University of Sassari

PhD Course in

Life Sciences and Biotechnologies

Curriculum: Microbiology and immunology

XXX cycle

PhD Course Director: Prof. Leonardo A. Sechi

Metaproteogenomic analyses of the gut microbiota in human and animal models: identification of changes induced by special diets in health and disease

Tutor: Prof. Sergio Uzzau

PhD candidate: Marcello Abbondio

PhD Course Director:

Prof. Leonardo A. Sechi

Università degli Studi di Sassari

Corso di Dottorato di ricerca in

Life Sciences and Biotechnologies

La presente tesi è stata prodotta durante la frequenza del corso di dottorato in Life Sciences and Biotechnologies dell'Università degli Studi di Sassari, a.a. 2016/2017 - XXX ciclo, con il sostegno di una borsa di studio cofinanziata 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 "Finanziamento di corsi di dottorato finalizzati alla formazione di capitale umano altamente specializzato, in particolare per i settori dell'ICT, delle nanotecnologie e delle biotecnologie, dell'energia e dello sviluppo sostenibile, dell'agroalimentare e dei materiali tradizionali".

La tesi è stata prodotta, altresì, grazie al contributo della Fondazione di Sardegna.

Acknowledgements

The present work has been possible thanks to the collaboration between the Department of Biomedical Sciences (University of Sassari, Italy) and Porto Conte Ricerche (Technology Park of Sardinia in Tramariglio, Alghero, Italy), where all the experiments were performed, including sample preparation, DNA sequencing, mass spectrometry identification, and metagenomic and metaproteomic analyses.

In addition, the collaboration with several other groups has occurred: i) the Department of Biomedical Sciences (University of Cagliari, Italy) for the rat model of caloric restriction; ii) the Institute of Food Sciences of the National Research Council (ISA-CNR, Avellino, Italy), for the mouse model of celiac disease; iii) the Institute for Genetic and Biomedical Research of National Research Council (IRGB-CNR, Cagliari), for the human healthy cohort study; iv) the Centre of Diabetology (San Giovanni University Hospital, Cagliari), for the dietary intervention study on T2D patients versus healthy controls; v) the Department of Agricultural Sciences (University of Sassari), for the ovine study.

Table of Contents

List of Figures and Tables

List of Abbreviations

ABC: ammonium bicarbonate

ACN: acetonitrile

AL: *ad libitum*

AL+F: *ad libitum* + functional bread

AL+S: *ad libitum* + standard bread

ANOVA: analysis of variance

BMI: body mass index

bps: base pairs

BW: body weight

CD: celiac disease

CFUs: colony-forming units

COG: Clusters of Orthologous Group

CR: caloric restriction

CR+F: caloric restriction + functional bread

CR+S: caloric restriction + standard bread

CR->AL: 1-week reversion from caloric restriction to *ad libitum*

CREA: Council for Agricultural Research and Analysis of Agricultural Economics

CV: coefficient of variation

DC: differential centrifugation

DL: direct lysis

DTT: dithiothreitol

ENA: European Nucleotide Archive

FASP: filter-aided sample preparation

F/B ratio: Firmicutes/Bacteroidetes ratio

FDR: false discovery rate

GH: glycoside hydrolase

GI: gastrointestinal

H_2: hydrogen

HCD: Higher Energy Collisional Dissociation

HLA: human leukocyte antigen

HMP: Human Microbiome Project

IAM: iodoacetamide

ibb: inverted beta binomial

IBD: inflammatory bowel disease

ISA-CNR: Institute of Food Sciences of the National Research Council

KEGG: Kyoto Encyclopedia of Genes and Genomes

LC: liquid chromatography

LCA: lowest common ancestor

LPS: lipopolysaccharide

MetaHIT: metagenomics of the human intestinal tract

MG: metagenomics

MP: metaproteomics

MRS: De Man-Rogosa-Sharpe

MS/MS: tandem mass

NAFLD: non-alcoholic fatty liver disease

NCBI-nr: National Center for Biotechnology Information non-redundant

NGS: next generation sequencing

NIH: National Institute of Health

ORF: open reading frame

OTUs: operational taxonomic units

PCA: Principal Components Analysis

PCoA: principal coordinate analysis

PCR: polymerase chain reaction

PSM: peptide-spectrum match

QIIME: Quantitative Insights Into Microbial Ecology

RLE: Relative Log Expression

rRNA: ribosomal RNA

SCFAs: short-chain fatty acids

SDS: sodium dodecyl sulfate

spp.: species

T2D: type 2 diabetes

TMAO: trimethylamine-N-oxide

V3: 16S rDNA variable region 3

V4: 16S rDNA variable region 4

WGS: whole genome shotgun

WHO: World Health Organization

Background

The term “microbiome", nowadays often substituted with "microbiota", was introduced for the first time in 2001 by Joshua Lederberg to define “the ecological community of commensal, symbiotic, and pathogenic microorganisms that literally share our body space and have been all but ignored as determinants of health and disease" (Lederberg and McCray 2001). Humans and animals, indeed, carry a huge load of microbes that populate most of their surfaces and outnumber the host cells by an estimated ratio of 10:1 (Savage 1977).

The human (and animal) microbiota include bacteria, fungi, archaea, protists, and viruses, distributed throughout the different body sites (i.e. oral cavity, vagina, upper respiratory tract, skin, and gastrointestinal tract).

Some of these microorganisms can cause illnesses, and are thus to be considered as pathogens, but they are only a minimal percentage. Rather, the majority are harmless commensal, often essential for the host health.

For this reason, and with the advent of the high-throughput next generation sequencing (NGS) techniques, the studies inspecting the microbiota are increased in an exponentially way. The project that, beyond the shadow of a doubt, marked the beginning of the era of microbiome studies is the Human Microbiome Project (HMP). This project was launched in 2007 by the National Institute of Health (NIH) of the United States of America with the ambitious purpose of describing the microbial diversity associated with health and disease (Peterson et al 2009). To this aim, 300 healthy adults were recruited and both 16S rDNA gene and microbiome genomes (also known as metagenome) were sequenced from 5 major body sites (18 subsites): nasal passages, oral cavities, skin, gastrointestinal (GI) tract, and urogenital/vaginal tract. In 2010, 178 genomes and 550,000 genes were sequenced and published (Nelson et al 2010), with the goal, today, to sequence 3,000 reference genomes.

HMP and several other studies highlighted the different composition of microbiota depending on the body localization (HMP Consortium 2012, Costello et al 2009, Grice et al 2008, Qin et al 2010, Ravel et al 2011, Willner et al 2011). As illustrated in Figure

1, the human microbiome is composed basically by four phyla, Actinobacteria, Bacteroidetes, Firmicutes, and Proteobacteria, but their relative abundance, and those of their lower taxa, differ tremendously according to the body site; for example the gut is dominated by Bacteroidetes and Firmicutes, while vagina by *Lactobacillus spp.*, and oral cavity by *Streptococcus spp.* (Morgan et al 2013).

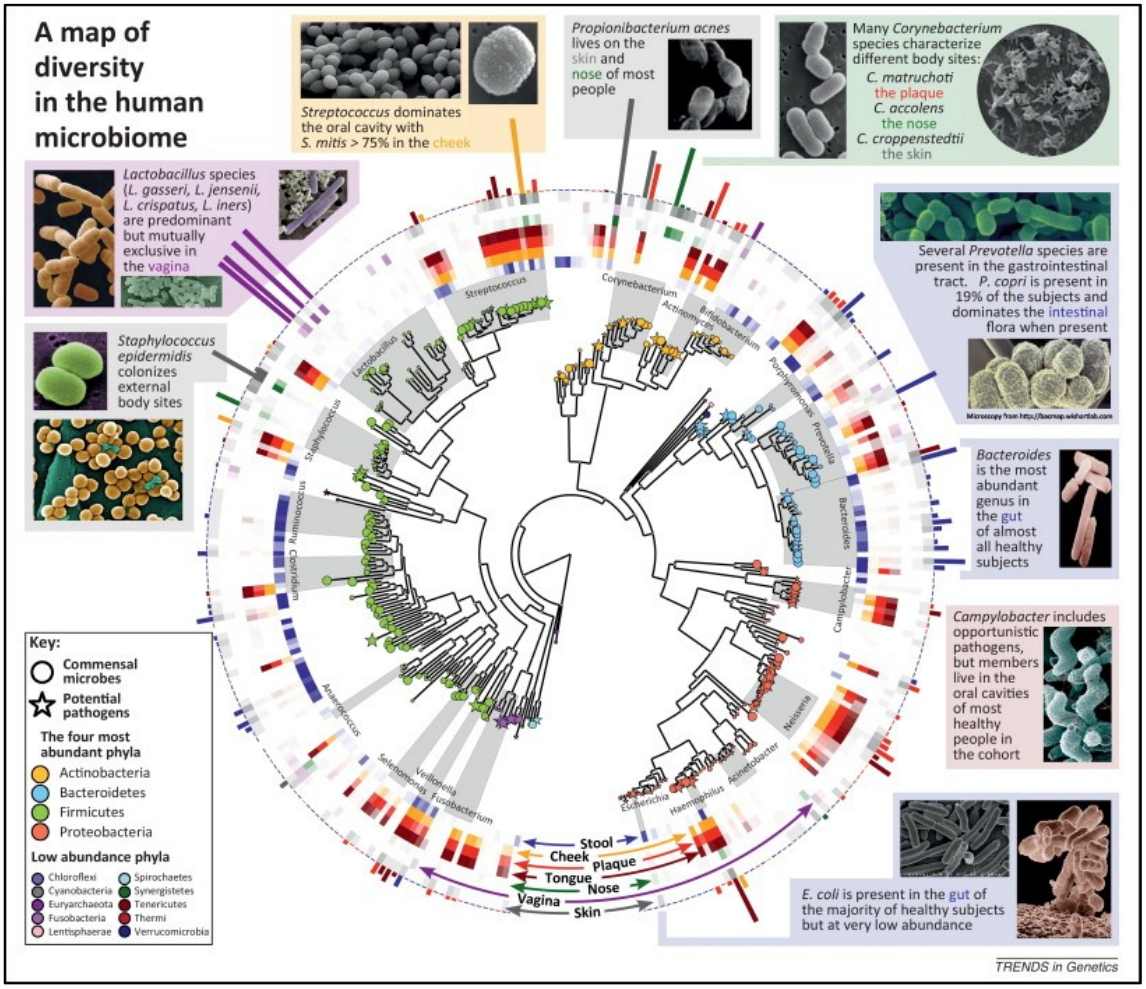

Figure 1: A map of microbial diversity in the human microbiota (Morgan et al 2013). A specific bacterial taxonomy distribution characterizes each body site. Districts with similar chemical and physical features are more related than others.

The diverse composition derived in the first place from chemical and physical features that characterize each district, such as pH, nourishment availability, aerobic or anaerobic conditions. Furthermore, after the colonization, key microorganisms can modify the habitat to which they adapt, triggering microbial co-occurrence and co-exclusion. The final result is that each body site is a highly specialized niche, where

members of the microbial community interact with each other and with the host environment (Grice and Segre 2012).

The gut microbiota

The gut microbiota is certainly one of the most studied microbial community among those that reside in the human body, as demonstrated by two important projects currently underway: the Metagenomics of the Human intestinal tract (MetaHIT), financed by European Union, and the American Gut Project, the world's largest crowd-funded citizen science project in existence. The GI tract was one of the first human ecosystems to be examined by whole genome shotgun (WGS) metagenomic analysis (Gill et al 2006). Because of their easy to be obtained and of the large amount of biomass which contain, fecal samples are regularly used as a proxy of the GI microbiota.

The human intestine is home for approximately up to 100 trillion (10^{14}) microbial cells, while the total number of microbial genes could be more than 100-fold superior than the total number of human genes (Backhed et al 2005, Ley et al 2006). Despite the gut core microbiota is composed mainly by Firmicutes and Bacteroidetes phyla and less by Actinobacteria and Proteobacteria, each individual has an exclusive intestinal flora signature, composed of up to 5000 microbial taxa (HMP Consortium 2012, Dethlefsen et al 2008, Tap et al 2009).

The gut microbiota is now considered just like a real organ, that contributes in an essential way to the development of the immune system (Ahern et al 2014), to the host metabolism and the energy production (Koropatkin et al 2012), and to the protection from enteropathogen invasion (Fukuda et al 2011).

Colonization of the GI tract begins at birth and evolves and shapes over a lifetime. The newborn's gut is characterized by a low diversity flora, composed mainly of mother inherited and environmental bacteria acquired at birth (Koenig et al 2011, Scholtens et al 2012).

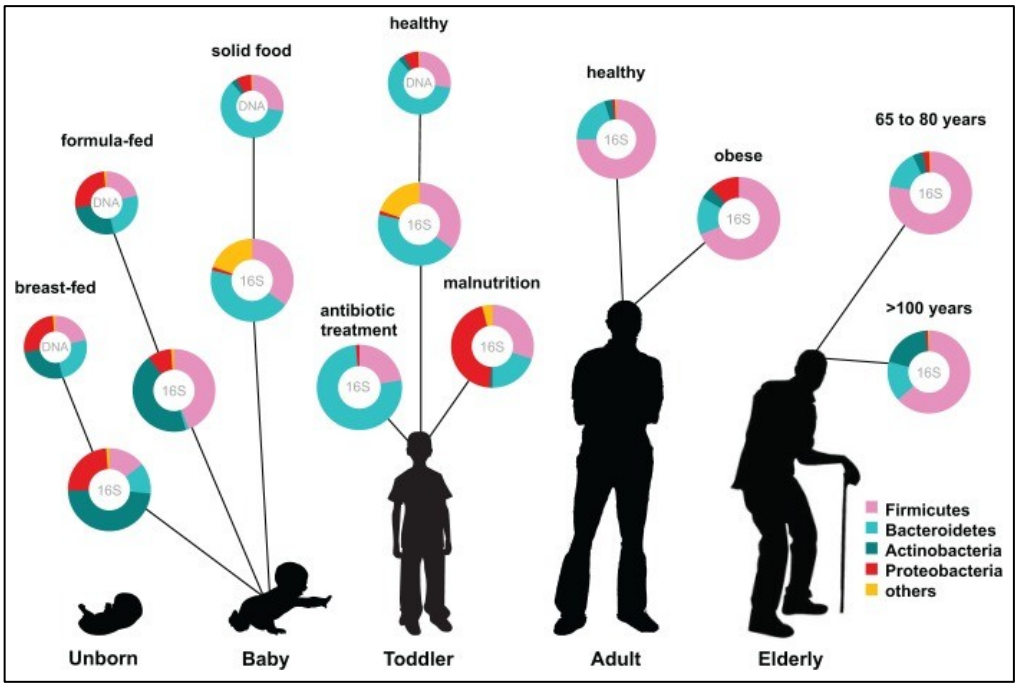

Figure 2: Human microbiota colonization and shaping through age progression and perturbations (Ottman et al 2012).

Gestational age, mode of delivery (natural or caesarean birth), diet (breast or formula feeding) and exposure to antibiotics are factors that influence the early colonization (Fouhy et al 2012, Marques et al 2010). Later on, gut microbiota complexity increases, with new microbes populating through feeding and other contacts, until becoming more stable, diverse and rich in adulthood, with the dominance of Firmicutes and Bacteroidetes (Figure 2) (Ottman et al 2012, Rajilic-Stojanovic et al 2009, Scholtens et al 2012).

Distinct microbes communities are present along the small intestine, caecum and large intestine (colon), given the diverse physiological variations along its length that include pH and nutrient gradients, mucus layer and host immune system (Donaldson et al 2016). First colonization, age progression, host genetics and countless environmental factors, such as diet, chemical exposition, lifestyle, can contribute to define the adult structure of the gut microbiota. Once composed, GI ecosystem results being partially stable in adulthood, as firstly noted by Caporaso and colleagues (Caporaso et al 2011a). Conservation of several taxa was confirmed by a number of long-term studies: from 75% of microbes in terms of relative abundance that persists after one year, up to 60% of strains detected in a 5 years monitoring (Faith et al 2013, Martinez et al

2013). Short- and long-terms stability is verified also after dietary change and antibiotic treatments, suggesting the resilience property of microbiota (David et al 2014a, Dethlefsen and Relman 2011, Jernberg et al 2007, Lozupone et al 2012, Wu et al 2011). However, changes in the intestinal milieu can take place resulting in a global remodeling of microbial hierarchy, and sometimes in an imbalance between the gut microbiota and the host, denominated as dysbiosis. These imbalances are known to be associated or predispose individuals to many diseases, starting from inflammatory bowel disease (IBD) and its two main manifestations, Crohn's disease and ulcerative colitis (Erickson et al 2012, Manichanh et al 2012), other gut intrinsic disorders such as colonic cancer and celiac disease (Arthur et al 2012, Girbovan et al 2017), metabolic syndromes, obesity (Turnbaugh et al 2009, Ussar et al 2015), diabetes (Karlsson et al 2013, Qin et al 2012), and non-alcoholic fatty liver disease (NAFLD) (Jiang et al 2015). In addition, the disruption of the intestinal community equilibrium can pave the way to pathogen invasion and/or pathobionts overgrowth, such as in *Clostridium difficile* infections (Rupnik et al 2009).

For these reasons, a better knowledge and characterization of the human microbiota is mandatory; firstly a comprehension of the interaction and alteration provoked by factors such as genotype, diseases, diet, age, and the exposition to drugs, and nutritional factors is necessary, then, in the near future, these data could allow scientists to selectively manipulate the gut microbiota in cases of dysbiosis-associated upsets.

Methods to characterize the gut microbiota

The investigations regarding the gut microbiota are almost always focused on its taxonomic characterization and, to a lesser extent, the metagenome analysis. Originally, taxonomic assignment was carried out through cultural techniques, making use of selective bacterial growth media and performing the identification on the base of morphological and metabolic features. The main drawback of this approach is that more than 80% of microbial species harbored by our gut could not be easily cultured, or are considered as uncultivable (Gevers et al 2012, Grice and Segre 2012).

Nowadays, the sequencing of 16S ribosomal RNA (rRNA) gene fragment is the most popular technique used for phylogenetic and taxonomic purposes. The 16S rRNA gene, which is present in all bacteria, is approximately 1,500 base pairs (bps) in length and contains both species-specific hypervariable regions and highly conserved sequences, which makes it well suited for polymerase chain reaction (PCR) amplification and sequencing (Hugenholtz and Pace 1996). However, 16S metagenomics presents many limitations, such as the huge variability in the number of 16S rRNA gene copies in bacterial genomes, PCR amplification bias, and difference in taxonomic assignment depending on the choice of the 16S rDNA variable region (Jumpstart Consortium Human Microbiome Project Data Generation Working Group 2012, Lamendella et al 2012, Schloss 2010). In addition, amplicon sequencing allows scientists to obtain information about the microbial taxonomy, but not about its gene assortment (Addis et al 2016, Ellegaard and Engel 2016), a limit that is partially overcome using specific tools to predict the functional profiling (Asshauer et al 2015, Langille et al 2013).

Shotgun metagenomics is another DNA-based approach, that consists in sequencing the whole microbial DNA, including bacteria, viruses and fungi, instead of amplifying a specific procariotic or eucariotic target locus. After extraction, DNA is subjected to fragmentation and sequenced, producing a great number of microbial sequences; sequences are then taxonomically and functionally annotated by aligning them to reference genomes, including bacteria, viruses and fungi, or with a "de novo" strategy (Lai et al 2012, Sharpton 2014). For these reasons, shotgun metagenomics provides information about both the taxonomic composition as well as functional genes. Nevertheless, metagenomics data presents several challenges. The first one is represented by the great complexity and dimension of the data generated, that affect the computational analysis, in terms of power and pipeline (Sharpton 2014). In addition, a lack or a scarce number of sequences deriving from poorly represented genomes could make difficult the assembly and alignment steps (Mende et al 2012). Furthermore, microbial DNA could be overwhelmed by unwanted host DNA, requiring molecular and bioinformatics strategy to filter it out (Garcia-Garcera et al 2013,

Schmieder and Edwards 2011). Finally, only genetic potential is investigated, while no information is provided about actively expressed functions.

In order to investigate the microbiota active functions, it is necessary to unveil the microbiota members expression profile (metatranscriptomics) or the protein products (metaproteomics).

Metatranscriptomics refers to the analysis of both rRNA and mRNA expressed by a microbial community. mRNAs provide an instantaneous image of the microbes responses to an external stimulus, that may alter the environmental conditions (Tveit et al 2014). After extraction, RNA is reverse transcripted to cDNA, and analyzed by high-throughput sequencing technologies (RNA-seq) (Giannoukos et al 2012). The main challenges in a metatranscriptomics analysis are the low stability of RNA and sequence assembly and annotation issues (Abram 2015).

Metaproteomics provides a direct measure of the active functions of the microbiota by studying the entire microbial proteomes. Host proteins are also detected and, although they may sometimes be considered as contaminants, they give researchers information about the host milieu and its response to the microbiota (Valles-Colomer et al 2016). In addition, post-translational modifications can also be analyzed (Ahrens et al 2010). Typically, a metaproteomic pipeline consists on protein extraction, their hydrolysis with proteolytic enzymes in order to generate a complex peptide mixture, which is analyzed by liquid chromatography (LC) separation systems with high-resolution mass spectrometers (MS). Metaproteomics requires appropriate databases for peptide identification, and their construction and annotation is of course the most complex bioinformatics task concerning this approach (Muth et al 2013). Since the majority of the gut microbial genomes are poorly or absolutely not characterized, the availability of metagenomic sequences from the community being studied is fundamental for a satisfactory protein identification (Tanca et al 2014).

Finally, the systematic analysis of the metabolites produced by microbial communities, named metametabolomics, is a currently established approach for functional microbiota analyses and it is growing together with the other methodologies in an integrated fashion.

Therefore, the choice of the best approach for studying the gut flora depends essentially on the question to be asked: 16S rDNA data provides exclusively information regarding the taxonomy, answering the question "who is there?"; metagenomics gives not only a understanding of the taxonomic composition of a microbial community, but also of its gene potential, answering the question "what could they do?"; finally, metaproteomics and metatranscriptomics, as well as metametabolomics, take a picture of what microbes are really doing in a specific moments (Figure 3) (Addis et al 2016, Sharpton 2014).

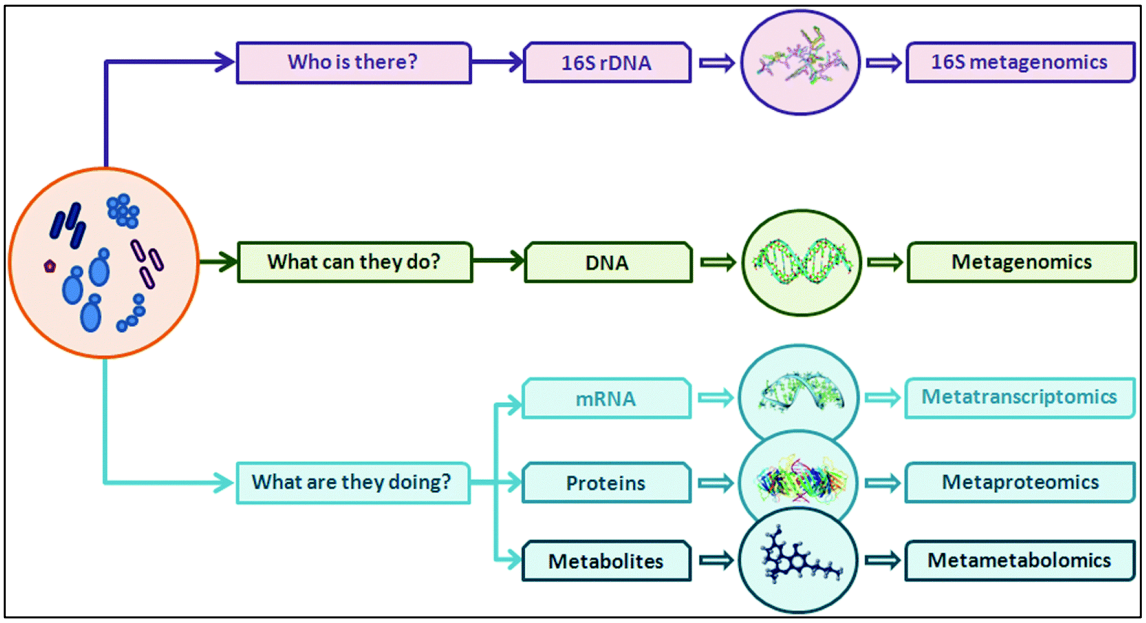

Figure 3: Outline of the approaches available for the study of the gut microbiota (Addis et al 2016).

Gut microbiota and diet

Diet has without a doubt an important role in modifying human gut microbiota as strong as medication. Since birth, maternal nutrition shapes the gut microbiota composition of breastfed newborns (Cabrera-Rubio et al 2012, Ma et al 2014). The amount, type and balance of the main dietary macronutrients (carbohydrates, proteins and fats) have a great impact on the intestinal flora. Many food components, such as plant-derived complex carbohydrates (i.e., fibers and resistant starch), that cannot be digested by the human digestive system, become primary substrates for our GI microbes, being capable of breaking down and fermenting them.

Dietary fiber provides a considerable amount of substrate to the intestinal microflora; it includes a broad array of nondigestible carbohydrates derived by plant cell wall components, such as cellulose and xylans, and storage polysaccharides, such as resistant starch and inulin (Salonen and de Vos 2014). Our gut microbiota is able to produce thousand of enzymes in order to ferment fiber into short-chain fatty acids (SCFAs), hydrogen (H_2), methane and carbon dioxide, that are the main metabolic end-products (El Kaoutari et al 2013, Slavin 2013).

SCFAs, mainly acetate, butyrate and propionate, are involved in various essential processes: energy source for colonocytes, anticancer and anti-inflammatory effects (butyrate) (Leonel and Alvarez-Leite 2012, Salonen and de Vos 2014); gluconeogenesis and protection from diet-induced obesity (propionate) (De Vadder et al 2014); lipogenesis in the liver (acetate) (Vipperla and O'Keefe 2012); histone deacetylase inhibition (acetate and butyrate) (Sonnenburg and Backhed 2016). SCFAs also reduce the luminal pH, contrasting the growth and the activity of pathogens (Kashtanova et al 2016); finally, they are involved in tissue repairing, epithelial integrity maintenance, secretion of IgA by B cells, and promotion of regulatory T cell response in the gut (Thorburn et al 2014).

In addition to provide fermentable substrates in the colon, dietary fiber promotes also the transit rate increase (Salonen and de Vos 2014). A fiber-rich diet is often associated to a healthy condition, both for the host and the intestinal community; several studies have shown that dietary fiber has a preventive and a therapeutic role in many disorders, including cardiovascular diseases, type II diabetes and obesity (Bodinham et al 2012, Hauner et al 2012, Murphy et al 2012).

Dietary proteins are an important part of a balanced diet, as they serve not only as the major source of nitrogen for the microbes growth, but also to enhance carbohydrates fermentation and assimilation, contributing to the health of the bowel. However, some end-products of protein, such as ammonia, hydrogen sulphide, amines, phenols, thiols and indoles, have been shown to be cytotoxic, genotoxic and carcinogenic (Conlon and Bird 2014, Toden et al 2005), while trimethylamine-N-oxide (TMAO), a metabolite from L-carnitine (abundant in red meat) and phosphatidylcholine (cheese, seafood eggs and

meat), have been associated with atherosclerosis promotion (Koeth et al 2013, Wang et al 2011).

An endogenous source of protein, alternative to those taken with the diet, are the mucins secreted from goblet cells, that assemble an extra GI layer important for the luminal protection (Johansson et al 2013). Mucins provide an always available protein resource and, with their high polysaccharide content, they may constitute half of the carbon flux in the intestinal tract (Salonen and de Vos 2014). Only few microorganisms are capable to utilize this nutrient source and most of all as alternative substrate (Png et al 2010); in case of scarcity of nutrients, dietary fiber in particular, some gut microbiota species can shift from polysaccharides to mucus glycan metabolism, causing the reduction of the mucus layer thickness and promoting access by mucosal pathogen (Desai et al 2016, Earle et al 2015).

Dietary fat also influences the composition and metabolic activity of the gut microbiota. High fat diet induces increased circulating levels of bacterial lipopolysaccharide (LPS) in humans (Moreira et al 2012), possibly as a consequence of increased intestinal permeability, and it has been associated with obesity, hepatic steatosis, non-alcoholic fat liver disease (NAFLD) (de Wit et al 2012, Devkota et al 2012, Kubeck et al 2016), and IBD (de Wit et al 2012, Devkota et al 2012, Kubeck et al 2016).

Fat intake influences the gut microbiota in an indirect manner too, through the hepatic production and release of bile acids. These molecules have antibacterial property and create a significant selection among the intestinal microbiota, but some microbial secondary bile acids products are potentially carcinogenic and are associated with the development of colorectal cancer and NAFLD (Mouzaki et al 2016, Ou et al 2012).

The exposure to a specific dietary pattern since youth affects tremendously our intestinal flora, selecting microbes on the basis of their capability to metabolize what the menu offers. For example, different studies have demonstrated that the gut microbiota of rural or hunter-gatherer Africans is equipped with a broad-spectrum of enzymes that are able to degrade dietary complex polysaccharides, while most of

these enzymes completely lacking in the Western populations that follow a high fat and protein diet (De Filippo et al 2010, Rampelli et al 2015, Yatsunenko et al 2012).

If long-term dietary habits have a strong impact on the intestinal ecosystem, human gut responds also to immediate changes in diet (David et al 2014b); however, the gut microbiome is generally resilient and short-term dietary intervention is unusually successful in treating obesity and malnutrition (Xu and Knight 2015).

Nowadays scientists are paying specific attention to dietary intervention, as demonstrated by the huge number of studies regarding probiotics, prebiotics and functional foods administration. Once the complex interactions between dietary components and gut microbes will be elucidated and deepened, diet could aim for becoming an easier and cheaper remedy against dysbiotic and pathological conditions than drugs and traditional therapy.

Aim of the research project

In keeping with the considerations outlined in the introduction, the main objective of this project is the study of the interactions of the gut microbiota with the diet.

Accordingly, the following aims are established:

Aim 1: to develop experimental animal models for the metaproteogenomic characterization of diet-induced changes of the gut microbiota.

Aim 2: to characterize the gut microbiota in human and livestocks through a metaproteogenomic approach.

Aim 3: to describe the gut microbiome taxonomic and functional features in human cohorts according to their dietary habits and to dietary interventions.

Chapter 1:

Animal models for the study of gut microbiota and diet interactions

1.1 Introduction

Animal models are commonly used by scientists to investigate human diseases in the event that experimentation on humans would be impracticable or unethical. Cancer, metabolic syndromes, autoimmune disorders, chronic diseases are only few examples of research branches that are investigated with the support of models, often in rodents.

Since the active role of the gut microbiota in human health and disease has been hypothesized, the gut microbiota of several animal models has become subject of study (Gkouskou et al 2014, Silverman et al 2017, Tomkovich et al 2017).

The advantages in using animals are innumerable: feces can be sampled easily and frequently; individuals can be sacrificed after the experiment to analyze different tissues or samples (i.e. cecal contents or liver); in disease models, sampling can take place before the onset and during the progression of the pathology; experimental design and parameters are more controlled, making the inter-individual variability smaller than in humans. In addition, there is a large use of animal without an intestinal flora - called germ-free - or with a controlled one - named gnotobiotic - that provide an invaluable experimental tool for investigating the interactions between host and microbiota, and are ideal for fecal transplantation or single strain administration experiments (Al-Asmakh and Zadjali 2015, Martin et al 2016).

In contrast, some factors need to be taken into account when dealing with animal models: firstly, the gut microbiota could not replicate the same development and

function occurring in the human one, neither the taxonomy could be comparable due to environmental and anatomical differences; several husbandry aspects, such as cohousing, degree of kinship, sex, feeding, handling, have an effect on microbiota and must be evaluated and if possible reduced; not least, especially in long-time experiments, animals with specific genotypes can get sick or die prematurely, resulting in a drop of the number of samples and consequently in a lower statistical power.

As for dietary treatments, mice and rats become good actors when ethical and compliance matters impede to test extreme diets (i.e. high-fat diet) or overfeeding in humans; moreover, studies with animals allow researchers to investigate both slight dietary intervention (supplement/removal of a single nutrient) and long-term diet habit (eating always the same food).

As an example, in diet-induced obesity scientists have shed light on the strong association of the gut microbiota to the disease when, after fecal transplantation from obese to germ-free mice or from lean to obese ones, the phenotype of the receiving was transformed in that of the donor (Ridaura et al 2013, Turnbaugh et al 2006).

Hence, animal models represent an important instrument to investigate the interaction among diet, gut flora and disease before moving to humans, if possible, and, when studies are well designed, to ensure a better control of the experimental variables and the inter-variability among individuals.

1.2 Caloric restriction promotes rapid expansion and long-lasting increase of Lactobacillus in the rat fecal microbiota

1.2.1 Aim of the study

In this study, it was investigated whether fecal microbiota variation is induced by short-term caloric restriction (CR) and whether such change perseveres in long-lasting CR, in a rat model of aging (Marongiu et al 2016). Animals were fed *ad libitum* (AL), as for the majority of murine experimental models. This condition might be assimilated to overfeeding, given the poor energy expenditure in the animal house settings, compared to wild rats in their natural environments.

The results reported in this chapter have been partially published on *Gut Microbes* (Fraumene et al 2017).

1.2.2 Experimental design

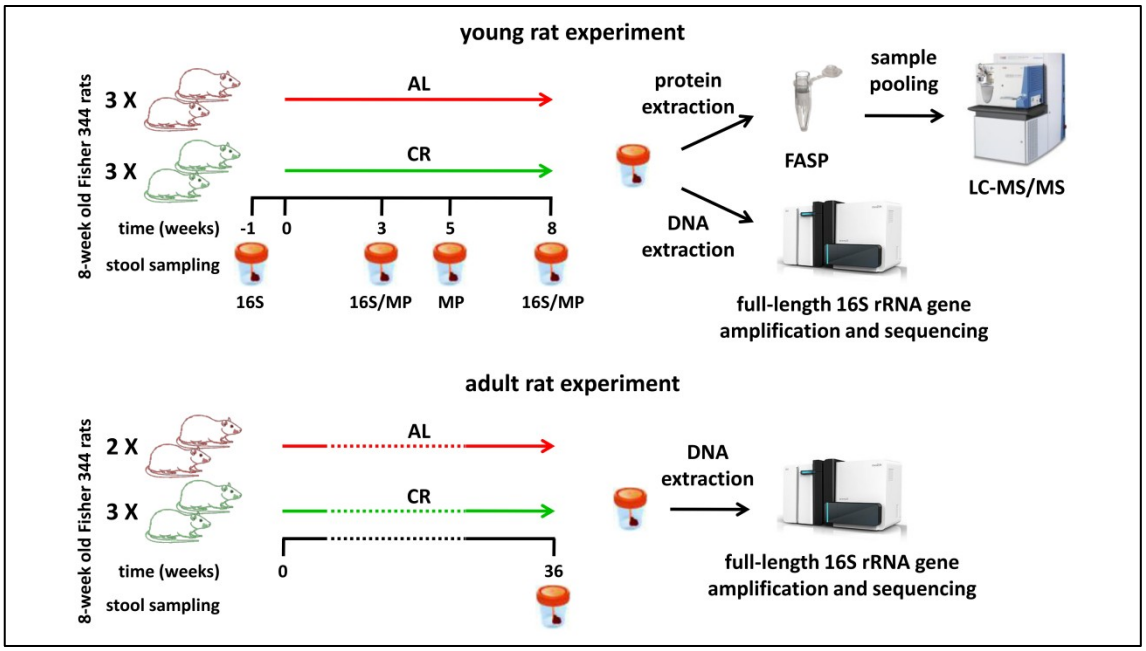

Figure 4: Schematic illustrating the experimental design of the study.

Fecal microbiota was evaluated in young growing and adult Fisher 344 rats, comparing animals fed AL and animals fed with 70% of the AL ratio (CR), in order to shed light on timing and nature of microbial composition and functional changes. Both 16S metagenomics and metaproteomics (MP) approaches were employed, as illustrated in Figure 4.

1.2.3 Material & Methods

Animal intervention and sampling

A total of 22 Fisher 344 male rats, from a colony available at the Department of Biomedical Sciences, University of Cagliari, were used. Rats were fed with Purina Rodent Lab Chow diet (3% of fat, Mucedola srl, Settimo Milanese, Italy) until the age of 8 weeks. Then animals were reared two per cage and split in two groups of 6: the first was maintained on the AL diet, the latter was provided 70% of the AL ratio (CR). Food to CR group was delivered every night at 1AM through a computer assisted automated food dispenser. Stool samples were collected one week before the beginning of the treatment, after 3, 5 and 8 weeks. In a parallel experiment, 10 rats were fed as described above since the age of 8 weeks and up to their mid-life. Feces were sampled after 36 weeks of CR treatment (6 rats) and from controls (4 rats) fed AL. Animal studies were reviewed and approved by the Institutional Animal Care and Use Committee of the University of Cagliari.

All collected fecal samples (N = 58) were immediately stored at -80°C until use. At the time of the analyses, feces were thawed at 4°C, and, in cases of both DNA and protein extraction, two portions were collected from each of them.

To evaluate the abundance of *Lactobacillus* spp. in fecal samples, bacterial cultures were conducted on De Man-Rogosa-Sharpe Agar (MRS, Oxoid, Basingstoke, UK) added with the fungicide cycloheximide (7 µg/mL, Sigma-Aldrich, Saint Louis, MO, USA) in anaerobic condition at 37°C. *Lactobacillus* colonies were identified according to their morphology and then by a genus-specific PCR amplification (Rinttila et al 2004).

Sequencing of the 16S rRNA gene

A total of 46 rat feces sampled one week before and 3 and 8 weeks after the beginning of the treatment, were subjected to DNA extraction with the QIAamp Fast Stool Kit protocol (QIAGEN, Hilden, Germany), following the manufacturer's instructions. After extraction, DNA was purified according to E.Z.N.A.® Soil DNA Kit (Omega Bio-Tek, Norcross, GA, USA) and quantified with a Qubit 2.0 Fluorometer (Life Technologies, Grand Island, NY, USA), using the Qubit® dsDNA High Sensitivity Assay Kit (Life Technologies). In the adult rats samples the full-length 16S rRNA gene was amplified using the universal primers 27F-1492R (AGAGTTTGATYMTGGCTCAG and TACGGYTACCTTGTTACGACTT, respectively) (Hugenholtz et al 2001). In the young rats samples the V4 region of 16S rRNA was amplified according to Caporaso and co-workers' protocol (Caporaso et al 2011b). PCR products were confirmed on 2% agarose gel (Sigma Aldrich). The gene amplification reaction was performed in duplicate for each experiment, pooled together, cleaned up with AMPure XP magnetic beads (Beckman Coulter, Brea, CA, USA) and quantified using the Qubit HS assay (Life Technologies). Sample libraries were made according to the Illumina Nextera XT sample preparation protocol (Illumina, San Diego, CA, USA) and then checked for quality on a 2100 BioAnalyzer with a DNA 1000 kit (Agilent Technologies, Santa Clara, CA, USA). The average insert size was around 500 bps. Libraries were quantified with the Qubit® dsDNA Broad-Range Assay Kits (Life Technologies), normalized and, after pooling, subjected to the cluster generation step with the cBOT cluster generation station, following the Illumina TruSeq Paired-End Cluster Kit protocol instructions. DNA sequencing was carried out using the Illumina HiScanSQ sequencer, using the paired-end method and 93 cycles of sequencing.

Shotgun metagenomics sequencing

A total of 12 fecal samples equally distributed among treatment groups and time points were subjected to whole metagenome sequencing in order to create an "in house" sequence database for the metaproteomic identification. DNA was extracted, purified and quantified as mentioned in the above section. Then DNA was randomly

tagmented and libraries were constructed according to the Nextera XT kit instruction. After quality control on a BioAnalyzer High Sensitivity DNA Chip (Agilent Technologies), quantification with the Qubit® dsDNA High Sensitivity, normalization and pooling, libraries were sequenced with the HiScanSQ sequencer, using also in this case the paired-end method and 93 cycles.

Bioinformatics and statistical analysis of sequencing data

After sequencing, all obtained reads were subjected to a demultiplexing step using Casava software (version 1.8.2) implemented in HiScan control software (Illumina). Then FASTX-Toolkit was used to trim reads for the first 20 bps, while sequences with Nextera adapter contamination were identified with the UniVec database (ftp://ftp.ncbi.nlm.nih.gov/pub/UniVec/) and removed.

The paired-end reads with a minimum overlap of eight base pairs were merged using the script join_paired_ends.py contained into the bioinformatics platform Quantitative Insights Into Microbial Ecology (QIIME) (version 1.9.1) (Caporaso et al 2010). Operational taxonomic units (OTUs) generation was performed with a QIIME pipeline following the USEARCH's OTU clustering recommendations (http://www.drive5.com/usearch/manual/otu_clustering.html); the closed-reference OTU picking was used to allow clustering of shotgun 16S sequences (Tanca et al 2017c). OTUs were obtained after clustering reads at 97% identity using UCLUST (Edgar 2010). Taxonomy assignment of resulting OTUs was performed using the Greengenes database (version 13_8) (DeSantis et al 2006). A number of OTUs were assigned to taxa given in square brackets, including *[Prevotella]*, *[Paraprevotellaceae] [Ruminococcus] gnavus*, and *[Clostridium] difficile*. These outputs from QIIME analyses refer to taxonomic assignments recommend by GreenGenes, mainly based on genome trees. As normalization step, rarefaction to equal sequencing depth was performed on all samples in each experiment. Richness was quantified as observed OTU counts, while alpha diversity was calculated based on Shannon diversity index; the statistical significance regarding differences in microbial community composition index between

sample categories was determined by non-parametric Monte Carlo permuted two-sided t-test (999 permutation).

Rarefied OTU tables were employed to perform Principal Components Analysis (PCA) and to generate PCA plots using the web platform MetaboAnalyst (version 3.0) (http:// http://www.metaboanalyst.ca/) (Xia and Wishart 2016); boxplots and scatter plots were created using GraphPad Prism (version 5.03), while heatmap was produced with the web application ClustVis (http://biit.cs.ut.ee/clustvis) (Metsalu and Vilo 2015) and edited using Inkscape (https://inkscape.org).

Sequence similarity search was performed via the BLAST [nucleotide] tool available in the European Nucleotide Archive (ENA) website, using default search parameters (http://www.ebi.ac.uk/Tools/sss/ncbiblast/nucleotide.html).

OTUs with differential abundance between the two sample groups were obtained by applying a negative binomial generalized linear model (Wald test within the DESeq2 module available in QIIME) (Love et al 2014) to the raw OTU counts. Differential abundance analysis at genus level was carried out between groups using Student's t test after log transformation and pareto scaling, while analysis within groups on paired data was performed with an established paired sample test for count data based on an inverted beta binomial model, available in ibb R package (http://www.oncoproteomics.nl/software/BetaBinomial.html) (Pham and Jimenez 2012). *P*-values were corrected for multiple inference using the Benjamini-Hochberg FDR procedure with an adjusted alpha cutoff value of 0.05.

Metagenome bioinformatics

Raw metagenomic reads were either filtered and clustered without assembly, or assembled into contigs. In the first case, read processing was performed with tools from the USEARCH suite (version 8.0.1623) (Edgar 2010). In particular, the following steps were carried out in succession: merging of paired reads (fastq_mergepairs command, setting parameters as follows: fastq_truncqual 3, fastq_minovlen 8, fastq_maxdiffs 0), sorting (sortbylength command), quality filtering (fastq_filter command, with fastq_truncqual 15 and fastq_minlen 100), and sequence clustering

(cluster_smallmem command, with 1 as identity threshold). In the second case, read assembly into contigs was performed using Velvet (version 1.2.10) (Zerbino and Birney 2008) with velveth command by setting 61 as k-mer length, and velvetg command by setting 200 as insert length and 300 as minimum contig length. FragGeneScan (version 1.30) (Rho et al 2010) was used for open reading frame (ORF) finding, with the training for Illumina sequencing reads with about 0.5 % error rate.

Protein sample preparation and mass spectrometry analysis

Proteins were extracted from a total of 36 fecal samples according to Tanca and colleagues' protocol (Tanca et al 2014). In detail, feces were resuspended in 100 µl of sodium dodecyl sulfate (SDS)-based reducing extraction buffer (2% SDS, 100 mM dithiothreitol, DTT, 20 mM Tris-HCl pH 8.8), incubated at 95°C for 20 min in agitation (500 rpm) in a Thermomixer Comfort (Eppendorf, Hamburg, Germany), and subsequently subjected to bead-beating combined with freeze-thawing as described below. To each sample a steel bead (5-mm diameter; Qiagen) was added; then, samples were incubated at −80°C for 10 min, subjected to bead-beating for 10 min (30 cycles/s in a TissueLyser LT mechanical homogenizer, Qiagen), incubated at −80°C for 10 min and then at 95°C for 10 min, and subjected to an additional bead-beating step of 10 min. Samples were lastly centrifuged at 20,000 × g for 10 min at 4°C, with the final supernatant being the protein extract. Extracted proteins were cleaned up, alkylated and subjected to trypsin digestion according to the filter-aided sample preparation (FASP) procedure (Wisniewski et al 2009), with minor modifications (Tanca et al 2013, Tanca et al 2015).

In detail, SDS protein extracts were diluted to 200 µL with a solution composed of 8M urea in 100 mM Tris-HCl, pH 8.8, filtrated by loading into Ultrafree® MC-GV centrifugal filters (Merck Millipore, Billerica, MA, USA) and then loaded into the Amicon® Ultra-0.5 centrifugal filter units with Ultracel-10 membrane (Merck Millipore). Samples were subjected to centrifugation at 14 000 × *g* for 15 min, then to dilution into the filter with 200 µL of the urea solution and again to centrifugation. Following centrifugation, the concentrates were mixed with 100 µL of iodoacetamide (IAM, 50 mM) in the urea

solution and incubated at 20°C for 20 min. After centrifugation, each concentrate was diluted with 100 µL of urea solution and concentrated again (this step was repeated twice). Next, concentrates were diluted with 100 µL of ammonium bicarbonate (ABC, 50 mM) and concentrated once more. This step was repeated another time. Subsequently, 40 µL of trypsin solution (150 ng in 50 mM ABC) were added to each filter, and the samples were incubated at 37°C overnight. Peptides were collected by centrifugation of the filter units, followed by a further wash step with 50 µL of a solution containing 70% acetonitrile (ACN) and 1% formic acid. Finally, the peptide mixtures were brought to dryness and resuspended in 0.2% formic acid to a final concentration of 1 mg/mL.

Peptide mixtures concentration was estimated by measuring absorbance at 280 nm with a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MS, USA), using dilutions of the MassPREP E. coli Digest Standard (Waters, Milford, MA, USA) to generate a calibration curve. After quantification, tryptic digests coming from rats reared in the same cage (2 rats per cage) were equally pooled.

Liquid chromatography (LC)-tandem mass (MS/MS) analyses were carried out using an LTQ-Orbitrap Velos mass spectrometer (Thermo Fisher Scientific) interfaced with an UltiMate 3000 RSLCnano LC system (Thermo Fisher Scientific). The single-run one-dimensional LC peptide separation was performed as detailed by Tanca and colleagues: (Tanca et al 2013, Tanca et al 2014): peptide mixtures (4 µg per run) were loaded, concentrated and desalted on a trapping pre-column (Acclaim PepMap C18, 75 µm × 2 cm nanoViper, 3 µm, 100 Å, Thermo Fisher Scientific), using 0.2% formic acid at a flow rate of 5 µl/min. Then, peptides were separated with a C18 column (Acclaim PepMap RSLC C18, 75 µm × 15 cm nanoViper, 2 µm, 100 Å, Thermo Fisher Scientific) at 35°C with a flow rate of 300 nL/min, using a 247-min gradient from 1% to 50% eluent B (0.2% formic acid in 95% ACN) in eluent A (0.2% formic acid in 5% ACN). The mass spectrometer was set up in a data dependent MS/MS mode under direct control of the Xcalibur software (version 1.0.2.65 SP2), where a full scan spectrum (from 300 to 1700 m/z) was followed by MS/MS spectra. The instrument was operated in positive mode with a spray voltage of 1.2 kV, a capillary temperature of 275°C, and was calibrated

before measurements. Full scans and MS/MS spectra were carried out in the Orbitrap with resolution of 30,000 and 7,500 at 400 m/z, respectively. The automatic gain control was regulated to 1,000,000 ions and the lock mass option was enabled on a protonated polydimethylcyclosiloxane background ion ($(Si(CH_3)_2O)_6$; m/z = 445.120025) as internal recalibration for accurate mass measurements (Olsen et al 2005). Peptide ions were chosen as the ten most intense peaks (Top 10) of the previous scan. 500 counts was set as the signal threshold for activating an MS/MS event. Higher Energy Collisional Dissociation (HCD) was performed at the far side of the C-trap and was selected as the method of fragmentation, with the following parameters: 40% value for normalized collision energy, isolation width of m/z 3.0, Q-value of 0.25, and activation time of 0.1 ms. Nitrogen was used as the collision gas.

Metaproteome bioinformatics

Microbial peptide identification was carried out using the Proteome Discoverer informatic platform (version 2.0; Thermo Fisher Scientific), with a workflow consisting of the following nodes (and respective parameters): Spectrum Selector for spectra preprocessing (precursor mass range: 350–5000 Da; S/N threshold: 1.5), SEQUEST-HT as search engine (enzyme: trypsin; maximum missed cleavage sites: 2; peptide length range 5–50 amino acids; maximum delta Cn: 0.05; precursor mass tolerance: 10 ppm; fragment mass tolerance: 0.02 Da; static modification: cysteine carbamidomethylation; dynamic modification: methionine oxidation), and percolator for peptide validation (false discovery rate, FDR, <1% based on peptide q-value). Only rank 1 peptides were kept after results filtering, and protein grouping was allowed according to the maximum parsimony principle.

A custom collection of metagenomic sequences obtained in house, as described in the previous section (7,422,716 sequences in total), was employed as sequence database, after joining ORF obtained both from non-assembled reads and from contigs. Host peptide identification was carried out with the Proteome Discoverer informatic platform (version 1.4; Thermo Fisher Scientific), using the protein sequences belonging to the order Rodentia and downloaded from UniProtKB/SwissProt (release 2017_05;

26,536 sequences in total) as database. Search engine and parameters were set as mentioned above for microbial peptide identification.

Taxonomic and functional annotation was carried out using multiple strategies. MEGAN (version 6.6.7) was used as first annotation option (Huson et al 2016). Protein sequences were preliminary subjected to a DIAMOND (version 0.8.22) search against the National Center for Biotechnology Information non-redundant (NCBI-nr) database (2016/09 update), using the blastp command with default parameters (Buchfink et al 2015); subsequently, DIAMOND outputs were loaded on MEGAN in order to perform both lowest common ancestor (LCA) classification and functional annotation (according to InterPro and eggNOG modules) using default parameters. Moreover, the Unipept web application (version 3.1; https://unipept.ugent.be) was used to carry out an LCA classification of the identified peptide sequences (Mesuere et al 2017). Finally, an additional functional annotation was accomplished by aligning with DIAMOND (blastp module, e-value threshold 10^{-5}) the identified protein sequences against a database containing all bacterial sequences from UniProtKB/Swiss-Prot (release 2016/09); UniProtKB/Swiss-Prot accession numbers were then utilized to retrieve protein name and Kyoto Encyclopedia of Genes and Genomes (KEGG) orthologous group information from the UniProt website via the 'retrieve' tool (Pundir et al 2016). Taxonomic information from different sources were combined, giving priority to MEGAN results; functional information from different sources were examined, merged and made uniform manually. Butyrate, propionate and acetate biosynthetic pathways were reconstructed based on the corresponding KEGG pathway maps (Kanehisa et al 2016), available at http://www.genome.jp/kegg/pathway.html.

Metaproteomics statistical analysis and graph generation

Spectral counts were uploaded to the web application MicrobiomeAnalyst (http://www.microbiomeanalyst.ca) to evaluate differential abundance through comparative statistical analysis (Dhariwal et al 2017). Features with prevalence in <10% of samples in a given comparison were filtered out. Prior to statistical testing, count data were subjected to transformation in accordance with the Relative Log

Expression (RLE) method (Anders and Huber 2010). Differential abundance analysis was then performed using the edgeR algorithm (Anders et al 2013), with an adjusted *P*-value (FDR) cutoff of 0.05.

Relative abundance data (obtained by dividing the count data by the total number of counts in a sample) were employed to create PCA plots using the web application ClustVis with default parameters. Heatmaps were generated starting from relative abundance data with the web application Morpheus (https://clue.io/morpheus). Relative abundance values were transformed by subtracting the median abundance of a given taxonomic/functional feature in the dataset, and subsequently dividing by the median absolute deviation, as one of the 'transform' options available in the 'color scheme' menu. Scatter plots were generated using GraphPad Prism starting from relative abundance data, and Student's t test was applied to calculate significant differences between total enzyme abundance means.

1.2.4 Results

Growth curves and serum lipids profile in rats fed AL or CR diet

Growth curves of AL and CR rats were well separated after only 3 weeks from the beginning of the CR diet (Figure 5).

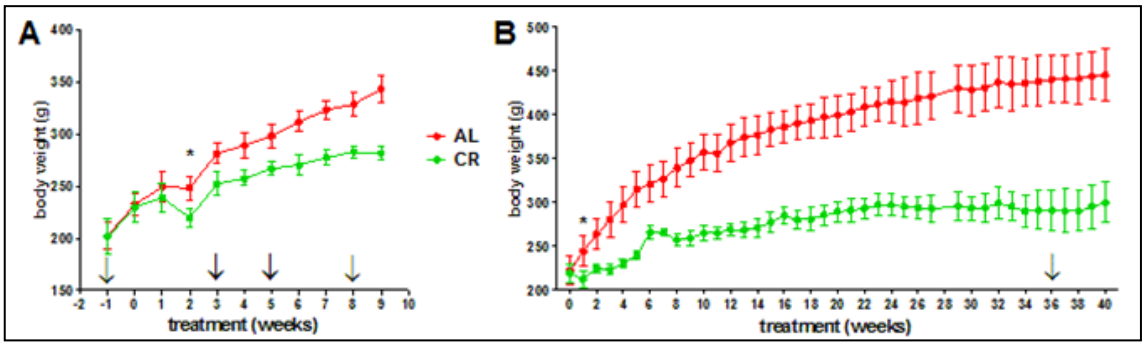

Figure 5: Growth curves of young growing (A) and adult (B) rats during *ad libitum* (AL, red) and caloric restriction (CR, green) treatment (Fraumene et al 2017). Arrows indicate the different weeks at which stools were collected, while the first point in which the two groups exhibit a statistically significant difference in weight is marked with an asterisk.

Animals fed AL showed initially a quick body weight increase and continued to gain weight until the 36th week, when their weight remained quite stable until the end of

the experiment. On the contrary, rats fed CR showed a slower growth compared to AL until 21 weeks of treatment, after which their weight remained stable. These data are in accordance with the literature for the Fisher 344 rat strain (Armbrecht et al 1988). Although the absolute food consumption was 30% less in CR animals, after 28 weeks of feeding CR rats were fed an equal amount of diet compared to AL group in relation to their actual size, as their food intake/body weight ratio became comparable (Figure 6). Significant changes in serum lipids profile after both short-term and long-term CR are well established (Choi et al 1988). Interestingly, significant decreases in total cholesterol and triglycerides were observed to be already evident after 8 weeks of CR regimen (Figure 7).

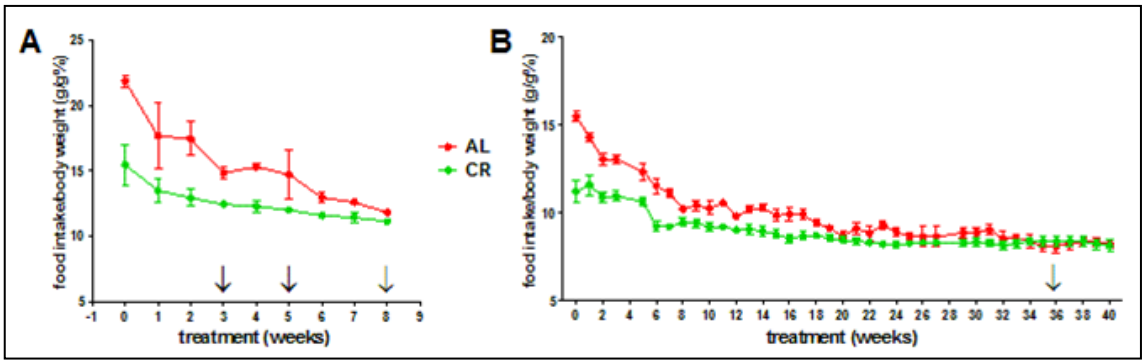

Figure 6: Food intake/body weight ratio curve of young growing (A) and adult (B) rats during *ad libitum* (AL, red) and caloric restriction (CR, green) treatment (Fraumene et al 2017). Arrows indicate the different weeks at which stools were collected.

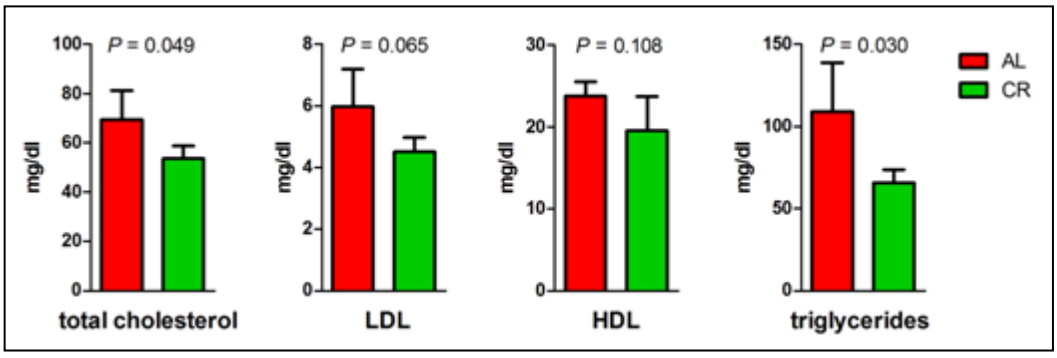

Figure 7: Lipid serum profile of young growing rats after 8 weeks of *ad libitum* (AL, red) and caloric restriction (CR, green) treatment (Fraumene et al 2017). Standard deviation and Student's t-test *P* value are reported.

Sequencing of 16S rDNA from the gut microbiota of young and adult Fisher 344 rats

To investigate the gut microbiota composition in the Fisher 344 rats subjected to either CR or AL chow diet, feces were collected from the 12 young rats one week before the beginning of dietary intervention (T0) and after 3 and 8 weeks. (Figure 4, young rat experiment). Feces were also collected from 10 adult rats after 36 weeks from the beginning of CR and analyzed with 16S metagenomics (Figure 4, adult rat experiment).

Sequencing of the 16S amplicons, from a total of 46 fecal samples, enabled to obtain 30,107,702 and 3,358,687 reads, respectively for the young and the adult rats experiments, after the merging of paired-end reads. Richness, calculated as the number of OTUs detected within each community, showed no significant difference comparing samples from the two groups of animals of the same age (Table 1).

As reported above for richness, Shannon's index analyses indicated similarity in microbial diversity between the CR and AL young growing rats, with the exception of feces sampled after 3 weeks of CR intervention, that showed a significant lower alpha-diversity compared to the AL control group (P = 0.015) (Table 1). Taking into account that the richness was not influenced by dietary intervention, this difference in alpha-diversity might be related to a reduced uniformity of the gut microbial community composition after the first three weeks of CR treatment. It is remarkable that this coincided with the time point when CR animals were subjected to a sudden change in average food intake per body weight (Figure 6). In the same way, alpha-diversity index did not differ between adult rats with or without CR treatment (Table 1).

	Richness (No. of detected OTUs)		Alpha diversity (Shannon index)	
Treatment (weeks)	**AL**	**CR**	**AL**	**CR**
-1	1649.8 ±6.2	1652.4 ±14.0	7.147 ±0.569	7.691 ±0.798
3	1664.1 ±3.4	1661.0 ±7.1	8.495 ±0.066	8.231 ±0.183
8	1667.1 ±3.1	1666.6 ±2.0	8.409 ±0.091	8.255 ±0.219
36	1168.8 ±8.4	1167.4 ±6.5	7.723 ±0.219	7.475 ±0.199

Table 1: OTU richness and alpha-diversity within AL (*ad libitum*) and CR (caloric restriction) groups at different time points. Means and standard deviation of the number of detected OTUs and Shannon index are reported.

CR treatment rapidly affects gut microbial community composition in young rats

Principal component analysis (PCA) was performed in order to evaluate the similarity between microbial communities in young rats with or without CR treatment.

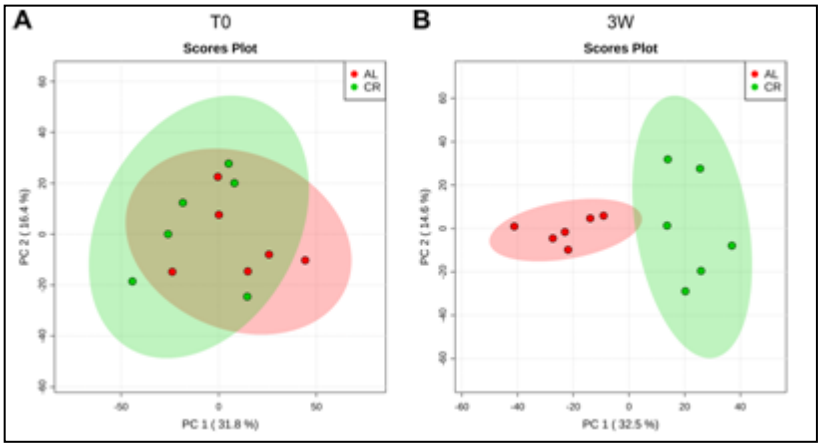

Figure 8: Beta-diversity at OTU level between *ad libitum* (AL, red) and caloric restriction (CR, green) groups before (T0, A) and after 3 weeks of treatment (3W, B) in young rats (Fraumene et al 2017). PCA was carried out on rarefied OTU table data with the web application MetaboAnalyst.

As illustrated in the PCA plots in the Figure 8, after only 3 weeks there was an evident clustering of the fecal microbiota of CR treated rats compared to AL controls.

Indeed, at this time point, a very large number of differentially abundant OTUs (n = 1,703) was observed between the two groups of rats, of which 1,175 were higher in AL feces and 528 were higher in CR feces. Analysis of the most represented genera showed significant differences regarding *[Ruminococcus] gnavus* (belonging to the family Lachnospiraceae), more abundant in AL rats, and *5-7N15* (family Bacteroidaceae), more abundant in CR rats (Figure 9).

After 8 weeks of CR treatment, a lower number of differentially abundant OTUs were recorded (n = 708), of which 364 were higher in AL samples and 344 were higher in CR treated samples. Any significant differences were observed when comparing the most abundant genera of AL and CR treated animals at 8 weeks (Figure 9), in line with the behavior of differential OTUs that belonged to the same taxon but with an opposite correlation to diet.

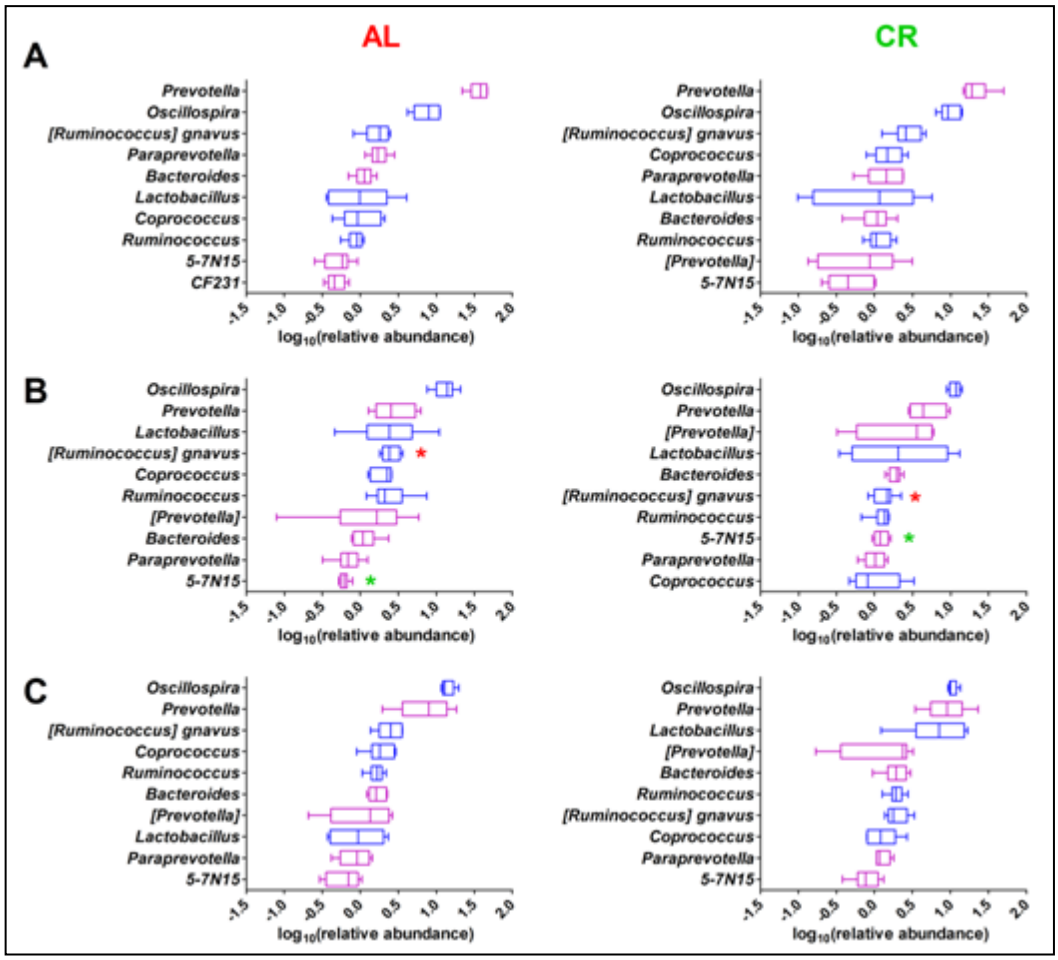

Figure 9: Top 10 genera within the young rats gut microbiota fed *ad libitum* (AL, left) and with caloric restriction (CR, right) (Fraumene et al 2017). Genera are ordered by decreasing median of the relative abundance among subjects at 1 week before treatment (*A*), 3 weeks (B), and 8 weeks of treatment (C). Boxes are colored based on the phylum membership (blue, Firmicutes; purple, Bacteroidetes). Asterisks indicate significant difference (FDR <0.05) between AL and CR rats at same time of treatment, with the asterisk corresponding to the group in which a given genus was more abundant.

Since these rats were in a fast growing phase, physiological changes depending on both diet (CR or AL) and rate of growth from 7 to 16 weeks of age (Figure 5A) could affect the gut microbiota composition at 8 weeks. Thus, to assess whether CR and AL diets induced specific changes during gut microbiota maturation in growing rats, the differential analysis on the relative abundance of some genera at different time points within the same group of animals was performed. In particular, rats fed AL diet for 8 weeks showed an increase of the relative abundance of genera *Oscillospira*, *Ruminococcus*, *Coprococcus*, and *Desulfovibrio* and a parallel decrease of genus *Prevotella*, variations already evident and significant after 3 weeks (Figure 10).

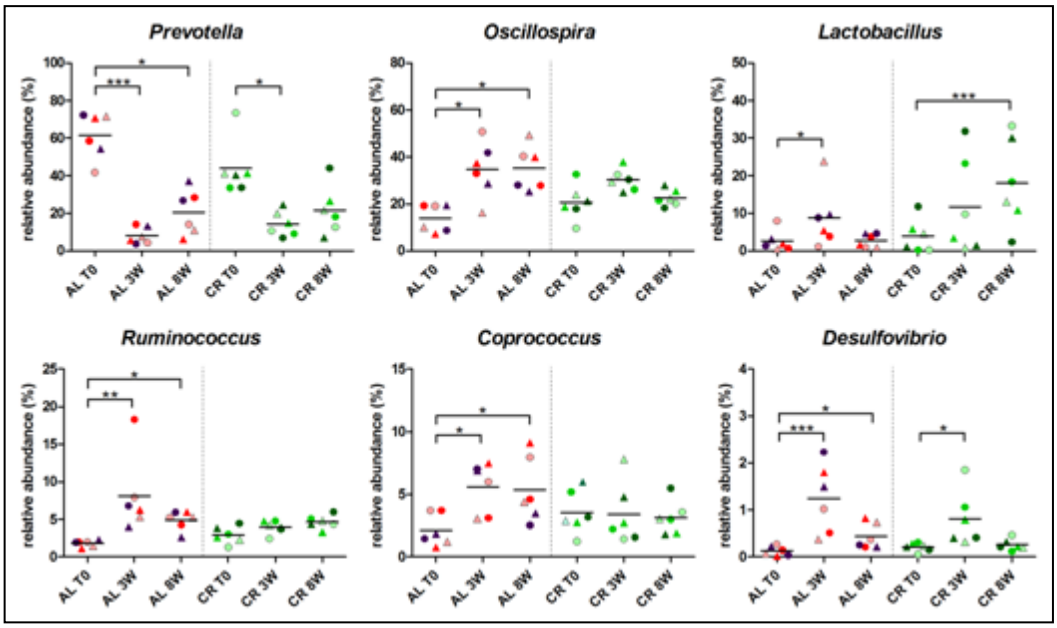

Figure 10. Relative abundance variation of genera in young rats fecal microbiota after 8 weeks of *ad libitum* (AL) or caloric restriction (CR) diets (Fraumene et al 2017). Only genera with a relative abundance >1% in at least one time point and significant variation at 8 weeks are represented. Each dot represents a different sample, with the same combination shape-color indicating the same rat. Means are also reported. Samples were evaluated 1 week before treatment (T0) and after 3 weeks (3W) and 8 weeks of treatment (8W). Asterisks indicate significant difference (**P* <0.05, ***P* <0.01, ****P* <0.001) within AL and CR groups between different time points.

Interestingly, *Prevotella* and *Desulfovibrio* (to a lesser extent) showed similar trends in fecal samples from CR fed rats, but not any significant variations after 8 weeks of CR. In the CR-treated group, in contrast, *Lactobacillus* genus showed a significant increase in young growing rats after 8 weeks of treatment; a significant increase was also observed in AL controls, but only after 3 weeks, while at week 8 its abundance lowered to the initial level (Figure 10). As a result, CR appears to attenuate the variations observed in AL rats and to promote, instead, *Lactobacillus* abundance in the young growing rats.

Composition of gut microbiota in adult rats

To evaluate the effect of long-term dietary intervention on the rat fecal microbiota, PCA was performed once more to describe beta-diversity in CR- and AL-fed animals. An even clearer clustering of the gut bacterial communities was observed according to the CR or AL diet (Figure 11, left).

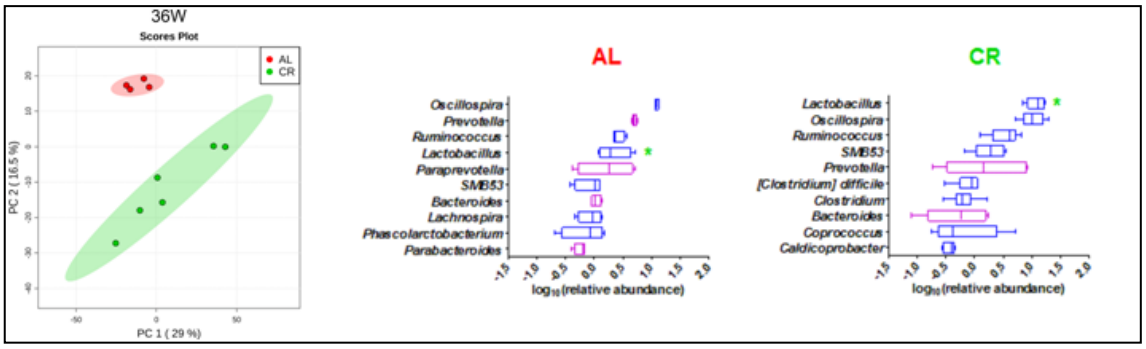

Figure 11: Beta-diversity and fecal microbiota composition of *ad libitum* (AL) and caloric restriction (CR) treated adult rats (Fraumene et al 2017). Left, PCA plot on rarefied OTU table data. Principal component analysis was carried out with the web application MetaboAnalyst Right, boxplots showing the top 10 microbial genera. Features are ordered by decreasing median of the relative abundance among subjects. Boxes are colored based on the phylum membership (blue, Firmicutes; purple, Bacteroidetes). Asterisks indicate significant difference (FDR <0.05) between AL and CR rats at same time of treatment, with the asterisk color corresponding to the group in which a given genus was more abundant.

However, the relative abundances of the most prevalent genera were similar comparing the two experimental groups (Figure 11, right). As reported previously for the young rat experiment, this result can be explained since OTUs belonging to the same genus might have opposite correlation with dietary intervention in both groups. Most interestingly, the only exception was *Lactobacillus*, that showed significantly higher relative abundance in CR fecal microbiota than in samples from "overfed" rats (Figure 11, right). Consequently, the high number (n = 32) of differential OTUs assigned to *Lactobacillus* that were associated with the 36 weeks long CR treatment also corresponded to an overall higher relative abundance of *Lactobacillus* genus in these animals. Bacterial cultures also confirmed the differential abundance of *Lactobacillus*. After culturing fecal samples in MRS agar plates, a significantly higher number of colony-forming units (CFUs) per gram of feces was detected in CR samples compared to AL controls (respectively $2.1 \pm 0.31 \times 10^9$ vs $5.8 \pm 0.38 \times 10^8$; *P* value <0.03).

Finally, attention was focused on the assortments of *Lactobacillus* OTUs identified both in the young (n = 446) and in the adult rats (n = 320). All the OTUs detected in the second group of animals were also present in the first group. This suggested that the very same *Lactobacillus* strains observed in juvenile animals persisted in aged animals. To delve into the trend of *Lactobacillus* relative abundance from young growing rats to adulthood, OTUs belonging to the genus *Lactobacillus* and present with >0.01% of

relative abundance in at least one group at one time point were selected and their relative abundance compared in all animals (Figure 12).

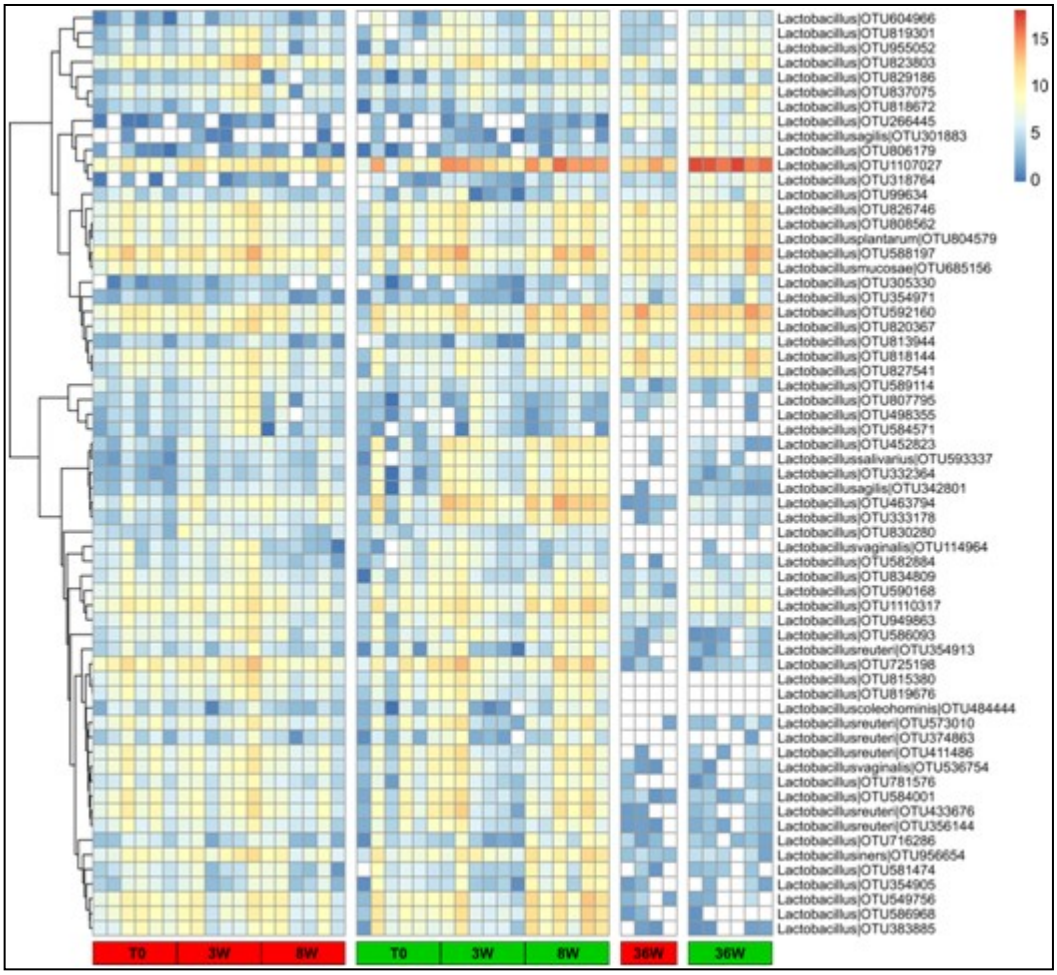

Figure 12: Heatmap showing the relative log-transformed abundance distribution of Lactobacillus spp. OTUs among all samples. OTUs with >0.01% of relative abundance in at least one group at one time point were selected. Columns are hierarchically ordered based on the experiment, treatment (*ad libitum*, red; caloric restriction, green), and time point.

A larger group of OTUs was scarcely represented in adult rats compared to the young ones regardless of the dietary intervention, while very few OTUs showed either low or high relative abundance at all time points, again in both groups. Conversely, a number of OTUs were more represented in CR animals at 8 and at 36 weeks of treatment than AL controls at the same time points. Among these, OTU 1107027 was the most abundant in the CR gut microbiota of samples collected at 8 weeks (mean = 0.17% and 2.39% in AL and CR treated rats, respectively) as well as in feces from adult rats (mean = 0.72% and 8.79% in AL and CR treated animals, respectively). A BLAST search against

the European Nucleotide Archive (ENA) database was performed to further refine the taxonomic assignment. OTU 1107027 exhibited sequence identity of 100% with an uncultured bacterial clone (HM124081) and 99.7-99.8% of identity with three *L. faecis* strains, a recently assigned species described as obligately homofermentative (Endo et al 2013).

Caloric restriction induces rapid and deep changes in the fecal microbiota metaproteome

The application of a shotgun metaproteomic approach was performed to verify if the significant compositional changes occurring in the fecal microbiota of young growing rats, after short-term administration of a CR diet, were associated to modification in its functional and metabolic profile. To this purpose, fecal samples were collected from rats after 3, 5 and 8 weeks of CR or AL feeding, and the metaproteomic profiles of their microbiota were characterized. A total of 142,942 mass spectra were obtained and could be matched with 878 different protein functions, belonging to over 250 different microbial genera.

First, focus was set on taxonomy, with the aim of investigating microbiota compositional changes based on the abundance of the proteins expressed by its microbial members. PCA was performed on genus abundance data, revealing a separation between CR and AL fed rats since 3 weeks of treatment, becoming even clearer at 8 weeks (P <0.001 PERMANOVA between groups considering all weeks), as illustrated in Figure 13, left.

Then, a differential abundance analysis was carried out to identify which genera were mainly responsible for the separation between the two groups. The heatmap in Figure 13 shows 24 genera with significantly differential distribution between AL and CR microbiota (relative abundance threshold 0.25%). Interestingly, 8 genera belonging to Bacteroidetes phylum, including *Prevotella* and *Bacteroides*, were found to be higher in CR fed rats compared to AL controls, while 11 Firmicutes genera, including *Clostridium*, *Eubacterium* and *Oscillibacter*, were found to be more abundant in AL compared to CR treated animals. The only genus belonging to Firmicutes to be enriched in CR fed rats

was *Lactobacillus*, with remarkable significance values (FDR = 9 x10^{-20}) and consistent with the 16S rDNA gene sequencing results reported above.

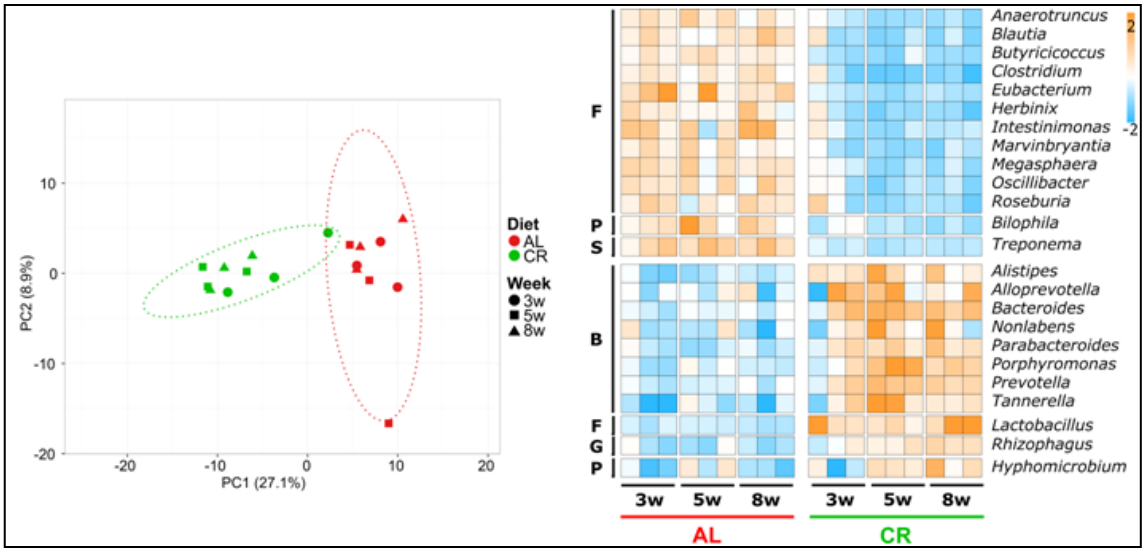

Figure 13: Changes in taxonomic composition at genus level based on metaproteomic results obtained upon caloric restriction treatment on young rats. Left, PCA plot based on microbial genera relative abundance data. Each dot indicates a sample (different time points, expressed in weeks, are illustrated with different shapes), while dotted ellipses indicate 95% confidence level. AL, *ad libitum*; CR, caloric restriction. Right, heatmap illustrating microbial genera with significantly differential abundance between AL and CR groups (edgeR analysis followed by Benjamini-Hochberg correction). Columns represent samples, while rows represent genera. Only genera with abundance >0.25% are shown, and ordered first according to the group in which they are significantly more abundant, and then based on the phylum to which they belong (B, Bacteroidetes; F, Firmicutes; G, Glomeromycota; P, Proteobacteria; S, Spirochaetes).

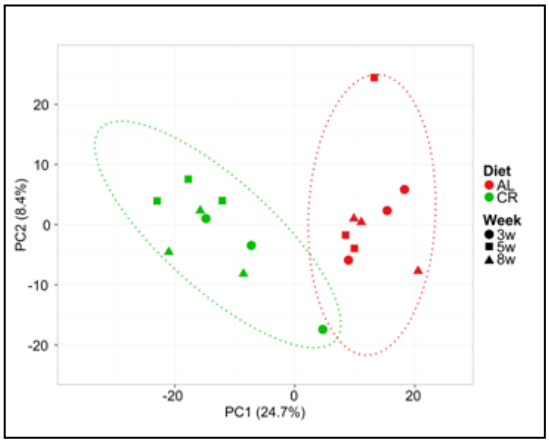

Figure 14: Beta-diversity at functional level observed upon caloric restriction treatment on young rats. PCA plot based on relative abundance of microbial protein functions. Each dot indicate a sample (different time points are illustrated with different shapes), while dotted ellipses indicate 95% confidence level. AL, *ad libitum*; CR, caloric restriction.

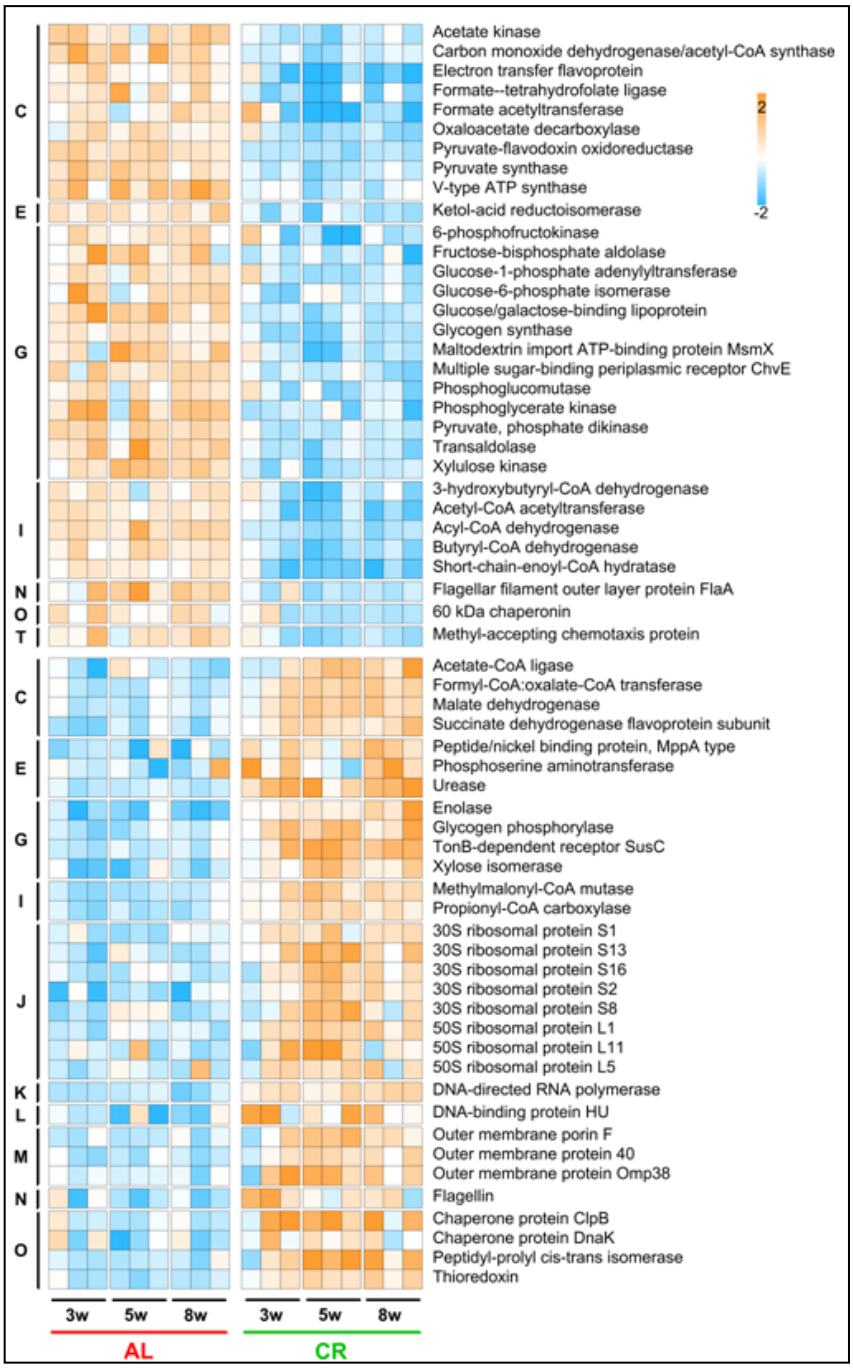

Figure 15: Changes in metaproteome functional expression observed upon caloric restriction treatment on young rats. Heatmap illustrating microbial functions with significantly differential abundance between AL and CR groups (edgeR, FDR <0.05). Columns represent samples, while rows represent functions. Only functions with abundance >0.25% are shown, and ordered first according to the group in which they are significantly more abundant, and then based on the COG category to which they belong.

Then, the functional traits discriminating the activity of the microbiota of CR and AL rats were investigated. As for taxonomic data, the PCA plot (Figure 14) clearly shows a clustering between CR and AL groups based on functional abundance data (*P* <0.001

PERMANOVA between the two groups considering all weeks). Furthermore, differential abundance analysis showed 167 functions with significant changes in the two groups (the 62 functions exceeding 0.25% of relative abundance in at least one group are listed in the Figure 15).

The CR intervention was observed to induce a deep rearrangement of the microbial proteins, involving both catalytic and structural/antigenic functions belonging to several different Clusters of Orthologous Group (COG) categories (with "carbohydrate metabolism and transport" being the most represented). Specifically, several enzymes implicated in carbohydrate degradation were significantly underrepresented after CR, contrary to the expression of various ribosomal, outer membrane, DNA-binding and stress-related proteins that appeared to be stimulated by the CR treatment. The differential trend was already appreciable after 3 weeks of treatment and reached top values after 5 weeks, that were kept also up to 8 weeks.

Caloric restriction promotes expression of propionogenic enzymes and limits abundance of butiryogenic and acetogenic enzymes

Several enzymes implicated in SCFAs biosynthesis were noticed to exhibit differential expression between CR and AL groups. Therefore, according to the orthologous genes listed in the corresponding KEGG pathways, all the functions related to butyrate, propionate and acetate metabolism were inspected.

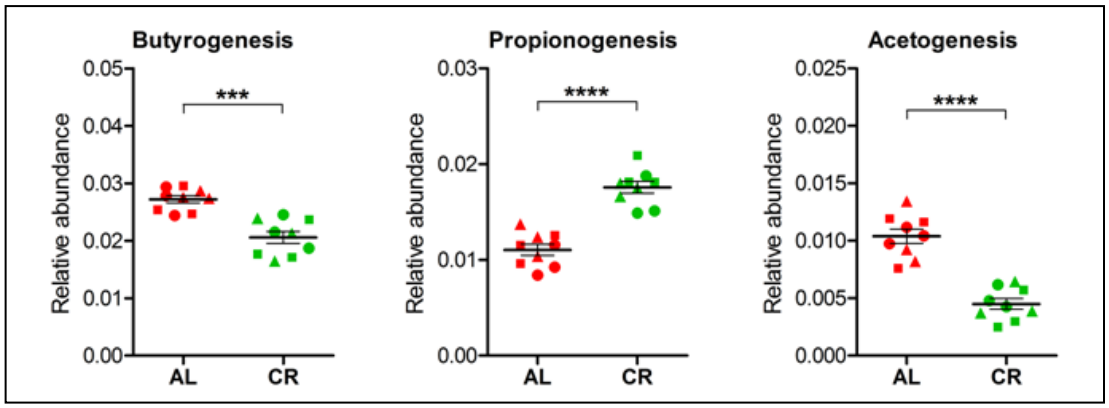

Figure 16: Scatter plots showing the relative abundance of enzymes involved in short-chain fatty acid biosynthesis. AL, *ad libitum*; CR, caloric restriction. Each dot indicate a sample; different time points are illustrated with different shapes (same as in Figure 13). Means and standard deviations are reported. ***P <0.001; ****P <10^{-5}.

First, the global relative abundance of all enzymes contributing to butyrogenesis, propionogenesis and acetogenesis in the different groups of animals was calculated. Thus, propionogenic functions were found significantly higher in the CR fecal microbiota, whereas a decrease in proteins having a role in butyrate and acetate biosynthesis was observed in CR fed rats (Figure 16).

Focus on host secreted proteins

In addition to the characterization of the microbial proteins represented in the stool sample (microbiota), metaproteomics allows also the identification of the host proteins secreted in the gut lumen and embodied within the feces. Therefore, a comparison between AL and CR fed rats concerning the host protein expression profile was performed.

Host differential functions	log_2FC(CR/AL)	FDR
Keratin, type II cytoskeletal 4	2.55	2.68E-04
Keratin, type I cytoskeletal 15	1.78	2.18E-02
Alpha amylase 1	1.75	2.68E-04
Keratin, type II cytoskeletal 1	1.73	2.25E-02
Bile salt activated lipase	1.53	3.27E-03
Neutral and basic amino acid transport protein rBAT	1.46	4.46E-04
Keratin, type II cuticular Hb4	1.45	2.18E-02
Lactase phlorizin hydrolase	1.44	1.20E-02
Complement C3	1.39	4.50E-05
Keratin, type I cytoskeletal 13	1.34	5.56E-04
Serine protease inhibitor A3N	1.29	2.68E-04
Acylamino acid releasing enzyme	1.24	3.28E-03
Neutral ceramidase	1.18	3.27E-03
Serine protease inhibitor A3L	0.95	5.23E-04
Cytochrome c, somatic	-1.12	1.71E-02

Table 2: Host differential functions upon caloric restriction treatment on young rats. The base 2 logarithm of CR/AL fold-change, log_2FC(CR/AL), and the edgeR *P* adjusted value, FDR, are reported. Functions are ordered by decreasing fold-change.

As a result, only 1 protein, namely the hemeprotein Cytochrome c, was found as significantly higher in the fecal proteome of the AL-fed young rats, while 14 in the CR animals (as listed in Table 2). Cytochrome c is well known as an indicator of cell death burden in any organ or tissue. It is released during mitochondrial damage that is

associated with processing of apoptosis, cell lysis during necrosis and even reversible mitochondrial and cell injury (Small and Gobe 2012). Among the proteins more abundant in CR-fed rats, 2 keratins of the basic type II (K1 and K4) and 2 of the acidic type I (K13 and K15) were detected. Colonic keratins levels have been reported to decrease in human colon during inflammatory stress, as observed in ulcerative colitis (Corfe et al 2015). In particular, keratin 1 has been strongly suggested to play a key role in maintaining epithelial barrier and intestinal mucosa permeability (Dong et al 2017). Other host proteins with significant higher abundance in the fecal samples of CR rats included catabolic enzymes (such as alpha-amylase, lactase, ceramidase), 2 serine inhibitors of proteases, an amino acid transporter, and, interestingly, the bile salt activated lipase, suggesting an increased bile abundance in the colonic lumen of CR-fed animals.

1.3 Caloric restriction-induced changes in the fecal microbiota are kept in the adulthood but start to be reversed just after 1 week of ad libitum diet

1.3.1 Aim of the study

As described by Fraumene and colleagues (Fraumene et al 2017) and in Section 1.2, compositional and functional changes, mainly involving *Lactobacillus* spp., were early produced in the fecal microbiota of CR-fed rats and maintained and heightened after 9 months of hypocaloric diet. Here, the purpose was to investigate, under a taxonomic and functional perspective, if these changes can be still observed in aged rats, and if a AL diet restoration after over a year of CR treatment could reverse the fecal microbiota structure and activity.

1.3.2 Experimental design

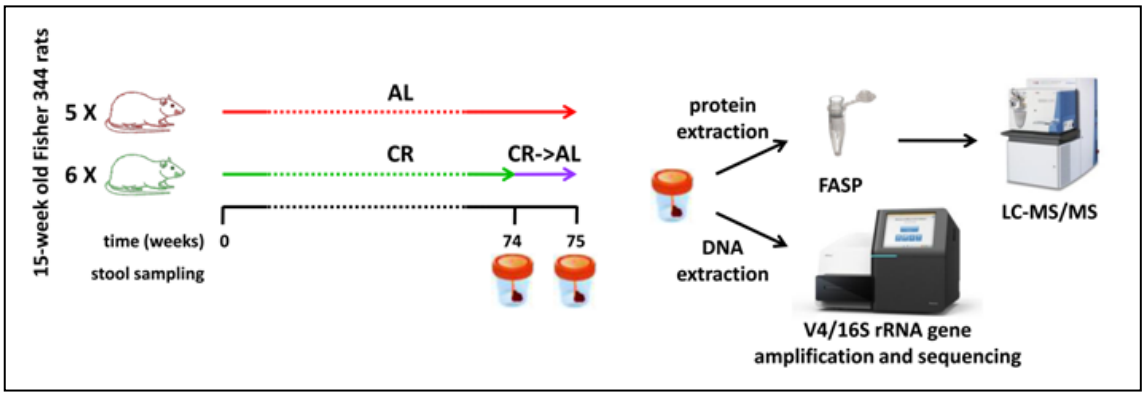

Figure 17: Schematic illustrating the experimental design of the study.

Fecal samples were collected from Fisher 344 rats treated both with CR and AL diets for 1.5 years (time point: 74 weeks); then, CR animals started to receive an AL diet, and their feces were sampled a week from diet change (CR->AL). DNA and proteins were extracted from fecal samples, with the aim of carrying out both 16S rRNA gene sequencing and metaproteome analysis, as illustrated in Figure 17.

1.3.3 Material & Methods

Animals and samples

A total of 11 Fisher 344 female rats were used from a colony available at the Department of Biomedical Sciences, University of Cagliari. Animals were fed with Purina Rodent Lab Chow diet with 3% of fat (Mucedola srl). Animal studies were reviewed and approved by the Institutional Animal Care and Use Committee of the University of Cagliari.

Animals were fed AL until 15 weeks of age; then, 6 of them were randomly selected to form a separate group and were given 70% of the AL ratio (CR); animals from the same group were housed two per cage. Feed to the CR group was delivered nightly at 1AM through a computer assisted automated food dispenser. After 74 weeks, CR group came back to an AL diet (CR->AL); fecal samples were collected at 75 weeks and 1 week after the diet change.

All collected fecal samples (n = 17) were immediately stored at -80°C until use. At the time of the analyses, stool samples were thawed at 4°C, and two portions were collected from each of them for protein and DNA extraction, respectively.

DNA extraction and 16S rDNA gene sequencing

17 fecal samples were subjected to DNA extraction, purification and quantification as already described in Materials & Methods of Section 1.2.

Libraries were constructed using Illumina's recommendations as implemented in 16S Metagenomic Sequencing Library Preparation guide. The 515F and 806R primers (GTGCCAGCMGCCGCGGTAA and GGACTACHVGGGTWTCTAAT, respectively), modified to contain adaptors for MiSeq sequencing, were used to amplify the variable region 4 (V4) of the 16S rRNA gene. Also in this case, two separate gene amplification reactions were performed for each experiment, pooled together and cleaned up with AMPure XP beads. Next, with a second PCR dual index barcodes and sequencing adapters were attached using the Illumina Nextera XT kit. Then, after library size check and quantification with a Bioanalyzer 2100, PCR products were pooled together. V4 amplicon sequencing was performed on the Illumina MiSeq platform using the MiSeq

Reagent Kit v2 according to the manufacturer's specifications to generate paired-end reads of 201 bases in length in each direction. Data quality control and analyses were performed using QIIME. The overlapping paired-end reads were merged using the script join_paired_ends.py inside the QIIME package, retaining for further analysis only reads with a length >200 bps. OTUs generation and their taxonomy assignment were carried out with the pipeline explained before (see Section 1.2.3).

Protein extraction and metaproteomic analysis

A total of 17 fecal samples were subjected to protein extraction and metaproteomic analysis as already described (see Section 1.2.3) with only one modification: before mass spectrometry analyses, peptide mixtures from animals of same cage were not pooled together, as happened for the young rats experiment, but run in LC-MS/MS singularly.

Statistical analysis and graph generation

Differential abundance analysis was performed on count data (spectral counts for metaproteomic results and read counts for 16S rDNA gene sequencing results) by uploading them to the web application MicrobiomeAnalyst. Features with prevalence in <10% of samples in each comparative analysis were filtered out and subjected to the RLE normalization. Statistical comparative analysis was performed with edgeR with an FDR cutoff of 0.05.

PCA plots were made using ClustVis with default parameters, while heatmaps with Morpheus, both starting from relative abundance data. Scatter plots were generated using GraphPad Prism on relative abundance data, and Student's t test was applied to calculate significance of differences between total enzyme abundance means.

1.3.4 Results

Gut microbiota taxonomic evaluation in CR treated adult rats compared to AL and after 1 week of reversion to AL

16S rRNA gene sequencing provided 1,429,643 reads, belonging to 217 different microbial genera. Regarding the analysis of the metaproteome, a total of 126,913 mass spectra could be matched with 1,000 different protein functions, assigned to 241 different microbial genera.

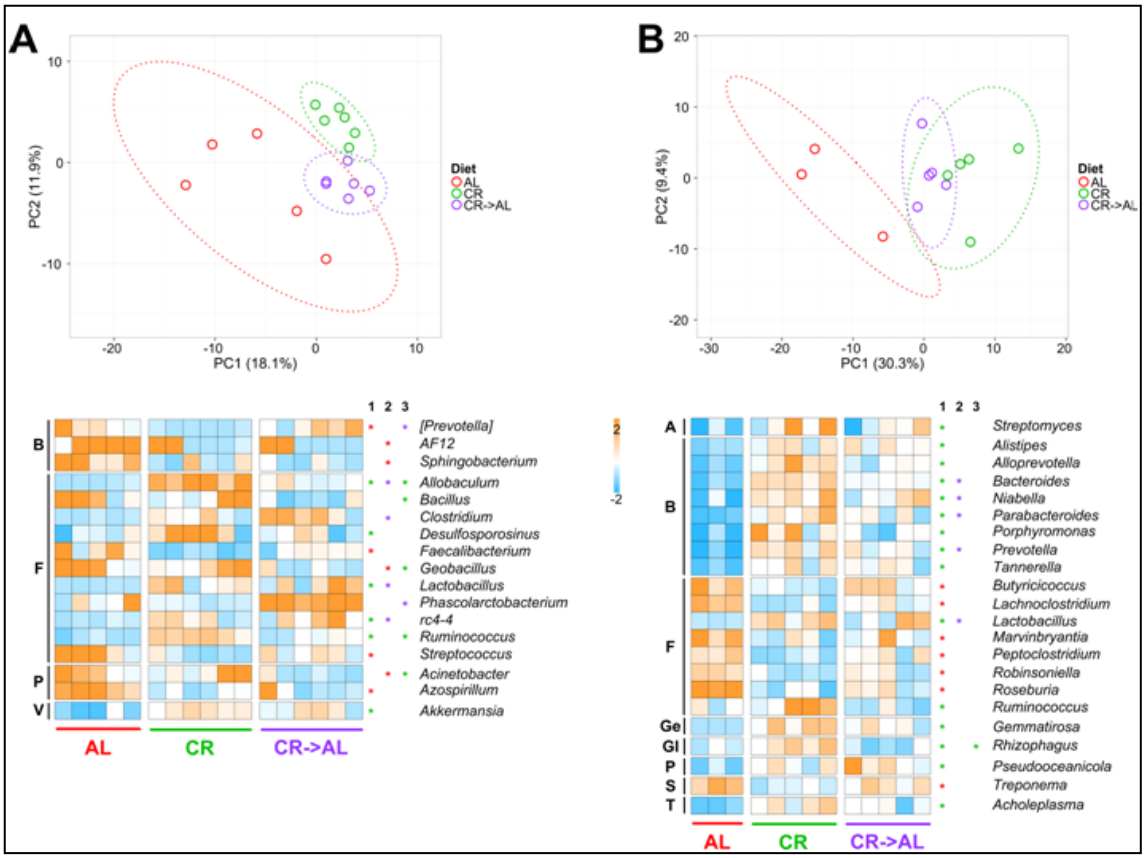

Figure 18: Adult rats (1.5 years of treatment): changes in fecal samples taxonomic composition at genus level based on 16S rDNA gene sequencing (A) and metaproteomics (B). AL, *ad libitum*; CR, caloric restriction; CR->AL, 1-week reversion from caloric restriction to *ad libitum*. (A) Top, PCA plot based on 16S rDNA gene sequencing relative abundance data at the genus level. Each dot indicate a sample, while dotted ellipses indicate 95% confidence level. Bottom, heatmap illustrating microbial genera with significantly differential abundance between groups (edgeR, FDR <0.05). Heatmap columns represent samples, while rows represent genera. Asterisks in supplementary columns 1, 2 and 3 indicate genera with significantly differential abundance upon AL vs CR, AL vs CR->AL, and CR vs CR->AL comparisons, respectively. The asterisk color refers to the group in which the genus was found as more abundant. Only genera with abundance >0.1% are shown, and ordered based on the phylum membership (B, Bacteroidetes; F, Firmicutes; P, Proteobacteria; V, Verrucomicrobia). (B) Same as in (A), but concerning metaproteomic data. Further phylum abbreviations: A, Actinobacteria; Ge, Gemmatimonadetes; Gl, Glomeromycota; S, Spirochaetes; T, Tenericutes.

Taxonomic results from 16S rRNA (A) and metaproteomic (B) analyses are illustrated in Figure 18. Long-term CR was found to induce structural changes that were kept after 1.5 years of dietary treatment, as PCA plots based on relative genus abundances clearly show (Figure 18, top). Even more interestingly, samples collected from rats after only one week of "diet reversion" (CR->AL) clustered in between AL and CR groups, indicating that an increase in the quantity (and not quality) of feed eaten could provoke a perturbation of the gut microbiota structure in adulthood. The two heatmaps of Figure 18 list the microbial genera (with relative abundance >0.1%) that showed a significantly differential abundance among the groups analyzed, according to 16S rRNA (left) and metaproteomic data (right). In general, genus abundances were more heterogeneously distributed among groups and among phyla based on 16S rRNA information; on the contrary, a more evident trend could be observed in metaproteomic data, with CR->AL genera exhibiting almost always intermediate abundance values if compared to AL and CR, and all Bacteroidetes differential genera being enriched in CR-fed animals, consistently with the results obtained on younger rats (see Section 1.2.4). *Lactobacillus* and *Ruminococcus* spp. were the only genera found to be significantly more represented in CR microbiota compared to AL controls according to both 16S rRNA and metaproteomic approaches.

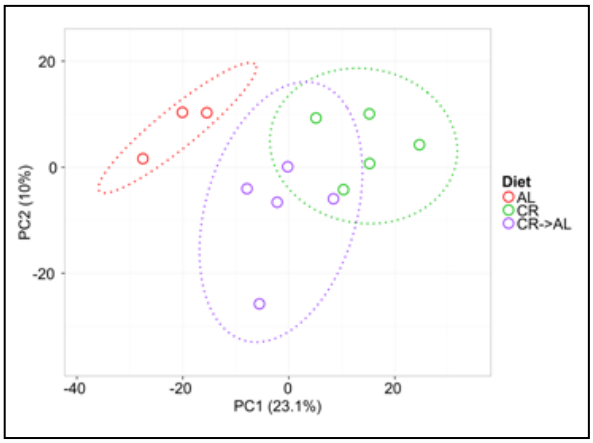

Figure 19: Beta-diversity at functional level observed upon caloric restriction treatment on adult rats (1.5 years of treatment). AL, *ad libitum*; CR, caloric restriction; CR->AL, 1-week reversion from caloric restriction to *ad libitum*. PCA plot based on relative abundance of microbial protein functions. Each dot indicate a sample, while dotted ellipses indicate 95% confidence level.

Gut microbiota functional evaluation in CR treated adult rats compared to AL and after 1 week of AL reinstatement

The metaproteome dataset was also investigated from the functional point of view to identify differences in the expression of proteins. PCA plot based on functional data showed a group separation consistent to that observed for taxonomic data (Figure 19).

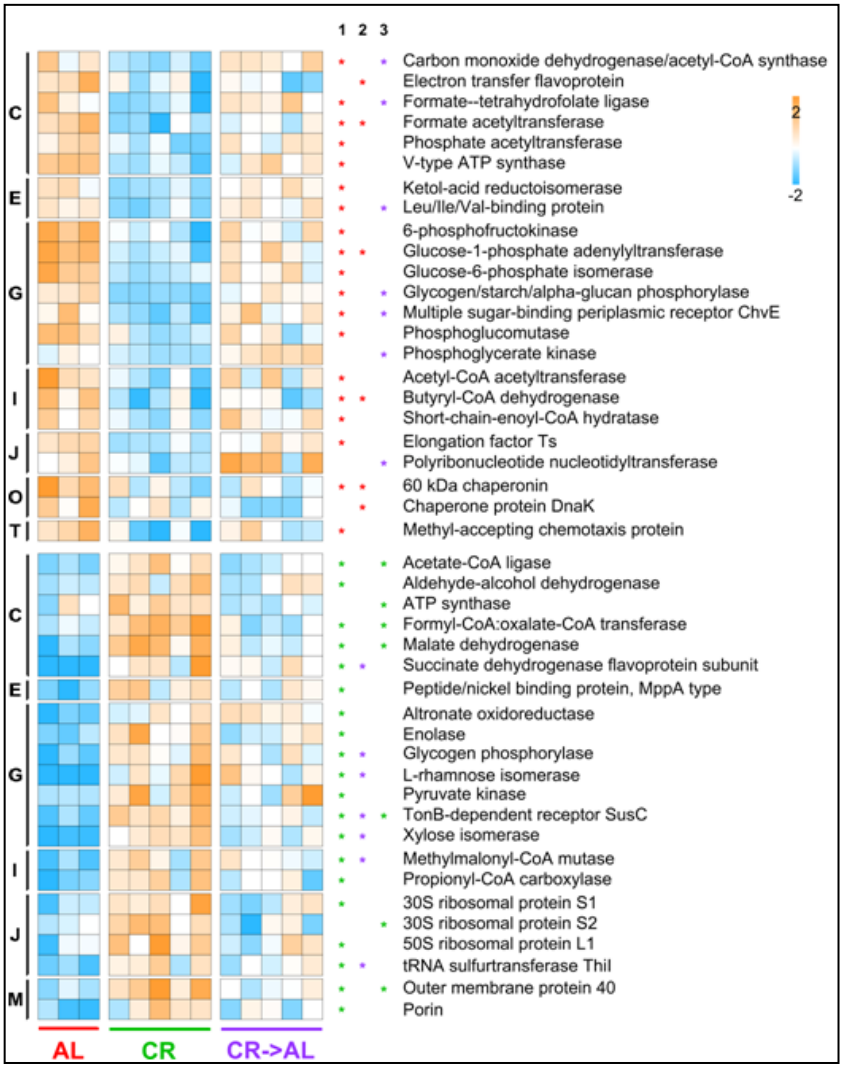

Figure 20: Changes in metaproteome functional expression observed upon caloric restriction treatment on adult rats (1.5 years of treatment). AL, *ad libitum*; CR, caloric restriction; CR->AL, 1-week reversion from caloric restriction to *ad libitum*. Heatmap illustrating microbial functions with significantly differential abundance between groups (edgeR, FDR <0.05). Heatmap columns represent samples, while rows represent functions. Asterisks in supplementary columns 1, 2 and 3 indicate functions with significantly differential abundance upon AL vs CR, AL vs CR->AL, and CR vs CR->AL comparisons, respectively. The asterisk color refers to the group in which the function was found as more abundant. Only functions with abundance >0.25% are shown, and ordered first according to the group in which they are significantly more abundant, and then based on the COG category to which they belong.

A total of 113 functions showed significantly differential abundance among the three sample groups; among them, 73 (65%) had been consistently found as differential in the young rat experiment (Section 1.2.4). Heatmap in Figure 20 lists 45 differential functions, based on an abundance threshold of >0.25% in at least one group, belonging to many different COG categories ("carbohydrate metabolism and transport" was also in this case the most represented). As previously observed in young rats, enzymes implicated in acetate/butyrate biosynthesis were significantly downregulated after CR, while the expression of those participating to propionogenesis appeared to be promoted by the CR treatment.

Several catalytic functions involved in pentose metabolism (including L-rhamnose and xylose isomerases and altronate oxidoreductase) were also revealed to be consistently higher in CR fecal metaproteome. Additionally, regarding functions that were observed as differentially abundant between AL and CR rats, in almost all cases in CR->AL rats the abundance value of the same function was intermediate between those measured in CR and AL rats, corroborating that a single week of "diet reversion" was sufficient to modify the functional profile of the microbiota.

Genus-specific functional analysis reveals peculiar functional shifts related to caloric restriction

The contribution of specific microbial genera to the functional activity of the gut microbiota was investigated combining functions with taxonomy annotations. In particular, data obtained both from adult rats and young rats (Section 1.2) were evaluated to illustrate the top proteins expressed by the most represented members of the fecal microbiota, including the Firmicutes members *Lactobacillus*, *Clostridium*, *Oscillibacter* and *Ruminococcus*, and *Bacteroides* and *Prevotella*, belonging to the Bacteroidetes phylum.

Lactobacillus was found to be significantly enriched in CR samples in adult rats, consistently with previous reports (Fraumene et al 2017) and in Section 1.2.4. Therefore, light was shed on which functions and metabolic activities related to

Lactobacillus spp. were increased in the CR microbiota at different time points (Figure 21).

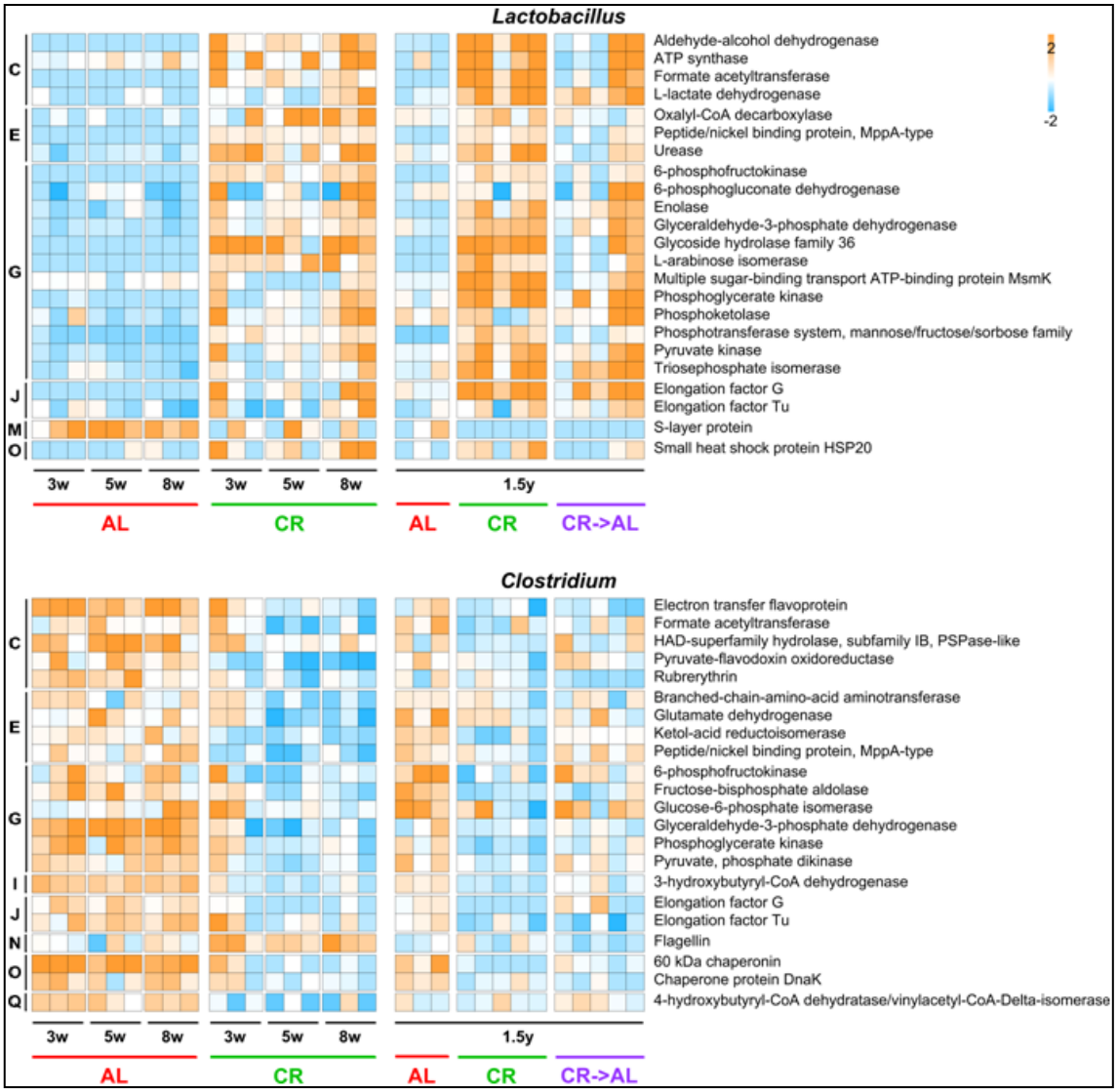

Figure 21: Functional expression profile of *Lactobacillus* (top) and *Clostridium* (bottom) metaproteomes. Relative abundance values concerning the young rat experiment (left, up to 8 weeks of treatment) and adult rat experiment (right, 1.5 years of treatment) are shown. AL, *ad libitum*; CR, caloric restriction; CR->AL, 1-week reversion from caloric restriction to *ad libitum*. Heatmap columns represent samples, while rows represent functions. Only functions with abundance >0.1% (*Lactobacillus*) and >0.05% (*Clostridium*) are shown (ribosomal proteins were excluded), and ordered according to the COG category to which they belong.

Looking at *Lactobacillus*-specific proteins, many enzymes were found to be significantly higher in the CR metaproteome; these enzymes were involved in pentose (belonging to pentose phosphate and pentose-glucuronate interconversion pathways), hexose

(such as glycolytic enzymes, but also alpha-galactosidases comprised in the glycoside hydrolase, GH, family 36), and pyruvate (to lactate and formate) metabolism, but also in oxalate and urea degradation. In addition, proteins responsible for transport and phosphorylation of carbohydrates were observed as more abundant in CR treated rats. Intriguingly, the S-layer protein was the only *Lactobacillus* protein function following the opposite trend (significantly higher in AL-fed rats). S-layer proteins are crystalline proteins that line over the bacterial membrane and, when released, adhere to the epithelial cell surface, enhancing bacterial-cell adhesion (Wang et al 2017b). Further, their expression increase when bacteria enter the stationary phase (Palomino et al 2016). Thus, their abundance in fecal sample of AL treated rats is in agreement with a reduced growth rate and relative abundance of this genus.

Clostridium spp. were instead reduced after CR treatment; as a consequence, many clostridial functions, including several enzymes mainly involved in glycolysis, but also in pyruvate and butyrate metabolism, were significantly decreased (Figure 21). Moreover, a clostridial protein with relevant antigenic properties, flagellin, was unexpectedly observed as more abundant in CR animals (young).

Bacteroides and *Prevotella*, as most of the Bacteroidetes genera, were higher in CR fed rats. When looking at proteins assigned to these genera (Figure 22), several proteins were found to be consistently increased in the CR rat fecal microbiota, principally involved in membrane transport, metabolism and protein folding. *Bacteroides* enzymes responsible for pentose catabolism and *Prevotella* proteins belonging to the TonB-dependent transport system were typically more abundant in the CR metaproteome.

Finally, the functions of *Oscillibacter* and *Ruminococcus* (Figure 23), two Firmicutes members exhibiting different behavior, were also examined. *Oscillibacter* proteins, according to most Firmicutes, were visibly depleted in the CR metaproteome, including several enzymes participating to butyrate, pyruvate and acetate metabolic pathways. On the contrary, *Ruminococcus* was enriched after long-term CR with respect to the control AL counterpart, with overexpression of cellulases (Glycosil hydrolase family 9) and a concurrent decrease of the oxidative stress-related protein rubrerythrin.

Figure 22: Functional expression profile of *Bacteroides* and *Prevotella* metaproteomes. Relative abundance values concerning the young rat experiment (left, up to 8 weeks of treatment, see Section 1.2) and adult rat experiment (right, 1.5 years of treatment) are shown. AL, *ad libitum*; CR, caloric restriction; CR->AL, 1-week reversion from caloric restriction to *ad libitum*. Heatmap columns represent samples, while rows represent functions. Only functions with abundance >0.25% (*Prevotella*) and >0.1% (*Bacteroides*) are shown (ribosomal proteins were excluded), and ordered according to the COG category to which they belong.

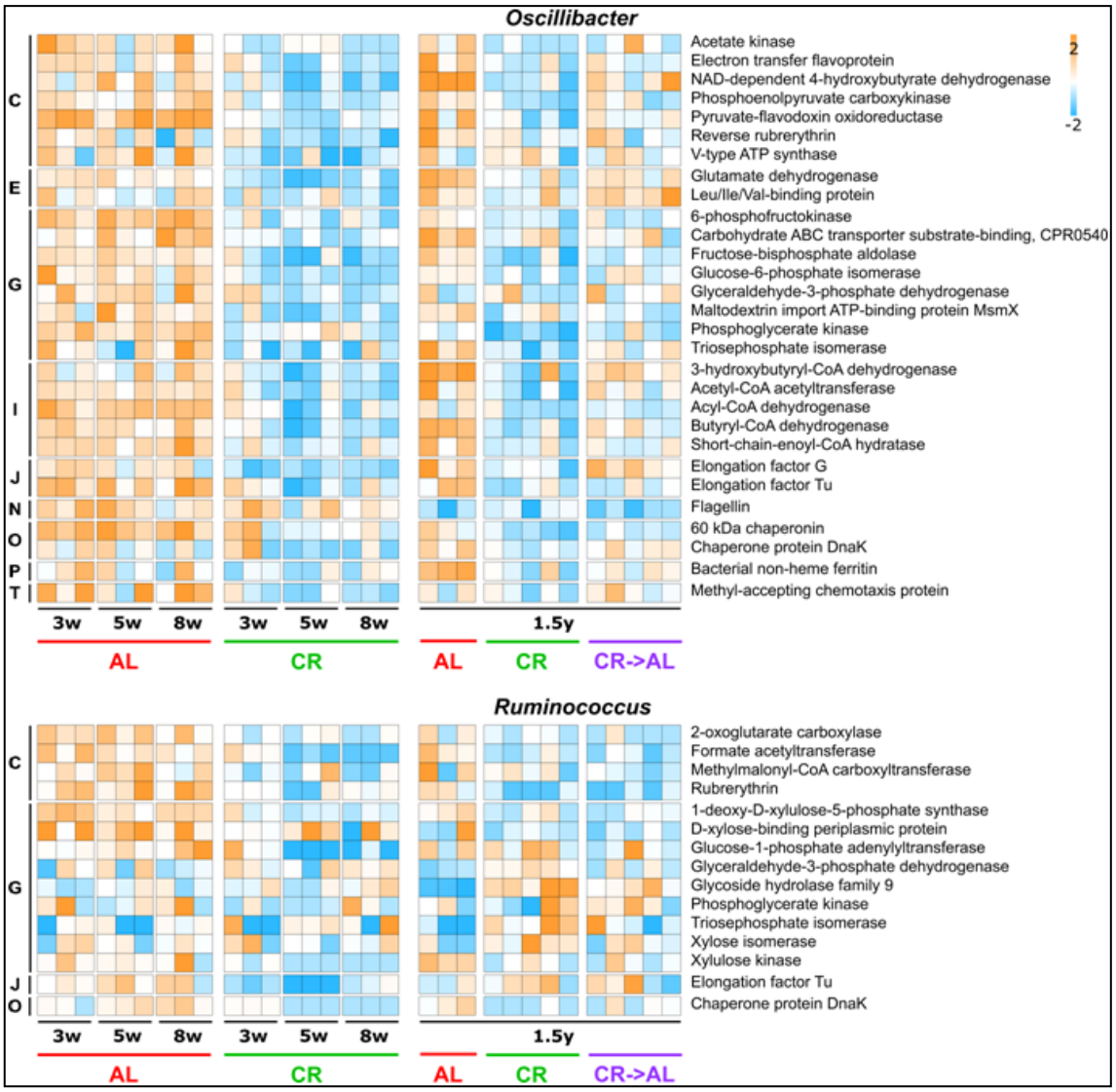

Figure 23: Functional expression profile of *Oscillibacter*, and *Ruminococcus* metaproteomes. Relative abundance values concerning the young rat experiment (left, up to 8 weeks of treatment, see Section 1.2) and adult rat experiment (right, 1.5 years of treatment) are shown. AL, *ad libitum*; CR, caloric restriction; CR->AL, 1-week reversion from caloric restriction to *ad libitum*. Heatmap columns represent samples, while rows represent functions. Only functions with abundance >0.1% (*Oscillibacter*) and >0.05% (*Ruminococcus*) are shown (ribosomal proteins were excluded), and ordered according to the COG category to which they belong.

Propionogenic enzymes promotion and limitation in the abundance of butyrogenic and acetogenic enzymes persisted after long-term caloric restriction treatment

As observed for young rats, several enzymes implicated in SCFA biosynthesis (mainly propionate, acetate and butyrate) exhibited differential expression between CR and AL fecal microbiota in adulthood. These pathways were again reconstructed according KEGG database.

Concerning the relative abundance of all enzymes contributing to butyrogenesis, propionogenesis and acetogenesis in the different study groups, functions involved in propionate production were again significantly higher in the CR metaproteome, whereas a decrease in proteins involved in butyrate and acetate biosynthesis was observed in CR fed rats (Figure 24), both results consistently with what observed in young rats (see Figure 16). Very interestingly, 1.5 years CR-treated rats showed a rapid restoration of the levels of acetogenic enzyme abundance, typical of an AL diet, after 1 week of AL diet reinstatement, while changes in butyrogenic and propionogenic enzymes appeared to be slower.

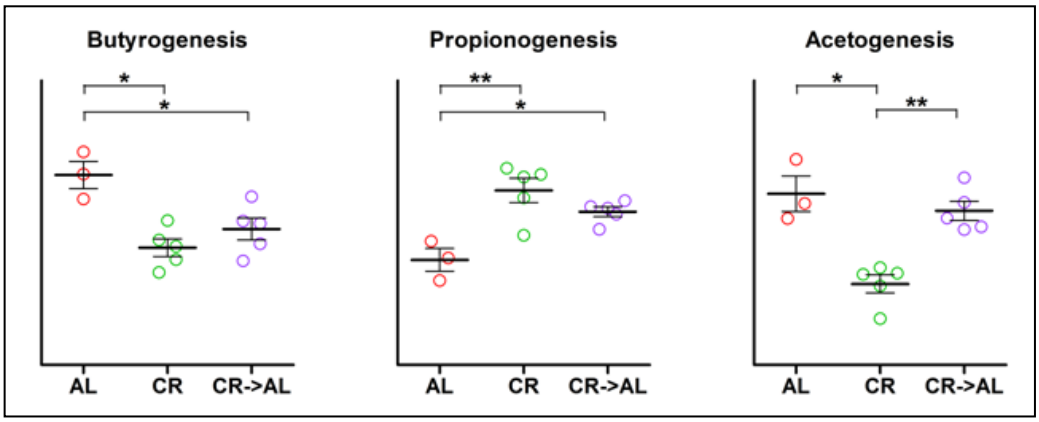

Figure 24: Scatter plots showing the relative abundance of enzymes involved in short-chain fatty acid biosynthesis in adult rats. AL, *ad libitum* (red); CR, caloric restriction (green); CR->AL, 1-week reversion from caloric restriction to *ad libitum* (violet). Each dot indicate a sample. Means and standard deviations are reported. **P* <0.05; ***P* <0.01.

To investigate the taxonomic contribution to the SCFA biosynthetic pathways, a map illustrating the relative abundance of the enzymes involved in buryrate (Figure 25), propionate (Figure 26) and acetate production (Figure 27), including the related taxonomic classification (at phylum and genus level), were generated, taking into account both the adult and the previously described young animals. A high level of cross-feeding appears to take place for butyrogenesis; indeed, each pathway step was found to be due to different Clostridia members (such as *Butyricicoccus*, *Butyrivibrio*, *Clostridium*, *Eubacterium*, *Faecalibacterium*, *Oscillibacter* and *Roseburia*).

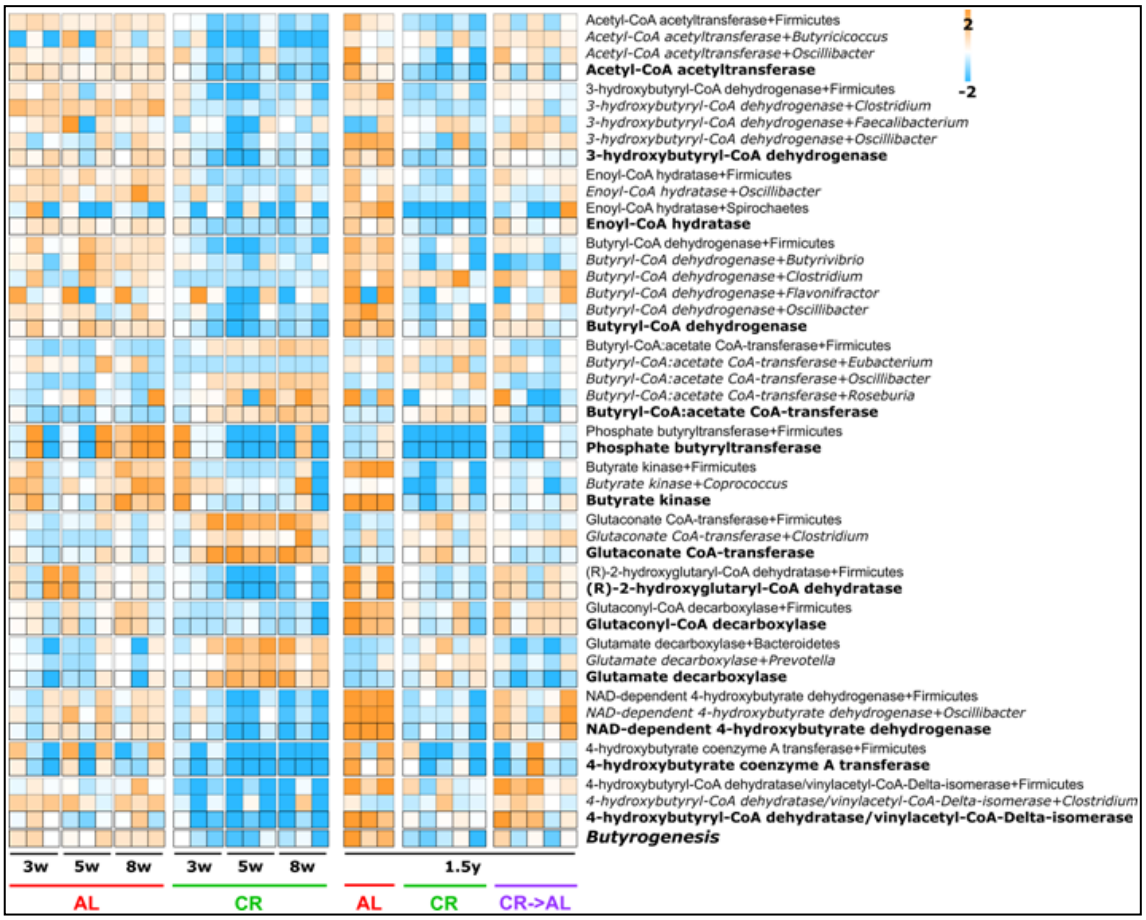

Figure 25: Expression profile of butyrogenic enzymes in young and adult rats. Relative abundance values concerning the young rat experiment (left, up to 8 weeks of treatment, see Section 1.2) and adult rat experiment (right, 1.5 years of treatment) are shown. AL, *ad libitum*; CR, caloric restriction; CR->AL, 1-week reversion from caloric restriction to *ad libitum*. Heatmap columns represent samples, while rows represent functions. Only functions and function-taxon combinations detected in at least half of AL or CR samples are shown. Functions/pathways in bold account for the total abundance of that function/pathway in the metaproteome, independently of the specific taxon to which the function/pathway was assigned.

Regarding propiogenesis, its central and characteristic reactions (catalyzed by methylmalonyl-CoA mutase and propionyl-CoA carboxylase) could be assigned essentially to *Bacteroides* and *Prevotella* spp. Of note, the abundance of acetogenic carbon monoxide dehydrogenase (assigned mostly to *Blautia*) was rapidly and strongly increased in previously CR-fed rats after just a week of AL feeding, restoring abundance levels comparable to those reached in the long-term AL fed group.

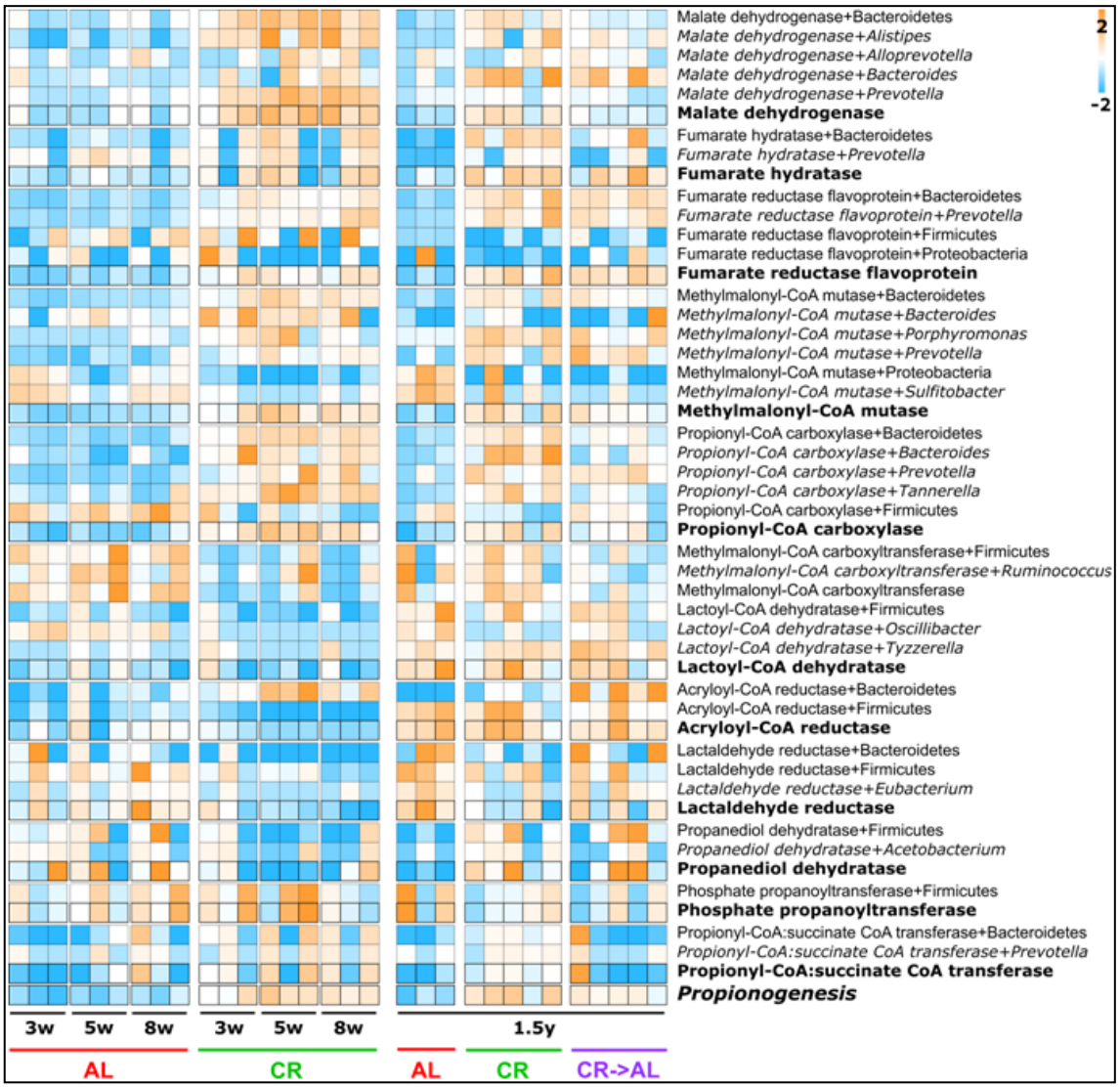

Figure 26: Expression profile of propionogenic enzymes in young and adult rats. Relative abundance values concerning the young rat experiment (left, up to 8 weeks of treatment, see Section 1.2) and adult rat experiment (right, 1.5 years of treatment) are shown. AL, *ad libitum*; CR, caloric restriction; CR->AL, 1-week reversion from caloric restriction to *ad libitum*. Heatmap columns represent samples, while rows represent functions. Only functions and function-taxon combinations detected in at least half of AL or CR samples are shown. Functions/pathways in bold account for the total abundance of that function/pathway in the metaproteome, independently of the specific taxon to which the function/pathway was assigned.

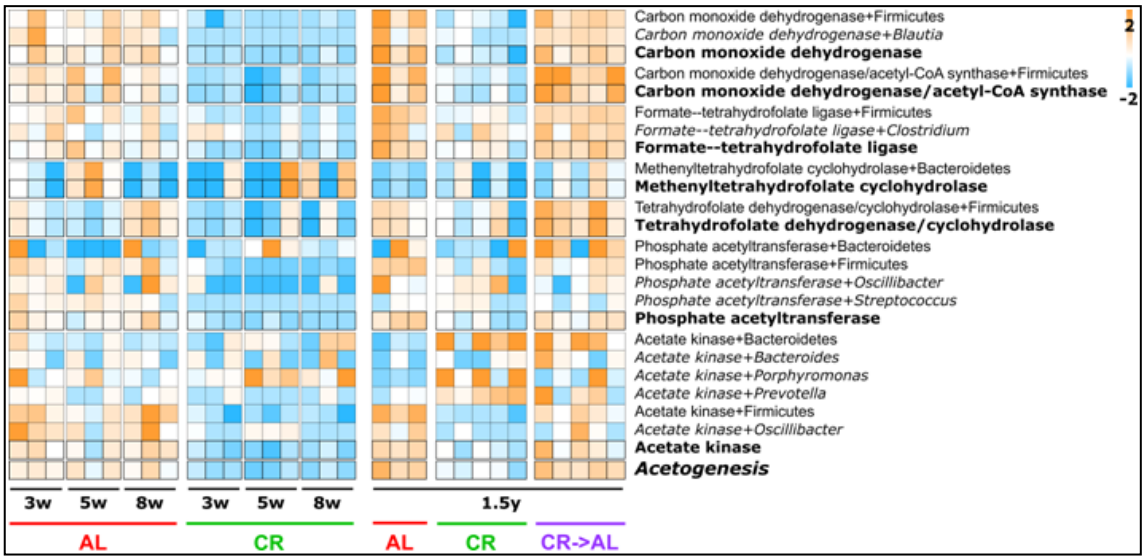

Figure 27: Expression profile of acetogenic enzymes in young and adult rats. Relative abundance values concerning the young rat experiment (left, up to 8 weeks of treatment, see Section 1.2) and adult rat experiment (right, 1.5 years of treatment) are shown. AL, *ad libitum*; CR, caloric restriction; CR->AL, 1-week reversion from caloric restriction to *ad libitum*. Heatmap columns represent samples, while rows represent functions. Only functions and function-taxon combinations detected in at least half of AL or CR samples are shown. Functions/pathways in bold account for the total abundance of that function/pathway in the metaproteome, independently of the specific taxon to which the function/pathway was assigned.

Host secreted proteins analysis

Also the expression profile of the detected host proteins was compared between AL- and CR-fed rats. As a result, 16 proteins were found as significantly higher in the fecal proteome of CR adult rats, while 4 in the AL animals (as listed in Table 3). The former were represented, consistently with the young rats experiment (Table 2), most exclusively by keratins (n = 10), of which 7 cuticular and 3 cytoskeletal; additionally, sucrase isomaltase and aminopeptidase Xaa Pro were overepresented in the CR fecal metaproteome, as well as enzymes involved in lipid degradation: inactive pancreatic lipase related protein 1, bile salt-activated lipase and neutral ceramidase (the last two showed this behavior also in younger rats). Neutral ceramidase catalyzes the hydrolysis of sphingolipids in the gut lumen (Ito et al 2014), while bile salt-activated lipase digests triglycerides, and changes in its expression have been reported in animal models of obesity fed high-fat diet (Bae et al 2016). Their higher abundance in CR animal feces is in line with the well-studied effect of caloric restriction and fasting on lipid metabolism (Higami et al 2004). On the contrary, Murinoglobulin-1, Calcium activated chloride

channel regulator 1 and Mucin-2 were detected in higher proportions in the control AL-fed rats. The former is known to be involved in the mucosal anti-inflammatory response as a protease inhibitor (Regler et al 1991). Accordingly, mucin production is also assessed as a common defense mechanism to protect the underlying mucosa against pathogens and calcium activated chloride channel regulator 1 may be involved in the regulation of mucus production and/or secretion by goblet cells (Toda et al 2002). Finally, actin, among the most abundant protein in the host cells, was released in higher amount in feces of the AL group, possibly due to more intense exfoliation processes. Together, these proteins may suggest a state of sub-clinical colonic inflammation, as confirmed also by the reduction of keratin content.

These data, hence, provide new insights on the metabolic state associated with intestinal homoeostasis and aging retardation by long term CR.

Host differential functions	log2FC(CR/AL)	FDR
Inactive pancreatic lipase related protein 1	2.93	3.20E-07
Keratin, type I cuticular Ha5	2.14	1.47E-03
Keratin, type II cuticular Hb2	1.93	9.14E-03
Keratin, type II cytoskeletal 2 epidermal	1.93	1.37E-02
Keratin, type II cuticular Hb5	1.46	1.81E-03
Keratin, type I cuticular Ha1	1.43	2.71E-02
Keratin, type II cuticular Hb4	1.38	6.98E-03
Bile salt activated lipase	1.36	1.59E-06
Keratin, type I cytoskeletal 10	1.22	9.14E-03
Xaa Pro aminopeptidase 2	1.19	1.47E-03
Keratin, type II cytoskeletal 1	1.11	2.47E-02
Keratin, type II cuticular Hb1	1.06	1.81E-03
Sucrase isomaltase, intestinal	1.03	1.90E-03
Keratin, type I cuticular Ha4	0.98	2.40E-02
Neutral ceramidase	0.87	1.37E-02
CUB and zona pellucida like domain containing protein 1	0.79	1.05E-02
Actin, cytoplasmic 1	-0.73	1.37E-02
Mucin 2 (Fragment)	-0.81	3.98E-03
Calcium activated chloride channel regulator 1	-0.90	4.14E-02
Murinoglobulin 1	-2.64	1.25E-03

Table 3: Host differential functions upon caloric restriction treatment on adult rats. The base 2 logarithm of CR/AL fold-change, $\log_2$FC(CR/AL), and the edgeR *P* adjusted value, FDR, are reported. Functions are ordered by decreasing fold-change.

1.4 Effects of sourdough-leavened bread diet integration on the fecal microbiota in a model of caloric restriction in rats

1.4.1 Aim of the study

As reported in 1.2 and 1.3 sections, a reduction of feed intake to 70% (caloric restriction, CR), even without a change in the feed composition, affected dramatically the structure and the functional activities of the fecal microbiota in both young and adult rats.

Here, this animal model of CR was employed to investigate the effect of diet integration with baker's yeast leavened bread or with sourdough-leavened bread on the gut microbiota of young rats. When the two processes are employed to prepare 2 bread products that are based on the very same raw material but different leavening microbes (*Saccharomyces cerevisiae* vs lactic bacteria), a lower glycemic index has been measured for the latter product (Maioli et al 2008, Poutanen et al 2009, Stamataki et al 2017).

A comparative analysis was then attempted by mean of a metaproteogenomic approach.

1.4.2 Experimental design

Fecal microbiota was evaluated in young Fisher 344 rats treated both with CR and AL, with commercially available VRF1 (P), VRF1 (P) substituted with 15% of baker's yeast "standard" (CR+S and AL+S), or sourdough "functional" (CR+F and AL+F) "Carasau" bread for 4 weeks. DNA and proteins were extracted from fecal samples, with the aim of carrying out 16S rRNA gene and shotgun metagenomic sequencing and metaproteome analysis, as illustrated in Figure 28.

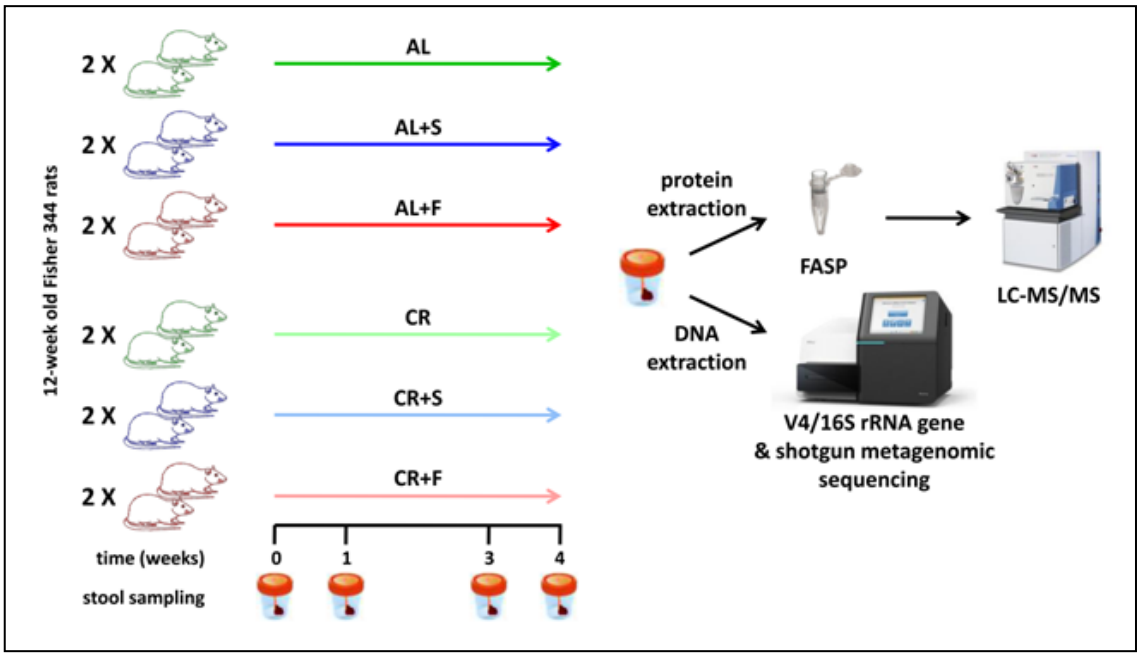

Figure 28: Schematic illustrating the experimental design of the study.

1.4.3 Material & Methods

Animals and samples

A total of 24 Fisher 344 10 weeks old male rats were bought from Charles River Laboratories Italia, SRL (Calco, Italy) together with the manufacturer's animal feed VRF1 (P) 811900 (4.5% of fats). Animals were reared, initially 4 per cage, and fed AL at the Department of Biomedical Sciences, University of Cagliari. After a first week of adaptation to the new habitat, rats were split in 2 per cage and subjected to a further week of adaptation to the meal timetable. Then, 6 experimental groups based on different diet composition were formed (4 rats, 2 per cage): a first groups of animals was fed *ad libitum* (AL), and a second with 70% of the AL ratio (CR); the other 4 groups were fed with either AL or CR ratio and with either "standard" or "functional" Carasau bread (AL+S, AL+F, CR+S, CR+F). Bread was produced by "L'Antico Forno" bakery company (Fonni, NU, Italy) and was added as the 15% of the total feed ratio. The amount of feed for all groups for each week was calculated on the base of AL group food intake. Feed was delivered manually every day at 11AM, whereas the day/night cycle was simulated by lightening the room at 11AM and turning off the lights at 11PM.

Fecal samples were collected before the beginning of the treatment (T0), after 1, 3 and 4 weeks. Feces from two rats (both CR group, one at week 1 and the other at week 4) could not be sampled. All collected fecal samples (n = 94) were immediately stored at -80°C until use. At the time of the analyses, stool samples were thawed at 4°C, and two portions were collected from each of them for protein and DNA extraction, respectively.

All animals were weighted each week, starting from their arrive in laboratory to the end of the treatment, apart from one week before the beginning of the experiment.

DNA extraction and 16S rDNA gene sequencing

94 fecal samples were subjected to DNA extraction, purification and quantification as already described in Materials & Methods of Section 1.2.

16S rRNA gene (V4 region) amplicons and libraries preparation were performed as described in Section 1.3.3.

V4 amplicon sequencing was performed with a MiSeq sequencer using the MiSeq Reagent Kit v3, the paired-end method and 201 cycles of sequencing. Data quality control and analyses were performed using QIIME. The overlapping paired-end reads were merged using the script join_paired_ends.py inside the QIIME package, retaining for further analysis only reads with a length >200 bps. OTUs generation and their taxonomy assignment were carried out with the pipeline previously explained (see Section 1.2.3).

Shotgun metagenomics sequencing and bioinformatics

A total of 15 fecal samples at the final time point (4 weeks) were subjected to whole metagenome sequencing in order to both perform metagenomic analysis and create an "in house" sequence database for the metaproteomic identification. DNA extraction and libraries preparation were already described in Section 1.2.3. Sequencing was performed with the MiSeq sequencer, using also in this case the v3 chemistry, the paired-end method and 201 cycles. For the custom database creation, metagenomic reads were subjected to the same bioinformatics workflow previously described

(Section 1.2.3). For the real metagenomic analysis, the already merged and filtered reads were subjected to a DIAMOND search against the NCBI-nr database (release 2016_09), for the taxonomic assignment, and against the UniProtKB/Swiss-Prot Bacteria database (release 2016_09), for the functional annotation. LCA classification was performed using MEGAN on DIAMOND outputs, with default parameters. For functional analysis protein family, KEGG orthology and pathway were retrieved from the UniProt website using the UniProtKB/Swiss-Prot accession numbers. Moreover, taxonomy and functions were manually combined together on Excel (Microsoft Corporation, Redmond, WA, USA).

Protein extraction and metaproteomic analysis

A total of 24 fecal samples at the final time points were subjected to protein extraction and metaproteomic analysis as already mentioned (see Section 1.2.3), and peptide mixtures were run singularly in LC-MS/MS.

Statistical analysis and graph generation

Differential abundance analysis at aggregated taxonomic level was carried out on counts data between groups using edgeR on the web application MicrobiomeAnalyst (after filtering out few represented features and RLE normalization), whereas analysis within groups on paired data was performed with the R package ibb. *P*-values were corrected for multiple inference using the Benjamini-Hochberg FDR procedure with an adjusted alpha cutoff value of 0.05. Alpha diversity (Shannon index), calculated starting from the OTU table, and richness (observed OTU counts) were compared among samples categories with one-way analysis of variance (ANOVA) following by Bonferroni comparison on all pairs of groups (alpha-value = 0.05) on GraphPad Prism. PCA plots were made at OTU level using the web application ClustVis with default parameters and edited with InkScape. Scatter plots, generated starting from relative abundance data, graphs and box plots were made using GraphPad Prism.

1.4.4 Results

Growth curves and glycemic profiles

Growth curve trends of AL and CR groups started to diverge after the first week of treatment (Figure 29). In CR groups (Figure 29, left) a drop in the body weight (BW) was observed immediately after the treatment, whereas BW remained nearly constant during the successive 3 weeks. It is of note that CR+S and CR+F rat curves were very similar, while CR rats had a lower weight at the beginning of the experiment.

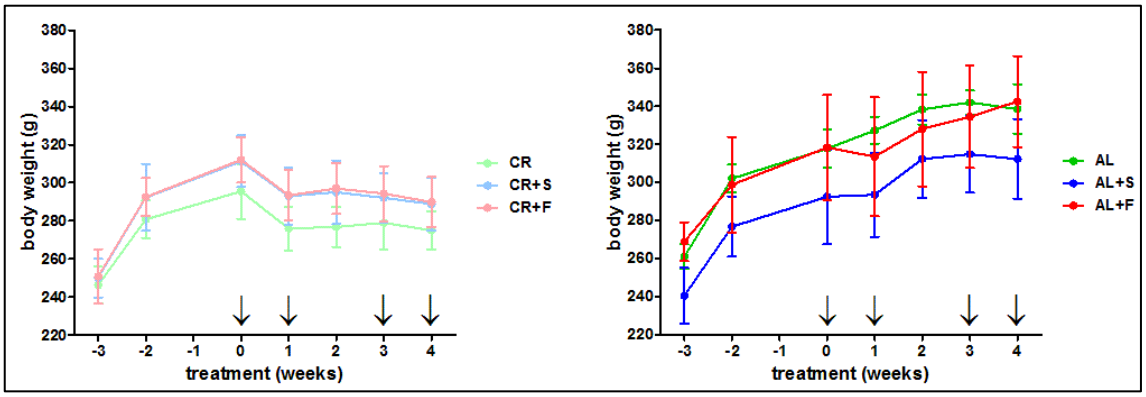

Figure 29: Growth curves of rats following different dietary treatments. Rats were fed *ad libitum* (AL) or with 70% of the AL ratio (CR), with integration of standard bread (CR+S and AL+S) or functional (sourdough) bread (CR+F and AL+F). Arrows indicate the different weeks at which stools were collected.

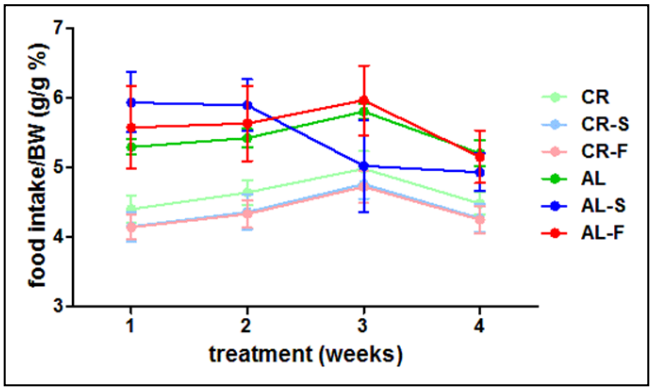

Figure 30: Food intake/body weight ratio curve of rats following different dietary treatments. Rats were fed *ad libitum* (AL) or with 70% of the AL ratio (CR), with integration of standard bread (CR+S and AL+S) or functional (sourdough) bread (CR+F and AL+F).

At week 4, the food intake/body weight ratio (Figure 30) began to be comparable among CR groups and between AL and AL+F rats, with the exception of AL+S-fed rats. This latter result was not unexpected because it was observed that rats following AL+S treatment reduced their food intake since the second week.

16S rDNA sequencing

To investigate the gut microbiota composition in the Fisher 344 rats subjected to the 4 different diet regimens, feces were collected from the 24 young rats at the beginning of dietary intervention (T0) and after 1, 3 and 4 weeks (Figure 28).

Sequencing of the of 16S rDNA V4 region enabled to obtain 3,313,132 reads from a total of 94 fecal samples, after the merging of paired-end reads.

Richness, calculated as the number of OTUs detected within each community, showed no significant difference comparing samples from the six groups of animals after 4 weeks of treatment (Figure 31, right). On the contrary, Shannon index analyses at the same week indicated a statistically significant lower alpha-diversity in AL-fed rats compared to all the other groups (one-way ANOVA/Bonferroni), except for CR-fed rats, that showed equally a low alpha-diversity but without significance (Figure 31, left); this suggested an effect of the bread supplementation in the increase of the alpha-diversity, not so much for enhancing the number of microbial species, as richness was similar among groups, but promoting a more even dispersion within the gut microbial community.

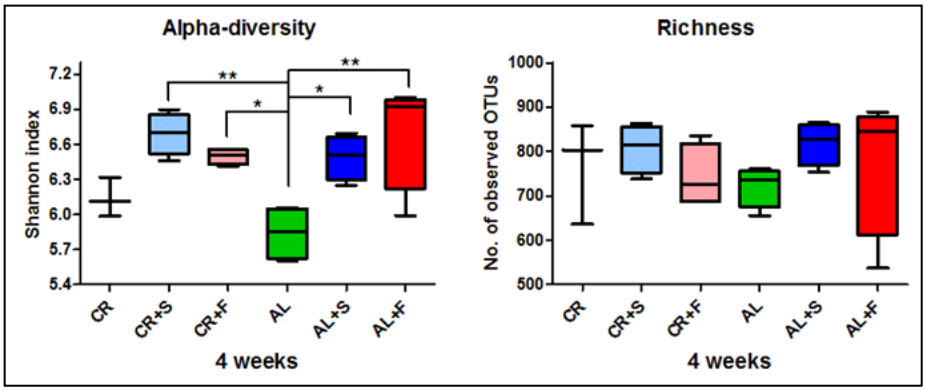

Figure 31: Alpha-diversity and richness within groups after 4 weeks of different dietary treatments. Box plots illustrated Shannon index values and the numbers of detected OTUs within rats fed *ad libitum* (AL) or with 70% of the AL ratio (CR), with integration of standard bread (CR+S and AL+S) or functional (sourdough) bread (CR+F and AL+F). Asterisks indicated statistically significance (one-way ANOVA/Bonferroni alpha-value = 0.05). **P* <0.05, ***P* <0.005.

Gut microbiota taxonomic evaluation after 4 weeks of dietary treatment

Principal component analysis (PCA) was performed in order to evaluate the similarity among microbial communities in rats fed AL and CR, with or without standard or

functional bread addition. As illustrated in Figure 32, there was no evident clustering until 3 weeks, whereas at week 4 there was a clear separation of chow diet rats (CR and AL groups) from the other groups with bread integration, entailing in this case a greater effect of the quality of the diet than the quantity on the gut microbiota composition.

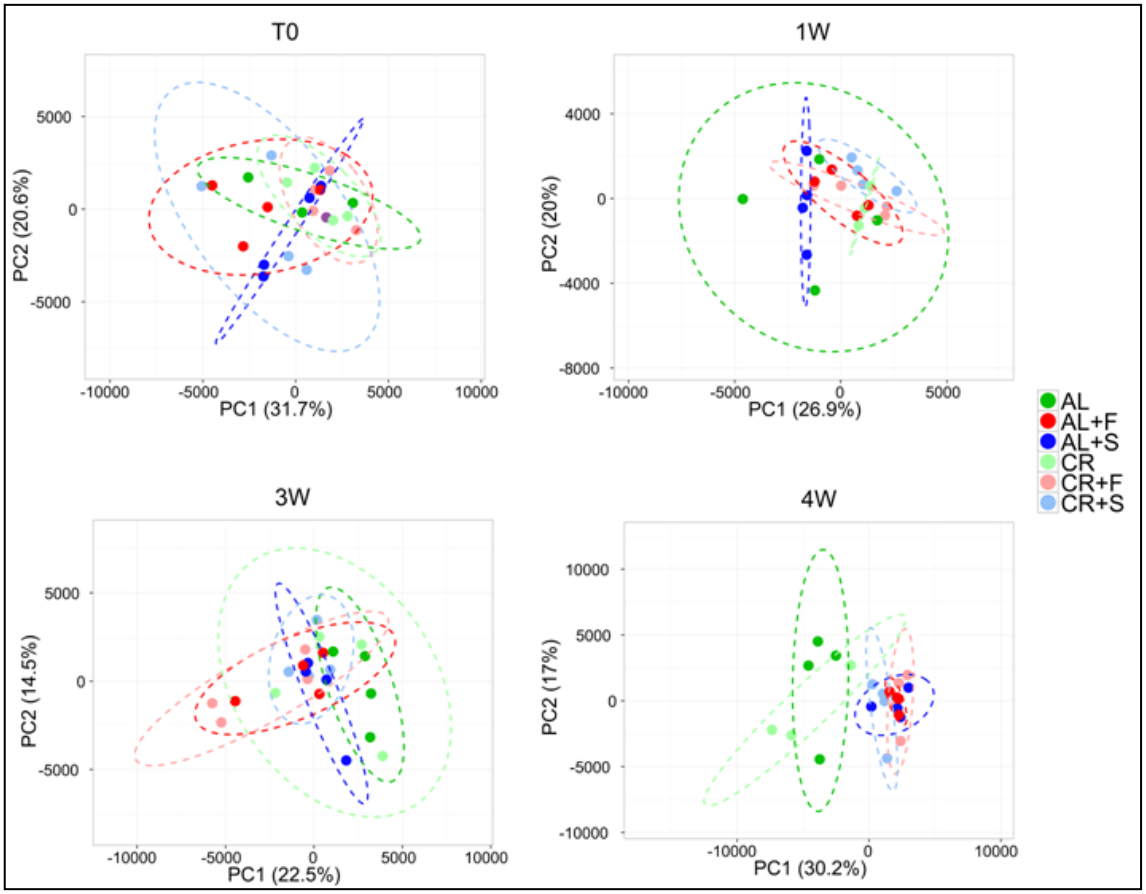

Figure 32: Beta-diversity at OTU level among groups at the beginning (T0), after 1 (1W), 3 (3W) and 4 weeks (4W) of different dietary treatments. Rats were fed *ad libitum* (AL) or with 70% of the AL ratio (CR), with integration of standard bread (CR+S and AL+S) or functional (sourdough) bread (CR+F and AL+F). PCA was carried out at OTU level with the web application ClustVis. Each dot indicate a sample, while dashed ellipses indicate 95% confidence level.

Differential analysis on the relative abundance at family and genus level was performed within each group between the starting (T0) and the final time point (4 weeks) to assess whether different diets induced specific changes during gut microbiota maturation in growing rats (Figure 33).

In particular, rats fed CR+S and CR+F for 4 weeks showed an increase of the relative abundance of genera *Lactobacillus*, *Coprococcus*, that was increased also in AL+S and CR groups, and especially *Bifidobacterium*, whose abundance was very significantly higher (FDR <0.00005) after the end of the treatment and was decreased in AL- and AL+S-fed animals. Finally, Verrucomicrobiaceae family was observed to decrease in CR and CR+S, and to be higher in CR+F, but without significance.

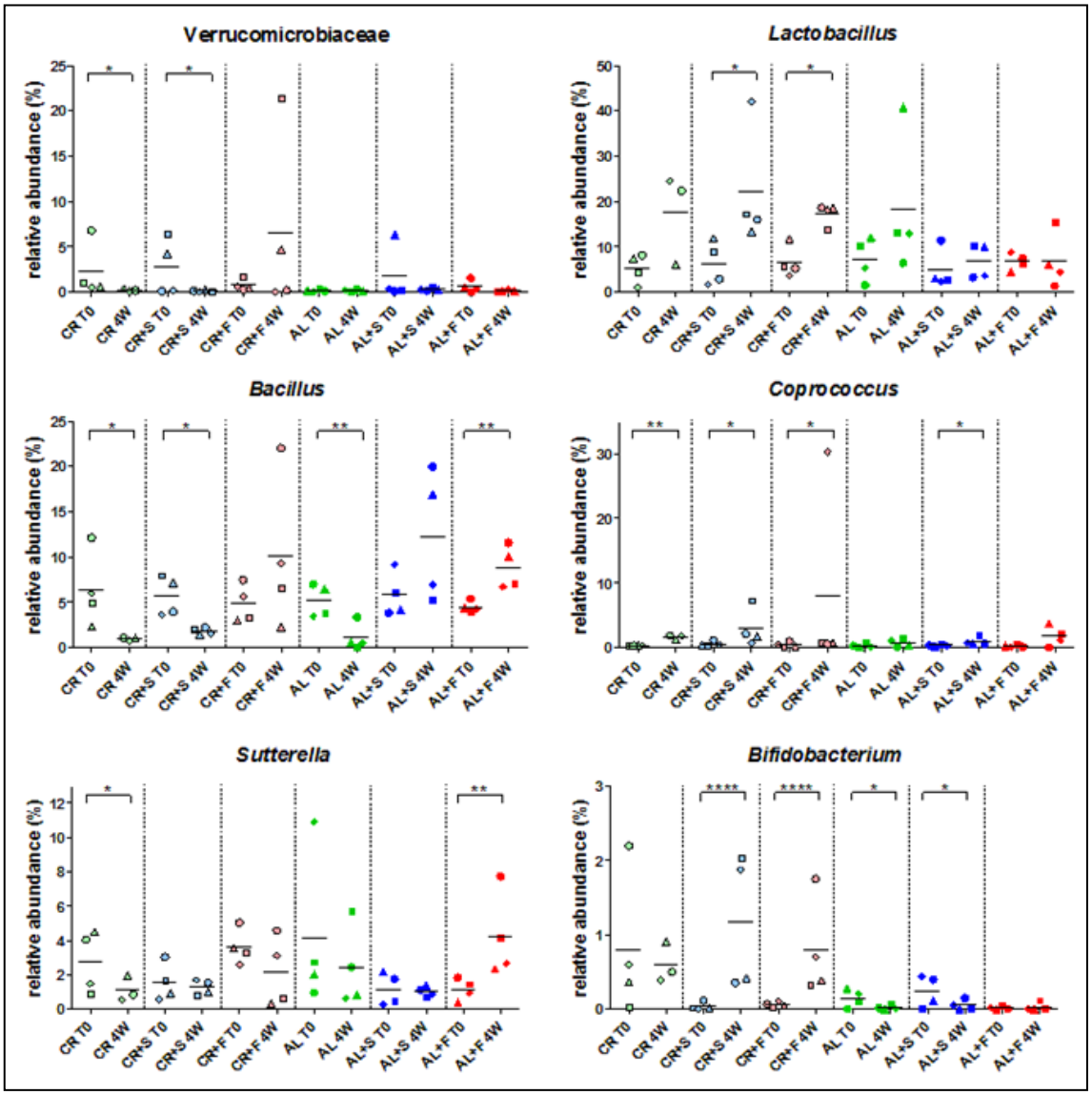

Figure 33: Relative abundance variation of families and genera in rats fecal microbiota after 4 weeks of different diets. Represented families and genera are selected based on their relative abundance (>1% in at least one time point). Rats were fed *ad libitum* (AL) or with 70% of the AL ratio (CR), with integration of standard bread (CR+S and AL+S) or functional (sourdough) bread (CR+F and AL+F). Each dot represents a different sample. Means are also reported. Samples were evaluated at the beginning of the treatment (T0) and after 4 weeks (4W). Asterisks indicate significant difference (ibb performed on paired count data, FDR correction, *<0.05, **<0.005, ****<10^{-5}) within groups between T0 and 4W.

Concerning the taxonomic comparison at the final time point (4 weeks), *Phascolarctobacterium* and *Bifidobacterium* showed significant different distribution among experimental groups. *Phascolarctobacterium*, a propionate producer, was more abundant in all three AL groups and in CR+S than in CR. Hence this firmicutes member could be related to an AL condition, consistently with a study on diet-induced obesity in rats, in which it was significantly associated with fat mass (Wang et al 2017a), and as illustrated in Figure 18, in which *Phascolarctobacterium* was the only genus to be differentially grown after one week of CR reversion into AL, although no significant differences were present between the two groups before the diet switching. In addition, in this experiment *Phascolarctobacterium* growth appears to be rescue in CR diet by the addition of standard bread. A similar combinatorial effect of diet ratio and composition occurred with *Bacillus* relative abundance levels, where a drop was observed in all animals fed only the standard feed VRF1, possibly as part of the gut microbiota maturation in these young rats. However, bread addition to diet was associated to *Bacillus* increase, but with significant values only in combination of AL and functional bread. Similarly, *Sutterella* appeared stable or decreased in all animal groups, but significantly increased when animal received functional bread together with an AL diet. Furthermore, *Bifidobacterium* was detected as more abundant in CR-fed rats than all the three AL-fed groups, and in CR+S and CR+F when compared with the respective group treated AL, although it is noteworthy that significant higher abundance in CR versus AL, CR+F, CR+S was detected already at T0 (Fig. 33). Previous studies reported an increase in the abundance of *Bifidobacterium* OTUs in CR-treated adult rats (Zhang et al 2013a), and an inverse correlation with its relative abundance and body weight, fat mass as well as metabolic endotoxemia and inflammation in mice (Cani et al 2007). Hence, in this experiment, addition of “Carasau” bread seems to accelerate the increase of *Bifidobacterium* associated to CR health-promoting diet.

All taken together, these data strongly suggest that, as expected, the fitness, replication rate and overgrowth of specific bacterial groups might be promoted by feed ratios (CR vs AL) or by feed composition as independent or combinatorial effects.

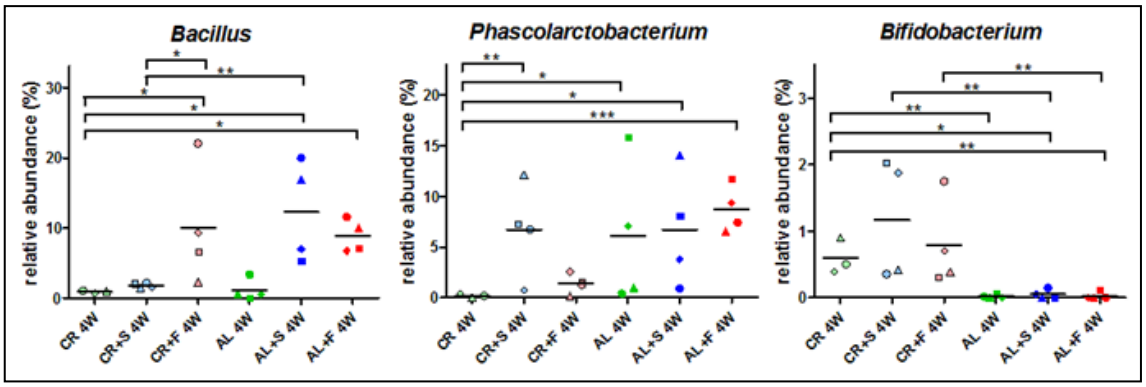

Figure 34: Genera with relative abundance variation among groups in rats fecal microbiota after 4 weeks of different diets. Represented genera are selected based on their relative abundance (>1% in at least one group). Rats were fed *ad libitum* (AL) or with 70% of the AL ratio (CR), with integration of standard bread (CR+S and AL+S) or functional (sourdough) bread (CR+F and AL+F). Each dot represents a different sample. Means are also reported. Asterisks indicate significant difference (edgeR performed on count data, FDR correction, *<0.05, **<0.005, ***<0.0005) between groups.

Finally, in order to inspect whether a sourdough-leavened bread could modulate differently the fecal microbiota, differential abundance analysis at 4 weeks was carried out on CR+S- and CR+F-treated animals. As shown before (Figure 30), the comparison between AL+S and AL+F groups revealed differences in BW and food intake/BW, partly due to an unexplained lack of appetite in AL+S rats; CR+S and CR+F animals were instead perfectly comparable in terms of both BW, growth curve, and food intake, allowing the hypothesis that, in the CR condition, actual differences in gut microbiota taxonomy could be due only to the type of bread leavening.

taxon	taxonomic level	log_2FC (CR+S/CR+F)	FDR
Verrucomicrobia	phylum	-5.40	1.49E-03
Verrucomicrobiae	order	-5.62	2.39E-03
Bacillales	class	-2.50	2.73E-02
Verrucomicrobiales	class	-5.62	1.54E-03
Bacillaceae	family	-2.75	8.67E-03
Verrucomicrobiaceae	family	-5.70	1.24E-03
Mycoplasma	genus	4.95	2.52E-04
Bacillus	genus	-2.98	3.05E-02
5-7N15	genus	-3.75	4.71E-03
Akkermansia	genus	-5.74	4.71E-03

Table 4: Differential taxa between sourdough-leavened and standard-leavened bread supplementation upon caloric restriction treatment on young rats. The base 2 logarithm of CR+S/CR+F fold-change, log_2FC(CR+S/CR+F), and the edgeR P adjusted value, FDR, are reported. Taxa are divided base on taxonomic hierarchy and ordered by decreasing fold-change.

Table 4 lists all the taxa that showed a significant differential abundance between sourdough-leavened and standard-leavened bread supplementation upon CR treatment in young rats. *Bacillus* and *Akkermansia* (Verrucomicrobiaceae) were observed to be higher in CR+F rats at diverse taxonomic levels; in addition *5-7N15*, belonging to Bacteroidaceae family, and *Mycoplasma*, the only taxon with an inverse trend, were detected as significantly distributed. As *Akkermansia* is known for its ability to degrade the intestinal mucin, its increased abundance could be related to a thickening of the gut mucus layer, while the expansion of *Bacillus* spp. is unclear. It is important to note that differential analysis performed at T0 did show no statistically significant diversity at any taxonomic levels, confirming that changes in fecal composition were due only to the type of bread given.

1.5 Gut microbiota profile in a mouse model of celiac disease

1.5.1 Aim of the study

Duodenal and colonic dysbiosis, in addition to the human leukocyte antigen (HLA) genotype, are proven to have a certain role in the pathogenesis of the celiac disease (CD), an immune-mediated disorder affecting the small intestine, caused by the ingestion of wheat gluten in genetically susceptible individuals (Caminero et al 2016, Galipeau et al 2015, Girbovan et al 2017). However, the mechanisms are still unclear, as the HLA genotype influences the gut colonization, while the gut microbiota alteration contributes to the risk of developing CD.

Recently, Mazzarella and colleagues presented a mouse model of celiac disease, in which transgenic mice for HLA-DQ8 developed both the histopathological and the immunological signs of the disease after ten days of intragastric administration of gliadin and simultaneous treatment with indomethacin, an inhibitor of cyclooxygenases (Mazzarella et al 2014).

Here, this CD model was employed to obtain knowledge on gut microbiota balance variations in celiac disease onset and progression, and to evaluate the microbial involvement in reconditioning toward a healthy gut by diet, by the 16S metagenomics of fecal microbiota.

1.5.2 Experimental design

Fecal microbiota was evaluated in HLA-DQ8 mice, before, during and at the end of 10 days of indomethacin treatment plus intragastic administration of gliadin (group G), detoxified gliadin (group D), the wheat fraction containing albumin and globulin (group AG), or the alcohol soluble fraction from quinoa (group Q); additionally, half of the G mice was subjected to a reinstatement of a gluten-free diet after the end of the

treatment described above. DNA was extracted from fecal samples, with the aim of performing 16S rRNA gene sequencing and analysis, as illustrated in Figure 35.

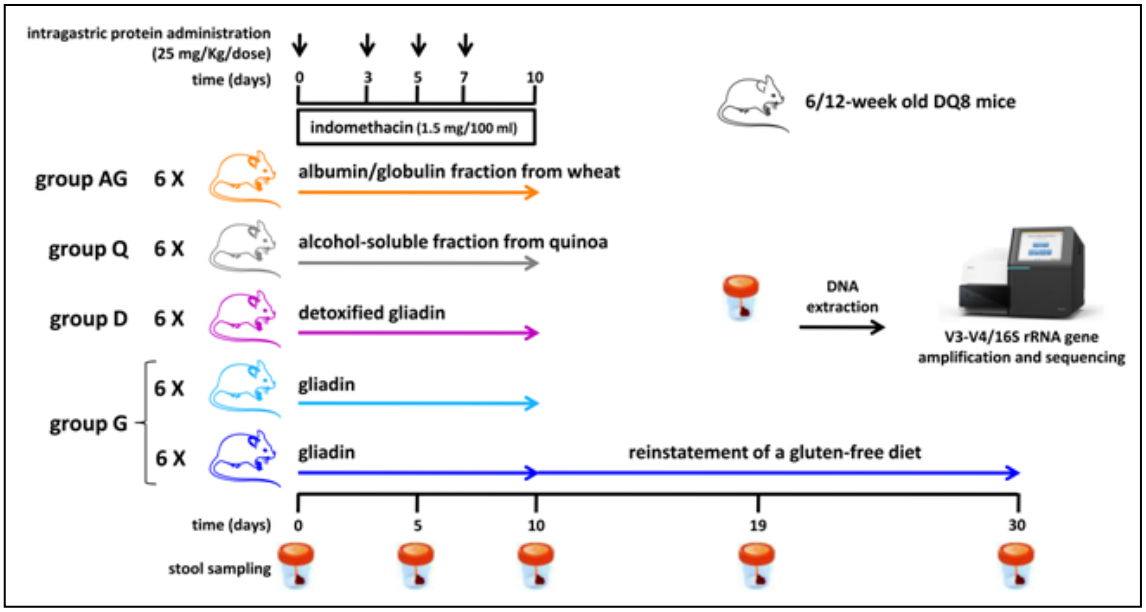

Figure 35: Schematic illustrating the experimental design of the study.

1.5.3 Material & Methods

Animals and samples

A total of 30 mice transgenic for the human leukocyte antigen (HLA)-DQ8, from a colony available at the Institute of Food Sciences of the National Research Council (ISA-CNR) in Avellino (Italy), were maintained in pathogen-free conditions and employed for this experiment.

Animals were reared singularly and fed with a gluten-free diet until the beginning of the experiment. Then, 4 experimental groups based on different administered proteins were formed: the wheat derived proteins gliadin (group G, 12 mice), detoxified gliadin via enzymatic treatment (group D, N = 6), albumin/globulin (group AG, N = 6), and the alcohol soluble fraction from quinoa (group Q, N = 6). Protein fractions were obtained as described in previous studies (Bergamo et al 2011, Mazzarella et al 2014) and intragastrically administered in mice (25 mg/kg doses) on days 0, 3, 5, and 7. At the same time, a subenteropathic dose of the cyclooxygenase inhibitor indomethacin (1.5 mg/100 ml, Sigma-Aldrich) was given in the drinking water and changed every 3-4

days. After 10 days, 6 mice belonging to G group were returned to a gluten-free diet for other 20 days, in absence of indomethacin. Fecal samples were collected before the beginning of the treatment (T0), after 5 (T5) and at the final time point of 10 days (T10). Stools were also sampled after 19 (T19) and 30 days (T30) in those mice which returned to the gluten-free diet. 3 feces could not be sampled (group D T5, group AG T5, group G T30). All collected fecal samples (n = 99) were immediately stored at -80°C and sent to Porto Conte Ricerche in dry ice. At the time of the analyses, stool samples were thawed at 4°C.

DNA extraction and 16S rDNA gene sequencing

99 fecal samples were subjected to DNA extraction and quantification as already described in Materials & Methods of Section 1.2, without the purification step with E.Z.N.A.® Soil DNA Kit.

Libraries were constructed using Illumina's recommendations as implemented in 16S Metagenomic Sequencing Library Preparation guide, using gene-specific primers with the addition of Illumina adapters with overhang nucleotide sequences. (TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG, forward, and GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC, reverse) to amplify the variable regions 3 and 4 (V3-V4) of the 16S rRNA gene (Klindworth et al 2013). V3-V4 amplicon sequencing was performed with a MiSeq sequencer using the MiSeq Reagent Kit v3, the paired-end method and 251 cycles of sequencing. Data quality control and analyses were performed with QIIME. The overlapping paired-end reads were merged using the script join_paired_ends.py inside the QIIME package, retaining for further analysis only reads with a length >200 bps. OTUs generation and their taxonomy assignment were carried out with the pipeline explained before (see Section 1.2.3).

Statistical analysis and graph generation

Differential abundance analysis at OTUs and aggregated taxonomic levels was carried out on count data between groups using edgeR on the web application

MicrobiomeAnalyst (after filtering out few represented features and RLE normalization), whereas analysis within groups on paired data was performed using the R package ibb. *P*-values were corrected for multiple inference using the Benjamini-Hochberg FDR procedure with an adjusted alpha cutoff value of 0.05. Alpha diversity (Shannon index), calculated starting from the OTU table, richness (observed OTU counts), were compared within samples categories with repeated measures ANOVA following by Bonferroni comparison on all pairs of groups (alpha-value = 0.05) on GraphPad Prism. PCA plots were made at OTU level using the web application ClustVis with default parameters and edited with InkScape. Scatter plots were made using GraphPad Prism.

1.5.4 Results

16S rDNA sequencing

Sequencing of the of 16S rDNA (V3-V4 regions) enabled to obtain 8,112,875 reads from a total of 99 fecal samples, after the merging of paired-end reads.

Richness, calculated as the number of OTUs detected within each community, showed a common trend in all groups with a rapid increase after 5 days, significant in AG and G mice, followed by a stabilization at day 10, with statistically significance between T0 and T10 still in AG and G (Figure 36). Interestingly, after gluten-free diet reinstatement, G animals showed a slow decrease of the richness, whose values, however, remained significantly higher than T0. No significant differences among the four groups of animals at T0, T5 and T10 were observed (data not shown).

Concerning alpha-diversity analysis, a trend similar to richness was observed, with a raise of the Shannon index at T5, significant in AG and G, that remained constant at T10 (significant between T0 and T10 still in AG and G) (Figure 37). As for richness, after 9 and 20 days of returning to the gluten-free diet alpha-diversity was decreased, this time with no significance compared to T0 values. Also, in this case, no significant differences among groups were observed at any time points (data not shown).

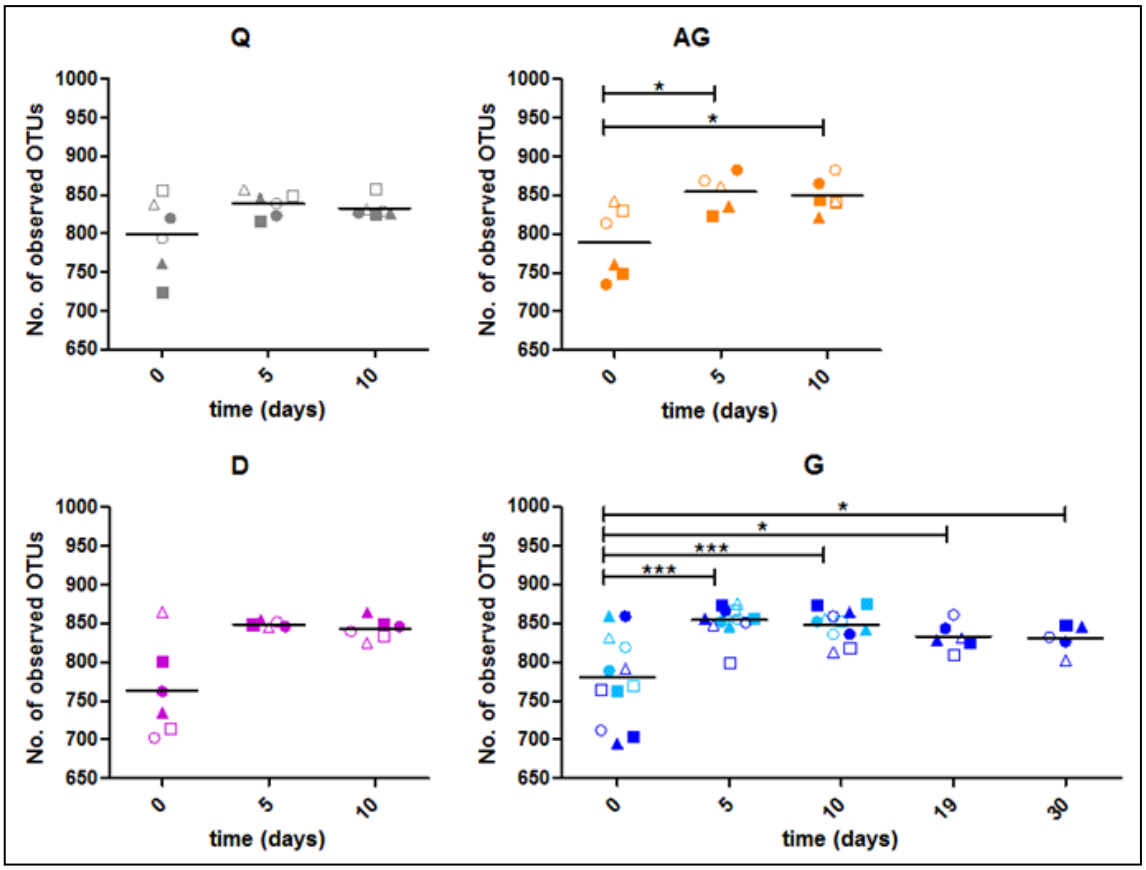

Figure 36: Richness comparison within groups at different time points. Scatter plots illustrated the number and the mean of detected OTUs within mice before (0), during (5) and after 10 days of indomethacin treatment, plus intragastric administration of gliadin (G), detoxified gliadin (D), albumin/globulin (AG), or alcohol soluble fraction from quinoa (Q). Each combination of shape and color represents a different sample. Means are also reported Asterisks indicated statistically significance (repeated measures ANOVA/Bonferroni alpha-value = 0.05). **P* <0.05, ****P* <0.0005.

This data suggested that a shaping of the microbial community occurred in all groups, and could be due to the common effect of the indomethacin, as reported by a recent study in which both richness and alpha-diversity, the latter with statistically significance, were increased in mice after only 6 hours of treatment with indomethacin (10 mg/kg) (Liang et al 2015).

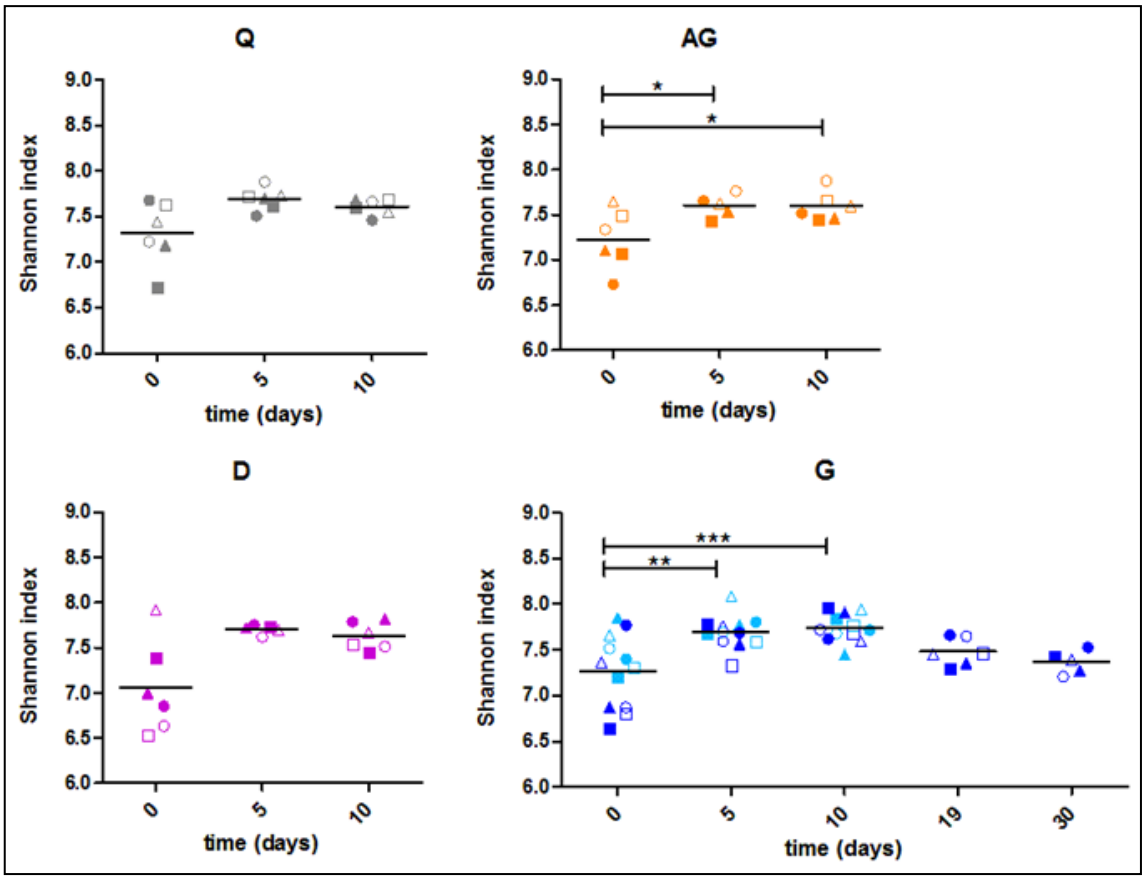

Figure 37: Alpha-diversity within groups at different time points. Scatter plots illustrated the Shannon index within mice before (0), during (5) and after 10 days of indomethacin treatment, plus intragastric administration of gliadin (G), detoxified gliadin (D), albumin/globulin (AG), or alcohol soluble fraction from quinoa (Q). Each combination of shape and color represents a different sample. Means are also reported Asterisks indicated statistically significance (repeated measures ANOVA/Bonferroni alpha-value = 0.05). *P <0.05, **P <0.005, ***P <0.0005.

Regarding the beta-diversity, PCA plots performed at OTU level illustrate no evident clustering among groups at T0 and T5, whereas at day 10 AG and G mice appeared as separated based on the principal components 2 (Figure 38). On the contrary, when PCA was carried out within groups, a clear variation of microbial community was evident during the treatment, as dots corresponding to individuals moved on the principal components 1 from T0 to T10 (Figure 39). In group G, mice that came back to gluten-free feeding were instead progressively closer to T0, indicating that a restoration of gut microbiota was ongoing.

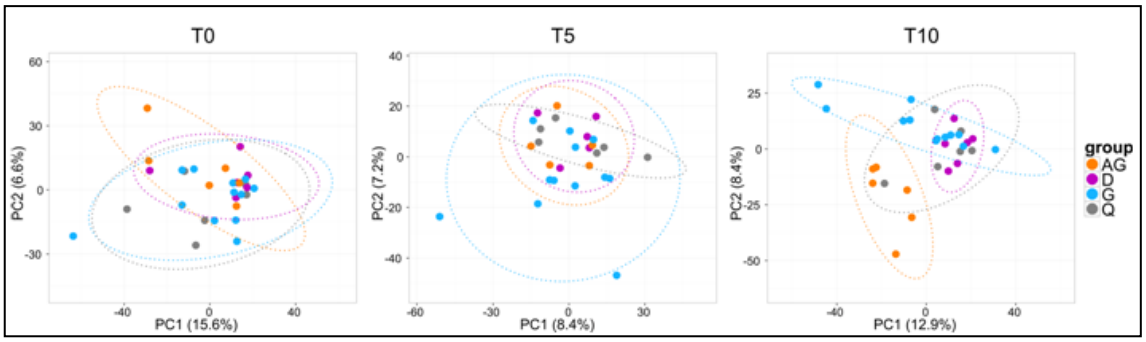

Figure 38: Beta-diversity at OTU level among groups at different time points. Mice before (T0), during (T5) and after 10 days (T10) of indomethacin treatment, plus intragastric administration of gliadin (G), detoxified gliadin (D), albumin/globulin (AG), or alcohol soluble fraction from quinoa (Q) PCA was carried out at OTU level with the web application ClustVis. Each dot indicate a sample, while dotted ellipses indicate 95% confidence level.

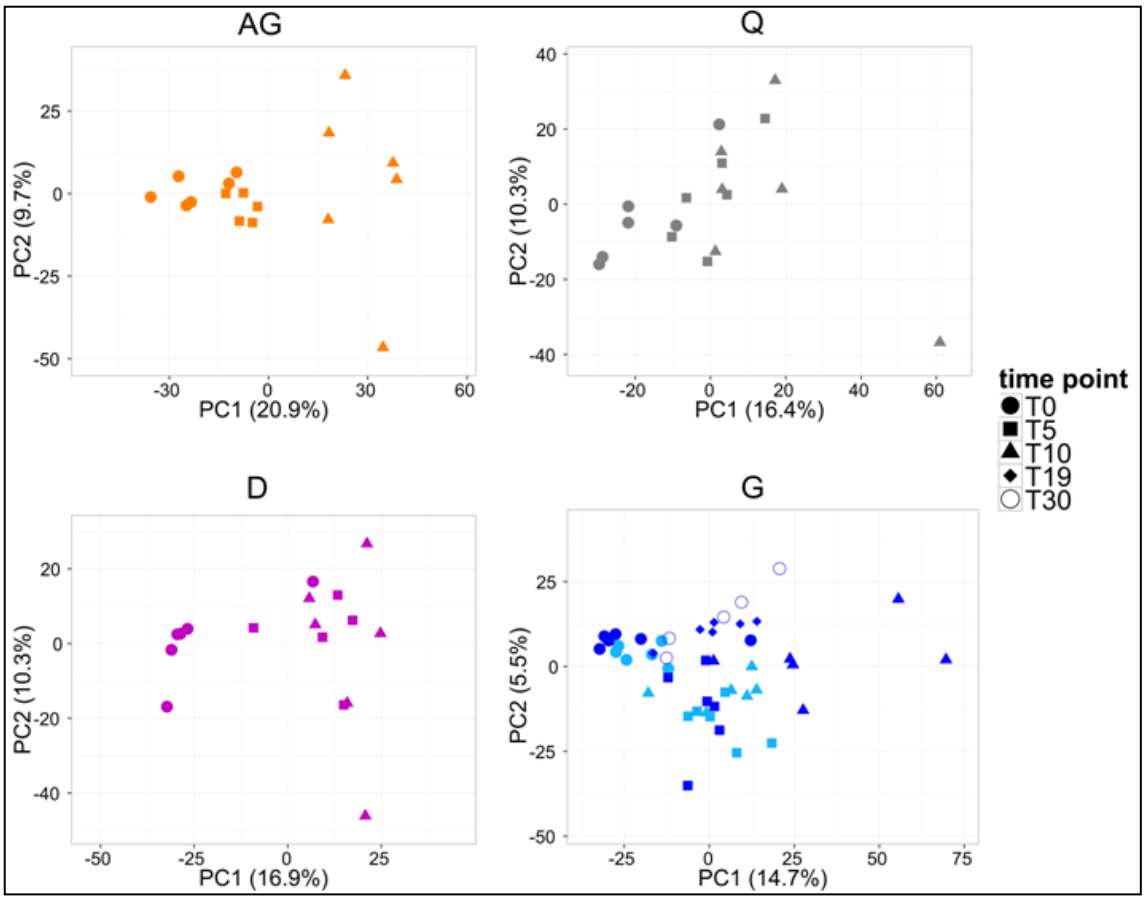

Figure 39: Beta-diversity at OTU level within groups at different time points. Mice before (T0), during (T5) and after 10 days (T10) of indomethacin treatment, plus intragastric administration of gliadin (G), detoxified gliadin (D), albumin/globulin (AG), or alcohol soluble fraction from quinoa (Q) Mice of G groups after 9 (T19) and 20 days (T30) of returning of gluten-free die. PCA was carried out at OTU level with the web application ClustVis. Each dot indicate a sample, while dotted ellipses indicate 95% confidence level.

Taxonomic evaluation among and within groups

Differential abundance analysis was performed both among groups at different time points and within groups, in order to identify taxonomic changes of the mice fecal microbiota occurred during the experiment based on the diverse types of administrated proteins, that triggered the CD onset or did not affect histopathologically the intestinal mucosa.

differential taxa T0	GvsAG		GvsQ		DvsAG		DvsQ		QvsAG	
	log_2fc G/AG	FDR	log_2fc G/Q	FDR	log_2fc D/AG	FDR	log_2fc D/Q	FDR	log_2fc Q/AG	FDR
Cyanobacteria	-1.38	4.44E-03								
Oscillatoriophycideae	-1.60	1.85E-02								
Chroococcales	-1.90	1.04E-02								
Xenococcaceae	-1.89	2.02E-02								
Butyrivibrio			2.96	4.41E-05			1.85	4.53E-02	-1.98	1.64E-02
differential taxa T5	**GvsAG**		**GvsQ**		**DvsAG**		**DvsQ**		**QvsAG**	
	log_2fc G/AG	FDR	log_2fc G/Q	FDR	log_2fc D/AG	FDR	log_2fc D/Q	FDR	log_2fc Q/AG	FDR
Myxococcales	3.22	2.59E-03			4.02	1.53E-05			3.48	1.95E-03
RB41			-1.49	4.80E-02						
OM27	3.52	3.88E-03			4.25	2.80E-05			3.72	2.21E-03
Butyrivibrio			2.64	1.31E-04			2.99	1.77E-04	-2.10	2.45E-03
differential taxa T10	**GvsAG**		**GvsQ**		**DvsAG**		**DvsQ**		**QvsAG**	
	log_2fc G/AG	FDR	log_2fc G/Q	FDR	log_2fc D/AG	FDR	log_2fc D/Q	FDR	log_2fc Q/AG	FDR
Butyrivibrio			2.46	2.00E-04			2.23	3.26E-02		

Table 5: Differential taxa among groups at different time points. Mice before (T0), during (T5) and after 10 days (T10) of indomethacin treatment, plus intragastric administration of gliadin (G), detoxified gliadin (D), albumin/globulin (AG), or alcohol soluble fraction from quinoa (Q). Only taxa with relative abundance >0.5% in at least 1 group are listed. The base 2 logarithm (log_2FC), and the edgeR *P* adjusted value, FDR, are reported for each group comparison. Taxa are ordered based on taxonomic level.

Taxonomic evaluation did not show substantial differences among groups. As reported in Table 5, only very few taxa (with relative abundance of at least 0.5%) showed significant abundance variations in some group comparisons, and most of which were resulted as differential at T0, including *Butyrivibrio* genus, the only taxa with statistical significance at T10. Digging deeper into data, very few differences were observed also at OTU level (Tables 6, 7 and 8). Indeed, only 10, 12 and 19 OTUs were observed as significant different in at least one group comparison at T0, T5 and T0, respectively (OTUs with relative abundance no less than 0.1%), and only 6 (T5) and 12 (T10) with no

significance already at T0. This data could explain why a small number of taxa was detected as differential abundant among groups, since few OTUs were resulted as modified in a certain experimental group, while all the other ones belonging to the same taxonomy did not show the similar behavior.

OTU No.	taxonomy	GvsAG		GvsQ		GvsD		QvsAG	
		log_2fc G/AG	FDR	log_2fc G/Q	FDR	log_2fc G/D	FDR	log_2fc Q/AG	FDR
4421998	f_Lachnospiraceae	5.37	4.47E-02	5.01	1.07E-02	4.97	4.23E-02		
22352	f_Ruminococcaceae							3.87	1.27E-02
267123	*g_Adlercreutzia*							2.83	4.35E-02
538907	*g_Butyrivibrio*			3.60	1.19E-04			-2.52	3.19E-02
298954	*g_Lactobacillus*					6.00	1.54E-02		
47365	*g_Lactobacillus*					5.78	4.83E-02		
173218	o_Clostridiales			5.27	3.47E-04	3.98	1.54E-02	-3.59	3.19E-02
173417	o_Clostridiales			3.61	4.75E-04			-2.73	3.19E-02
260001	o_Clostridiales							4.61	2.30E-03
262179	o_Clostridiales							3.85	1.31E-02

Table 6: Differential OTU among groups at T0. Mice before (T0), during (T5) and after 10 days (T10) of indomethacin treatment, plus intragastric administration of gliadin (G), detoxified gliadin (D), albumin/globulin (AG), or alcohol soluble fraction from quinoa (Q). Only OTUs with relative abundance >0.1% in at least 1 group are listed. The base 2 logarithm (log_2FC), and the edgeR *P* adjusted value, FDR, are reported for each group comparison. OTUs are ordered based on assigned taxonomy.

OTU No.	taxonomy	GvsQ		GvsD		DvsAG		DvsQ		QvsAG	
		log_2fc G/Q	FDR	log_2fc G/D	FDR	log_2fc D/AG	FDR	log_2fc D/Q	FDR	log_2fc Q/AG	FDR
4421998	f_Lachnospiraceae	4.31	2.32E-03	4.75	2.70E-02						
3563387	f_Lachnospiraceae	3.65	3.80E-02					5.12	5.52E-07	-4.62	1.78E-03
509119	f_OM27					4.35	8.20E-03			3.72	2.51E-02
320169	f_S24-7					4.34	1.19E-02			3.74	3.12E-02
538907	*g_Butyrivibrio*	3.69	6.26E-07					3.96	5.66E-06	-3.09	1.98E-04
173218	o_Clostridiales	5.01	6.30E-08					2.75	1.34E-02	-2.58	3.52E-03
173417	o_Clostridiales	3.95	6.30E-08					4.23	2.58E-07	-3.37	2.53E-05

262179	o_Clostridiales	-3.09	8.59E-03					4.08	1.63E-02
260001	o_Clostridiales	-3.28	8.88E-03					5.08	1.12E-03
186464	o_Clostridiales	-1.69	3.18E-02						
3919792	o_Clostridiales	-1.67	4.45E-02						
828708	o_Clostridiales	1.72	4.54E-02						

Table 7: Differential OTU among groups at T5. Mice before (T0), during (T5) and after 10 days (T10) of indomethacin treatment, plus intragastric administration of gliadin (G), detoxified gliadin (D), albumin/globulin (AG), or alcohol soluble fraction from quinoa (Q). Only OTUs with relative abundance >0.1% in at least 1 group are listed. The base 2 logarithm (log_2FC), and the edgeR *P* adjusted value, FDR, are reported for each group comparison. OTUs are ordered based on assigned taxonomy.

OTU No.	taxonomy	GvsAG		GvsQ		GvsD		DvsQ		QvsAG	
		log_2fc G/AG	FDR	log_2fc G/Q	FDR	log_2fc G/D	FDR	log_2fc D/Q	FDR	log_2fc Q/AG	FDR
4421998	f_Lachnospiraceae	4.86	4.00E-03	4.90	8.97E-04	4.82	4.76E-03				
3563387	f_Lachnospiraceae			3.94	3.91E-03			4.68	3.38E-03	-4.32	3.99E-04
22352	f_Ruminococcaceae	-2.88	1.24E-02								
345126	f_S24-7	-1.23	3.94E-02								
309363	f_S24-7			-2.00	2.66E-03						
162639	f_S24-7			-1.85	6.28E-03						
228601	*g_Bacteroides*	-2.55	1.24E-02								
181719	*g_Bacteroides*	-2.53	1.24E-02								
158211	*g_Blautia producta*	-2.44	3.03E-02								
538907	*g_Butyrivibrio*			3.34	2.13E-05			3.13	3.54E-03	-2.21	1.94E-02
4459156	*g_Coprococcus*			4.40	4.11E-03			5.05	3.38E-03	-4.78	3.99E-04
1684221	*g_Desulfovibrio C21_c20*	-1.56	3.03E-02								
173218	o_Clostridiales			4.80	1.97E-05			2.24	4.96E-02	-3.74	3.99E-04
173417	o_Clostridiales			3.27	1.97E-05			3.17	2.36E-03	-2.35	4.06E-03
569581	o_Clostridiales			-5.90	2.41E-05						
260001	o_Clostridiales			-3.58	2.21E-04			-3.96	4.54E-02		
262179	o_Clostridiales			-3.60	3.57E-04			-5.41	3.38E-03		
275895	o_Clostridiales							3.13	3.95E-02	-2.88	1.11E-02
296064	o_RF32			-3.48	2.04E-03	-3.87	1.80E-03				

Table 8: Differential OTU among groups at T10. Mice before (T0), during (T5) and after 10 days (T10) of indomethacin treatment, plus intragastric administration of gliadin (G), detoxified gliadin (D), albumin/globulin (AG), or alcohol soluble fraction from quinoa (Q). Only OTUs with relative abundance >0.1% in at least 1 group are listed. The base 2 logarithm (log_2FC), and the edgeR *P* adjusted value, FDR, are reported for each group comparison. OTUs are ordered based on assigned taxonomy.

Differential abundance analysis within groups showed instead a large number of taxa that modified their amount during the experiment. In Figure 39 are illustrated some selected genera and families exhibiting significant quantity fluctuations. More interestingly, for all these taxa a common trend in all the experimental groups could be described: the abundance of *Akkermansia*, *Desulfovibrio*, *Blautia* and *Adlercreutzia* decreased dramatically after 5 days of treatment, and the same trend, with a softer fall, was observed for the Bacteroidetes family S24-7; *Bacillus* and *Ruminococcus* showed the opposite trend, being significantly more abundant at T5 than at T0, whereas *[Prevotella]*, belonging probably to *Paraprevotella* genus, after a rapid increase from 0 to 5 days, decreased their abundance from T5 to T10. This taxonomic turmoil affected the Firmicutes/Bacteroidetes ratio F/B too, with a consistent increase along the 10 days, principally due to Clostridiales (increased) and S24-7 (decreased). It is important to note that after the gluten-free diet restoration (G mice, 19 and 30 time points), some contrary trends could be observed, more evident regarding F/B, with a significant drop of Firmicutes and increase of S24-7, whereas no changes regarded other genera as *Akkermansia*, *Desulfovibrio*, *Prevotella* and *Blautia*. All these results suggested that a common perturbation of the gut microbiota composition occurred in all groups, independently from the type of administrated proteins and, more important, from the eventual occurrence of the CD signs.

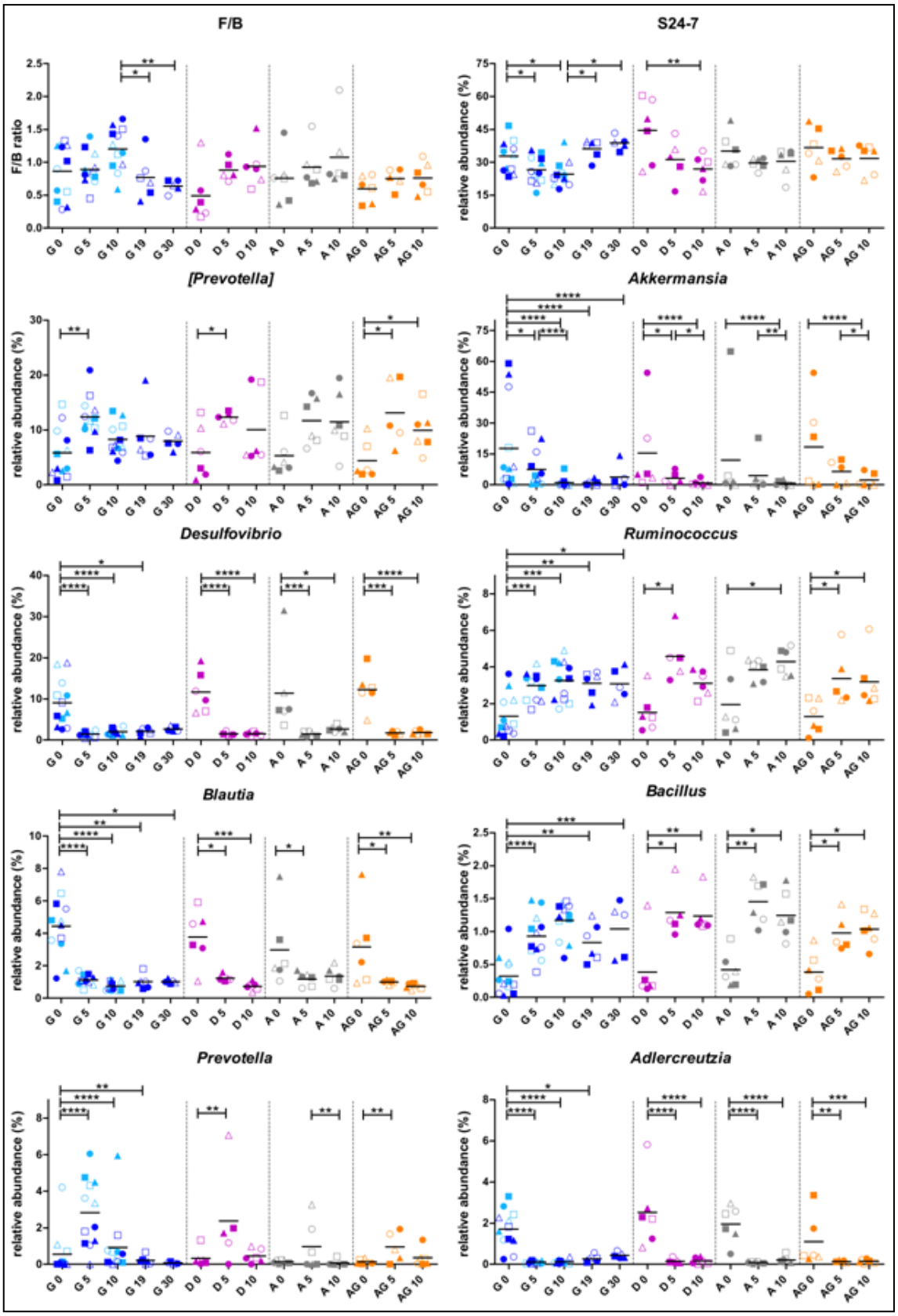

Figure 39: Taxa with relative abundance variation among groups in mice fecal microbiota during the treatment. Represented taxa are selected based on their relative abundance (>1% in at least one group). Mice before (0), during (5) and after 10 days (10) of indomethacin treatment, plus intragastric

administration of gliadin (G), detoxified gliadin (D), albumin/globulin (AG), or alcohol soluble fraction from quinoa (Q). Mice of G groups after 9 (19) and 20 days (30) of returning of gluten-free die Each dot represents a different sample. Means are also reported. Asterisks indicate significant difference (ibb test on paired data performed on counts, FDR correction, *<0.05, **<0.005, ***<0.0005, ****<0.00005) between time points comparison within groups. For F/B ratio comparison, the repeated measures ANOVA followed by Bonferroni comparison between all pairs of groups was performed.

The common denominator in the treatment was the use of indomethacin, an inhibitor of cyclooxygenases, that recently was observed to alter dramatically the composition of the gut microbiota (Liang et al 2015, Xiao et al 2017). Some of the changes observed in this experiment are consistent with the literature, such as the decrease of S24-7 and the increase of F/B ratio and *Ruminococcus* after 6 hours or 2 days of indomethacin administration on mice (10 mg/kg). So the indomethacin appeared to exert a more predominant effect on the gut flora than the gluten or gluten-free intake. Concerning the CD model, it remained unclear why very similar changes were observed in the intestinal microbial composition of all animals, regardless the development of the disease. One possible explanation could be that indomethacin was indispensable to disrupt the microbiota balance, preparing the environment to enteropathy when gluten was concurrently administrated. For instance, *Akkermansia* is known to be an intestinal mucin degrader (Derrien et al 2004), and its drop could be related to the thinning of gut mucus layer during the indomethacin treatment, as well as *Blautia*, that was observed to be reduced in the mucosa-adherent microbiota in patients with colorectal cancer (Chen et al 2012). *Akkermansia* was also associated to a gluten-free diet in NOD mice, on the contrary of *[Prevotella]* (Hansen et al 2014), and with *Adlercreutzia* to a fed fish oil diet in mice, that prevented inflammation than a lard diet (Caesar et al 2015).

Further investigations are needed in order to shed light on this model and clarify the possible implication of the microbiota in the onset and progressive of the celiac disease. The microbiota analyses performed in this study suggest, however, that the cycloxygenase-inhibitor indomethacin triggers a rapid change of the mucosal environment that, in turn, affects the microbiota structure. Changes in host and microbiota functions may enable the onset of CD, hence, when gluten is provide in this mice model diet.

1.6 Conclusion and perspectives

Animal models are a very useful tool to study the gut microbiota interaction with the diet. The reduced and more controlled number of experimental variables, reflecting in a lower inter-individual variability enables, consequently, to employ a relatively small number of individuals and to easily replicate the experiments, compared to human studies.

In this chapter, a rat model of caloric restriction and a mouse model of celiac disease were employed for investigating the implication and the variation of the gut microbiota based on different diet treatment.

Caloric restriction models was used in order to study the effect of different amount of the same food in young, mid-life and adult rats, also after diet reversion, a type of experiment almost impossible to replicate in human. 16S metagenomics allowed the detection of changes in the microbial composition related principally to a blooming of *Lactobacillus* in CR rats and, more in general, with an augmented F/B ratio in AL animals. Metaproteomics not only confirmed the 16S results, but also permitted the identification of specific pathways with differential abundance between groups, regarding in particular the biosynthesis of SCFAs, otherwise difficultly conceivable based on 16S data speculations.

The same model was employed to test the effect of sourdough-leavened bread supplementation to the CR diet. After 16S sequencing several genera were detected as statistically varied after 4 weeks of feeding and on the basis of different diet treatments, while metaproteomics analysis, to date not yet completed on the very same samples, could confirm and/or delve into these data.

The mouse model of celiac disease, on the contrary, did not reveal itself as suitable to study the gut microbiota variation and role during the onset and the progression of the CD. While the CD model was observed to be very efficient from immunological and histopathological point of view, it could present some possible biases in metagenomic studies. 16S results obtained in this study showed that the treatment with indomethacin, used to promote the CD signs in gliadin-treated DQ8-mice, dramatically altered the microbiota balance, although administered with subenteropathic dose and

not affecting the intestine of mice without gliadin administration. Regrettably, the experimental design lacked two further control groups (mice treated only with indomethacin or only with gliadin), in order to subtract independently the effect of these molecules. Nevertheless, potential future metaproteomics analysis on the same sample dataset might give a complete picture of the microbial functional variations, elucidating the microbiota active functions at the onset and at progression of CD.

Further investigations are needed in order to shed light on this model and clarify the possible implication of the microbiota in the onset and progressive fate of the celiac disease. Also according to previous results reported by Liang and colleagues (Liang et al 2015), most of (if not all) the variations occurring in the gut microbiota of our experimental mice depended on indomethacin and were toward an "inflammatory" microbial profile (i.e., increased F/B and lowered *Akkermansia* abundance). It is conceivable that these changes are due to indomethacin-dependent effects on the mucosal immunity and physiology that, in turn, affected the gut microbial ecology. To disentangle the host and microbial effector(s), further experimental studies are required, including the treatment of experimental mice with antibiotics to evaluate the efficacy of the indomethacin-gluten treatment in a microbiota-independent fashion. However, even with this control group, failure to induce CD might still be dependent on the altered metabolism of indomethacin by microbial enzymes. Another possible experimental initiative could be aimed at evaluate the power of feed formulas to keep higher the abundance of specific taxa (i.e. Bacteroidetes, Actinobacteria and Verrucomicrobia) and/or daily administration of fecal samples from untreated animals, to evaluate the possible counteraction of this microbial "prophylaxis" against the onset of CD in indomethacin/gluten treated animals.

Chapter 2:

Integrated metaproteogenomic analyses of the human and the ovine gut microbiomes

2.1 Introduction

As mentioned in Section 1.1, several animal models have been employed to study the intestinal microbiota and its interactions with host, disease, diet and other environmental factors.

Things get more complicated when these aspects are studied in humans, as some experiments in animal models are not translatable due to mainly ethical and logistical issues.

Most of the human gut microbiota studies are focused on the comparison of microbial communities between healthy and diseased patients, in order to identify potential biomarkers associated to the pathology (Cosorich et al 2017, Erickson et al 2012, Kasai et al 2016).

This type of studies usually cannot ascertain whether dysbiotic states are cause or consequence of a given pathology, as it is only possible to monitor the microbial community variations during the various stages of the disease development, except few examples, such as large observational longitudinal studies (Vatanen et al 2016), or studies based on human microbiota transplantation in animal models (Berer et al 2017).

Concerning diet-microbiota interactions, metabolic syndromes, such as obesity (Kong et al 2014, Rebello et al 2015) and type 2 diabetes (Forslund et al 2017, Karlsson et al

2013, Qin et al 2012), have been widely studied, as well as comparisons between populations with very dissimilar dietary habits (De Filippo et al 2010, Lang et al 2014, Schnorr et al 2014, Zimmer et al 2012). Additionally, another significant and rapidly increasing research field regards intervention trials often using probiotics and prebiotics (Hjorth et al 2017, Staudacher et al 2017, Watson et al 2017).

As well documented by the scientific literature (HMP Consortium 2012, Lloyd-Price et al 2016), gut microbial communities present a very high inter-individual variability in humans, also when considering a homogeneous and healthy population, due to differences in lifestyle, diet, environment, genetics or other uncountable factors (Falony et al 2016, Goodrich et al 2014).

As a result of this huge variability, a much larger number of subjects is needed in human studies than in animal models (Falony et al 2016, Pascal et al 2017), even more when 16S rDNA analysis are performed, as taxonomy presents higher variability than gene potentials (HMP Consortium 2012).

Thus, given the huge amount of samples to examine and possible covariates to take into consideration, data analysis demands considerable computational and statistical efforts, and, more often, the need to stratify the population (Falony et al 2015).

Furthermore, also participants recruitment is not trivial, as individuals must be adequately well instructed in order to reduce possible analysis biases due to, for instance, contamination during sampling procedure, or uncorrected/incomplete information regarding drug consumption and diet habits. Finally, repeated fecal samplings can become problematic for the compliance of the subjects and their recall.

Despite all the issues reported above, studies on the human microbial communities present also several advantages. One of the most important is the presence of specific and constantly updated human gut microbiota databases, that allow the identification of a high amount of microbial species and functions, and a wealth of information that can be reinvestigated in the future with more robust analysis pipelines and bioinformatics tools. As an example, microbiota population data could be scrutinized in order to identify and define what composition and, more challenging, functionalities are associated with health; this could become the next goal for the meta-omics field,

as such information will be vital for formulating personalized therapeutic interventions aimed at managing microbiota-mediated health (Greenhalgh et al 2016).

Some of what has been described above for humans is valid also for farm animals, whose microbiotas are becoming a focus of interest for their importance to improve animal health, productivity and nutrition strategies (Deusch et al 2015, Tilocca et al 2017).

In fact, despite studied animals share the same diet and spaces, less in the case of free-range grazing, a high inter-individual variability has been similarly described (Stanley et al 2013, Weimer 2015). Furthermore, close attention must be paid during the samplings to avoid environmental and cross contaminations.

Ruminants are one of the most studied livestocks, in addition to pigs, more often used as an appropriate animal model for human (Yin et al 2017), and chickens. Cattle, sheep and goats represented in 2015 over 3.5 billion of the domesticated ruminants worldwide, constituting a noteworthy source of earnings (Deusch et al 2015).

In the last years, the attention of researchers has been focused mainly on the rumen microbiota and its widest range of biochemical activities, starting from complex carbohydrates digestion to the isomerization and saturation of the dietary unsaturated fatty acids (Hess et al 2011, Huws et al 2011), which are of great interest for the biotechnological industry. Moreover, several studies have regarded the ruminal microbial communities and pathways involved in methane emission, for its considerable impact on the environment (Tapio et al 2017), and changed in response to dietary treatment to improve the animals productivity (Carberry et al 2012, Klevenhusen et al 2017).

Regarding the intestinal microbiota, few data have been collected, mainly on the taxonomic composition and in bovine (Kim et al 2017); very few studies have been reported concerning small ruminants and investigating the microbiota functional activities (Al-Masaudi et al 2017), despite the crucial role of the microbial community in the gut metabolism.

2.2 Characterization of human gut microbiota functions and metabolic pathways of healthy human cohort

2.2.1 Aim of the study

In the latest years, a growing number of studies have investigated the gut microbiota functional potential by shotgun metagenomics (MG) approaches, in addition to the mere taxonomic description that can be obtained through the application of 16S rRNA gene sequencing. However, since sequenced genes are not necessarily expressed, MG cannot provide reliable information about the microbial functional traits that are actually changing in response to different stimuli, such as host metabolism, immunity, neurobiology, diet, or other environmental factors; instead, this type of information can be gathered by functional meta-omics, as metatranscriptomics and metaproteomics (MP).

Nowadays, few pioneering studies have presented the analysis of paired metagenomes and metaproteomes in disease-related human cohorts (Erickson et al 2012, Heintz-Buschart et al 2016), whereas a systematic, comparative investigation of taxonomic and functional features potentially and actually expressed by the gut microbiota of a healthy population has not been described so far.

Here, fecal samples were collected from a cohort of healthy subjects of a clinically monitored Sardinian population and subjected to both shotgun metagenomics (MG) and metaproteomics (MP). MG and MP data were then mined in a comparative fashion in order to: i) identify the actively and consistently expressed functions of the gut microbiota (GM), which are therefore needful for the GM-host homeostasis; ii) describe the conserved and variable GM features within the population; iii) investigate the specific functional and metabolic contribution of specific gut microbial taxa.

The results reported in this section have been recently published on *Microbiome* (Tanca et al 2017a).

2.2.2 Experimental design

Fifteen healthy adult individuals (7 males and 8 females) were selected from a clinically monitored Sardinian population. Fecal samples were collected from each subject and undergone both to shotgun DNA sequencing (metagenome profiling) and shotgun mass spectrometry analysis (metaproteome profiling). The metagenomes were also employed as sequences databases and subjected to taxonomic and functional annotation.

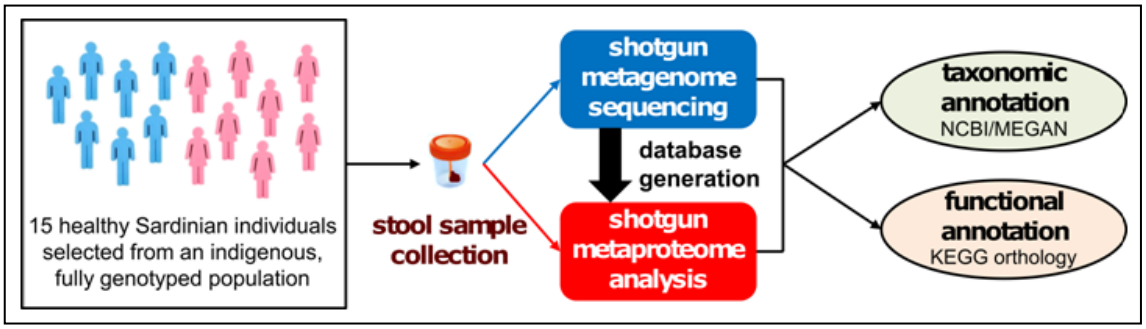

Figure 40: Schematic illustrating the experimental design of the study (Tanca et al 2017a).

2.2.3 Material & Methods

Samples

A stool sample was collected from each of 15 healthy Sardinian volunteer (8 females and 7 males) from a cohort population belonging to the SardiNIA study (Pilia et al 2006). In detail, the SardiNIA is a longitudinal study that investigates genotypic and phenotypic aging-related traits. All participants come from 4 Ogliastra villages (Lanusei, Arzana, Ilbono, and Elini) in Sardinia (Italy). Since November 2001, a total of 6,921 individuals aged 18–102 (>60% of the population eligible for employment in the area) were recruited and the majority of them (N = 6,602) have been assessed for ≈ 13.6 million genetic variants (Sidore et al 2015).

Stool samples were collected from subjects self-reporting the absence of antibiotics treatment during the preceding six months from sample collection, inflammatory

bowel disease and other autoimmune conditions, atypical fluctuation in body weight during the last three years before sample collection, and significant variations of body temperature during the last two weeks. Participants were chosen avoiding gender, age and body mass index (BMI) biases (Table 9), and all followed an omnivorous diet.

All samples were immediately stored at -80°C, next transferred to the Porto Conte Ricerche laboratories in dry ice, and stored again at -80°C until use. Then, samples were thawed at 4°C, and from each of them two equal fecal portions (weighing approximately 250 mg each) were collected, the first subjected to DNA extraction for MG analysis, and the second to protein extraction for MP analysis.

DNA sample preparation and metagenome sequencing

DNA extraction was performed with the QIAamp DNA Stool Mini Kit, while libraries were constructed according to the Nextera XT kit (average insert size ≈ 700 bps) and sequenced with the HiScanSQ sequencer, using the paired-end method and 93 cycles of sequencing.

Metagenome bioinformatics

Paired reads were subjected to merging and filtering with USEARCH (version 8.1.1861) as previously reported (Section 1.2.3). The mean length of the paired-end merged reads was 134 bps. Since sequencing depth may affect estimation of the relative abundances of gene categories, after filtering reads were subjected to random subsampling using the fastx_subsample command (sample_size 200000). A control evaluation of the taxonomic and functional information depth revealed that 96% of taxa and 98% of KEGG functions (relative abundance >0.01%) in the non-subsampled dataset were maintained upon subsampling. Taxonomic annotation was carried out using MEGAN (version 5.11.3) on read sequences undergone to DIAMOND (version 0.7.1) search against the NCBI-nr database (2016/03 update). Functional annotation was instead undertaken through a DIAMOND blastx search (top hit and e-value threshold 10^{-5}) against the UniProt/Swiss-Prot database with (taxonomy Bacteria,

2015/12 update), and then retrieving KEGG orthologous group information associated with each UniProt/Swiss-Prot accession number (UniProt Consortium 2015).

Protein sample preparation and metaproteome bioinformatics

Fecal samples were subjected to protein extraction, LC-MS/MS analysis, and peptide identification as detailed in Section 1.2.3. The sequence database used for peptide identification was composed of the ORFs found using FragGeneScan (version 1.19) starting from the MG reads obtained in this study, upon clustering at 100% using the dedicated USEARCH tool (25,328,860 sequences in total).

All ORFs matched with at least an MS spectrum upon database searching (average length 42 amino acids) were subjected to taxonomic and functional classification, following the same procedure described above for the whole metagenome sequences, except using the DIAMOND blastp command instead of blastx.

Statistical analysis and graph generation

The relative abundance of a taxon/function in a subject was calculated by adding the number of reads (for MG) or peptide-spectrum matches (PSMs) (for MP) assigned to that taxon/function, and then by dividing the count by the total read/PSM count of the subject (so that the sum of the abundances of all taxa/functions detected in each subject is 1). Only taxa and functions with a relative abundance higher than 0.01% were considered for differential analysis.

Bray-Curtis dissimilarity values were computed using the R package 'vegan'. The Wilcoxon signed rank test (R package 'stats') was utilized with continuity correction for Bray-Curtis dissimilarity values comparison between MG and MP. The size of differential abundance of each feature between MG versus MP or Firmicutes versus Bacteroidetes was calculated for each subject and expressed as a relative abundance log ratio, using a correction factor (10^{-5}) in order to remove discontinuity caused by missing values, and the global log ratio was obtained as the mean of the log ratios calculated for each subject. The sets of log ratios were further tested for significant deviation from zero using the one-sample t test, and the nominal two-tailed *P* values

were corrected for multiple testing following the Benjamini-Hochberg method (α = 0.05), using the SGoF+ tool (version 3.8) (Carvajal-Rodriguez and de Una-Alvarez 2011). ClustVis was employed to generate PCA plots and heatmap, boxplots were created using BoxPlotR (http://boxplot.tyerslab.com) (Spitzer et al 2014), bar graph were produced with GraphPad Prism, whereas cladograms were generated using GraPhlAn (Asnicar et al 2015) and then edited using Inkscape. KEGG pathway maps were customized by uploading KO numbers through the 'user data mapping' function on the KEGG website (http://www.kegg.jp).

2.2.4 Results

General metrics

Fecal samples were collected from a total of 15 healthy individuals from a clinically monitored Sardinian population, and analyzed both with metagenomics and metaproteomics approach. Table 9 illustrates the population metrics regarding sex, age and BMI. Studied individuals were equally distributed for gender (8 female and 7 male), with the median age of 39 years (range: 22-48), and their median BMI value of 23.2 (range: 18.4-31.2) at the time of sampling, widely comparable to that of the general Italian population (source: ISTAT 2014).

Sample ID	gender	age	BMI
S1	F	39	26.5
S2	M	39	20.5
S3	M	33	26.5
S4	M	48	29.0
S5	F	45	27.7
S6	F	27	20.7
S7	M	44	27.0
S8	F	44	19.3
S9	F	23	23.2
S10	M	45	24.2
S11	F	48	22.8
S12	F	38	18.4
S13	F	24	20.1
S14	M	24	31.2
S15	M	22	22.0

Table 9: Gender, age and BMI of the human subjects selected for the study (Tanca et al 2017a).

Concerning data analysis, a total of 25,993,645 MG reads and 107,069 PSMs were globally obtained in this study, with a mean of 2,077,370 reads and 7,138 PSMs per sample. Since the wide inter-individual variability in the total number of reads, MG reads were randomly subsampled to allow a better comparison among samples. Taxonomic and functional annotation yields varied between MG and MP (Table 10), with MG data that exhibited a larger depth of information than MP, both in taxonomic and functional terms, as expected and already observed (Erickson et al 2012, Heintz-Buschart et al 2016).

		MG		MP	
		mean	**SD**	**mean**	**SD**
Taxonomy	% of reads/peptides annotated as Bacteria/Archaea	81.31%	4.92%	95.37%	1.10%
	% of reads/peptides annotated as Bacteria/Archaea and further assigned to a specific phylum	96.62%	1.18%	87.43%	1.69%
	% of reads/peptides annotated as Firmicutes and further assigned to a specific genus	50.11%	6.68%	40.20%	7.19%
	% of reads/peptides annotated as Bacteroidetes and further assigned to a specific genus	71.13%	8.30%	59.33%	5.69%
Function	% of reads/peptides assigned to a specific KOG	13.31%	0.98%	39.51%	5.21%

Table 10: Taxonomic and functional annotation yields (Tanca et al 2017a).

Potential and active functions in the gut microbiota

Figure 41 illustrates a much clearer separation between MG and MP patterns based on functional data (C and D) than taxonomic data (A and B), after a preliminary, unsupervised multivariate analysis.

The most represented phyla (A), genera (B) and functions (KEGG orthologous groups, KOGs; panel C) detected by MG and MP are reported in Figure 42.

Concerning taxonomy, the most abundant phyla and genera widely overlapped between MG and MP, on the contrary of functions that presented considerable differences, highlighting a divergence between functional potential and activity. For instance, even if enzymes belonging to catabolic pathways were commonly massively abundant, they were not included among the genes with the highest number of copies in the metagenome.

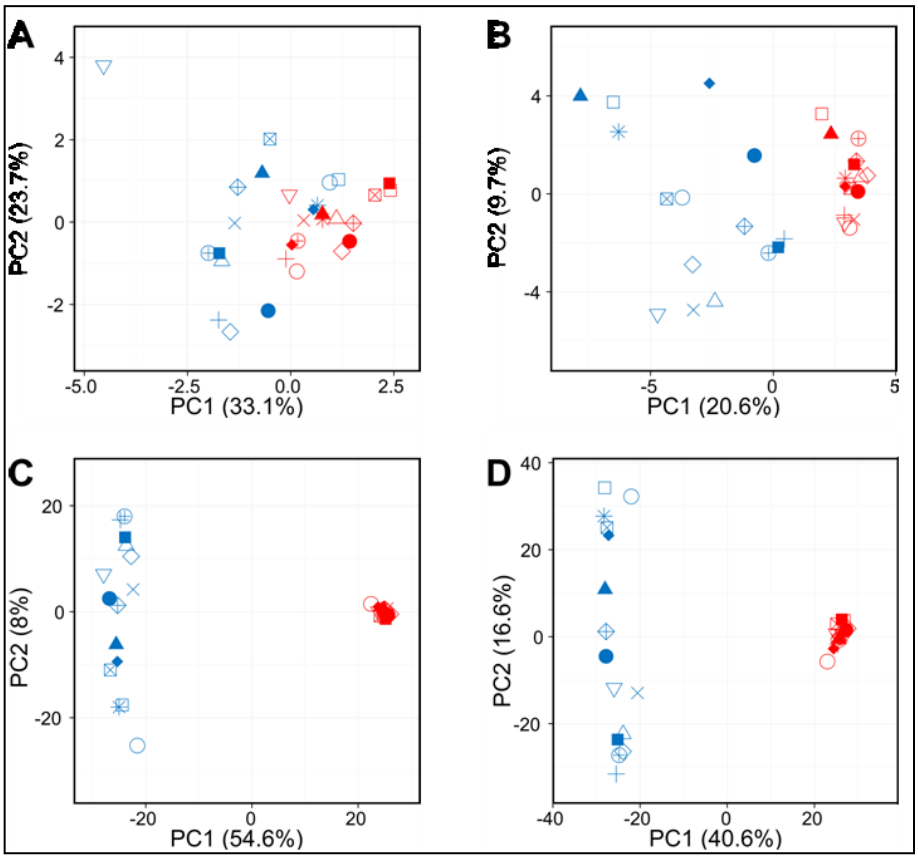

Figure 41: Principal Component Analysis plots related to taxonomic and functional features (Tanca et al 2017a). MG data are in blue, while MP data are in red. Each dot (with different shape) represent a different human subject. (A) phyla; (B) genera; (C) KOGs; (D) KOG-phylum combinations.

The most abundant protein function detected in this study was the glutamate dehydrogenase, consistently with what previously reported by Kolmeder and coworkers (Kolmeder et al 2012, Kolmeder et al 2016). This enzyme plays a pivotal role in the intermediary metabolism in bacteria as well as in animals, providing a chief biosynthetic pathway for glutamate production. Given its high abundance, the impact of the glutamate dehydrogenase might be relevant for colonization and survival of many other taxa that inhabit the human intestine. Moreover, the glutamate circuit has been proposed as key to the neuro-endocrinological role of gut microbiota, the signaling to the central nervous system through the intestinal epithelial cells glutamate receptors, and the activation of the vagal route (Mazzoli and Pessione 2016).

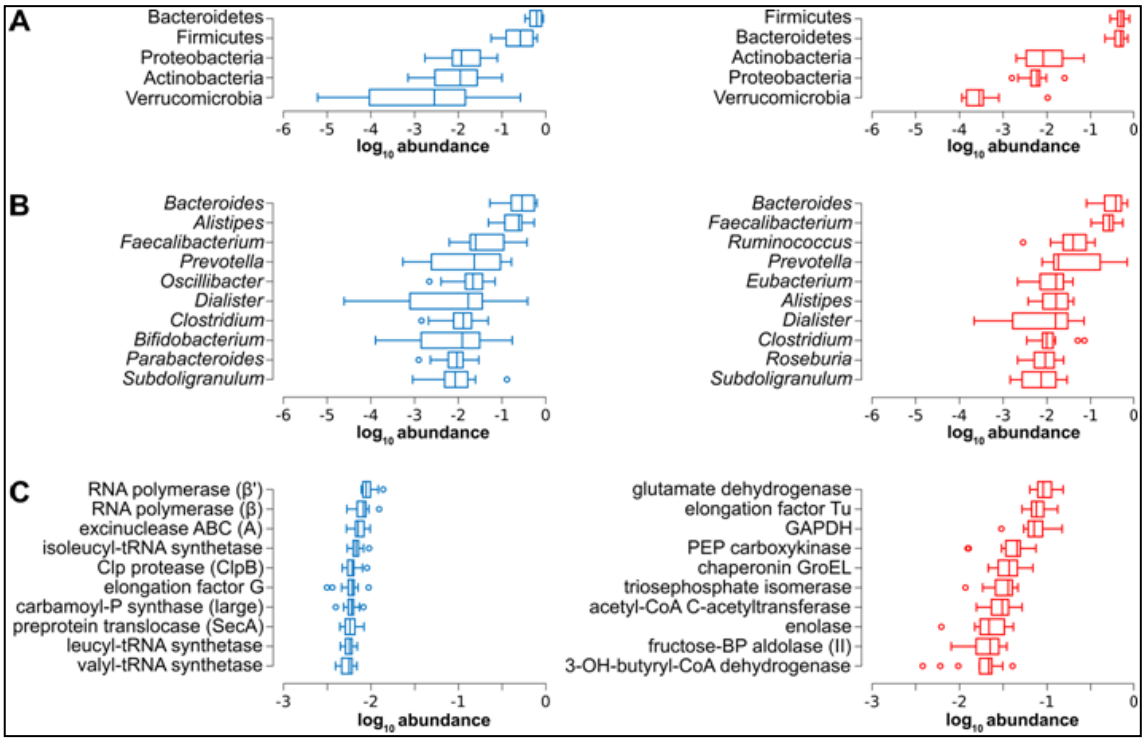

Figure 42: Main metagenome and metaproteome features of the gut microbiota (Tanca et al 2017a). MG data are in blue (left), while MP data are in red (right). Data are ordered by decreasing median of the relative abundance among subjects. Tukey's boxplots showing the top 5 microbial phyla (A), the top 10 microbial genera (B), and the top 10 gene/protein functions (KEGG orthology groups) (C). Subunits name are shown into brackets. GAPDH, glyceraldehyde 3-phosphate dehydrogenase; PEP, phosphoenolpyruvate; P, phosphate; BP, bisphosphate; OH, hydroxy.

Glyceraldehyde 3-phosphate dehydrogenase, phosphoenolpyruvate carboxykinase, acetyl-CoA C-acetyltransferase, enolase, and many other enzymes responsible for essential steps of glycolysis and butyrogenesis were detected also in earlier studies as abundant functions (Kolmeder et al 2016, Verberkmoes et al 2009), supporting the hypothesis that these functions and pathways are essential for the intestinal homeostasis.

Correlation analysis between MG and MP profiles revealed a higher correlation when considering taxa abundances, with a linear decrease when going down to lower taxonomic levels (Spearman's $\rho = 0.90 \pm 0.06$ (mean ± standard deviation) at phylum level, $\rho = 0.68 \pm 0.07$ at genus level), than functional features, that presented a considerably lower correlation ($\rho = 0.21 \pm 0.06$).

Figure 42 provides also an indication on the inter-individual variability (expressed by box width). Dissimilarity (Bray-Curtis index) at the genus and functional level, for both MG and MP data, was computed between-subject to quantify this observation. As

expected, a higher inter-individual variability at taxonomic level could be measured in MG compared to MP (Wilcoxon signed-rank test with continuity correction, two-tailed $P = 4.7 \times 10^{-8}$), whereas functional data analysis revealed a higher variability in MP than in MG ($P < 2.2 \times 10^{-16}$).

MG and MP features were then compared to identify functions consistently expressed by the gut microbes within the healthy human cohort under study, taking into account the gene potential of the same microbiota. To this aim, the log MP/MG abundance ratio was computed for each taxonomic and functional feature for each individual, and then it was tested the difference of the log ratios from zero through the one-sample t test with Benjamini-Hochberg correction, as already described in a comparison between metagenome and metatranscriptome (Franzosa et al 2014).

As a confirm of the divergence between potential and active gut microbiota functions, the percentage of differential functional features was extremely higher than taxonomic features (Table 11).

		>in MP	***nd***	**>in MG**
taxa	*MG dataset*	6%	35%	59%
	MP dataset	9%	55%	36%
KOG	*MG dataset*	6%	5%	89%
	MP dataset	26%	28%	46%

Table 11: Percentage of taxa and functions with differential abundance between the human gut metagenomes and metaproteomes analyzed in this study (Tanca et al 2017a). The extent of differential abundance of each feature between two groups (MG vs MP) was calculated for each individual and expressed as a relative abundance log ratio. The sets of log ratios were further tested for significant deviation from zero using the one-sample t test with FDR correction for multiple testing. nd, not differential.

Focusing on taxonomy (Figure 43, A), many key intestinal microbial taxa showed their relative abundances significantly different in gene potential and expressed proteins comparison across the cohort. For example, Proteobacteria, Spirochaetes, Verrucomicrobia and Coriobacteriales showed a significantly low log MP/MG ratio, Firmicutes and Bacteroidetes taxa behaved, in contrast, more heterogeneously.

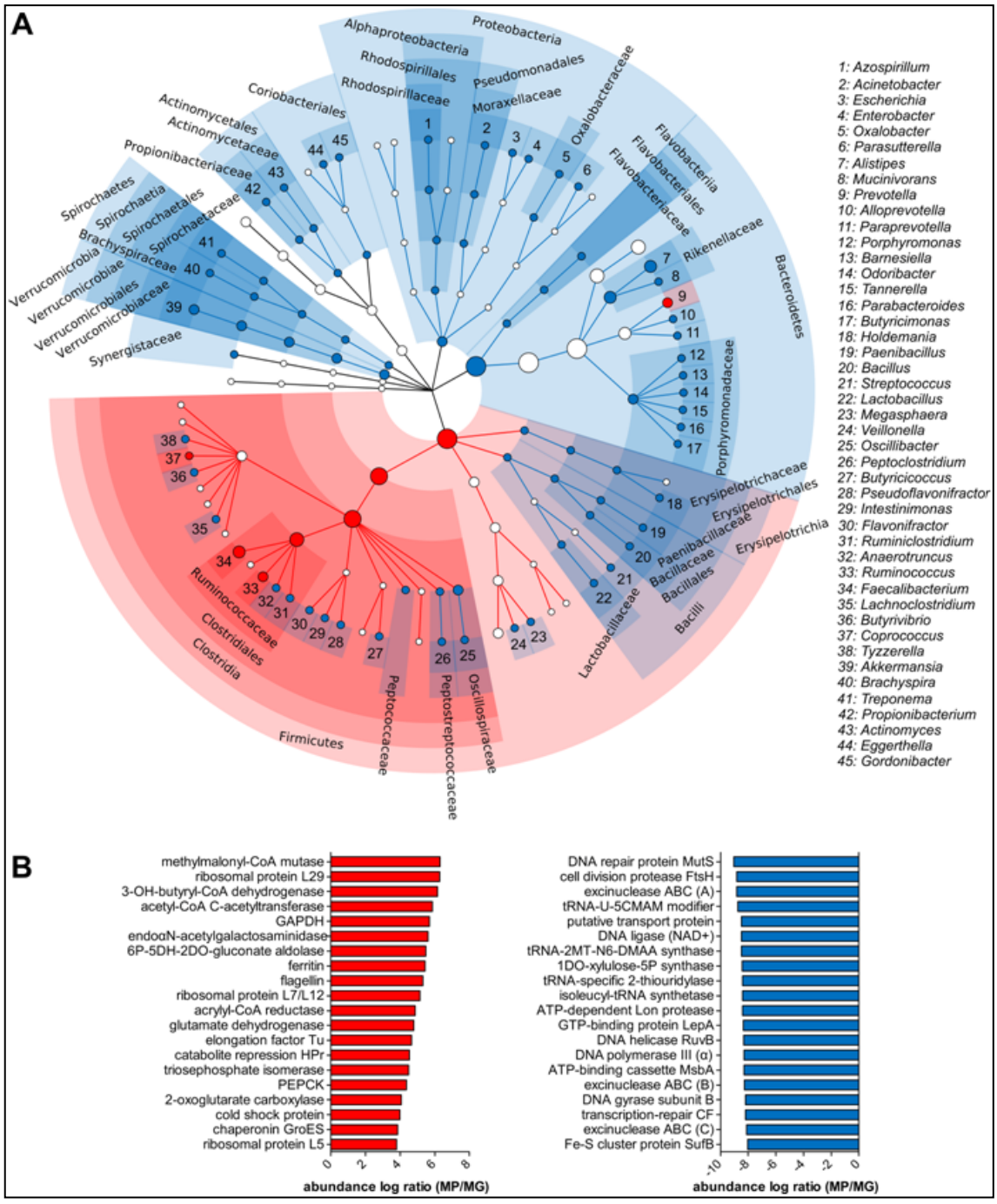

Figure 43: Features with significantly differential abundance between gut metaproteome and metagenome (Tanca et al 2017a). Data were filtered based on the mean relative abundance (>0.01%). of each feature in the sample cohort. (A) Cladogram illustrating differentially abundant taxa (blue, more abundant in MG; red, more abundant in MP). Dot size is proportional to the mean relative abundance of the corresponding taxon. (B) Bar graphs illustrating the KEGG orthology functional groups with higher MP/MG log ratio (top 20, left) and those with lower MP/MG log ratio (top 20, right). Subunits name are shown into brackets. OH, hydroxy; GAPDH, glyceraldehyde 3-phosphate dehydrogenase; 6P-5DH-2DO-gluconate, 6-phospho-5-dehydro-2-deoxy-D-gluconate; PEPCK, phosphoenolpyruvate carboxykinase; U-5CMAM modifier, uridine 5-carboxymethylaminomethyl modification enzyme; 2MT-N6-DMAA, 2-methylthio-N6-dimethylallyladenosine; 1DO-xylulose-5P, 1-deoxy-D-xylulose-5-phosphate; CF, coupling factor.

In detail, a significantly high log MP/MG ratio was observed for *Prevotella*, belonging to Bacteroidetes, and *Faecalibacterium* and *Ruminococcus* (Firmicutes), whereas Rikenellaceae and Porphyromonadaceae (Bacteroidetes), as well as Bacilli and Erysipelotrichia (Firmicutes), displayed a significantly low log MP/MG ratio.

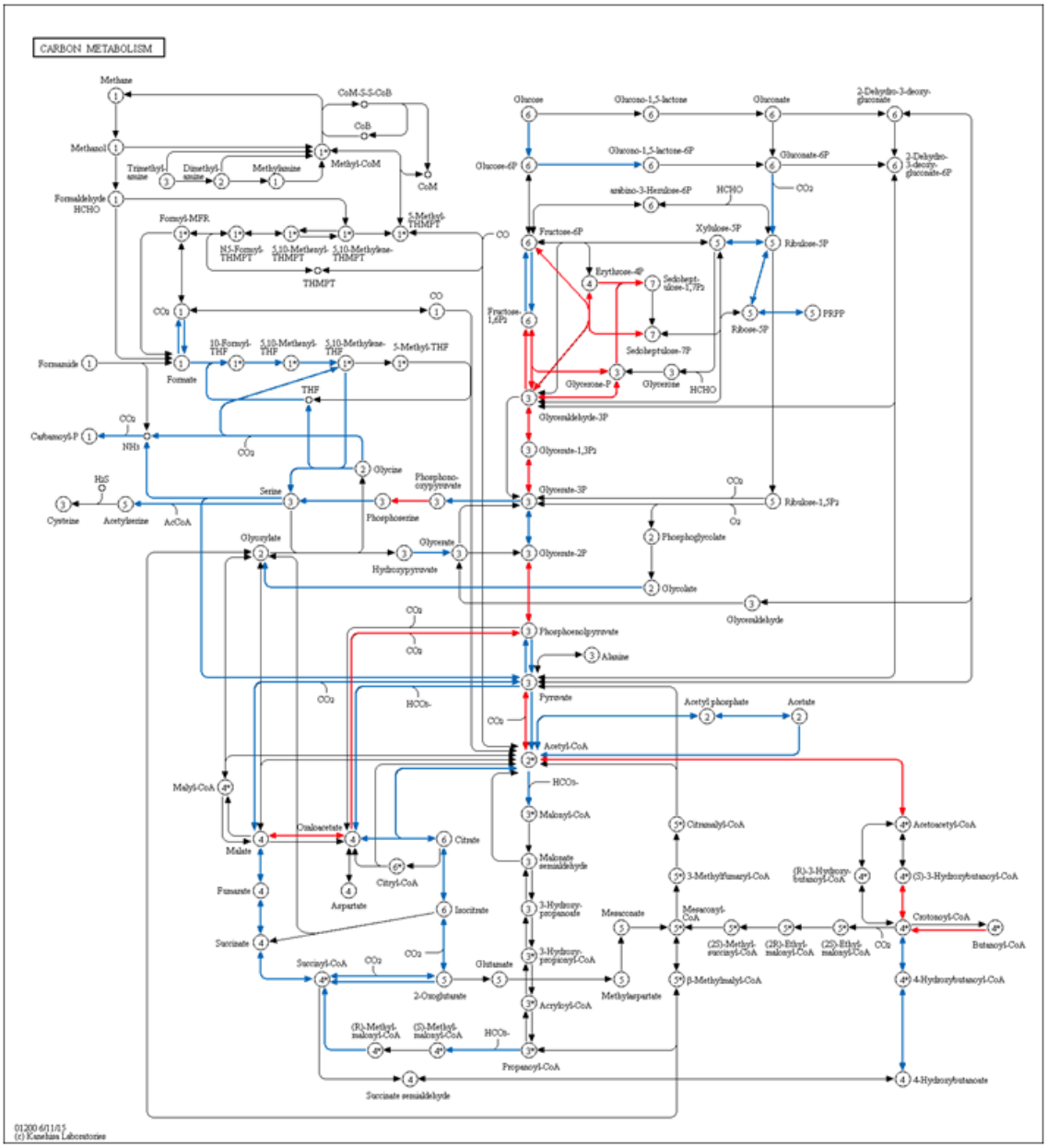

Figure 44: Metabolic functions with differential abundance between MP and MG datasets mapped in the KEGG carbon metabolism pathway (Tanca et al 2017a). Red arrows indicate enzymes with significantly higher abundance in the MP dataset, while blue arrows indicate enzymes with significantly higher abundance in the MG dataset.

Regarding the functional analysis, several enzymatic functions of the studied cohort gut microbiota exhibited a significantly high log MP/MG ratio, as those involved in SCFAs metabolism, including propionate and, mostly, butyrate, as well as in carbohydrates, polyols and organic acids degradation (Figure 43B shows the top differential functions). In addition, also two non-enzymatic proteins, ferritin and flagellin, were included in the 20 functions with the highest MP/MG log ratio.

Carbon metabolism KEGG map (Figure 44) confirms and visually illustrates that the most active metabolic activities performed by the gut microbiota are related to glycolysis, gluconeogenesis, pentose phosphate pathway and biosynthesis of butyrate.

Conversely, functions with the lowest MP/MG log ratios were related to the biosynthesis of amino acid, tRNA and cell wall, as well as to DNA replication and repair.

Conserved and variable features in the gut microbiota

To investigate which taxa and functions were more conserved and variable within the gut microbiota of human cohort under study, the abundance coefficient of variation (CV) was calculated for each taxon and function measured by MG and/or MP across the fifteen individuals. Two arbitrary thresholds (CV >150% and <60%) were set to delineate features with high and low inter-individual variability, respectively.

		number of features		
		CV <60%	**60%< CV <150%**	**CV >150%**
taxa	*MG dataset*	10%	56%	34%
	MP dataset	12%	57%	31%
KOG	*MG dataset*	59%	39%	2%
	MP dataset	19%	49%	32%

Table 12: Percentage distribution of conserved and variable features within the human gut metagenomes and metaproteomes analyzed in this study (Tanca et al 2017a).

The amount of high and low variability features was comparable between MG and MP (about 30% and 10%, respectively); on the other hand, expressed functions (MP) were globally much more variable in abundance within the population than the potential functions (MG), consistently with dissimilarity data (PCA analyses), confirming that the metaproteome displays a higher plasticity, being thus a preferential indicator of

functional changes in the gut microbiota when compared to MG approaches. (Table 12).

Focusing on taxonomy (Figure 45, A), a quite weak correlation could be observed between MG and MP in relation to taxa abundance variability ($\rho = 0.33$); however, and most importantly, no taxa exhibited opposite trends (e.g. high variability with MG and low variability with MP).

The taxonomic lineage from Verrucomicrobia to *Akkermansia* displayed high variability within the subjects consistently with both MG and MP results, as well as, *Bifidobacterium*, *Prevotella* and *Butyrivibrio*, although with small differences in the degree of variability between MG and MP, suggesting a possible higher responsiveness to variables like diet or other environmental factors. Consistently, *Prevotella* abundance has been observed to be modulated by increase in fibers in the diet and to glucose metabolism and tolerance (De Vadder et al 2016, Kovatcheva-Datchary et al 2015), changes in *Akkermansia* levels have been recently related to several different foods and dietary variables (Dao et al 2016, Roopchand et al 2015), whereas many bifidobacteria are widely and long used as probiotics in order to induce/restore gut microbiota homeostasis (O'Callaghan and van Sinderen 2016, Tojo et al 2014).

On the contrary, the taxonomic lineage from Bacteroidetes to *Bacteroides*, as well as *Alistipes* and *Faecalibacterium*, were found to be rather conserved in abundance among the subjects analyzed.

The most conserved and variable functions (KOGs) are reported in Figure 45B. A very weak correlation was observed between MG and MP concerning function abundance variability ($\rho = 0.12$). On the whole, the abundance of genes related to tRNA and peptidoglycan synthesis, besides DNA replication and repair, exhibited low variability among subjects, contrary to some potential activities (including transposases, antibiotic resistance genes and enzymes related to glycans and biogenic amines catabolism) showing a higher variability.

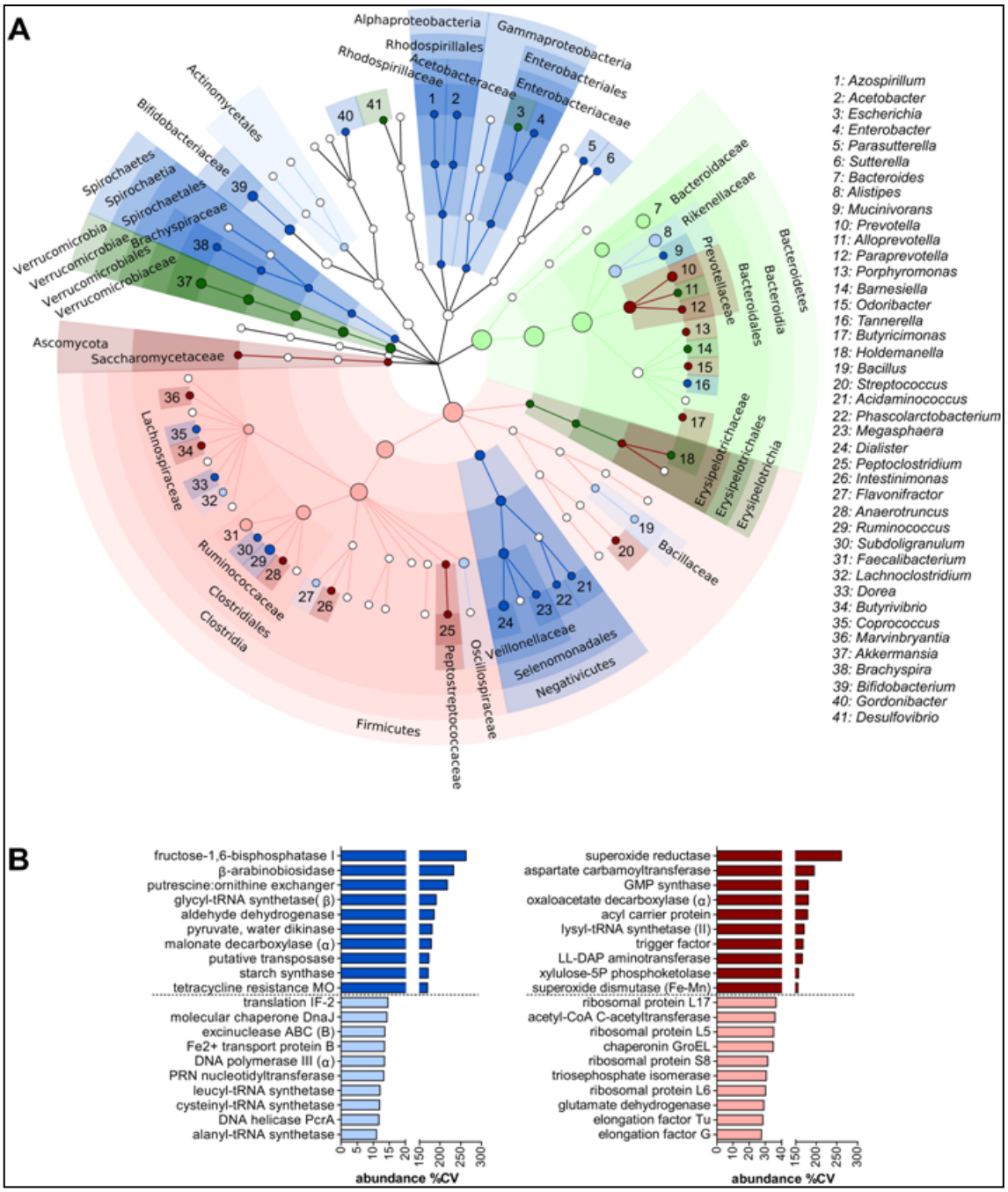

Figure 45: Inter-individual variability of gut microbiota features (Tanca et al 2017a). Data were filtered based on the mean relative abundance of the features in the sample cohort (threshold >0.01%). (A) Cladogram showing taxa with CV >150% (variable, darker color) or <60% (conserved, lighter color) across subjects, according to MG (blue) and MP (red) data. Green dots represent taxa found conserved (dark) or variable (light) based on both MG and MP data. Dot size is proportional to the mean relative abundance of the corresponding taxon. (B) Bar graphs showing the 10 KEGG orthology functional groups with higher CV (variable, darker color) and the 10 with lower CV (conserved, lighter color) across subjects, according to MG (left, blue) and MP (right, red) data. Subunits name are shown into brackets. Only functional groups detected in at least half of the subjects are illustrated. MO, monooxygenase; IF, initiation factor; PRN, polyribonucleotide; GMP, guanosine monophosphate; LL-DAP, L,L-diaminopimelate; 5P, 5-phosphate.

Conversely, functions involved in glutamate degradation and butyrate biosynthesis, as well as 'housekeeping' glycolytic enzymes and translation factors, appeared to be consistently active in all subjects (with high abundance and low variability) based on MP data. Moreover, many stress-related proteins (such as superoxide scavengers and a trigger factor) were found to be interestingly among the most variable gut microbiota features within the population.

Specific functional contribution of Firmicutes and Bacteroidetes

In order to find phylum-specific functions, i.e. activities mainly or exclusively due to one of the major gut microbiota phyla, the log Firmicutes/Bacteroidetes (F/B) abundance ratio was computed for each function on a subject-by-subject basis, as described above for the MG vs MP comparison. Functions with the highest and lowest log F/B ratios, according to MG (left, blue) and MP (right, red) data are shown in Figure 46. Phylum-specific genes within the metagenome of the cohort under study, belonged to very different activities (including sporulation, cell wall biogenesis and ion transport), and mapped to several relevant biosynthetic and degradative pathways.

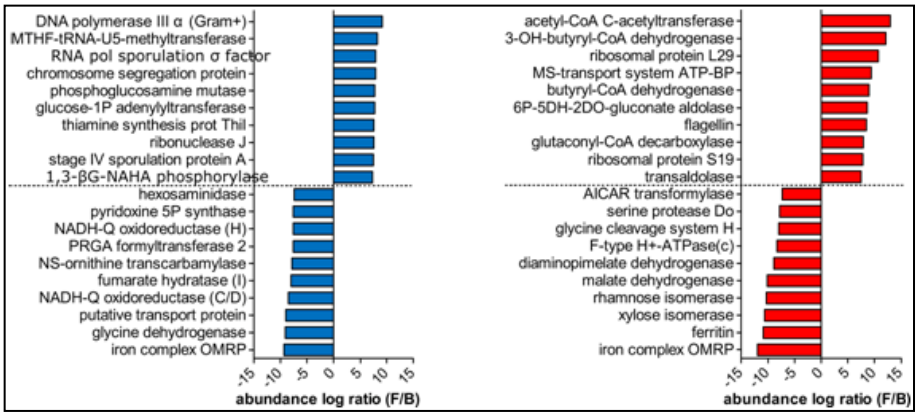

Figure 46: KEGG orthology functional groups with significantly differential abundance between Firmicutes and Bacteroidetes (Tanca et al 2017a). Data were filtered based on the mean relative abundance of the features in the sample cohort (threshold >0.01%). Functions with higher (top 10) and lower (top 10) Firmicutes/Bacteroidetes (F/B) log ratio according to MG data are shown in the left bar graph (blue); functions with higher (top 10) and lower (top 10) F/B log ratio according to MP data are shown in the right bar graph (red). Subunits name are shown into brackets. MTHF-tRNA-U5, methylenetetrahydrofolate--tRNA-(uracil-5-); 1P, 1-phosphate; 1,3-β G-NAHA, 1,3-beta-galactosyl-N-acetylhexosamine; 5P, 5-phosphate; NADH-Q, NADH-quinone; PRGA, phosphoribosylglycinamide; NS-ornithine, N-succinyl-L-ornithine; OMRP, outermembrane receptor protein; OH, hydroxy; MS-transport, multiple sugar transport; ATP-BP, adenosine triphosphate-binding protein; 6P-5DH-2DO-gluconate, 6-phospho-5-dehydro-2-deoxy-D-gluconate; AICAR, 5-aminoimidazole-4-carboxamide ribonucleotide.

egarding the metaproteome, the specific contribution of the two main gut microbiota phyla appeared to be better delineated, and pointed towards more interrelated metabolic activities. Bacteroidetes were found to be specifically involved in multiple activities, such as iron homeostasis, catabolism of non-glucose monosaccharides (rhamnose, xylose) and folate metabolism, while the specific contribution of Firmicutes was mainly in butyrate biosynthesis, being most of these specific enzymes (including acetyl-CoA C-acetyltransferase, 3-hydroxybutyryl-CoA dehydrogenase, butyryl-CoA dehydrogenase, glutaconyl-CoA decarboxylase and enoyl-CoA hydratase) finally converging on butyrate production.

Active role of main gut microbiota members in the carbohydrate metabolism

The specific role of the main gut microbial members within carbohydrate metabolism was further investigated. For this aim, functional and taxonomic annotations of transporters and enzymes identified by MP and involved in processes ranging from polysaccharide degradation to SCFA production were manually parsed. As schematized in Figure 47, complex polysaccharides are generally degraded in the extracellular space, the resultant oligo- and monosaccharides are transported inside the microbial cell, where they are subsequently degraded through carbohydrate catabolic pathways (converging on glycolysis); finally, pyruvate and related intermediates are utilized for the biosynthesis of SCFAs, including acetate, propionate and butyrate.

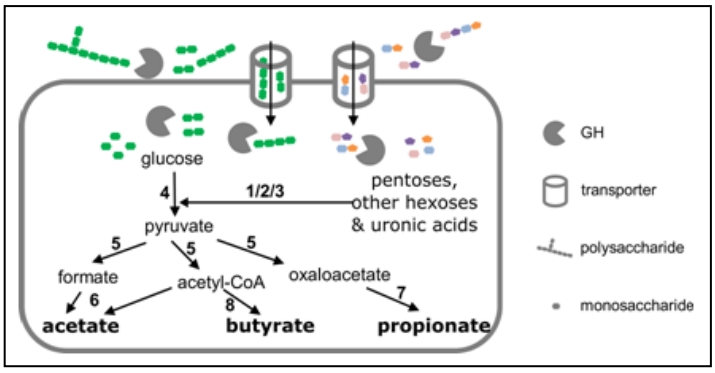

Figure 47: Schematic overview of gut microbiota metabolic pathways from carbohydrate uptake and degradation to the production of SCFAs (Tanca et al 2017a). Number in bold correspond to the metabolic pathways listed in Figure 48. GH, glycosyl hydrolase.

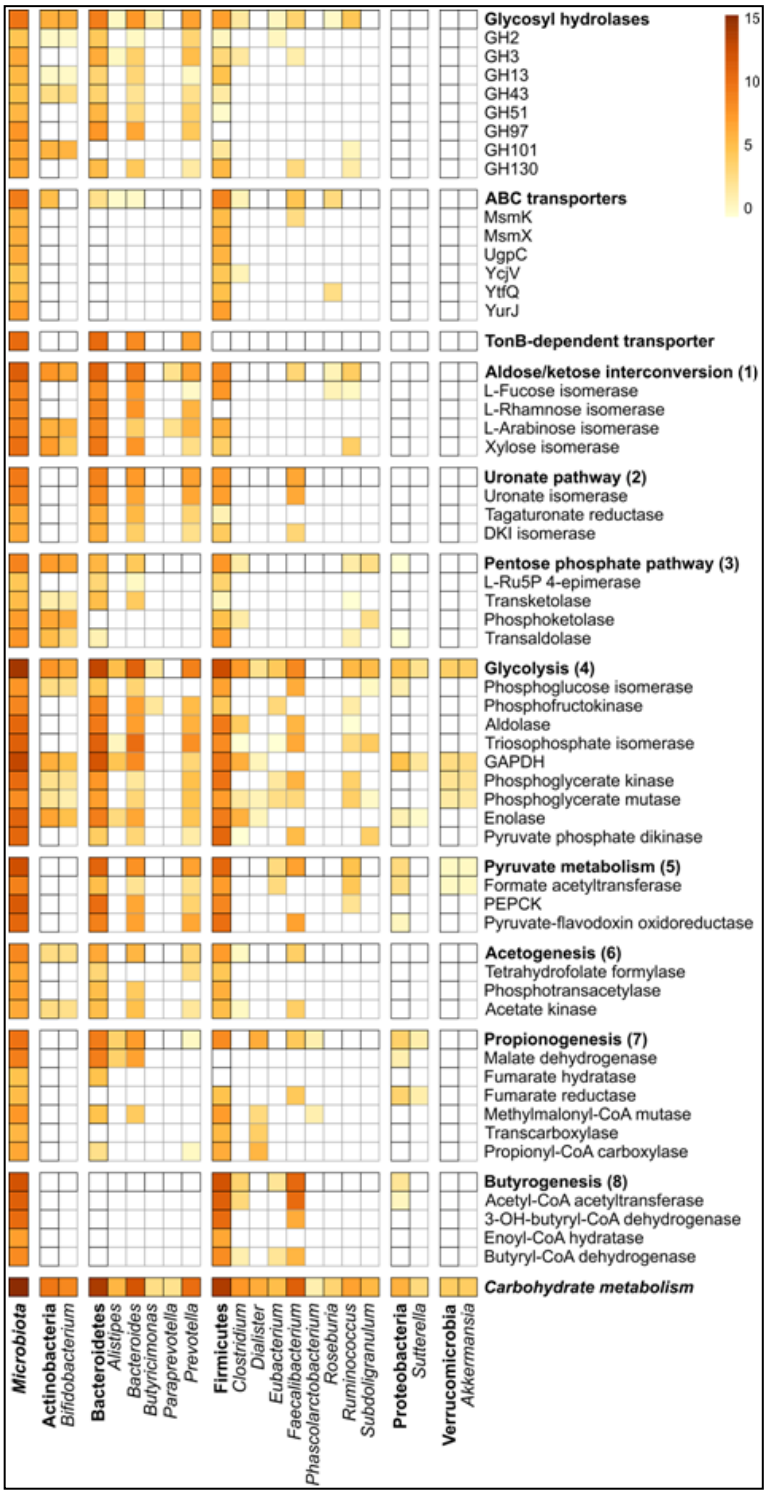

Figure 48: Active carbohydrate metabolism pathways and related taxonomic assignments (Tanca et al 2017a). Combination of carbohydrate metabolism pathways/enzymes (rows) and specific gut microbiota phyla/genera (columns) found by MP analysis. Heatmap color scale is based on the logarithmized relative abundance (average of 15 subjects) of each function-taxon combination. For each pathway (rows), only enzymes detected in at least half of the subjects are reported, while the top row (in bold, corresponding to black-bordered squares) accounts for the total abundance of all enzymes (found in at least one subject) belonging to the pathway. For each phylum (columns), only genera expressing a function in at least two subjects are reported, and the phylum column (in bold, corresponding to black-bordered squares) accounts for the total abundance of all functions assigned to that given phylum.

"Carbohydrate metabolism" and "microbiota" report the total of rows and columns, respectively. GH, glycosyl hydrolase; ABC, ATP-binding cassette; MsmK, multiple sugar-binding transport ATP-binding protein MsmK; MsmX, maltodextrin import ATP-binding protein MsmX; UgpC, sn-glycerol-3-phosphate import ATP-binding protein UgpC; YcjV, uncharacterized ABC transporter ATP-binding protein YcjV; YtfQ, ABC transporter periplasmic-binding protein YtfQ; YurJ, uncharacterized ABC transporter ATP-binding protein YurJ; DKI, 4-deoxy-L-threo-5-hexosulose-uronate ketol; Ru5P, ribose-5-phosphate; GAPDH, glyceraldehyde 3-phosphate dehydrogenase; PEPCK, phosphoenolpyruvate carboxykinase; OH, hydroxy.

Figure 48 shows the expression level of each function-taxonomy combination, with functions grouped according to the reference pathway (or functional family), and microbial genera grouped according to the corresponding phylum. Overall pathway results were retrieved from MP expression data of 81 functional groups; 51 of them, detected in at least half of the subjects, were considered as single functions.

On the whole, the expression of glycolytic enzymes accounted for about half of the total carbohydrate metabolism of the gut microbiota, while the relative contribution of butyrate, propionate and acetate biosynthesis was 12%, 3% and 1%, respectively. This data demonstrated a high and, as described above, constant butyrogenetic activity within a healthy human cohort, in line with previous reports (Kolmeder et al 2016), thus sustaining butyrate production as a pivotal requirement for intestinal health.

Aldose/ketose interconversion (7%) was a further relevant metabolic activity, as well as sugar transporters (comprising TonB-dependent transporters from Bacteroidetes and ABC transporters from Firmicutes) accounted in total for 6% of carbohydrate metabolism-related proteins. Carbohydrate metabolism appeared to be due to Bacteroidetes and Firmicutes almost exclusively (51% and 46% of the total, respectively), with a minor role of Actinobacteria in quantitative terms (3%).

Focusing on the taxa-specific metabolic activities, *Bifidobacterium* spp. was observed to get involved considerably into mucin glycoprotein degradation (endo-α-N-acetylgalactosaminidase activity), pentose hydrolysis (beta-xylosidase), interconversion (pentose isomerases) and catabolism (phosphoketolase and transaldolase belonged to the pentose phosphate pathway). *Bacteroides* provided a peculiar and key role in starch and glucomannan degradation and uptake through the starch utilization system (Sus) enzymes and transporters, and were also shown to play an active contribution in fucose, rhamnose and uronate metabolism, and in the glycolytic pathway (especially in the preparatory phase); *Prevotella*, another

Bacteroidetes member, was instead primarily involved in xyloglucan and arabinan degradation. Among Firmicutes genera, *Faecalibacterium* was found to be an important and almost exclusive involvement in butyrogenesis (as well as in oligosaccharide membrane transport and pyruvate phosphate dikinase activity), while *Dialister* in the final part of propionogenesis; moreover, a considerably high formate C-acetyltransferase activity assigned to *Ruminococcus* genus was detected. Of note, a high level of sequence homology was observed for many orthologous genes within the same phylum, making it difficult to assign taxonomy down to the genus level.

These data provide useful insights into the catabolism and cross-feeding networks actually occurring in a healthy human gut microbiota. Additionally, they suggest that caution should be used before drawing conclusions on the actual gut microbial functional activity based on metagenomic data, supporting MP as a valuable approach to investigate the functional role of the gut microbiota in health and disease.

2.3 Human gut microbiome variations according to their dietary habits

2.3.1 Aim of the study

Once described and thoroughly investigated which were the functional features potentially and actively expressed in a human healthy population (Section 2.2), the focus was shifted on evaluating the effect of the diet on the gut microbiota functionalities.

In view of this, dietary diaries and fecal samples were collected from a cohort of healthy subjects of a clinically monitored Sardinian population (the same as Section 2.2). On the basis of dietary data a subgroup of individuals were selected and the corresponding stools were subjected to both shotgun metagenomics (MG) and metaproteomics (MP). Differential abundance was evaluated on MG and MP data and used to perform statistical comparative analysis between the more distant dietary habits within this healthy population, in order to identity taxa/functions that correlate with diet specific features.

2.3.2 Experimental design

One-week dietary diary was collected from 61 selected healthy adult individuals, in order to compute the caloric intake and macronutrients diet composition for each participant.

Then, fifteen healthy adult individuals (7 males and 8 females) were further selected on the basis of their differences in macronutrients intake. Fecal samples were collected from each subject and undergone both to shotgun DNA sequencing (metagenome profiling) and shotgun mass spectrometry analysis (metaproteome profiling) (Figure 49). Finally, differential abundance analysis was performed both on MG and MP data through statistical comparison between the extremes of the different groups of interest.

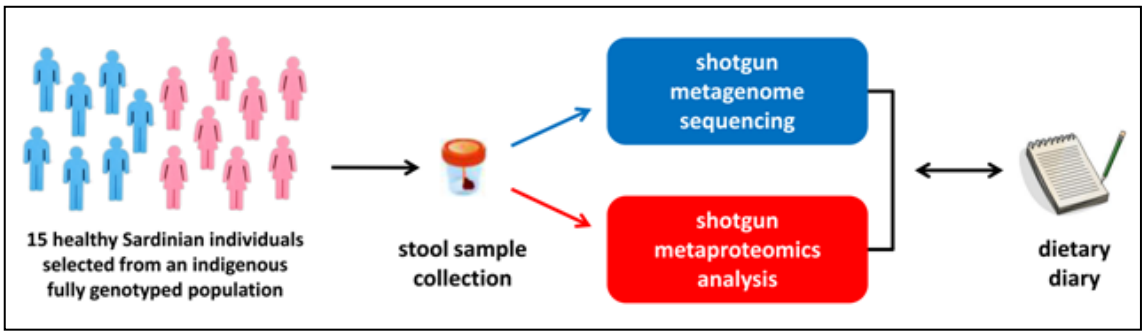

Figure 49: Schematic illustrating the experimental design of the study.

2.3.3 Material & Methods

Dietary diary and sample collection

A one-week dietary diary (Figure 53) was collected from 61 healthy Sardinian volunteers (33 females and 28 males) from a cohort population belonging to the SardiNIA study (Pilia et al 2006) (see Section 2.2.3 for major details). Selected individuals self-reported the absence of antibiotics treatment during the preceding six months from sample collection, inflammatory bowel disease and other autoimmune conditions. Participants were instructed to annotate each food and beverage consumed in each meal, including their quantity, every day for a week, at the end of which a fecal sample was collected from each individual.

Data from dietary diaries were computerized, then the macronutrient composition and the caloric intake were manually computed for each food according to the Council for Agricultural Research and Analysis of Agricultural Economics (CREA) database http://www.crea.gov.it/, or in the case of brand foods by inspecting their nutrition facts labels. Next, the daily average percentage intake of carbohydrate, protein, fat, fiber and alcohol was calculated for each subject.

Stools were collected from all 61 participants and immediately stored at -80°C, next transferred to the Porto Conte Ricerche laboratories in dry ice, and stored again at -80°C until use. Then, samples were thawed at 4°C, and from each of them two equal fecal portions (weighing approximately 250 mg each) were collected, the first subjected to DNA extraction for MG analysis, and the second to protein extraction for MP analysis.

Then, a subgroup of 15 subjects were selected based on high divergence in carbohydrate, protein, fat, fiber and alcohol intake, in order to perform differential analysis between the extremes of each group. Only fecal samples of the 15 selected subjects were then subjected to both DNA and protein extraction.

DNA sample preparation and metagenome sequencing

DNA extraction was performed with the QIAamp DNA Stool Mini Kit, while libraries were constructed according to the Nextera XT kit and sequenced with the MiSeq sequencer, using the MiSeq Reagent Kit v3 with the paired-end method and 201 cycles of sequencing.

Metagenome bioinformatics

Paired reads were subjected to merging and filtering with USEARCH (version 8.1.1861) as previously reported (Section 1.2.3). Taxonomic annotation was carried out using MEGAN (version 6.6.7) on read sequences undergone to DIAMOND (version 0.8.22) search against the NCBI-nr database (2016/09 update). Functional annotation was performed through a DIAMOND blastx search (top hit and e-value threshold 10^{-5}) against the UniProt/Swiss-Prot database with (taxonomy Bacteria, 2016/09 update), and then retrieving protein family group information associated with each UniProt/Swiss-Prot accession number. To implement the "in house" sequence database for the metaproteomic analysis filtered reads were subjected to sorting and sequence clustering, while raw reads were assembled into contigs, with tools and parameters describe in Section 1.2.3. Finally, FragGeneScan was used for open reading frame (ORF) finding both on clustered reads and assembled contigs, with the training for Illumina sequencing reads with about 0.5 % error rate.

Protein sample preparation and metaproteome bioinformatics

Due to the lack of fecal material, it was not possible analyzed with MP one of the 15 samples subjected to MG analysis; in view of this a fecal sample from the most comparable individual on the basis of the parameter analysis was chosen for the

protein extraction. Fecal samples were undergone to protein extraction and peptide mixtures preparation as detailed in Section 1.2.3.

LC-MS/MS analyses were carried out using a Q-Exactive Orbitrap mass spectrometer (Thermo Fisher Scientific), operating with a EASY-spray source, interfaced with an UltiMate 3000 RSLCnano LC system.

Peptide mixtures (4 µg per run) were loaded, concentrated and desalted on a trapping pre-column (Acclaim PepMap C18, 75 µm × 2 cm nanoViper, 3 µm, 100 Å), using 0.2% formic acid at a flow rate of 5 µl/min. Then, peptides were separated with a C18 Easy-spray column (Acclaim PepMap RSLC C18, 75 µm × 50 cm nanoViper, 2 µm, 100 Å) at 35°C with a flow rate of 250 nL/min for 135 min, using the following three-step gradient of eluent B (0.2% formic acid in 95% ACN) in eluent A (0.2% formic acid in 5% ACN): 1-30% for 115 min, 30-60% for 10 min and 60-95% for 10 min. The mass spectrometer was set up in a data dependent MS/MS mode under direct control of the Xcalibur software (version 3.1.66.10), where a full scan spectrum (from 300 to 1700 m/z) was followed by MS/MS spectra. The instrument was operated in positive mode. The temperature of ion transfer capillary, spray voltage, and S-lens RF level were set to 250 °C, 1.85 kV, and 60, respectively. The instrument was tuned and calibrated in positive and negative mode once a week using the calibration solutions, including caffeine, MRFA, and a mixture of fluorinated phosphazines ultramark 1621. The mass spectra were acquired with full MS mode at a resolution of 70,000 within a mass range of 300–1,700 m/z with 1.0 $\times 10^6$ of Automatic Gain Control target and 120 ms of maximum ion injection time. After ion activation/dissociation, the 12 most abundant peaks (Top12 method) were measured with higher energy C-trap dissociation (HCD) at a normalized collision energy of 25%. MS/MS spectra were carried out with resolution of 17,500 with 5.0 $\times 10^5$ of Automatic Gain Control target and 60 ms of maximum ion injection time. The lock mass option was enabled on a protonated polydimethylcyclosiloxane background ion ($(Si(CH_3)_2O)_6$; m/z = 445.120025) as internal recalibration for accurate mass measurements (Olsen et al 2005). The dynamic exclusion was 30 s and the collision gas was nitrogen.

Peptide identification was performed as described in Section 1.2.3. The sequence database used for peptide identification (23,600,215 sequences in total) was composed by human MG reads and contigs obtained in house (comprised those related to this study) along with a publicly deposited catalog of human gut microbial genes (http://meta.genomics.cn/) (Li et al 2014), upon clustering at 100% using VSEARCH (Rognes et al 2016).

All ORFs matched with at least an MS spectrum upon database searching (average length 42 amino acids) were subjected to taxonomic and functional classification, following the same procedure described above for the whole metagenome sequences, except using the DIAMOND blastp command instead of blastx.

Statistical analysis and graph generation

Reads and PSMs counts were uploaded to the web application MicrobiomeAnalyst to evaluate differential abundance through comparative statistical analysis between the extremes of the different groups of interest. Features with prevalence in <10% of samples in each comparison were filtered out, and count data were subjected to RLE transformation prior to statistical testing. Differential abundance analysis was then performed using the edgeR algorithm, with an adjusted *P*-value (FDR) cutoff of 0.05. Principal coordinate analysis (PCoA) were carried out on counts data using MicrobiomeAnalyst based on Bray-Curtis dissimilarity index and beta-diversity was tested with PERMANOVA, whereas the generated PCoA plots were then edited with Inkscape.

2.3.4 Results

General metrics

Fecal samples were collected from a total of 15 healthy individuals from a clinically monitored Sardinian population, and analyzed both with metagenomics and metaproteomics approach. Table 13 illustrates the population metrics regarding sex, age, BMI and diet.

Studied individuals were equally distributed for gender (8 females and 7 males), with the median age of 36 years (range: 25-49), and their median BMI value of 24.3 (range: 18.2-35.1) at the time of sampling. Regarding dietary data, all participants followed an omnivorous diet (no vegetarians and/or vegans were present). Daily caloric intake, measured as the average value of the monitored week intake, was comprised between 1215 and 3279 kcal (median 1621), while the median distribution of the macronutrients in the diet was observed to be 44% of carbohydrates (3% of fibers), 3% of alcohol, 14% of proteins and 35% of fats, whose value is slightly larger than the highest percentage contribution recorded in the typical Mediterranean diet (Willett et al 1995).

It is important to note that these data were computed based on what declared by participants and their good faith in annotating accurately each type and quantity of food eaten for 7 days; so some of these values, such as the caloric intake, could have been overestimated or, more probably, underestimated.

sample ID	gender	age	BMI	caloric intake (kcal)	carbohydrates	alcohol	fats	proteins	fibers	notes
S1	F	31	19.81	1690.8	54.80%	0.00%	33.46%	11.74%	2.99%	
S2	M	36	20.48	3279.7	45.67%	8.31%	32.59%	13.43%	2.40%	
S3	F	48	30.78	1420.4	44.03%	6.56%	35.13%	14.28%	5.10%	
S4	F	42	24.56	1525.4	43.94%	0.00%	42.22%	13.83%	3.55%	
S5	M	34	26.45	1623.0	48.38%	6.59%	34.10%	10.92%	2.56%	
S6	M	48	23.12	1215.5	40.96%	4.56%	38.58%	15.90%	2.18%	
S7	F	36	19.96	1573.3	51.01%	0.00%	34.48%	14.51%	5.20%	
S8	M	49	29.02	2084.2	45.83%	3.39%	38.25%	12.54%	2.51%	
S9	F	49	18.18	1303.2	39.33%	0.00%	42.79%	17.88%	3.68%	only MG
S10	F	37	26.56	1226.2	42.57%	0.00%	35.38%	22.05%	5.66%	
S11	F	30	35.11	1679.5	53.37%	1.15%	34.15%	11.33%	3.52%	
S12	M	33	21.16	2429.8	50.90%	3.30%	32.58%	13.21%	2.42%	
S13	M	27	30.78	1610.2	37.36%	19.52%	31.38%	11.74%	3.15%	
S14	M	25	21.51	1619.3	26.12%	9.05%	48.05%	16.77%	2.14%	
S15	F	48	24.35	1750.2	35.69%	2.19%	42.38%	19.74%	4.81%	
S16	F	34	20.83	1920.2	39.07%	0.00%	46.31%	14.62%	3.06%	only MP

Table 13: Metrics and macronutrients intake of the human subjects selected for the study. Data concerning the caloric and macronutrients intake were collected from one-week dietary diary. Daily caloric intake was computed as the average of the week, while carbohydrates, alcohol, fats, proteins and fibers intake are expressed as their relative percentage contribution to the caloric intake.

Concerning data analysis, a total of 11,060,841 MG reads and 170,472 PSMs were globally obtained in this study, with a mean of 691,302 reads and 11,364 PSMs per sample.

Differential abundance analysis of taxa and functions in the gut microbiota of a healthy population based on the diet

Figures 50, 51 and 52 illustrate the beta-diversity among samples based on their gender, age and BMI, respectively, at taxonomic, functional and taxonomy/function combined level. The population under study did not appear to separate for both gender and sex, whereas a separation based on age classes could observed in MP (PERMANOVA $P = 0.019$ at order level, protein families $P = 0.02$, protein families-order $P = 0.001$).

Differential abundance analysis was carried out at taxonomic (order and genus), functional (protein families) and taxonomy/function combined level (protein families-order) on the basis of diverse population metrics and dietary habits. Concerning caloric and alcohol intake, samples to compare were separated according to the gender; important differences were indeed noted in the two sexes, for instance all males have consumed alcohol, while more than half of females have not drunk it. Finally, fiber intake was very similar among males, on the contrary of females, so the differential analysis was performed only within the latter.

In Table 14 are listed all the statistical comparative analyses performed, including the numbers of individuals, the tested groups, and the features with significant different abundance for each comparison. On the whole, few differential features were observed at taxonomic level, more at functional and combined levels, with the majority regarding MP data. An exception was then high number (N = 80) of the metagenome functions detected as diversely abundant in the fiber intake statistical analysis. Lastly, no or very few taxonomic and functional features were detected as differential after performing the analyses according to diversity in BMI, gender, and carbohydrates and proteins intake.

Focusing on taxonomy (Table 15), several features were detected as differential when compared fecal microbiota based on the subject's age.

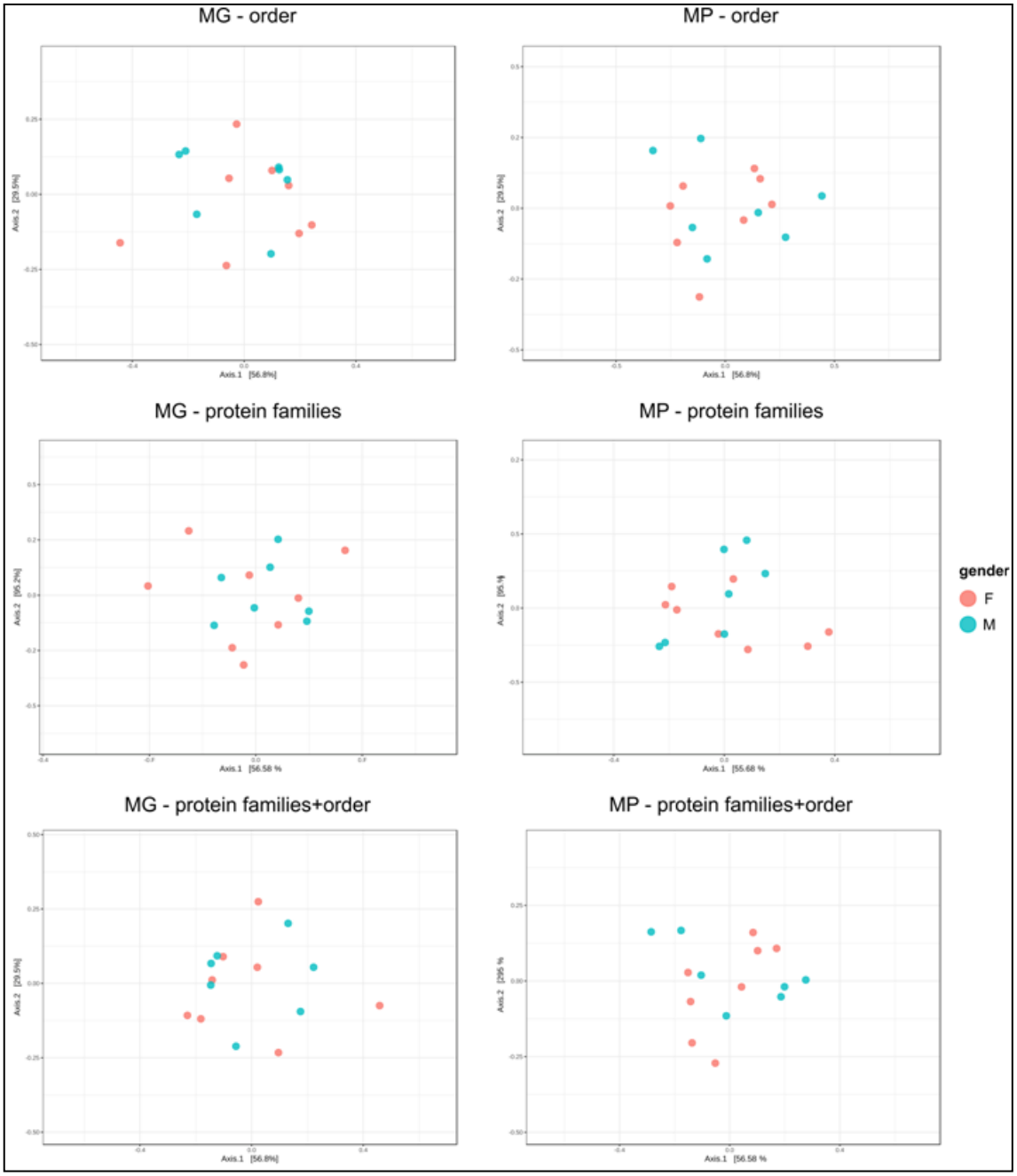

Figure 50: Principal coordinate analysis plots related to taxonomic, functional and taxa/function combined features based on the studied population gender. PCoA plots were generated with the web application MicrobiomeAnalyst based on Bray-Curtis dissimilarity.

The orders Selenomonadales and Veillonellales, as well as the genera *Megasphaera*, *Dialister*, *Prevotella* and *Mitsuokella* were detected as more abundant in the

metaproteome of the young (under 32) subjects, the latter also in the metagenome and again with an elevated fold-change; *Subdoligranulum* and *Bifidobacterium* (and the corresponding order Bifidobacteriales) showed a significant higher abundance in the older subjects (over 47).

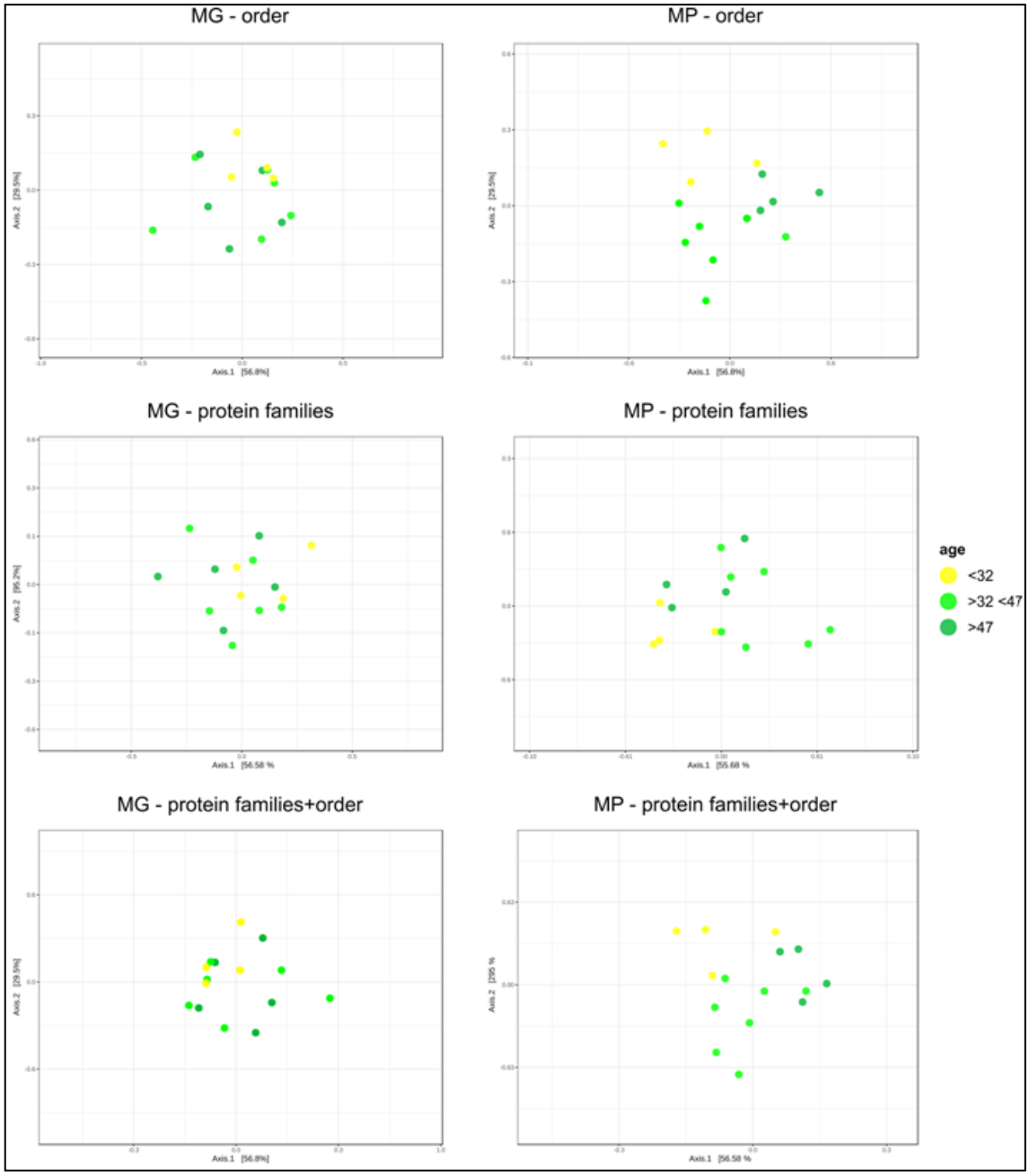

Figure 51: Principal coordinate analysis plots related to taxonomic, functional and taxa/function combined features based on the studied population age. PCoA plots were generated with the web application MicrobiomeAnalyst based on Bray-Curtis dissimilarity.

Selenomonadales and *Mitsuokella* were more abundant in female with high caloric intake too, while *Prevotella* in alcohol high-consumer males both in MG and MP, in line with what already reported (Queipo-Ortuno et al 2012). Furthermore, high levels of *Phascolarctobacterium* were observed in the metagenome of high consumers of fats and in males with low caloric intake.

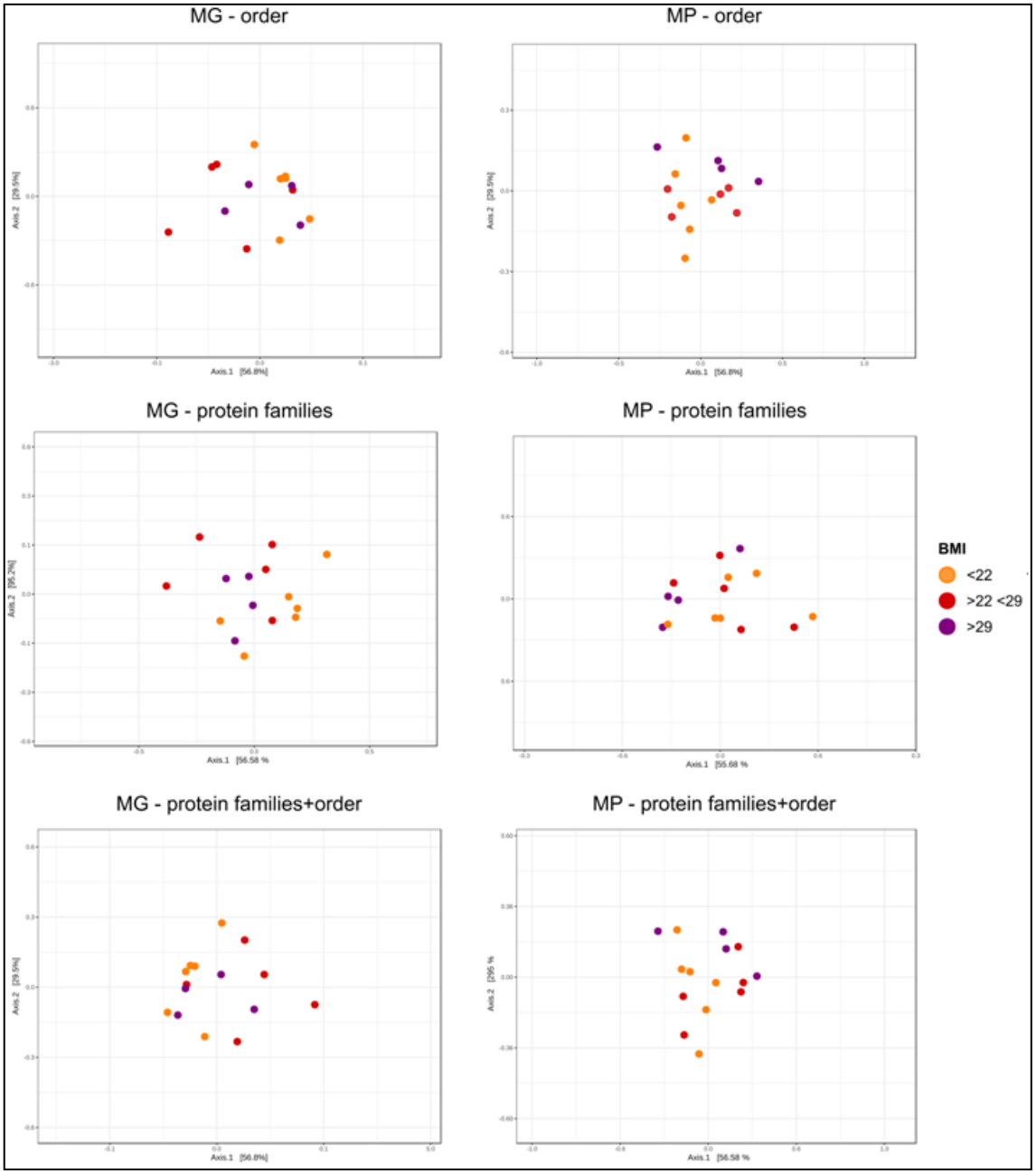

Figure 52: Principal coordinate analysis plots related to taxonomic, functional and taxa/function combined features based on the studied population BMI. PCoA plots were generated with the web application MicrobiomeAnalyst based on Bray-Curtis dissimilarity.

Focusing on functions, Tables 16 and 17 show the differential abundant protein families detected in the various comparison in MP and MG, respectively.

category	No. of individuals	comparison groups	differential abundant features							
			order		genus		protein families		protein families+order	
			MP	MG	MP	MG	MP	MG	MP	MG
gender	7 vs 7	M vs F	1	0	0	0	0	0	0	0
age	4 vs 4	<32 vs >47 (years)	3	3	7	5	17	1	56	7
BMI	4 vs 4	<22 vs >29	0	0	0	0	0	0	0	0
caloric intake (M)	4 vs3	<1700 vs >2000 (kcal)	0	2	0	5	0	0	0	0
caloric intake (F)	4 vs 3	<1600 vs >1600 (kcal)	2	2	2	0	2	6	2	0
carbohydrates	4 vs 6	<40 vs >45 (%)	0	0	1	0	0	0	0	0
alcohol (M)	3 vs 3	<5 vs >8 (%)	0	0	1	4	7	9	11	6
alcohol (F)	3 vs 4	yes vs no	1	1	0	1	7	0	21	0
fats	6 vs 6	<35 vs >38 (%)	0	0	0	10	1	0	1	2
proteins	4 vs 4	<12 vs >15 (%)	0	0	0	0	1	0	1	0
fibers (F)	3 vs 3	<3.5 vs 5 (%)	0	0	0	0	1	80	1	3

Table 14: Summary of the differential abundant features detected in each comparison. Statistical comparative analysis was performed using edgeR and Benjamini-Hochberg correction (FDR <0.05). M = male, F = female.

It is of interest the high number of differential potential functions detected when comparing females with high and low fiber consumption. A total of 45 protein families with at least 0.05% of their relative abundance within the population exhibited significant difference in abundance, 37 higher in elevated fiber consumers and 8 in low eaters of fibers. Interestingly, 6 glycosyl hydrolase families (9 if considering all differential features without filtering for their relative abundance, including 2 cellulase) were more abundant in the high-fiber group, against only 1 (2 without filtering) in the low-fiber group. In addition, arabinose and xylose isomerase families, a polysaccharide lyase, and the enzyme that catalyzed the epimerization of cellobiose, the structural unit of cellulose, showed a high abundance in the fiber consumers, whereas oleate hydratase and amidase families in the low-fiber group. This important divergence observed in the metagenome between the low- and the high-fiber consumers, yet not detected in the metaproteome, might lead to suppose that subjects had eaten a rich-fiber diet during the week-long observation might have owned and selected a wider set of hydrolytic enzyme genes in their metagenome thanks to their long-term history of dietary habits. Furthermore, almost no differential

functions were detected when combined them with the order (see Table 18), suggesting no associations with the taxonomy.

Concerning the other comparative analysis, deserve a mention the Salmonella virulence plasmid 65kDa B protein (SpvB) family, that was detected as different abundant in many comparison; actually, this protein family was present with a very high abundance in only two subjects (1.8% and 0.9%, while the average percentage of the other 13 subjects was <0.05%), thus influencing the analysis.

taxonomy	category	MP		MG	
		log_2FC M/F	**FDR**	**log_2FC M/F**	**FDR**
Coriobacteriales	gender	2.34	3.29E-02		
taxonomy	**category**	**log_2FC <32/>47**	**FDR**	**log_2FC <32/>47**	**FDR**
Selenomonadales	age	4.29	1.37E-05		
Veillonellales		2.56	2.90E-04		
Bifidobacteriales		-3.06	9.70E-04		
Mitsuokella		7.25	9.26E-07	9.99	3.77E-04
Megasphaera		3.99	1.35E-02		
Dialister		2.80	1.35E-02		
Prevotella		2.66	1.12E-02		
Subdoligranulum		-2.46	1.60E-02		
Bifidobacterium		-2.98	1.12E-02		
taxonomy	**category**	**log_2FC low/high**	**FDR**	**log_2FC low/high**	**FDR**
Acidaminococcales	caloric intake (M)			7.32	3.36E-02
Phascolarctobacterium				9.45	1.31E-02
taxonomy	**category**	**log_2FC low/high**	**FDR**	**log_2FC low/high**	**FDR**
Selenomonadales	caloric intake (F)	-3.04	4.95E-02		
Mitsuokella		-4.95	1.22E-02		
taxonomy	**category**	**log_2FC yes/no**	**FDR**	**log_2FC low/high**	**FDR**
Acidaminococcus	carbohydrates	4.39	7.93E-03		
taxonomy	**category**	**log_2FC low/high**	**FDR**	**log_2FC low/high**	**FDR**
Akkermansia	alcohol (F)			10.50	2.14E-03
taxonomy	**category**	**log_2FC low/high**	**FDR**	**log_2FC low/high**	**FDR**
Prevotella	alcohol (M)	-4.26	3.20E-02	-8.63	1.21E-02
Sutterella				-8.65	3.63E-02
taxonomy	**category**	**log_2FC low/high**	**FDR**	**log_2FC low/high**	**FDR**
Butyricicoccus	fats			-4.74	2.18E-02
Phascolarctobacterium				-6.17	2.18E-02

Table 15: Differential abundant features at taxonomy level in each comparison. Statistical comparative analysis was performed using edgeR and Benjamini-Hochberg correction (FDR <0.05). M = male, F = female. Only orders and genera with at least 0.5% of their relative abundance in one of the two groups for each comparison were reported. Taxa are ordered based on taxonomic level and then by decreasing fold change.

Finally, focusing on taxonomy/function combinations, Tables 18 and 19 show the differential abundant protein families detected in the various comparison in MG and MP, respectively.

The main evidence was the opposite trends of Bacteroidales and Clostridiales, with functions assigned to the former order more abundant in the under 32 years old, males high-alcohol consumers and females that have not drunk alcohol, and the latter order in the opposite groups. Concerning alcohol consumption, it would appear that a specific Clostridiales/Bacteroidales ratio (and consequently F/B ratio) was associated with a moderate drinking level, whereas both the absence and the elevated consumption were associated with higher levels of Bacteroidales.

Taken together, these data, although preliminary and obtained on a small number of individuals, suggested that differences in the potential and in the active functionalities can be detected within a healthy and homogeneous population based on diverse dietary habits.

Further in-depth analyses, extended to a larger cohort, are required to validate both functional and taxonomic information, as well as new computations of the present data, taking into consideration some relevant covariates (as an example, age) or with other algorithms.

Day 1 DATE____/____/ _____

Food	Type and quantity	time
BREAKFAST		
Drinks (milk, tea, juice, coffee)		
Fruits		
Yogurt		
Sweets		
other/comments		
MORNING LUNCHEON		
LUNCH		
Carbohydrates/grain (pasta, rice)		
Vegetable		
Cheese		
Legumes		
Fruits		
Meat		
Fish		
Eggs		
Drinks (wine, water, beer; ...)		
Condiments (olive oil, butter, cheese)		
Sweets		
other/comments		
AFTERNOON LUNCHEON		
DINNER		
Carbohydrates/grain (pasta, rice)		
Vegetable		
Cheese		
Legumes		
Fruits		
Meat		
Fish		
Eggs		
Drinks (wine, water, beer; ...)		
Condiments (olive oil, butter, cheese)		
Sweets		
other/comments		
EVENING LUNCHEON		

Figure 53: One-week dietary diary facsimile.

differential MP protein families - age	log_2FC <32/>47	FDR
SpvB family	5.59	1.31E-04
SusF family	2.84	3.78E-02
Ribosomal protein L31P family	2.63	4.49E-02
MurCDEF family	2.63	1.00E-02
NusA family	2.50	3.78E-02
Disproportionating enzyme family	2.23	4.49E-02
MotB family	2.11	9.88E-03
Beta-ketoacyl-ACP synthases family	1.99	4.36E-02
FKBP-type PPIase family	1.55	2.55E-02
Ketol-acid reductoisomerase family	-1.39	2.41E-02
Xylose isomerase family	-1.91	4.49E-02
Transaldolase family	-1.98	1.00E-02
Bacterial/plant glucose-1-phosphate adenylyltransferase family	-2.56	1.52E-03
Ni-containing carbon monoxide dehydrogenase family	-2.86	2.55E-02
XFP family	-2.88	5.09E-04
differential MP protein families - caloric intake (F)	**log_2FC low/high**	**FDR**
MotB family	-2.81	3.89E-02
SpvB family	-5.21	1.32E-03
differential MP protein families - alcohol (F)	**log_2FC yes/no**	**FDR**
Cysteine synthase/cystathionine beta-synthase family	3.22	7.12E-04
Glycosyl hydrolase 9 (cellulase E) family	3.16	4.37E-03
Glycosyl hydrolase 94 family	2.80	5.11E-03
UPRTase family	2.44	3.17E-02
GPI family	1.80	2.90E-02
ATPase C chain family	-2.17	3.17E-02
Ribosomal protein L31P family	-3.51	2.90E-02
differential MP protein familie - alcohol (M)s	**log_2FC low/high**	**FDR**
Xanthine dehydrogenase family	3.92	4.51E-02
XFP family	3.50	1.03E-02
Peptidase S8 family	2.98	1.97E-02
Bacterial microcompartments protein family	2.75	2.35E-02
Aldehyde dehydrogenase family	2.69	4.51E-02
Transaldolase family	2.35	4.51E-02
MurCDEF family	-3.46	1.71E-02
differential MP protein families - fats	**log_2FC low/high**	**FDR**
SpvB family	4.67	5.09E-04
differential MP protein families - proteins	**log_2FC low/high**	**FDR**
SpvB family	5.67	4.59E-05
differential MP protein families - fibers (F)	**log_2FC low/high**	**FDR**
SpvB family	5.70	1.87E-03

Table 16: Differential abundant active functions in each comparison. Statistical comparative analysis was performed using edgeR and Benjamini-Hochberg correction (FDR <0.05). M = male, F = female. Only protein families with at least 0.05% of their average relative abundance among subjects are reported. Features are ordered by decreasing fold change.

differential MG protein families - fibers (F)	log_2FC low/high	FDR
Glycosyl hydrolase 1 family	2.23	1.12E-02
Oleate hydratase family	2.03	1.59E-02
Transposase IS3/IS150/IS904 family	1.90	2.98E-02
UPF0371 family	1.84	5.08E-03
Amidase family	1.80	1.80E-02
Transposase 11 family	1.74	2.54E-03
ThiI family	1.66	2.07E-02
UPF0210 family	1.47	1.45E-02
Transferase hexapeptide repeat family	-0.97	3.21E-02
Class-II pyridoxal-phosphate-dependent aminotransferase family	-0.99	2.71E-02
Class-I pyridine nucleotide-disulfide oxidoreductase family	-1.02	3.78E-02
Orn/Lys/Arg decarboxylase class-II family	-1.05	3.78E-02
Phosphofructokinase type A (PFKA) family	-1.07	2.54E-02
Polysaccharide synthase family	-1.17	4.92E-02
AccD/PCCB family	-1.22	1.45E-02
SurE nucleotidase family	-1.24	4.04E-02
Resistance-nodulation-cell division (RND) (TC 2.A.6) family	-1.24	3.02E-03
UxaA family	-1.25	3.07E-02
Peroxiredoxin family	-1.32	2.71E-02
Complex I subunit 5 family	-1.33	8.39E-03
DegT/DnrJ/EryC1 family	-1.34	2.02E-02
OXA1/ALB3/YidC family	-1.35	2.31E-02
Complex I subunit 1 family	-1.36	4.95E-02
Glycosyl hydrolase 3 family	-1.41	1.80E-02
Diaminopimelate dehydrogenase family	-1.43	4.77E-02
MscS (TC 1.A.23) family	-1.43	5.08E-03
Arabinose isomerase family	-1.46	4.13E-02
Mannonate dehydratase family	-1.49	4.13E-02
Transposase IS21/IS408/IS1162 family	-1.49	4.92E-02
Glycosyl hydrolase 51 family	-1.56	1.04E-02
Glycosyl hydrolase 2 family	-1.60	7.30E-03
KduI family	-1.73	4.39E-02
TonB-dependent receptor family	-1.77	2.71E-02
Phenylacetyl-CoA ligase family	-1.79	1.23E-02
Glycosyl hydrolase 130 family	-1.80	3.21E-02
6-phosphogluconate dehydrogenase family	-1.86	1.45E-02
Peptidase S46 family	-1.87	5.18E-04
KdpA family	-1.94	1.70E-02
Glycosyl hydrolase 31 family	-1.95	3.35E-05
Methylmalonyl-CoA mutase family	-2.05	1.29E-03
Complex I subunit 2 family	-2.12	6.15E-03
Glycosyl hydrolase 43 family	-2.15	3.85E-05
LipB family	-2.29	3.85E-05
Xylose isomerase family	-2.52	4.91E-04
Polysaccharide lyase 8 family	-2.99	8.75E-10
differential MG protein families - caloric intake (F)	**log_2FC low/high**	**FDR**
UPF0371 family	-1.73	2.60E-02
Oleate hydratase family	-1.99	3.06E-02

differential MG protein families - alcohol (M)	log_2FC low/high	FDR
LpxC family	-1.68	1.17E-02
Glycosyl hydrolase 67 family	-2.39	1.17E-02
Transposase mutator family	-3.04	7.93E-05

Table 17: Differential abundant potential functions in each comparison. Statistical comparative analysis was performed using edgeR and Benjamini-Hochberg correction (FDR <0.05). M = male, F = female. Only protein families with at least 0.05% of their average relative abundance among subjects are reported. Features are ordered by decreasing fold change.

differential MG protein families-order - age	log2FC (U32vsO47)	FDR
Transposase 11 family+Bacteroidales	2.89	1.44E-02
differential MG protein families-order - fats	**log2FC (low vs high)**	**FDR**
ABC transporter superfamily+Enterobacterales	-3.83	3.67E-02
differential MG protein families-order - fibers (F)	**log2FC (low vs high)**	**FDR**
SusD family+Bacteroidales	-2.98	3.58E-02
Polysaccharide lyase 8 family+Bacteroidales	-3.03	3.58E-02

Table 18: Differential abundant protein family/order combinations on MG data in each comparison. Statistical comparative analysis was performed using edgeR and Benjamini-Hochberg correction (FDR <0.05). M = male, F = female. Only protein families with at least 0.05% of their average relative abundance among subjects are reported. Features are ordered by decreasing fold change.

differential MP protein families+order - age	log2FC <32/>O47	FDR
SpvB family+Bacteroidales	5.42	9.75E-04
SecA family+Bacteroidales	3.32	6.77E-03
MurCDEF family+Bacteroidales	3.15	6.77E-03
Isocitrate and isopropylmalate dehydrogenases family+Bacteroidales	3.13	1.33E-02
NusA family+Bacteroidales	3.09	2.35E-02
ATCase/OTCase family+Bacteroidales	2.98	3.61E-02
Class-I aminoacyl-tRNA synthetase family+Bacteroidales	2.75	6.69E-03
FKBP-type PPIase family+Bacteroidales	2.65	1.34E-02
Ribosomal protein S2P family+Bacteroidales	2.63	6.77E-03
Methylmalonyl-CoA mutase family+Veillonellales	2.61	6.77E-03
SusF family+Bacteroidales	2.61	4.88E-02
CarB family+Bacteroidales	2.53	2.39E-02
Phosphohexose mutase family+Bacteroidales	2.50	1.75E-02
Ketol-acid reductoisomerase family+Bacteroidales	2.47	1.75E-02
Triosephosphate isomerase family+Selenomonadales	2.42	4.72E-02
Glycosyl hydrolase 2 family+Bacteroidales	2.32	1.92E-02
Class-II aminoacyl-tRNA synthetase family+Bacteroidales	2.31	4.16E-02
Glycosyl hydrolase 51 family+Bacteroidales	2.30	1.75E-02
Phosphoglycerate kinase family+Bacteroidales	2.24	2.59E-02
FGAMS family+Bacteroidales	2.09	3.61E-02
Transketolase family+Bacteroidales	2.04	4.88E-02
MotB family+Bacteroidales	1.94	4.16E-02
PEP-utilizing enzyme family+Bacteroidales	1.87	4.88E-02
NifJ family+Clostridiales	-1.43	4.88E-02
ETF alpha-subunit/FixB family+Clostridiales	-1.66	3.61E-02

ABC transporter superfamily+Clostridiales	-1.68	4.90E-02
Glyceraldehyde-3-phosphate dehydrogenase family+Clostridiales	-1.76	4.97E-02
DapA family+Clostridiales	-1.78	4.16E-02
Uronate isomerase family+Clostridiales	-1.86	3.61E-02
Ketol-acid reductoisomerase family+Clostridiales	-1.95	2.35E-02
Transaldolase family+Clostridiales	-2.10	3.61E-02
TelA family+Clostridiales	-2.12	2.57E-02
Glycogen phosphorylase family+Clostridiales	-2.16	3.86E-02
Glycosyl hydrolase 130 family+Clostridiales	-2.17	4.16E-02
Ribosomal protein L13P family+Clostridiales	-2.18	3.71E-02
Phosphofructokinase type A (PFKA) family+Clostridiales	-2.38	3.61E-02
Zinc metallo-hydrolase group 3 family+Clostridiales	-2.41	2.46E-02
Ribosomal protein L21P family+Clostridiales	-2.46	4.97E-02
Xylose isomerase family+Bifidobacteriales	-2.54	4.88E-02
Aldolase class II family+Clostridiales	-2.56	3.85E-03
Acetyl-CoA hydrolase/transferase family+Clostridiales	-2.58	2.23E-02
Enolase family+Bifidobacteriales	-2.72	4.16E-02
Bacterial/plant glucose-1-phosphate adenylyltransferase family+Clostridiales	-3.15	1.17E-03
XFP family+Bifidobacteriales	-3.52	1.73E-02
XFP family+Clostridiales	-4.06	2.10E-05
differential MP protein families+order - caloric intake (F)	**log2FC low/high**	**FDR**
Glycosyl hydrolase 3 family+Clostridiales	-3.58	3.93E-03
SpvB family+Bacteroidales	-5.28	3.93E-03
differential MP protein families+order - alcohol (F)	**log2FC yes/no**	**FDR**
Polyribonucleotide nucleotidyltransferase family+Clostridiales	3.72	1.11E-03
Glycosyl hydrolase 9 (cellulase E) family+Clostridiales	3.21	7.17E-03
XFP family+Clostridiales	2.92	4.93E-02
Glycosyl hydrolase 94 family+Clostridiales	2.85	1.86E-02
Class-V pyridoxal-phosphate-dependent aminotransferase family+Clostridiales	2.47	7.17E-03
Class-II aminoacyl-tRNA synthetase family+Clostridiales	2.44	1.38E-02
FAD-dependent oxidoreductase 2 family+Clostridiales	2.40	1.86E-02
UPRTase family+Clostridiales	2.37	4.09E-02
Zinc metallo-hydrolase group 3 family+Clostridiales	2.36	3.58E-02
DapA family+Clostridiales	2.26	4.52E-02
GPI family+Clostridiales	2.14	2.87E-02
Ribosomal protein L10P family+Clostridiales	2.02	4.85E-02
ATPase C chain family+Bacteroidales	-2.22	3.58E-02
Gfo/Idh/MocA family+Bacteroidales	-2.33	4.85E-02
Ribosomal protein L5P family+Bacteroidales	-2.48	3.58E-02
Glycosyl hydrolase 130 family+Bacteroidales	-2.55	3.97E-02
4Fe4S bacterial-type ferredoxin family+Bacteroidales	-2.91	3.97E-02
Glycosyl hydrolase 31 family+Bacteroidales	-3.41	7.17E-03
Ribosomal protein L13P family+Bacteroidales	-3.46	7.17E-03
differential MP protein families+order - alcohol (M)	**log2FC low/high**	**FDR**
XFP family+Clostridiales	4.74	3.54E-02
Glycosyl hydrolase 3 family+Clostridiales	4.20	4.60E-02
XFP family+Bifidobacteriales	4.18	4.60E-02
Glycosyl hydrolase 130 family+Clostridiales	3.60	3.54E-02
Bacterial solute-binding protein 5 family+Bifidobacteriales	3.57	3.54E-02
FAD-dependent oxidoreductase 2 family+Clostridiales	3.24	3.54E-02

Glycogen phosphorylase family+Clostridiales	3.18	4.60E-02
Class-I fumarase family+Bacteroidales	-3.09	4.60E-02
MurCDEF family+Bacteroidales	-3.30	4.60E-02
Glycosyl hydrolase 13 family+Bacteroidales	-3.48	3.54E-02
differential MP protein families+order - fats	**log2FC low/high**	**FDR**
SpvB family+Bacteroidales	4.52	3.81E-03
differential MP protein families+order - proteins	**log2FC low/high**	**FDR**
SpvB family+Bacteroidales	5.57	6.00E-03
differential MP protein families+order - fibers	**log2FC low/high**	**FDR**
SpvB family+Bacteroidales	5.72	7.84E-03

Table 19: Differential abundant protein family/order combinations on MP data in each comparison. Statistical comparative analysis was performed using edgeR and Benjamini-Hochberg correction (FDR <0.05). M = male, F = female. Only protein families with at least 0.05% of their average relative abundance among subjects are reported. Features are ordered by decreasing fold change.

2.4 Human gut microbiome variations according to dietary interventions

2.4.1 Aim of the study

Type 2 diabetes (T2D) is a complex disorder influenced by both genetic and environmental components, and has become one of the challenging public health issue throughout the world; the World Health Organization (WHO) reported that the worldwide prevalence of diabetes, the majority concerning the type 2, has risen from 4.7% in 1980 to 8.5% in 2014 among adults over 18 years of age (about 422 billions of people) (http://www.who.int/mediacentre/factsheets/fs312/en/).

Recently, evidences of an altered gut microbiota in T2D subjects were provided (Candela et al 2016, Karlsson et al 2013, Qin et al 2012, Zhang et al 2013b), suggesting a possible role for intestinal microbes in the disease onset.

In a recent study from Costantino Palmas and coworkers (Cereal 14/20, Università di Cagliari) a dietary intervention, based on the substitution of industrial bread and pasta with a sourdough-leavened bread and "functional" pasta both produced from ancient wheat cultivar, milled with traditional methods, was demonstrated to ameliorate clinical conditions in a cohort of Sardinian T2D patients (Palmas et al., unpublished data).

Here, by means of 16S metagenomic and metaproteomic analyses, this study aimed to investigate the changes induced on the gut microbiota of T2D patients and healthy controls after "challenge" with the above mentioned functional foods. To date, clinical data from University of Cagliari is not yet available and the results presented here are related only on gut microbiota structural and functional changes according to the dietary interventions.

2.4.2 Experimental design

Gut microbiota was evaluated in 12 T2D patients (8 males and 4 females), and 10 healthy controls (Ctrl) (8 males, 2 females), subjected to dietary intervention by substituting common bread with a functional one for 2 months. Further, the Ctrl group was subjected to an additionally month of intervention by substituting both bread and pasta with functional ones. Fecal samples were collected from each subject at the beginning (T0) and after 2 months (2M) of the bread treatment (+B), and from Ctrl at the end of the further treatment with functional pasta (+P). DNA and proteins were extracted from fecal samples, with the aim of carrying out 16S rRNA gene and metaproteome analysis, as illustrated in Figure 54.

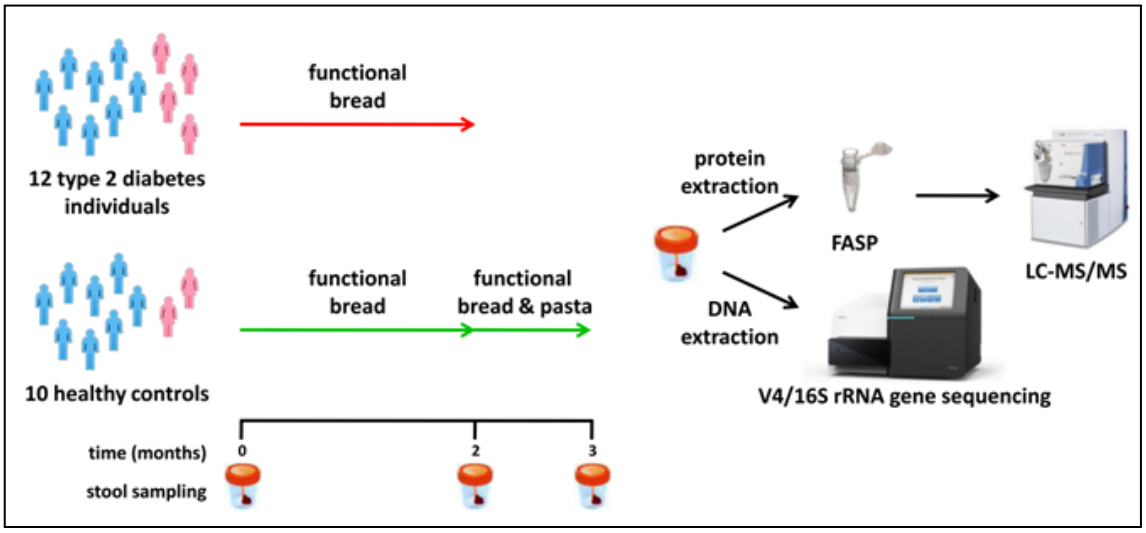

Figure 54: Schematic illustrating the experimental design of the study.

2.4.3 Material & Methods

Sample collection

Stool samples were collected from 12 T2D individuals (8 males, 4 females) selected from the Centre of Diabetology, San Giovanni University Hospital, Cagliari (Italy), and 10 healthy controls (8 males, 2 females). All 22 individuals were subjected to dietary treatment by substituting for 2 months the common industrial bread they were used to with a functional one. The functional bread was obtained starting from the Senatore Cappelli and the Karalis varieties of durum wheat, through cold milling on stone and

sourdough leavening. Fecal samples were collected at the beginning and at the end of the dietary intervention. In addition, healthy participants continued for an extra month the treatment with the simultaneous substitution of pasta with the functional pasta, obtained from the same wheat varieties. Feces were collected at the end of the month.

All collected feces (N = 54) were immediately stored at -80°C, next transferred to the Porto Conte Ricerche laboratories in dry ice, and stored again at -80°C until use. Then, samples were thawed at 4°C, and from each of them two equal fecal portions (weighing approximately 250 mg each) were collected, the first subjected to DNA extraction and the second to protein extraction.

DNA extraction and 16S rDNA gene sequencing

DNA extraction was performed with the QIAamp DNA Stool Mini Kit, while 16S rRNA gene (V4 region) amplicons and libraries preparation were performed as described in Section 1.3.3. V4 amplicon sequencing was performed with a MiSeq sequencer using the MiSeq Reagent Kit v3, the paired-end method and 201 cycles of sequencing.

Data quality control and analyses were carried out with QIIME. The overlapping paired-end reads were merged using the script join_paired_ends.py inside QIIME, retaining for further analysis only reads with a length >200 bps. OTUs generation and their taxonomy assignment were performed with the pipeline explained before (see Section 1.2.3).

Protein sample preparation and metaproteome bioinformatics

Fecal samples were undergone to protein extraction, LC-MS/MS analysis, and peptide identification as detailed in Section 1.2.3.

The sequence database used for peptide identification was the same as Section 2.3.

All ORFs matched with at least an MS spectrum upon database searching (average length 42 amino acids) were subjected to taxonomic and functional classification, following the same procedure described in Section 2.3.3.

Statistical analysis and graph generation

Differential abundance analysis at aggregated taxonomic level was carried out on count data between groups using edgeR on the web application MicrobiomeAnalyst (after filtering out few represented features and RLE normalization), whereas analysis within groups on paired data was performed the R package ibb. *P*-values were corrected for multiple inference using the Benjamini-Hochberg FDR procedure with an adjusted alpha cutoff value of 0.05. Alpha diversity (Shannon index), calculated starting from the OTU table, and richness (observed OTU counts) were compared among groups with one-way ANOVA, and within groups using repeated measures ANOVA, both followed by Bonferroni comparison on all pairs of groups (alpha-value = 0.05) on GraphPad Prism. PCA plots were generated starting from the relative abundance of OTUs using the web application ClustVis with default parameters and edited with InkScape. Scatter plots, generated starting from relative abundance data were made using GraphPad Prism.

2.4.4 Results

General metrics

A total of 12 T2D patients from the Centre of Diabetology of Cagliari and 10 healthy controls (Ctrl) were subjected to dietary treatment by substituting totally for 2 months the common bread in the diet with a bread obtained with flour from two ancient varieties of durum wheat, and sourdough leavening; in addition, Ctrl were subjected to an extra month of dietary treatment by substituting also pasta with a functional one, made starting from the same wheat varieties and milling procedures. Fecal samples were collected from each subject at the beginning (T0), after 2 months of bread intervention (2M) and, from Ctrl, after 1 month of further pasta intervention (3M), and analyzed both with 16S metagenomics and metaproteomics in order to evaluate changes in the gut microbiota.

T2D and Ctrl subjects metrics were reported in Table 20 and 21, respectively. T2D individuals (8 males and 4 females) presented a median age of 63 years (range: 56-74), a diabetes duration of 3-16 years (median = 7), and a median BMI value of 28.5 (range:

22.3-33.8) at T0 and 28.6 (range: 23.3-33.8) after 2 months of functional bread intervention (2M).

Ctrl subjects presented a median age of 56 years (range: 51-64), and a constant median BMI value of 24.6 (range: 18.7-34.7) at T0, 2M (range: 18.7-34.4) and after 1 month (3M) of further treatment with functional pasta (range: 18.7-34.4).

Sample_ID	sex	age	duration of diabetes (years)	BMI T0	BMI 2M
S1	M	68	12	31.7	31.5
S2	M	63	5	22.3	23.3
S3	F	61	7	33.1	32.7
S4	M	66	3	28.4	28.8
S5	F	72	16	33.8	33.8
S6	M	74	15	31.9	32.3
S7	F	56	5	27.4	27.4
S8	M	57	7	24.9	24.9
S9	M	65	5	29.0	28.8
S10	M	59	10	25.5	25.5
S11	M	58	6	25.8	23.9
S12	F	63	14	28.6	28.4

Table 20: Metrics of the T2D subjects selected for the study. Individuals were selected from the Centre of Diabetology, San Giovanni University Hospital, Cagliari (Italy). BMI was measured at the beginning (T0) and after 2 months (2M) of dietary treatment with functional bread.

Sample_ID	sex	age	BMI T0	BMI 2M	BMI 3M
S13	M	51	23.2	23.2	23.2
S14	M	55	26.4	26.4	26.1
S15	M	52	20.7	20.7	20.7
S16	M	60	32.8	32.6	32.4
S17	M	53	18.7	18.7	18.7
S18	F	57	21.9	21.9	21.9
S19	M	60	34.7	34.4	34.4
S20	F	61	24.8	24.8	24.8
S21	M	64	24.8	24.8	24.8
S22	M	53	24.3	24.3	24.3

Table 21: Metrics of the healthy controls selected for the study. BMI was measured at the beginning (T0), after 2 months (2M) of dietary treatment with functional bread, and after 1 month (3M) of further treatment with functional pasta (+P).

16S rDNA sequencing

Sequencing of the of V4 region of the 16S rDNA from a total of 54 fecal samples, enabled to obtain 1,115,500 reads, after the merging of paired-end reads.

Richness, calculated as the number of OTUs detected, and alpha-diversity (Shannon index) values within groups were reported in Table 22. No significant differences both among and within groups were detected.

PCA plots generated starting from the relative abundance of OTUs did not show any separations between T2D and Ctrl, both before and at the end of the dietary intervention with functional bread (Figure 55), and within groups at the different time points (Figure 56).

group	Richness (No. of observed OTU)	alpha-diversity (Shannon index)
T2D T0	457 ±123	5.334 ±0.814
Ctrl T0	423 ±147	5.138 ±1.447
T2D+B 2M	456 ±130	5.221 ±0.970
Ctrl+B 2M	456 ±135	4.884 ±1.368
Ctrl+B+P 3M	480 ±139	5.155 ±1.096

Table 22: OTU richness and alpha-diversity within groups at the beginning and after the end of the dietary treatment. Type 2 diabetes (T2D) and healthy individuals (Ctrl), at the beginning (T0), after 2 months (2M) of dietary treatment with functional bread (+B), and after 1 month (3M) of further treatment with functional pasta (+P). Means and standard deviation of the number of detected OTUs and Shannon index are reported.

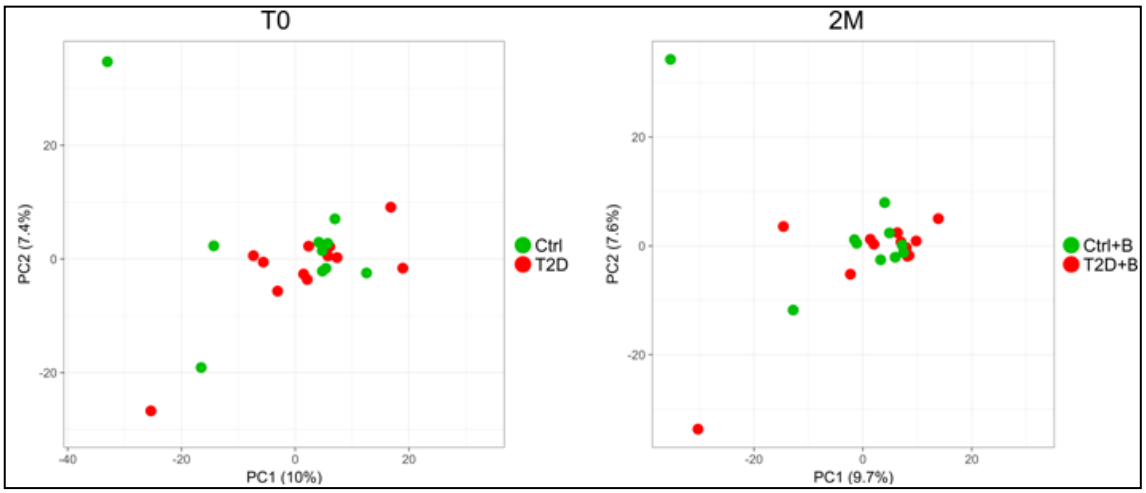

Figure 55: Beta-diversity at OTU level among groups at the beginning and after the dietary intervention. Type 2 diabetes (T2D) and healthy individuals (Ctrl), at the beginning (T0) and after 2 months (2M) of dietary treatment with functional bread (+B). PCA was carried starting from OTUs (relative abundance) with the web application ClustVis. Each dot indicate a sample.

Differences in abundance of families and genera were tested among and within groups through statistical comparative analysis. In Figure 57 are represented the more abundant taxa (relative abundance >0.5%) observed as differential.

On the whole, few differential features were observed with the main differences observed between T2D and Ctrl at T0. In detail, Prevotellaceae, Bacteroidaceae and the genus *Lactobacillus*, were observed as higher in T2D, whereas Verrucomicrobiaceae higher in Ctrl.

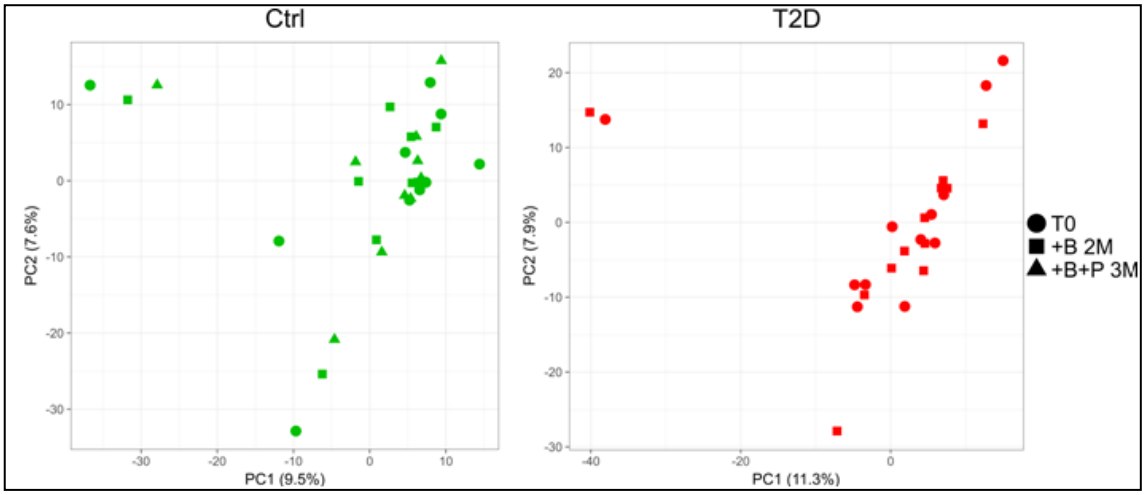

Figure 56: Beta-diversity at OTU level within groups at different time points. Type 2 diabetes (T2D) and healthy individuals (Ctrl), at the beginning (T0), after 2 months (2M) of dietary treatment with functional bread (+B), and after 1 month (3M) of further treatment with functional pasta (+P). PCA was carried out at OTU level with the web application ClustVis. Each dot indicate a sample, while each shape corresponds to a different time point.

Some of these dissimilarities could be related to the metabolic disorder, as *Lactobacillus* was recently reported to be higher in the microbiota of T2D patients after V3-V4 sequencing (Candela et al 2016) and by Karlsson and colleagues after shotgun metagenomics analysis of European women gut microbiota (Karlsson et al 2013), while Verrucomicrobiaceae (i.e. *Akkermansia*) were observed as more abundant in healthy individuals in another study, also based on 16S metagenomics (V3-V5 regions) (Zhang et al 2013b). However, some discordances between the present study results and the literature were noted, as in the work of Candela and colleagues *Prevotella* (belonging to the Prevotellaceae family) and *Bacteroides* (Bacteroidaceae) were higher in healthy controls (Candela et al 2016); moreover, no similarities were found with the shotgun metagenomics data reported by the group of Qin on Chinese individuals (Qin et al 2012), that observed also an opposite trend for *Akkermansia* (higher in T2D).

Concerning differences registered during the dietary experimentation, interestingly, *Dialister* showed a significant increase in T2D subjects and in general when comparing T2D plus Ctrl at T0 vs 2M (data not shown). As reported in Section 2.2.4, sequences of *Dialister* were associated to the latter three enzymes of the propionogenesis in the gut microbiota of a healthy cohort, suggesting its almost exclusively role in this biosynthetic pathways. As propionate was associated to protection from diet-induced obesity in mice (Liou et al 2013) and to weight gain reduction in a randomized study of overweight adults (Chambers et al 2015), this result could be of great interest and further investigation will be needed to confirm this hypothesis, starting from the metaproteomics analysis on the present data, that are currently underway.

To complete this study, clinical traits variations (i.e., glycemia, lipid profiling, etc.) will be collected and their possible correlation with microbiota features will be analyzed.

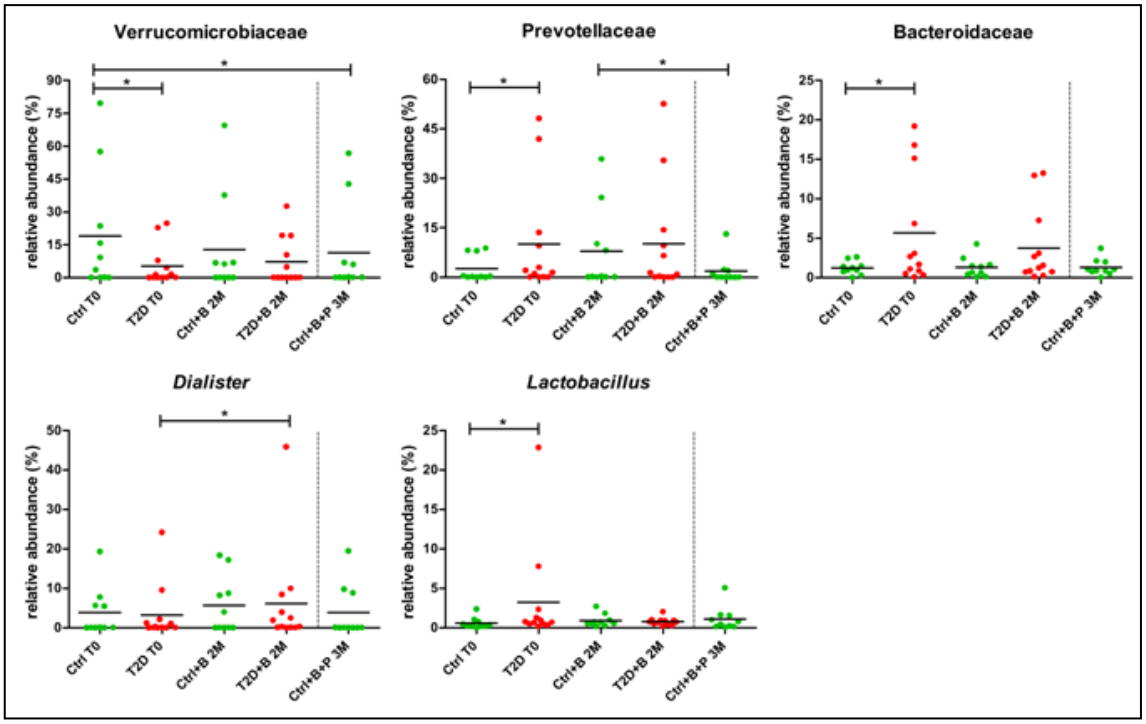

Figure 57: Families and genera with relative abundance variation among groups at the beginning and after the dietary intervention. Represented families and genera are selected based on their relative abundance (>0.5% in at least one group). Type 2 diabetes (T2D) and healthy individuals (Ctrl), at the beginning (T0), after 2 months (2M) of dietary treatment with functional bread (+B), and after 1 month (3M) of further treatment with functional pasta (+P). Each dot represents a different sample. Means are also reported. Asterisks indicate significant difference (edgeR performed on count data, FDR correction, *<0.05) between groups.

2.5 Functions and metabolic pathways in the gastrointestinal tracts of an economically relevant livestock species: the ovine

2.5.1 Aim of the study

Sheep farming is widespread worldwide for the purpose of meat, milk, skin and/or wool production.

Despite their crucial role in the sheep gut metabolism, much less is currently known on the microbial populations that colonize the intestine, while a considerable amount of data have been collected on the composition and functions of sheep rumen microbiota (Brilhante et al 2015, Morgavi et al 2015, Shi et al 2014, Zeng et al 2015).

As in other non-ruminant mammalian, including human, the microbial population that colonizes the sheep large intestine is expected to be key in providing energy, antigens, and metabolites that positively affect host metabolism, physiology and immunity, and, consequently, a productive organism.

In keeping with these premises, the composition and functions of the microbial populations associated to the final tract of the sheep large intestine, where the last stage of plant mass digestion occurs with a significant potential contribution to host energy harvesting and physiology homeostasis, were investigated.

In addition, to a better understand of the microbial communities inhabited the diverse digestive compartments, both the luminal and the mucosa-associated microbiota collected from ten gastrointestinal tracts of a pre-weaned lamb were examined.

To reach this aim, an integrated multi-omic strategy, comprising 16S rDNA, shotgun metagenomic sequencing and metaproteomics, was employed to unravel structure, genetic potential, and functions and pathways actively expressed by the sheep fecal microbiota, as well as to describe microbial taxa and their active functionalities presented in the different tracts of the lamb digestive system.

The results reported in this section have been recently published on *Microbial Biotechnology* (Tanca et al 2017b) and *Proteomes* (Palomba et al 2017).

2.5.2 Experimental design

Fecal samples were collected from 5 lactating sheep from the same lock. Additionally, at necropsy, the luminal and the mucosal contents of ten different gastrointestinal tracts (rumen, reticulum, omasum, abomasum, duodenum, jejunum, ileum, cecum, colon, and rectum) were collected from a pre-weaned lamb.

Fecal samples undergone to both 16S and shotgun metagenome sequencing, as well as shotgun mass spectrometry analysis, while luminal and mucosal samples to 16S metagenomic and metaproteomic analyses (Figure 58).

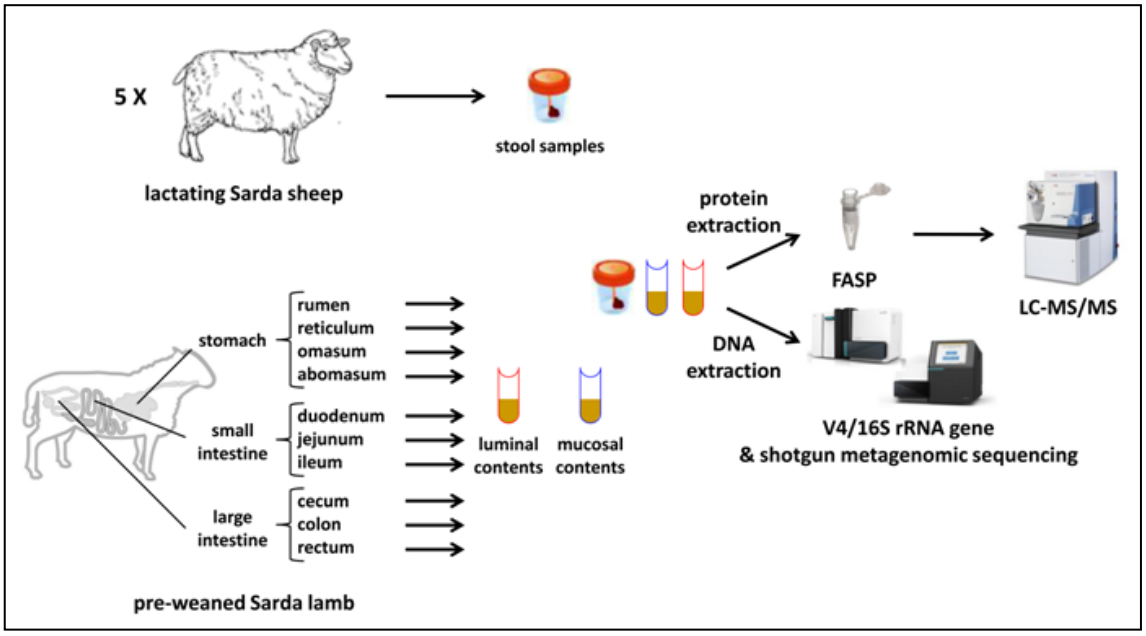

Figure 58: Schematic illustrating the experimental design of the study.

2.5.3 Material & Methods

Animal description and sample collection

Fecal samples were collected from the rectal ampulla of 5 lactating Sarda sheep belonging to the same flock. To minimize the contact with the surrounding environment feces were immediately transferred from the rectum to the collection

tube. Sheep were free-grazing and fed a limited amount of commercial feed only during milking (max. 400 g/day), and were apparently healthy.

All samples were immediately stored at -80°C until use. At the time of the analyses, samples were thawed at 4°C, and from each of them two stool fragments were collected for DNA and protein extraction, respectively.

Additionally, a pre-weaned Sarda lamb (30 days of age) fed almost exclusively mother's milk was sacrificed, and ten different gastrointestinal (GI) tracts (rumen, reticulum, omasum, abomasum, duodenum, jejunum, ileum, cecum, colon, and rectum) were collected. At necropsy, each tract was isolated with stitches to avoid the loss of the luminal content, cut, and gently washed externally with saline solution; then immediately frozen and stored at -80°C until use. Concerning the longest intestinal tracts, such as duodenum and jejunum, only the central section (measuring approximately 10-15 cm) was kept.

Luminal and mucosal contents were collected from GI tracts as follows: each tract was thawed at 4°C, washed externally with saline solution, unrolled, and opened by cutting an extremity on a sterile petri dish to collect the possibly leaking luminal fluid. When tract contents were liquid, the external wall was "squeezed" by flattening using a glass slide. In case of too little amount of luminal material for the subsequent analyses (omasum, ileum and cecum), the inner part was delicately washed with saline and the "washing" content was collected and considered as luminal. Furthermore, as abomasum, colon and rectum presented a more compact content, in this case a central portion of the solid matter was collected.

Subsequently, residual luminal material was washed out with saline trying to prevent the estrangement of mucosa-adhering material and tracts were opened by longitudinal cut and stretched on a clean petri dish. Next, mucosal content was collected by scraping the internal wall using a glass scale.

All luminal (N = 10) and mucosal (N = 10) collected samples were split into two tubes for DNA and protein extraction, respectively.

DNA sample preparation

Fecal samples were subjected to direct lysis (DL) or differential centrifugation (DC) as previously described (Apajalahti et al 1998, Tanca et al 2015). In detail, stool samples (weighing approximately 100 mg each) were resuspended in PBS reach a final volume of 50 ml, vortexed, shaken in a tube rotator for 45 min, and centrifuged at low-speed (500 x *g*) for 5 minutes to eliminate gross particulate and insoluble material. Then, the supernatants were carefully transferred to a clean centrifuge bottle in polyallomer (Beckman Coulter) and kept at 4°C, while the pellets were suspended again in PBS. The whole procedure was repeated for a total of three rounds. Finally, the supernatants (one per round, therefore three per sample) were subjected to centrifugation at 20,000 x *g* for 15 min, and DNA extraction was performed on the derivative pellets in parallel using the QIAamp Fast DNA Stool Kit and the E.Z.N.A. Soil DNA Kit. DC pretreated samples were additionally subjected to DNA extraction according to the standard phenol/chloroform/isoamyl alcohol (25:24:1) method.

Luminal and mucosal samples were undergone directly to DNA extraction using PowerSoil® DNA Isolation Kit (Mo Bio Laboratory Inc. - now part of Qiagen, Carlsbad, CA USA), following the manufacturer's instruction.

16S rDNA analysis

DNA from fecal samples was undergone to V4 region of 16S rDNA amplification, while libraries were constructed according to the Nextera XT kit, and DNA sequencing was performed with the Illumina HiScanSQ sequencer, using the paired-end method and 93 cycles of sequencing.

Concerning luminal and mucosal samples, extracted DNA was subjected also in this case to V4 amplification, but when PCR products were checked on 2% agarose gel, very few amount of microbial DNA regarding several tracts was detected. In view of this, the full-length 16S amplification was performed and in this case a positive and satisfactory amplification for all samples was confirmed by the agarose gel control. A further amplification of V4 was performed on full-length 16S amplicons. V4 libraries were constructed according to the Nextera XT kit, and DNA sequencing was performed

in duplicate with the MiSeq sequencer, using the MiSeq Reagent Kit v3, the paired-end method and 201 cycles of sequencing.

The Illumina demultiplexed paired-reads obtained from fecal samples were trimmed for the first 20 bp using FASTX and the sequences with Nextera adapter contamination were identified using the UniVec database and removed. Therefore, the paired-reads with a minimum overlap of eight bases were merged using a specific QIIME script. Concerning luminal and mucosal sample sequencing, the overlapping paired-end reads were merged using the script join_paired_ends.py inside the QIIME package, retaining for further analysis only reads with a length >200 bps. OTUs generation and their taxonomy assignment were done on all samples using the pipeline described in Section 1.2.3.

The relative proportion of read counts was used as a quantitative estimation of the abundance of each taxon.

Richness (observed OTU counts) and alpha diversity (Shannon index) were calculated starting from the OTU table. PCA plot was generated starting from the relative abundance of OTUs using the web application ClustVis with default parameters and edited with InkScape.

Metagenome analysis

Libraries were constructed starting from fecal samples according to the Nextera XT kit and sequenced with the HiScanSQ sequencer, using the paired-end method and 93 cycles of sequencing.

Paired reads were subjected to merging and filtering with USEARCH (version 8.1.1861) as previously reported (Section 1.2.3).

Taxonomic annotation was performed using MEGAN (version 5.10) on read sequences subjected to DIAMOND (version 0.7.1) search against the NCBI-nr DB (2014/05 update), using the blastx command with default parameters.

Functional annotation was accomplished by DIAMOND blastx search (top hit and e-value threshold 10^{-5}) against UniProt/Swiss-Prot database (taxonomy Bacteria,

2014/12 update) and subsequent retrieval of protein family, KEGG orthologous group and pathway information associated with each UniProt/Swiss-Prot accession number.
The relative proportion of read counts was used as a quantitative estimation of the abundance of each taxon or function.
Moreover, 15 metagenomes selected from luminal and mucosal were obtained to create a custom sequence database for the metaproteomic analysis of the ten lamb gastrointestinal tracts. In this case libraries, constructed with the Nextera XT kit, were sequenced with the MiSeq sequencer, using the MiSeq Reagent Kit v3, with the paired-end method and 201 cycles of sequencing.

Protein sample preparation and metaproteome analysis

Stool samples from sheep (average weight 356 ± 31 mg) and luminal and mucosal samples from lamb were subjected to the protein extraction procedure as detailed in Section 1.2.3.
LC-MS/MS analysis was carried out using an LTQ-Orbitrap Velos an UltiMate 3000 RSLCnano LC system. The single-run 1D LC peptide separation was performed loading 4 μg of the peptide mixture obtained per each sample and applying a 485 min separation gradient. Peptide identification was performed using the Proteome Discoverer (version 1.4), with Sequest-HT as search engine and Percolator for peptide validation (FDR <1%). Search parameters were set as reported previously (Section 1.2.3).
For fecal samples parallel searches were performed using three different sequence databases. The first database was composed by the metagenomic sequences obtained in this study starting from stool samples, both as raw reads and assembled contigs (8,595,757 sequences). Paired reads were merged using the script join_paired_ends.py included in the Qiime package, (version 1.9) with a minimum overlap of 8 base pairs. The output sequences were filtered (with a fastq_truncqual option = 15) and clustered at 100% using USEARCH (version. 5.2.236). Read assembly into contigs was carried out using Velvet with commands and parameters already described in Section 1.2.3. ORFs were found from both reads and contigs using FragGeneScan, with the training for Illumina sequencing reads with about 0.5% error rate.

The second database was a selection of all bacterial, archaeal, fungal and gut microbiota sequences (79,203,800 sequences in total) from the of the UniProtKB database (2015/02 update).

The metaproteomic data (regarding the microbial component of the fecal material) were obtained by merging results of searches against the two above mentioned databases.

A third database (specifically, the whole UniProtKB database, release 2014/12, 89,136,540 sequences) was finally employed to achieve information only concerning the non-microbial components of the sheep microbiota.

The relative proportion of spectral counts (peptide-spectrum matches, PSMs) was used as a quantitative estimation of the abundance of each taxon or function.

Taxonomic and functional assignments were performed as described above for metagenome sequences, except using the DIAMOND blastp command instead of blastx.

Concerning luminal and mucosal samples parallel searches were performed using two different sequence databases. The first database was composed by the metagenomic sequences obtained in this study starting from luminal and mucosal samples, both as raw reads and assembled contigs (3,474,764 sequences). After merging and filtering with USEARCH, paired reads were subjected to sorting, quality filtering and sequence clustering. Raw reads were assembled in parallel into contigs using MetaVelvet (version 1.2.01) (Namiki et al 2012), with velveth command by setting 61 as k-mer length, velvetg command by setting 200 as insert length and 300 as minimum contig length, and meta-velvetg command. Finally, FragGeneScan was employed as describe above for ORF finding both on clustered reads and assembled contigs.

The second database was a selection of taxa on the base of 16S metagenomics analysis (24,350,176 sequences). In detail, all sequences belonging to microbial genera whose relative abundance was >0.1% in at least one of the luminal and mucosal samples were selected and downloaded from UniProtKB database (2017/07 update).

The metaproteomic data (regarding the microbial component of the luminal and mucosal material) were obtained by merging results of searches against the two above cited databases.

In this case, taxonomic and functional annotation was performed using MEGAN (version 6.9) on read sequences subjected to DIAMOND (version 0.8.22) search against the NCBI-nr DB (2016/09 update), using the blastp command with default parameters; subsequently, DIAMOND outputs were loaded on MEGAN in order to perform both LCA classification and InterPro functional annotation with default parameters.

Finally, a third database (UniProtKB taxonomy Ruminantia, 2017/10 update, 115,553 sequences) was employed to achieve information concerning the host.

2.5.4 Results

16S rDNA analysis of pre-weaned lamb digestive microbiota

A total of 1,147,915 reads were obtained after V4 region of the 16S rDNA sequencing, performed in duplicate after full-length 16S amplification on 20 luminal and mucosal samples.

Richness, calculated as the number of detected OTUs, and alpha-diversity (Shannon index) distributions along the gastrointestinal tract of a pre-weaned lamb are showed in Figure 59. As illustrating by the line graphs, the highest values of alpha-diversity and richness were observed in the fore-stomachs, while a remarkable drop was observed in the abomasum and in all the intestinal tracts, with a slow increase of the richness in the large intestine. Concerning the comparison between the luminal and the mucosa-adhered communities, both sample types showed analogous trends.

PCA plot generated starting from the relative abundance of OTUs (Figure 60) showed a clear separation between the fore-stomachs and the other GI tracts based on the first component; in addition, rumen and reticulum were separated from omasum on the second component. Moreover, luminal and mucosal contents were generally more similar between them.

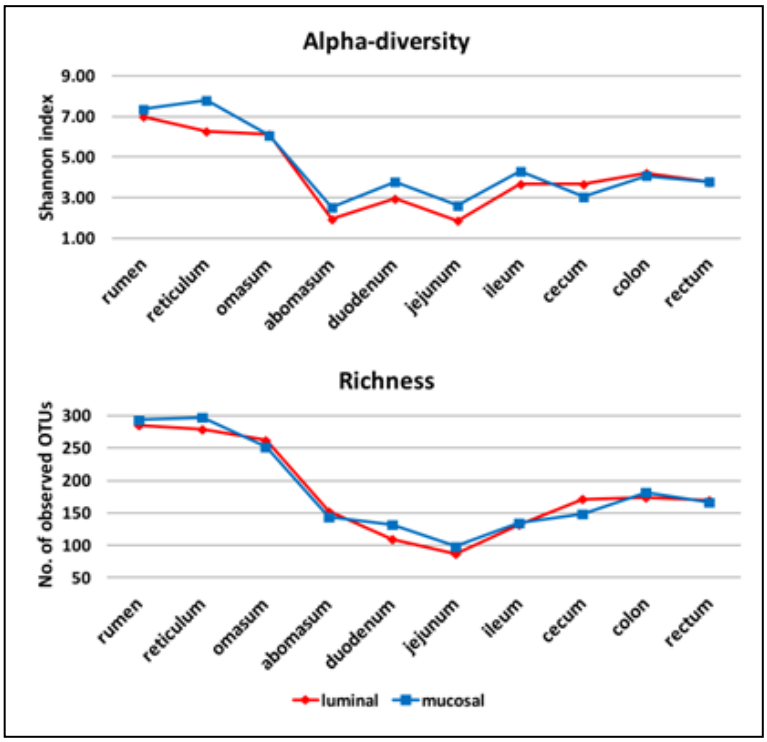

Figure 59: Alpha-diversity and richness distribution along the gastrointestinal tracts of a pre-weaned lamb. Line graphs illustrates Shannon index values (alpha-diversity, top) and the number of observed OTUs (Richness, bottom) observed for each luminal (red) and mucosal (blue) GI sample.

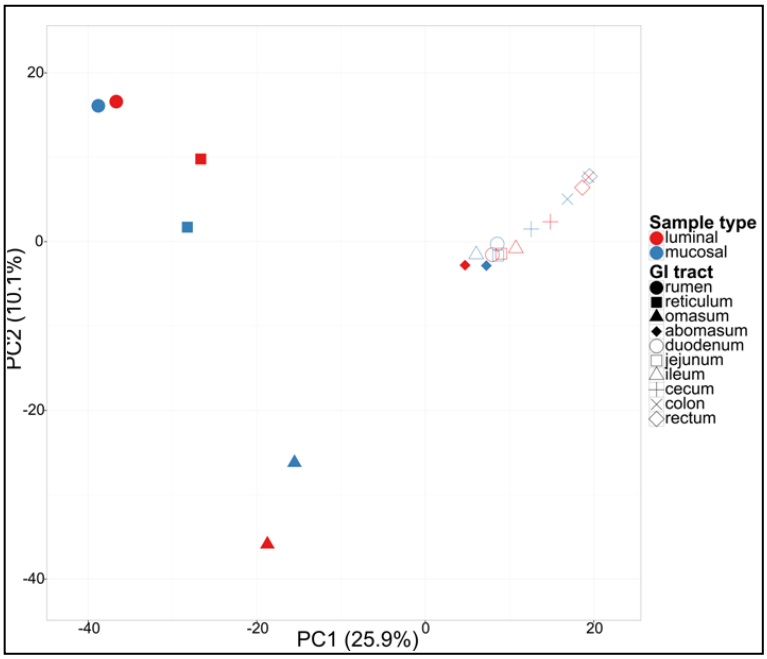

Figure 60: Beta-diversity at OTU level within the gastrointestinal tracts of a pre-weaned lamb. PCA plot was generated starting from the relative abundance of OTUs. Each dot indicate a sample, each shape corresponds to a different gastroinstenal tract, while colours corresponds to luminal (red) and mucosal (blue) samples.

Focusing on taxonomy, important differences were detected along the gastrointestinal tracts. Figure 61 reported the top 20 families identified in all sections and their relative abundances, computed by aggregating OTUs at family level.

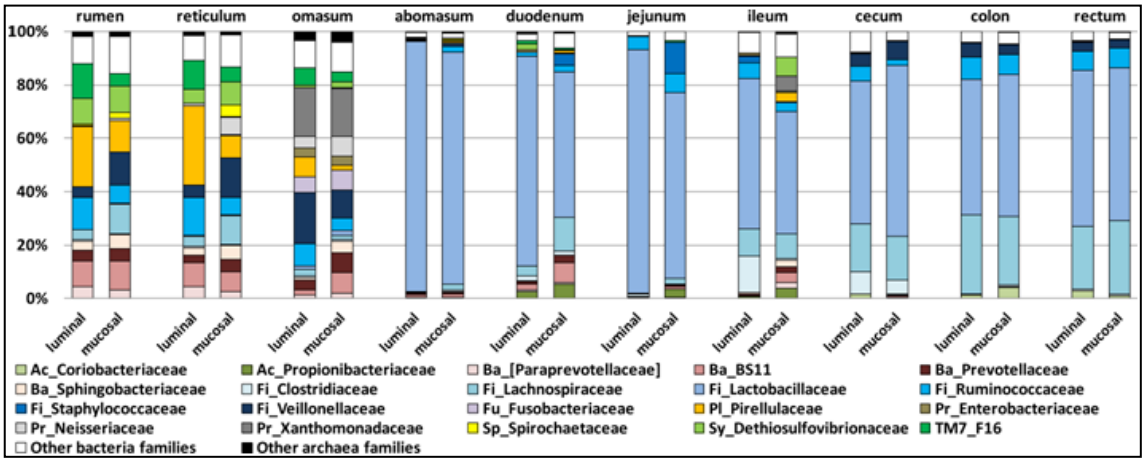

Figure 61: Top 20 microbial families distribution along the gastrointestinal tracts of a pre-weaned lamb. The relative abundance of families identified in all sections according to 16S data is reported. Families are ordered by phyla they belong to. Ac: Actinobacteria; Ba: Bacteroidetes; Fi: Firmicutes; Fu: Fusobacteria; Pl: Planctomycetes; Pr:Proteobacteria; Sp: Spirochaetes; Sy: Synergistetes.

At first glance, a strong divergence was observed between the first three compartments of the lamb stomach and all the other tracts, in accordance with alpha-diversity and richness, and what illustrated in the Figure 60.

Rumen and reticulum showed the presence of high levels of Pirellulaceae (22% and 30%, respectively), Ruminococcaceae (12% and 14%), F16 family belonging to TM7 (13% and 11%), the Bacteroidetes family BS11 (10% and 9%) and Dethiosulfovibrionaceae (10% and 5%) in the lumen, plus some other families with their relative abundance of approximately 3-4%; the respective mucosal communities were quite similar, but with a drop of Pirellulaceae counterbalanced by higher levels of Veillonellaceae (12% in rumen and 15% in reticulum) and Lachnospiraceae (both 11%). Veillonellaceae (19% in the luminal and 11% in the mucosal community) and Xanthomonadaceae (both 18%) were the most abundant families observed in the omasum microbiota, where some families observed in the first two tract were also detected (such as Ruminococcaceae, Pirellulaceae and F16). In general, rumen, reticulum and omasum presented a well assorted microbiota, with several families

overtaking the 3% of their relative abundance; furthermore, similar values of Firmicutes/Bacteroides (F/B) ratio were observed in the luminal and mucosal content of rumen and reticulum (0.77 vs 0.96, and 0.94 vs 1.15), while a higher F/B ratio was detected in the luminal vs the mucosal content of omasum (2.46 vs 0.82). Archaea families were well represented (around 1% and up to 3% in the omasum).

The remaining seven GI tracts showed instead a completely different structure, with the Firmicutes family of Lactobacillaceae that prevailed almost exclusively in the abomasum (94% lumen and 87% mucosa) and slowly decreased along the intestine reaching a minimum of 46% in the ileum mucosal microbiota, and in general remaining around the 50% in the large intestine tracts. This fact affected also the F/B ratio that was totally moved towards the Firmicutes. Other families exhibiting a large abundance were Lachnospiraceae (mucosal duodenum, 12%, ileum, 10%, and large intestine, up to 29% in luminal colon), Clostridiaceae (13% and 8% in the lumen of ileum and cecum) and Ruminococcaceae in colon and rectum (around 8%). Finally, very few Archaea were detected.

Metaproteomics analysis of pre-weaned lamb digestive microbiota

Metaproteomic analysis of the 20 luminal and mucosal samples enabled to obtained a total of 528,927 PSMs, of which 413,295 related to the host and only 115,632 as microbial. As reported in Table 23, in a lot of tracts/samples few PSMs were functionally and taxonomically annotated.

In keeping of this, the microbiota functions were inspected in the tracts presenting a better results in microbial PSMs identification and annotation, i.e. luminal content of rumen, cecum, colon and rectum.

On the whole, several enzymes involved in the glycolysis and pyruvate metabolism, as well as ribosomal proteins, were detected in the top 20 functions (InterPro database) considering the four selected GI tracts (Table 24). Glyceraldehyde/erythrose phosphate dehydrogenase family (including glyceraldehyde-3-phosphate dehydrogenase and D-erythrose-4-phosphate dehydrogenase) was observed as the

most abundant functions in rumen and cecum, and the second most one in colon and rectum, where EF-Tu/EF1A proteins showed the highest abundance.

GI tract	sample type	microbial PSMs	host PSMs	host to microbial PSMs ratio	function	phylum	family	function + phylum	function + familyf
rumen	luminal	11,138	9,154	0.82	3,100	4,888	2,458	1,909	922
	mucosal	3,964	20,527	5.18	446	620	319	331	170
reticulum	luminal	8,126	12,467	1.53	1,956	3,000	1,526	1,193	583
	mucosal	4,588	21,490	4.68	336	539	310	259	183
omasum	luminal	3,197	6,572	2.06	879	1,579	1,101	782	584
	mucosal	5,183	23,357	4.51	604	901	623	521	395
abomasum	luminal	520	3,502	6.73	60	97	90	54	48
	mucosal	4,184	23,852	5.70	215	296	159	178	106
duodenum	luminal	4,700	26,559	5.65	227	340	193	195	111
	mucosal	5,815	30,496	5.24	254	376	203	205	120
jejunum	luminal	4,600	23,578	5.13	351	504	376	308	238
	mucosal	5,382	26,114	4.85	304	474	204	223	123
ileum	luminal	6,058	30,539	5.04	327	510	235	251	143
	mucosal	5,000	26,589	5.32	223	301	162	177	104
cecum	luminal	8,627	15,954	1.85	3,542	4,999	3,598	3,353	2,336
	mucosal	6,070	30,404	5.01	427	547	267	330	170
colon	luminal	9,410	13,453	1.43	4,135	6,152	3,710	3,878	2,113
	mucosal	5,019	27,327	5.44	270	394	218	226	129
rectum	luminal	8,938	12,957	1.45	3,863	5,846	3,612	3,659	2,051
	mucosal	5,113	28,404	5.56	237	344	180	194	108

Table 23: Peptide spectrum matches (PSMs) detected and taxonomic and functional annotated in each GI tract of a pre-weaned lamb.

In general, the latter three gastrointestinal tracts were very similar in terms of abundant functions, while some differences were observed when comparing them with the rumen. As an example, glutamate dehydrogenase overtook the 8% of relative abundance in rumen, whereas was around 2% in the large intestine. Phosphoserine aminotransferase exhibited the same trend, almost disappearing in cecum. On the contrary, rumen was less rich in several ribosomal proteins and a few enzymes of glycolyis, while pyruvate kinase was not identified.

function (InterPro)	rumen	cecum	colon	rectum
Glyceraldehyde/Erythrose phosphate dehydrogenase family	10.10%	6.92%	5.10%	5.80%
Glutamate dehydrogenase	8.32%	1.50%	2.20%	1.71%
Translation elongation factor EFTu/EF1A, bacterial/organelle	7.39%	4.94%	6.22%	6.06%
Pyruvate-flavodoxin oxidoreductase	2.65%	3.90%	4.59%	5.02%
Pyruvate, phosphate dikinase	1.55%	2.68%	4.11%	3.00%
Enolase	2.06%	2.91%	1.50%	1.79%
Fructose-1,6-bisphosphate aldolase, class 2	1.68%	2.17%	2.85%	2.43%
Ribosomal protein L7/L12	2.77%	1.64%	1.91%	2.07%
Triosephosphate isomerase, bacterial/eukaryotic	1.26%	2.00%	2.52%	2.12%
Phosphoglycerate kinase	0.71%	1.98%	2.15%	2.30%
Phosphoenolpyruvate carboxykinase, ATP-utilising	1.84%	2.12%	2.25%	1.89%
Translation elongation factor EFG/EF2	0.52%	1.72%	2.18%	2.17%
Pyruvate kinase	0.00%	1.98%	0.58%	0.83%
Phosphoserine aminotransferase	1.87%	0.06%	0.27%	0.21%
Ribosomal protein S7, bacterial/organellar-type	0.45%	1.69%	1.09%	1.42%
Ribosomal protein S2, bacteria/mitochondria/plastid	0.68%	1.55%	1.38%	1.53%
Ribosomal protein L5, bacterial-type	0.65%	1.55%	1.21%	1.24%
Ribosomal protein S3, bacterial	0.87%	1.44%	1.11%	1.09%
Ribosomal protein S8	0.97%	0.85%	1.23%	1.32%

Table 24: Top 20 microbial functions detected in the lumen of rumen and large intestine of a pre-weaned lamb. The relative abundance of functions identified in all sections according to MP data is reported. Functions are ordered by decreasing of the maximum value observed among the four tracts.

Optimization of protocols for DNA extraction and sample cleanup for the analysis of the sheep fecal microbiota

Five different sample preparation/extraction methods from sheep fecal samples were compared, and 16S rRNA gene amplification efficiency was used as a probe to determine the DNA suitability for downstream metagenomic analysis. In particular, as described in "DNA sample preparation" in Section 2.5.3., samples were subjected to differential centrifugation (DC) or direct lysis (DL); then, DNA was extracted from both preparations after enzymatic and mechanical lysis with QIAamp Fast DNA Stool or with the E.Z.N.A. Soil DNA Kit. Finally, samples were also subjected to DC and then extracted with the standard phenol/chloroform/isoamyl alcohol (25:24:1) method.

As a result, the combination of stool DC preparation followed by DNA extraction and purification with the E.Z.N.A. soil DNA kit, designed to remove with highest efficiency PCR inhibitors, was the only protocol capable of providing a satisfactory quantity (25 ± 0.28 ng/g stool sample) and the best quality of extracted DNA (100% of samples providing a 16S rRNA gene PCR amplification product; data not shown).

Sheep fecal microbiota composition

16S and shotgun metagenomic and shotgun metaproteomic analyses were performed to assess the fecal microbiota composition in sheep. The taxonomic composition of the prokaryotic microbiota according to the three method results is shown in Figure 62.

As expected, Firmicutes and Bacteroidetes made over 80% of total bacteria in all cases. Furthermore, Firmicutes was detected as the most adundant phylum in all animals and with all approaches, followed by Bacteroidetes (Figure 62). However, the average Firmicutes/Bacteroidetes (F/B) ratio ranged from 6.0 for 16S data down to 1.6 for MP, through 3.4 for MG. The archaeal Euryarchaeota was the fifth, seventh and third most abundant phylum according to 16S, MG and MP results, respectively. Whereas Firmicutes levels were very similar among individuals (CV <10% with all approaches), a higher variation could be observed for other important phyla, particularly for Actinobacteria and Verrucomicrobia. These data was in line with the existing studies concerning a metagenomic analysis of fecal samples from other ruminants, that reported a general predominance of Firmicutes over Bacteroidetes in cattle, with Ruminococcaceae and Lachnospiraceae being the most representative microbial families of the former phylum (Durso et al 2010, Durso et al 2011, Kim et al 2014, Shanks et al 2011). On the contrary, Bacteroidetes was the most represented phylum in the sheep rumen according to the literature (Castro-Carrera et al 2014, Lopes et al 2015, Morgavi et al 2015). Consistently, a marked increase in the F/B ratio from rumen to colon was recently described in cattle (Mao et al 2015). Regarding minor phyla, a remarkable amount of functionally active Spirochaetes was found, principally belonging to the genus *Treponema*, especially *T. saccharophilum*, previously described as a large pectinolytic spirochete present in the rumen (Paster and Canale-Parola 1985). Moreover, this phylum was recently described as the fourth most represented within the microbiota of the ruminant digestive tract (Peng et al 2015).

Finally, fungal taxa were also identified with MG and MP approaches, accounting for about 0.05% and 1.1% of the microbiota, respectively. Ascomycota was detected in both cases as the most abundant fungal phylum.

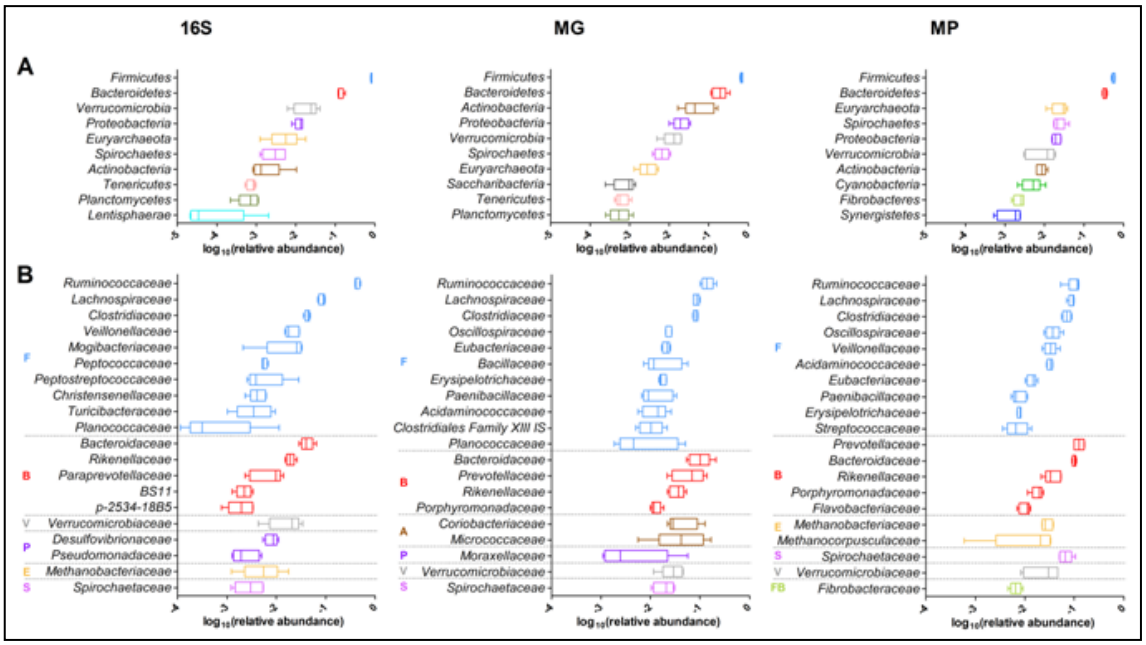

Figure 62: Taxonomic composition of the sheep fecal prokaryotic microbiota, according to V4-16S rRNA (16S, left), metagenomic (MG, center) and metaproteomic (MP, right) results (Tanca et al 2017b). A) Tukey's box plot showing the microbiota composition at phylum level. The top 10 phyla are shown, ordered by decreasing mean relative abundance. B) Tukey's box plot showing the microbiota composition at family level. The top 20 families are shown, grouped based on the relative phylum (A, Actinobacteria; B, Bacteroidetes; E, Euryarchaeota; F, Firmicutes; FB, Fibrobacteres; P, Proteobacteria; S, Spirochaetes; V, Verrucomicrobia) and further ordered by decreasing mean relative abundance.

Going down to the family level, 16S, MG and MP allowed for the identification of 76, 385 and 171 different microbial families, respectively. Considering the 'core microbiota' (i.e., taxa found consistently in all the analyzed animals), a total of 45, 168 and 50 microbial families were found after 16S, MG and MP analyses, respectively, of which 19 in common between the three techniques. As illustrated in Figure 62B Ruminococcaceae, Lachnospiraceae and Clostridiaceae were consistently found with all approaches to be the first, second and third most abundant Firmicutes family, respectively, according to what observed in cows (Durso et al 2010, Durso et al 2011, Kim et al 2014, Shanks et al 2011). Families belonging to Bacteroidetes exhibited instead more variable distributions, with Prevotellaceae becoming the principal family overall when considering protein expression data. Families from other phyla, such as Spirochaetaceae from Spirochaetes and Verrucomicrobiaceae from Verrucomicrobia, were also detected among the top 20 abundant families with all approaches.

Assessment of functions potentially and actively expressed by the sheep fecal microbiota

Figure 63A shows the 20 most abundant genes and protein families identified upon MG and MP analyses, respectively. In the metagenome several genes related to membrane transport of molecules (ABC transporter, ATPase, permease, SecA), DNA replication and repair (helicase, DNA polymerase, topoisomerase, MutS), transcription (RNA polymerase), translation (tRNA synthetases and translation factors), and protein folding (chaperones) were detected, plus a few encoding for metabolic enzymes.

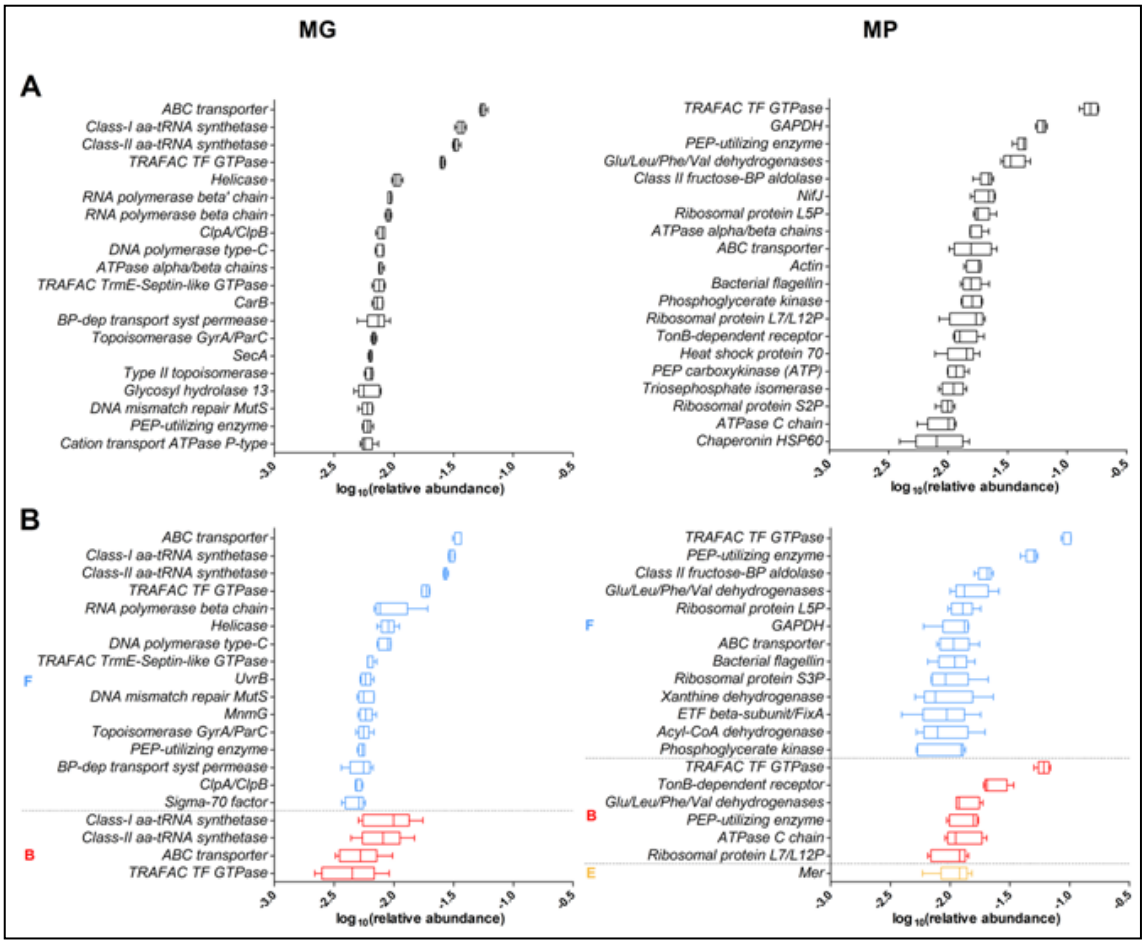

Figure 63: Functional potential and activity of the sheep fecal microbiota, as measured by metagenomics (MG, left) and metaproteomics (MP, right), respectively (Tanca et al 2017b). A) Tukey's box plot showing the 20 most abundant gene (left) and protein families. B) Tukey's box plot showing the 20 most abundant gene family-phylum (left) and protein family-phylum (right) combinations, grouped based on the relative phylum (B, Bacteroidetes; E, Euryarchaeota; F, Firmicutes) and further ordered by decreasing mean relative abundance.

On the contrary, the metaproteome was found to be rich in functions related to metabolism (8 enzymes), in particular carbohydrate degradation, followed by protein synthesis and folding (translation factor, ribosomal proteins, chaperones). Protein families involved in transport and signaling, such as ABC transporters, TonB-dependent receptor, ATPases, flagellin, were also present. On the whole, 2097 gene families and 441 protein families were identified by MG and MP, respectively. Considering the 'core functions' (namely, associated to the microbiota of all the analyzed animals), MG and MP analyses allowed for the detection of a total of 904 and 191 functional families, respectively, of which 152 were identified both as functional potential and expressed proteins.

Focusing on function-taxonomy combinations at the phylum level (Figure 63B), a higher abundance of functions encoded/expressed by Firmicutes was found compared to those encoded/expressed by Bacteroidetes, consistently with the taxonomic information. According to MP data, some functions from other phyla (i.e., from Euryarchaeota) were also noticeably expressed. More interestingly, the most represented genes assigned to Firmicutes according to MG results were in most cases the same of the most represented genes assigned to Bacteroidetes, while, when considering the main expressed protein functions, several of them were not overlapping between these two phyla. To further investigate this "phylum-specific" functional contribution, those protein families exclusively and unambiguously assigned to a specific phylum, and detected in all samples, were sought (42 in total).

As illustrated in Table 25, several relevant and considerably abundant protein functions were actually phylum-specific, while phylum-specific 'core' MG functions (N = 60) were all detected at very low abundance. Among the detected protein functions, the presence of the TonB-dependent receptor as specific for Bacteroidetes, the enzyme 5,10-methylenetetrahydromethanopterin reductase (Mer) involved in methanogenesis for Euryarchaeota, in addition to aldehyde oxidoreductase (belonging to the xanthine dehydrogenase family), formate-tetrahydrofolate ligase and carbon monoxide dehydrogenase (both involved in one-carbon metabolism) for Firmicutes were noted. Among other expressed protein functions of possible interest flagellins

could be cited, belonging mainly to Firmicutes (*Clostridium* and *Selenomonas* genera) and Spirochaetes (again, essentially *Treponema*).

Protein family	Phylum	*mean*	*CV%*
TonB-dependent receptor	Bacteroidetes	***2.40%***	***26.08%***
Group II decarboxylase	Bacteroidetes	***0.26%***	***50.81%***
Ribosomal protein S1P	Bacteroidetes	***0.26%***	***37.56%***
NagA	Bacteroidetes	***0.16%***	***43.86%***
GHMP kinase	Bacteroidetes	***0.12%***	***44.78%***
ExbB/TolQ	Bacteroidetes	***0.11%***	***53.80%***
Class-I fumarase	Bacteroidetes	***0.09%***	***51.93%***
Gfo/Idh/MocA	Bacteroidetes	***0.06%***	***61.38%***
Eukaryotic mitochondrial porin	Basidiomycota	***0.05%***	***38.84%***
RuBisCO large chain	Cyanobacteria	***0.21%***	***47.87%***
Reaction center PufL/M/PsbA/D	Cyanobacteria	***0.08%***	***39.93%***
Mer	Euryarchaeota	***1.16%***	***29.77%***
[NiFe]/[NiFeSe] hydrogenase large subunit	Euryarchaeota	***0.56%***	***68.70%***
MTD	Euryarchaeota	***0.18%***	***50.39%***
MtrA	Euryarchaeota	***0.18%***	***40.18%***
FrhB	Euryarchaeota	***0.14%***	***52.48%***
Archaeal histone HMF	Euryarchaeota	***0.08%***	***34.82%***
Ribosomal protein L12P	Euryarchaeota	***0.07%***	***52.38%***
N-Me-Phe pilin	Fibrobacteres	***0.06%***	***55.55%***
Xanthine dehydrogenase	Firmicutes	***1.07%***	***66.76%***
Formate--tetrahydrofolate ligase	Firmicutes	***0.66%***	***21.79%***
Ni-containing carbon monoxide dehydrogenase	Firmicutes	***0.53%***	***25.85%***
ETF alpha-subunit/FixB	Firmicutes	***0.45%***	***29.82%***
Complex I 51 kDa subunit	Firmicutes	***0.40%***	***38.67%***
FldB/FldC dehydratase beta subunit	Firmicutes	***0.38%***	***75.26%***
Acetyl-CoA hydrolase/transferase	Firmicutes	***0.27%***	***16.74%***
Elongation factor P	Firmicutes	***0.24%***	***43.65%***
Diol/glycerol dehydratase small subunit	Firmicutes	***0.22%***	***44.07%***
Glycosyltransferase 1	Firmicutes	***0.16%***	***37.55%***
Glycosyl hydrolase 101	Firmicutes	***0.13%***	***60.81%***
Glyoxalase I	Firmicutes	***0.13%***	***31.50%***
Bacterial solute-binding protein 1	Firmicutes	***0.13%***	***28.51%***
Acyl-CoA mutase large subunit	Firmicutes	***0.12%***	***8.25%***
Diol/glycerol dehydratase medium subunit	Firmicutes	***0.11%***	***29.30%***
V-ATPase proteolipid subunit	Firmicutes	***0.10%***	***50.73%***
Glutamine synthetase	Firmicutes	***0.08%***	***33.96%***
Aldolase class II	Firmicutes	***0.08%***	***57.15%***
GSP E	Firmicutes	***0.04%***	***25.28%***
Hfq	Firmicutes	***0.04%***	***25.28%***
Peptidase S41A	Planctomycetes	***0.05%***	***38.22%***
Resistance-nodulation-cell division	Proteobacteria	***0.04%***	***25.28%***

Table 25: Protein families assigned exclusively to a single phylum and detected in all samples. Functions are ordered by phylum and then by mean percentage abundance. The coefficients of variation (percentage) are also reported.

Remarkably, half of the functionally annotated peptides identified assigned to Spirochaetes were from flagellar proteins. Members of this phylum are indeed known to have long flagella, enclosed in the periplasm and capable to confer them a unique motility (Wolgemuth 2015). Furthermore, SASPs (small, acid-soluble spore proteins), mainly of clostridial origin, were identified in 4 out of 5 samples, indicating the presence of endospores within the fecal microbiota of these sheep.

Microbial metabolic pathways in sheep gut

Gene and protein functional data were additional grouped according to the UniProt 'pathway' annotation.

The main metabolic pathways potentially and actively functioning in the sheep fecal microbiota based on MG and MP data, respectively, are shown in Figure 64A. Metagenome and metaproteome were both highly represented by several amino acid, nucleoside and carbohydrate biosynthetic routes. However, whereas most of the main MG functions were involved in biosynthetic pathways, MP data, hence concerning active functionalities, were mainly related to catabolic activities. Moreover, enzymes involved in methanogenesis were considerably more abundant than expected, considering their gene content as assessed by MG analysis.

Figure 64B illustrates pathway-taxonomy combinations at phylum level, revealing that several different phyla actually participated to the metabolism at similar extents, based on MP results, although the large majority of pathways covered in the metagenome were associated to Firmicutes. Phylum-specific pathways were also detected, i.e. butanoate metabolism and 1,2-propanediol degradation as specific for Firmicutes, starch degradation for Bacteroidetes, and methanogenesis for Euryarchaeota.

Then carbon metabolism was investigated by mapping the identified proteins into the corresponding KEGG pathway, in order to inspect which microbial players were mainly involved in each specific enzymatic step. While glycolysis was revealed to be performed in parallel by several different phyla, other metabolic functions within the

sheep fecal microbiota were exerted in a phylum-specific fashion, as clearly illustrated in Figure 65.

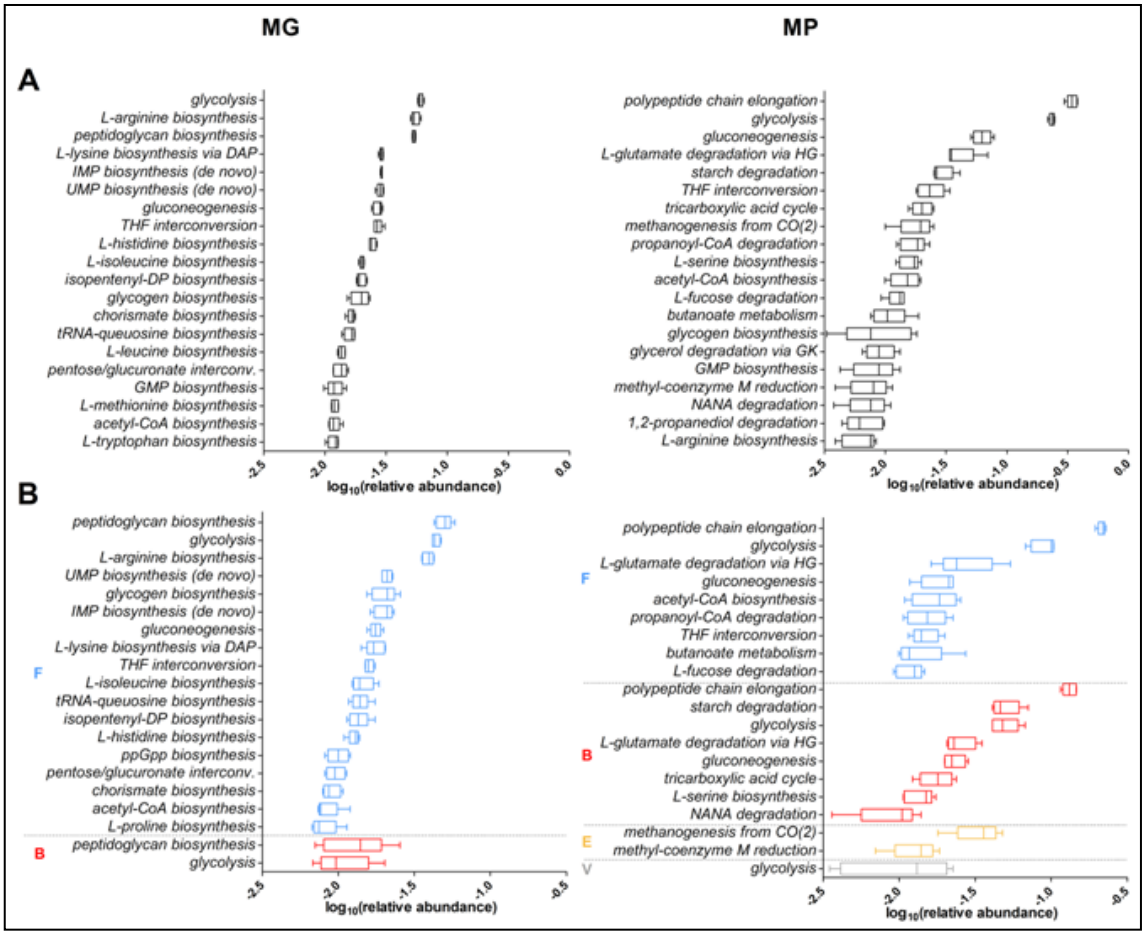

Figure: 64. Metabolic pathway potential and activity of the sheep fecal microbiota, as measured by metagenomics (MG, left) and metaproteomics (MP, right), respectively (Tanca et al 2017b). A) Tukey's box plot showing the 20 most relevant pathways, based on the related gene (left) and protein (right) abundance. B) Tukey's box plot showing the 20 most relevant pathway-phylum combinations, based on the related gene (left) and protein (right) abundance. Pathways are grouped based on the relative phylum (B, Bacteroidetes; E, Euryarchaeota; F, Firmicutes; V, Verrucomicrobia) and further ordered by decreasing mean relative abundance.

In the first place, the unique contribution of Archaea to methanogenesis was evident, as expected. Due to zootechnical, environmental and ecological reasons, the characterization and monitoring of methanogenesis in ruminants is receiving growing attention (Kumar et al 2014, Shi et al 2014). Furthermore, has been reported that diets have an effect on the activity of methanogens in lambs cecum, but not in the rumen, suggesting a possible compensation of rumen methane production with cecum methanogenesis (Popova et al 2013). In this study, the methanogenic route (from 5,10-

methenyltetrahydromethanopterin to methane, including those responsible for methyl-coenzyme M reduction) was almost entirely reconstructed in the microbiota of all animals analyzed based on MP data, and the corresponding enzymatic functions were assigned specifically to *Methanobrevibacter ruminantium* (belonged to Methanobacteriaceae family), and *Methanocorpusculum labreanum* (Methanocorpusculaceae) families.

Moreover, the acetogenic Wood-Ljungdahl pathway (from carbon dioxide to acetate, including the tetrahydrofolate interconversion steps) was found to be entirely covered by Firmicutes members, mainly Clostridiales. This pathway is used by acetogens to convert hydrogen and carbon dioxide into acetic acid, and its pivotal importance is related to the oxidation of the hydrogen generated during the fermentation of dietary macromolecules (Koropatkin et al 2012). In this study, all enzymatic players involved in this pathway were identified, supporting its key relevance within microbial metabolism in sheep colon. More specifically, a high level of conservation within the Firmicutes members were observed concerning most enzymes involved in the tetrahydrofolate interconversion steps, as the majority were taxonomically assigned not lower than the phylum level. On the contrary, the key players of the last two reactions (from acetyl-CoA to acetate, consecutively catalyzed by phosphate acetyltransferase and acetate kinase) could be identified as being members of Clostridiales (including Lachnospiraceae and Oscillospiraceae), although a few peptides were assigned to *Bacteroides* and *Prevotella* genera from Bacteroidetes.

In addition, Bacteroidales were found to be involved in most steps of the tricarboxylic acid cycle, consistently with a recent work on the effects of diet-induced obesity on the mouse gut microbiota (Denou et al 2016), while the enzyme RuBisCo, as expected, was detected only as expressed by photosynthetic Cyanobacteria. Finally, the galactokinase, one key enzyme in the galactose metabolism, was found to be associated to Bacteroidetes members only. This enzyme was recently demonstrated as essential for a *Bacteroides* species to accomplish early colonization of the colonic microbiota (Yaung et al 2015).

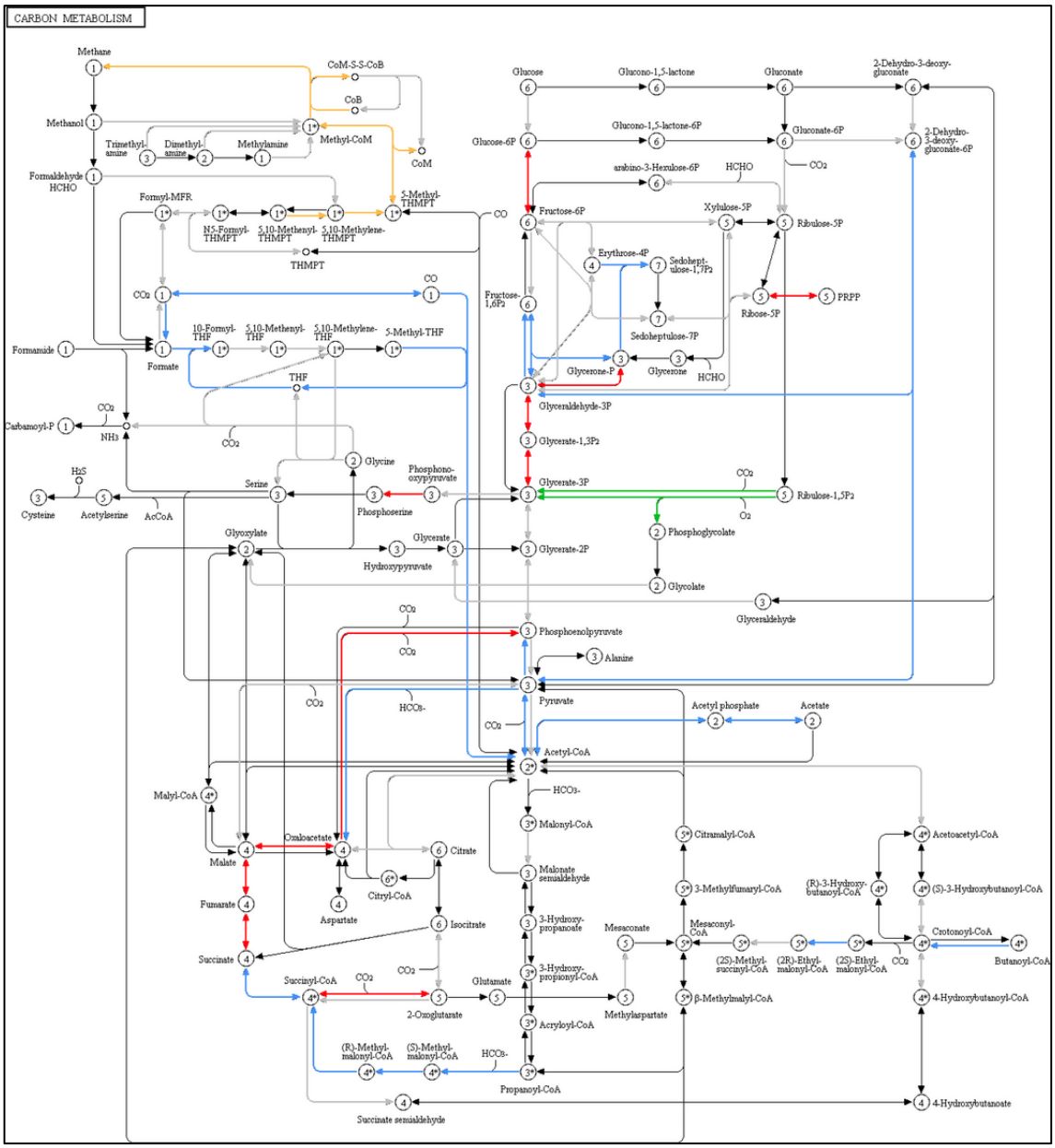

Figure 65: Enzymatic functions identified by metaproteomics and mapped in the KEGG carbon metabolism pathway. Colored arrows indicate enzymes detected in all animals, with the color corresponding to the main phylum to which the function was assigned (red, Bacteroidetes; blue, Firmicutes; orange, Euryarchaeota; green, Cyanobacteria). Grey arrows indicate enzymes detected in at least one but not all animals, or not assigned unambiguously to at least one phylum.

Focus on glycan import and degradation: ABC transporters, starch utilization system and glycosyl hydrolases

The functions responsible for import and degradation of glycans were also evaluated, in view of the high content in glycans in the plant-based sheep diet, and to investigate

the (residual) relevance of such activities in the sheep colon after the massive digestion of plant material carried out in the rumen.

Table 2 lists genes and proteins classified as ABC transporters, especially those capable of transporting mono- and oligosaccharides. These comprise both generic multiple sugar transport systems, such as msmX, and more specific transporters, carrying mono- and disaccharides as ribose and maltose. In most cases, consistently with both MG and MP data, these functions were related to Firmicutes members (several different Clostridiales genera), followed by Actinobacteria (in particular Actinomycetales) for MG and Spirochaetes (mainly *Treponema*) for MP.

Transported molecule	Identified gene(s)	Associated phyla	Identified protein(s)	Associated phyla
aldouronate	*lplB, lplC*	Firmicutes	lplA	Firmicutes
alpha-glucoside	*aglK*		aglK	Spirochaetes
arabinogalactan	*ganQ*	Firmicutes		
arabinosaccharide	*araQ*	Firmicutes		
D-allose	*alsA*			
D-xylose	*xylG*			
L-Arabinose	*araG*			
maltose/maltodextrin	*malK*		malK	Firmicutes
methyl-galactoside	*mglA*	Firmicutes	mglB	Firmicutes
multiple mono- and oligosaccharides	*msmX*	Firmicutes, Actinobacteria	msmX	Firmicutes
multiple oligosaccharides	*gguA, gguS*	Firmicutes, Actinobacteria	gguS	Firmicutes
myo-Inositol	*iatA*	Firmicutes		
rhamnose	*rhaT*	Firmicutes		
ribose/D-xylose	*rbsA, rbsC*	Firmicutes, Actinobacteria	rbsB	Firmicutes
sn-Glycerol-3-phosphate	*ugpC*	Firmicutes, Actinobacteria	ugpC	Firmicutes

Table 26: Carbohydrate ABC transporter genes and proteins identified in the fecal microbiota of all sheep by metagenomics and metaproteomics, respectively.

As mentioned above, a large amount of peptides were assigned to the Bacteroidetes-specific TonB-dependent receptor family, namely to the starch utilization system (Sus) protein C, essential for complex carbohydrate degradation in Bacteroidetes (Martens et al 2009, Reeves et al 1997). The Sus proteins, located in the periplasm and the outer membrane, binds the starch to the cell surface, subsequently degrades it into oligosaccharides which are transported into the periplasmic space, where they are further digested into simpler sugars and imported into the bacterial cell (Koropatkin et al 2012). Noteworthy, the percentage of identified peptides functionally assigned to the Ton-B dependent receptor family was 20-fold higher than the percentage of the

corresponding genes sequenced, indicating a very strong expression rate for this gene family.

Finally, attention was focused on glycosyl hydrolases (GHs), due to their key role in degradation of plant biomass. A total of 56 different GH families were detected in the metagenome, of which 28 were found in all animals. The most represented family was GH 13, mainly composed by alpha-amylases, with 70% of genes belonging to Firmicutes; then, GH 3 (mainly beta-glucosidases, 57% from Firmicutes and 41% from Bacteroidetes), GH 2 (mainly beta-galactosidases, assigned at 51% to Firmicutes and 44% to Bacteroidetes) and GH 51 (64% Firmicutes and 31% Bacteroidetes) were found. Conversely, three GH families were found in the metaproteome of all animals, namely GH 101 (clostridial endo-alpha-N-acetylgalactosaminidase), GH 13 (pullulanase from Lachnospiraceae) and GH 94 (cellobiose phosphorylase, assigned to various phyla). Interestingly, the relative abundance of GH peptides identified was about 6-fold lower than that of the corresponding genes sequenced, indicating a likely poor expression of this functional gene family, maybe because of the relatively low amount of complex (and still undigested) polysaccharides that reach the colon after the extensive degradation occurred within the upper tracts of the ruminant digestive system (Huntington et al 2006).

Non-microbial components detected in the fecal material through metaproteomics

The mass spectra generated in this study were also searched against a "generic" database (without taxonomic filters towards microbial sequences) to achieve information about all organisms contained in the fecal samples. As a result, it was found that, in addition to spectra of microbial origin (about 30% of the total), approximately 45% of spectra were assigned to the host (phylum Chordata), while 13% could be attributed to plant material (with those assigned to Poaceae and Fabaceae families accounting for about 70% of them). Less than 1% of identified peptides were classified as belonging to further eukaryotic phyla as – in decreasing abundance order – Arthropoda, Mollusca, Nematoda, Annelida, and Platyhelminthes (including the *Fasciola* genus, comprising parasites of the small ruminant intestine). Even more

interestingly, peptides from several protists known for intestinal tropism in sheep were also detected, such as the genera *Entamoeba*, *Blastocystis*, *Andalucia*, *Giardia* and the ciliate protozoan *Entodinium* with a well known capability to digest and ferment starch, producing SCFAs (Belzecki et al 2013).

In addition, based on MG data, a few sequences assigned to *Haemonchus* were detected in four animals (about 0.005% of annotated reads; an identified peptide sequence was also assigned to this genus). Of note, *Haemonchus contortus* has been described as the most important gastroenteric nematode of sheep in many regions of the world, causing haemonchosis (Getachew et al 2007).

2.6 Conclusion and perspectives

In this chapter, the interaction between diet and the gut microbiota was investigated in humans and an economically relevant livestock species, namely sheep.

As already mentioned, planning experiments inspecting the role of the gut microbiota and its association/variation in response to different dietary habits is not trivial, both for humans and non-model animals, as well as the subsequent data analysis.

As a huge inter-individual variability is expected, the collection of subjects' metadata is mandatory to stratify the population for data analysis.

In view of this, the human and sheep populations under study were firstly subjected to gut microbiota characterization both taxonomically and functionally.

The integrated metaproteogenomic analysis led to take a detailed picture of the main potential and active functionalities and the major actors involved, up to the characterization of taxon-specific expressed functions. These data allowed, for instance, the identification of the most conserved (such as *Faecalibacterium* and the butyrate biosynthesis), and variable (such as *Akkermansia*, *Prevotella* and *Bifidobacterium* genera, and oxidative stress-related functions) features within the human healthy gut microbiota, supporting the former as a key requirement for intestinal health, while the latter as possibly highly responsive to variables like diet or other environmental factors. In addition, the microbial carbon metabolism was thoroughly reconstructed, as well as its major players, providing useful insights into the catabolism and cross-feeding networks. Concerning the ovine, the first multi-omic characterization of its fecal microbiota was reported, in addition to the description of the luminal and mucosa-associated microbiotas along the gastrointestinal tract of a pre-weaned lamb. These findings could be the starting point for further studies investigating the relationship between the sheep gut microbiota dynamics and dietary variables, with the aim of improving the wealth and, consequently, the productivity of animals.

Finally, two pioneering studies investigating the gut microbiota-diet interaction were carried out on two human populations. The first regarded a highly homogeneous healthy cohort, whose microbiota was studied based on dietary habits through the

collection of nutrition diaries; the second was a case-control study where type 2 diabetes patients and healthy individuals were subjected to a dietary intervention with sourdough-leavened bread, which had been demonstrated to ameliorate clinical conditions in a cohort of a group of T2D patients. The data collected from the two studies are not still definitive and need further analysis, for instance metaproteomics, currently underway for the T2D study. Although the small population examined, some preliminary indications could be reported, as the important divergence observed in the microbiota genetic set of low- and high-fiber consumers, or the increase of the propionate producer *Dialister* after 2 months of sourdough-leavened bread intervention in T2D patients. However, future in-depth analyses, extended to a larger cohort, are required to validate both functional and taxonomic information.

Conclusion and perspectives

In the last years, the huge number of researches investigating the gut microbiota has bumped into the growing interest in personalized medicine and nutrition. As well known, a "healthy" gut microbiota is a sign of a healthy individual, as the microbial community has a role in immune system development, metabolism, and protection from enteropathogen invasion. The identification and characterization of the structure and functionalities of the health-associated gut microbiota is of paramount importance and will be essential for formulating targeted therapeutic interventions. On the other hand, as diet is one of the major factors that influence the intestinal microbial communities, new insights concerning the mechanisms through which diet is able to modulate the gut milieu are also crucial, in order to utilize food as an alternative to drugs in the therapy of dysbiotic disease states.

In view of this, several studies performed to shed light on diet-gut microbiota interactions were reported here.

First, the in-depth characterization of the fecal microbiota of healthy humans and sheep was carried out, improving information related to gut microbiota taxonomy and active functionalities, as well as highlighting taxon-specific functions and cross-feeding reactions that occur in the carbon metabolism pathways in the intestine. In addition, concerning the human gut microbiota characterization, also the most variable and conserved features were identified, the latter with the potentiality to become, after further investigation, possible biomarkers of intestinal health.

Second, two different animal models were employed in order to evaluate the effect of caloric restriction feeding on the microbiota, and the involvement of gut microbial communities in the celiac disease onset and progression. The first model was revealed as suitable for studying the intestinal variation caused by dietary treatments, and several changes due to the sole amount of food were observed, such as the blooming of *Lactobacillus spp.* in the caloric restriction-treated animals, as well as the simultaneous increase of the enzymes involved in propionogenesis, evidences that might be soon investigated also in human. Concerning the celiac disease study, the

sole 16S data did not allow for the identification of microbiota changes clearly related to the disorder, and further functional investigations are needed.

Lastly, two pilot studies focused on the diet were performed in human, the first based on the analysis of different diet habits through the collection of dietary diaries, the second as an actual dietary intervention. Although preliminary and regarding a small population, data obtained from these two studies could pave the way for further in-depth analysis and studies on a wider population.

On the whole, these data confirmed the integrated metaproteogenomics, and in general the use of multi-omic techniques, as one of the most appropriate approach in order to deeply characterize the composition and the metabolic activities occurring in the gut microbiota.

Indeed, when taken individually, metagenomics (16S and shotgun) and metaproteomics showed intrinsic limitations since they allow for obtaining deep and specific information but restricted to few aspects, such as a detailed microbiota structure for 16S, the microbial gene assortments for shotgun metagenomics, and the microbial and host active functions occurring in the intestine. Although new bioinformatics tools can predict the potential functionalities from the taxonomy or reconstruct the microbial composition based on the peptide sequence, an experimental pipeline that integrates different approaches is recommended in order to obtain a more reliable understanding of the gut microbiota. In addition, as metaproteomics requires suitable databases for peptide identification, the availability of incomplete metagenomic sequences from the community being studied represents a major constraint for a satisfactory protein detection.

To fully characterize structure and function of a complex microbial community, hence, the main challenges to be tackled in the near future concern bioinformatic tasks, including i) sequence reconstruction and functional annotation of the majority of gut microbial genomes, being currently insufficiently or absolutely not characterized; ii) appropriate and uniform sequence and functional databases; and iii) standardized and widely approved pipelines for the analysis of the gut microbiota.

References

Abram F (2015). Systems-based approaches to unravel multi-species microbial community functioning. Comput Struct Biotechnol J 13: 24-32.

Addis MF, Tanca A, Uzzau S, Oikonomou G, Bicalho RC, Moroni P (2016). The bovine milk microbiota: insights and perspectives from -omics studies. Mol Biosyst 12: 2359-2372.

Ahern PP, Faith JJ, Gordon JI (2014). Mining the human gut microbiota for effector strains that shape the immune system. Immunity 40: 815-823.

Ahrens CH, Brunner E, Qeli E, Basler K, Aebersold R (2010). Generating and navigating proteome maps using mass spectrometry. Nat Rev Mol Cell Biol 11: 789-801.

Al-Asmakh M, Zadjali F (2015). Use of Germ-Free Animal Models in Microbiota-Related Research. J Microbiol Biotechnol 25: 1583-1588.

Al-Masaudi S, El Kaoutari A, Drula E, Al-Mehdar H, Redwan EM, Lombard V et al (2017). A Metagenomics Investigation of Carbohydrate-Active Enzymes along the Gastrointestinal Tract of Saudi Sheep. Front Microbiol 8: 666.

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

Anders S, McCarthy DJ, Chen Y, Okoniewski M, Smyth GK, Huber W et al (2013). Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. Nat Protoc 8: 1765-1786.

Apajalahti JH, Sarkilahti LK, Maki BR, Heikkinen JP, Nurminen PH, Holben WE (1998). Effective recovery of bacterial DNA and percent-guanine-plus-cytosine-based analysis of community structure in the gastrointestinal tract of broiler chickens. Appl Environ Microbiol 64: 4084-4088.

Armbrecht HJ, Strong R, Boltz M, Rocco D, Wood WG, Richardson A (1988). Modulation of age-related changes in serum 1,25-dihydroxyvitamin D and parathyroid hormone by dietary restriction of Fischer 344 rats. J Nutr 118: 1360-1365.

Arthur JC, Perez-Chanona E, Muhlbauer M, Tomkovich S, Uronis JM, Fan TJ et al (2012). Intestinal inflammation targets cancer-inducing activity of the microbiota. Science 338: 120-123.

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

Asshauer KP, Wemheuer B, Daniel R, Meinicke P (2015). Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data. Bioinformatics 31: 2882-2884.

Backhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI (2005). Host-bacterial mutualism in the human intestine. Science 307: 1915-1920.

Bae YJ, Kim SE, Hong SY, Park T, Lee SG, Choi MS et al (2016). Time-course microarray analysis for identifying candidate genes involved in obesity-associated pathological changes in the mouse colon. Genes Nutr 11: 30.

Belzecki G, Miltko R, Kwiatkowska E, Michalowski T (2013). The ability of rumen ciliates, Eudiplodinium maggii, Diploplastron affine, and Entodinium caudatum, to use the murein saccharides. Folia Microbiol (Praha) 58: 463-468.

Berer K, Gerdes LA, Cekanaviciute E, Jia X, Xiao L, Xia Z et al (2017). Gut microbiota from multiple sclerosis patients enables spontaneous autoimmune encephalomyelitis in mice. Proc Natl Acad Sci U S A.

Bergamo P, Maurano F, Mazzarella G, Iaquinto G, Vocca I, Rivelli AR et al (2011). Immunological evaluation of the alcohol-soluble protein fraction from gluten-free grains in relation to celiac disease. Mol Nutr Food Res 55: 1266-1270.

Bodinham CL, Smith L, Wright J, Frost GS, Robertson MD (2012). Dietary fibre improves first-phase insulin secretion in overweight individuals. PLoS One 7: e40834.

Brilhante RS, Silva ST, Castelo-Branco DS, Teixeira CE, Borges LC, Bittencourt PV et al (2015). Emergence of azole-resistant Candida albicans in small ruminants. Mycopathologia 180: 277-280.

Buchfink B, Xie C, Huson DH (2015). Fast and sensitive protein alignment using DIAMOND. Nat Methods 12: 59-60.

Cabrera-Rubio R, Collado MC, Laitinen K, Salminen S, Isolauri E, Mira A (2012). The human milk microbiome changes over lactation and is shaped by maternal weight and mode of delivery. Am J Clin Nutr 96: 544-551.

Caesar R, Tremaroli V, Kovatcheva-Datchary P, Cani PD, Backhed F (2015). Crosstalk between Gut Microbiota and Dietary Lipids Aggravates WAT Inflammation through TLR Signaling. Cell Metab 22: 658-668.

Caminero A, Galipeau HJ, McCarville JL, Johnston CW, Bernier SP, Russell AK et al (2016). Duodenal Bacteria From Patients With Celiac Disease and Healthy Subjects Distinctly Affect Gluten Breakdown and Immunogenicity. Gastroenterology 151: 670-683.

Candela M, Biagi E, Soverini M, Consolandi C, Quercia S, Severgnini M et al (2016). Modulation of gut microbiota dysbioses in type 2 diabetic patients by macrobiotic Ma-Pi 2 diet. Br J Nutr 116: 80-93.

Cani PD, Neyrinck AM, Fava F, Knauf C, Burcelin RG, Tuohy KM et al (2007). Selective increases of bifidobacteria in gut microflora improve high-fat-diet-induced diabetes in mice through a mechanism associated with endotoxaemia. Diabetologia 50: 2374-2383.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK et al (2010). QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7: 335-336.

Caporaso JG, Lauber CL, Costello EK, Berg-Lyons D, Gonzalez A, Stombaugh J et al (2011a). Moving pictures of the human microbiome. Genome Biol 12: R50.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ et al (2011b). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci U S A 108 Suppl 1: 4516-4522.

Carberry CA, Kenny DA, Han S, McCabe MS, Waters SM (2012). Effect of phenotypic residual feed intake and dietary forage content on the rumen microbial community of beef cattle. Appl Environ Microbiol 78: 4949-4958.

Carvajal-Rodriguez A, de Una-Alvarez J (2011). Assessing significance in high-throughput experiments by sequential goodness of fit and q-value estimation. PLoS One 6: e24700.

Castro-Carrera T, Toral PG, Frutos P, McEwan NR, Hervas G, Abecia L et al (2014). Rumen bacterial community evaluated by 454 pyrosequencing and terminal restriction fragment length polymorphism analyses in dairy sheep fed marine algae. J Dairy Sci 97: 1661-1669.

Chambers ES, Viardot A, Psichas A, Morrison DJ, Murphy KG, Zac-Varghese SE et al (2015). Effects of targeted delivery of propionate to the human colon on appetite regulation, body weight maintenance and adiposity in overweight adults. Gut 64: 1744-1754.

Chen W, Liu F, Ling Z, Tong X, Xiang C (2012). Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer. PLoS One 7: e39743.

Choi YS, Goto S, Ikeda I, Sugano M (1988). Age-related changes in lipid metabolism in rats: the consequence of moderate food restriction. Biochim Biophys Acta 963: 237-242.

Conlon MA, Bird AR (2014). The impact of diet and lifestyle on gut microbiota and human health. Nutrients 7: 17-44.

Consortium HMP (2012). A framework for human microbiome research. Nature 486: 215-221.

Consortium UniProt (2015). UniProt: a hub for protein information. Nucleic Acids Res 43: D204-212.

Corfe BM, Majumdar D, Assadsangabi A, Marsh AM, Cross SS, Connolly JB et al (2015). Inflammation decreases keratin level in ulcerative colitis; inadequate restoration associates with increased risk of colitis-associated cancer. BMJ Open Gastroenterol 2: e000024.

Cosorich I, Dalla-Costa G, Sorini C, Ferrarese R, Messina MJ, Dolpady J et al (2017). High frequency of intestinal TH17 cells correlates with microbiota alterations and disease activity in multiple sclerosis. Sci Adv 3: e1700492.

Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R (2009). Bacterial community variation in human body habitats across space and time. Science 326: 1694-1697.

Dao MC, Everard A, Aron-Wisnewsky J, Sokolovska N, Prifti E, Verger EO et al (2016). Akkermansia muciniphila and improved metabolic health during a dietary intervention in obesity: relationship with gut microbiome richness and ecology. Gut 65: 426-436.

David LA, Materna AC, Friedman J, Campos-Baptista MI, Blackburn MC, Perrotta A et al (2014a). Host lifestyle affects human microbiota on daily timescales. Genome Biol 15: R89.

David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE et al (2014b). Diet rapidly and reproducibly alters the human gut microbiome. Nature 505: 559-563.

De Filippo C, Cavalieri D, Di Paola M, Ramazzotti M, Poullet JB, Massart S et al (2010). Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. Proc Natl Acad Sci U S A 107: 14691-14696.

De Vadder F, Kovatcheva-Datchary P, Goncalves D, Vinera J, Zitoun C, Duchampt A et al (2014). Microbiota-generated metabolites promote metabolic benefits via gut-brain neural circuits. Cell 156: 84-96.

De Vadder F, Kovatcheva-Datchary P, Zitoun C, Duchampt A, Backhed F, Mithieux G (2016). Microbiota-Produced Succinate Improves Glucose Homeostasis via Intestinal Gluconeogenesis. Cell Metab 24: 151-157.

de Wit N, Derrien M, Bosch-Vermeulen H, Oosterink E, Keshtkar S, Duval C et al (2012). Saturated fat stimulates obesity and hepatic steatosis and affects gut microbiota composition by an enhanced overflow of dietary fat to the distal intestine. Am J Physiol Gastrointest Liver Physiol 303: G589-599.

Denou E, Marcinko K, Surette MG, Steinberg GR, Schertzer JD (2016). High-intensity exercise training increases the diversity and metabolic capacity of the mouse distal gut microbiota during diet-induced obesity. Am J Physiol Endocrinol Metab 310: E982-993.

Derrien M, Vaughan EE, Plugge CM, de Vos WM (2004). Akkermansia muciniphila gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. Int J Syst Evol Microbiol 54: 1469-1476.

Desai MS, Seekatz AM, Koropatkin NM, Kamada N, Hickey CA, Wolter M et al (2016). A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. Cell 167: 1339-1353 e1321.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K et al (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72: 5069-5072.

Dethlefsen L, Huse S, Sogin ML, Relman DA (2008). The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. PLoS Biol 6: e280.

Dethlefsen L, Relman DA (2011). Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. Proc Natl Acad Sci U S A 108 Suppl 1: 4554-4561.

Deusch S, Tilocca B, Camarinha-Silva A, Seifert J (2015). News in livestock research - use of Omics-technologies to study the microbiota in the gastrointestinal tract of farm animals. Comput Struct Biotechnol J 13: 55-63.

Devkota S, Wang Y, Musch MW, Leone V, Fehlner-Peach H, Nadimpalli A et al (2012). Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in Il10-/- mice. Nature 487: 104-108.

Dhariwal A, Chong J, Habib S, King IL, Agellon LB, Xia J (2017). MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. Nucleic Acids Res.

Donaldson GP, Lee SM, Mazmanian SK (2016). Gut biogeography of the bacterial microbiota. Nat Rev Microbiol 14: 20-32.

Dong X, Liu Z, Lan D, Niu J, Miao J, Yang G et al (2017). Critical role of Keratin 1 in maintaining epithelial barrier and correlation of its down-regulation with the progression of inflammatory bowel disease. Gene 608: 13-19.

Durso LM, Harhay GP, Smith TP, Bono JL, Desantis TZ, Harhay DM et al (2010). Animal-to-animal variation in fecal microbial diversity among beef cattle. Appl Environ Microbiol 76: 4858-4862.

Durso LM, Harhay GP, Bono JL, Smith TP (2011). Virulence-associated and antibiotic resistance genes of microbial populations in cattle feces analyzed using a metagenomic approach. J Microbiol Methods 84: 278-282.

Earle KA, Billings G, Sigal M, Lichtman JS, Hansson GC, Elias JE et al (2015). Quantitative Imaging of Gut Microbiota Spatial Organization. Cell Host Microbe 18: 478-488.

Edgar RC (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26: 2460-2461.

El Kaoutari A, Armougom F, Gordon JI, Raoult D, Henrissat B (2013). The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. Nat Rev Microbiol 11: 497-504.

Ellegaard KM, Engel P (2016). Beyond 16S rRNA Community Profiling: Intra-Species Diversity in the Gut Microbiota. Front Microbiol 7: 1475.

Endo A, Irisawa T, Futagawa-Endo Y, Salminen S, Ohkuma M, Dicks L (2013). Lactobacillus faecis sp. nov., isolated from animal faeces. Int J Syst Evol Microbiol 63: 4502-4507.

Erickson AR, Cantarel BL, Lamendella R, Darzi Y, Mongodin EF, Pan C et al (2012). Integrated metagenomics/metaproteomics reveals human host-microbiota signatures of Crohn's disease. PLoS One 7: e49138.

Faith JJ, Guruge JL, Charbonneau M, Subramanian S, Seedorf H, Goodman AL et al (2013). The long-term stability of the human gut microbiota. Science 341: 1237439.

Falony G, Vieira-Silva S, Raes J (2015). Microbiology Meets Big Data: The Case of Gut Microbiota-Derived Trimethylamine. Annu Rev Microbiol 69: 305-321.

Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K et al (2016). Population-level analysis of gut microbiome variation. Science 352: 560-564.

Forslund K, Hildebrand F, Nielsen T, Falony G, Le Chatelier E, Sunagawa S et al (2017). Corrigendum: Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. Nature 545: 116.

Fouhy F, Guinane CM, Hussey S, Wall R, Ryan CA, Dempsey EM et al (2012). High-throughput sequencing reveals the incomplete, short-term recovery of infant gut microbiota following parenteral antibiotic treatment with ampicillin and gentamicin. Antimicrob Agents Chemother 56: 5811-5820.

Franzosa EA, Morgan XC, Segata N, Waldron L, Reyes J, Earl AM et al (2014). Relating the metatranscriptome and metagenome of the human gut. Proc Natl Acad Sci U S A 111: E2329-2338.

Fraumene C, Manghina V, Cadoni E, Marongiu F, Abbondio M, Serra M et al (2017). Caloric restriction promotes rapid expansion and long-lasting increase of Lactobacillus in the rat fecal microbiota. Gut Microbes: 1-11.

Fukuda S, Toh H, Hase K, Oshima K, Nakanishi Y, Yoshimura K et al (2011). Bifidobacteria can protect from enteropathogenic infection through production of acetate. Nature 469: 543-547.

Galipeau HJ, McCarville JL, Huebener S, Litwin O, Meisel M, Jabri B et al (2015). Intestinal microbiota modulates gluten-induced immunopathology in humanized mice. Am J Pathol 185: 2969-2982.

Garcia-Garcera M, Garcia-Etxebarria K, Coscolla M, Latorre A, Calafell F (2013). A new method for extracting skin microbes allows metagenomic analysis of whole-deep skin. PLoS One 8: e74914.

Getachew T, Dorchies P, Jacquiet P (2007). Trends and challenges in the effective and sustainable control of Haemonchus contortus infection in sheep. Review. Parasite 14: 3-14.

Gevers D, Knight R, Petrosino JF, Huang K, McGuire AL, Birren BW et al (2012). The Human Microbiome Project: a community resource for the healthy human microbiome. PLoS Biol 10: e1001377.

Giannoukos G, Ciulla DM, Huang K, Haas BJ, Izard J, Levin JZ et al (2012). Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes. Genome Biol 13: R23.

Gill SR, Pop M, Deboy RT, Eckburg PB, Turnbaugh PJ, Samuel BS et al (2006). Metagenomic analysis of the human distal gut microbiome. Science 312: 1355-1359.

Girbovan A, Sur G, Samasca G, Lupan I (2017). Dysbiosis a risk factor for celiac disease. Med Microbiol Immunol 206: 83-91.

Gkouskou KK, Deligianni C, Tsatsanis C, Eliopoulos AG (2014). The gut microbiota in mouse models of inflammatory bowel disease. Front Cell Infect Microbiol 4: 28.

Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R et al (2014). Human genetics shape the gut microbiome. Cell 159: 789-799.

Greenhalgh K, Meyer KM, Aagaard KM, Wilmes P (2016). The human gut microbiome in health: establishment and resilience of microbiota over a lifetime. Environ Microbiol 18: 2103-2116.

Grice EA, Kong HH, Renaud G, Young AC, Bouffard GG, Blakesley RW et al (2008). A diversity profile of the human skin microbiota. Genome Res 18: 1043-1050.

Grice EA, Segre JA (2012). The human microbiome: our second genome. Annu Rev Genomics Hum Genet 13: 151-170.

Group JCHMPDGW (2012). Evaluation of 16S rDNA-based community profiling for human microbiome research. PLoS One 7: e39315.

Hansen CH, Krych L, Buschard K, Metzdorff SB, Nellemann C, Hansen LH et al (2014). A maternal gluten-free diet reduces inflammation and diabetes incidence in the offspring of NOD mice. Diabetes 63: 2821-2832.

Hauner H, Bechthold A, Boeing H, Bronstrup A, Buyken A, Leschik-Bonnet E et al (2012). Evidence-based guideline of the German Nutrition Society: carbohydrate intake and prevention of nutrition-related diseases. Ann Nutr Metab 60 Suppl 1: 1-58.

Heintz-Buschart A, May P, Laczny CC, Lebrun LA, Bellora C, Krishna A et al (2016). Integrated multi-omics of the human gut microbiome in a case study of familial type 1 diabetes. Nat Microbiol 2: 16180.

Hess M, Sczyrba A, Egan R, Kim TW, Chokhawala H, Schroth G et al (2011). Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. Science 331: 463-467.

Higami Y, Pugh TD, Page GP, Allison DB, Prolla TA, Weindruch R (2004). Adipose tissue energy metabolism: altered gene expression profile of mice subjected to long-term caloric restriction. FASEB J 18: 415-417.

Hjorth MF, Roager HM, Larsen TM, Poulsen SK, Licht TR, Bahl MI et al (2017). Pre-treatment microbial Prevotella-to-Bacteroides ratio, determines body fat loss success during a 6-month randomized controlled diet intervention. Int J Obes (Lond).

Hugenholtz P, Pace NR (1996). Identifying microbial diversity in the natural environment: a molecular phylogenetic approach. Trends Biotechnol 14: 190-197.

Hugenholtz P, Tyson GW, Webb RI, Wagner AM, Blackall LL (2001). Investigation of candidate division TM7, a recently recognized major lineage of the domain Bacteria with no known pure-culture representatives. Appl Environ Microbiol 67: 411-419.

Huntington GB, Harmon DL, Richards CJ (2006). Sites, rates, and limits of starch digestion and glucose metabolism in growing cattle. J Anim Sci 84 Suppl: E14-24.

Huson DH, Beier S, Flade I, Gorska A, El-Hadidi M, Mitra S et al (2016). MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. PLoS Comput Biol 12: e1004957.

Huws SA, Kim EJ, Lee MR, Scott MB, Tweed JK, Pinloche E et al (2011). As yet uncultured bacteria phylogenetically classified as Prevotella, Lachnospiraceae incertae sedis and unclassified Bacteroidales, Clostridiales and Ruminococcaceae may play a predominant role in ruminal biohydrogenation. Environ Microbiol 13: 1500-1512.

Ito M, Okino N, Tani M (2014). New insight into the structure, reaction mechanism, and biological functions of neutral ceramidase. Biochim Biophys Acta 1841: 682-691.

Jernberg C, Lofmark S, Edlund C, Jansson JK (2007). Long-term ecological impacts of antibiotic administration on the human intestinal microbiota. ISME J 1: 56-66.

Jiang W, Wu N, Wang X, Chi Y, Zhang Y, Qiu X et al (2015). Dysbiosis gut microbiota associated with inflammation and impaired mucosal immune function in intestine of humans with non-alcoholic fatty liver disease. Sci Rep 5: 8096.

Johansson ME, Sjovall H, Hansson GC (2013). The gastrointestinal mucus system in health and disease. Nat Rev Gastroenterol Hepatol 10: 352-361.

Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M (2016). KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 44: D457-462.

Karlsson FH, Tremaroli V, Nookaew I, Bergstrom G, Behre CJ, Fagerberg B et al (2013). Gut metagenome in European women with normal, impaired and diabetic glucose control. Nature 498: 99-103.

Kasai C, Sugimoto K, Moritani I, Tanaka J, Oya Y, Inoue H et al (2016). Comparison of human gut microbiota in control subjects and patients with colorectal carcinoma in adenoma: Terminal restriction fragment length polymorphism and next-generation sequencing analyses. Oncol Rep 35: 325-333.

Kashtanova DA, Popenko AS, Tkacheva ON, Tyakht AB, Alexeev DG, Boytsov SA (2016). Association between the gut microbiota and diet: Fetal life, early childhood, and further life. Nutrition 32: 620-627.

Kim M, Kim J, Kuehn LA, Bono JL, Berry ED, Kalchayanand N et al (2014). Investigation of bacterial diversity in the feces of cattle fed different diets. J Anim Sci 92: 683-694.

Kim M, Park T, Yu Z (2017). Metagenomic investigation of gastrointestinal microbiome in cattle - A review. Asian-Australas J Anim Sci.

Klevenhusen F, Petri RM, Kleefisch MT, Khiaosa-Ard R, Metzler-Zebeli BU, Zebeli Q (2017). Changes in fibre-adherent and fluid-associated microbial communities and fermentation

profiles in the rumen of cattle fed diets differing in hay quality and concentrate amount. FEMS Microbiol Ecol 93.

Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M et al (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res 41: e1.

Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R et al (2011). Succession of microbial consortia in the developing infant gut microbiome. Proc Natl Acad Sci U S A 108 Suppl 1: 4578-4585.

Koeth RA, Wang Z, Levison BS, Buffa JA, Org E, Sheehy BT et al (2013). Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. Nat Med 19: 576-585.

Kolmeder CA, de Been M, Nikkila J, Ritamo I, Matto J, Valmu L et al (2012). Comparative metaproteomics and diversity analysis of human intestinal microbiota testifies for its temporal stability and expression of core functions. PLoS One 7: e29913.

Kolmeder CA, Salojarvi J, Ritari J, de Been M, Raes J, Falony G et al (2016). Faecal Metaproteomic Analysis Reveals a Personalized and Stable Functional Microbiome and Limited Effects of a Probiotic Intervention in Adults. PLoS One 11: e0153294.

Kong LC, Holmes BA, Cotillard A, Habi-Rachedi F, Brazeilles R, Gougis S et al (2014). Dietary patterns differently associate with inflammation and gut microbiota in overweight and obese subjects. PLoS One 9: e109434.

Koropatkin NM, Cameron EA, Martens EC (2012). How glycan metabolism shapes the human gut microbiota. Nat Rev Microbiol 10: 323-335.

Kovatcheva-Datchary P, Nilsson A, Akrami R, Lee YS, De Vadder F, Arora T et al (2015). Dietary Fiber-Induced Improvement in Glucose Metabolism Is Associated with Increased Abundance of Prevotella. Cell Metab 22: 971-982.

Kubeck R, Bonet-Ripoll C, Hoffmann C, Walker A, Muller VM, Schuppel VL et al (2016). Dietary fat and gut microbiota interactions determine diet-induced obesity in mice. Mol Metab 5: 1162-1174.

Kumar S, Choudhury PK, Carro MD, Griffith GW, Dagar SS, Puniya M et al (2014). New aspects and strategies for methane mitigation from ruminants. Appl Microbiol Biotechnol 98: 31-44.

Lai B, Ding R, Li Y, Duan L, Zhu H (2012). A de novo metagenomic assembly program for shotgun DNA reads. Bioinformatics 28: 1455-1462.

Lamendella R, VerBerkmoes N, Jansson JK (2012). 'Omics' of the mammalian gut--new insights into function. Curr Opin Biotechnol 23: 491-500.

Lang JM, Eisen JA, Zivkovic AM (2014). The microbes we eat: abundance and taxonomy of microbes consumed in a day's worth of meals for three diet types. PeerJ 2: e659.

Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA et al (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat Biotechnol 31: 814-821.

Lederberg J, McCray AT (2001). Ome SweetOmics--A Genealogical Treasury of Words. The Scientist 15: 8-8.

Leonel AJ, Alvarez-Leite JI (2012). Butyrate: implications for intestinal function. Curr Opin Clin Nutr Metab Care 15: 474-479.

Ley RE, Peterson DA, Gordon JI (2006). Ecological and evolutionary forces shaping microbial diversity in the human intestine. Cell 124: 837-848.

Li J, Jia H, Cai X, Zhong H, Feng Q, Sunagawa S et al (2014). An integrated catalog of reference genes in the human gut microbiome. Nat Biotechnol 32: 834-841.

Liang X, Bittinger K, Li X, Abernethy DR, Bushman FD, FitzGerald GA (2015). Bidirectional interactions between indomethacin and the murine intestinal microbiota. Elife 4: e08973.

Liou AP, Paziuk M, Luevano JM, Jr., Machineni S, Turnbaugh PJ, Kaplan LM (2013). Conserved shifts in the gut microbiota due to gastric bypass reduce host weight and adiposity. Sci Transl Med 5: 178ra141.

Lloyd-Price J, Abu-Ali G, Huttenhower C (2016). The healthy human microbiome. Genome Med 8: 51.

Lopes LD, de Souza Lima AO, Taketani RG, Darias P, da Silva LR, Romagnoli EM et al (2015). Exploring the sheep rumen microbiome for carbohydrate-active enzymes. Antonie Van Leeuwenhoek 108: 15-30.

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

Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R (2012). Diversity, stability and resilience of the human gut microbiota. Nature 489: 220-230.

Ma J, Prince AL, Bader D, Hu M, Ganu R, Baquero K et al (2014). High-fat maternal diet during pregnancy persistently alters the offspring microbiome in a primate model. Nat Commun 5: 3889.

Maioli M, Pes GM, Sanna M, Cherchi S, Dettori M, Manca E et al (2008). Sourdough-leavened bread improves postprandial glucose and insulin plasma levels in subjects with impaired glucose tolerance. Acta Diabetol 45: 91-96.

Manichanh C, Borruel N, Casellas F, Guarner F (2012). The gut microbiota in IBD. Nat Rev Gastroenterol Hepatol 9: 599-608.

Mao S, Zhang M, Liu J, Zhu W (2015). Characterising the bacterial microbiota across the gastrointestinal tracts of dairy cattle: membership and potential function. Sci Rep 5: 16116.

Marongiu F, Serra MP, Doratiotto S, Sini M, Fanti M, Cadoni E et al (2016). Aging promotes neoplastic disease through effects on the tissue microenvironment. Aging (Albany NY) 8: 3390-3399.

Marques TM, Wall R, Ross RP, Fitzgerald GF, Ryan CA, Stanton C (2010). Programming infant gut microbiota: influence of dietary and environmental factors. Curr Opin Biotechnol 21: 149-156.

Martens EC, Koropatkin NM, Smith TJ, Gordon JI (2009). Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. J Biol Chem 284: 24673-24677.

Martin R, Bermudez-Humaran LG, Langella P (2016). Gnotobiotic Rodents: An In Vivo Model for the Study of Microbe-Microbe Interactions. Front Microbiol 7: 409.

Martinez I, Muller CE, Walter J (2013). Long-term temporal analysis of the human fecal microbiota revealed a stable core of dominant bacterial species. PLoS One 8: e69621.

Mazzarella G, Bergamo P, Maurano F, Luongo D, Rotondi Aufiero V, Bozzella G et al (2014). Gliadin intake alters the small intestinal mucosa in indomethacin-treated HLA-DQ8 transgenic mice. Am J Physiol Gastrointest Liver Physiol 307: G302-312.

Mazzoli R, Pessione E (2016). The Neuro-endocrinological Role of Microbial Glutamate and GABA Signaling. Front Microbiol 7: 1934.

Mende DR, Waller AS, Sunagawa S, Jarvelin AI, Chan MM, Arumugam M et al (2012). Assessment of metagenomic assembly using simulated next generation sequencing data. PLoS One 7: e31386.

Mesuere B, Van der Jeugt F, Willems T, Naessens T, Devreese B, Martens L et al (2017). High-throughput metaproteomics data analysis with Unipept: A tutorial. J Proteomics.

Metsalu T, Vilo J (2015). ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. Nucleic Acids Res 43: W566-570.

Moreira AP, Texeira TF, Ferreira AB, Peluzio Mdo C, Alfenas Rde C (2012). Influence of a high-fat diet on gut microbiota, intestinal permeability and metabolic endotoxaemia. Br J Nutr 108: 801-809.

Morgan XC, Segata N, Huttenhower C (2013). Biodiversity and functional genomics in the human microbiome. Trends Genet 29: 51-58.

Morgavi DP, Rathahao-Paris E, Popova M, Boccard J, Nielsen KF, Boudra H (2015). Rumen microbial communities influence metabolic phenotypes in lambs. Front Microbiol 6: 1060.

Mouzaki M, Wang AY, Bandsma R, Comelli EM, Arendt BM, Zhang L et al (2016). Bile Acids and Dysbiosis in Non-Alcoholic Fatty Liver Disease. PLoS One 11: e0151829.

Murphy N, Norat T, Ferrari P, Jenab M, Bueno-de-Mesquita B, Skeie G et al (2012). Dietary fibre intake and risks of cancers of the colon and rectum in the European prospective investigation into cancer and nutrition (EPIC). PLoS One 7: e39361.

Muth T, Benndorf D, Reichl U, Rapp E, Martens L (2013). Searching for a needle in a stack of needles: challenges in metaproteomics data analysis. Mol Biosyst 9: 578-585.

Namiki T, Hachiya T, Tanaka H, Sakakibara Y (2012). MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. Nucleic Acids Res 40: e155.

Nelson KE, Weinstock GM, Highlander SK, Worley KC, Creasy HH, Wortman JR et al (2010). A catalog of reference genomes from the human microbiome. Science 328: 994-999.

O'Callaghan A, van Sinderen D (2016). Bifidobacteria and Their Role as Members of the Human Gut Microbiota. Front Microbiol 7: 925.

Olsen JV, de Godoy LM, Li G, Macek B, Mortensen P, Pesch R et al (2005). Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. Mol Cell Proteomics 4: 2010-2021.

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

Ou J, DeLany JP, Zhang M, Sharma S, O'Keefe SJ (2012). Association between low colonic short-chain fatty acids and high bile acids in high colon cancer risk populations. Nutr Cancer 64: 34-40.

Palomba A, Tanca A, Fraumene C, Abbondio M, Fancello F, Atzori AS et al (2017). Multi-Omic Biogeography of the Gastrointestinal Microbiota of a Pre-Weaned Lamb. Proteomes 5.

Palomino MM, Waehner PM, Fina Martin J, Ojeda P, Malone L, Sanchez Rivas C et al (2016). Influence of osmotic stress on the profile and gene expression of surface layer proteins in Lactobacillus acidophilus ATCC 4356. Appl Microbiol Biotechnol 100: 8475-8484.

Pascal V, Pozuelo M, Borruel N, Casellas F, Campos D, Santiago A et al (2017). A microbial signature for Crohn's disease. Gut 66: 813-822.

Paster BJ, Canale-Parola E (1985). Treponema saccharophilum sp. nov., a large pectinolytic spirochete from the bovine rumen. Appl Environ Microbiol 50: 212-219.

Peng S, Yin J, Liu X, Jia B, Chang Z, Lu H et al (2015). First insights into the microbial diversity in the omasum and reticulum of bovine using Illumina sequencing. J Appl Genet 56: 393-401.

Peterson J, Garges S, Giovanni M, McInnes P, Wang L, Schloss JA et al (2009). The NIH Human Microbiome Project. Genome Res 19: 2317-2323.

Pham TV, Jimenez CR (2012). An accurate paired sample test for count data. Bioinformatics 28: i596-i602.

Pilia G, Chen WM, Scuteri A, Orru M, Albai G, Dei M et al (2006). Heritability of cardiovascular and personality traits in 6,148 Sardinians. PLoS Genet 2: e132.

Png CW, Linden SK, Gilshenan KS, Zoetendal EG, McSweeney CS, Sly LI et al (2010). Mucolytic bacteria with increased prevalence in IBD mucosa augment in vitro utilization of mucin by other bacteria. Am J Gastroenterol 105: 2420-2428.

Popova M, Morgavi DP, Martin C (2013). Methanogens and methanogenesis in the rumens and ceca of lambs fed two different high-grain-content diets. Appl Environ Microbiol 79: 1777-1786.

Poutanen K, Flander L, Katina K (2009). Sourdough and cereal fermentation in a nutritional perspective. Food Microbiol 26: 693-699.

Pundir S, Martin MJ, O'Donovan C (2016). UniProt Tools. Curr Protoc Bioinformatics 53: 1 29 21-15.

Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C et al (2010). A human gut microbial gene catalogue established by metagenomic sequencing. Nature 464: 59-65.

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

Queipo-Ortuno MI, Boto-Ordonez M, Murri M, Gomez-Zumaquero JM, Clemente-Postigo M, Estruch R et al (2012). Influence of red wine polyphenols and ethanol on the gut microbiota ecology and biochemical biomarkers. Am J Clin Nutr 95: 1323-1334.

Rajilic-Stojanovic M, Heilig HG, Molenaar D, Kajander K, Surakka A, Smidt H et al (2009). Development and application of the human intestinal tract chip, a phylogenetic microarray: analysis of universally conserved phylotypes in the abundant microbiota of young and elderly adults. Environ Microbiol 11: 1736-1751.

Rampelli S, Schnorr SL, Consolandi C, Turroni S, Severgnini M, Peano C et al (2015). Metagenome Sequencing of the Hadza Hunter-Gatherer Gut Microbiota. Curr Biol 25: 1682-1693.

Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SS, McCulle SL et al (2011). Vaginal microbiome of reproductive-age women. Proc Natl Acad Sci U S A 108 Suppl 1: 4680-4687.

Rebello CJ, Burton J, Heiman M, Greenway FL (2015). Gastrointestinal microbiome modulator improves glucose tolerance in overweight and obese subjects: A randomized controlled pilot trial. J Diabetes Complications 29: 1272-1276.

Reeves AR, Wang GR, Salyers AA (1997). Characterization of four outer membrane proteins that play a role in utilization of starch by Bacteroides thetaiotaomicron. J Bacteriol 179: 643-649.

Regler R, Sickinger S, Schweizer M (1991). Differential regulation of the two mRNA species of the rodent negative acute phase protein alpha 1-inhibitor 3. FEBS Lett 282: 368-372.

Rho M, Tang H, Ye Y (2010). FragGeneScan: predicting genes in short and error-prone reads. Nucleic Acids Res 38: e191.

Ridaura VK, Faith JJ, Rey FE, Cheng J, Duncan AE, Kau AL et al (2013). Gut microbiota from twins discordant for obesity modulate metabolism in mice. Science 341: 1241214.

Rinttila T, Kassinen A, Malinen E, Krogius L, Palva A (2004). Development of an extensive set of 16S rDNA-targeted primers for quantification of pathogenic and indigenous bacteria in faecal samples by real-time PCR. J Appl Microbiol 97: 1166-1177.

Rognes T, Flouri T, Nichols B, Quince C, Mahe F (2016). VSEARCH: a versatile open source tool for metagenomics. PeerJ 4: e2584.

Roopchand DE, Carmody RN, Kuhn P, Moskal K, Rojas-Silva P, Turnbaugh PJ et al (2015). Dietary Polyphenols Promote Growth of the Gut Bacterium Akkermansia muciniphila and Attenuate High-Fat Diet-Induced Metabolic Syndrome. Diabetes 64: 2847-2858.

Rupnik M, Wilcox MH, Gerding DN (2009). Clostridium difficile infection: new developments in epidemiology and pathogenesis. Nat Rev Microbiol 7: 526-536.

Salonen A, de Vos WM (2014). Impact of diet on human intestinal microbiota and health. Annu Rev Food Sci Technol 5: 239-262.

Savage DC (1977). Microbial ecology of the gastrointestinal tract. Annu Rev Microbiol 31: 107-133.

Schloss PD (2010). The effects of alignment quality, distance calculation method, sequence filtering, and region on the analysis of 16S rRNA gene-based studies. PLoS Comput Biol 6: e1000844.

Schmieder R, Edwards R (2011). Quality control and preprocessing of metagenomic datasets. Bioinformatics 27: 863-864.

Schnorr SL, Candela M, Rampelli S, Centanni M, Consolandi C, Basaglia G et al (2014). Gut microbiome of the Hadza hunter-gatherers. Nat Commun 5: 3654.

Scholtens PA, Oozeer R, Martin R, Amor KB, Knol J (2012). The early settlers: intestinal microbiology in early life. Annu Rev Food Sci Technol 3: 425-447.

Shanks OC, Kelty CA, Archibeque S, Jenkins M, Newton RJ, McLellan SL et al (2011). Community structures of fecal bacteria in cattle from different animal feeding operations. Appl Environ Microbiol 77: 2992-3001.

Sharpton TJ (2014). An introduction to the analysis of shotgun metagenomic data. Front Plant Sci 5: 209.

Shi W, Moon CD, Leahy SC, Kang D, Froula J, Kittelmann S et al (2014). Methane yield phenotypes linked to differential gene expression in the sheep rumen microbiome. Genome Res 24: 1517-1525.

Sidore C, Busonero F, Maschio A, Porcu E, Naitza S, Zoledziewska M et al (2015). Genome sequencing elucidates Sardinian genetic architecture and augments association analyses for lipid and blood inflammatory markers. Nat Genet 47: 1272-1281.

Silverman M, Kua L, Tanca A, Pala M, Palomba A, Tanes C et al (2017). Protective major histocompatibility complex allele prevents type 1 diabetes by shaping the intestinal microbiota early in ontogeny. Proc Natl Acad Sci U S A.

Slavin J (2013). Fiber and prebiotics: mechanisms and health benefits. Nutrients 5: 1417-1435.

Small DM, Gobe GC (2012). Cytochrome c: potential as a noninvasive biomarker of drug-induced acute kidney injury. Expert Opin Drug Metab Toxicol 8: 655-664.

Sonnenburg JL, Backhed F (2016). Diet-microbiota interactions as moderators of human metabolism. Nature 535: 56-64.

Spitzer M, Wildenhain J, Rappsilber J, Tyers M (2014). BoxPlotR: a web tool for generation of box plots. Nat Methods 11: 121-122.

Stamataki NS, Yanni AE, Karathanos VT (2017). Bread making technology influences postprandial glucose response: a review of the clinical evidence. Br J Nutr 117: 1001-1012.

Stanley D, Geier MS, Hughes RJ, Denman SE, Moore RJ (2013). Highly variable microbiota development in the chicken gastrointestinal tract. PLoS One 8: e84290.

Staudacher HM, Lomer MCE, Farquharson FM, Louis P, Fava F, Franciosi E et al (2017). A Diet Low in FODMAPs Reduces Symptoms in Patients With Irritable Bowel Syndrome and A Probiotic Restores Bifidobacterium Species: A Randomized Controlled Trial. Gastroenterology 153: 936-947.

Tanca A, Biosa G, Pagnozzi D, Addis MF, Uzzau S (2013). Comparison of detergent-based sample preparation workflows for LTQ-Orbitrap analysis of the Escherichia coli proteome. Proteomics 13: 2597-2607.

Tanca A, Palomba A, Pisanu S, Deligios M, Fraumene C, Manghina V et al (2014). A straightforward and efficient analytical pipeline for metaproteome characterization. Microbiome 2: 49.

Tanca A, Palomba A, Pisanu S, Addis MF, Uzzau S (2015). Enrichment or depletion? The impact of stool pretreatment on metaproteomic characterization of the human gut microbiota. Proteomics 15: 3474-3485.

Tanca A, Abbondio M, Palomba A, Fraumene C, Manghina V, Cucca F et al (2017a). Potential and active functions in the gut microbiota of a healthy human cohort. Microbiome 5: 79.

Tanca A, Fraumene C, Manghina V, Palomba A, Abbondio M, Deligios M et al (2017b). Diversity and functions of the sheep faecal microbiota: a multi-omic characterization. Microb Biotechnol 10: 541-554.

Tanca A, Manghina V, Fraumene C, Palomba A, Abbondio M, Deligios M et al (2017c). Metaproteogenomics reveals taxonomic and functional changes between cecal and fecal microbiota in mouse. Frontiers in Microbiology 8.

Tap J, Mondot S, Levenez F, Pelletier E, Caron C, Furet JP et al (2009). Towards the human intestinal microbiota phylogenetic core. Environ Microbiol 11: 2574-2584.

Tapio I, Snelling TJ, Strozzi F, Wallace RJ (2017). The ruminal microbiome associated with methane emissions from ruminant livestock. J Anim Sci Biotechnol 8: 7.

Thorburn AN, Macia L, Mackay CR (2014). Diet, metabolites, and "western-lifestyle" inflammatory diseases. Immunity 40: 833-842.

Tilocca B, Burbach K, Heyer CME, Hoelzle LE, Mosenthin R, Stefanski V et al (2017). Dietary changes in nutritional studies shape the structural and functional composition of the pigs' fecal microbiome-from days to weeks. Microbiome 5: 144.

Toda M, Tulic MK, Levitt RC, Hamid Q (2002). A calcium-activated chloride channel (HCLCA1) is strongly related to IL-9 expression and mucus production in bronchial epithelium of patients with asthma. J Allergy Clin Immunol 109: 246-250.

Toden S, Bird AR, Topping DL, Conlon MA (2005). Resistant starch attenuates colonic DNA damage induced by higher dietary protein in rats. Nutr Cancer 51: 45-51.

Tojo R, Suarez A, Clemente MG, de los Reyes-Gavilan CG, Margolles A, Gueimonde M et al (2014). Intestinal microbiota in health and disease: role of bifidobacteria in gut homeostasis. World J Gastroenterol 20: 15163-15176.

Tomkovich S, Yang Y, Winglee K, Gauthier J, Muhlbauer M, Sun X et al (2017). Locoregional Effects of Microbiota in a Preclinical Model of Colon Carcinogenesis. Cancer Res 77: 2620-2632.

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

Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE et al (2009). A core gut microbiome in obese and lean twins. Nature 457: 480-484.

Tveit AT, Urich T, Svenning MM (2014). Metatranscriptomic analysis of arctic peat soil microbiota. Appl Environ Microbiol 80: 5761-5772.

Ussar S, Griffin NW, Bezy O, Fujisaka S, Vienberg S, Softic S et al (2015). Interactions between Gut Microbiota, Host Genetics and Diet Modulate the Predisposition to Obesity and Metabolic Syndrome. Cell Metab 22: 516-530.

Valles-Colomer M, Darzi Y, Vieira-Silva S, Falony G, Raes J, Joossens M (2016). Meta-omics in Inflammatory Bowel Disease Research: Applications, Challenges, and Guidelines. J Crohns Colitis 10: 735-746.

Vatanen T, Kostic AD, d'Hennezel E, Siljander H, Franzosa EA, Yassour M et al (2016). Variation in Microbiome LPS Immunogenicity Contributes to Autoimmunity in Humans. Cell 165: 842-853.

Verberkmoes NC, Russell AL, Shah M, Godzik A, Rosenquist M, Halfvarson J et al (2009). Shotgun metaproteomics of the human distal gut microbiota. ISME J 3: 179-189.

Vipperla K, O'Keefe SJ (2012). The microbiota and its metabolites in colonic mucosal health and cancer risk. Nutr Clin Pract 27: 624-635.

Wang L, Jacobs JP, Lagishetty V, Yuan PQ, Wu SV, Million M et al (2017a). High-protein diet improves sensitivity to cholecystokinin and shifts the cecal microbiome without altering brain inflammation in diet-induced obesity in rats. Am J Physiol Regul Integr Comp Physiol 313: R473-R486.

Wang R, Jiang L, Zhang M, Zhao L, Hao Y, Guo H et al (2017b). The Adhesion of Lactobacillus salivarius REN to a Human Intestinal Epithelial Cell Line Requires S-layer Proteins. Sci Rep 7: 44029.

Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, Dugar B et al (2011). Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. Nature 472: 57-63.

Watson H, Mitra S, Croden FC, Taylor M, Wood HM, Perry SL et al (2017). A randomised trial of the effect of omega-3 polyunsaturated fatty acid supplements on the human intestinal microbiota. Gut.

Weimer PJ (2015). Redundancy, resilience, and host specificity of the ruminal microbiota: implications for engineering improved ruminal fermentations. Front Microbiol 6: 296.

Willett WC, Sacks F, Trichopoulou A, Drescher G, Ferro-Luzzi A, Helsing E et al (1995). Mediterranean diet pyramid: a cultural model for healthy eating. Am J Clin Nutr 61: 1402S-1406S.

Willner D, Furlan M, Schmieder R, Grasis JA, Pride DT, Relman DA et al (2011). Metagenomic detection of phage-encoded platelet-binding factors in the human oral cavity. Proc Natl Acad Sci U S A 108 Suppl 1: 4547-4553.

Wisniewski JR, Zougman A, Nagaraj N, Mann M (2009). Universal sample preparation method for proteome analysis. Nat Methods 6: 359-362.

Wolgemuth CW (2015). Flagellar motility of the pathogenic spirochetes. Semin Cell Dev Biol 46: 104-112.

Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA et al (2011). Linking long-term dietary patterns with gut microbial enterotypes. Science 334: 105-108.

Xia J, Wishart DS (2016). Using MetaboAnalyst 3.0 for Comprehensive Metabolomics Data Analysis. Curr Protoc Bioinformatics 55: 14 10 11-14 10 91.

Xiao X, Nakatsu G, Jin Y, Wong S, Yu J, Lau JY (2017). Gut Microbiota Mediates Protection Against Enteropathy Induced by Indomethacin. Sci Rep 7: 40317.

Xu Z, Knight R (2015). Dietary effects on human gut microbiome diversity. Br J Nutr 113 Suppl: S1-5.

Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M et al (2012). Human gut microbiome viewed across age and geography. Nature 486: 222-227.

Yaung SJ, Deng L, Li N, Braff JL, Church GM, Bry L et al (2015). Improving microbial fitness in the mammalian gut by in vivo temporal functional metagenomics. Mol Syst Biol 11: 788.

Yin L, Yang H, Li J, Li Y, Ding X, Wu G et al (2017). Pig models on intestinal development and therapeutics. Amino Acids.

Zeng Y, Zeng D, Zhang Y, Ni X, Tang Y, Zhu H et al (2015). Characterization of the cellulolytic bacteria communities along the gastrointestinal tract of Chinese Mongolian sheep by using PCR-DGGE and real-time PCR analysis. World J Microbiol Biotechnol 31: 1103-1113.

Zerbino DR, Birney E (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18: 821-829.

Zhang C, Li S, Yang L, Huang P, Li W, Wang S et al (2013a). Structural modulation of gut microbiota in life-long calorie-restricted mice. Nat Commun 4: 2163.

Zhang X, Shen D, Fang Z, Jie Z, Qiu X, Zhang C et al (2013b). Human gut microbiota changes reveal the progression of glucose intolerance. PLoS One 8: e71108.

Zimmer J, Lange B, Frick JS, Sauer H, Zimmermann K, Schwiertz A et al (2012). A vegan or vegetarian diet substantially alters the human colonic faecal microbiota. Eur J Clin Nutr 66: 53-60.