is washed in two centrifugation steps. Wash conditions using two wash buffers ensure complete removal of any residual impurities without affecting DNA binding. Purified, concentrated DNA is eluted from the QIAamp Mini spin column in 50 ul Nuclease-free water (Ambion, Waltham, MA).

DNA extraction was performed according to the E.Z.N.A. Soil DNA Kit manufacturer's instructions. Briefly, sample was resuspended in 1 ml SLX - Mlus buffer and lysed using 500 mg glass beads; after a first incubation at 70°C for 10 minutes, a second incubation at 95°C for 2 minutes was performed for DNA isolation from Gram-positive bacteria. Then, 270 ul P2 buffer was added to remove proteins and inhibitors and 0.7 volumes isopropanol to precipitate DNA. Subsequently, 100 ul HTR reagent was added to the pellet resuspended in Elution Buffer to remove inhibitors. Finally, DNA bound to the HiBind DNA Mini Column membrane is washed in centrifugation steps to complete removal of any residual impurities without affecting DNA binding. Purified, concentrated DNA is eluted from the HiBind DNA Mini Column in 50 ul Nuclease-free water (Ambion, Waltham, MA).

In order to extract the microbial DNA from ovine fecal samples we applied the protocol published by Hildebrand and colleagues (Hildebrand *et al.*, 2012), that combine beads pretreatment with QIAamp DNA Stool protocol (Qiagen, Hilden, Germany); also we tested the phenol/chloroform/Isoamyl Alcohol (25:24:1) organic extraction method on sheep samples.

In the first extraction, the fecal sample was dissolved in 1 ml TE-buffer (10 mM Tris–HCl, 1 mM EDTA, pH 8) and centrifuged at 500 x g for 2 minutes. The supernatant was centrifuged for 5 minutes at 19000 x g and pellet was dissolved in 1.2 ml TE-buffer. The sample was transferred to a tube containing 0.5 ml stainless steel beads (5 mm diameter, Qiagen, Hilden, Germany) and 30 µl 10% sodium dodecyl sulphate (SDS). Bacterial cells were lysed by shaking for 4 minutes on a bead-beater (TissueLyser LT mechanical homogenizer, Qiagen) and centrifuged at 2,300 x g for 1 minute. Finally, DNA was extracted using the QIAamp DNA stool Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions.

In the second extraction, we used phenol that is a powerful denaturing of the proteins and lipid solvent. The chloroform is added to increase the action denaturing of the phenol on the proteins, improve the solubilization of the lipids and accentuate the difference in density between the organic phase and the aqueous phase, facilitating the separation. Briefly, the sample was homogenized to equal volume of phenol/ chloroform/Isoamyl Alcohol (25:24:1) and centrifuged for 10 minutes at 12000 rpm; to the supernatant (aqueous phase) was added 0.7 volume isopropanol to precipitate DNA; then, the precipitate was centrifuged at 12000 rpm for 15 minutes, and the pellet washed with 1 ml ethanol. After centrifugation at 12000 rpm for 5 minutes the supernatant was eliminated and the pellet was dried on air for 10-15 minutes at room temperature. Finally, the pellet was resuspended in 100 ul Nuclease-free water (Ambion, Waltham, MA).

The extracted DNA was quantified on a Qubit 2.0 Fluorometer (Life Technologies), using the Qubit ds DNA High Sensitivity Assay Kit (Life Technologies). DNA integrity was confirmed on 0.8% agarose gel (Sigma Aldrich, St. Louis, MO).

4.3 16S rRNA gene amplification

Primer design for universal amplification of the V4 region of 16S rDNA was based on a protocol published by Caporaso and co-workers (Caporaso *et al*., 2011). Amplification of the entire 16S-rRNA genes was performed using the universal primers 27F-1492R (AGAGTTTGATYMTGGCTCAG and TACGGYTACCTTGTTACGACTT, respectively). In both cases, PCR cycling conditions were as follows: 2 minutes at 94°C; 28 cycles of 30 seconds at 94°C, 30 seconds at 55°C, 2 minutes at 68°C; finally, 7 minutes at 72°C. PCR products were confirmed on 2% agarose gel (Sigma Aldrich, St. Louis, MO). Two

separate 16S rRNA gene amplification reactions were performed, pooled together, cleaned up using AMPure XP (Beckman Coulter, Brea, CA) magnetic beads and quantified with the Qubit HS assay (Invitrogen) using Qubit fluorometer 2.0 (Life technologies, Grand Island, NY).

Library construction

Libraries were constructed with the Illumina Nextera XT kit (Illumina, San Diego, CA). Simultaneous amplicon fragmentation and adaptor sequence ligation (also called tagmentation) were performed according to manufacturer protocol. Briefly, 5 µl of input DNA at 0.2 ng/µl (1 ng total) was mixed with tagmentation buffer (10 µl) and Amplicon Tagment Mix (5 µl), then incubated at 55° C for 5 minutes. A limited-cycle PCR was carried out to enrich and perform dual indexing on the tagmented DNA: 72° C for 3 minutes, 98° C for 30 seconds, 5 cycles of 95° C for 10 seconds, 55° C for 30 seconds, 72° C for 30 seconds), and a final cycle of 72°C for 5 minutes. Indexed libraries were purified using AMPure XP beads (Beckman Coulter, Brea, CA) and validate using High Sensitivity DNA chip on an 2100 Bioanalyzer instrument (Agilent Technologies, Santa Clara, CA). Sequence-ready libraries were normalized to ensure equal library representation in the pooled samples.

4.4. DNA sequencing

Pooled libraries were diluted and loaded at 20 pM on a V3 flow cell, with 1% phiX control using CBot instrument to generate the clusters, according to the V3 Illumina TruSeq Paired End Cluster Kit protocol (Illumina, San Diego, CA). After clusters generation (density 750–850 K/mm²), DNA sequencing was performed with the Illumina HiScanSQ sequencer, using the paired-end method and 93 cycles of sequencing. Basecalling was done using Illumina's Real Time Analysis (RTA) software version 1.14.21. Obtained BCL files were converted

into QSeq format using Bcl2Qseq 1.9.3, then converted to fastQs. After, all reads were subjected to a demultiplexing step using Casava software version 1.8.2 implemented in HiScan control software (Illumina, San Diego, CA).

4.5 Sequencing data analysis

The Illumina demultiplexed paired-reads were trimmed for the first 20 bp using FASTX and the sequences with Nextera adapter contamination were identified using the UniVec database (ftp://ftp.ncbi.nlm.nih.gov/pub/UniVec/) and removed. Therefore, the paired-reads with a minimum overlap of eight bases were merged using a specific QIIME script. OTU generation was done using a QIIME pipeline based on USEARCH's OTU clustering recommendations (http://www.drive5.com/usearch/manual/otu_clustering.html). Reads were clustered at 97% identity using UCLUST to produce OTUs (Edgar, 2010). Taxonomy assignment of resulting OTUs was performed using the Greengenes 13_8 database (Desantis *et al*., 2006). With taxonomic lineages in hand, OTU tables were generated and used for downstream analysis. In particular, alpha (Shannon index) and beta (weighted or unweighted UniFrac distances) diversity metrics and taxonomic classifications were computed using the QIIME software suite (Caporaso *et al*., 2010; Kuczynski *et al*., 2011).

Differential analysis were performed using DeSeq2 (Love MI *et al*., 2014; Anders S *et al*., 2010), a statistical tool available in QIIME. DeSeq2 implements a method based on the negative binomial distribution and it is extensively used in RNA-Seq studies. Indeed, in early RNA-Seq studies lacking biological replicates, the distribution of feature counts across technical replicates was reported to fit well to a Poisson distribution where the variance is equal to the mean (Marioni JC *et al*., 2008; Bullard JH *et al*., 2010), but when biological replicates are included, it has been noted that the Poisson distribution

underestimates the variation seen in the data (Robinson MD *et al*., 2007), a problem known as overdispersion. Therefore, the resulting statistical test based on Poisson assumptions does not control type-I error. As economical considerations typically do not allow large numbers of biological replicates, the Negative Binomial (NB) distribution has been proposed because of its ability to deal with the overdispersion problem (Anders S *et al*., 2010; Robinson MD *et al*., 2010) and it has achieved a dominant position in the methodologies to model feature counts for RNA-Seq data. The data generated from DNA-seq analysis are similar to those obtained by RNA-seq, in both cases are obtained a series of raw counts. Furthermore, in both approaches it's not easy to accurately detect differentially expressed genes or OTUs between groups. Difficulties faced by researchers in DNA-seq study design and analysis, as well as RNA-Seq are: general biases and errors inherent in the NGS technology (e.g. biases introduced during library preparation, specific biases in sequence quality and error rate), undetermined effects of sequencing depth and the number of replicates, the combination of technical and biological variation as well as biases within and between treatment groups that make it difficult to accurately discriminate real biological differences between groups. Therefore, it is possible to use DeSeq2 approach for the differential abundance analysis obtained from DNA-seq data.

Chapter 5

Results

According to the aims of this doctoral project, a number of results have been obtained that are related to a search for a pragmatic solution toward methodological issues and, also, to investigate specific biological issues.

For the sake of clarity, the results obtained are illustrated below taking into account methodological issues (5.1 – 5.3) and biological issues (5.4 – 5.7).

5.1 Validation of robust methods of genomic DNA preparation from complex samples

Feces is a complex sample matrix with regards to DNA extraction. Several substances may be co-extracted having inhibitory effects on downstream analysis, and the quality and quantity of the extracted DNA will greatly influence the outcome of Next Generation Sequencing. In extracting DNA from a fecal matrix, two courses can be taken; direct lysis of microbial cells in the fecal matrix, or pretreatment of the stool sample to promote the DNA extraction. Usually, direct lysis (chemical, mechanical or a combination of these) is the most widely applied strategy.

5.1.1 Quality of methods for murine fecal and cecal samples

We evaluated the efficacy of QIAamp DNA Stool protocol (Qiagen, Hilden, Germany), DNA extraction method widely used for gut microbiota studies, on murine fecal samples and cecal contents. For this purpose, we extracted DNA from four stool samples and four cecal content of two mice. The final DNA concentration was higher in cecal samples compared to stool (mean: 9.08 ± 3.77 ng/µl vs 3.78 ± 2.03 ng/µl, respectively), however, the quality of the DNA was such as to permit the amplification of the 16S rRNA gene and, therefore, suitable for the preparation of libraries for Next Generation Sequencing.

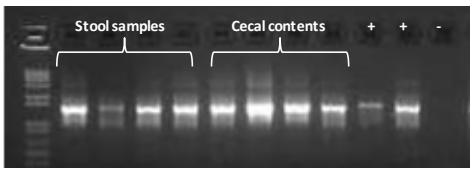

Figure 6. agarose gel (2%) of stool and cecal amplicons

5.1.2 Quality of methods for ovine stool samples

We tested a total of 5 approaches to extract DNA from five fecal samples: DNA extraction directly from the fecal matrix using two commercial kits, QIAamp DNA Stool protocol (Qiagen, Hilden, Germany) and E.Z.N.A. Soil DNA Kit (Omega Bio-Tek, Norcross, GA); DNA extraction with different methods after pretreatment of the sample as described in the paragraph 4.1.1 (differential centrifugation) in order to enrich for microbial cells, including phenol/chloroform/Isoamyl Alcohol (25:24:1) method, the protocol described by Hildebrand and collaborators (Hildebrand *et al.*, 2012), and the commercial kit protocols previously evaluated directly on the stool sample. Integrity evaluation determined by 0.8% agarose gel electrophoresis showed a remarkable degradation of the DNA extracted with E.Z.N.A DNA Kit protocol.

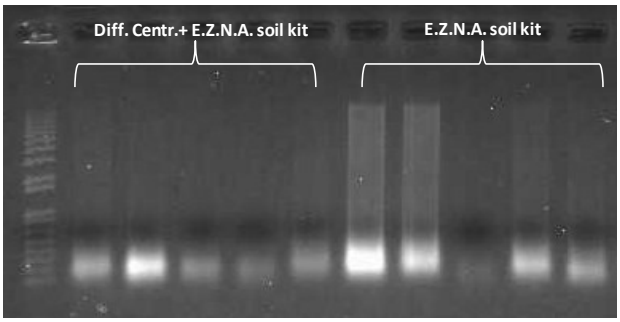

Figure 7. 0.8% agarose gel of DNA extracted by E.Z.N.A DNA Kit protocol

DNA concentration was higher in the samples extracted with phenol/chloroform/Isoamyl Alcohol method (94.8 ± 18.66 ng/µl), while it was lower in those extracted with protocol described by Hildebrand and E.Z.N.A.

soil DNA kit (mean: 2.5 ± 0.28 ng/µl). To determine the suitability of extracted DNA for downstream Next Generation Sequencing we used 16S rRNA gene amplification assays. The results demonstrated that only the DNA extracted with E.Z.N.A. soil DNA kit protocol, applied after differential centrifugation, was suitable for subsequent PCR amplification and Next Generation Sequencing.

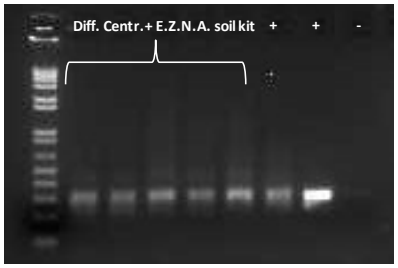

Figure 8. 16S rRNA amplification of samples pretreated with differential centrifugation and extracted with E.Z.N.A. soil DNA

5.1.3 Quality of methods for human stool samples

This study examined stool samples from 52 individuals using two DNA extraction approaches to identify and eventually quantify the extent of the DNA extraction method (technical variation) impact upon inter-subject variation (biological variation). For this purpose we extracted the DNA with the QIAamp DNA Stool protocol (Qiagen, Hilden, Germany); then, an aliquot of extracted DNA was subjected to purification according to E.Z.N.A. soil DNA kit protocol. Subsequently, DNA was subjected to amplification of the 16S rRNA gene and sequenced. PCoA Analysis showed that samples clustered according to the DNA extraction approach used.

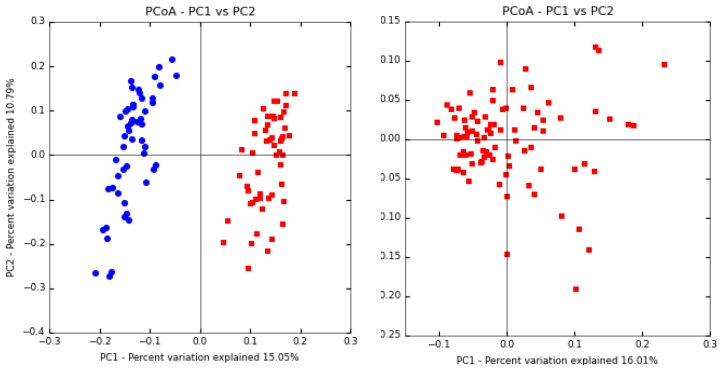

Figure 9. On the left the PCoA with in blue the samples extracted with Stool kit QIAGEN and in red the samples purified with E.Z.N.A. soil DNA Kit protocol; on the right the PCoA with samples purified with E.Z.N.A. soil DNA Kit protocol

Indeed, while the ratio Firmicutes / Bacteroidetes does not change regardless the used method, the purification with E.Z.N.A. soil kit showed an enrichment of Actinobacteria (4.1% ± 8.2% vs 0.68% ± 1.3%; $P = 0.005$) and Proteobacteria (3% ± 3.8% vs 0.83% ± 2.8%; $P = 0.07$) phyla compared to samples extracted only with QIAamp DNA Stool protocol; despite this, the value of alpha diversity (Shannon index) was similar (7.38 ± 0.65 vs 7.20 ± 0.95 respectively). Furthermore to determine the suitability of extracted DNA for downstream Next Generation Sequencing, we subjected to amplification of 16S rRNA gene, 89 samples treated with both methods; the results showed that only the 58.42% of samples extracted with the QIAamp DNA Stool protocol were suitable for the NGS libraries preparation, whereas 100% of the samples purified according E.Z.N.A. soil DNA kit protocol were suitable.

5.1.4 Discussion and conclusion

Sampling and DNA extraction methods that result in adequate yields of microbial DNA are also required to accurately represents the microbial community. The DNA yield and quality, and its suitability for downstream PCR amplifications varied considerably, depending on the DNA extraction method used. In recent years, numerous studies have been conducted to assess the potential of different DNA extraction methods from fecal samples of different hosts (human, mice, rats and avian samples). However, Ferrand and colleagues evaluated different methods from caecal contents and feces of mice with classical techniques such as qPCR (Ferrand J *et al.*, 2014); other studies utilize the amplification of different hypervariable regions of 16S rRNA gene as V4 and V6 (Josefsen M H *et al*., 2015; Xin Peng *et al*., 2013). Moreover, in most studies researchers evaluated the effectiveness of commercial kits, including the QIAamp DNA Stool kit which is one of the most used methods. In this context, our study provides, for the first time a comparative evaluation between different extraction methods of microbial DNA from sheep fecal sample, offering innovative approaches in the sample preparation as a pretreatment with differential centrifugation and identifying the best method of DNA extraction which it is constituted by a combined approach (differential centrifugation followed by DNA extraction with Protocol of E.Z.N.A. soil DNA kit).

We assessed the validity of QIAamp DNA Stool kit protocol in murine fecal and cecal samples for intestinal microbiota studies performed with Next Generation Sequencing platforms. Furthermore, we observed that, despite being very commonly used method, the QIAamp DNA Stool kit might not as robust as required; in fact only 50% of human fecal specimens treated with this protocol had a good quality; we, therefore, investigated an alternative method that has allowed a yield of 100% of samples suitable for the amplification of full 16S rRNA gene; in this way we have proposed for the first time the use of

purification with E.Z.N.A. soil DNA kit protocol to increase the quality of the DNA extracted with the QIAamp DNA Stool kit. In conclusion, it is essential to continually evaluate nucleic acid extraction methods, identifying those that are the most efficient, accurate and reproducible based on the type and origin of the sample to be treated (i.e. human, mouse, sheep) and that reflects the real composition of the microbiota.

5.2 Comparative assessment of Illumina technology methods for the characterization of the gut microbiota

In this work, we have compared the impact of different methods for metagenomic sequencing and data analysis, making use of real gut microbiome from human and mouse. Different gut microbiome samples were obtained from intestinal contents and feces of adult mice and from fecal samples of a healthy human volunteer. DNA was then amplified with specific primers for the hypervariable region V4 and for the complete 16S rRNA gene. The resulting V4, full length 16S and whole metagenome datasets were compared in terms of microbial alpha-diversity, beta-diversity and community structure.

5.2.1 Mice samples sequencing pipelines

Full metagenome, complete 16S and V4 sequencing of 6 murine samples (three stools and three cecal contents) resulted in an average of 9,672 reads (0.13% of filtered total sequences), 199,954 reads (46.73% of filtered sequences), and 217,406 reads (72.41% of filtered total sequences) assigned for sample, respectively. In fecal samples, a significant greater number of filtered reads obtained with the 16S amplicons sequencing (means: 203,538 ± 8,890.79) compared to full metagenome (means: 8,427 ± 1,993), but lower than V4

amplicons (means: 221,638 ± 4,961; $P = 0.0511$); similarly, in cecal content samples, a larger number of reads with the V4 amplicons sequencing (means: 213,175 ± 2,335) were observed compared to metagenome (means: 10,917 ± 1,075), and 16S amplicons (means: 196,371 ± 152,923; $P = 0.194953$). On the contrary, OTU's number recorded in fecal samples and cecal content samples was lower with V4 than with 16S sequencing (means: 751 ± 96 vs 909 ± 73; $P = 0.090456$ and 767 ± 89.6 vs 904 ± 44.8; $P = 0.1003$, respectively).

Samples	16S Reads assigned	V4 Reads assigned	Meta Reads assigned	16S OTUs assigned	V4 OTUs assigned	Meta OTUs assigned
stool						
281	193376	218851	7102	825	640	461
301A	209884	227366	10719	956	807	774
301B	207353	218696	7460	947	806	807
mean	**203538**	**221638**	**8427**	**909**	**751**	**680**
St.dev.	**± 8891**	**± 4961**	**± 1993**	**± 73**	**± 96**	**± 190.95**
Intestinal content						
281	183538	212550	10735	888	664	518
301A	213292	211216	12072	870	815	753
301B	192282	215759	9945	955	823	711
mean	**196371**	**213175**	**10917**	**904**	**767**	**661**
St.dev.	**± 15292.5**	**± 2335**	**± 1075.2**	**± 44.8**	**± 89.5**	**± 125.32**

Table 2. OTUs and reads numbers for 16S, V4 and full metagenome samples

Shannon's diversity index was higher in 16S and shotgun metagenome samples than V4 samples (7.04 ± 0.49 and 7.24 ± 1.30 vs 5.88 ± 1.24, respectively).

16S Samples	16S Shannon index	V4 Shannon index	Meta Shannon index
stool			
281	6.47	4.50	7.59
301A	7.32	6.23	5.81
301B	7.32	6.90	8.33
mean	**7.04**	**5.88**	**7.24**
St.dev.	**± 0.49**	**± 1.24**	**± 1.30**
Intestinal content			
281	6.35	5.01	5.58
301A	7.23	6.45	7.35
301B	7.71	6.48	7.35
mean	**7.10**	**5.98**	**6.76**
St.dev.	**± 0.69**	**± 0.84**	**± 1.02**

Table 3. Shannon index value from stool and cecal samples

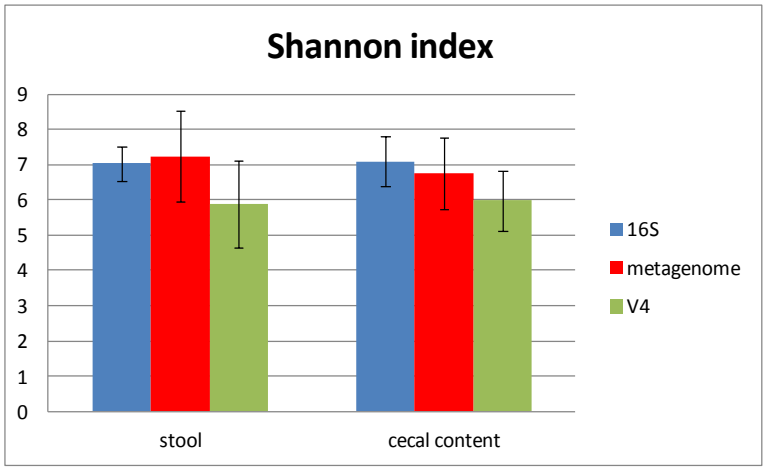

Figure 10. bar chart of Shannon index in stool and cecal content

Beta-diversity, investigated by PCoA analysis (unweighted UniFrac distance metric), showed an evident clustering of stool and cecal content samples and between 16S, V4 and full metagenome samples.

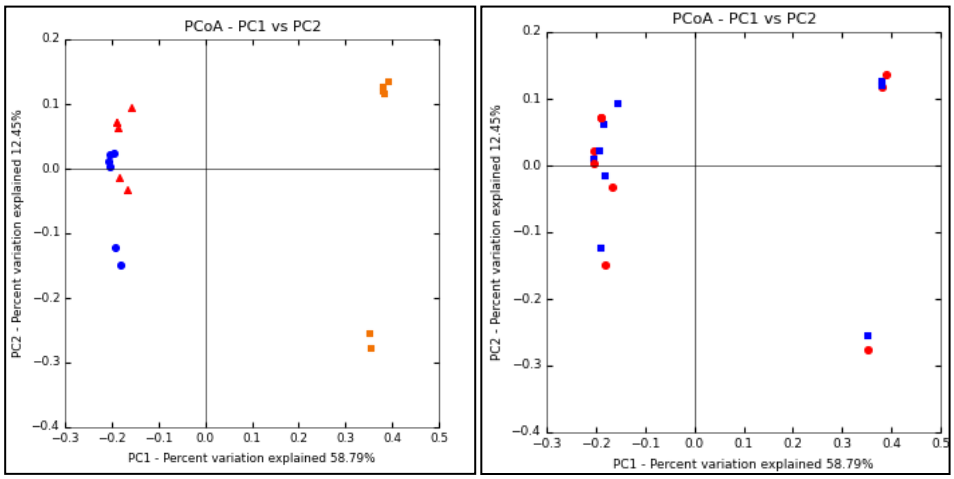

Figure 11. PCoA plot sequencing methods: on the left,full metagenome (orange), 16S (red) and V4 amplicons (blue); on the right, stool (red) and cecal content (blue)

Differential analysis performed by DESeq2 showed that a number of OTUs belonging to either Firmicutes and Bacteroidetes were dramatically underrepresented or absent in the V4 datasets compared to the full length 16S. In stool samples, at the phylum level, the percentage of Bacteroidetes was higher in the 16S than V4 and the entire metagenome (68.58% ± 16.25% vs 47.28% ±

22.42% and 54.90% ± 16.53%, respectively); in particular, the genus *Bacteroides* was absent in V4 datasets compared to 16S (0.0% vs 0.38% ± 0.11%, $P = 0.00026$ in stool and vs 0.30% ± 0.11%, $P = 0.00048$ in cecal content). Firmicutes was higher in V4 (48.95% ± 25.26) compared to 16S (30.47% ± 15.07) and metagenome (40.39% ± 19.64). The phylum of Actinobacteria was underrepresented in 16S samples (0.46% ± 0.47) compared to V4 (3.39% ± 2.69) and metagenome (3.77% ± 3.14); among members of Actinobacteria, *Bifidobacterium* was differently abundant between 16S (0.047% ± 0.08), V4 (1.79% ± 1.69; $P = 0.0058$) and metagenome datasets (0.52% ± 0.53; $P = 0.00017$), while *Adlercreutzia* only between 16S and metagenome samples (0.21% ± 0.06% vs 0.82% ± 0.42%; $P = 0.049911$). On the contrary, in the intestinal contents the abundance of Bacteroidetes was higher in entire metagenome (73.05% ± 4.90%; $P = 0.040978$), while 16S and V4 amplicons had similar values of the main phyla (Figure 11). An important aspect is that V4 approach in the stool samples showed a ratio Firmicutes/Bacteroidetes >1 while the very same stool samples showed a <1 value with 16S and full metagenome analyses (Bacteroidetes higher than Firmicutes).

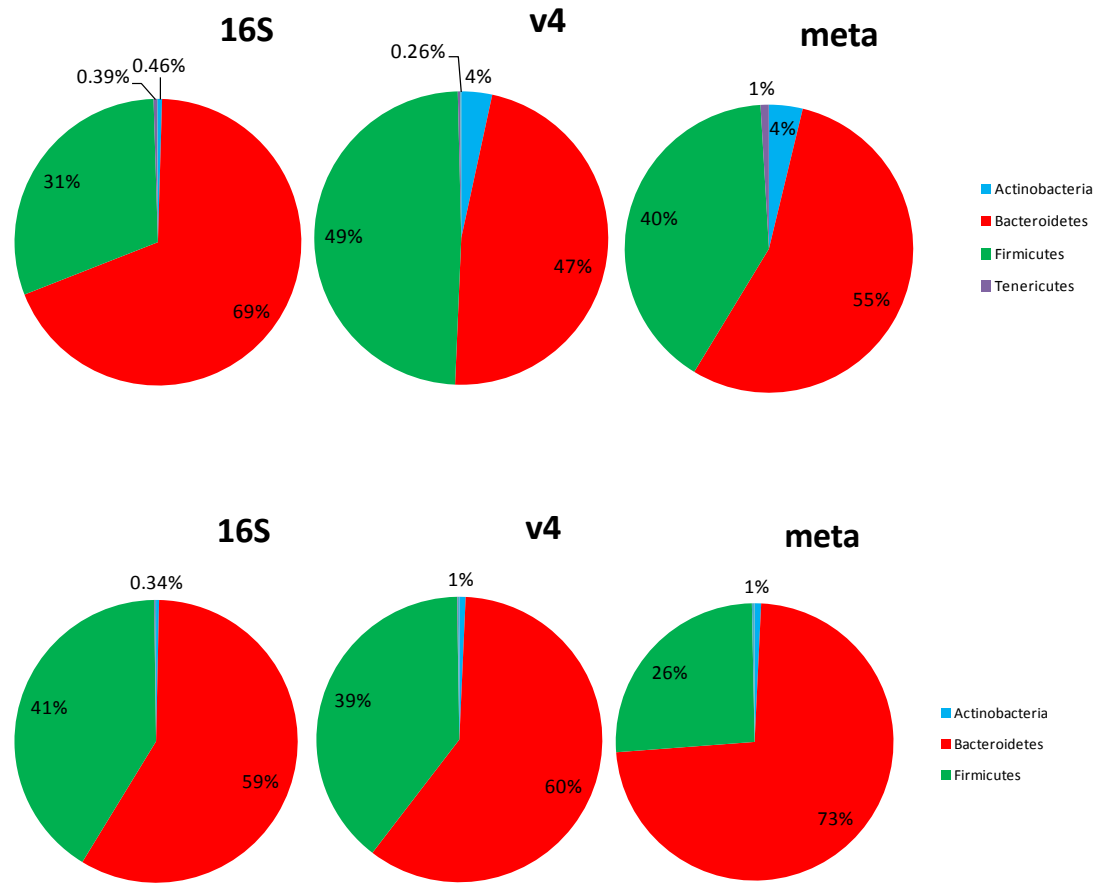

Figure 12. Pie Charts of principal phyla in stool samples (above) and cecal contents (below)

5.2.2 Human sample sequencing pipelines

V4 and complete 16S sequencing of the human samples resulted in an average 229,233 and 126,386 reads assigned for sample, respectively. As expected, the OTUs numbers and alpha diversity values were higher for 16S samples than for V4 amplicons (means: 1,460 ± 36.77 vs 1,343 ± 4.24; P = 0.154677 and 8.40 ± 0.007 vs 7.88 ± 0.003; P = 0.026081).

Samples	16S Reads assigned	16S OTUs assigned	16S Shannon index	V4 Reads assigned	V4 OTUs assigned	V4 Shannon index
HumanA	82311	1434	8,42	232402	1346	7.91
HumanB	170460	1486	8,41	226063	1340	7.86
mean	**126386**	**1460**	**8,41**	**229233**	**1343**	**7.87**
St.dev.	**± 62331**	**± 36.77**	**± 0.007**	**± 4482**	**± 4.24**	**± 0.037**

Table 4. OTUs, reads numbers and Shannon index for 16S, V4 samples

Taking into account that significantly lower number of reads obtained by full metagenome sequencing, as compared to 16S amplicon sequencing, allow to detect similar alpha-diversity indexes in murine samples (see Table 2), we also evaluated three different runs of shotgun metagenome sequencing of human samples performed at different coverages. A metagenomic library from the very same DNA extracted sample was sequenced making use of a v3 flowcell lane (~ 18.75Gb expected capacity for single lane), a third of a lane (~ 6.25Gb expected), and a sixth of a lane (~ 3.1Gb expected).

	Entire lane	1/3	1/6
Reads assigned	109458	61810	36271
OTUs number	2301	2247	2195
Shannon index	9.29	9.15	9.13

Table 5. OTUs, reads numbers and Shannon index for metagenome samples

The amount of reads (merged and filtered) generated by NGS was 75,892,265 (assigned: 109,458) for "entire lane", 36,618,279 (assigned: 61,810) for "1/3", and 28,196,488 (assigned: 36,271) for "1/6". Nonetheless, the amount of OTUs identified was similar in all outputs (2,301 vs 2,247, and 2,195 respectively); also, the Shannon index was similar for the three different coverages (9.29, 9.15, and 9.13, respectively).

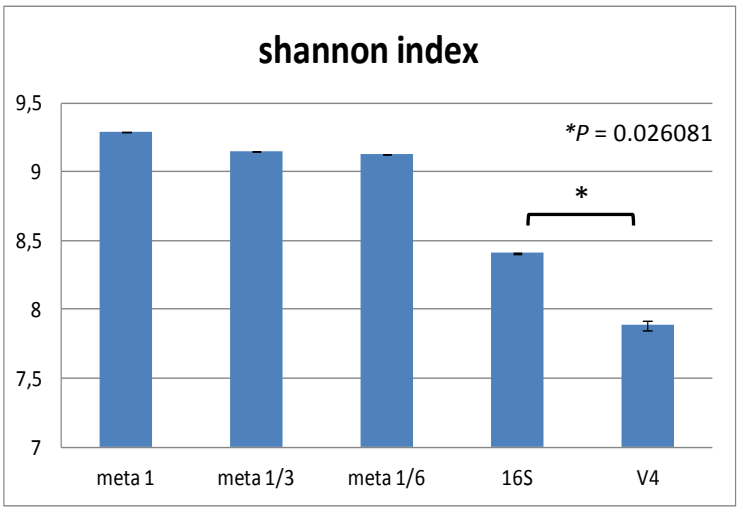

Figura 13. Bar chart of Shannon index values in human samples

Differential analysis performed by DESeq2 showed several OTUs belonging to the phylum Bacteroidetes differentially abundant between 16S and V4 datasets. In fact, even at the phylum level it is evident a different taxonomic distribution of the 3 main phyla (Bacteroidetes mean: 48.40 ± 2.43% in 16S, 41 ± 2.5% in V4 and 56.93 ± 0.20% in full metagenome; Firmicutes mean: 48.81 ± 2.6% in 16S, 54.48 ± 2.9% in V4 and 35.83 ± 0.41% in full metagenome. Proteobacteria were also differently abundant when comparing the three approaches (1.27% ± 0.11 in 16S, 2.55% ± 0.28 in V4 and 3.41% ± 0.05 in metagenome) Among the genera belonging to the Bacteroidetes phylum, *Prevotella* was absent in V4 compared to 16S dataset (0.62% ± 0.07; P = 1.04E-09), while *Bacteroides fragilis* was underrepresented (0.87% ± 0.04 vs 1.70% ± 0.18, respectively). Among Proteobacteria, in particular, the *Sutterella* genus was underrepresented in the 16S (0.41% ± 0.05) compared to V4 (1.22% ± 0.12; P = 0.001182) and to metagenome (1.15% ± 0.04; P = 0.039243).

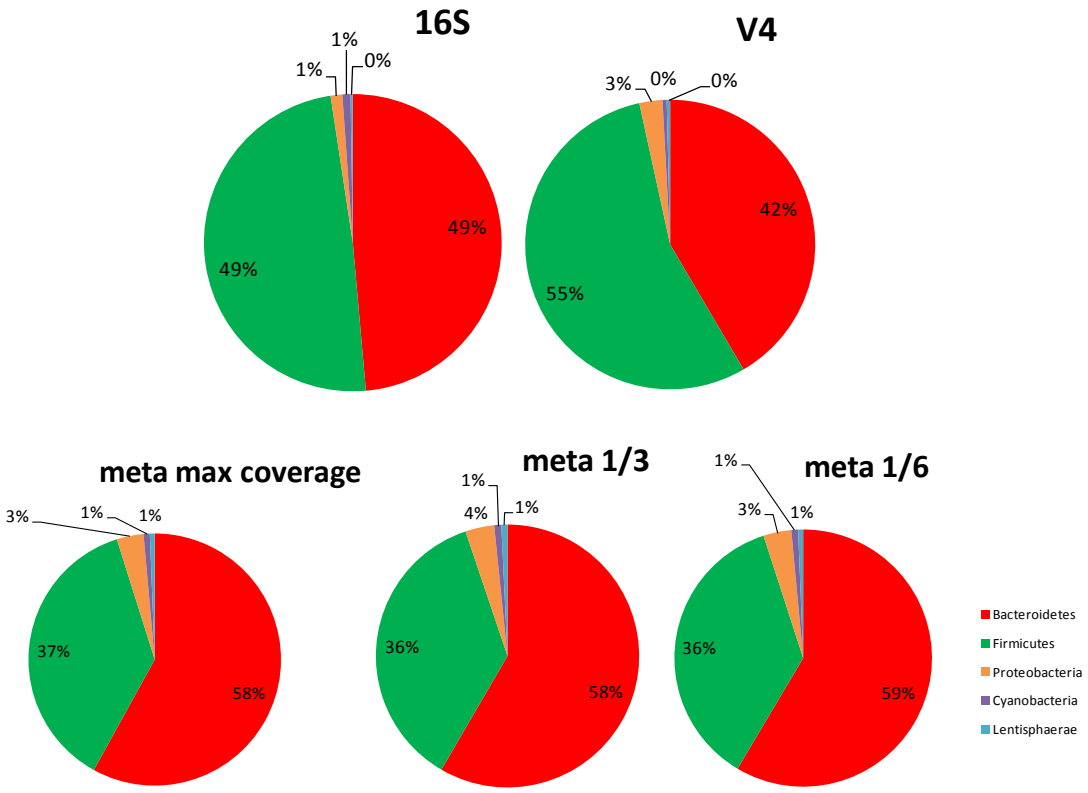

Figura 14. B Pie Charts of principal phyla in different approaches.

Finally, as shown in Figure 13 there are not significant differences in terms of taxonomic composition obtained by the analysis of the metagenome at different coverages.

5.2.3 Discussion and conclusion

This study represents the first evaluation of 3 different approaches (V4, 16S and full metagenome) performed on different types of sample (feces and cecal contents) and, simultaneously, on different hosts (mice and human). In fact, to date, no methodological studies of this type have been reported. We also performed an assessment on the impact of different coverage on the taxonomic characterization. This allowed us to obtain different results that suggested several conclusions. The numbers of reads assigned does correlate to numbers of OTUs identified and consequently to alpha diversity value; indeed, the number of reads is higher in amplicons V4 but this does not corresponds to a higher amount of OTUs and higher values of alpha diversity. On the contrary, for the 16S and, especially, for the full metagenome, greater values of alpha diversity and number of OTUs are not due to a higher number of reads assigned. This inverse correlation between number of reads, number of OTUs and Shannon Index, when comparing V4 vs 16S sequencing, would suggest that V4 datasets might be biased toward a more limited number of OTU and, therefore, showing less richness in the composition of microbial communities. In addition, the full metagenome, even with the lowest coverage tested (a sixth of lane), showed a high Shannon index, despite a low number of reads assigned. However, it is difficult to compare the metagenome with 16S amplicons due to the different number of reads assigned (significantly higher in 16S). In addition, the 16S approach allows a greater depth of sequencing (sequencing these rRNA encoding regions a highest multiple of times) compared to shotgun metagenome, due to a lower number of sequences in 16S than the entire metagenome, which, in turn, determines a greater number of reads. An important finding has emerged in the coverage analysis carried on the human sample: while the number of reads assigned is proportional to the coverage, the taxonomic richness (in terms of Shannon index and number of OTUs) was very similar. Also, unlike the murine samples, the number of OTUs and Shannon index are higher in the samples of the metagenome compared to 16S. Among the taxa that were recorded differently according to sequencing strategies, Proteobacteria are normal commensal in the human gut but they also represent an important group of

pathobionts; thus, their careful detection and changes monitoring is of paramount importance. Therefore, a special attention should be paid to the methods that might bias toward over- or toward underrepresentation of this group. However, the possibility of using 16S amplicons to obtain functional information is an important aspect to consider for the lowest cost and less effort bioinformatics required compared to metagenome. This suggests the need to evaluate the best approach in view of the microbial community to be investigated (i.e. human or mice), the possibility to assess the impact of lower coverage (less than 1/6) on the microbiome study and deepen the evaluation of the 16S amplicons for the functional studies. Nevertheless, these data relating principally to the bias of the V4 respect the 16S observed in mice and in human, must be validated using the MiSeq sequencer (Illumina, CA) which is becoming the most used Next Generation Sequencing instrument to sequence metagenomic samples in recent years.

5.3 Comparative assessment of bioinformatic methods for metagenomic data analysis. A study model with murine fecal samples

5.3.1 OTU table settings and comparative generation of sample diversity metrics

As introduced in the section "Methods to study the microbiome - Metagenomics", a critical step following the reads acquisition is their assignment to OTUs. Once generated, OTUs are the features that will allow to assess taxonomy structure and stability or fluctuation of the bacterial community. In order to handle the large numbers of OTUs generated and to make use of them to characterize the gut microbiota, we evaluated three possible approaches. To this extent, we compared the alpha and beta diversity values obtained i) from raw OTU table generated from the QIIME analysis, ii) from raw OTU table filtered to values higher than 0.5% of OTU abundance or iii) from OTU table normalized through a sub-sampling (86970 reads). These evaluations were carried out on samples collected at 3 weeks and 10 weeks of age from 8 mice (n = 16) to investigate the changes of the intestinal microbiota

in the different growth stages. In this study model, as illustrated in Table 6, all three approaches measured alpha-diversity values (Shannon index) with no significant changes between 3 weeks and 10 weeks.

	3 weeks old	10 weeks old
Raw OTU table	8.06 ± 0.56	7.83 ± 0.36
Filtered OTU table	7.42 ± 0.30	7.43 ± 0.22
Sub-sampling OTU table	8.66 ± 0.24	8.65 ± 0.89

Table 6. Shannon index values in 3 weeks and 10 weeks old mice groups

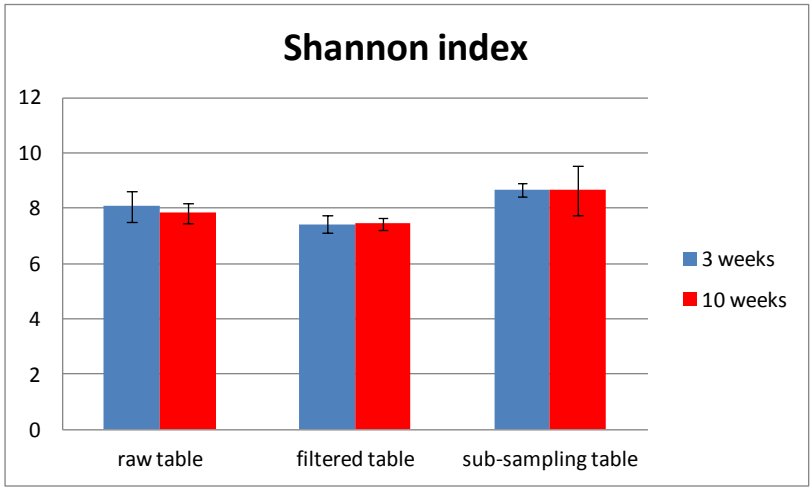

Figure 15. Bar chart of Shannon index at 3 weeks and 10 weeks old

Despite the high presence of singleton and doubleton in raw table, the Shannon index value was similar in all tables; this can be explained by the poor heterogeneity of laboratory mice. Analysis of β-diversity (unweighted Unifrac distance metric) illustrates, as expected, differences in the taxonomic composition between the two groups according to the development stages.

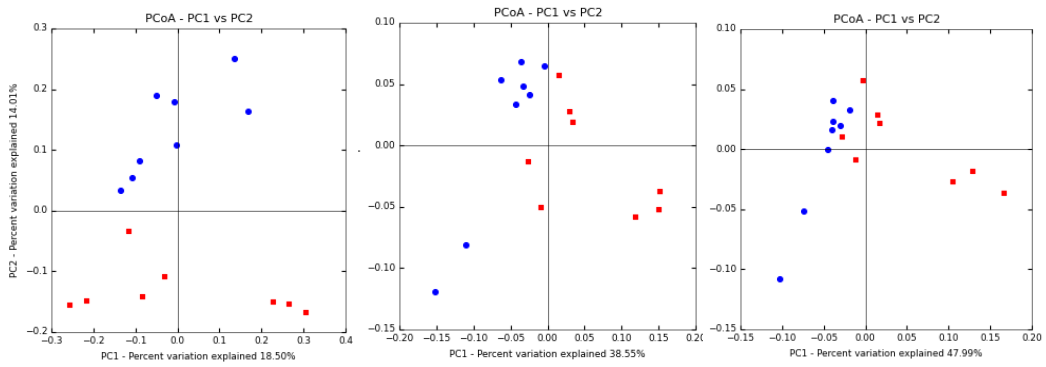

Figure 16. PCoA plots in three different approaches. In blue the samples at 10 weeks old; in red the samples at 3 weeks old.

Moreover, we tested three different statistical tools designed to define features that are differentially abundant between two or more groups of multiple samples: DeSeq2, metagenomeSeq and STAMP software. DeSeq2 and metagenomeSeq are software packages available in QIIME. STAMP (statistical analysis of taxonomic and functional profiles) is a graphical software package that provides statistical hypothesis tests and exploratory plots for analyzing taxonomic and functional profiles. For DeSeq2 and metagenomeSeq we considered the OTUs with FDR multiple test correction value less than 0.05, while to STAMP we used a Welch's test with FDR multiple test correction ($\leq$ 0.05). As shown by the bar chart below, metagenomeSeq did not identify significant differences in the microbial composition of the two groups, while STAMP and DeSeq2 identified 186 and 915 differential OTUs respectively, of which 179 are in common between them (Venn diagram).

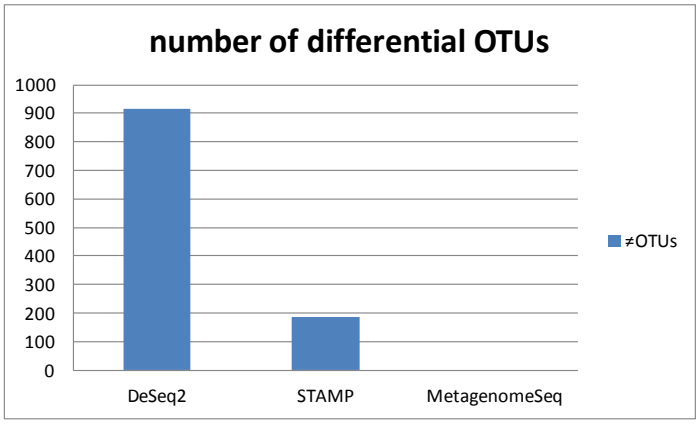

Figure 17. Bar chart of differential OTUs in the three different approaches

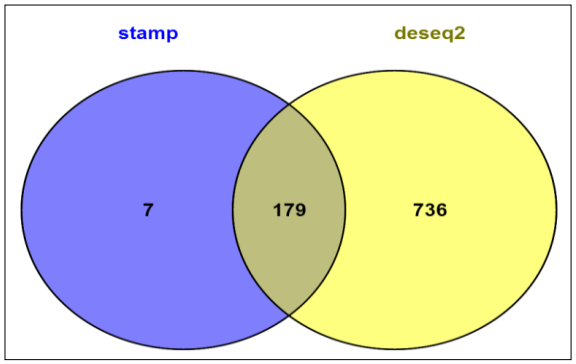

Figure 18. Venn diagram of OTUs in common between STAMP and DeSeq2

5.3.2 Conclusion and discussion

Next-generation sequencing (NGS) technology has extraordinarily enhanced the scope of research in the life sciences, and in particular, it has revolutionized the human gut microbiome research. However, this high throughput sequencing enabled obtaining thousands to millions of reads per run at decreasing costs but raised new challenges, notably for computation analysis of this biological data. Bioinformatics is faced with the problem of how to handle and analyze these datasets in an efficient and useful way. To date, there is still a serious need for bioinformatics pipelines that can efficiently process many large datasets enabling a precise and sensitive description of the complex microbial diversity of an ecological community. In this context, my doctoral project aimed to define a straightforward and optimized pipeline; in particular, in light of the reported results, the pipeline should include the use of all the approaches described above to assess and validate samples differences in terms of alpha and beta diversity. However, the use of different OTU tables, may be of little importance when the microbiome are less different from each other and, instead, more importantly, when they are most heterogeneous.Noteworthy, for the differential analysis performed with DeSeq2 is more sensitive than that achieved with the two used methods. This was not completely unexpected since this is a widely used software for the analysis of counts in RNA sequencing studies and, in this study, it has been applied to similar datasets (DNA reads).

5.4 Sample model for gut microbiota: does the fecal sample model for intestinal contents in mice?

Since most of the gut microbiota studies make use of feces to characterize the bacteriological content that populate the colonic mucosa, we intended to evaluate the possible bias in taxonomy definition driven by the use of fecal samples (stool) instead of intestinal (cecal) contents. For this purpose, we have collected 8 samples, 4 stool and 4 cecal contents, from 4 mice. 16S analysis performed by QIIME showed that in both types of samples, Bacteroidetes are more prevalent compared to Firmicutes (Bacteroidetes: 75.17% ± 21.07% in stool and 55.61% ± 10.16% in cecal content; Firmicutes: 24.15% ± 19.97% in stool and 44.04% ± 10.45% in cecal content), although the rate of Bacteroidetes

is higher in stool; however, the differential analysis performed with DeSeq2 did not show significant differences between the two groups according to the major phyla.

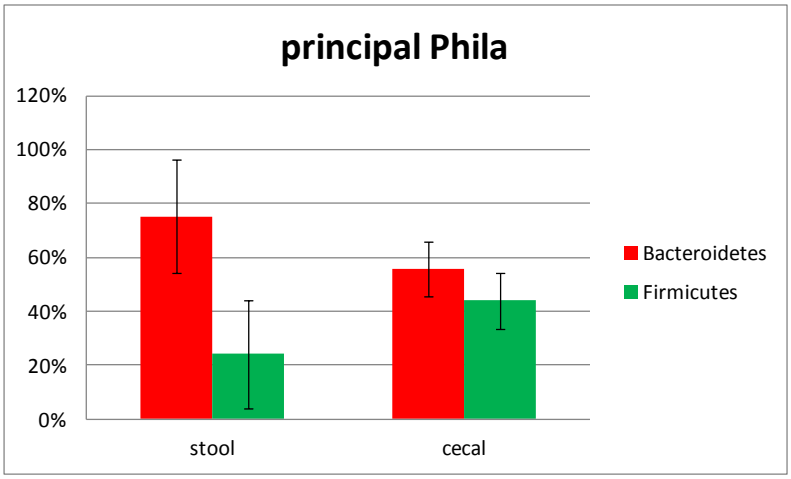

Figure 19. Bar charts of principal Phyla distribution in stool and cecal content

Concerning the metrics for community diversity, the values of alpha diversity did not reveal significant differences between feces and intestinal contents, although the Shannon index is higher in cecal content (7.62 ± 0.68 vs 7.19 ± 0.40; $P = 0.2449$) in keeping with the higher OTUs number (1081 ± 82.5 vs 1065 ± 132; $P = 0.6897$, respectively).

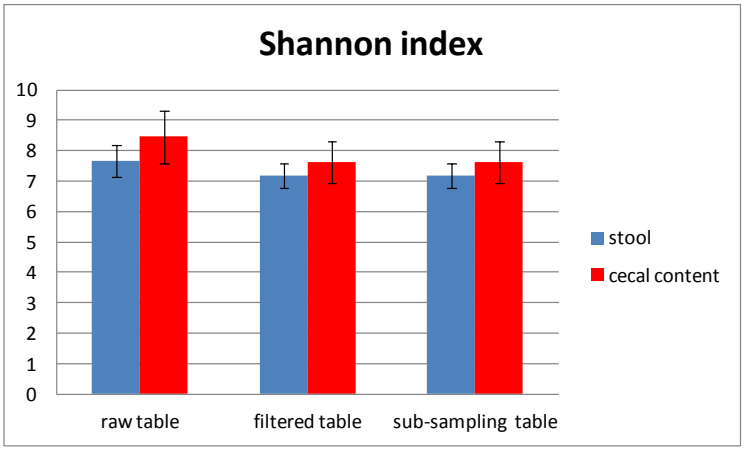

Figure 20. bar chart of Shannon index value in stool and cecal content

Analysis of β-diversity (unweighted Unifrac distance metric) indicates that samples grouped according to individual, not for type of sample.

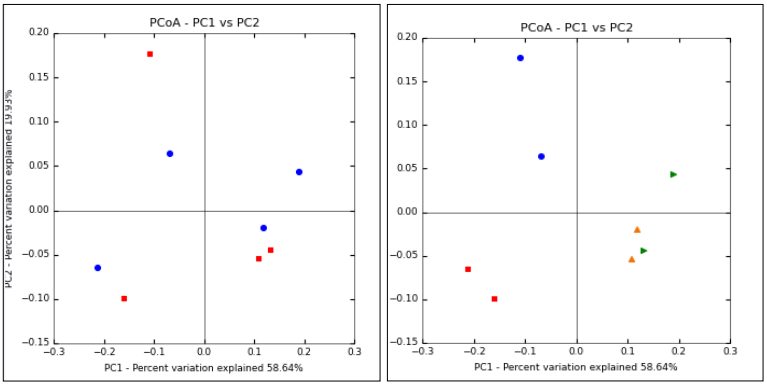

Figure21. In the PCoA plot on the left in blue color the stool samples and in red the cecal content; in the PCoA plot on the right the samples is colored by mouse.

5.4.1 Discussion and conclusion

The choice of starting sample is the first critical step in the gut microbiota study. However, an important aspect in the choice between feces and intestinal contents, is that stool sample can be collected easily anytime, but sample results only represent what is being shed at that particular time; the use of intestinal contents has the advantage of representing a real gut microbial composition, but samples can only be collected after animal sacrifice (in model studies) or else with invasive procedure (in human patients or volunteers), turning the cecal content not ideal to be used in most studies (i.e. human gut microbiota). In a recent study, Stanley and collaborators comparing fecal and cecal microbiota in chicken (Stanley *et al*. ,2015), demonstrated that fecal microbiota is qualitatively similar to cecal microbiota but quantitatively different (OTUs number), yet stool samples can be effectively used to detect some shifts and responses of cecal microbiota. Our study, carried with mice samples, show a modest reduction of complexity in the bacterial community when passing from the gut milieu (cecal content) to the cage litter (stool). This difference in OTUs number is not significant ($P = 0.6897$), but might suggest that in the "new" environment (i.e., lower temperature, atmospheric oxygen, humidity) might change the relative abundance of some species and partially select against some non adapted to the new condition. All taken together, in addition, our data showed that there are not significant differences in microbial community structure and, then, fecal samples can be used to model for the real microbial population structure in the colonic mucosa.

Nonetheless, while control/treatment differences could be detected using either cecal or fecal samples, conclusions are likely to differ. Therefore, when drawning wider conclusion through metanalyses originated from studies carried with both type of samples might be misleading. Thus, in conclusion, the choice of sampling site remains critical in experimental design on gut microbiota studies.

5.5 First time characterization of the ovine colonic gut microbiota

Background. Sarda breed sheep is the most consistent breed in Italy and is mainly bred in Sardinia. In this region it is important both economically and socially being the most relevant species of Sardinian livestock and having represented for centuries the sustenance of people living in rural areas. Its breeding throughout the past century had a decisive turning point linked to the market of sheep dairy products, like cheese. Indeed, Sardinia, with more than 3 million animals raised and distributed in 12,718 companies, is the leader Region of sheep farming in Italy; in fact, it holds more than 40% of the sheep population and about 60% of national milk production.

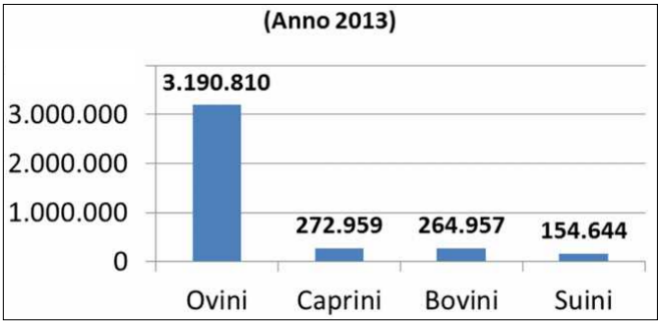

Figure 22. livestock numbers in the year 2013. Font: Laore Sardegna

Therefore, the achievement of further significant information concerning the physiology of these animals has a crucial economic relevance. Except for studies aimed to investigate the rumen, the microbial component of sheep has been poorly analyzed so far. In larger ruminants, the microbiota has been investigated for two important reasons: its role in environmental pollution due to methane production (Wang *et al*., 2012), and the possibility to obtain ideal candidates for industrial applications connected to the microbial ability to break down lignocelluloses (Yue *et al*., 2013). However, several intestinal diseases of

paramount importance affecting sheep and other ruminants alter the systemic health status, causing loss of animals and decreasing productivity on farm with a profound socio-economic impact. Some of these disease are related to bacteria, such as *Mycobacterium avium* subsp. Paratuberculosis in Johne's disease (Windsor PA 2014), *Campylobacter* spp. and E.coli O175:H7 (Lacasta D *et al*., 2015) as cause of enterocolitis, *Salmonella enterica* and in particular, serovar Abortusovis that is the most common causative agent of ovine salmonellosis in southern Europe (Uzzau S *et al*., 2001). To date, no studies have yet performed on the colonic microbiota in sheep and a dataset from healthy animals might be useful to add knowledge on the gut physiology of this important animal species.

Results. In the light of these premises, we have characterized the gut (colonic) microbiota of healthy Sarda breed sheep, in order to identify a dataset describing the sheep "core" microbiome that populated the colonic mucosa. For this purpose, stool samples from five sheep were pretreated with differential centrifugation and extracted as described in chapter 5.1.2. Sequencing of samples resulted in an average of 611329 ± 170798 total sequences and 148969 ± 36368 reads assigned.

	Total reads	Reads assigned
Ovine 1	715296	162654
Ovine 2	721181	174444
Ovine 3	549000	143126
Ovine 4	337529	88349
Ovine 5	733641	176270

Table 7. Reads numbers obtained from NGS and reads assigned by QIIME analysis

Number of observed OTUs and alpha diversity values was different according to different input OTU table. As expected, given the high presence of singleton and doubleton in raw table, the OTUs number and the Shannon index value was higher in raw table than others two input tables.

	Shannon index	Observed OTUs
Raw table	10.67 ± 0.17	9091 ± 1322
Filtered table	9.71 ± 0.11	2217 ± 84
Sub-sampling table	9.58 ± 0.10	1568 ± 35

Table 8. Shannon index and OTUs number from different tables

The QIIME analysis showed a similar taxonomic distribution in five sheep investigated. The bacterial community was dominated by Firmicutes and Bacteroidetes (mean: 7.99% ± 2.5% and 15.36% ± 2.86%, respectively), followed by Verrucomicrobia (mean: 2.67% ± 1.62%), Proteobacteria (1.82% ± 0.59%), Euryarchaeota and Actinobacteria (mean: 0.98% ± 0.77% and 0.36% ± 0.53%, respectively).

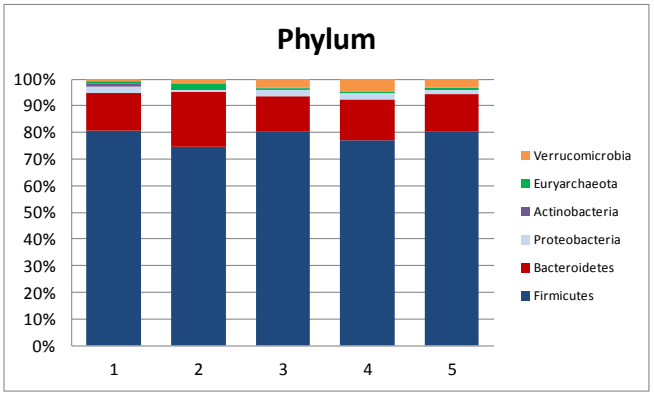

Figure 23. Bar charts of main phyla (> 0.5%)

Phylum	ov1	ov2	ov3	ov4	ov5
Firmicutes	79,47%	74,15%	79,98%	76,76%	79,58%
Bacteroidetes	13,97%	20,32%	13,34%	15,28%	13,91%
Proteobacteria	2,12%	0,91%	2,33%	2,20%	1,54%
Actinobacteria	1,31%	0,08%	0,14%	0,15%	0,10%
Euryarchaeota	0,85%	2,26%	0,68%	0,20%	0,92%
Verrucomicrobia	0,74%	1,56%	3,02%	4,94%	3,11%
Spirochaetes	0,57%	0,33%	0,18%	0,12%	0,51%
Lentisphaerae	0,41%	0,01%	0,01%	0,02%	0,01%
Tenericutes	0,34%	0,28%	0,18%	0,23%	0,16%
Fibrobacteres	0,08%	0,01%	0,00%	0,01%	0,01%
Elusimicrobia	0,06%	0,01%	0,01%	0,02%	0,00%
Cyanobacteria	0,04%	0,02%	0,02%	0,01%	0,02%
Planctomycetes	0,02%	0,05%	0,10%	0,05%	0,08%
Synergistetes	0,01%	0,00%	0,00%	0,01%	0,01%

Table 9. Bacterial phyla detected in ovine stool samples

At the class level, Clostridia (mean: 76.81% ± 2.12%), Bacteroidia (15.36% ± 2.86%), Verrucomicrobiae (2.40% ± 1.45%), Bacilli (1.12% ± 1.19%), Gammaproteobacteria (0.97% ± 0.46%), Methanobacteria (0.92% ± 0.79%) and Deltaproteobacteria (0.77% ± 0.22%) were the most abundant bacteria (mean > 0.5%).

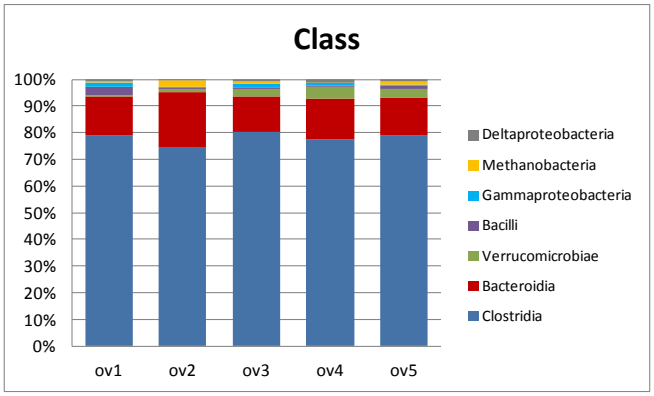

Figure 24. Bar charts of main classes (> 0.5%)

Class	ov1	ov2	ov3	ov4	ov5
Clostridia	76,34%	73,82%	79,41%	76,32%	78,18%
Bacteroidia	13,97%	20,32%	13,34%	15,28%	13,91%
Verrucomicrobiae	0,54%	1,50%	2,79%	4,36%	2,82%
Bacilli	3,13%	0,26%	0,53%	0,38%	1,32%
Gammaproteobacteria	1,31%	0,41%	1,47%	1,07%	0,57%
Methanobacteria	0,55%	2,26%	0,68%	0,19%	0,92%
Deltaproteobacteria	0,67%	0,45%	0,85%	1,00%	0,89%
Spirochaetes	0,57%	0,33%	0,18%	0,12%	0,51%
Verruco-5	0,20%	0,06%	0,23%	0,58%	0,29%
Actinobacteria	1,29%	0,00%	0,03%	0,01%	0,02%
CK-1C4-19	0,25%	0,24%	0,17%	0,22%	0,14%
[Lentisphaeria]	0,41%	0,01%	0,01%	0,02%	0,01%
Coriobacteriia	0,01%	0,08%	0,11%	0,14%	0,08%
Planctomycetia	0,02%	0,05%	0,10%	0,05%	0,08%
Methanomicrobia	0,26%	0,00%	0,00%	0,00%	0,00%
Erysipelotrichi	0,01%	0,07%	0,04%	0,05%	0,07%
Epsilonproteobacteria	0,02%	0,03%	0,01%	0,10%	0,06%
Alphaproteobacteria	0,07%	0,02%	0,01%	0,02%	0,02%
RF3	0,09%	0,01%	0,00%	0,00%	0,00%
Fibrobacteria	0,08%	0,01%	0,00%	0,01%	0,01%
Elusimicrobia	0,06%	0,01%	0,01%	0,02%	0,00%
Mollicutes	0,00%	0,03%	0,01%	0,01%	0,02%

Betaproteobacteria	0,04%	0,00%	0,00%	0,01%	0,01%
Thermoplasmata	0,04%	0,00%	0,00%	0,00%	0,00%
Synergistia	0,01%	0,00%	0,00%	0,01%	0,01%

Table 10. Classes detected in ovine stool samples

Considering the families, large part of OTUs was attributable to members of Firmicutes: *Ruminococcaceae* (mean: 40.11% ± 6.08%), followed by *Lachospiraceae* (7.83% ± 1.14%), *Clostridiaceae* (4.18% ± 0.62%), *Mogibacteriaceae* (2.65% ± 1.46%), *Veillonellaceae* (2.11% ± 0.82%). Instead, *Bacteroidaceae*, *Rikenellaceae* and *Paraprevotellaceae* (mean: 4.90% ± 1.37%, 2.24% ± 0.38% and 1.03% ± 0.63%, respectively) were the most abundant members of Bacteroidetes. In addition, the families *Verrucomicrobiaceae*, *Methanobacteriaceae* and *Desulfovibrionaceae,* respectively, belonging to the Verrucomicrobia, Euryarchaeota and Proteobacteria phyla, showed an abundance higher than 0.5%.

Family	ov1	ov2	ov3	ov4	ov5
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae	49,31%	36,47%	37,19%	43,17%	34,43%
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;Other	12,20%	15,79%	24,18%	17,04%	18,05%
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae	7,70%	9,23%	6,39%	7,14%	8,67%
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;Other	6,58%	7,75%	5,25%	5,88%	5,84%
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae	3,37%	7,08%	4,33%	5,12%	4,59%
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae	3,59%	4,99%	4,43%	3,53%	4,35%
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Mogibacteriaceae]	0,28%	3,10%	3,90%	2,31%	3,67%
k__Bacteria;p__Verrucomicrobia;c__Verrucomicrobiae;o__Verrucomicrobiales;f__Verrucomicrobiaceae	0,54%	1,50%	2,79%	4,36%	2,82%
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae	2,16%	2,81%	1,74%	2,26%	2,20%
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae	1,40%	2,78%	1,63%	1,53%	3,18%
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptostreptococcaceae	0,83%	0,33%	0,45%	0,35%	4,13%
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__[Paraprevotellaceae]	0,45%	1,84%	1,31%	1,25%	0,33%
k__Archaea;p__Euryarchaeota;c__Methanobacteria;o__Methanobacteriales;f__Methanobacteriaceae	0,55%	2,26%	0,68%	0,19%	0,92%
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae	1,03%	0,37%	1,42%	0,53%	0,41%
k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__Desulfovibrionales;f__Desulfovibrionaceae	0,63%	0,43%	0,84%	0,98%	0,81%
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptococcaceae	0,45%	0,40%	0,44%	0,61%	0,52%
k__Bacteria;p__Firmicutes;c__Bacilli;o__Turicibacterales;f__Turicibacteraceae	0,58%	0,11%	0,34%	0,22%	0,91%
k__Bacteria;p__Spirochaetes;c__Spirochaetes;o__Spirochaetales;f__Spirochaetaceae	0,57%	0,33%	0,18%	0,12%	0,51%
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Christensenellaceae	0,15%	0,24%	0,36%	0,31%	0,40%
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae	0,79%	0,05%	0,12%	0,10%	0,27%
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Micrococcaceae	1,23%	0,00%	0,02%	0,01%	0,02%
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__RF16	0,62%	0,07%	0,10%	0,29%	0,11%

k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__p-2534-18B5	0,07%	0,35%	0,18%	0,16%	0,36%
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__BS11	0,13%	0,15%	0,34%	0,20%	0,29%
k__Bacteria;p__Tenericutes;c__CK-1C4-19;Other;Other	0,25%	0,24%	0,17%	0,22%	0,14%
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Paenibacillaceae	1,00%	0,00%	0,00%	0,00%	0,01%
k__Bacteria;p__Verrucomicrobia;c__Verruco-5;o__WCHB1-41;f__RFP12	0,19%	0,05%	0,21%	0,28%	0,24%
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Gracilibacteraceae	0,13%	0,21%	0,12%	0,11%	0,25%
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Porphyromonadaceae	0,55%	0,12%	0,04%	0,02%	0,06%
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Planococcaceae	0,67%	0,01%	0,03%	0,02%	0,06%
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae	0,05%	0,01%	0,04%	0,51%	0,08%
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Dehalobacteriaceae	0,04%	0,08%	0,18%	0,07%	0,13%
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Tissierellaceae]	0,17%	0,11%	0,04%	0,06%	0,07%
k__Bacteria;p__Lentisphaerae;c__[Lentisphaeria];o__Victivallales;f__Victivallaceae	0,41%	0,01%	0,01%	0,02%	0,01%
k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae	0,01%	0,08%	0,11%	0,14%	0,08%
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Prevotellaceae	0,02%	0,13%	0,06%	0,08%	0,10%
k__Bacteria;p__Verrucomicrobia;c__Verruco-5;o__WCHB1-41;f__WCHB1-25	0,01%	0,00%	0,02%	0,30%	0,04%
k__Bacteria;p__Planctomycetes;c__Planctomycetia;o__Pirellulales;f__Pirellulaceae	0,02%	0,05%	0,10%	0,05%	0,08%
k__Archaea;p__Euryarchaeota;c__Methanomicrobia;o__Methanomicrobiales;f__Methanocorpusculaceae	0,26%	0,00%	0,00%	0,00%	0,00%
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae	0,02%	0,09%	0,04%	0,03%	0,06%
k__Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae	0,01%	0,07%	0,04%	0,05%	0,07%
k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales;f__Campylobacteraceae	0,02%	0,03%	0,01%	0,10%	0,06%
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Aeromonadales;f__Succinivibrionaceae	0,08%	0,02%	0,01%	0,03%	0,08%
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__[Acidaminobacteraceae]	0,03%	0,03%	0,04%	0,03%	0,08%
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__SBYG_4172	0,03%	0,04%	0,04%	0,03%	0,05%
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Eubacteriaceae	0,00%	0,01%	0,01%	0,01%	0,15%
k__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__RF32;Other	0,07%	0,02%	0,01%	0,02%	0,02%
k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;o__GMD14H09;Other	0,03%	0,01%	0,01%	0,01%	0,07%
k__Bacteria;p__Tenericutes;c__RF3;o__ML615J-28;Other	0,09%	0,01%	0,00%	0,00%	0,00%
k__Bacteria;p__Fibrobacteres;c__Fibrobacteria;o__Fibrobacterales;f__Fibrobacteraceae	0,08%	0,01%	0,00%	0,01%	0,01%
k__Bacteria;p__Cyanobacteria;c__4C0d-2;o__YS2;Other	0,04%	0,02%	0,02%	0,01%	0,02%
k__Bacteria;p__Elusimicrobia;c__Elusimicrobia;o__Elusimicrobiales;f__Elusimicrobiaceae	0,06%	0,01%	0,01%	0,02%	0,00%
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Xanthomonadaceae	0,09%	0,00%	0,00%	0,00%	0,00%
k__Bacteria;p__Tenericutes;c__Mollicutes;o__RF39;Other	0,00%	0,03%	0,01%	0,01%	0,02%
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Microbacteriaceae	0,06%	0,00%	0,01%	0,00%	0,00%
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;Other	0,04%	0,00%	0,01%	0,00%	0,02%
k__Bacteria;p__WPS-2;Other;Other;Other	0,00%	0,00%	0,01%	0,03%	0,03%
k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Alcaligenaceae	0,04%	0,00%	0,00%	0,01%	0,01%
k__Bacteria;p__Firmicutes;c__Clostridia;Other;Other	0,01%	0,01%	0,01%	0,01%	0,01%
k__Archaea;p__Euryarchaeota;c__Thermoplasmata;o__E2;f__[Methanomassiliicoccaceae]	0,04%	0,00%	0,00%	0,00%	0,00%
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae	0,04%	0,00%	0,00%	0,00%	0,00%
k__Bacteria;p__Firmicutes;c__Clostridia;o__MBA08;Other	0,01%	0,00%	0,00%	0,00%	0,03%
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;Other;Other	0,02%	0,01%	0,01%	0,01%	0,00%
k__Bacteria;p__Proteobacteria;c__Deltaproteobacteria;Other;Other	0,01%	0,01%	0,00%	0,00%	0,02%
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__S24-7	0,00%	0,00%	0,00%	0,00%	0,04%
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__[Barnesiellaceae]	0,00%	0,01%	0,00%	0,01%	0,01%

k__Bacteria;Other;Other;Other;Other	0,01%	0,00%	0,01%	0,01%	0,01%
k__Bacteria;p__Synergistetes;c__Synergistia;o__Synergistales;Other	0,01%	0,00%	0,00%	0,01%	0,01%
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__[Exiguobacteraceae]	0,03%	0,00%	0,00%	0,00%	0,00%

Table 11. Families detected in ovine gut samples

Within the 97 genera identified, microbiota faecal composition showed a prevalence of *Clostridium*, *Ruminococcus, Akkermansia* and *Oscillospira* (mean: 3.28% ± 0.65%, 3.21% ± 1.77%, 2.39% ± 1.45% and 2.19% ± 0.83%, respectively). Instead, sulfate-reducing bacteria as *Desulfovibrio* and another cellulolytic bacteria as *Butyrivibrio* showed an abundance lower than 0.5%. Finally, Methanobrevibacter genus is the most dominant component of methanogen populations (mean: 0.71% ± 0.41%).

Genus	ov1	ov2	ov3	ov4	ov5
f__Bacteroidaceae;g__5-7N15	2,69%	5,39%	3,35%	3,85%	3,89%
f__Clostridiaceae;g__Clostridium	2,95%	4,20%	3,68%	2,96%	2,60%
f__Ruminococcaceae;g__Ruminococcus	1,24%	2,44%	2,47%	5,81%	4,07%
f__Verrucomicrobiaceae;g__Akkermansia	0,54%	1,49%	2,75%	4,35%	2,82%
f__Ruminococcaceae;g__Oscillospira	2,15%	1,47%	2,14%	3,57%	1,60%
f__[Mogibacteriaceae];g__Mogibacterium	0,07%	2,19%	2,50%	1,43%	2,39%
f__Veillonellaceae;g__Phascolarctobacterium	1,28%	2,14%	1,09%	1,27%	1,01%
f__Lachnospiraceae;g__Coprococcus	2,04%	1,20%	0,43%	0,66%	0,73%
f__Ruminococcaceae;g__Ruminococcusflavefaciens	0,07%	0,72%	0,44%	2,50%	0,91%
f__[Paraprevotellaceae];g__CF231	0,43%	1,58%	1,24%	0,86%	0,28%
f__Lachnospiraceae;g__Butyrivibrio	0,29%	0,69%	0,87%	0,61%	1,28%
f__Methanobacteriaceae;g__Methanobrevibacter	0,52%	1,46%	0,61%	0,16%	0,78%
f__Lachnospiraceae;g__Dorea	0,64%	0,63%	0,61%	0,54%	0,40%
f__Turicibacteraceae;g__Turicibacter	0,58%	0,11%	0,34%	0,22%	0,91%
f__Bacteroidaceae;g__Bacteroides	0,21%	0,72%	0,18%	0,38%	0,27%
f__Veillonellaceae;g__Succiniclasticum	0,06%	0,29%	0,28%	0,07%	1,07%
f__Spirochaetaceae;g__Treponema	0,57%	0,33%	0,18%	0,12%	0,51%
f__Pseudomonadaceae;g__Pseudomonas	0,48%	0,14%	0,42%	0,18%	0,09%
f__Peptococcaceae;g__rc4-4	0,24%	0,22%	0,21%	0,39%	0,26%
f__Bacillaceae;g__Bacillus	0,63%	0,04%	0,09%	0,08%	0,16%
f__Desulfovibrionaceae;g__Desulfovibrio	0,01%	0,09%	0,46%	0,25%	0,15%
f__Methanobacteriaceae;g__Methanosphaera	0,03%	0,68%	0,06%	0,03%	0,11%
f__Paenibacillaceae;g__Paenibacillus	0,82%	0,00%	0,00%	0,00%	0,01%
f__Micrococcaceae;g__Arthrobacter	0,75%	0,00%	0,01%	0,01%	0,01%
f__Porphyromonadaceae;g__Paludibacter	0,54%	0,09%	0,02%	0,01%	0,05%
f__[Paraprevotellaceae];g__[Prevotella]	0,01%	0,21%	0,05%	0,37%	0,04%
f__Clostridiaceae;g__SMB53	0,08%	0,11%	0,08%	0,06%	0,33%

f__[Mogibacteriaceae];g__Anaerovorax	0,02%	0,09%	0,14%	0,13%	0,17%
f__Veillonellaceae;g__Selenomonas	0,01%	0,11%	0,08%	0,04%	0,31%
f__Lachnospiraceae;g__Roseburia	0,04%	0,18%	0,12%	0,09%	0,11%
f__Veillonellaceae;g__Anaerovibrio	0,01%	0,12%	0,03%	0,05%	0,27%
f__Lachnospiraceae;g__Anaerostipes	0,12%	0,16%	0,03%	0,04%	0,06%
f__Prevotellaceae;g__Prevotella	0,02%	0,13%	0,06%	0,08%	0,10%
f__Pseudomonadaceae;g__Pseudomonasstutzeri	0,35%	0,00%	0,01%	0,01%	0,00%
f__[Tissierellaceae];g__Sedimentibacter	0,07%	0,11%	0,04%	0,06%	0,07%
f__Pseudomonadaceae;g__Pseudomonasveronii	0,08%	0,03%	0,16%	0,05%	0,03%
f__Lachnospiraceae;g__Blautia	0,06%	0,08%	0,07%	0,06%	0,07%
f__Planococcaceae;g__Solibacillus	0,33%	0,00%	0,00%	0,00%	0,00%
f__Lachnospiraceae;g__[Ruminococcus]	0,04%	0,07%	0,05%	0,05%	0,04%
f__Clostridiaceae;g__02d06	0,05%	0,03%	0,04%	0,04%	0,09%
f__Micrococcaceae;g__Microbisporarosea	0,20%	0,00%	0,01%	0,00%	0,00%
;f__Campylobacteraceae;g__Campylobacter	0,02%	0,03%	0,01%	0,10%	0,06%
f__Lachnospiraceae;g__Epulopiscium	0,20%	0,00%	0,00%	0,00%	0,00%
f__Streptococcaceae;g__Streptococcus	0,02%	0,07%	0,03%	0,03%	0,05%
f__SBYG_4172;Other	0,03%	0,04%	0,04%	0,03%	0,05%
f__Eubacteriaceae;g__Pseudoramibacter_Eubacterium	0,00%	0,01%	0,01%	0,01%	0,15%
f__Veillonellaceae;g__Selenomonasruminantium	0,00%	0,04%	0,02%	0,01%	0,12%
f__Lachnospiraceae;g__[Ruminococcus]gnavus	0,02%	0,06%	0,03%	0,03%	0,04%
f__Paenibacillaceae;g__Cohnella	0,18%	0,00%	0,00%	0,00%	0,00%
f__Clostridiaceae;g__Oxobacter	0,00%	0,05%	0,04%	0,04%	0,03%
f__Erysipelotrichaceae;g__Bulleidia	0,01%	0,03%	0,04%	0,04%	0,05%
f__Bacteroidaceae;g__Bacteroidesplebeius	0,02%	0,05%	0,02%	0,03%	0,03%
f__Planococcaceae;g__Lysinibacillusboronitolerans	0,14%	0,00%	0,00%	0,00%	0,00%
f__Ruminococcaceae;g__Ethanoligenens	0,04%	0,02%	0,02%	0,02%	0,04%
f__Succinivibrionaceae;g__Ruminobacter	0,07%	0,01%	0,00%	0,01%	0,05%
f__Peptococcaceae;g__Desulfosporosinusmeridiei	0,05%	0,01%	0,02%	0,02%	0,03%
f__Planococcaceae;g__Kurthiagibsonii	0,13%	0,00%	0,00%	0,00%	0,00%
f__[Acidaminobacteraceae];g__Acidaminobacter	0,02%	0,02%	0,02%	0,01%	0,05%
f__Bacillaceae;g__Bacillusmuralis	0,11%	0,00%	0,00%	0,00%	0,01%
f__Pseudomonadaceae;g__Pseudomonaspseudoalcaligenes	0,01%	0,01%	0,05%	0,02%	0,02%
f__Bacillaceae;g__Bacilluscereus	0,03%	0,01%	0,01%	0,01%	0,05%
f__Xanthomonadaceae;g__Stenotrophomonas	0,09%	0,00%	0,00%	0,00%	0,00%
f__Clostridiaceae;g__Alkaliphilus	0,08%	0,00%	0,00%	0,00%	0,00%
f__Clostridiaceae;g__Caloramator	0,01%	0,02%	0,01%	0,01%	0,03%
f__Pseudomonadaceae;g__Pseudomonasbalearica	0,08%	0,00%	0,00%	0,00%	0,00%
f__Ruminococcaceae;g__Faecalibacteriumprausnitzii	0,03%	0,01%	0,01%	0,00%	0,01%
f__Methanocorpusculaceae;g__Methanocorpusculum	0,07%	0,00%	0,00%	0,00%	0,00%
f__Peptococcaceae;g__Peptococcus	0,01%	0,02%	0,01%	0,01%	0,02%
f__Clostridiaceae;g__Proteiniclasticum	0,00%	0,01%	0,02%	0,01%	0,03%
f__Enterobacteriaceae;g__Erwiniachrysanthemi	0,01%	0,00%	0,00%	0,04%	0,01%
f__Ruminococcaceae;g__Ruminococcuscallidus	0,00%	0,00%	0,01%	0,02%	0,03%

f__Lachnospiraceae;g__Roseburiafaecis	0,00%	0,01%	0,02%	0,01%	0,02%
f__Verrucomicrobiaceae;g__Akkermansiamuciniphila	0,00%	0,02%	0,04%	0,01%	0,00%
f__Bacteroidaceae;g__BF311	0,00%	0,01%	0,01%	0,02%	0,01%
f__Alcaligenaceae;g__Sutterella	0,04%	0,00%	0,00%	0,01%	0,01%
f__Clostridiaceae;g__Alkaliphilustransvaalensis	0,01%	0,01%	0,01%	0,01%	0,01%
f__Streptococcaceae;g__Streptococcusalactolyticus	0,01%	0,02%	0,02%	0,01%	0,01%
f__Lachnospiraceae;g__Pseudobutyrivibrio	0,00%	0,03%	0,00%	0,00%	0,02%
f__Fibrobacteraceae;g__Fibrobactersuccinogenes	0,04%	0,00%	0,00%	0,01%	0,00%
f__[Paraprevotellaceae];g__YRC22	0,00%	0,02%	0,01%	0,01%	0,01%
f__Planococcaceae;g__Planomicrobium	0,01%	0,01%	0,01%	0,01%	0,02%
f__[Methanomassiliicoccaceae];g__vadinCA11	0,04%	0,00%	0,00%	0,00%	0,00%
f__Peptostreptococcaceae;g__Tepidibacter	0,00%	0,00%	0,01%	0,00%	0,03%
f__Clostridiaceae;g__Clostridiumperfringens	0,01%	0,00%	0,00%	0,00%	0,03%
f__Dehalobacteriaceae;g__Dehalobacterium	0,01%	0,00%	0,01%	0,01%	0,02%
f__[Paraprevotellaceae];g__Paraprevotella	0,01%	0,02%	0,02%	0,00%	0,00%
f__Moraxellaceae;g__Acinetobacter	0,04%	0,00%	0,00%	0,00%	0,00%
f__Erysipelotrichaceae;g__Bulleidiahoa12_73A10	0,00%	0,01%	0,00%	0,01%	0,02%
f__Lachnospiraceae;g__[Ruminococcus]torques	0,00%	0,01%	0,01%	0,01%	0,01%
f__Planococcaceae;g__Rummeliibacillus	0,00%	0,00%	0,01%	0,01%	0,02%
f__Lachnospiraceae;g__Moryella	0,00%	0,00%	0,00%	0,00%	0,03%
f__Victivallaceae;g__Victivallisvadensis	0,03%	0,00%	0,00%	0,00%	0,00%
f__Gracilibacteraceae;g__Lutispora	0,01%	0,00%	0,00%	0,01%	0,01%
f__Microbacteriaceae;g__Leucobacter	0,03%	0,00%	0,00%	0,00%	0,00%
f__Porphyromonadaceae;g__Parabacteroides	0,00%	0,02%	0,01%	0,00%	0,00%
f__Gracilibacteraceae;g__Gracilibacter	0,01%	0,00%	0,00%	0,01%	0,00%
f__Succinivibrionaceae;g__Succinivibrio	0,00%	0,01%	0,00%	0,00%	0,01%

Table 12. List of genera detected in ovine stool samples

5.5.1 Discussion and conclusion

Bacterial diversities within the rumen of sheep have been investigated in recent years as a result of development of 16S rRNA analysis performed by Next Generation Sequencing, yet similar data on the gut microbiome diversity in the lower gastrointestinal tract of sheep are limited and obtained through methods classified as "classical", such as PCR and culturing, while new approaches have been less used. However, new technologies have broadened our appreciation of the diversity and complexity of the microbiome. A deeper understanding of the entire intestinal microbiome in ruminants (bovine), including those taxa that are not core members, is reshaping research questions and hypotheses addressing societal and economical pressures to decrease methane emission or nitrogen excretion while improving fiber digestibility, feed intake, feed efficiency and animal health. In this contest, our work is the first study performed on stool samples using Next Generation Sequencing techniques to characterize the gut microbiome populating the colon of healthy sheep. These results, then, represent the first description of the ovine fecal microbiome and demonstrate its outstanding biological diversity. Moreover, we have identified in stool samples bacteria involved in the process of nitrate reduction in the rumen as *Selemonas ruminantium* and *Campylobacter* (Zhao *et al*., 2015), sulfate reduction bacteria as *Desulfovibrio* (Howard and Hungate, 1976), cellulolytic bacteria as *Ruminococcus flavefaciens*, *Fibrobacter succinogenes*, *Butyrivibrio* and *Clostridium* genera (Ransom-Jones *et al*., 2012), methanogens Archaea as *Methanobrevibacter*, *Methanosphaera* (Jansenn and Kirs, 2008; Hook *et al.*, 2010). Thanks to their ability to efficiently digest fibrous materials, the cellulolytic bacteria have gained attention from researchers in animal husbandry and biology energy, while the interest for methanogenic bacteria was renewed in the past decade because enteric methane generation contributes to global anthropogenic greenhouse gas emissions and represents a 2–12% loss of feed energy for the animal (Ripple W *et al*., 2014; Johnson D and Ward, 1996). Therefore, even from the fecal samples, it appear possible to gain information of the bacterial player that transform feed (plant material) in the sheep rumen from a given individual under study or under feeding trial controls. Moreover, based on these data, and on the methods we have optimized to sequence the metagenome of the colonic sheep microbiota, large-scale studies could be carried out to correlate changes in sheep gut microbiota to zootechnical and production variables, and with the final aims of optimizing livestock animals productivity, their protection from the numerous communicable and not

communicable disease and, finally, the modern post-genomic approaches for their continuous genetic selection.

5.6 A metagenomic dataset to perform association studies linking gut microbiota to T1D host traits in a murine model

Background. Type 1 diabetes (T1D) is a multifactorial disease that results from an inflammation of the pancreatic islets of Langerhans, termed insulitis, and subsequent destruction of the insulin producing β cells in genetically predisposed individuals upon environmental stimulation (Lehuen A *et al.*, 2010). Although genetic factors can predispose an individual to T1D, twin and family studies show that only a fraction of those genetically predisposed individuals (less than 50%) will develop the disease (Todd JA *et al*.,2007; Barret JC *et al*., 2009; Redondo MJ *et al*., 2001). Thus, it is strongly believed that environmental factors are important for the initiation and development of T1D, including viral infection and diet (Bach JF *et al*., 2002). Further during the last decade, a number of studies showed an association with gut microbiota in the pathogenesis of T1D; indeed, gut microbiota play an important role in the regulation of autoimmunity, tolerance and intestinal permeability that are several mechanisms by which gut microbiota could affect the development of T1D (Changyun H *et al*., 2015). In this contenst, non-obese Diabetic (NOD) mice spontaneously develop T1D and are model frequently used for studying the etiology of this disease (Castano *et al*., 1990). Indeed, NOD mice develop insulitis by the age of approximately 10 weeks, shortly after being weaned from their mothers and their (On average, 60-80% of female and 10-20% of male NOD) tend to develop overt hyperglycemia indicating diabetes onset between 12-20 weeks of age (peak = 16-18 weeks). Moreover, the NOD mouse has a very special MHC class II region; it lacks the E-complex owing to a deletion in the E_{α} gene (Hattori *et al*., 1986) and the A-complex is expressed in a form unique to the NOD strain, where aspartic acid, at position 57 of the β chain, is replaced by serine (Achea-Orbea H *et al*., 1987). Both restoration of the E complex (Nishimoto H *et al*., 1987; Bohme J *et al*, 1990) or insertion of a normal A complex (Slattery RM *et al*., 1990) by creation of class II transgenic NOD mice was shown to protect from disease. In particular, the transgenic

undisrupted E_α gene gives NOD mice total protection from insulitis at least up to 1 year of age.

Results. In the light this background, we begun a collaboration with the group of Diane Mathis at Harvard Medical School (Boston, MA, USA) to assess possible microbiota biomarkers of TD1 susceptibility at the onset time of insulitis (pre-clinical phase) and in clinical onset timing. For this purpose, we received stool samples collected at 3 weeks and 10 weeks of age from female wild type NOD mice and class II transgenic NOD mice that express E_α transgene (E_α16 mice).

To correct for any possible effect due to cage environment and maternal group, E_α16 and NOD mice were obtained crossing NOD dam with E_α16 mice sire. Progeny was housed mixing in the same cage both genotypes.

All approaches, as listed in paragraph 5.3, to measure alpha-diversity in 3 weeks old mice, and differential analysis performed by DeSeq2, consistently assessed no changes associated to genetic background (7.48 ± 0.27 in NOD mice vs 7.36 ± 0.36 in Eα16 mice; $P = 0.6134$).

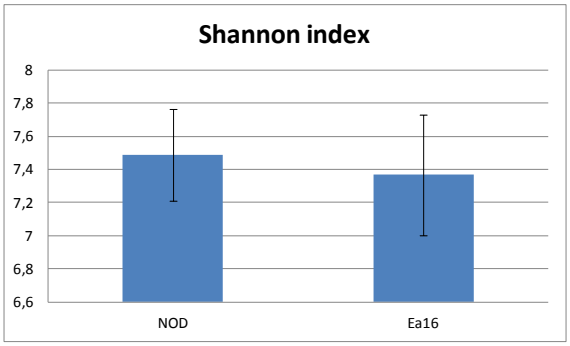

Figure 25. Shannon index value between NOD and Ea16 mice at 3 weeks old

On the contrary, beta diversity analysis (PCoA unweighted Unifrac distance metric) demonstrated a maternal effect dominating the microbial taxonomy distribution.

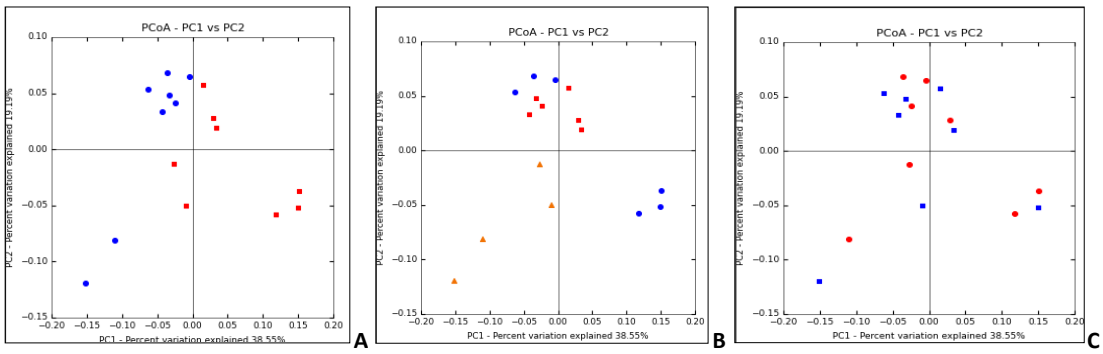

Figure 26. Plots showing the diversity among 16 fecal samples. **A)** red squares = 3 weeks samples, blue circles = 10 weeks samples, **B)** each color defines a different cage group; **C)** red circles = Eα16 mice, blue squares = NOD mice

Gut microbiota of mice at 10 weeks of age, instead, showed different diversity metrics and taxonomy according to genotype. Alpha diversity value was higher in Eα16 mice compared to NOD mice (8.81 ± 0.23 vs 8.61± 0.12; $P = 0.05339$, respectively).

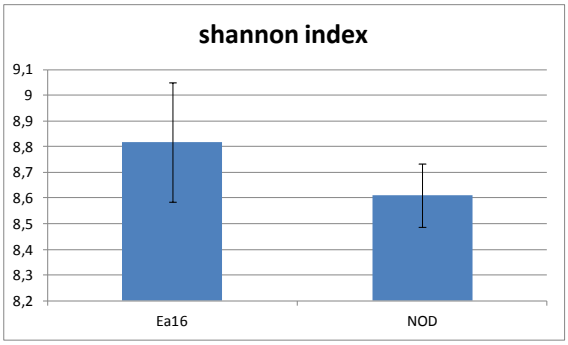

Figure 27. Shannon index values in Eα16 and NOD mice

As illustrated in the pie charts below, the Firmicutes group was more abundant in Eα16 mice compared to NOD (12% ± 9.2% vs 5% ± 2.6%; $P = 0.1075$, respectively); among the members of this phylum, most of the OTUs classified at the family level as *Ruminococcaceae* were significantly differently abundant in the two groups (Eα16 mean: 3.88% ± 2.02%; NOD mean: 1.81% ± 0.83%; $P = 0.037756$); the abundance of *Oscillospira* genus, representing 90% of this family group, was significantly different between the two genotype (Eα16 mean: 2% ± 0.89%; NOD mean: 0.94 ± 0.34%; $P = 0.029076$).

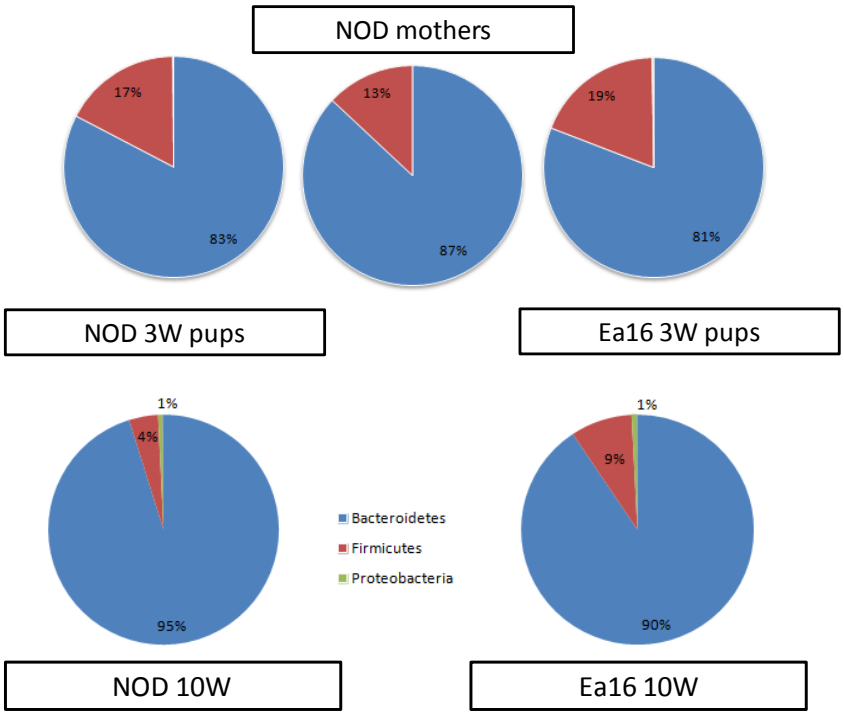

Figure 28. Pie charts of principal phyla between different genotype in mice at 3 and 10 weeks old (16 fecal samples)

Noteworthy, also genera members of *Lachnospiraceae* family as *Blautia* and *Dorea* (phylum Firmicutes), showed variation in its relative abundance according to genetic background (0.09% ± 0.08% in Eα16 vs 0.03% ± 0.02% NOD, $P = 0.048691$ and 0.06% ± 0.08% in Eα16 vs 0.01% ± 0.02% in NOD mice, $P = 0.033131$; respectively).

5.6.1 Discussion and conclusion

Several studies have demonstrated that patients with T1D are characterized by a moderate degree of gut microbial dysbiosis. However, there is a paucity of studies investigating the gut microbiota changes occurring before the clinical onset in genetically susceptible hosts, leaving substantial controversies regarding the relevance of the effect of the body's metabolism on the microbiota and/or the effects of the latter over this immunological and metabolic disorder. A better understanding of how gut bacteria-induced immunoregulation contributes to the pathogenesis of T1D is, therefore, necessary.

In our study, we found no differences in the microbial composition of the microbiota between the two groups with different genotypes at 3 weeks of age.

Yet, a close relationship was found with the maternal microbiota, attributable to the recent weaning of the pups from the mother. Instead, seven weeks later, at the time of insulitis development, but still before the clinical onset, mice gut microbiota showed a different taxonomy according their genetic susceptibility to T1D. This appears important because at that age, detection of microbial changes might serve to monitor the pre-clinical phase of the diabetes. Heightened gut permeability has been demonstrated to be one of the phenomena that precede the clinical onset of T1D in both animal models of autoimmune diabetes, as well as in patients with T1D and prediabetic individuals (Neu J *et al*., 2005; Bosi E *et al*., 2006; Vaarala O *et al.*, 2008). It has been suggested that the imbalance of bacteria, such as Bacteroidetes, which ferment short-chain fatty acid (SCFA), can affect the gut permeability (Schwartz RF *et al*., 2007; Brugman S *et al*., 2006). Our work, in accordance with other published studies (Giongo A *et al*., 2011; Schwartz RF *et al*.,2007; Roesch LF *et al*., 2009), showed a higher abundance of bacteroidetes in mice susceptible to develop diabetes; on the contrary, Eα16 mice showed a blooming of Firmicutes, including *Lachnospiraceae* and *Ruminococceae* families. However, we have to be cautious in thinking that a modulation of the gut microbiota may be useful as therapeutic intervention. Nevertheless, the advantages of microbial therapies are obvious: less expensive, less invasive and potentially long-lasting beneficial effects. Thus, extending knowledge on specific host and gut microbial composition and functional pathways involved in the development of T1D, may shed lights on the potential of novel microbiota-targeted therapeutic approaches to prevent or treat T1D.

5.7 Dietary and immunogenetic background impact over gut microbiota in a translational murine model of NAFLD

Background. Non-alcoholic fatty liver disease (NAFLD) is the most common liver disease worldwide (Sattar N *et al*., 2014). Its importance is due to not only its prevalence but also its evolution toward non-alcoholic steatohepatitis (NASH) and its association with increased risk of hepatic (e.g. cirrhosis and hepatocellular carcinoma HCC) and extra-hepatic (e.g. type 2 diabetes mellitus, cardiovascular disease, chronic kidney disease and cancer) complications, and therefore tends to be associated with increased morbidity/mortality and health expenditure (Vernon G et *al.*, 2011; Vuppalanchi R and Chalasani N, 2009). Immunogenetics traits might play a significant role in development of NAFLD.

Inflammasomes are intracellular multiprotein complexes, expressed in both parenchymal and non-parenchymal cells of the liver that, in response to cellular danger signals, activate caspase-1, with the release of the pro-inflammatory cytokines interleukin-1β (IL-1β) and interleukin-18 (IL-18) (Szabo and Csak 2012). The NLRP3 inflammasome, of the NOD-like receptor family, is one of the most extensively studied inflammasome component and is able of sensing a wide variety of alarm signals, both exogenous as pathogen-associated molecular patterns (PAMPS) and endogenous as danger-associated molecular patterns (DAMPS) (Martinon and Tschopp 2005). The presence of NLRP3 and/or inflammasome activation has been shown in several liver cell phenotype, such as hepatic stellate cells (Watanabe, Sohail et al. 2009), macrophages (Stienstra, Saudale *et al*. 2010) and hepatocytes (Csak, Ganz *et al*. 2011). Its role in the pathophysiology of NASH is under investigation, since NLRP3 inflammasome components are increased in various diet-induced NASH models in mice and in NASH patients (Csak, Ganz *et al*. 2011; De Minicis, Agostinelli *et al*. 2014; De Minicis, Rychlicki *et al*. 2014). Recent studies have implicated that the gut microbiota might play a role in the development of NAFLD (Mehal 2013). Oral treatment of lean germ-free mice with the cecal microbiota of obese mice caused an increase in hepatic triglyceride accumulation (Backhed, Ding *et al*. 2004). Further, obese humans are enriched in the microbial energy-harvesting phylum Firmicutes, which can directly improve energy yield from intestinal contents leading to obesity-associated NAFLD (Turnbaugh, Ley *et al*. 2006). Dietary habits influences the composition of gut microflora, can promote bacterial overgrowth and translocation (De Minicis, Rychlicki *et al*. 2014), allowing intestinal bacteria, or their products, to reach the liver through the portal circulation. Bacterial products are sensed by the toll-like receptor (TLRs) systems to elicit a pro-inflammatory and pro-fibrogenic response at least partially mediated by the inflammasome system (Miura, Seki *et al*. 2010; Tilg and Moschen 2015). Moreover, in the gut the inflammasome has an important role in regulating epithelial permeability (Zaki, Boyd *et al*. 2010) and microbial ecology (Elinav, Strowig *et al*. 2011). On this regard genetic inflammasome deficiency-associated dysbiosis has been hypothesized to result in abnormal accumulation of bacterial products in the portal circulation and in increased severity of NASH (Henao-Mejia, Elinav *et al*. 2012).

Results. To validate our methodological approaches in another translational model of gut microbiota – traits relationships, they were applied to elucidate the dietary and immunogenetic background effects on microbiota community. Specifically, to assess the influence of a "westernized" high fat and high carbohydrate diet with free fructose and sucrose (HFHC diet) and the lack of NLRP3 inflammasome on gut microbial diversity, we evaluated the structural differences that possibly occur in WT and NLRP3-/- mice fed with either HFHC or chow (standard) diet. In this model, cecal contents were analyzed since a single time point at the end of the experiment was evaluated. This activities were part of a collaboration with the group of Prof. Gianluca Svegliati Baroni at Unversità Politecnica delle Marche (Gastroenterology Department).

The microbial communities associated to each group of 3-5 animals, for a total of 15 cecal samples, were compared according to their α-diversity and β-diversity. Significant differences in α-diversity were detected on either WT and Nlrp3$^{-/-}$ mice when fed different diets. Although significant, a less pronounced difference was observed between WT and Nlrp3$^{-/-}$ chow fed mice, while similarly low α-diversity values were recorded in either WT and Nlrp3$^{-/-}$ mice when fed with HFHC.

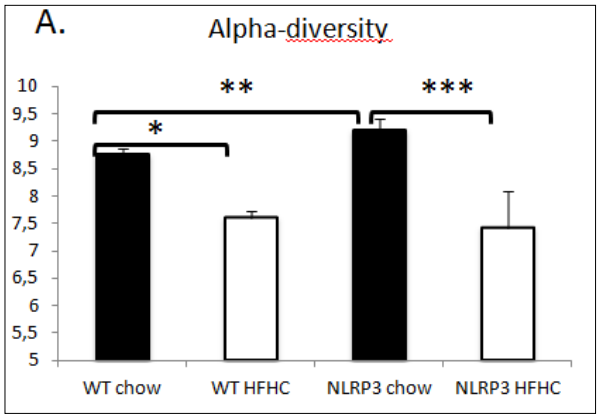

Figure 29. Gut microbiota alpha-diversity values in the 4 groups: calculated Shannon's Index was statistically different as indicated by stars:*P < 0.0001; **P= 0.04;***P= 0.01.

Analysis of β-diversity clearly illustrates group-level differences in the taxonomic composition, with the highest variation according to diet treatment, and the variation due to genetic background being increased in HFHC fed mice.

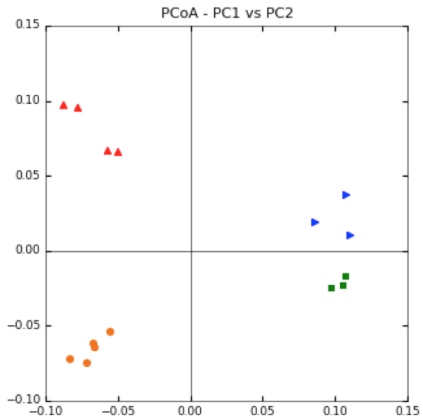

Figure 30. Gut microbiota beta-diversity values in the 4 groups: PC1 and PC2 explained, respectively, 50.82% and 24.91% of variation; Light triangles = NLRP3-/- HFHC-fed, dark triangles = NLRP3-/- Chow-fed, circles = wt HFHC-fed, squares = WT Chow-fed

We also found significant correlations between microbial taxa, lack of NLRP3 inflammasome and diet. Firmicutes largely outcompeted Bacteroidetes in HFHC fed WT mice. In WT mice, HFHC diet appeared also to promote an increased abundance of Proteobacteria (3.21% ± 0.88% vs 0.61% ± 0.07% in chow fed mice; $P = 0.009$). However, a blooming of Proteobacteria occurred in an even more dramatic fashion in Nlrp3$^{-/-}$ HFHC fed mice compared to chow fed mice (averaging 28.48% ± 5.14%; $P = 0.001$). The increased abundance concerned mostly OTUs classified at the family level as *Desulfovibrionaceae*, including *Desulfovibrio* and *Bilophila* pathobiont genera, representing 73% and 22% of this family group, respectively. Noteworthy, another genus, *Akkermansia* (phylum Verrucomicrobia), showed great variation in its relative abundance according to diet and genetic background. *Akkermansia* accounted for 2.58% ± 0.08% of the microbial community in chow fed wt mice, while it was almost undetectable (0.006%) in HFHC WT fed mice. In mice lacking NLRP3, while treated with chow diet, *Akkermansia* abundance dropped to 0.63 ± 0.26%. Interestingly, relative abundance of this genus showed a dramatic increase in Nlrp3$^{-/-}$ mice when fed with HFHC diet (12.11% ± 7.02%).

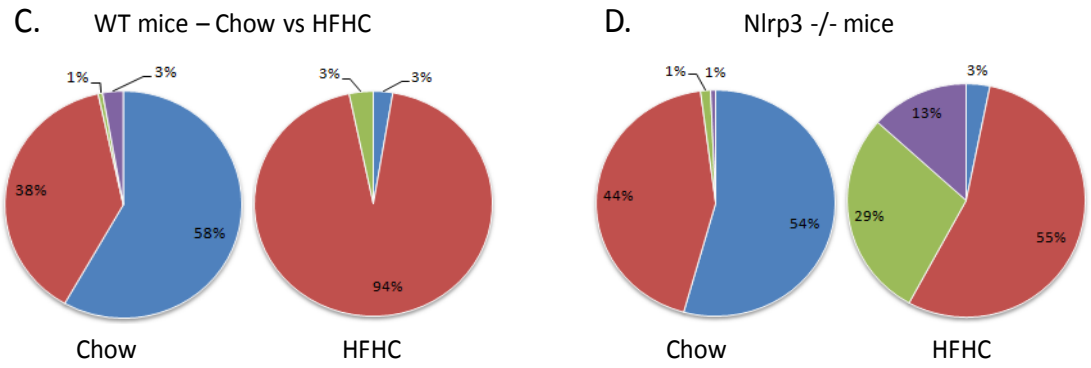

Figure 31. Pie charts of more abundant phyla. In red, Firmicutes; in blue, Bacteroidetes; in purple, Verrucomicrobia; in green, Proteobacteria

All taken together, these data reveal an altered gut microbiota in response to HFHC diet, with an expected blooming of energy harvesting microbiota members (i.e. Firmicutes) and displacement of the Bacteroidetes phylum. Further, our data showed changes of abundance of pathobionts like *Bilophila* (> 4 fold change in HFHC diet) whose increase is not dependent on host genetic background, as well as an overwhelming increase of *Akkermansia* and *Desulfovibrio* in HFHC diet that, instead, appeared strongly dependent on NLRP3 deficiency.

5.7.1 Discussion and conclusion

Bacteroidetes/Firmicutes ratio is frequently reported to be reduced in mice models with high fat diet and in obese human subjects (Kim KA, 2012; Turnbaugh, 2006; Verdam FJ, 2013). Thus, our observation on blooming of Firmicutes in HFHC fed mice was expected. In addition, Proteobacteria are less represented in eubiotic gut microbiota and a wealth of studies support the hypothesis that a bloom of this bacterial taxa is associated to an unstable gut microbial community and might actively promote intestinal inflammation (Shin NR, 2015). To this extent, we have reported that the cecal contents from HFHC fed mice, particularly in the NLRP3$^{-/-}$ group, carry an increase abundance of Proteobacteria and, in particular, of the mucus degrading *Desulfovibrio* genus. In our murine model of NAFLD, this data is of interest since, in addition to tight junctions, bacterial translocation across the intestinal wall is limited by the mucus layer. The whole intestinal surface is covered by a mucus layer attached to the enterocytes, making them poorly accessible to luminal bacteria. Mucins,

the mucus 'building blocks', are the substrate for mucus-degrading bacteria such as *A. muciniphila*, a gut commensal that has been proposed to modulates host's cellular pathways involved in basal metabolism homeostasis and immune tolerance toward commensal microbiota (Derrien M, 2011). However, exaggerate mucus degradation by *A. muciniphila* might contributes to intestinal inflammation due to the increase layer crossing by luminal antigens (Ganesh BP, 2013). Mucus layer is composed by two mucin chemotypes, distinguished as sulphated and sialyated. Sulphomucins are mature forms predominant in healthy condition and derived by sialyated mucins through posttranslational modification. A reduction on sulphated mucin forms has been shown in acute inflammation and this variation reflected the degree of mucosal inflammation, suggesting that a lowered amount of sulphomucins might correlate with an excessive degradation of the mucus layer and with its increased permeability to PAMPS. In addition, the reduction of sulphomucin abundance has been associated to increase relative abundance of *Desulfovibrio*, a species that as well as *A. muciniphila*, is capable to harvest sulphate from intestinal mucins dissimilatory sulphate reduction (G. Lennon, 2013). In keeping with these previous observations, we showed a dramatic increase in *Desulfovibrio* and *Akkermansia* in the cecal samples of NLRP3$^{-/-}$ mice following treatment with a HFHC diet. Therefore, it could be conceived that, in addition to tight junction loosening, mucus layer chemotype variation and degradation by sulphate reducing bacterial species might concur to the increased permeability to PAMPS and TLR activation in NLRP3$^{-/-}$ mice fed with HFHC diet. Indeed, intestinal permeability (G. Svegliati Baroni personal communication) was increased significantly in HFHC fed mice compared to Chow fed mice.

Noteworthy, while its presence has so far been consistently associated to healthy mucosa, *Akkermansia* increased abundance has been recently reported in Rag$^{-/-}$ mice, a model that lacks all mature lymphocytes (Zhang H, ISME J 2015). Thus, we report here a similar effect in a different mouse model of adaptive immunity deficiency, strengthening the hypothesis that *Akkermansia* may serve as a biomarker of immunodeficiency and that its uncontrolled blooming might worsen mucosa inflammation and the leaky gut condition.

References

Abu Al-Soud and Rådström P. Purification and Characterization of PCR-Inhibitory Components in Blood Cells. J Clin Microbiol. 2001; 39(2): 485–493.

Abu-Shanab A, Quigley EM. The role of the gut microbiota in nonalcoholic fatty liver disease. Nat Rev Gastroenterol Hepatol. 2010;7:691–701.

Acha-Orbea, H., and H. O. McDevitt.. The first external domain of the nonobese diabetic mouse class II I-A 3 chain is unique. Proc. Natl. Acad. Sci. USA. 1987; 84:2435-2439.

Ahern, P P, Faith, J J, Gordon, J I. Mining the Human Gut Microbiota for Effector Strains that Shape the Immune System. Immunity.2014; 05.012.

Alam C, Bittoun E, Bhagwat D, Valkonen S, Saari A, Jaakkola U, Eerola E, Huovinen P, Hänninen A. Effects of a germ-free environment on gut immune regulation and diabetes progression in non-obese diabetic (NOD) mice. Diabetologia. 2011;54(6):1398-406.

al-Waiz M, Mikov M, Mitchell SC, Smith RL. The exogenous origin of trimethylamine in the mouse. Metabolism. 1992; 41: 135–136.

Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11(10):R106.

Arthur JC, Perez-Chanona E, Mühlbauer M, Tomkovich S, Uronis JM, Fan TJ, Campbell BJ, Abujamel T, Dogan B, Rogers AB, Rhodes JM, Stintzi A, Simpson KW, Hansen JJ, Keku TO, Fodor AA, Jobin C. Intestinal inflammation targets cancer-inducing activity of the microbiota. Science. 2012; 338(6103):120-3.

Atarashi K, Tanoue T, Shima T, Imaoka A, Kuwahara T, Momose Y, Cheng G, Yamasaki S, Saito T, Ohba Y, Taniguchi T, Takeda K, Hori S, Ivanov II, Umesaki Y, Itoh K, Honda K. Induction of colonic regulatory T cells by indigenous Clostridium species. Science. 2011; 331(6015):337-41.

Bach JF. The effect of infections on susceptibility to autoimmune and allergic diseases. N Engl J Med. 2002;347:911–20.

Backhed F, Ding H, Wang T, Hooper LV, Koh GY, Nagy A, *et al*. The gut microbiota as an environmental factor that regulates fat storage. Proc Natl Acad Sci USA. 2004; 101(44):15718–15723.

Bäckhed F, Ding H, Wang T, Hooper LV, Koh GY, Nagy A, Semenkovich CF, Gordon JI. The gut microbiota as an environmental factor that regulates fat storage. Proc Natl Acad Sci U S A. 2004;101(44):15718-23.

Backhed F, Manchester JK, Semenkovich CF, Gordon JI. Mechanisms underlying the resistance to diet-induced obesity in germ-free mice. Proc Natl Acad Sci USA. 2007; 104(3):979–984.

Bäckhed, F., Ley, R.E., Sonnenburg, J.L., Peterson, D. a, and Gordon, J.I. Host-bacterial mutualism in the human intestine. Science. 2005;307, 1915–1920.

Barrett JC, Clayton DG, Concannon P, Akolkar B, Cooper JD, Erlich HA, Julier C, Morahan G, Nerup J, Nierras C, Plagnol V, Pociot F, Schuilenburg H, Smyth DJ, Stevens H, Todd JA, Walker NM, Rich SS. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat Genet. 2009;41(6):703-7.

Bentley DR *et al*., Accurate whole human genome sequencing using reversible terminator chemistry. Nature. 2008;456(7218):53-9.

Böhme J, Schuhbaur B, Kanagawa O, Benoist C, Mathis D. MHC-linked protection from diabetes dissociated from clonal deletion of T cells. Science. 1990 Jul;249(4966):293-5.

Bosi E, Molteni L, Radaelli MG, Folini L, Fermo I, Bazzigaluppi E, Piemonti L, Pastore MR, Paroni R. Increased intestinal permeability precedes clinical onset of type 1 diabetes. Diabetologia. 2006;49(12):2824-7.

Bouskra D, Brézillon C, Bérard M, Werts C, Varona R, Boneca IG, Eber G. Lymphoid tissue genesis induced by commensals through NOD1 regulates intestinal homeostasis. Nature. 2008; 456(7221):507-10.

Brestoff JR & Artis D. Commensal bacteria at the interface of host metabolism and the immune system. Nature Immunology. 2013;14, 676–684.

Brown C T, Davis-Richardson A G, Giongo A, Gano K A, Crabb D B, Mukherjee N, Casella G, Drew J C, Ilonen J, Knip M, Hyöty H, Veijola R, Simell T, Simell O, Neu J, Wasserfall C H, Schatz D, Atkinson M A, Triplett E W. Gut microbiome metagenomics analysis suggests a functional model for the development of autoimmunity for type 1 diabetes. PLoS One. 2011; 6(10):e25792.

Brugman S1, Klatter FA, Visser JT, Wildeboer-Veloo AC, Harmsen HJ, Rozing J, Bos NA. Antibiotic treatment partially protects against type 1 diabetes in the Bio-Breeding diabetes-prone rat. Is the gut flora involved in the development of type 1 diabetes?. Diabetologia. 2006; 49(9):2105-8.

Bullard JH, Purdom E, Hansen KD and Dudoit S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics. 2010; 11:94.

Cahenzli J, Köller Y, Wyss M, Geuking MB, and McCoy KD. Intestinal Microbial Diversity during Early-Life Colonization Shapes Long-Term IgE Levels. Cell Host Microbe. 2013; 14(5): 559–570.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7(5):335-6.

Chakravorty S, Helb D, Burday M, Connell N, and Alland D. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. J Microbiol Methods. 2007; 69(2): 330–339.

Clarridge JE. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. Clin Microbiol Rev. 2004 Oct;17(4):840-62.

Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, *et al*. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res. 2009; 37(Database issue): D141–D145.

Cryan JF, Dinan TG. Mind-altering microorganisms: the impact of the gut microbiota on brain and behaviour. Nat Rev Neurosci. 2012;13(10):701-12.

Csak T, Ganz M, Pespisa J, Kodys K, Dolganiuc A, Szabo G. Fatty acid and endotoxin activate inflammasomes in mouse hepatocytes that release danger signals to stimulate immune cells. Hepatology. 2011; 54(1):133-44.

de Goffau M C, Fuentes S, van den Bogert B, Honkanen H, de Vos W M, Welling G W, Hyöty H, Harmsen H J. Aberrant gut microbiota composition at

the onset of type 1 diabetes in young children. Diabetologia. 2014; 57(8):1569-77.

de Goffau M C, Luopajärvi K, Knip M, Ilonen J, Ruohtula T, Härkönen T, Orivuori L, Hakala S, Welling G W, Harmsen H J, Vaarala O. Fecal microbiota composition differs between children with β-cell autoimmunity and those without. Diabetes. 2013; 62(4):1238-44.

De Minicis S, Agostinelli L, *et al*. HCC Development Is Associated to Peripheral Insulin Resistance in a Mouse Model of NASH. PLoS One. 2014; 9(5): e97136.

Derrien M, Van Baarlen P, Hooiveld G, Norin E, Müller M, de Vos WM. Modulation of Mucosal Immune Response, Tolerance, and Proliferation in Mice Colonized by the Mucin-Degrader Akkermansia muciniphila. Front Microbiol. 2011;2:166.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol. 2006;72(7):5069-72.

Douglas-Escobar M. Effect of Intestinal Microbial Ecology on the Developing Brain. JAMA pediatrics. 2013; 167(4):1-6.

Dumas ME, Barton RH, Toye A, Cloarec O, Blancher C, Rothwell A, Fearnside J, Tatoud R, Blanc V, Lindon JC, Mitchell SC, Holmes E, McCarthy MI, Scott J, Gauguier D, Nicholson JK. Metabolic profiling reveals a contribution of gut microbiota to fatty liver phenotype in insulin-resistant mice. Proc Natl Acad Sci U S A. 2006; 103(33):12511-6.

Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26(19):2460-1.

Elinav E, Strowig T, Kau AL, Henao-Mejia J, Thaiss CA, Booth CJ, Peaper DR, Bertin J, Eisenbarth SC, Gordon JI, Flavell RA. NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. Cell. 2011;145(5):745-57.

Ferrand J, Patron K, *et al*. Comparison of seven methods for extraction of bacterial DNA from fecal and cecal samples of mice. J Microbiol Methods. 2014;105:180-5.

Fleissner CK, Huebel N, Abd El-Bary MM, Loh G, Klaus S, Blaut M. Absence of intestinal microbiota does not protect mice from diet-induced obesity. Br J Nutr. 2010; 104(6):919–929.

Fleissner CK, Huebel N, Abd El-Bary MM, Loh G, Klaus S, Blaut M. Absence of intestinal microbiota does not protect mice from diet-induced obesity. Br J Nutr. 2010; 104(6):919–929.

Fouhy F, Guinane C, Hussey S, Wall R, C. Ryan A, Dempsey E, Murphy B, Ross P, Fitzgerald G, Stanton C. High-Throughput Sequencing Reveals the Incomplete, Short-Term Recovery of Infant Gut Microbiota following Parenteral Antibiotic Treatment with Ampicillin and Gentamicin. Antimicrobial Agents and Chemotherapy. 2012; 56(11): 5811–5820.

Fukuda S, Toh H, Hase K, Oshima K, Nakanishi Y, Yoshimura K, Tobe T. Bifidobacteria can protect from enteropathogenic infection through production of acetate. Nature. 2011; 469, 543–547.

Ganesh BP, Klopfleisch R, Loh G, Blaut M. Commensal Akkermansia muciniphila exacerbates gut inflammation in Salmonella Typhimurium-infected gnotobiotic mice. PLoS One. 2013;8(9):e74963.

Giongo A, Gano K A, Crabb D B, Mukherjee N, Novelo L L, Casella G, Drew J C, Ilonen J, Knip M, Hyöty H, Veijola R, Simell T, Simell O, Neu J, Wasserfall C H, Schatz D, Atkinson M A, Triplett E W. Toward defining the autoimmune microbiome for type 1 diabetes. ISME J. 2011; 5(1):82-91.

Han E, Sinsheimer JS, Novembre J. Characterizing bias in population genetic inferences from low-coverage sequencing data. Mol Biol Evol. 2014; 31(3):723-35.

Handelsman J. Metagenetics: spending our inheritance on the future. Microb Biotechnol. 2009;2(2):138–9.

Handelsman J. Metagenetics: spending our inheritance on the future. Microb Biotechnol. 2009;2(2):138-9.

Hattori M, Buse JB, Jackson RA, Glimcher L, Dorf ME, Minami M, Makino S, Moriwaki K, Kuzuya H, Imura H, *et al.* The NOD mouse: recessive diabetogenic gene in the major histocompatibility complex. Science. 1986;231(4739):733-5.

Henao-Mejia J, Elinav E, Jin C, Hao L, Mehal WZ, Strowig T, Thaiss CA, Kau AL, Eisenbarth SC, Jurczak MJ, Camporez JP, Shulman GI, Gordon JI, Hoffman HM, Flavell RA. Inflammasome-mediated dysbiosis regulates progression of NAFLD and obesity. Nature. 2012;482(7384):179-85.

Holmes E, Wang Y & Marchesi JR. Culture-independent analysis of the gut microbiota in colorectal cancer and polyposis. Environl Microbiol. 2008; 10, 789-798.

Hook SE, Wright AD, McBride BW. Methanogens: methane producers of the rumen and mitigation strategies. Archaea. 2010 Dec;2010:945785.

Howard BH, Hungate RE. Desulfovibrio of the sheep rumen. Appl Environ Microbiol. 1976;32(4):598-602.

Hsin-Jung Wu and Eric Wu. The role of gut microbiota in immune homeostasis and autoimmunity. Gut Microbes. 2012; 3(1): 4–14.

Hu C, Ding H, Li Y, Pearson JA, Zhang X, Flavell RA, Wong FS, Wen L. NLRP3 deficiency protects from type 1 diabetes through the regulation of chemotaxis into the pancreatic islets. Proc Natl Acad Sci U S A. 2015; 112(36):11318-23.

Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. Genome Res. 2007;17(3):377-86.

Johnson, D.E. and Ward, G.M. Estimates of animal methane emissions. Environ Monit Assess. 1996; 42, 133–141.

Josefsen MH, Andersen SC, Christensen J, Hoorfar J. Microbial food safety: Potential of DNA extraction methods for use in diagnostic metagenomics. J Microbiol Methods. 2015;114:30-4.

Kau AL, Ahern PP, Griffin NW, Goodman AL, Gordon JI. Human nutrition, the gut microbiome and the immune system. Nature. 2011;474(7351):327-36.

Kim KA, Gu W, Lee IA, Joh EH, Kim DH. High fat diet-induced gut microbiota exacerbates inflammation and obesity in mice via the TLR4 signaling pathway. PLoS One. 2012;7(10):e47713.

Koeth RA, Wang Z, Levison BS, Buffa JA, Org E, Sheehy BT, Britt EB, Fu X, Wu Y, Li L, Smith JD, DiDonato JA, Chen J, Li H, Wu GD, Lewis JD, Warrier M, Brown JM, Krauss RM, Tang WH, Bushman FD, Lusis AJ, Hazen SL. Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. Nature Medicine. 2013;19(5):576-85.

Kuczynski J, Costello EK, Nemergut DR, Zaneveld J, Lauber CL, Knights D, et al. Direct sequencing of the human microbiome readily reveals community differences. Genome Biol. 2010;11:210.

Kuitunen M, Kukkonen K, Juntunen-Backman K, Korpela R, Poussa T, Tuure T, Haahtela T, Savilahti E. Probiotics prevent IgE-associated allergy until age 5 years in cesarean-delivered children but not in the total cohort. Journal of Allergy and Clinical Immunology. 2009;123(2):335-41.

L Monteiro, D Bonnemaison, A Vekris, K G Petry, J Bonnet, R Vidal, J Cabrita, and F Mégraud. Complex polysaccharides as PCR inhibitors in feces: Helicobacter pylori model. J Clin Microbiol. 1997; 35(4): 995–998.

Lamendella R, Strutt S, Borglin S, Chakraborty R, Tas N, Mason OU, *et al*.. Assessment of the Deepwater Horizon oil spill impact on Gulf coast microbial communities. Front. Microbiol. 2014;5:130 10.3389/fmicb.2014.00130.

Lederberg, J; McCray, AT 'Ome sweet' omics - A genealogical treasury of words. SCIENTIST; 2001; 15(7): 8-8.

Lehuen A, Diana J, Zaccone P, Cooke A. Immune cell crosstalk in type 1 diabetes. Nat Rev Immunol. 2010;10(7):501-13.

Ley RE, Turnbaugh PJ, Klein S, Gordon JI. Microbial ecology: Human gut microbes associated with obesity. Nature. 2006c; 444:1022–1023.

Longman R S, Yang Y, Diehl G,. Kim S V., Littman D R. Microbiota-Host Interactions in Mucosal Homeostasis and Systemic Autoimmunity. Cold Spring Harb Symp Quant Biol. 2013; 78: 193–201.

Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.

Lozupone C, Lladser ME, Knights D, Stombaugh J, and Knight R. UniFrac: an effective distance metric for microbial community comparison. ISME J. 2011; 5(2): 169–172.

Maha Al-Asmakh, Farhana Anuar, Fahad Zadjali, Joseph Rafter, and Sven Pettersson. Gut microbial communities modulating brain development and function. Gut Microbes. 2012; 3(4): 366–373.

Mardis ER. Next-generation DNA sequencing methods. Annu. Rev. Genomics Hum. Genet. 2008b; 9:387–402.

Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. 2008;18(9):1509-17.

Marques TM, Wall R, Ross RP, Fitzgerald GF, Ryan CA, Stanton C. Programming infant gut microbiota: influence of dietary and environmental factors. Current Opinion in Biotechnology. 2010; 21(2):149-56.

Martinon F, Tschopp J. NLRs join TLRs as innate sensors of pathogens. Trends Immunol. 2005;26(8):447-54.

Maslowski KM, Vieira AT, Ng A, Kranich J, Sierro F, Yu D, Schilter HC, Rolph MS, Mackay F, Artis D, Xavier RJ, Teixeira MM, Mackay CR. Regulation of inflammatory responses by gut microbiota and chemoattractant receptor GPR43. Nature. 2009; 461(7268):1282-6.

Maukonen J, Simões C, Saarela M. The currently used commercial DNA-extraction methods give different results of clostridial and actinobacterial populations derived from human fecal samples. FEMS Microbiol Ecol. 2012;79(3):697-708.

Mazmanian SK, Liu CH, Tzianabos AO, Kasper DL. An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. Cell. 2005;122(1):107-18.

McDonald D, Price MN, Goodrich J, Nawrocki EP, *et al*. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. The ISME Journal. 2012; 6, 610–618.

McLoughlin RM, Mills KH. Influence of gastrointestinal commensal bacteria on the immune responses that mediate allergy and asthma. Journal of Allergy and Clinical Immunology. 2011;127(5):1097-107.

McOrist AL, Jackson M & Bird AR. A comparison of five methods for extraction of bacterial DNA from human faecal samples. J Microbiol Methods. 2002; 50: 131–139.

Metzker L. Sequencing technologies — the next generation. Nature Reviews Genetics. 2010; 11, 31-46.

Milani C, Hevia A, Foroni E, Duranti S, Turroni F, et al. Assessing the fecal microbiota: an optimized ion torrent 16S rRNA gene-based analysis protocol. PLoS One. 2013; 8: e68739.

Moreau MC, Ducluzeau R, Guy-Grand D, Muller MC. Increase in the population of duodenal immunoglobulin A plasmocytes in axenic mice associated with different living or dead bacterial strains of intestinal origin. Infect. Immun. 1978; 21, 532–539.

Musso G, Gambino R, Cassader M. Interactions between gut microbiota and host metabolism predisposing to obesity and diabetes. Annu Rev Med. 2011; 62:361–380.

Nava GM, Stappenbeck TS. Diversity of the autochthonous colonic microbiota. Gut Microbes. 2011; 2(2):99-104.

Nechvatal JM, Ram JL, Basson MD, Namprachan P, Niec SR, Badsha KZ, Matherly LH, Majumdar APN, Kato I. Fecal collection, ambient preservation, and DNA extraction for PCR amplification of bacterial and human markers from human feces. J Microbiol Meth. 2008;72:124–132.

Nell S, Suerbaum S, Josenhans C. The impact of the microbiota on the pathogenesis of IBD: lessons from mouse infection models. Nat. Rev. Microbiol. 2010; 8 564–577.

Nelson K.E., Weinstock G.M., Highlander S.K., Worley K.C., Creasy H.H., Wortman J.R. A catalog of reference genomes from the human microbiome. Science. 2010;328:994–999. *et al*.

Neu J, Reverte CM, Mackey AD, Liboni K, Tuhacek-Tenace LM, Hatch M, Li N, Caicedo RA, Schatz DA, Atkinson M: Changes in intestinal morphology and

permeability in the biobreeding rat before the onset of type 1 diabetes. J Pediatr Gastroenterol Nutr. 2005; 40:589–595.

Nishimoto, H., Kikutani, H., Tamamura, K. & Kishimoto, T. Nature. 1987; 328, 432−434.

Ochoa-Reparaz J, Mielcarz D, Begum-Haque S, Kasper L. Gut, bugs, and brain: role of commensal bacteria in the control of central nervous system disease. Ann Neurol. 2011; 69: 240–247.

Oh JZ, Ravindran R, Chassaing B, Carvalho FA, Maddur MS, Bower M, Hakimpour P, Gill KP, Nakaya HI, Yarovinsky F, Sartor RB, Gewirtz AT, Pulendran B. TLR5-mediated sensing of gut microbiota is necessary for antibody responses to seasonal influenza vaccination. Immunity. 2014; 41(3):478-92.

O'Hara A and Shanahan F. The gut flora as a forgotten organ. EMBO Reports. 2006; 7(7): 688–693.

Olszak T, An D, Zeissig S, Vera MP, Richter J, Franke A, Glickman JN, Siebert R, Baron RM, Kasper DL, Blumberg RS. Microbial exposure during early life has persistent effects on natural killer T cell function. Science. 2012;336(6080):489-93.

Ottman N, Smidt H, de Vos WM, Belzer C. The function of our microbiota: who is out there and what do they do?. Front Cell Infect Microbiol. 2012;2:104.

Peng X, Yu KQ, Deng GH, Jiang YX, Wang Y, Zhang GX, Zhou HW. Comparison of direct boiling method with commercial kits for extracting fecal microbiome DNA by Illumina sequencing of 16S rRNA tags. J Microbiol Methods. 2013;95(3):455-62.

Peter H. Janssen and Marek Kirs. Structure of the Archaeal Community of the Rumen. Appl Environ Microbiol. 2008; 74(12): 3619–3625.

Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. 2007;35(21):7188-96.

Qin J, Li Y, Cai Z, Li S, Zhu J. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature. 2012;490(7418):55-60.

Qin J., Li R., Raes J., Arumugam M., Burgdorf K.S., Manichanh C. A human gut microbial gene catalogue established by metagenomic sequencing. Nature. 2009;464:59–65. *et al*.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41(Database issue):D590-6.

Redondo MJ, Fain PR, Eisenbarth GS. Genetics of type 1A diabetes. Recent Prog Horm Res. 2001;56: 69–89.

Riesenfeld CS, Schloss PD, Handelsman J. Metagenomics: genomic analysis of microbial communities. Annu Rev Genet. 2004;38:525–52.

Ripple WJ, Smith P, Haberl H, *et al*. Ruminants, climate change and climate policy. Nat. Clim. Change.2014;4, 2–5.

Robinson MD, and Smyth GK. Moderated statistical tests for assessing differences in tag abundance. Bioinformatics. 2007; 23 (21): 2881-2887.

Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139-40.

Roesch LF, Lorca GL, Casella G, Giongo A, Naranjo A, Pionzio AM, Li N, Mai V, Wasserfall CH, Schatz D, Atkinson MA, Neu J, Triplett EW. Culture-independent identification of gut bacteria correlated with the onset of diabetes in a rat model. ISME J. 2009; 3(5):536-48.

Romijn JA, Corssmit EP, Havekes LM, Pijl H. Gut-brain axis. Current opinion in clinical nutrition and metabolic care. 2008;11:518–521.

Ronaghi M. Pyrosequencing sheds light on DNA sequencing. Genome Res. 2001; 11(1):3-11.

Round JL, Lee SM, Li J, Tran G, Jabri B, Chatila TA, Mazmanian SK. The Toll-like receptor 2 pathway establishes colonization by a commensal of the human microbiota. Science. 2011; 332(6032):974-7.

S Graham, P Courtois, W J Malaisse, J Rozing, F W Scott, A Mc I Mowat. Enteropathy precedes type 1 diabetes in the BB rat. Gut. 2004; 53:1437-1444.

Salonen A, Nikkilä J, Jalanka-Tuovinen J, Immonen O, Rajilić-Stojanović M, Kekkonen RA, Palva A, de Vos WM. Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: effective recovery of bacterial and archaeal DNA using mechanical cell lysis. J Microbiol Methods. 2010;81(2):127-34.

Samuel BS, Shaito A, Motoike T, Rey FE, Backhed F, Manchester JK, Hammer RE, Williams SC, Crowley J, Yanagisawa M, Gordon JI. Effects of the gut microbiota on host adiposity are modulated by the short-chain fatty-acid binding G protein-coupled receptor, Gpr41. Proc Natl Acad Sci U S A. 2008; 105(43):16767-72.

Santamaria M, Fosso B, Consiglio A, De Caro G, Grillo G, Licciulli F, Liuni S, Marzano M, Alonso-Alemany D, Valiente G, Pesole G. Reference databases for taxonomic assignment in metagenomics. Brief Bioinform. 2012; 13(6):682-95.

Sapone A, de Magistris L, Pietzak M, et al. Zonulin up-regulation is associated with increased gut permeability in subjects with type 1 diabetes and their relatives. Diabetes. 2006; 55:1443–1449.

Sattar N, Forrest E, and Preiss D. Non-alcoholic fatty liver disease. BMJ. 2014; 349: g4596.

Scanlan PD, Shanahan F, Clune Y, Collins JK, O'Sullivan GC, O'Riordan M,

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 2009;75(23):7537-41.

Schwartz RF, Neu J, Schatz D, Atkinson MA, Wasserfall C. Comment on: Brugman S et al. (2006) Antibiotic treatment partially protects against type 1 diabetes in the Bio-Breeding diabetes-prone rat. Is the gut flora involved in the development of type 1 diabetes? Diabetologia 49:2105-2108. Diabetologia. 2007; 50(1):220-1.

Seki E and Schnab B. Role of innate immunity and the microbiota in liver fibrosis: crosstalk between the liver and gut. J Physiol. 2012; 590(Pt 3): 447–458.

Sekirov I, Russell SL, Antunes LC, Finlay BB. Gut microbiota in health and disease. Physiol Rev. 2010; 90(3):859-904.

Seksik P, Rigottier-Gois L, Gramet G, Sutren M, Pochart P, Marteau P, Jian R, and Doré J. Alterations of the dominant faecal bacterial groups in patients with Crohn's disease of the colon. Gut. 2003; 52(2): 237–242.

Shin A, Lee E, Jeon D, Park YG, Bang JK, Park YS, Shin SY, Kim Y. Peptoid-Substituted Hybrid Antimicrobial Peptide Derived from Papiliocin and Magainin 2 with Enhanced Bacterial Selectivity and Anti-inflammatory Activity. Biochemistry. 2015;54(25):3921-31.

Slattery, R. M. *et al*. Nature. 1990; 345, 724−726.

Sonnenburg JL1, Xu J, Leip DD, Chen CH, Westover BP, Weatherford J, Buhler JD, Gordon JI. Glycan foraging in vivo by an intestine-adapted bacterial symbiont. Science. 2005;307(5717):1955-9.

Stanley D, Geier MS, Chen H, Hughes RJ, Moore RJ.Comparison of fecal and cecal microbiotas reveals qualitative similarities but quantitative differences. BMC Microbiol. 2015 Feb 27;15:51.

Stienstra R, Saudale F, Duval C, Keshtkar S, Groener JE, van Rooijen N, Staels B, Kersten S, Müller M. Kupffer cells promote hepatic steatosis via interleukin-1beta-dependent suppression of peroxisome proliferator-activated receptor alpha activity. Hepatology. 2010;51(2):511-22.

Suzuki T, Yamada T, Fujimura T, Kawamura E, Shimizu M, Yamashita R. Diabetogenic effects of lymphocyte transfusion on the NOD or NOD nude mouse. In: Rygaard J, Brunner N, Groem N, Spang-Thomsen M, editors. Immune-Deficient Animals in Biomedical Research, Copenhagen 1985. Karger; Basel: 1987. pp. 112–116.

Szabo G, Csak T. Inflammasomes in liver diseases. J Hepatol. 2012;57(3):642-54.

Tap, J. *et al*. Towards the human intestinal microbiota phylogenetic core. Environ. Microbiol. 2010; 11, 2574-2584.

Thomas T, Gilbert J, Meyer F. Metagenomics - a guide from sampling to data analysis. Microb Inform Exp. 2012;2(1):3.

Tilg H, Moschen AR. Food, immunity, and the microbiome. Gastroenterology. 2015;148(6):1107-19.

Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, Plagnol V, Bailey R, Nejentsev S*, et al*. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. Nat Genet. 2007;39(7):857-64.

Tringe SG, von Mering C, Kobayashi A, et al. Comparative metagenomics of microbial communities. Science. 2005;308(5721):554–7.

Tringe SSG, Hugenholtz P. A renaissance for the pioneering 16S rRNA gene. Curr Opin Microbiol. 2008;11(5):442–6.

Turnbaugh PJ, Bäckhed F, Fulton L, Gordon JI. Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. Cell Host Microbe. 2008; 3(4):213-23.

Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, Egholm M, Henrissat B, Heath AC, Knight R, Gordon JI. A core gut microbiome in obese and lean twins. Nature. 2009;457(7228):480-4.

Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature.2006; 444(7122):1027–1031.

Uzzau S, Leori GS, Petruzzi V, Watson PR, Schianchi G, Bacciu D, Mazzarello V, Wallis TS, Rubino S. Salmonella enterica serovar-host specificity does not correlate with the magnitude of intestinal invasion in sheep. Infect Immun. 2001;69(5):3092-9.

Vaarala O, Atkinson M A, and Josef Neu. The “Perfect Storm” for Type 1 Diabetes. The Complex Interplay Between Intestinal Microbiota, Gut Permeability, and Mucosal Immunity. Diabetes. 2008; 57(10): 2555–2562.

Vanhoutte, T, Huys G, De Brandt E and Swings J. Temporal stability analysis of the microbiota in human feces by denaturing gradient gel electrophoresis using universal and group-specific 16S rRNA gene primers. FEMS Microbiology Ecology. 2004.48(3), 437-446.

Vasileiadis S, Puglisi E, Arena M, Cappa F, Cocconcelli PS, Trevisan M. Soil bacterial diversity screening using single 16S rRNA gene V regions coupled

with multi-million read generating sequencing technologies. PLoS One. 2012;7(8):e42671.

Verdam FJ, Fuentes S, de Jonge C, Zoetendal EG, Erbil R, Greve JW, Buurman WA, de Vos WM, Rensen SS. Human intestinal microbiota composition is associated with local and systemic inflammation in obesity. Obesity (Silver Spring). 2013;21(12):E607-15.

Vernon G, Baranova A, Younossi ZM. Systematic review: the epidemiology and natural history of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis in adults. Aliment Pharmacol Ther. 2011;34(3):274-85.

Vijay-Kumar M, Aitken JD, Carvalho FA, Cullender TC, Mwangi S, Srinivasan S, Sitaraman SV, Knight R, Ley RE, Gewirtz AT. Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. Science. 2010; 328(5975):228-31.

Vuppalanchi R, Chalasani N. Nonalcoholic fatty liver disease and nonalcoholic steatohepatitis: Selected practical issues in their evaluation and management. Hepatology. 2009;49(1):306-17.

Wang JK., Ye JA, and Liu JX. Effects of tea saponins on rumen microbiota, rumen fermentation, methane production and growth performance--a review. Trop. Anim. Health Prod.2012; 44, 697–706.

Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, Dugar B, Feldstein AE, Britt EB, Fu X, Chung YM, Wu Y, Schauer P, Smith JD, Allayee H, Tang WH, DiDonato JA, Lusis AJ, Hazen SL. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. Nature. 2011;472(7341):57-63.

Watanabe A, Adnan Sohail M, Assis Gomes D. Inflammasome-mediated regulation of hepatic stellate cells. Am J Physiol Gastrointest Liver Physiol. 2009; 296(6): G1248–G1257.

Watts, T, Berti I, Sapone A, *et al*. Role of the intestinal tight junction modulator zonulin in the pathogenesis of type I diabetes in BB diabetic-prone rats. Proc. Natl. Acad. Sci. 2005; 102: 2916–2921.

Wen L, Ley RE, Volchkov PY, Stranges PB, Avanesyan L, Stonebraker AC, Hu C, Wong FS, Szot GL, Bluestone JA, Gordon JI, Chervonsky AV. Innate

immunity and intestinal microbiota in the development of Type 1 diabetes. Nature. 2008;455(7216):1109-13.

Wen L, Ley RE, Volchkov PY, Stranges PB, Avanesyan L, Stonebraker AC, Hu C, Wong FS, Szot GL, Bluestone JA, Gordon JI, Chervonsky AV. Innate immunity and intestinal microbiota in the development of Type 1 diabetes. Nature. 2008; 455(7216):1109-13.

Wilson IG. Inhibition and facilitation of nucleic acid amplification. Applied and Environmental Microbiology. 1997;63: 3741–3751.

Windsor PA, Eppleston J, Dhand NK, Whittington RJ. Effectiveness of Gudair™ vaccine for the control of ovine Johne's disease in flocks vaccinating for at least 5 years. Aust Vet J. 2014;92(7):263-8.

Wolever TMS, Jenkins DJA, Jenkins AL & Josse RG. The glycemic index: methodology and clinical implications. Am. J. Clin. Nutr. 1991; 54, 846 – 854.

Wommack KE, Bhavsar J, Ravel J. Metagenomics: read length matters. Appl Environ Microbiol. 2008 Mar;74(5):1453-63.

Wu GD, Lewis JD, Hoffmann C, Chen Y-Y, Knight R, *et al*. Sampling and pyrosequencing methods for characterizing bacterial communities in the human gut using 16S sequence tags. BMC Microbiol. 2010; 10: 206.

Xia LC, Cram JA, Chen T, Fuhrman JA, Sun F. Accurate genome relative abundance estimation based on shotgun metagenomic reads. PLoS One. 2011;6(12):e27992–e27992.

Yatsunenko T1, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP, Heath AC, Warner B, Reeder J, Kuczynski J, Caporaso JG, Lozupone CA, Lauber C, Clemente JC, Knights D, Knight R, Gordon JI. Human gut microbiome viewed across age and geography. Nature. 2012; 486(7402):222-7.

Yue ZB, Li WW, Yu HQ. Application of rumen microorganisms for anaerobic bioconversion of lignocellulosic biomass. Bioresour Technol. 2013;128:738-44.

Zaki MH, Boyd KL, Vogel P, Kastan MB, Lamkanfi M, Kanneganti TD. The NLRP3 inflammasome protects against loss of epithelial integrity and mortality during experimental colitis. Immunity. 2010;32(3):379-91.

Ziegler AG and Nepom GT. Prediction and Pathogenesis in Type 1 Diabetes. Immunity. 2010; 32(4): 468–478.

Acknowledgements

This PhD thesis is the result of collaboration between the Department of Biomedical Sciences (University of Sassari) and Porto Conte Ricerche (Technology Park of Sardinia in Tramariglio, Alghero). At the Park was essential the interaction with researchers from Porto Conte Ricerche, active in the Laboratory of Molecular Genetics and Proteomics, which in recent years have developed important projects on the intestinal microbiota.

First and foremost I want to thank my advisor Prof. Sergio Uzzau for supporting me during these past three years.

I am also very grateful to Dr. Cristina Fraumene (Molecular Genetics Laboratory, Porto Conte Ricerche) for her scientific advice and knowledge and many insightful discussions, suggestions and encouragement during the rough road of this formative experience.

Also thank you to Dr. Antonio Palomba and Dr. Alessandro Tanca (Proteomics Laboratory, Porto Conte Ricerche) who has helped with the project and supported me in many ways.

I would like to thank Dr. Massimo Deligios (Department of Biomedical Sciences, University of Sassari) for the support during the past three years.

I thank all the present members of the Porto Conte Ricerche and the staff of International PhD School in Biomolecular and Biotechnological Sciences (University of Sassari).

Without they precious support it would not be possible to conduct this research.

Last but not the least, I would like to thank my family for all their love, encouragement and support in all my pursuits.