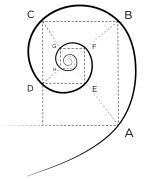
UNIVERSITÀ DEGLI STUDI
DI MILANO



DOTTORATO DI MEDICINA MOLECOLARE E TRASLAZIONALE

CICLO XXXIII

Anno Accademico 2019/2020

TESI DI DOTTORATO DI RICERCA

MED09

EPILEPSY AND MICROBIOTA: TIME- AND DRUG-RELATED SHIFTS IN THE BACTERIAL COMMUNITY

Dottoranda : Camilla CECCARANI

Matricola N° : R12048

TUTORE : prof.ssa Elisa BORGHI

CO-TUTORE: Marco SEVERGNINI

COORDINATORE DEL DOTTORATO: prof. Michele SAMAJA

INDEX

SOMMARIO

L'epilessia è una nota condizione caratterizzata da convulsioni, episodi di instabilità elettrica che provocano scuotimenti involontari nei pazienti. Oggigiorno ci sono molte strategie, di cui quella farmacologica è la principale, ma circa un terzo dei pazienti mostra resistenza ai farmaci e, nonostante i diversi approcci terapeutici per i pazienti resistenti alla terapia, molti non riescono a raggiungere il controllo delle crisi. È, quindi, importante cercare nuove strategie terapeutiche complementari che possano influenzare il quadro clinico e migliorare la qualità della vita del paziente. La connessione tra il microbiota umano e le malattie neurologiche è stata studiata a fondo negli ultimi anni. In particolare, si ritiene che l'asse intestino-cervello sia un fattore chiave nello sviluppo della malattia e nella suscettibilità alle crisi. Ciononostante, le informazioni finora disponibili, in letteratura, sulla composizione del microbiota intestinale nei pazienti con epilessia sono limitate.

Al fine di verificare l'influenza dei batteri sullo sviluppo e il trattamento dell'epilessia, durante il mio dottorato di ricerca ho analizzato 3 set di dati di pazienti epilettici: 1) un set di dati di bambini ricoverati in ospedale dopo la loro prima crisi epilettica e campionati altre 2 volte, dopo 4 e 12 mesi sottoposti a monoterapia farmacologica; 2) dataset di adulti che utilizzano quotidianamente da più di 2 anni il farmaco più comune disponibile senza più convulsioni; 3) un set di dati di ragazze affette dalla sindrome di Rett, una malattia genetica che include disturbi del linguaggio e della coordinazione e movimenti ripetitivi. Le complicanze dei pazienti con sindrome di Rett includono crisi epilettiche, e i pazienti assumono gli stessi farmaci dei pazienti epilettici. Questa coorte è stata aggiunta al progetto come controllo negativo, opposti ai controlli sani arruolati. Questi diversi set di dati gettano le basi per speculazioni sui cambiamenti e lo sviluppo del microbiota durante l'esordio dell'epilessia e una terapia farmacologica costante in termini di diversità batterica, composizione e interazione.
Il DNA batterico estratto dai campioni fecali raccolti dai pazienti è stato sequenziato tramite tecniche di sequenziamento di nuova generazione su una piattaforma

MiSeq Illumina, e le letture risultanti sono state analizzate attraverso una pipeline bioinformatica utilizzando il software QIIME, R, e Matlab.

È stato osservato che la biodiversità era diminuita in condizioni patologiche e durante l'assunzione di farmaci in tutti i confronti. Di particolare interesse è il picco di *Akkermansiaceae* nel microbiota intestinale dei bambini non in terapia. I campioni dei bambini hanno mostrato un'interessante firma intra-individuale e una separazione di diversità nel tempo di trattamento farmacologico, probabilmente correlata a una ridotta abbondanza di gruppi benefici come *Faecalibacterium*. Dall'analisi della matrice di co-abbondanza è emerso un trend opposto dei generi batterici legati a *Subdoligranulum* e di quelli associati a *Bacteroides*. Gli adulti in terapia farmacologica stabile hanno mostrato alcune tendenze simili ai pazienti Rett come se, nonostante il diverso quadro clinico, la comunità batterica intestinale fosse più uniformata e simile in termini di biodiversità. L'analisi di selezione delle varianti eseguita sui pazienti che non assumevano farmaci e sui pazienti in terapia ha confermato le famiglie *Akkermansiaceae* e *Christensenellaceae* come rilevanti per l'alterazione della composizione batterica durante il trattamento farmacologico. La separazione dei batteri Gram-positivi e Gram-negativi ha mostrato un interessante cambiamento delle frazioni con un andamento a favore dei Gram-negativi: i diversi strati protettivi potrebbero influenzare o essere influenzati dall'assunzione del farmaco.

I risultati mostrati in questa tesi evidenziano l'esistenza di differenze in termini di diversità microbica generale e tassonomia. Ulteriori studi condotti su un numero maggiore di pazienti in tempi diversi e sviluppo della malattia, e un'implementazione dell'analisi metabolica che tenga conto delle assunzioni nutrizionali, potrebbero contribuire a rafforzare questi risultati in termini di interazioni tra il microbiota e meccanismi innati di difesa dell'ospite, come così come la discriminazione a livello di genere e al di sotto di esso.

ABSTRACT

Epilepsy is a well-known condition characterized by seizures, episodes of electric instability provoking involuntary vigorous shaking in patients. Nowadays there are many strategies, of which the pharmaceutical is the main one, but around one-third of the patients show drug resistance and, despite the different therapeutic approaches for drug-resistant patients, many cannot reach seizure control. Indeed, it is important to look for new complementary therapeutic strategies that can influence the clinical picture and improve the patient's quality of life. The connection between human microbiota and neurological diseases has been thoroughly studied over the last years. In particular, the gut-brain axis has been addressed to be a key player in the development of the disease and to seizure susceptibility. Nonetheless, there is limited information so far, in literature, about the composition of the intestinal microbiota in patients with epilepsy.

To check for bacterial influence on epilepsy development and treatment, during my Ph.D. I've analyzed 3 datasets of epileptic patients: 1) a dataset of children admitted to the hospital after their first seizure and sampled 2 more times, after 4 and 12 months subjected to drug monotherapy; 2) a dataset of adults that are using daily for more than 2 years the most common drug available with no more seizures; 3) a dataset of girls affected by the Rett syndrome, a genetic disorder that comprises impairments in language and coordination and repetitive movements. Complications of Rett syndrome patients include seizures, and the patients take the same drugs as the epileptic dataset. This cohort has been added to the project as a negative control, as opposed to the healthy controls enrolled. This diverse dataset set the stage for speculations on the microbiota changes and development during epilepsy onset and constant drug therapy in terms of bacterial diversity, composition, and interaction.

The bacterial DNA extracted from the fecal samples collected from the patients has been sequenced through next-generation sequencing techniques on a MiSeq Illumina platform, and the resulting reads have been analyzed through a bioinformatic pipeline using the QIIME software, R, and Matlab.

Biodiversity was observed to be decreased in pathological conditions and along with the drug assumption in all comparisons. Of particular interest is the peak of *Akkermansiaceae* in the children's drug-naive gut microbiota. The children showed an interesting intra-individual signature and a diversity separation over pharmacological treatment time, possibly related to a reduced abundance of beneficial groups such as *Faecalibacterium*. From the co-abundance network analysis, the opposite trend of the group of genera related to *Subdoligranulum* and the ones related to *Bacteroides* have emerged. The adults in stable drug therapy showed some trends similar to the Rett patients, as if, despite the different clinical picture, the gut bacterial community was comparable and similar in terms of biodiversity. The feature selection analysis performed on the drug-naive and the drug-assuming patients confirmed the families *Akkermansiaceae* and *Christensenellaceae* to be relevant for the bacterial composition alteration during the pharmacological therapy. The separation of Gram-positive and Gram-negative bacteria showed an interesting change of the fractions with a trend in favor of the Gram-negative: the different protective layers could influence or be influenced by the drug assumption.

The results shown in this thesis highlight the existence of differences in terms of general microbial diversity and taxonomy. Further studies carried out on a larger number of patients at different times and disease development, and a metabolic analysis implementation taking into account the nutritional intakes, could contribute to reinforcing these findings in terms of interactions between the microbiota and innate mechanisms of host defense, as well as the discrimination at and below the genus level.

LIST OF ABBREVIATIONS

- *Clinical*

CNS: Central Nervous System
ENS: Enteric Nervous System
CT: Computed Tomography
MRI: Magnetic Resonance Imaging
EEG: Electroencephalograph
PET: Positron Emission Tomography
SPECT: Single Photon Emission Computed Tomography
MECP2: Methyl CpG-binding Protein 2

- *Drugs*

AEDs: Anti-Epileptic Drugs
CBZ: Carbamazepine
LTG: Lamotrigine
ETS: Etosuximide
LEV: Levetiracetam
PB: Phenobarbital
TPM: Topiramate
VPA: Valproic Acid

- *Analysis*

GM: Gut Microbiota
PCR: Polymerase chain reaction
KD: Ketogenic Diet
PCoA: Principal Coordinates Analysis
NGS: Next Generation Sequencing
OTU: Operational Taxonomic Unit
SCFA(s): Short Chain Fatty Acid(s)
RDP: Ribosomal Database Project
rRNA: ribosomal RNA
CAG: Co-Abundant Groups
PC1/PC2/PC3: first/second/third Principal Coordinate

- *Experimental labels*

HC: Helthy Controls
EPI: epileptic patients
DT12: Drug-Therapy 12 months
DN: Drug-Naive
SDT: Stable Drug Therapy
DT4: Drug-Therapy 4 months
RTT: Rett syndrome
DA: Drug Assumption

RESEARCH INTEGRITY DECLARATION

I'm pleased to affirm that the results reported in this present study have been conducted following the four fundamental principles of research integrity stated by The European Code of Conduct for Research Integrity (ALLEA, Berlin, 2018):

- *Reliability* in ensuring the quality of research, reflected in the design, the methodology, the analysis and the use of resources;
- *Honesty* in developing, undertaking, reviewing, reporting and communicating research in a transparent, fair, full and unbiased way;
- *Respect* for colleagues, research participants, society, ecosystems, cultural heritage and the environment;
- *Accountability* for the research from idea to publication, for its management and organization, for training, supervision and mentoring, and for its wider impacts.

These very same values have been taken into account also for each of the projects I participated to during the Ph.D. school years.

EXTERNAL REVIEW DECLARATION

This thesis was reviewed by four external independent reviewers, and has been corrected afterwards.

For their time, their useful notes, and all the suggestions, I would like to thank: dr. Marco Candela (*University of Bologna - Pharmacy and Biotechnology*); dr. Edoardo Pasolli (*University of Naples Federico II - Department of Agricultural Sciences*); dr. Clelia Peano (*National Research Council - Institute of Genetic and Biomedical Research*); dr. Martino Pengo (*IRCCS Istituto Auxologico Italiano - Department of Cardiovascular, Neural and Metabolic Sciences*).

INTRODUCTION

1. Epilepsy

Epilepsy is one of the first human diseases ever described, with a rich and distinguished history. Written records dated epilepsy as early as in the 4000 BC [1] and many famous historical figures have been reported, or suggested through literature analysis, to had had it: Vincent Van Gogh, Napoleon, Michelangelo - to name a few. [2] Today, epilepsy affects around 50 million people all around the world, and it is estimated that five million people are diagnosed with epilepsy each year. [3, 4]

1.1 Disease characteristics

Epilepsy, or seizure disorder, is a chronic non-communicable disease of the brain. The term “epilepsy syndrome” refers not to a specific condition but to similar groups of features occurring in patients: one can be diagnosed with epilepsy based on seizure characteristics, on the electroencephalography findings, on external or internal triggering factors, such as genetic mechanisms and prognosis or response to anti-epileptic drugs. [5]

Generally, the main identifier of epilepsy is the recurrence of unprovoked seizures: episodes of involuntary movement that may involve a part of the body or the entire body. Seizure episodes are a result of electric instability or excessive discharges in a group of brain cells. Depending on whether the epileptic crisis involves the whole brain in several areas or just a part, the patient is said to have, respectively, a generalized or a focal/partial seizure (*Figure 1*). Seizures can vary from the briefest lapses of attention or muscle jerks to severe and prolonged convulsions; over 60% of the seizures are

convulsive. [6] Seizures can also vary in frequency of occurrence, from less than 1 per year to several per day. Seizures recurrence or the need for treatment are conditions that define a patient with “active epilepsy”; the estimated proportion of the general population with active epilepsy at a given time is around 4-10 per 1000 people. [7]

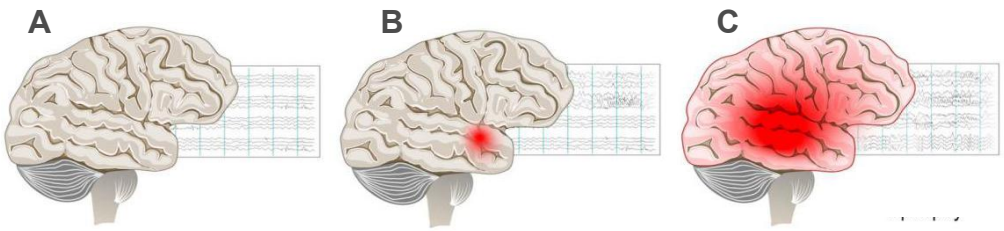


Figure 1: Activity in the brain. Compared to the brain of a non-epileptic subject (A), the different electric activity indicate whether a seizure is limited in one region (focal, B) or taking place in several areas of the brain (generalized, C).
(*Image credit: adapted from Shutterstock*)

Characteristics of seizures vary and depend on where the disturbance first starts in the brain, and how far it spreads. Temporary symptoms, such as loss of awareness or consciousness, and disturbances of movement, sensation (including vision, hearing, and taste), mood, or other cognitive functions often occur. Sometimes seizures are accompanied by a loss of consciousness and control of bladder or bowel functions. Usually, after experiencing these episodes, people do not remember what happened and the recovery period may last from minutes to hours (“post-ictal phase”); confusion, headache, and difficulty in speaking are generally the main symptoms ensuing but, in some cases, psychosis and behavioral disorders may happen. [5]

People with epilepsy tend to have more physical problems (such as fractures and bruises from falls and injuries related to seizures), as well as higher rates of psychological conditions, including anxiety and depression.

Similarly, the risk of premature death in people with epilepsy is up to three times higher than in the general population, with the highest rates of premature mortality found in low- and middle-income countries and rural areas. [7] During convulsive seizures some first-aid measures need to be employed: positioning the patient on his/her side in order to avoid fluids from getting into the lungs and, whether the seizure worsens or lasts more than five minutes, administering benzodiazepines, such as midazolam or diazepam. Usually, a period of clinical observation should follow to assess possible brain damage or to avoid sudden relapses. [5]

1.2 Causes and diagnosis

Although many underlying molecular mechanisms can lead to epilepsy, the cause of the disease is still unknown in many global cases.

The causes of epilepsy can be divided into categories: structural, genetic, infectious, metabolic, immune, unknown. Since epilepsy is a non-communicable disease, it is not contagious. Epilepsy could be provoked by brain damage from prenatal or perinatal causes (for example, a loss of oxygen or trauma during birth) or by congenital abnormalities or genetic conditions with associated brain malformations; it can be caused by a severe head injury, a stroke that restricts the amount of oxygen to the brain, or by an infection of the brain, such as meningitis or encephalitis; moreover, certain genetic syndromes can lead to epilepsy, and so do brain tumors. [5]

Whichever cause leads to the disorder onset, patients often display abnormal brain wave characteristics, due to the nature of the seizures, even when they are not experiencing them. Proper diagnosis of seizures and epilepsy is essential for effective treatment. Diagnostic tests can help determine if and where a lesion in the brain is causing seizures. Thus, there

are six main forms of generalized seizures according to the characterizing symptoms of each type: tonic-clonic, tonic, clonic, myoclonic, absence, and atonic seizures. The recognition of the right one is essential to determine the adequate drug treatment. [8]

The tool used to measure and identify abnormal electrical activity in the brain is the electroencephalograph (EEG), usually performed in a specialized outpatient clinic. The electrical signals in the brain are monitored with electrodes (sensors) attached to the scalp, within a range of time that goes from minutes to a couple of hours. Neurologist and clinicians read the EEG result, in order to find evidence of abnormal electrical activity in the brain and eventually figure out the type of seizure a patient is having. Brain imaging such as positron emission tomography (PET), magnetic resonance imaging (MRI), Single Photon Emission Computed Tomography (SPECT), and computed tomography (CT) scans are used to observe the structure of the brain and map out damaged areas or abnormalities, such as tumors and cysts, which can be the underlying origins of seizures.

Nonetheless, an epilepsy diagnosis cannot be made based on the EEG findings only: the history and anamnesis of the patients are needed in order to provide a reasonable clinical supposition for a correct diagnosis. [9]

Furthermore, seizures do not occur frequently as random events in the population: the subjects that will be diagnosed to be affected by epilepsy often had their symptoms triggered as they experience disturbances or altered behaviors such as, among others, stress, lack of sleep, alcohol abuse, flickering light, sudden noises, emotional upheavals. The limit beyond which seizures occur is known as “seizure threshold”, which varies among individuals and is lower in people that would be affected by epilepsy. [10]

1.3 A rare disease characterized by seizure onset: the Rett syndrome

Epilepsy can be a condition *per se*, but can also be a secondary effect in other syndromes, as they can be characterized by seizure episodes.

This happens, for example, in the Rett syndrome (OMIM: 312750), a genetic disorder that typically becomes apparent after 6–18 months of age. This condition affects about 1 in 15000 girls, due to mutations in the *MECP2* gene (encoding for the methyl CpG-binding protein 2) linked to the X chromosome. The typical Rett symptoms include impairments in language and coordination and repetitive movements, along with slower growth, a smaller head size, and difficulty in walking. [11]

Complications of Rett syndrome can include seizures; this, along with the absence of a known cure for the condition, explains why these patients are often assuming an anti-epileptic therapy.

1.4 Therapy

Seizures can be controlled through drug therapy, prescribed after the assessment tests, shortly after the onset of the disease. Up to 70% of people living with epilepsy or syndromes epilepsy-related could become seizure-free with the appropriate use of anti-epileptic drugs (AEDs).

The main treatments currently available in Italy include valproic acid (VPA), ethosuximide (ETS), lamotrigine (LTG), levetiracetam (LEV), carbamazepine (CBZ), topiramate (TPM), and phenobarbital (PB). All of these drugs have a low cost, come in form of a pill and have to be taken daily, with the dosages prescribed by the patient's physician after an electroencephalogram analysis. [12]

Despite the many anti-epileptic drug therapies available, around one-third of the epileptic patients develop drug resistance. [3, 7]

The several AEDs in use have different sites of action, therefore the molecular mechanisms that lead to drug resistance are likely to be different as well. There are several molecular reasons hypothesized to be responsible for this resistance: the “transporter hypothesis” says that the drug may fail to reach its target [4]; the “target hypothesis” is about a reduction in the target’s sensitivity to the treatment [13]; a hypothesis about the drug missing its target, since the drug’s scope is to prevent seizures but, as the seizure trigger is not due to the same pathogenic process in every patient, it may not hit the specific mechanism involved for the individual. [4] Novel targets for future drug development could cover these kinds of situations, in order to find the right therapy for every epileptic patient.

When monotherapy fails, prescribing the combination of two or more drugs together is a common and fast solution for gaining a moderate reduction of seizure attacks. Research studies in mice suggest that administering two AEDs acting on the same pharmacologic mechanism is less effective than administering two AEDs with different paths of action. [14] In fact, the drugs combined are usually chosen based on their different modes of action: for example, among the most used combinations, VPA and LTG are assumed for partial onset and generalized seizures cases [15, 16], while VPA and ETS are indicated for patients with absence seizures. [17]

In case the pharmacological treatment proves to be insufficient for the patient, the surgical approach should be often considered: in 60-70% of cases, it is proven to be resolutive by removing a mass (e.g. a tumor) or cutting some brain parts (e.g. the hippocampus). [18] For those who are not eligible for surgery, other strategies are the neurostimulation, which is the

application of electric stimuli to the vagus nerve, in order to suppress the abnormal discharges in the brain [19], and the ketogenic diet (as reported in *Introduction, 2.1.4 Diet Influence,* page 17).

Despite different therapeutic approaches currently available for patients with drug-resistant epilepsy, many patients cannot reach seizure control. Indeed, it is important to look for new complementary therapeutic strategies that can influence the clinical picture and improve the patient's quality of life.

2. Microbiota

In the last decade, there has been an increasing excitement on the potential contribution of the bacteria species living in our gastrointestinal system to human diseases, and their possible role in a cure as part of the treatment.

2.1 Our forgotten organ

The living organisms, on a micrometric scale, are called microbes (from the Ancient Greek words μικρός, “small” and βίος, “life”). This term includes a huge variety of species living on the planet, either single-celled or as a colony, many of them still unknown: it comprehends all bacteria, archaea, protozoa, and fungi species. They live in every part of the biosphere, and live in every macroorganism, such as plants and animals - humans included.

As the term “microorganism” concerns the singular individual of a species, the term “microbiota” represents the community of organisms that reside in a specific environment. The microorganisms living in our body, including bacteria, viruses, and archaea, are called collectively “human microbiota”, with distinct district names depending on which specific site is the subject; among them, the gastrointestinal tract hosts the most studied one: the gut microbiota (GM).

Collectively, the human microbiota species outnumber eukaryotic cells by as many as ten to one and their collective genome is estimated at 150 times larger than the human gene complement. [20, 21] However, because of their very small size (which is about one-tenth the size of eukaryotic cells, with a range of 0.5-5.0 μm), the microorganisms make up a small portion of human body mass. This portion can weight up to 2 kilograms and has a common composition in the human population for about one-third; the remaining two-thirds are specific for each individual.

Microbiota species live and act in ecological communities, have many roles in the normal activity of the host's body and several metabolic and immune functions. All of these elements lead the human microbiota to gain the definition of an "organ", generated after birth with the first bacterial colonization [22, 23] and evolved throughout the entire lifespan of the host, as a result of all the environmental, dietary and health influences. After one century of molecular biology, the existence and importance of the human microbiota have been studied only in the last two decades, also due to the difficulties in cultivating the anaerobes species. That is why, many times, our microbiota is called "a forgotten organ". [24]

Within the gastrointestinal tract, the microbial population varies from district to district. (*Figure 2*) In the stomach, bacterial abundance is lower than in the intestine due to the extremely acid pH from the gastric juices, but it is still present. [25] The duodenum has a limited microbiota as well, this time due to the rapid transit of nutritional compounds and to the secretion of biliary and pancreatic fluids that kill the majority of the bacterial species ingested; furthermore, its propulsive movement activity prevents a stable lumen colonization. [26]

There is a progressive increment in the number of bacterial species from the jejunum to the ileum, in particular of Gram-negative and obligate anaerobes. In contrast with these two areas, the colon swarms a dynamic and complex microbial ecosystem mainly made up of strict anaerobes [20, 27], particularly abundant due to the great substrate availability, the suitable surfaces, and a good environment for bacterial growth, with slow transit and a tolerable pH. [28, 29] The colon tract is the main contributor for the total microbiota abundance; hence, it is the hub for estimating the number of bacteria in the human body.

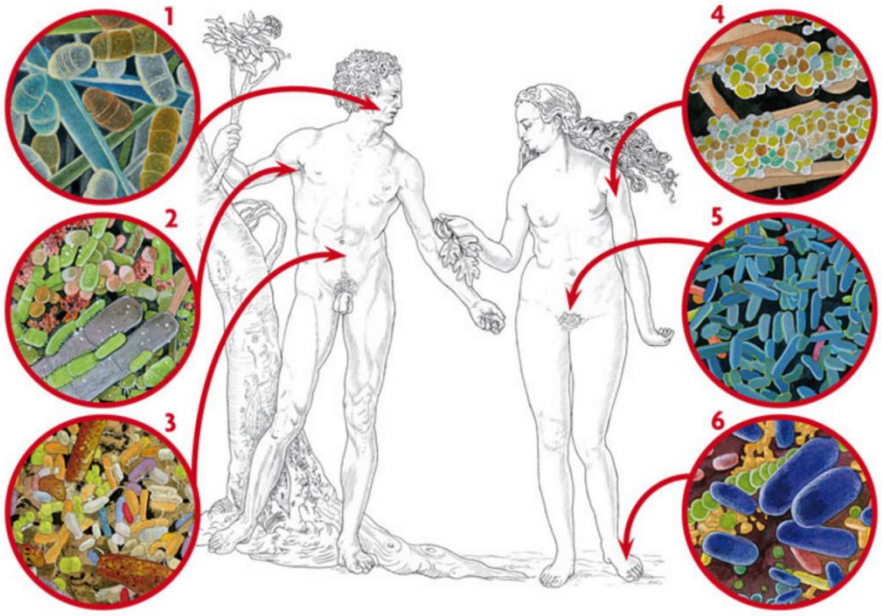


Figure 2. The human body offers a great diversity of habitats. For example: 1) Oral microbiota, 2) Microbiota of the armpit, 3) Gut microbiota, 4) Skin microbiota, 5) Vaginal microbiota, 6) Microbiota of the foot, between the toes. (*Source: [30]*)

2.1.1 Functions

While every human has a unique gut microbiota composition, its physiologic functions are basically the same. [31]

The term "microbiome" indicates the overall genome of the ecosystem, in this case in the intestinal environment. These codes for a metabolome (set of all metabolites produced by the community) that is more multifaceted than that of man, since the combined genomes of all organisms present in the microbiota (bacteria, archaea, eukaryotes, and viruses) constitute a wealth of millions of genes capable of carrying out metabolic functions that our organism has not developed in the course of evolution, precisely because they are already supplied by symbiont bacteria. Taken altogether, the microbiota has an overall genome 100 times bigger than the human one,

which allows a great contribution to the host metabolism and an essential role in food conversion into nutrients and energy. [32, 33]

Each bacterial community will communicate with other microbes by several mechanisms: they can exchange molecular and genetic material via horizontal transfer, and alter their collective behavior on a population-wide level in response to surrounding species or stimuli, like temperature or pH variations. Among others, these mechanisms remodel the microbial composition of the community. [34, 35, 36] Furthermore, a great quantity and variety of extracellular enzymes, toxins, antimicrobial compounds, inflammatory cytokines, and metabolites generated by the bacterial community within the host can modulate host physiology. [37, 38]

Through dynamic microbe-microbe and/or microbe-host interactions, the diverse microbiota is not only established and maintained but also contributes beneficial or pathological influences to host health.

One of the main microbiota functions is its crucial role in the immune system. The delicate balance maintenance in the immune system is crucial for the health status of a host, and it occurs with the invasive pathogen elimination. A self-tolerance towards the beneficial bacteria is kept in order to avoid autoimmunity issues (tolerance that often lacks in patients with autoimmune disorders). [39]

Resident bacteria are a fundamental line of resistance to colonization by exogenous microbes. They actively regulate the production of nutrients by the host through a negative feedback mechanism, in order to prevent the availability of nutrients for potential pathogens. Besides, they compete for attachment sites on the brush border of intestinal epithelial cells and produce antimicrobial substances, inhibiting the growth of their pathogenic competitors.

Furthermore, the gut microbiota and its metabolites stimulate the immune system. Given the bacterial ability to affect intestinal permeability [40], mucosal immune function [41], and the activity in the enteric nervous system [42], the gut microbiota has also been shown to be involved in modulating gastrointestinal functions. Additionally, several studies suggest that the microbiota and its metabolites are likely to be involved in modulating behaviors and brain processes such as repressiveness to stress [43], anxiety, and depression [44]; eating behaviors [45]; and brain biochemistry. [46]

The intestinal mucosa represents the main interface between the immune system and the external environment, and the collaboration between the host and the bacteria seems to play a role in the development of the immune system itself. As a result of this interaction, there is a protective response on the part of commensal bacteria, i.e. an inflammatory response to pathogenic organisms or the triggering of apoptosis. The commensal bacteria of the gastrointestinal tract therefore play an active role in the development and homeostasis of the immune system. [33, 47]

2.1.2 Composition

The gut microbiota is made up of 500-1000 species, dominated by anaerobic bacteria. [48] They have many different lifestyles and abilities, and their composition and distribution vary depending on the age, health status, and diet of the host. Each specific site taken in exam is characterized by a different microbiota composition one from each other and, thus, since the gastrointestinal tract extends from mouth to anus, there are several microbiota depending on the sample collected. [49]

In the stomach, microbiota composition is characterized mainly by *Lactobacillus* and *Streptococcus,* which reside in the mucosal layer that

covers the gastric epithelium. Among the Proteobacteria in this organ, of particular importance is *Helicobacter pylori*, whose ideal habitat is the gastric mucus, because it produces metabolites that are toxic for the mucosa and cause the onset of several diseases like gastritis and ulcer. [50]

Bacteria and fungi increase their number, density, and complexity proceeding along the gastrointestinal tract: from the proximal gastric and duodenal, with a prevalence of aerobic organisms, the population in the cecum and the colon becomes mainly made up of anaerobic bacteria, due to the decreased oxygen availability. Three bacterial phyla dominate the gut microbiota in western adults: Firmicutes, Bacteroidetes, Proteobacteria, and Actinobacteria. [51]

As for the small intestine, few studies have been conducted due to the difficult sampling. However, some molecular analysis studies have been carried out on bacteria associated with mucous membranes, and these have reported the enrichment of the species of *Streptococceae* and *Lactobacillales* (*Bacillus* subgroup of Firmicutes), of *Actinomycinaeae* and *Corynebacteriaeceae* (an *Actinobacteria* group), with a reciprocal decrease of the species *Clostridia* and *Bacteroides*. [52]

In the colon/gut microbiota, the Firmicutes phylum is the most abundant one, comprising more than 200 genera, followed in a lesser amount by the other phyla. In fact, the typical colon microbiota of a healthy patient is generally composed for more than half (about 65%) of Firmicutes, 25% of Bacteroidetes, 8% of Proteobacteria, about 5% of Actinobacteria, and then from other minor phyla [49, 53], although the relative abundance of these phyla can vary greatly from individual to individual, especially in response to various environmental factors or differences in the diet and lifestyle of the individual. [54]

Many studies have demonstrated the beneficial effects, direct or indirect, on the host by prebiotics and probiotics (respectively, nutrition metabolized by the microbiota and living microorganisms consumed with the diet), such as the recovery of the microbial composition following the use of antibiotics or other medications. As fermentable carbohydrates and, therefore, nutrients for beneficial bacteria, prebiotics help improve the functionality of the microbiota by stimulating the growth and activity of probiotic bacteria (*Bifidobacteria* and lactic acid bacteria) and, in turn, having beneficial effects on the gastrointestinal and immune systems. [55] Probiotics, on the other hand, are present in some fermented products such as yogurt and help the intestinal microbiota to maintain its balance, integrity, and diversity. [28] However, there is a lack of evidence of the impact of probiotic supplements on the composition of the fecal microbiota in healthy adults, as reported in a recent study. [56]

On the other hand, a main factor that can perturb the composition of the microbiota is the use of antibiotics. These have a profound effect on the microbiota and their misuse or excessive use is associated with an increase in antibiotic-resistant pathogens. Studies have been conducted that prove large alterations of the microbiota following treatment with antibiotics. [57] Although specific taxonomic groups affected may vary among individuals, some of these do not recover even months after treatment and, in general, there is a decrease in bacterial diversity in the long term. The bacterial community regenerates after antibiotic treatments, but a reduced resistance to colonization is observed, which allows external microbes to be able to supplant the commensal ones causing permanent changes in the structure of the microbiota and changing the disease state. [58]

2.1.3 Development through life

For many years, it has been assumed that the human fetus was microbiologically sterile and that bacterial colonization began during and after birth, when the newborn comes into contact with the vaginal canal microbes, the environment, the mother's skin and her breast milk. [23] The thought that the womb is sterile and, accordingly, that a neonate's microbiota starts at birth was an accepted fact. However, recent studies suggest microbial communities exist in the placenta, amniotic fluid, and meconium: the human fetus could not reside in a sterile environment. [59]

As for the bacterial colonization of the intestine, data about the fetus microbiota is still lacking, due to the difficulties in collecting a sample before birth. As for the newborn's gut microbiota after birth, several studies show how it is a gradual process. Typically the first colonizers are facultative anaerobic bacteria, followed by strict anaerobes such as *Bifidobacterium*, *Bacteroides*, *Clostridium,* and *Eubacterium* [60, 61]; the change in the breathing mechanism and consequently the change of bacteria occurs thanks to the progressive decrease in oxygen present in the gastrointestinal tract caused by the metabolism of facultative anaerobes.

There are many factors that have an impact on the development of the baby's microbiota, such as the mode of delivery, the age of the pregnant woman, infant hospitalization, antibiotic therapy, the mother's microbiota and the mode of feeding during the early stages of life [62, 63]; geographic location is also very important, as demonstrated in a study among European countries. [64] It seems that newborns born with complete gestation and vaginally and who were then fed only by breastfeeding during the first months of life have the healthiest microbiota, while children born by cesarean section have a delayed colonization and fewer *Bacteroides* and *Bifidobacteria* [65]; the latter are also more colonized and to a greater extent

by *Clostridium*. The differences between the two groups of newborns persist up to one year of life. [66]

During this first year, a large fluctuation in the composition of the microbiota is observed, mainly due to the weaning and gradual introduction of new foods and solid foods, but also to the gradual exposure to environmental bacteria. The microbiota reaches stability only between the first years of life, after weaning is completed, when it becomes similar in composition to that of an adult individual. [67]

The complex microbial community of an adult individual typically consists of a few hundred species, but in some individuals the species richness can run into the thousands. [68, 69] The estimate of the richness and complete microbial diversity in the human population varies from several thousand up to 40,000 species [70, 71] increasing with development and reaching the highest complexity in the individual. [72]

Each individual has a growing microbial composition that remains relatively stable throughout most of an adult's life. Although the individual microbial composition has an "individual core" that varies at the bacterial phylum level and depends on the depth of the analysis [73, 74], the general phylogenetic profile can be categorized into a limited number of well balanced symbiotic states between microbes and host, the so-called enterotypes. [75]

In the later stages of life, the composition of the microbiota changes again and becomes less diverse and more dynamic, characterized by an imbalance from Bacteroides to Firmicutes, an increase in Proteobacteria and a decrease in *Bifidobacterium*. [76] As a result, older people have substantial microbial differences with younger people and it is also interesting that the microbiota varies significantly even beyond the age of

70-80, with differences in the composition in centenarians and semi super-centenarians (105-108 years old). [77, 78]

2.1.4 Diet influence

Diet is a major factor shaping the composition of the intestinal microbial community, as what the host ingests can provide the correct nutrients for bacterial growth. [79, 80] For example, certain types of dietary fibers, known as microbiota-accessible carbohydrates, represent an essential energy source to a healthy intestinal microbiota. [81]

The several different dietary habits around the world shape in different ways the population gut microbiota:for example, variation across human populations have been observed between the “Western diet” typical of peope living in metropolitan areas and the dietary habits of rural areas. [82] Of particular notice is the comparison of the bacterial community of a native population, the Hadza of Tanzania, and Italian subjects [83], which lead to the hypothesis that the peculiar hunter-gathers GM could enhance their ability to digest and metabolize better the nutrients from fibrous plant foods.

Apart from these dramatic composition shifts, the gut microbiota within a person responds rapidly to dietary changes throughout the entire lifespan. Bacteria are regularly purged through the colon but have the ability to double in number within one hour; within 24-48 hours of a dietary intervention, rapid changes are thought to alter the microbial composition on a genus and family level. [84]

The extensive dietary changes commonly adopted in the population, such as vegetarian or vegan diets, can alter the gut microbiota composition in specific and peculiar ways, primarily influenced by a higher fiber amount derived from fruits, vegetables, and plant-based foods. [85]

The ketogenic diet (KD) is extremely fiber-deprived, with high fat and low carbohydrates consumption. This diet has proven efficacy for refractory epilepsy [86, 87], and has been reported that it could significantly reshape the gut microbiota of epileptic infants. [88] Recent studies have investigated changes in the gut microbiota in adult patients with epilepsy during KD, showing a diminished relative abundance of *Bifidobacteria* species and other fiber-consuming bacteria. [89]

The mechanism activated by the KD is the ketosis, i.e. the conversion in the liver of the dietary fat in a high amount of ketone bodies circulating in the blood. These ketone bodies are used as an “alternative energy substrate for ATP production in the cells of the body including the brain”. [90] Many biochemical, metabolic and hormonal changes are induced by this energy shift, and these changes may contribute to the number of seizures reduction, decreasing the neuronal excitability. It is currently unknown whether fiber intake change or influence on the gut microbiome contributes to the anti-seizure effect, or whether this is only a potentially problematic consequence for the gut microbiome when increasing the dietary fat intake. [90]

2.2 Gut-brain axis

Comprising ~150 times more genes than the human genome [91], the gut microbiota offers a multitude of enzymatic reactions. Health-promoting metabolites, such as vitamins and short-chain fatty acids (SCFAs), are produced by bacteria and impact both gut and peripheral health via several gut-organ-axes. One of such axes, the microbiota-gut-brain axis has gained tremendous attention in the research community in the past years.

The gut microbiota and the brain can communicate bidirectionally through the central and enteric nervous systems as well as endocrine, immune, and

metabolic pathways. [92] (*Figure 3*) It has been shown that the intestinal microbiota influences the central nervous system (CNS) physiology and neurochemistry impacting behavior, cognition, mood, anxiety, and depression. [93] Intestinal dysbiosis has been associated with a variety of neurological disorders, e.g. autism [94], multiple sclerosis [95], Parkinson's [96], and Alzheimer's. [97] The composition of the gut microbiota is influenced by environmental factors and to a lesser extent by host genetics. [98]

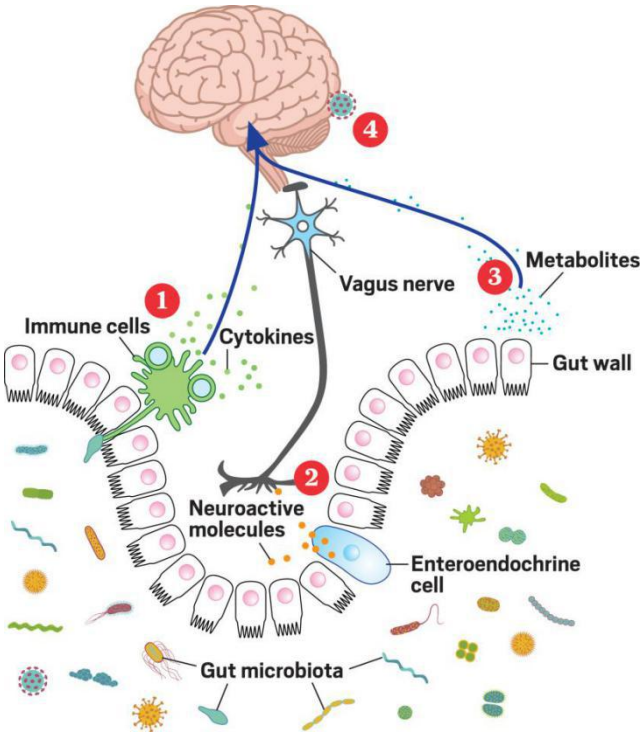


Figure 3: Possible mechanisms of gut-brain axis interactions. 1. The interaction of the GM with the immune cells in the gut, prompts them to make cytokines circulating from the blood to the brain. **2.** Enteroendocrine cells produce neuroactive molecules and peptides, interacting the vagus nerve, which sends signals to the brain. This mechanism could be stimulated by the GM. **3.** Neurotransmitters can be produced by the bacterial species in the gut. These metabolites circulate to the brain, where some of them are small enough to cross the blood-brain barrier, while others alter cell activity at the barrier itself. **4.** Recently gut bacteria have been found in human brain tissue, suggesting that microbes might somehow be making their way into the brain. (*Source: Adapted from [99]*)

Recent advances in microbiology have characterized the functional interactions between microbiota and host. This “forgotten organ” is now considered not only as a key player in human homeostasis but also as a direct/indirect causal agent in influencing various diseases such as allergy, irritable bowel syndrome, type 2 diabetes mellitus, obesity, cancer, and neurological diseases. [100] The gut microbiota is thought to be connected to the disease through the so-called “gut-brain axis”, a way of interactions between the brain and the microbial community living in the gastrointestinal tract.

In the last few years, the potential contribution of the microbiota in our gastrointestinal system to CNS diseases has opened many research and future opportunities. The CNS and the enteric nervous system (ENS) of the gastrointestinal tract are tightly connected by hormones, neuromodulators, and neurotransmitters related to efferent/ afferent nerves including the vagus nerve. As most (>95%) of the microbiota in our body are gastrointestinal residents, these microbes are perfectly situated to react to and influence neuronal, humoral, metabolic, or immune signaling underlying the gut-brain relationship.

On the one hand, as an example, elevated levels of noradrenaline in the gut-lumen (caused by stress, for example) can influence gene expression or abundance of some bacteria. [101, 102] On the other hand, mice grown in a germ-free environment demonstrate exaggerated stress responses and anxiety behaviors related to changes in the hypothalamic-pituitary-adrenal axis. [103] Similarly, disruption of the maternal gut microbiome perturbs the neurodevelopment of their offspring, which show autism-like behaviors that can be rescued by introducing the commensal bacteria species *Lactobacillus reuteri*. [104, 105] These studies emphasize this bidirectional

relationship and support disruption of gut homeostasis as a potential risk factor in psychiatric or neurological disorders.

To date, high-quality evidence about alterations of microbial ecology or the production of microbial-derived metabolic products in human patients with brain or gut-brain disorders is limited. [106] There is still little clinical information about the gut-brain axis mechanisms in diseases, and if acute alteration of the intestinal microbiota actually effect clinical symptoms. [107, 108]

Furthermore, it is still up to debate whether the patients with these disorders have gut bacterial alterations due to an imbalance from the gut microbial interface (bottom-up theory) or changes the signal transmission from brain to gut (top-down). [109]

2.3 Gut bacteria, diseases, and epilepsy

In public opinion, bacteria have been seen as harmful to our health and associated with dirty environments. Modern microbiology recognizes the existence of a bacterial species spectrum, depending on their role and interactions with or inside the host. In fact, the majority of bacterial species that make the gut microbiota composition are either harmless or beneficial (respectively called commensal bacteria and symbionts); relatively few types of bacteria (called pathobionts) expose the host to a certain risk, infections, or diseases. [110, 111]

Most of the time, in a complex environment as the intestinal one, the pathobiont bacteria are usually suppressed thanks to the other huge amount of beneficial gut bacteria that promote the regular balance. [112] There is no clear definition for the characteristics of a “healthy” human gut microbiota

community, despite the several and important research projects aimed to this issue, such as the Metagenomics of the Human Intestinal Tract, MetaHIT, and the Human Microbiome Project, HMP. An average composition can be deduced by healthy donors, but the multitude of factors that influence the microbiota is so important that defining a sample or patient truly healthy can be complicated.

Nonetheless, the composition described before (*Introduction, 2.1.2 Composition*, page 12) is a general information about the gut microbiota composition in non-diseased patients. Alteration in the host's health and/or lifestyle may lead to the loss of balance in the microbiota, which can remain altered during the entire period of alteration; when it reaches a modification so deeply different from the normal composition, this kind of gut microbiota is called to be in a state of "dysbiosis". This status can be connected to chronic diseases, intestinal function disorder, intestinal inflammation, stress condition, allergies, and many other situations.

For what concerns epilepsy patients, there is still little information about the gut bacterial alteration that can occur, and how it could be linked to seizures or drug therapy. [90] The majority of research studies focus on case-reports or small cohorts, for example with a fecal microbiota transplant, [113] and on patients in KD [89]. Since this diet has been proven to profoundly influence the gut microbiota with or without a disease, [114] despite its recognized role in reducing the number of seizures, further investigations are needed to understand which bacterial species are characteristic of an epileptic patient with no man-made alterations.

3. The analysis

The microbiota is the subject of the metagenomic analysis, a term that indicates the usage of modern genomic techniques for studying the entire gene content of all bacterial communities present in a sample, avoiding the necessity for the isolation and cultivation of the singular species. This has been particularly relevant for anaerobic bacteria, which make up the majority of the microbiota.

Given the huge number of bacterial groups present in each microbiota sample, singular species determination is not possible: to obtain the microbiota composition, it is necessary to use the so-called “next-generation sequencing” techniques and a bioinformatic procedure, essential for the identification of specific patterns in patients.

3.1 Sample types

Metagenomic analysis in humans can be performed from potentially any kind of sample: studies have been conducted on saliva, tissues (epidermis, mucosal), nasal swabs, genitalia, but the most common sampling is feces.

For the great influence of diet and the numerous niches and species present in the gastrointestinal tract, fecal samples are the most studied and commonly used for gut-brain axis studies to characterize bacterial presence and differences among hosts. Furthermore, fecal samples are easy for the patients to collect and provide, via non-invasive and inexpensive procedures, and the bacterial DNA extraction has several solid protocols to follow that minimize the human error and ensure a clearer result. For these reasons, for this present study, fecal samples have been used.

3.2 The sequencing necessity

The main objective of human microbiota research is to observe and measure the microbial community structure and dynamics, which are the relationships among its members, which substances are produced and consumed, and, eventually, the interaction with the human host and what distinguish one microbiota to another considering the patients' different status (health, nutrition, drugs, etc.).

Until over a decade ago, gut microbiota comprehension was limited by technical issues, as bacterial community members have always been identified in vitro through comparisons among colonies using their physiologic/phenotypical characteristics. This culture-dependent methods needed the microorganism to be isolated in a lab to be studied on selective media. Given that the human microbiota hosts several hundreds of species and that there are more than 10^4 bacterial cells per gram or milliliter of luminal content, this culture method became insufficient for the metagenomic study; furthermore, many bacterial species in the gut are strictly anaerobic and have never been cultivated.

New genetic technologies allow the identification of the complex gut metagenome and its influence on the health and diseased status of the host. One of the main limitations of this sequencing method is that by identifying the DNA present in the sample, we cannot tell whether it belongs to a living or dead microorganism, if it was present due to the host defense or if it was at the end of its life cycle. To have a clearer picture of the bacterial presence and activity, metagenomic data has to be integrated with transcriptomic, metabolomic or proteomic data in order to see which bacteria were actually active in the sample analyzing the transcripts, the metabolites produced or the protein expressed, respectively.

3.3 Obtaining the base sequence

After extracting the bacterial genetic material from the patient's sample, the microorganisms can be identified through determination of their nucleic acid composition. This process is called "sequencing", as it allows to collect the nucleotide sequence of the bacterial DNA or RNA.

The most popular sequencing method used nowadays is the so-called "next-generation sequencing" (NGS)[115], as it followed the first sequencing reactions made by Frederik Sanger in 1977 [116] and his strategy for base reading and error percentage: the speed, the simplicity and the output (from thousands to hundreds millions of sequences per sample) of the new processes determined the huge success of many companies and platforms, each based on different working strategies. The platforms on the market differ for the molecular mechanisms used for obtaining the nucleotide sequence, but they all have the common idea of the sequencing by synthesis: Roche 454 uses the pyrosequencing, while Illumina exploits the bridge amplification for emitting the fluorescent light; Ion Torrent uses the detection of hydrogen ions released during the polymerization of a template DNA strand; among the newest instruments, PacBio an uninterrupted template-directed synthesis of an entire DNA molecule in real time. [117, 118, 119]

Each of these and of the non-listed sequencing approaches and platforms have strengths and weaknesses, and must be accurately chosen on the basis of the sample type, the quality and quantity of output desired, the sequencing depth and the budget in disposal. [120] For example, for a more accurate and wide data, as species or strain taxonomic assignments and functional pathways identification, a Whole Genome Sequencing (WGS) is commonly used. [121] This technique allows to obtain data about the entire bacterial genome, not only from a selected portion as with the NGS methods,

and the functional characterization of the microorganisms in the samples. Despite the greater features, WGS is less frequently used than the NGS as it has a much higher cost (about 10 times more expensive): the NGS choice provides a rapid and economic choice for characterizing the microbiota composition, with a good resolution, even if not enough to get reliable information below the genus level.

The platform used in this present study is the MiSeq instrument, produced by Illumina (San Diego, CA, USA), released on the market between 2012 and 2013. It can process up to 384 samples per time, obtaining around 15 GB of sequences. [122] The high amount of samples that can be submitted in one analytic process (also called a “run”) is due to the multiplexing strategy, which consists in the identification of the single samples through short known nucleotide sequences ligated at the end of the fragments generated. The short nucleotide sequences attached are called “index” or “barcode”, and the fragments produced by the bridge amplification sequencing happening in the MiSeq are called “reads”. By separating the indexes, it is possible to extract and identify the genomic content of each sample.

What is assembled with the MiSeq machine, though, is not the entire genomic content of all the bacteria present in the samples, despite the gigabytes size would suggest a consistent data amount. Each run produces up to 15-20M reads, with a maximum length of 300 nucleotides, in a so-called paired-end method, which consists of the sequencing of the two terminal ends of the amplified fragment. The sequencing target is a ribosomal protein, a characterizing portion of the bacterial genome, that allows the most possible identification and quantification of the organisms in the sample: the 16S sub-unit of the ribosomal RNA (rRNA) gene.

3.3.1 The target gene

The 16S rRNA gene is now considered the cornerstone of bacteria and archaea classification after the American biologist Carl Woese, in 1977, used molecular phylogenetic techniques applied to this component and to the gene that codes for it. The 16S rRNA is 15,9 kilobases long and has 9 hypervariable regions, denominated from V1 to V9. These regions are alternated to highly conservative ones, common to all bacteria and archeae. The conserved regions are targeted by “universal” primers that allow the sequencing and identification of the hypervariable ones, which are amplified through a polymerase chain reaction (PCR) before sequencing: they are peculiar for each species and, therefore, allow the taxonomic distinction - even for unknown species.

This step produces a library of double-stranded DNA fragments that can be recognized by the MiSeq sequencer through specific adapters. The DNA amplified from all the samples, then, is pooled together and submitted in the machine with the reagents needed for the sequencing reactions. The sequencing itself takes place on a flow cell, a thin glass chip with a lane containing billiards of templates deposited in fixed positions, which grant a stochastic distribution of the reads clusters, in order to provide a better and clearer base signal, producing a bigger data output. [123]

The DNA pool is heated and one of the double strands is cut and washed off, leaving a single strand on the flow cell. At every sequencing cycle, the four nucleotides (A, T, C, G), marked with a fluorescent molecule, are added on the flow cell and they pair with the complementary base on the single-strand DNA fragment. The sequencing reaction stops at every base added, and the picture derived by the laser excitation of the last fluorophore added is collected through a CCD camera (Charge-Couple Device). The fluorescent marker is removed from the nucleotide paired and a new base is added.

This step is repeated at every cycle, until the sequence is over and the sequenced strand gets washed away and the single-strand hangs down on another template, flipping the nucleotide molecule, and the sequencing process starts again, from the other end of the fragment.

The MiSeq sequencer, in the end, delivers the conversion of the luminous signal in an analyzable output and removes the multiplexing barcode, preparing the reads for the following step. The reads length obtained in the sequencing process is limited by several factors that cause a decadent signal: a lower read quality could be due to incomplete cuts of the fluorescent markers or of the terminal portions anchored on the flow cell; nucleotide mis-call or mis-assignment, rather than a base insertion or deletion is the most common error type.

3.4 *The bioinformatic approach*

All the analysis performed for the study presented in this thesis have been carried with the Quantitative Insights Into Microbial Ecology software (QIIME): an open-source software designed for the transition from the raw sequencing data to its interpretation and its deposit into a public database. [124] It has been developed a decade ago and, with its several releases and new features that cover the majority of researchers biological questions, it is still considered the main software for 16S analysis, despite the competition of similar packages.

Once the sequencing reads have been obtained from the MiSeq, the data analysis with QIIME consists of three steps: 1) reads quality assessment; 2) clustering and alignment; 3) use of these data to produce representations of the similarity of communities such as abundance curves, biodiversity plots, and ecological and statistical descriptors of the bacterial community structure. [125]

3.4.1 Reads quality assessment

The first step comprises all the processing and filtering procedures necessary to prepare the data for the bacterial analysis and interpretation.

16S rRNA gene data analysis quality is crucial for a correct classification of the microorganisms involved. The reads quality and length are detected in the first place. Pre-processing methods for trimming the bases read by the sequencing platform are commonly used for filtering the reads and discarding the ones with a lower quality, reducing the number of wrong reads. [126, 127] One of those methods consists in the usage of a *quality score* for trimming the length of low quality reads or for cutting a fixed number of bases at the end of the reads, where the sequencing quality is typically lower. Quality, in Illumina-based sequencing, similarly to the Sanger-based one, is expressed according to Phred score (usually in the range of 0-40), and is determined as the overall probability of having an incorrect base call in a given position along the sequence. Phred scores are logarithmically related to base-calling error probability ($Q = -10 \log_{10}(P)$). [128, 129]

Putative chimeras are detected along with the quality of the reads. These particular reads are made up of fragments coming from several 16S genes, produced by the artificial recombination of two or more DNA molecules during the PCR amplification. They are created when the PCR product from a sequence is ligated to another sequence, acting like an extension primer for it, and thus generating a DNA molecule from two templates. Similarly to sequencing errors, chimeric sequences can be erroneously interpreted as new amplicons and, therefore, could be wrongly used for bacterial diversity estimation. [130]

3.4.2 Clustering and annotation

The second step generates the data tables needed for subsequent interpretation.

Before proceeding taxonomic assignment and quantification, sequencing reads are subjected to a clustering procedure, aimed at the identification of clusters of (uncultivated or unknown) organisms, by DNA sequence similarity. These clusters are defined as “Operational Taxonomic Units”

(OTUs), which are pragmatic proxies for species at different taxonomic levels. These OTUs are the base of a clustering theory: given all the possible mutation and sequencing errors that can happen during the PCR reactions, it is not possible to assume that each read represents a different bacterial species. To remedy to this issue, a comparison among reads is carried out to align them and to see the most similar ones. If a group of reads present a homology out of the entire sequence higher than a given threshold, these reads will be clustered and will form an OTU.

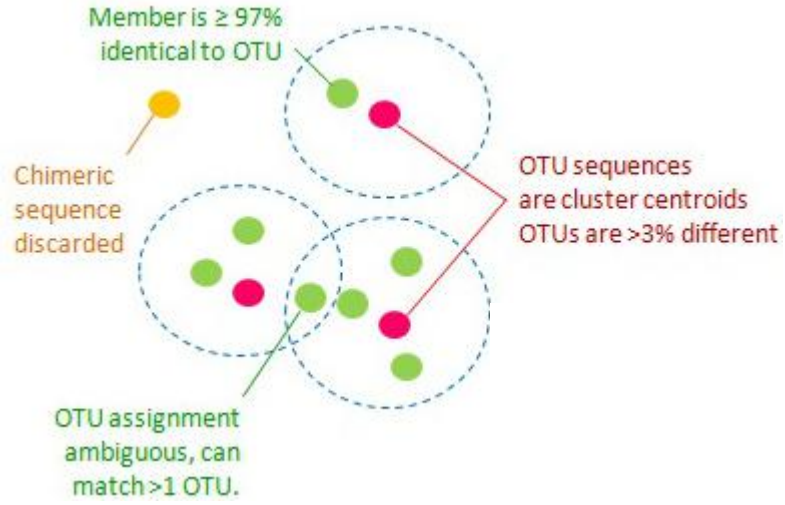


Figure 4. OTU clustering. The reads are clustered together according to a given similarity threshold (97%) in order to have OTU sequences with less than 3% dissimilarity. Ambiguity rises when a read is not clearly assigned to one OTU; reads not clustered in any OTU are labeled as “chimera” and discarded. (*Source: [131])*

A percentage threshold of 97% is nowadays accepted as a standard value and as a good approximation following the conventional rule that 97% identity approximately defines members of the same bacterial species: a lower value would group more reads together and lower the OTU number. [132] This clustering is crucial in the analysis as it could allow the distinction of more species than the ones actually present, grouping “fake species” derived from the mutation and sequencing errors mentioned above in the *Introduction* (*3.3.1 The target gene*, page 27). In other words, clustering reads for their 97% homology means that in one OTU there will be only reads that have no more than 3% difference between them. (*Figure 4*)

Once the OTUs have been determined, taxonomic annotation follows: this is a procedure necessary to understand which species the grouped sequences come from and to give them a unique taxonomic identification name.

The taxonomic assignment is what allows a proper bacterial classification and identification. “Taxonomy” (from the Ancient Greek ταξις, taxis, “sorting”, and νομος, nomos, “rule”) is the classification discipline. Biology taxonomy is part of the systemic biology, which defines taxon groups and studies the relationships between organisms and their evolution history. A taxon, in biology, is the grouping of organisms real, morphologically and genetically distinguishable from others and recognizable as a systemic unit, placed in a hierarchic structure of the scientific classification.

From general to particular, the most used taxonomic categories are: domain, kingdom, phylum, class, order, family, genus, species. These categories are hierarchically ordered: each genus integrally includes one or more species, each family one or more genera and so on. Moving up in the hierarchy, the characteristics common to group members are fewer in number but more

and more basic, and likewise. [133] The purpose of classifying organisms into formally defined taxa, rather than through informal groups, is to provide groups whose constituency (i.e., which organisms comprise them) remains within a narrow scope and whose Latin name has universal value, regardless of language used for communication.

The alignment of the reads to the reference genome dataset produces tables and depth of the sequencing coverage for the reads of each genome and return taxonomic data and abundances. [134] Not all reads match sequences in databases because not all organisms have reference genomes, due to culture difficulties or to the fact that many species are still unknown, or to difficulties in the assignment of sequenced bacteria to known clades. Choosing a good database is important, as there are several available choices, differing for the approaches used to perform taxonomic establishment. The most widely used in microbiota analysis is GreenGenes database, [135] but its taxonomic terms proposed from phylogenetic methods have not been maintained since 2013 (version 13.8). The most updated databases are: SILVA ribosomal RNA database project, [136] RDP, [137] and that developed by the National Center for Biotechnology Information (NCBI BLASTN, [138]). The one used for this present study is SILVA, which counts on yearly updates and is, therefore, considered accurate and reliable.

In this study, the Ribosomal Database Project (RDP) classifier, a naive Bayesian classifier, was used to assign sequences derived from bacterial 16S genes to the corresponding taxonomy model, providing taxonomic assignment for phylogenetic level down to the genus. [139] As species are reevaluated year after year and microbiology discrepancies are resolved or risen, the classifier is constantly updated to correctly assign the 16S sequences.

3.4.3 Data usage

The third and last step uses the data derived from the previous ones and produces phylogenetic trees, needed for diversity analysis and ecology descriptors. Statistical tools to describe the differences among bacterial population in patients are used as well. [140]

An important concept to address when dealing with OTUs and taxonomic classifications is the diversity among populations, such as the number of distinct groupings in a sample or in the original population. Many bacterial species are abundant in some individuals and many organisms prevail in the majority of individuals, as stated before (*Introduction, 2.1.2 Composition*, page 12). Although microorganism composition can vary over time and show some similarity between individuals, our intestinal content appears to be highly personalized when considered in terms of microbial presence, absence and abundance.

The description of an ecological community (such as that of human microbiota) can be conducted using two of the terms introduced by Whittaker [141] for measuring biodiversity over spatial called: alpha-and beta-diversity.

Within the same community (in this kind of studies, the sample taken from an individual), alpha-diversity is measured: this is an estimate of the wealth and distribution of taxa in a single population. Richness is the description of the number of species in a sample (determined by the number of OTUs), while diversity is the measure that combines richness and how the different species make up the gut microbiome. [141] The rarefaction tables (i.e. estimation of the alpha-diversity of the dataset on reads subsets, at increasing reads depth) are the basis for measuring alpha-diversity,

because they reflect the difference between samples based on the abundance of taxonomic varieties in a community. Moreover, they are used to determine if the sampling was sufficient to accurately characterize the bacterial community studied, evaluating whether or not the alpha-diversity measures at increasing depths reach a plateau.

The QIIME pipeline provides various metrics for this analysis. The ones used in this present study are: Chao1, Observed Species, Shannon and PD_whole_tree. Chao1 provides an estimate of the minimum number of species in a sample and is an abundance-based coverage estimator of bacterial richness that gives more weight to the low abundant species. [143] Observed Species counts the unique OTUs found in the sample. Shannon provides information on observed OTU abundances taking into account both wealth and uniformity and calculates the population diversity in samples, measuring bacterial diversity based on OTU counts; it has a greater weight on species richness even if it provides more inference about the community composition, contemplating also the relative abundance of the species present. [144] PD_whole_tree is the Faith's Phylogenetic Distance, the only phylogenetic metric used in this study, and it requires a phylogenetic tree to provide results on the microbial differences in the community; it calculates the phylogenetic distance between OTUs by measuring the total branch length for the OTUs in each sample. [140] When comparing sample groups, permutation-based tests can be used to evaluate if a certain experimental condition is characterized by a significantly higher or lower biodiversity.

The term beta-diversity, on the other hand, is used for cross-sample diversity analyzes. Samples are separated by comparing common OTUs or by viewing relative abundances of different OTUs. To compare the samples, a distance matrix is constructed that quantifies the differences between them, looking at the taxonomic groups present and/or the phylogeny or

abundances of different species. Among phylogenetic distances, the most commonly used is the Unique Fraction method (“Unifrac” distance, [145]), which measures the phylogenetic distance between sets of taxa in a phylogenetic tree as the fraction of the branch length of the tree that leads to descendants from either one environment or the other, but not both. A quantitative measure (weighted Unifrac) is ideally suited to revealing community differences that are due to changes in relative taxon abundance; the qualitative measure (unweighted Unifrac) is most informative when communities differ primarily by what can live in them, in part because abundance information can obscure significant patterns of variation in which taxa are present. [146]

The Principal Coordinates Analysis (PCoA) is a technique of linear decomposition analysis that is used to make comparisons between samples, finding linear combinations of different markers and separating the points of the graph (the samples) into clusters corresponding to the different groups, representing inter-object dissimilarity in a low-dimensional Euclidean space. The points are plotted in a multidimensional space in such a way that the distance between them is as close as possible to that calculated in the Unifrac matrix, the most commonly used phylogeny-dependent similarity metric: this distance can be “unweighted” if it considers the similarity between the sequences of the samples but does not compare the abundances between them; it can be “weighted” if the distance considers both the similarity between sequences and their abundance. [147]

The coordinate that separates the points most is the first main coordinate (PC1), the second best is PC2 and so on. Each principal coordinate has a percentage value that represents the percentage of variance expressed by that dimension. With two quantitative variables such as the principal coordinates, the graphical representation is a scatter plot. The points (the

samples) are plotted in a graph where the axes represent two out of the coordinates (or, in a three-dimensional graph, three coordinates together). In order to better visualize similarities and differences among groups (i.e.: healthy subjects from diseased ones), data points are usually colored according to categories and a statistical test, such as the “Adonis” test [148]: the analysis of variance using distance matrices, for partitioning distance matrices among sources of variation, and fitting linear models to distance matrices, which uses a permutation test with pseudo-F ratios, implemented in the R language.

Taxonomic abundances can provide a consistent variety of applications. Relative abundances can be compared among experimental groups usually by employing non-parametric tests such as the Mann-Whitney U-test, in order to find increased or depleted taxa in a group of samples. Moreover, they can easily show the microbiota composition, but also help suggest interactions among them. This can be achieved through a *co-abundance analysis*, in which statistical correlation (such as Pearson or Spearman) is implemented to hypothesize if certain groups of bacterial species frequently occur together and similarly in the samples. This type of analysis generates a correlation matrix and a hierarchical tree of interaction, that can be used to identify bacterial clusters. In each cluster, called co-abundant group (CAG), the bacterial species grouped together share more patterns and traits based on their common abundance in the samples. The interactions in terms of common presence or absence can be extrapolated from the matrix as well, and a co-abundance network can be generated. [149]

The correlation matrix can be generated not only to compare the relative abundances of bacterial species, but also to have a glimpse of how they are related to non-taxonomic variables such as nutrient intake from the host, SCFAs detected or other metadata collected from the patient, for example

the body mass index (BMI), the day of sampling (in case of a longitudinal study), or physical values. A linear regression model can be implemented for a *feature selection analysis* able to discriminate whether these metadata can be informative in the research for the bacteria more probably crucial or characteristic in the dataset. Machine learning is the study of computer algorithms that improve automatically through experience. [150] This kind of analysis is one of the core concepts in machine learning, a term coined in 1959 by Arthur Samuel, pioneer in the artificial intelligence field. Its methods are particularly useful for recognizing patterns in complex datasets such as the gut microbiota. [151]

The model works with the features (e.g.: the high number of bacterial genera present in the samples), and determines which ones are more informative distinguishing between them accordingly to importance scores. [152] The model works dividing samples according to a status label and creating two subsets of data: a training set, used to train the model, and a test set, on which the model is applied to determine whether it works by classifying the samples correctly.

AIM OF THE PROJECT

Recently, changes in the gut microbiota have been linked to a possible role in epilepsy development and to seizure susceptibility. [90] To date, the connection between microbiota and epilepsy is still unclear as there is limited information, in literature, about the bacterial composition and eventual changes in patients with epilepsy, in particular in children.

During my Ph.D., thanks to the collaboration with Dr. Vignoli from the Regional Epilepsy Center at ASST Santi Paolo e Carlo in Milan, I had the opportunity to analyze three datasets of epileptic patients at different disease and therapy stages.

The first dataset consisted of children admitted to the hospital after their first seizure; diagnosed with epilepsy, these children were sampled and followed up with 2 more sampling after 4 and 12 months subjected to drug monotherapy. The project was aimed at monitoring the gut microbiota differences from a naive bacterial state, influenced only by the seizure occurrence, to a gradually pharmacologically-treated one.

The second dataset was made up of adult epileptic patients with a stable years-long monotherapy that resulted in no more seizure episodes. The hypothesis is that gut microbiota in these patients should have reached a new balance after many years with a constant drug assumption.

The third dataset consisted of children affected by the Rett syndrome and seizure episodes, for which they assume the same drug prescribed to epileptic patients. They have been added to this thesis dataset as a “negative control”, opposed to the healthy subjects enrolled for a disease- and drug-free microbiota: given the additional condition to the seizures and

the different lifestyle and cognitive impairments that often affect the Rett patients, their microbiota is thought to be extremely diverse from the others.

This thesis deals an exploratory project that looks for connections and influences between gut bacteria and the epilepsy development through time and drug therapy. All the samples have been collected only after making sure that the patients were not in any particular diet, such as the KD, nor taking any antibiotic or probiotic in the six months preceding the sampling, in order to see a more realistic bacterial picture (i.e.: not altered by non-natural influences).

Taken together, this resulting dataset aimed at providing a broad comparison of: 1) patients with a drug-naive bacterial community (affected only by the epilepsy onset); 2) patients assessing their gut microbiota after introducing the drug therapy; 3) patients with a well-assessed bacterial balance after years of therapy; 4) patients dealing with an established drug therapy, reduced mobility and other symptoms. The addition of the adults in drug therapy to the timeline with several samples from the same subject could provide a glimpse of the progression of an eventual shift in the gut microbiota, allowing a long-term observation that often is not clinically possible or easy to accomplish.

This diverse datasets set the stage for speculations on the microbiota changes and development during epilepsy onset and a constant drug therapy in terms of bacterial diversity, composition, and interaction. In case the hypothesis of a gut microbiota impairment due to seizures and/or drug assumption has a solid base, after confirmation by further research with huger cohorts, this could pave the way to a probiotic or prebiotic therapy aimed at restoring a healthier bacterial community and, possibly, easing the seizure occurrence complementing or substituting the drug therapy.

MATERIALS AND METHODS

1. The dataset

The total dataset derived by the union of the datasets described in *Introduction, 4. Aim* (page 40) and consisted of 65 fecal samples.

The mentally and physically healthy subjects enrolled as positive gut microbiota control (healthy control, HC) were 28, a number comparable to that of patients. The age range of the HC group was from 7 to 28 years. For a clearer image of the native and non-altered bacterial community, all the HC subjects that were under any other medication or assumed antibiotic or probiotic treatments in the previous three months, have been discarded.

Samples from the children before and after therapy dataset (9.2 ± 4.6 years) were taken at their first onset of seizure (time zero, t0; labeled drug-naive DN), after 4 months of therapy (time one, t1; labeled drug-therapy 4 months DT4) and after 12 months from t0/DN (time two, t2; labeled drug-therapy 12 months DT12). The drugs assumed were either VPA or ETS. The samples collected at each time-point were 8 as DN, 8 at DT4, 6 at DT12. The healthy subjects enrolled for the time-line comparison between these time-points were a subset of 9 HC, age (7.3 ± 3.9) and sex matched.

The Rett syndrome subjects (RTT) enrolled as negative gut microbiota control were 7, with a mean age range of 23 ± 8.7 years. They all had different degrees of clinical severity, all carried the *MECP2* gene mutation and were all assuming one of the AEDs prescribed to the epileptic patients involved in this study. Three of them still had seizures, while four were seizure-free. Two patients were on monotherapy (VPA and CBZ, respectively), whereas five were on polytherapy (three including VPA, two

including CBZ). Gastrointestinal discomfort and constipation were present in all subjects, despite regular feeding ability. This dataset has been published in Borghi et al. [153]

The adult epileptic patients in a stable drug therapy (labeled SDT) were all taking the same VPA drug, for at least two years.

All participants were Caucasian, living in Northern Italy, and without any dietary restriction. The following table summarizes the dataset composition:

Group	N°	Age	Sex (F/M)	BMI	Drug
DN	8	8.5 ± 4.4	6 / 2	*NA*	-
DT4	8	8.8 ± 4.2	6 / 2	*NA*	VPA / ETS
DT12	6	10.3 ± 4.5	5 / 1	*NA*	VPA / ETS
SDT	8	30.3 ± 5.2	8 / 0	*NA*	VPA
RTT	7	23.0 ± 8.7	7 / 0	17.2 ± 3.9	VPA / CBZ
HC	28	21.0 ± 7.9	19 / 9	20.9 ± 2.2	-
	subset: 9	7.3 ± 3.9	6 / 3	*NA*	-

1.2 Subjects recruitment

All subjects were enrolled at the ASST Santi Paolo e Carlo Hospital (Milan, Italy), in the Epilepsy and Sleep Disturbances Center (adult and children after therapy patients) and in the Child Neuropsychiatry Department (children before therapy, Rett patients). The diagnosis of RTT was made according to the Rett diagnostic criteria defined in 2010. [154] Inclusion criteria were clinical diagnosis of RTT and demonstrated *MECP2* mutation, while exclusion criteria were the use of antibiotics or probiotics in the three months before. All these studies were approved by the Local Ethics Committee (protocol number 2016/ST/199, 28 July 2016). A written informed consent was obtained from healthy subjects and from the parents or legal guardians of the enrolled patients.

2. DNA extraction and sequencing

The following procedures were performed in the Microbiology laboratory, at the Department of Health Sciences of the University of Milan (Milan, Italy).

The fecal samples collected by the patients, whenever not immediately processed, have been stored at -20°C until use. From each fecal sample, an aliquot of 200mg was taken in order to perform the genetic material extraction.

Total bacterial DNA extraction was performed using the Spin stool DNA kit (Stratec Molecular, Berlin, Germany), according to the manufacturer's instructions. The extraction kit provides the homogenization of the fecal samples in a lysis buffer and its incubation at 95°C for 10 minutes. The lysate samples were mixed with the InviAdsorb matrix to remove PCR inhibition components. After incubating at 70°C for 10 minutes, Proteinase K was added to the supernatant, for protein digest and degradation. Bacterial DNA purification is obtained through the addition of suitable buffers, aimed at the elimination of eventual foreign particles. The filtrated product was eluted with the kit buffer to a volume of 100 µl.

The bacterial DNA concentration and degree of purity have been evaluated through spectrophotometer (NanoDrops, Spectrophotometer ND-1000). For each sample, 1 µl has been read at 240 nm wavelength, and the optic density value has been automatically converted in ng/µl for evaluating the extraction quality.

25 ng of DNA from each stool sample was utilized to construct a sequencing library. 16S rRNA gene amplicon libraries were performed with a two-step barcoding approach according to Illumina 16S Metagenomic Sequencing Library Preparation (Illumina, San Diego, CA, USA), which targets the

V3-V4 hypervariable regions of 16S rRNA. In the first-step PCR, 16S rRNA gene of all bacteria was amplified as follows.

PCR reactions were prepared in a final volume of 25 µl with the following reagents:

	Volume
Microbial Genomic DNA (5 ng/µl in 10 mM Tris pH 8.5)	2.5 µl
Amplicon PCR Reverse Primer (1 µM)	5 µl
Amplicon PCR Forward Primer (1 µM)	5 µl
2x KAPA HiFi HotStart ReadyMix	12.5 µl

The amplification conditions were: 95°C for 3 minutes; 25 cycles of: 95°C for 30 seconds, 55°C for 30 seconds, 72°C for 30 seconds; 72°C for 5 minutes; hold at 4°C. To verify the size of PCR amplicon, the PCR product were run on electrophoresis gel and ~550 bp were obtained for all the samples.

Before proceeding with the second PCR for indexing the samples, a clean-up step was performed to purify the samples. The purified amplicons resulting from first PCR step were amplified with dual-index primers using Nextera DNA Library Preparation Kit (Illumina). Each sample possessed specific barcode sequences at the 5′- and 3′-end of the PCR amplicon to discriminate among each other in the pooled library.

The second PCR was obtained in a final volume of 50 µl with the following reagents:

	Volume
DNA	5 µl
Nextera XT Index Primer 1 (N7xx)	5 µl
Nextera XT Index Primer 2 (S5xx)	5 µl
2x KAPA HiFi HotStart ReadyMix	25 µl
PCR Grade water (Dnase-Rnase free)	10 µl

The amplification conditions were: 95°C for 3 minutes; 8 cycles of: 95°C for 30 seconds, 55°C for 30 seconds, 72°C for 30 seconds; 72°C for 5 minutes; hold at 4°C.

After the second PCR and a subsequent clean-up, library concentration and exact product size were measured using a KAPA Library Quantification Kit (Kapa Biosystems, Woburn, MA, USA) and an Agilent 2100 Bioanalyzer System (Agilent, Santa Clara, CA, USA), respectively.

A pooled library (20 nM) and a PhiX control v3 (20 nM) (Illumina) were mixed with 0.2 N fresh NaOH and hybridization buffer HT1 (Illumina) to produce the final concentration at 12 pM each. The resulting library was mixed with the PhiX control v3 (5%, v/v) (Illumina) and 600 µL loaded on a MiSeq v2 (500 cycle) Reagent cartridge for obtaining a paired-end 2 × 250 bp sequencing. All sequencing procedures were monitored through the Illumina BaseSpace application. De-multiplexed FASTQ files were generated by Illumina MiSeq Reporter.

3. Data analysis

Raw sequencing reads from the datasets have been pooled and processed together.

The read pairs have been merged by using PandaSeq software (version 2.5, [155]), which performs a local assembly between two overlapping pairs, generating a single fragment covering the whole V3–V4 amplicon. Fragments of length <250 bases or >900 bases, as well as non-overlapping sequences, were discarded. Then, fragments were filtered using the "split_libraries_fastq.py" utility of the QIIME suite, which filters out sequences having more than 25% nucleotides with a Phred score of 3 or less.

Quality-filtered reads were, then, analyzed with the standard QIIME pipeline (version 1.9.0). Sequences were grouped into OTUs (Operational Taxonomic Units) by using UCLUST [156] with 97% similarity threshold and taxonomy was assigned against the SILVA bacterial 16S rRNA database (release 132) by RDP classifier at 50% confidence. Singleton OTUs (i.e., clusters made up of only 1 read) were discarded as possible artifacts or unlikely bacterial sequences.

Sample biodiversity (alpha-diversity) was estimated according to different microbial diversity metrics: Chao1, Shannon index, observed species and Faith's phylogenetic distance (PD whole tree). Statistical comparisons between groups have been conducted with the QIIME pipeline, through the Monte Carlo non-parametric permutation test provided by the "compare_alpha_diversity.py" function.

Beta-diversity matrices were calculated using both weighted and unweighted Unifrac metrics [147] and a Principal Coordinates Analysis were

conducted. Data separation was tested through the PERMANOVA test (adonis function) in the R package vegan (version 2.0-10 [157]).

The RStudio software (version 1.2.1335 [158]; R version 3.6.3 [159]) has been used for: the statistical evaluations in taxonomy analysis, using the pairwise t-test from the package “rstatix” (version 0.6.0 [160]); the feature selection analysis, via the generalized linear model from the “stats” R package (version 3.6.3); plotting all these analysis results using *ggplot2* from the package “tidyverse” (version 1.3.0 [161]). The co-abundance correlation heatmap and clustering were performed in Matlab (version 2008; Natick, MA, USA). The networks of co-abundant groups were plotted through the Cytoscape software [162] using the Matlab input for nodes and edges.

A p-value <0.05 was chosen as the threshold for statistical significance. Unless otherwise stated, a Bonferroni correction for multiple comparisons was applied.

RESULTS

1. General observations

Overall, the patients had a different biodiversity and composition from the healthy controls, as reported in the alpha-diversity and in the beta-diversity shown in *Figure 5A* and *5B*, respectively. For each alpha metric and for the unweighted Unifrac distance, the comparison was found significant: Chao1, p=0.031; observed species, p=0.016; Shannon index, p=0.026; PD whole tree, p=0.024; unweighted Unifrac distance, p=0.03.

This analysis resulted in a separation at phylum level of the Tenericutes (p=0.0498), reduced in the patients, while at family and genus level there were significant distinctions only for bacterial with relative abundances <1%. (*Supplementary Table 1* and *Supplementary Figure 1*, pages 89 and 94).

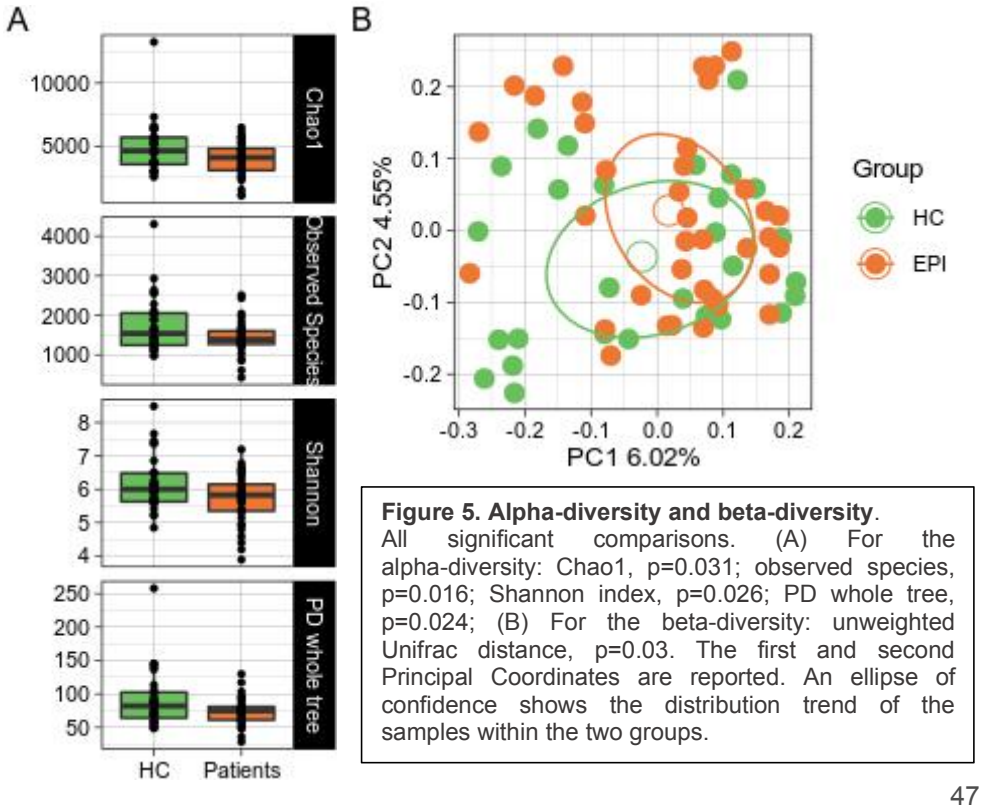


Figure 5. Alpha-diversity and beta-diversity.
All significant comparisons. (A) For the alpha-diversity: Chao1, p=0.031; observed species, p=0.016; Shannon index, p=0.026; PD whole tree, p=0.024; (B) For the beta-diversity: unweighted Unifrac distance, p=0.03. The first and second Principal Coordinates are reported. An ellipse of confidence shows the distribution trend of the samples within the two groups.

2. Seizure naive, drug stabilized and over-modified GM

The previous results have been deepened by narrowing the dataset, establishing several categories for the analysis. A group called “EPI” was created to comprise the 22 samples assuming AEDs (i.e.: the adults in stable therapy, plus DT4 and DT12 subjects). This group was meant to represent the drug-assuming patients, in comparison with the sample with a drug-naive GM (the children at time zero, drug-naive, DN; 8 samples) and with the control groups, HC (28 samples) and RTT (7).

Given the peculiar clinical picture of the RTT, body mass index (BMI) was calculated using the formula *weight (kg)/height (m)*^{2}, and and compared to the HC. BMI was 17.2 ± 3.9 (mean ± SD) in RTT patients, and 20.9 ± 2.2 in HC ($p = 0.073$).

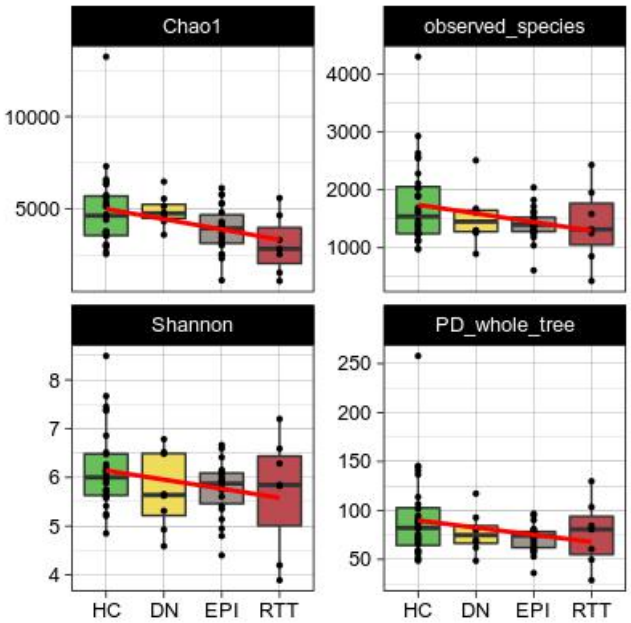


Figure 6. Alpha-diversity along the patient groups: healthy controls (HC), drug-naive epileptic patients (DN), patients in drug therapy for more than 1 year (DT4, DT12, SDT), and Rett patients (RTT). The linear regression computed shows the conditional mean trend along the pathological condition. No significant values were observed.

Biodiversity was found altered along the pathological conditions (*Figure 6*). The smoothed conditional mean trend elaborated through a linear regression shows that there is a clear trend towards reduced biodiversity with disease progression for every alpha-diversity metric, even though with no statistical significance. The Chao1 metric, an abundance-based coverage estimator of bacterial richness, reported the major different biodiversity within the samples. The drug-naive GM was found altered compared to HC, while the EPI group was found to be more similar to the bacterial diversity present in the RTT patients. The presence of outliers and data point dispersion was observed for all the groups.

Differences among these groups were seen also through the PCoA plot showing the beta-diversity, the inter-sample biodiversity (*Figure 7*). An interesting observation is that the EPI group and the drug-naive children seem to be overlapping, as the drug assumption didn't influence significantly the GM composition; the extremely high p-values point in that direction

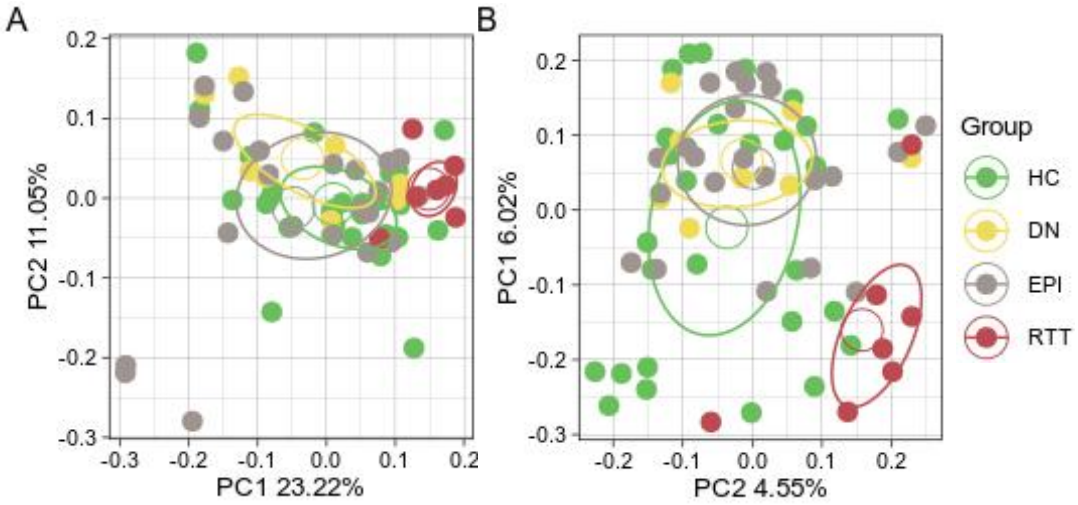


Figure 7. Beta-diversity for the experimental groups. Healthy controls (HC), drug-naive epileptic patients (DN), patients in drug therapy for more than 1 year (DT4, DT12, SDT), and Rett patients (RTT) are compared. Weighted (A) and unweighted (B) Unifrac Distances are represented, using the first and second principal coordinates. An ellipse of confidence shows the distribution trend of the samples within the several groups. The first and second Principal Coordinates are reported for both plots.

(p=0.82 and p=0.95 for weighted and unweighted Unifrac distances). The RTT patients are clearly separated from the other groups, as those samples have their own composition; the comparisons between RTT and all the other groups had <0.01.

The taxonomic comparison revealed several significant differences in the bacterial abundance among the four groups.

At phylum level (*Figure 8*) a profound reduction of the Firmicutes phyla in the Rett patients was observed, in particular compared to the HC group (rel. ab: 54.7% and 35.6% in HC and RTT, respectively; p=0.044). The Bacteroidetes phyla showed to be depleted in the drug-naive GM (29% compared to 38.2% in HC), while RTT group had an opposite trend with a

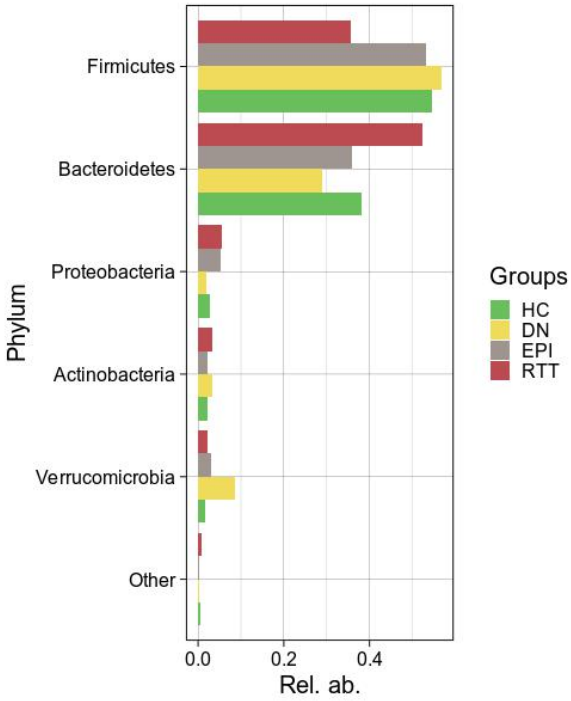


Figure 8. Relative abundance distribution of the main phyla among the experimental groups. Only the phyla with more than 1% of relative abundance in all the groups are represented; those with less than the set threshold were clustered into the "Other" group. Bacterial phyla are ordered after HC abundances.

higher abundance (52.4%, RTT vs DN: p=0.047); the group of epileptic patients in drug therapy, instead, showed a Bacteroidetes abundance similar to the HC one (35.9%). Verrucomicrobia, instead, was found to be characteristic of the microbiota of DN patients, accounting for 8.7% of the relative abundance, compared to 1.7% in HC subjects (p=0.0205).

The EPI and RTT group have a Verrucomicrobia abundance slightly higher but similar to HC (3.1% and 2.2%, respectively). Total relative abundances are reported in *Supplementary Table 2*, page 90.

As for the family level, *Ruminococcaceae* was found profoundly decreased in the RTT group: only 8.8%, compared to HC (28.2%, p=0.0065), drug-naive (30.9%, p=0.0134) and EPI patients (30.0%, p=0.0324). (Figure

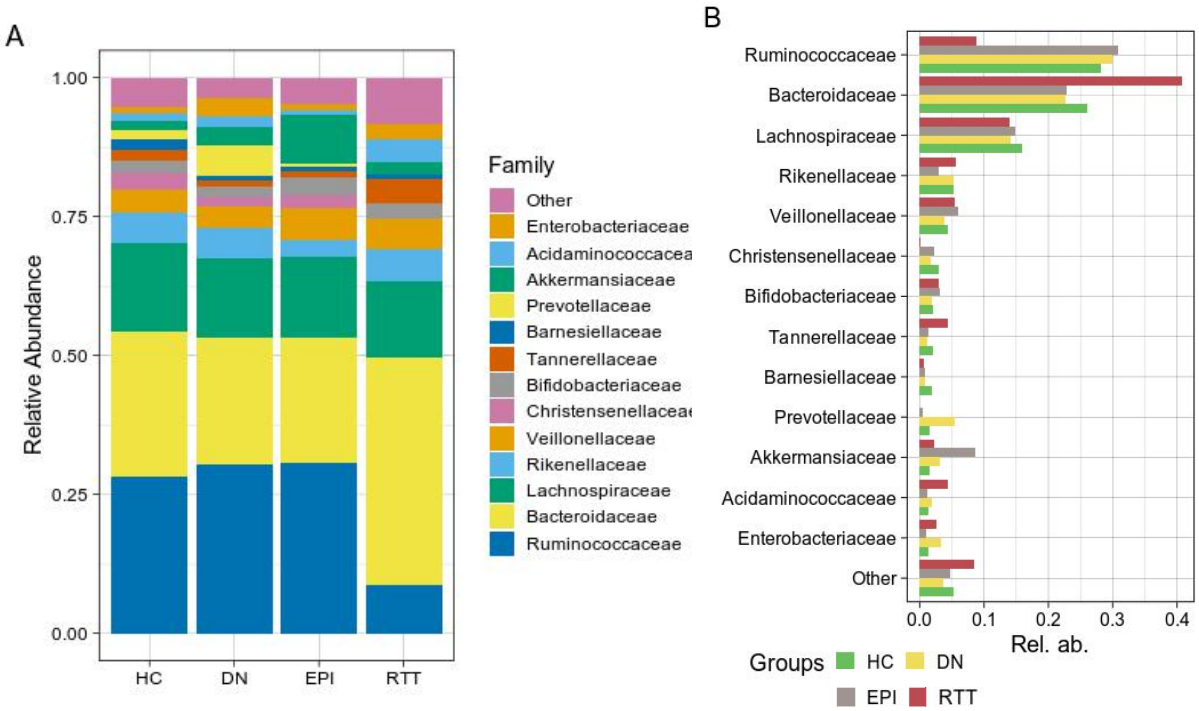


Figure 9. Relative abundance histograms at family level through a stacked (A) and a bar plot (B). Only the phyla with more than 1% of relative abundance in all the groups are represented; those with less than the set threshold were clustered into the "Other" group. Bacterial families are ordered after HC abundances.

Among the main genera, i.e. with a relative abundance higher than 0.1%, some bacterial groups were found to be significantly different among the patients. As seen for the *Bacteroidaceae* family, *Bacteroides* genus was

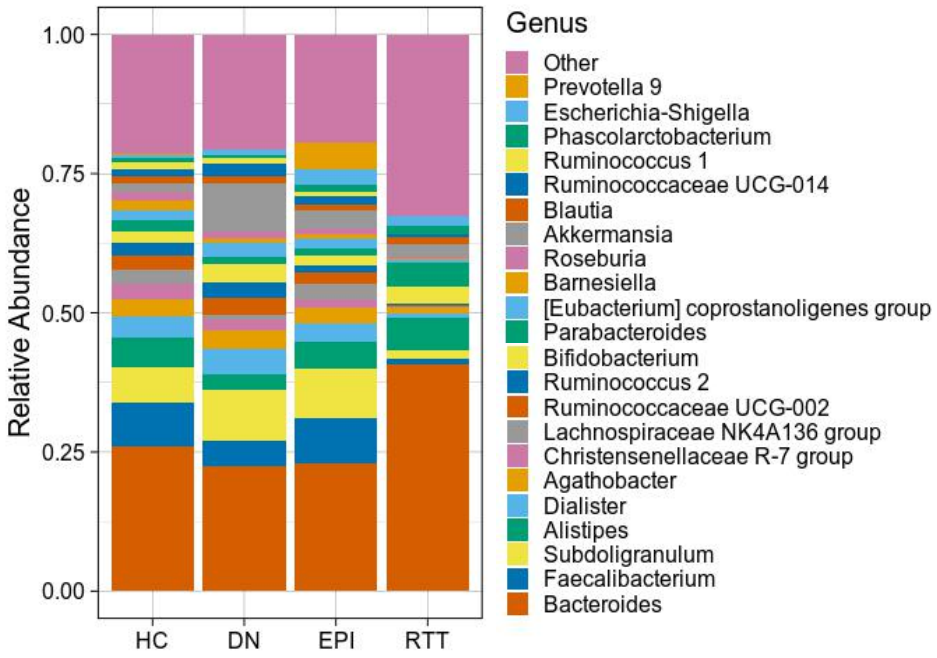


Figure 10. Relative abundance at genus level. The main 20 genera, with a relative abundance higher than 1% among all the groups, are represented; those with less than the set threshold were clustered into the "Other" group. Bacterial families are ordered after HC abundances.

found in a much higher quantity in RTT patients than in all the other groups (40.7% vs 22.6% in EPI, p=0.0134); *Parabacteroides* followed a similar trend, 4.4% in RTT patients compared to 1.1% in the EPI group (p=0.0232) and to 2.0% in HC and 1.3% in the drug-naive children. *Akkermansia* genus confirmed what was observed at higher phylogenetic levels: a clear and significant increment compared to all the other groups (8.7% vs 1.6% in HC, p=0.0201). (*Figure 10*)

A co-abundance analysis has been conducted with the purpose to see if and how the bacterial genera were found abundant or absent together. (Figure 11) The main 25 genera, having an average relative abundance higher than

1% in at least 30% of the samples within each group, have been included. The red tiles indicate a positive relationship, while the blue ones a negative. Along the diagonal line, the clusterization, according to Spearman's correlation and Ward linkage metric, allowed the distinction of four co-abundant groups (CAG). Each of these has been named after the most abundant genus: Allistipes CAG(green), Subdoligranulum CAG (pink), Faecalibacterium CAG (yellow), Bacteroides CAG (red).

A co-abundance matrix and corresponding symmetrical heatmap have been generated also for the four groups of patients, separately: HC, drug-naive, EPI, and RTT. (*Supplementary Figure 2*, page 95) The bacterial abundances of each genus present in the CAGs have been added up to generate the relative abundance of the CAG themselves and compared to see how they were different among the experimental groups. (*Supplementary Table 4* and *Supplementary Figure 3*, page 92 and 96)

A network of co-abundance has been made for each patient group, in order to observe the relationships of the several members of the CAGs, combined with the different relative abundances (the size of the nodes) and the intensity of the relationship (the transparency of the arches). The width of the arches indicates the intensity of the correlation, while the color indicates whether this correlation was negative (blue) or positive (red).

Despite the not-so-different taxonomy abundances (represented by the size of the nodes), the DN network (*Figure 12B*) seemed to have stronger negative correlations between the bacterial genera within the CAGs compared to the HC one (*Figure 12A*), in particular between the Faecalibacterium and the Subdoligranulum CAGs. Many Subdoligranulum CAG members were found negatively correlated to the Bacteroides CAG ones as well, while the group was overall more positively correlated to the

Allistipes CAG. The networks for EPI and RTT groups are shown in *Supplementary Figure 4A* and *4B* (page 97), respectively. While the EPI correlations were weaker (R value closer to 0 than to 1) but quite comparable with the DN ones, the RTT network shows a completely different pattern, with a majority of positive and strong correlations among all the CAGs, even if with a strongly reduced relative abundance of many of the genera involved. A remarkable role was played by *Subdoligranulum* and the genera in its CAG. This genus was the only one within the CAG to maintain a relatively high abundance in the Rett group, while all the others were dramatically reduced.

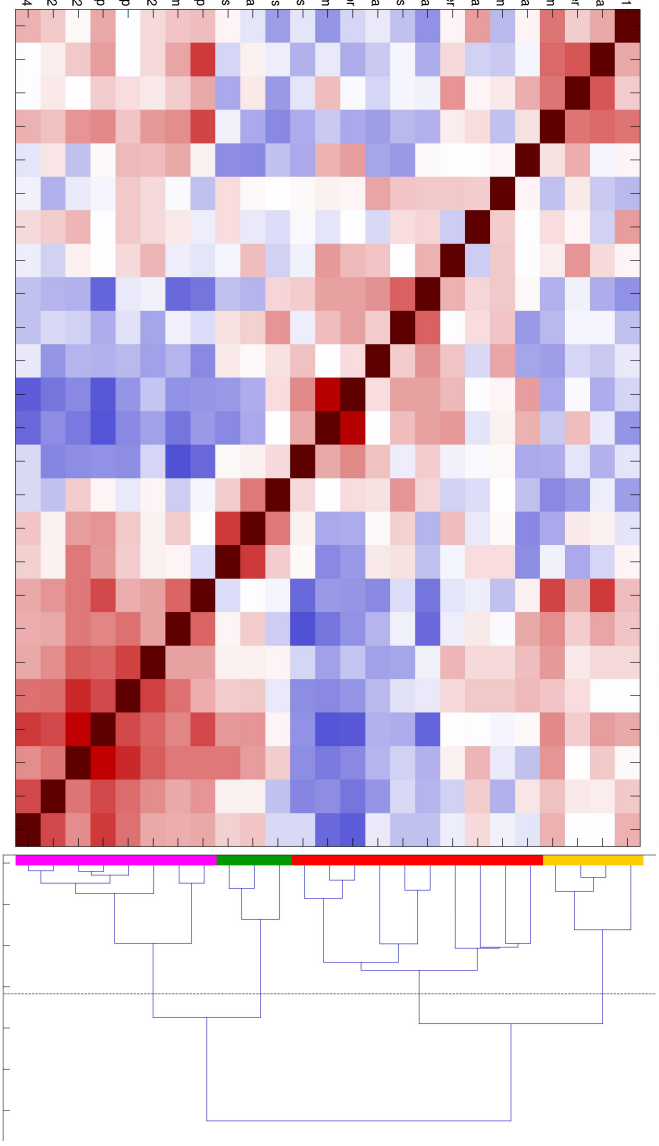


Figure 11. Heatmap of the co-abundances occurring in the total dataset. Red tiles indicate a positive relationship, blue ones a negative. According to the Spearman correlation, 4 clusters were identified as summarizing the genera. X- and Y-axis are the same, but mirrored as it is a symmetrical heatmap. The partition of this heatmap in HC, DN, EPI, and RTT groups is shown in *Supplementary Figure 2* (page 95).

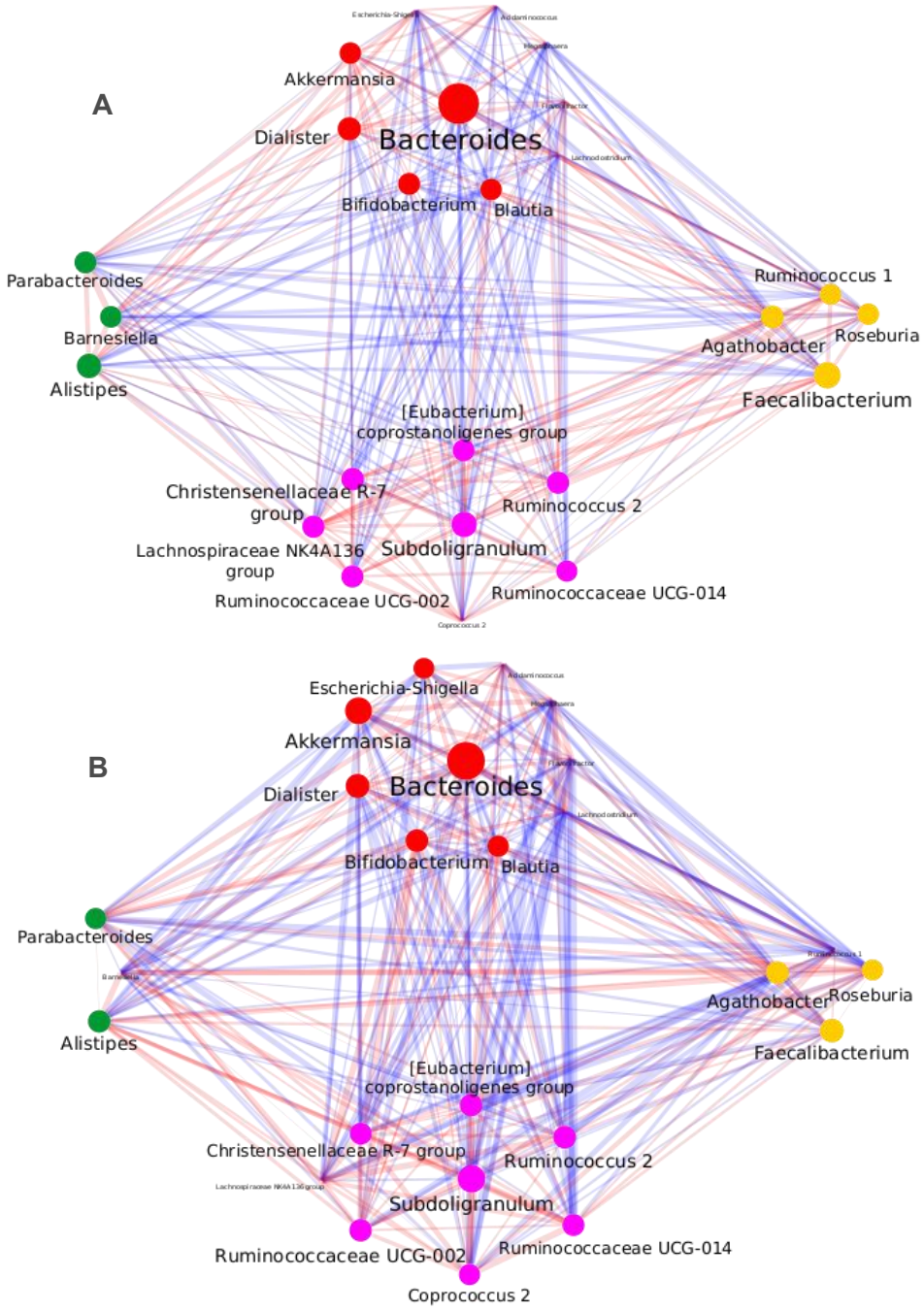


Figure 12. Correlation among bacterial genera. The networks for HC (A) and DN (B) are shown. Relative abundances of the singular genera are represented by the size of the nodes and of the labels. The width of the arches indicates a higher or lower correlation value according to the Spearman test. The arches colors indicate a positive (red) or negative (blue) interaction. Bacterial genera are divided into the CAG deduced by the co-abundance heatmap in *Figure 10*. EPI and RTT networks are shown in *Supplementary Figure 4A* and *4B* (page 97).

3. Timeline

As 22 samples were from a longitudinal study throughout one year, with three sampling aimed at looking for the impact of the drug therapy on the gut microbiota, I focused the analysis by dividing the EPI groups into the time categories: of the 8 drug-naive samples collected at time-zero (DN) from patients, all of them provided a sampling after 4 months of drug assumption (DT4), and 6 of them provided a subsequent follow-up after 12 months (DT12). The HC samples used for the comparisons were 9 subjects age- and sex-matched with the patients, in order to reduce a possible confounding factor in this younger subset.

As a preliminary glimpse, the children's samples show a firm signature across the months and the different sampling, as shown in the PCoA in *Figure 13*. Through time, many samples showed an overall shifting trend along the first principal coordinate (PC1): 5/8 patients (62.5%) have their DN time-point at lower PC1 values, while 4/6 patients (66.7%) have the DT12 samples at the higher values.

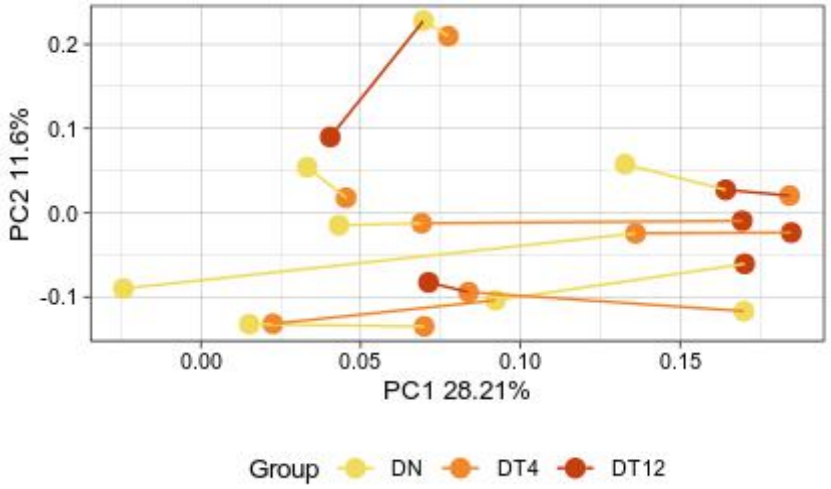


Figure 13. Beta-diversity of the samples grouped by patient of origin. The lines connect the samples from the single patient, while the colors indicate the time-point of sampling collection. The first and second Principal Coordinates are reported within the unweighted Unifrac metric.

This trend could be observed also in the alpha-diversity analysis, which revealed how the children's biodiversity was reduced gradually and constantly over the months of pharmacological treatment. (*Figure 14*)

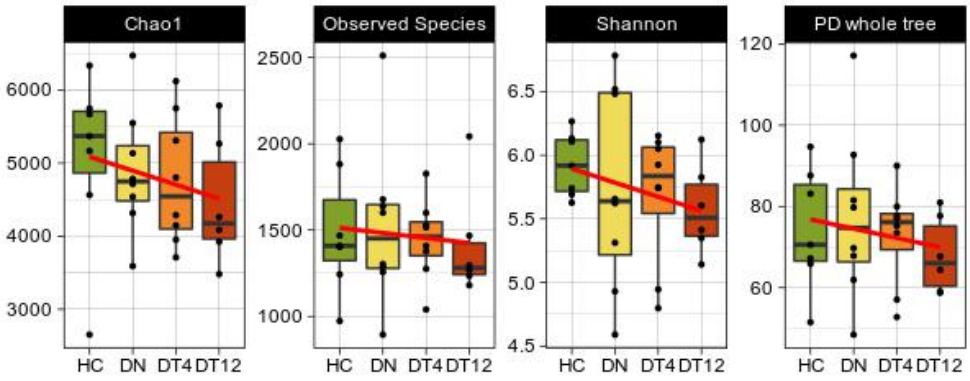


Figure 14. Alpha-diversity along the time-points. Healthy subjects (HC) were used as controls for epileptic children with a drug-naive gut microbiota (DN) sampled after the first seizure, and the same patient sampled after 4 (DT4) and 12 (DT12) months of drug therapy. The linear regression shows the conditional mean trend decreasing along the months in therapy. No significant values were observed.

Taxonomic abundances switches along time can reveal what happens when epilepsy and a AED therapy starts. (Figure 15) Relative abundances at phylum and family level are reported in *Supplementary Table 5*.

At genus level, it is worth noting the sudden peak of *Akkermansia* in patients at the onset of the seizure disorder (8.7% vs 2.1% in HC, p=0.0462). After 4 months of drug assumption, this relative abundance drops rapidly (0.5% in DT4) and returns to a normal amount after 12 months (3.2% in DT12), comparable to the HC abundance. On the other hand, *Faecalibacterium* was found decreasing tragically at DN (4.7% *vs* 11% in HC), going back to a normal level after the drug therapy (8.6% at DT4, 11.1% at DT12). '*Prevotella 9*' has entirely the opposite trend, as it is mainly absent before the therapy (1.0% in DN patients) and then, after 4 months of drug assumption, increasing rapidly (4.7% at DT4, 5.9% at DT12) to the healthy amount (4% in HC).

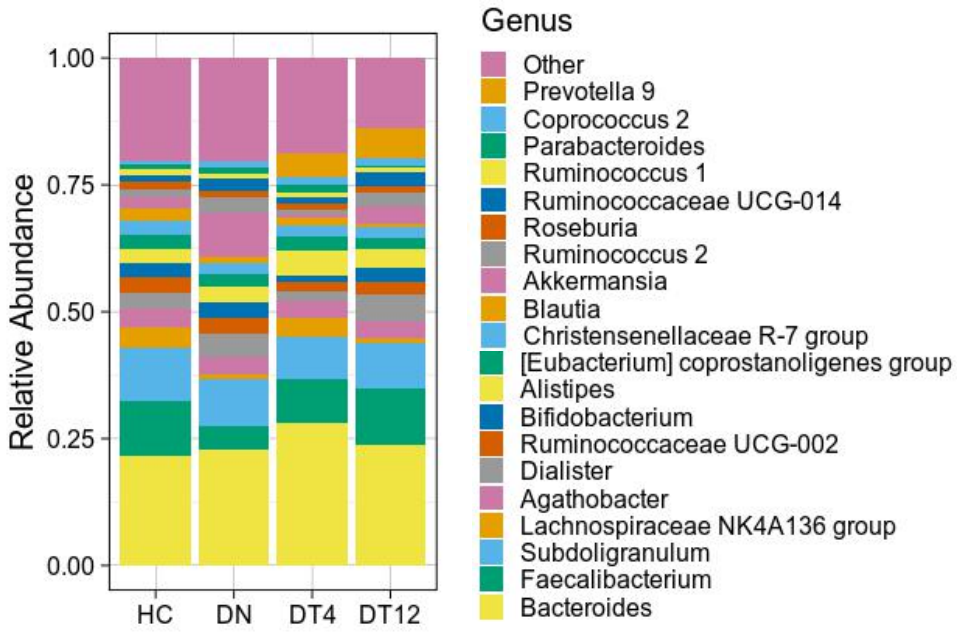


Figure 15. **Relative abundance at genus level.** The main 20 genera, with a relative abundance higher than 1% among all the groups, are represented; those with less than the set threshold were clustered into the "Other" group. Bacterial families are ordered after HC abundances.

3.1 A speculation tool

In order to speculate on how these changes could stabilize or still mutate in the future years of drug therapy, the samples from the adults continuously assuming an analogous pharmacological treatment have been added up, as an artificial new time-point. The linear regression applied on the alpha-diversity shows a decreasing pattern over time and therapy assumption. It is worth noting that the adult group and the RTT group have an overall highly similar biodiversity (with a p-value=1 for all the metrics). (*Figure 16A*)

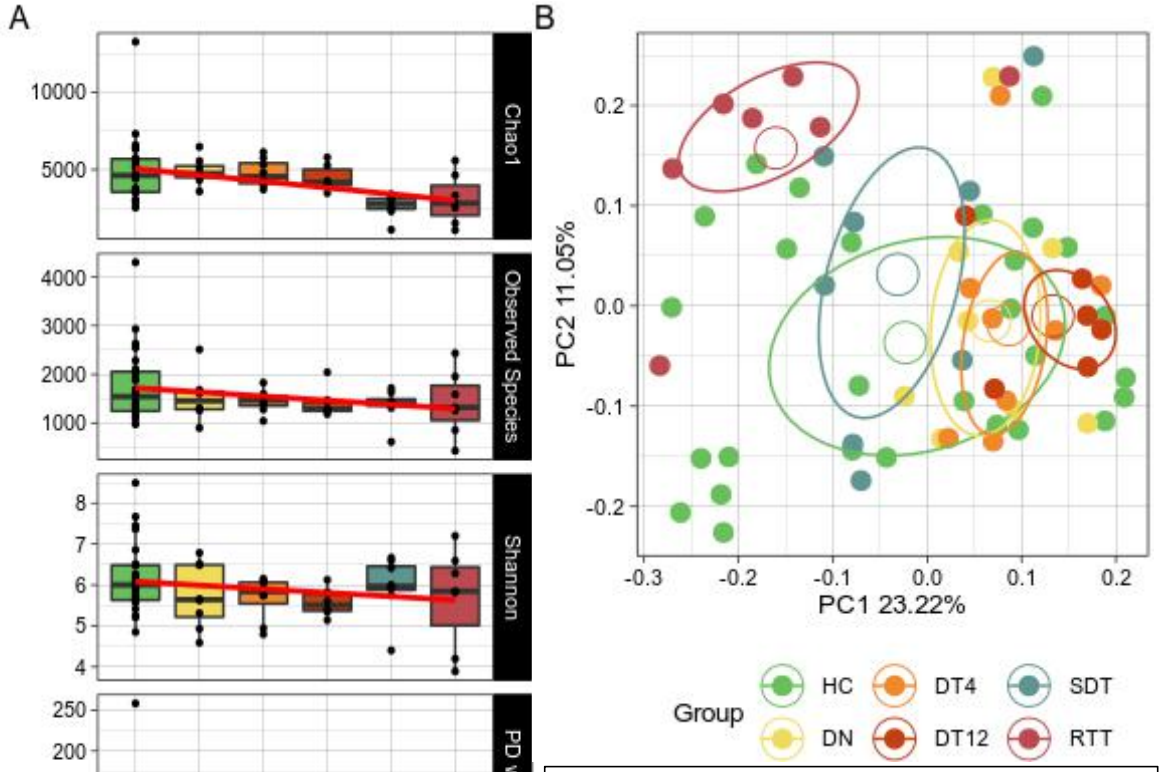


Figure 16. Biodiversity distribution over the time and the therapy assumed. (A) Alpha-diversity: the linear regression computed (the red line) shows that the biodiversity decreases along with the time of drug assumption increment. (B) Beta-diversity: the distribution trend of the samples within the several groups is shown using the first and second Principal Coordinates; an ellipse of confidence shows the distribution trend of the samples within the groups.

In the PCoA it is evidenced how the kids cluster together, while adults in years of therapy have accomplished some similarity to the HC in terms of overall diversity since their distributions and mean centroids are close one another. (*Figure 16B*)

Since not all bacteria react the same way to drugs due to their different cell wall, the gram coloration could show some interesting difference among groups. Gram-positive decrease gradually as Gram-negative abundances rise in time and therapy assumption. (*Figure 17*)

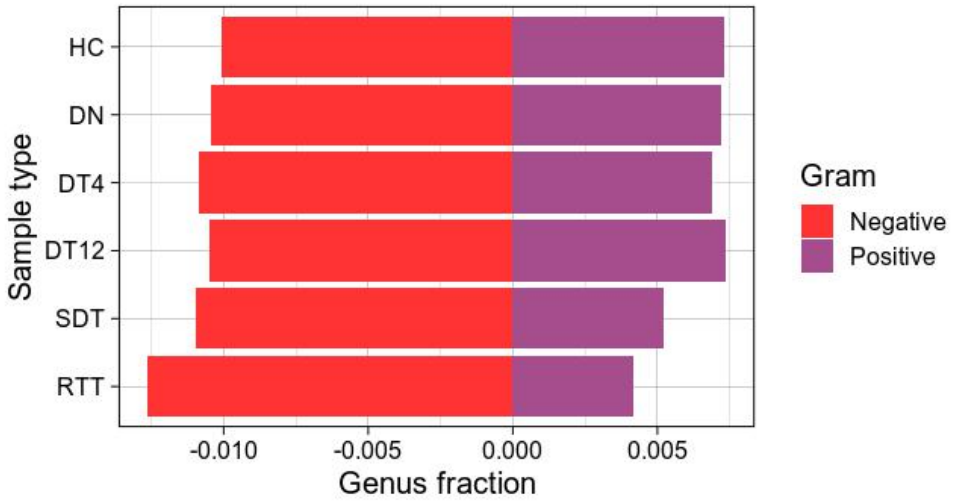


Figure 17. Gram coloration separation among bacterial genera. The first 100 genera have been included, corresponding to the 99.99% of the cumulative abundance of the total genera.

In the taxonomy abundances distribution (*Supplementary Figure 5*, page 98), '*Prevotella 9*', that was already found interesting among the real time-line with the children dataset, with this amplified time-line it seems to be linked to the drug assumption: as it was mainly absent in HC an DN samples, it increases over the first months and year of therapy, keeping a stable amount in the adult patients; the RTT group, instead, appears to be not affected by it. We also see a higher abundance in *Escherichia-Shigella* among the adult group, and for a little amount also in RTT.

4. Pharmacological treatment effects

As a focus only on the drug assumption in epileptic patients, the subjects have been selected and divided into those who were not currently taking any drug (the drug-naive group, DN) and those who have been taking the same drug for more than one year (DT12 and adult group), labeled DA ("drug-assumption"). In order to look only for drug-stabilized GM compositions, DT4 has been excluded from the analysis, as 4 months could be not enough time for microbiota to be changed in a stable way.

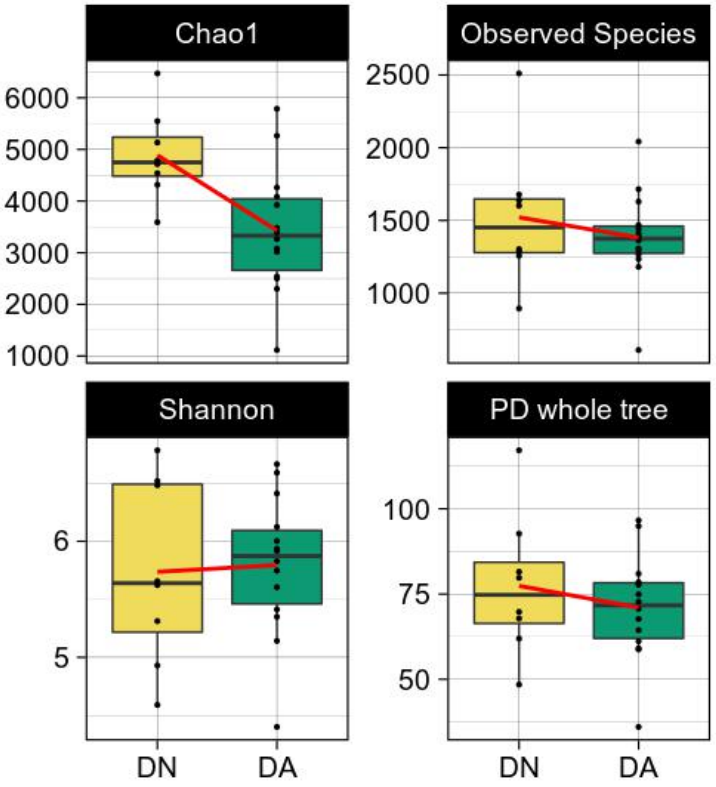


Figure 18. Alpha-diversity distribution for epileptic patients according to the drug assumption. Patients have been divided in non assuming drugs (the drug-naive group, DN) and in pharmacological treatment for more than one year (DA). Chao1 metric was found significantly reduced over the groups (p=0.0072). The linear regression computed (the red line) shows that the biodiversity was reduced also with the Observed Species and PD whole tree metrics, while bacterial richness (Shannon index) was increased.

Only the Chao1 metric had a significant trend over the drug assumption (p=0.00712, *Figure 18*), while the other metrics did not. Biodiversity, furthermore, has different trends with regards to the alpha-diversity metric used: Chao1, Observed Species and PD whole tree had a decreasing values in the DA group, while Shannon index had an increment.

A feature selection analysis to see which of the families could be more imputed to highlight the dysbiosis suggested by the alpha-diversity distribution was performed. Using the values from each alpha-diversity metric as a co-variate to discriminate the bacterial groups, the most abundant 30 out of 143 families were tested. Since the feature selection analysis doesn't perform well with too many bacterial absences (i.e. relative abundances at 0%), only the bacterial groups present for at least 0.001% of relative abundance in all the samples have been added for selection.

For Chao1 metric, the family relevant to the model was *Akkermansiaceae*; for Observed Species metric, *Akkermansiaceae* and *Lachnospiraceae*; for Shannon index, *Tannerellaceae*; for PD whole tree, *Christensenellaceae*.

These suggestions are in line with the different relative abundances found among the experimental groups (*Lachnospiraceae* and *Christensenellaceae* are part of the Subdoligranulum CAG seen in *Figure 12* (page 57), while *Akkermansiaceae*, from the Bacteroides CAG, was already noticed in Figure 14) and comply with the alpha-diversity drug-assumption findings.

As reported in *Figure 19A*, these bacteria were all decreased during the drug assumption. With a distinction among the groups (*Figure 19B*), though, this constant decreasing trend was observed only in the *Christensenellaceae* family, while the others did not show clear tendencies.

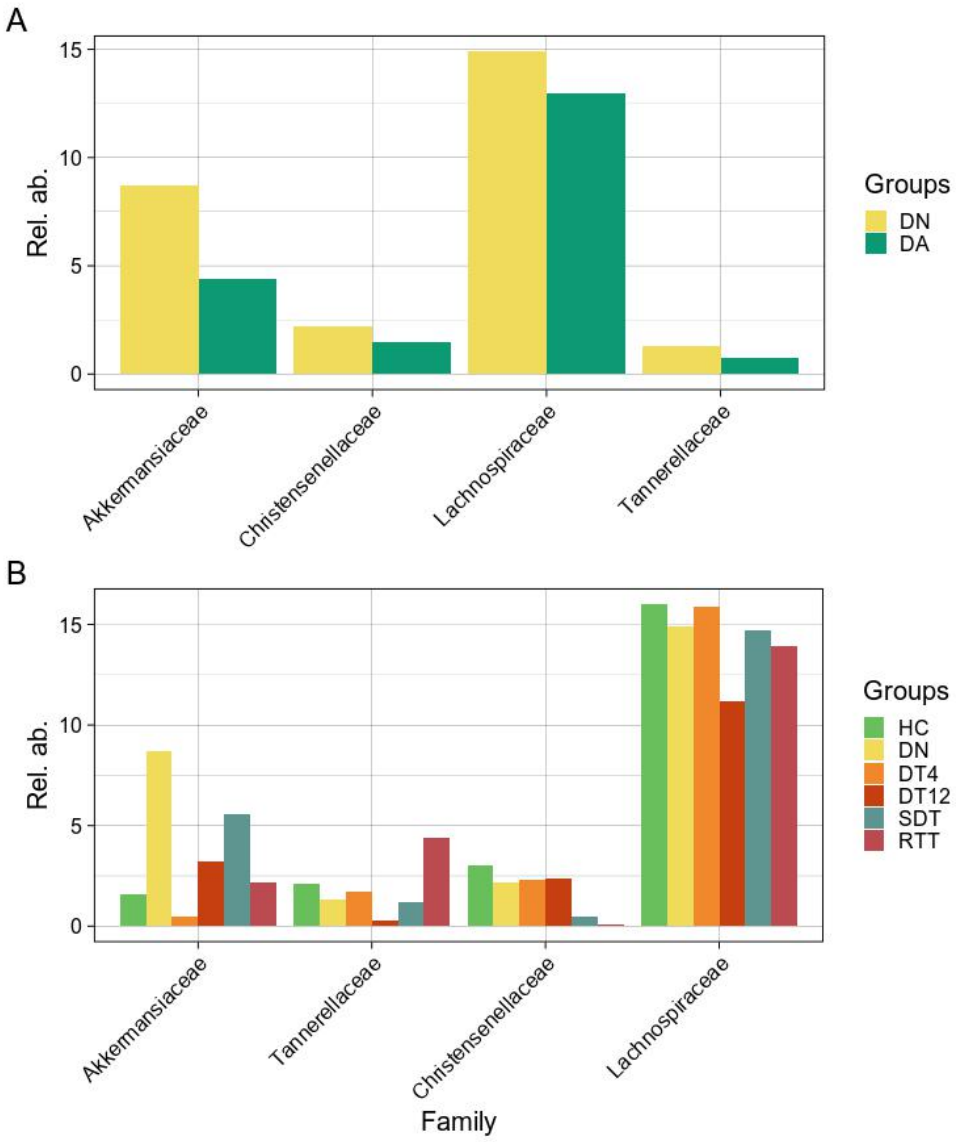


Figure 19. Relative abundances of selected bacterial families. (A) Divided by drug-assumption, as in the feature selection analysis; (B) for all the experimental groups in the dataset.

DISCUSSION

This exploratory thesis revealed some interesting aspects that could be used for speculations on the gut microbiota changes in patients affected by epilepsy seizures and in a consequential pharmacological treatment.

Drug-naive as an only-epileptic gut microbiota

Drug-naive (DN) patients have been studied in order to have an idea of a gut microbiota composition altered only by epilepsy, with in a picture cleared of potential confounders and focused on the diseased and unaltered by drug treatments.

DN subjects were found to have some interesting and peculiar trends: for example, a conspicuous increase in the *Akkermansia* genus compared to all the other groups (from 2.1% in the healthy controls (HC), to 8.7% in DN), further confirmed by highlighting the *Akkermansiaceae* family through the feature selection analysis. *Akkermansia* is known to have a protective effect on microbiota, as a mucin degrader, and is normally considered a good bacteria with a beneficial role: the high abundance could be caused by a protecting role of this bacteria after the occurrence of the first seizure of the children. In the co-abundant network, *Akkermansia* was grouped together with the Bacteroides co-abundant group (CAG), which had an increasing abundance with the progression of the disease, with a higher overall abundance in the non-healthy subjects.

On the other hand, *Faecalibacterium*, was found decreased in DN (4.7% *vs* 7.9% in HC): the seizure seems to be connected to the loss of a key beneficial player in the health of the gut microbial community.

Microbiota development in children

When the pharmacological therapy starts, the gut microbiota could be altered by the constant drug assumption. [169] As seen in the separation between Gram-positive and negative genera, the proportion of bacteria changes gradually, from a certain distribution in the healthy status to a higher proportion of Gram-negative along with the drug assumption. Since all the subjects were drug-respondents and did not experience further seizures following the initial recruitment, the assumption that the observed alterations are due to the drug therapy and not to the disease development could be valid.

The relative abundance of Proteobacteria rises from a healthy status quantity (0.5% in HC) over time (1.9% in DN patients, 2.9% at DT4, 1.3% at DT12), and was partially reflected in the higher abundance of the *Escherichia-Shigella* genus: 0.6% for HC rose at 1.1% in DN, stabilized at 0.8% at DT4 and 0.5% at DT12.

The high increase in the *Akkermansia* was observed to be stabilized over the months of drug therapy, with lower relative abundances, more similar to the HC ones (from 8.7% in DN to 0.5% at DT4 and 3.2% at DT12, compared to 2.1% of HC). The pharmacological treatment seems to have had a balancing role in the expression of this marker of good health.

A positive trend was also observed for the *Faecalibacterium* genus, which was greatly depleted after the first seizures (4.7% in DN *vs* 7.9% in HC) but, however, recovered over the first year of drug assumption (8.6% at DT4 and 11.1% at DT12). This genus is the most common indicator of microbial health in the gut, and its recover through drug therapy after the untreated seizures could be worth noting as a positive outcome.

On the other hand, there was a slight decrease over time in the *Lachnospiraceae* family (14.9% in DN patients becoming 11.2% at DT12, compared to 16.0% in HC): this family hydrolyzes starch and other sugars to produce butyrate and SCFAs. [164] From the findings of this study, it appears evident that the drug therapy affected but not dramatically this butyrate-producer, a useful and needed metabolite linked to gut microbiome health.

These alterations could be the reason why the analysis showed such a decreased biodiversity within the children samples, although with a similar bacterial composition, as evidenced by the beta-diversity analysis, compared to the other experimental groups. Biodiversity reduction has been generally related to a pro-inflammatory status within the gut. Nonetheless, Figure 13 (page 57) highlighted how this dissimilarity followed an overall common trend among the children studied: along the first principal coordinate of the unweighted Unifrac metric, the samples were separated between the DN and the DT12 sampling, indicating a bacterial diversity influenced by the drug assumption in the same individual over time.

This separation could be highlighted also by the '*Prevotella 9*' abundance, relatively absent in HC and DN subjects (<0.4% in both groups) but highly incremented after drug therapy (4.7% at DT4 and 5.9% at DT12). This group is named after its designation in SILVA database and defines a cluster mainly made up of *Prevotella copri*. Intriguingly, the main function of this species is to promote inflammation and to alter gut permeability, as reported by a recent study. [165] *Prevotella copri* is also thought to maintain the production of IL-6 in the gut, which could trigger an inflammatory response in the host. [167] A higher sequencing resolution could help unraveling the species characterization to address with more confidence this finding, but it is, nonetheless, an interesting trend.

SDT addition, a future insight

The adults in a stable drug therapy served as a parameter of future gut microbiota development following years of pharmacological treatment. Given the difficulties in keeping track of a single patient over such a huge time window, these findings could be useful to have an overall idea of possible future alterations; however, they must be taken into consideration with care and represent only a speculation on the basis of the observed evidences. This addition has a main bias, which serves as a dataset limitation: the age range diversity between the children and the adults could be an issue in stating how much of the differences found in the study could be influenced by the age, and not only by the microbiota behaviour with respect to disease and drug assumption.

This set of patients, however, could be seen as a bridge between the GM in epileptic patients at the early therapy stages and at a worse condition, with the Rett syndrome's different symptoms. While the Chao1 metric captured a constant decreasing biodiversity along the artificial timeline, SDT patients showed an interesting bacterial similarity with the RTT group in all of the other alpha-diversity measures used. Considering the extremely different clinical picture of these types of patients, their gut microbiota composition was overall similar and comparable. This could indicate that an elongated and constant assumption of a AEDs assumption shaped the gut microbiota composition reducing the differences among groups.

An imbalance toward the higher abundance of Gram-negative bacteria with a corresponding reduction of Gram-positive was observed. This is most probably due to the increased relative abundance of Proteobacteria, in SDT and RTT groups. Gram-negative bacteria, such as *Escherichia-Shigella* (dramatically increased in the SDT group, 6.0%, compared to the 1% and below relative abundances in the children dataset and in HC), are thought to

be carriers of pro-inflammatory molecules thanks to their lypopolisaccharide outer layer. In fact, these bacteria can be protected from the elimination and ingestion of white blood cell by the additional layer, allowing them to keep producing endotoxins that alter the gut microbiota composition and set the stage for a dysbiosis.

RTT trend in a different picture

The Rett syndrome dataset was the most peculiar in this present study, and showed the most profound alterations with respect to both HC and the epileptic-only patients. While the comparison with a healthy control was already described in Borghi et al., [153] the goal of the RTT addition to the dataset was to enhance similarities and differences of patients undergoing an analogous pharmacological treatment. Alpha- and beta-diversity evaluation only reported a firm signature of the RTT group, and taxonomy analysis could suggest something in that direction.

Ruminococcus, on the other hand, appears to be a hallmark of the HC group, decreasing dramatically since the first seizure in DN patients. The database used for classification (i.e.: SILVA) has a fragmented classification for the *Ruminococcus* genus (with several groups as *Ruminococcaceae UCG-002, UCG-005* and *UCG-014*, *Ruminococcus 1, Ruminococcus 2*,…) but this observation is sustained at a higher level by the *Ruminococcaceae* family, which has a 28.2% relative abundance in HC and only 3.1% in DN patients, 3% in the EPI group and 0.8% in RTT. This family is often found reduced in diseased patients, for example in people affected by inflammatory bowel disease [163] or Parkinson's disease. [96]

The *Christensenellaceae* family is emerging as an important player in human health, being related to a healthy status in a number of other different disease contexts, including obesity and inflammatory bowel

disease. [167] The relative abundance of *Christensenellaceae* in the human GM is inversely related to the body mass index (BMI) of the host, [167] and the same correlation, not inversely, is valid for *Bacteroides* [168]: these two genera linked to BMI were indeed found, respectively, reduced and profoundly increased in the Rett patients, who are thinner due to their lifestyle and syndrome with a BMI significantly lower (17.2 ± 3.9 in RTT patients, compared to 20.9 ± 2.2 in HC). On the other hand, *Christensenellaceae* has been found relevant in the feature selection analysis: the drug-assuming patients showed a reduced biodiversity but an increased relative abundance compared to the DN patients. The assumption about *Christensenellaceae* and the BMI could be valid for the epileptic patients as well, in the long term, as the SDT group has a profoundly decreased relative abundance of this family as well (0.01% in RTT and 0.5% in SDT *vs* 3% in HC, 2.2 in DN, 2.3% at DT4, and 2.4% in DT12).

From the network of bacterial interactions, a remarkable role was played by *Subdoligranulum* and the genera related to it. This genus was the only one within the homonym CAG to maintain a relative abundance above 1.6% in the Rett group, while all the others were dramatically reduced (<0.6%). The Subdoligranulum CAG was observed to be mostly depleted in RTT, whereas all the members of the Bacteroides CAG were increased. The relationships within the network for RTT patients were profoundly altered compared to the HC and the other experimental groups, sharing some similarities with the EPI group. Due to the small number of samples in each subset of patients, it was not possible to elaborate a co-abundance matrix and reliable network of interactions for all the groups, so any similarity or difference between the RTT and the SDT groups cannot be confuted through this analysis.

Limitations

The most relevant limitation of this research is related to the fact that the number of patients is relatively low and all these findings are to be considered only as first hints of a behavior that needs to be verified in larger cohorts of patients.

Due to the unpredictable nature of the seizure onset in children, the enrollment of a higher number of drug-naive patients in the recruitment period has been really difficult; moreover, often, consecutive samples from the same patient are not easy to obtain. As for the adults in stable drug therapy, many of them are in polytherapy, and the diverse mechanisms of action of the drugs assumed could lead to different confounding results. The epileptic patients sequenced and presented here were selected for monotherapy, with no other pharmacological treatments, as an attempt to have a clearer gut microbiota picture, possibly influenced only by known factors. Lastly, the Rett dataset is small as this syndrome is, luckily, rare to find and, even more rare and luckily for the patients, not every subject has epileptic seizure and assume AEDs.

The limitations due to the small number of patients could reduce the data reliability. The taxonomic relative abundances and significant findings may change with a higher amount of samples per category, as well as the alpha- and beta-diversity. As for the feature selection analysis, since the samples used in the present study were already correctly classified with clinical records and were not in such a great number that could have been divided into training and test sets, the interpretation of the result is not of a prediction but a statement, a descriptive model of the data, in order to identify a subset of predictive taxa. [151] For example, in a larger and longitudinal dataset there could have been a status label about an epileptic seizure happening or not in conjunction with the sampling, and the model

would have predicted for the dataset whether some bacterial genera were particularly informative for a seizure happening in some days during a treatment or dietary intervention.

The union of the different types of patients could lead to the hypothesis that sex and, in particular, age (as the SDT was of only adults and the RTT patients were older than the children sample-set) could have influenced the findings, obscuring the drug-related observations. As for the sex issue, the Rett and SDT patients enrolled were only females, as well as the majority of the overall subjects, so the only criticism that could be held in this issue is the imbalance toward one gender. For the importance of the influence of these differences among the datasets, both doubts have already been addressed by other studies: the age difference is thought to not interfere with the observations made, as the microbial diversity develops and converges toward an adult-like microbiota by the end of the first years of life; [67] as for the sex issue, the 16S resolution is thought to not being sufficient to unravel human GM differences linked to sex-dependent bacteria, [170] as it doesn't allow enough confidence to characterize the bacterial community at a species or strain level, where a difference between metabolite-producers that could interact with sex hormones could be seen.

CONCLUSION

This present study aimed at explore different aspects of epileptic observation in the light of gut microbiota alterations.

Despite the already discussed limitations of the study, many interesting speculations have been made that could set the stage for a bigger project that could lead to confirmations or afterthoughts about the diversities we observed with longitudinal sampling and focusing on drug assumption.

The results shown in this thesis highlight the existence of differences in terms of general microbial diversity and taxonomy. In addition to the influence from the diseased status itself, those bacterial differences might also be impacted by the pharmacological treatment that the patients were following. Of particular interest are the observations made about the time-line during the drug assumption, with a separation over time often clear and worthy of a deeper sequencing technique.

Further studies could implement the findings presented in this thesis. A larger number of patients at different times and disease development could decrease the amount of confounders due to the heterogeneity of the subjects, especially for age, sex, type of epilepsy, and drug assumed; a metabolomic analysis implementation could take into account the nutritional intakes and explore the impact of bacterial metabolites on epilepsy frequency and response to treatment; a whole metagenomic analysis could contribute to reinforce these findings with a better discrimination below the genus level and with the identification of biomarkers associated to the onset of the disease.

BIBLIOGRAPHY

[1] Gross R.A., “A brief history of epilepsy and its therapy in the western hemisphere”, *Epilepsy Research*. Volume 12, Issue 2, Pages 65-74. (1992) ISSN 0920-1211. https://doi.org/10.1016/0920-1211(92)90028-R.

[2] Epilepsy Toronto
https://epilepsytoronto.org/about-epilepsy/learn-about-epilepsy/famous-people-with-epilepsy/

[3] Banerjee P.N., Filippi D., Hauser W.A., “The descriptive epidemiology of epilepsy - a review.” *Epilepsy Research,* vol. 85,1 (2009) 31-45. doi:10.1016/j.eplepsyres.2009.03.003

[4] Kwan P., et al., “Drug-resistant epilepsy.” *The New England journal of medicine* vol. 365,10 (2011) 919-26. doi:10.1056/NEJMra1004418

[5] Stafstrom C.E., Carmant L., “Seizures and epilepsy: an overview for neuroscientists”, *Cold Spring Harb Perspect Med*. (2015) 5(6)a022426. doi:10.1101/cshperspect.a022426

[6] National Institute for Health and Clinical Excellence, “Epilepsies: diagnosis and management”. *National Clinical Guideline Centre.* (2012) pp. 21–28

[7] World Health Organization, "Epilepsy Fact sheet", February 2016. Archived from the original on 11 March 2016. Retrieved 4 March 2016. https://www.who.int/news-room/fact-sheets/detail/epilepsy

[8] Bromfield E.B., Cavazos J.E., Sirven J.I, editors, “An Introduction to Epilepsy [Internet]. West Hartford” (CT) *American Epilepsy Society* (2006). Chapter 2, Clinical Epilepsy. Available from: https://www.ncbi.nlm.nih.gov/books/NBK2511/

[9] van Donselaar C.A., Stroink H., Arts W Dutch Study Group of Epilepsy in Childhood, “How confident are we of the diagnosis of epilepsy?”, *Epilepsia*. (2006) 47 Suppl 1:9-13. doi:10.1111/j.1528-1167.2006.00653.x

[10] Caplan R., “Psychopathology and seizure threshold”, *Epilepsy currents*, *13*(3), (2013)141–142. https://doi.org/10.5698/1535-7511-13.3.141

[11] Ehinger Y., Matagne V., Villard L., Roux J.C.. “Rett syndrome from bench to bedside: recent advances". *F1000Res*. 2018;7:398. (2018) doi:10.12688/f1000research.14056.1

[12] Glauser T., Ben-Menachem E., Bourgeois B, Cnaan A., Guerreiro C., Kälviäinen R., Mattson R., French J.A., Perucca E., Tomson T., et al. “Updated ILAE evidence review of antiepileptic drug efficacy and effectiveness as initial monotherapy for epileptic seizures and syndromes”, *Epilepsia*. (2013) 54(3)551-563. doi:10.1111/epi.12074

[13] Remy S., and Heinz B., “Molecular and cellular mechanisms of pharmacoresistance in epilepsy.” *Brain: a journal of neurology* vol. 129,Pt 1 (2006) 18-35. doi:10.1093/brain/awh682

[14] Luszczki J.J., Czuczwar S.J., “Preclinical profile of combinations of some second generation antiepileptic drugs: an isobolographic analysis”., *Epilepsia (*2004) Aug;45(8):895-907. doi: 10.1111/j.0013-9580.2004.05104.x.

[15] Brodie M.J., Yuen A.W., “Lamotrigine substitution study: evidence for synergism with sodium valproate?” *Epilepsy Research,* (1997) Mar;26(3)423-32. doi: 10.1016/s0920-1211(96)01007-8.

[16] Pisani F., Oteri G., Russo M.F., Di P.R., Perucca E., Richens A., “The efficacy of valproate-lamotrigine comedication in refractory complex partial seizures: evidence for a pharmacodynamic interaction”, *Epilepsia,* (1999) Aug;40(8)1141-6. doi: 10.1111/j.1528-1157.1999.tb00832.x.

[17] Rowan A.J., Meijer J.W., de Beer-Paw-likowski N., van der Geest P., Meinardi H. “Valproate-ethosuximide combination therapy for refractory absence seizures”, *Arch Neurol,* (1983) Dec;40(13)797-802. doi: 10.1001/archneur.1983.04050120047006.

[18] Miller J.W., & Hakimian S., “Surgical treatment of epilepsy”, *Continuum (Minneapolis, Minn.)*, *19* (3 Epilepsy), (2013) 730–742. doi:10.1212/01.CON.0000431398.69594.97

[19] Lin Y., & Wang Y., “Neurostimulation as a promising epilepsy therapy”, *Epilepsia open*, *2*(4), (2017)371–387. https://doi.org/10.1002/epi4.12070

[20] Qin J., Li R., Raes J., et al.; MetaHIT Consortium. “A human gut microbial gene catalogue established by metagenomic sequencing”, *Nature* (2010) 464:59–65. doi: 10.1038/nature08821

[21] Bhattacharjee S., Lukiw W.J. “Alzheimer’s disease and the microbiome”, *Front Cell Neurosci*. (2013) 7:153. doi: 10.3389/fncel.2013.00153

[22] Baquero F. & Nombela C., “The microbiome as a human organ”, *Clinical Microbiology and Infection*, *18*(s4), 2–4. (2012) http://doi.org/10.1111/j.1469-0691.2012.03916.x

[23] Cheng J., Palva A. M., de Vos W. M., & Satokari R., “Contribution of the Intestinal Microbiota to Human Health: From Birth to 100 Years of Age” *Springer Berlin Heidelberg*. pp. 323–346. (2011) http://doi.org/10.1007/82_2011_189

[24] O’Hara, A. M., & Shanahan, F., “The gut flora as a forgotten organ”. *EMBO Reports*, *7*(7), 688–93. (2006) http://doi.org/10.1038/sj.embor.7400731

[25] Sender R., Fuchs S., et al., “Are We Really Vastly Outnumbered? Revisiting the Ratio of Bacterial to Host Cells in Humans.” *Cell* vol. 164,3 (2016) 337-40. doi:10.1016/j.cell.2016.01.013

[26] Nardone G. & Compare D., “The human gastric microbiota: Is it time to rethink the pathogenesis of stomach diseases?” *United European Gastroenterology Journal.,* (2015) 3(3), 255–60. http://doi.org/10.1177/2050640614566846

[27] Backhed F., Ley R.E., Sonnenburg J.L., Peterson D.A., “Gordon Host-bacterial mutualism in the human intestine. *Science”.* (2005) 307: 1915-1920

[28] Bengmark S., “Pre-, pro- and synbiotics”. *Curr Opin in Clin Nutr Metab Care,* (2001) 4:571-579.6.

[29] Savage D C. “Microbial ecology of the gastrointestinal tract.” *Annual review of microbiology* vol. 31 (1977) 107-33. doi:10.1146/annurev.mi.31.100177.000543

[30] Puche C., Guerrero R. and Berlanga M;
https://metode.es/revistas-metode/secciones/casa-microbios/yo-soy-yo-y-mi-microbiota.html

[31] Bengmark S., “Gut microbiota, immune development and function”, *Pharmacological Research*, *69*(1), (2013) 87–113. http://doi.org/10.1016/j.phrs.2012.09.002

[32] Dethlefsen L., McFall-Ngai M. & Relman D.A., “An ecological and evolutionary perspective on human-microbe mutualism and disease”, *Nature*, 449(7164), 811–8. (2007) http://doi.org/10.1038/nature06245

[33] Vrieze A., Holleman F., Zoetendal E. G., de Vos W. M., Hoekstra J. B. L., & Nieuwdorp, M., “The environment within: how gut microbiota may influence metabolism and body composition”, *Diabetologia*, 53(4), 606–613. (2010) http://doi.org/10.1007/s00125-010-1662-7

[34] McNally L., Brown S.P., “Building the microbiome in health and disease: niche construction and social conflict in bacteria2”, *Philos Trans R Soc Lond B Biol Sci*. (2015) 370:20140298. doi: 10.1098/rstb.2014.0298

[35] Waters C.M., Bassler B.L., “Quorum sensing: cell-to-cell communication in bacteria”, *Annu Rev Cell Dev Biol*. (2005) 21:319–346. doi: 10.1146/annurev.cellbio.21.012704.131001

[36] Kokou F., Sasson G., Nitzan T., Doron-Faigenboim A., Harpaz S., Cnaani A., Mizrahi I., “Host genetic selection for cold tolerance shapes microbiome composition and modulates its response to temperature”, *Elife*. (2018) 7:e36398. doi: 10.7554/eLife.36398

[37] Engevik M.A., Versalovic J. “Biochemical features of beneficial microbes: foundations for therapeutic microbiology”, *Microbiol Spectr.* (2017) vol.5.5:1–35. doi:10.1128/microbiolspec.BAD-0012-2016

[38] De Vadder F., Kovatcheva-Datchary P., Goncalves D., et al., “Microbiota generated metabolites promote metabolic benefits via gut-brain neural circuits”, *Cell*. (2014) 156:84–96. doi: 10.1016/j.cell.2013.12.016

[39] Goodnow C.C., Sprent J., Fazekas de St Groth B., Vinuesa C.G. “Cellular and genetic mechanisms of self tolerance and autoimmunity” *Nature (*2005) 435:590-7; PMID: 15931211; http://dx.doi.org/10.1038/nature03724

[40] Matricon J, et al., “Review article: Associations between immune activation, intestinal permeability and the irritable bowel syndrome”, *Aliment Pharmacol Ther*. (2012) Dec;36(11-12)1009-31. doi: 10.1111/apt.12080.

[41] Hughes P.A., Zola H., Penttila I.A., Blackshaw L.A., Andrews J.M., Krumbiegel D., “Immune activation in irritable bowel syndrome: can neuroimmune interactions explain symptoms?” *Am J Gas-troenterol*. (2013) Jul;108(7)1066-74. doi: 10.1038/ajg.2013.120.

[42] Forsythe P., Kunze W.A., “Voices from within: gut microbes and the CNS”, *Cell Mol Life Sci.* (2013) Jan;70(1)55-69. doi: 10.1007/s00018-012-1028-z.

[43] Dinan T.G., Cryan .J.F., “Regulation of the stress response by the gut microbiota: implications for psychoneuroendocrinology”. *Psychoneuroendocrinology,* (2012) Sep; 37(9)1369-78. doi: 10.1016/j.psyneuen.2012.03.007.

[44] Foster J.A., McVey Neufeld K.A., “Gut-brain axis: how the microbiome influences anxiety and depression”, *Trends Neurosci*. (2013) May;36(5)305-12. doi: 10.1016/j.tins.2013.01.005.

[45] Alcock J., Maley C.C., Aktipis C.A., “Is eating behavior manipulated by the gastrointestinal microbiota? Evolutionary pressures and potential mechanisms”. *Bioessays. (*2014) 36(10)940–949. doi: 10.1002/bies.201400071.

[46] Stilling R.M., Dinan T.G., Cryan J.F., “Microbial genes, brain & behaviour - epigenetic regulation of the gut-brain axis”, *Genes Brain Behav*. (2014) 13(1)69–86. doi: 10.1111/gbb.12109.

[47] Ramanan D., & Cadwell K., “Review Intrinsic Defense Mechanisms of the Intestinal Epithelium”, *Cell host & microbe,* vol. 19,4 (2016) 434-41. doi:10.1016/j.chom.2016.03.003

[48] Xu J., and Jeffrey I.G., “Honor thy symbionts.” *Proceedings of the National Academy of Sciences of the United States of America* vol. 100,18 (2003) 10452-9. doi:10.1073/pnas.1734063100

[49] Costello E.K., Lauber C.L., Hamady M., Fierer N., Gordon J.I., and Knight R. “Bacterial Community Variation in Human Body Habitats Across Space and Time”, *Science* (2009) Dec 18;326(5960)1694-7. doi: 10.1126/science.1177486.

[50] Lahner E. et al. “Gastric precancerous conditions and Helicobacter pylori infection in dyspeptic patients with or without endoscopic lesions.” *Scandinavian journal of gastroenterology* vol. 51,11 (2016) 1294-8. doi:10.1080/00365521.2016.1205129

[51] Clemente J.C., Ursell L.K., Parfrey L.W., Knight R. “The impact of the gut microbiota on human health: an integrative view”. *Cell*. (2012) Mar 16;148(6):1258-70. doi: 10.1016/j.cell.2012.01.035.

[52] Frank D.N., St. Amand A.L., Feldman R.A., et al., “Molecular-phyloge- netic characterization of microbial community imbalances in human inflammatory bowel diseases”, *Proc Natl Acad SciUSA* (2007) Aug 21;104(34)13780-5. doi: 10.1073/pnas.0706625104.

[53] Muegge B.D., Kuczynski J., Knights D., Clemente J.C., Gonzalez A., Fontana L., Henrissat B., Knight R., and Gordon J.I. “Diet Drives Convergence in Gut Microbiome Functions Across Mammalian Phylogeny and Within Humans”. *Science* (2011) May 20;332(6032)970-4. doi: 10.1126/science.1198719.

[54] Tidjani Alou M., Lagier J.C., Raoult D. “Diet influence on the gut microbiota and dysbiosis related to nutritional disorders”, *Hum. Microbiome J*. (2016) 1:3–11. doi: 10.1016/j.humic.2016.09.001.

[55] Al-Sheraji S.H., Ismail A., Manap M.Y., Mustafa S., Yusof R.M., & Hassan F.A. “Prebiotics as functional foods: A review.” *Journal of Functional Foods,* 5(4), 1542–1553. (2013) http://doi.org/10.1016/j.jff.2013.08.009

[56] Kristensen N.B., Bryrup T., Allin K.H., Nielsen T., Hansen T.H., Pedersen O., Cole J. “Alterations in fecal microbiota composition by probiotic supplementation in healthy adults: a systematic review of randomized controlled trials.” *Genome Medicine* 8(1), 52. (2016) http://doi.org/10.1186/s13073-016-0300-5

[57] Zarrinpar A., Chaix A., Xu Z.Z. *et al.,* “Antibiotic-induced microbiome depletion alters metabolic homeostasis by affecting gut signaling and colonic metabolism” *Nat Commun* 9**,** 2872, (2018) https://doi.org/10.1038/s41467-018-05336-9

[58] Bäumler A.J., & Sperandio V. “Interactions between the microbiota and pathogenic bacteria in the gut”, *Nature*, 535(7610), (2016) 85–93. http://doi.org/10.1038/nature18849

[59] Perez-Munoz M.E., Arrieta M.C, .Ramer-Tait A.E., Walter J., “A critical assessment of the ‘sterile womb’ and ‘in utero colonization’ hypotheses: implications for research on the pioneer infant microbiome”, *Microbiome* 5, 48. (2017) doi:10.1186/s40168-017-0268-4

[60] Favier C.F., Vaughan E.E., De Vos W.M., et al., “Molecular monitoring of succession of bacterial communities in human neonates”, *Applied and environmental microbiology* vol. 68,1 (2002) 219-26. doi:10.1128/aem.68.1.219-226.2002

[61] Palmer C., Bik E.M., DiGiulio D.B., et al 2007, “Development of the human infant intestinal microbiota” *PLoS Biol* vol. 5,7 (2007) e177. doi:10.1371/journal.pbio.0050177

[62] Penders J., Thijs C., Vink C., et al., “Factors influencing the composition of the intestinal microbiota in early infancy”, *Pediatrics* vol. 118,2 (2006) 511-21. doi:10.1542/peds.2005-2824

[63] Dominguez-Bello M.G., et al., “Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns.” *Proceedings of the National Academy of Sciences of the United States of America* vol. 107,26 (2010) 11971-5. doi:10.1073/pnas.1002601107

[64] Fallani M. et al., “Intestinal microbiota of 6-week-old infants across Europe: geographic influence beyond delivery mode, breast-feeding, and antibiotics.” *Journal of pediatric gastroenterology and nutrition* vol. 51,1 (2010) 77-84. doi:10.1097/MPG.0b013e3181d1b11e

[65] Shin H., Pei Z., Martinez K.A., Rivera-Vinas J.I., Mendez K., Cavallin H., & Dominguez-Bello M.G. “The first microbial environment of infants born by C-section: the operating room microbes.” *Microbiome*, 3**,** 59 (2015). https://doi.org/10.1186/s40168-015-0126-1

[66] Arboleya S., Suárez M., Fernández N., Mantecón L., Solís G., Gueimonde M., de los Reyes-Gavilán C.G., “C-section and the Neonatal Gut Microbiome Acquisition: Consequences for Future Health”. *Ann Nutr Metab,* 73(suppl 3)17-23. (2018) doi: 10.1159/000490843

[67] Rodríguez J.M., Murphy K., Stanton C. et.al. “The composition of the gut microbiota throughout life, with an emphasis on early life”. *Microbial ecology in health and disease 26*,26050. (2015) https://doi.org/10.3402/mehd.v26.26050

[68] Eckburg P.B., Bik E.M., Bernstein C.N., et al. “Diversity of the human intestinal microbial flora.” *Science (New York, N.Y.)* vol. 308,5728 (2005) 1635-8. doi:10.1126/science.1110591

[69] Tap J., et al., “Towards the human intestinal microbiota phylogenetic core.” *Environmental microbiology* vol. 11,10 (2009) 2574-84. doi:10.1111/j.1462-2920.2009.01982.x

[70] Frank D.N., et al. “Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases.” *Proceedings of the National Academy of Sciences of the United States of America* vol. 104,34 (2007) 13780-5. doi:10.1073/pnas.0706625104

[71] Scholtens P.A., Oozeer R., Martin R., Amor K.B., Knol J., “The early settlers: intestinal microbiology in early life”, *Annu. Rev. Food Sci. Technol*. Vol.3, (2012) 425–447 10.1146/annurev-food-022811-101120

[72] Rajilić-Stojanović M., Shanahan F., Guarner F., & de Vos W.M., “Phylogenetic analysis of dysbiosis in ulcerative colitis during remission. *Inflammatory Bowel Diseases”*, *19*(3), (2013) 481–8. http://doi.org/10.1097/MIB.0b013e31827fec6d

[73] Zoetendal E.G., Rajilic-Stojanovic M., De Vos W.M., “High-throughput diversity and functionality analysis of the gastrointestinal tract microbiota”, *Gut* 57, (2008) 1605–1615 10.1136/gut.2007.133603

[74] Jalanka-Tuovinen J., Salonen A., Nikkila J., Immonen O., Kekkonen R., Lahti L., Palva A., De Vos W.M., “Intestinal microbiota in healthy adults: temporal analysis reveals individual and common core and relation to intestinal symptoms”, *PLoS ONE* (2011) 6:e23035 10.1371/journal.pone.0023035

[75] Arumugam M., Raes J., Pelletier E., Le Paslier D., Yamada T., Mende D.R., Fernandes G.R., Tap J., Bruls T., Batto J.M., “Enterotypes of the human gut microbiome, *Nature* (2011) 473, 174–180 10.1038/nature09944

[76] Biagi E., Nylund L., Candela M., Ostan R., Bucci L., Pin E., Nikkïl J., Monti D., Satokari R., Franceschi C., et al., “Through Ageing, and Beyond: Gut Microbiota and Inflammatory Status in Seniors and Centenarians” *PLoS ONE* (2010) vol. 5,5 e10667. (2010), doi:10.1371/journal.pone.0010667

[77] Biagi E., Candela M., Turroni S., Garagnani P., Franceschi C., and Brigidi P., “Ageing and gut microbes: Perspectives for health maintenance and longevity” *Pharmacological Research* vol. 69,1 (2013) 11-20. doi:10.1016/j.phrs.2012.10.005

[78] Ottman N., Smidt H., de Vos W.M., & Belzer C., “The function of our microbiota: who is out there and what do they do?” *Frontiers in Cellular and Infection Microbiology* (2012) 2, 104. http://doi.org/10.3389/fcimb.2012.00104

[79] Falony G., Joossens M., Vieira-Silva S., Wang J., Darzi Y., Faust K., et al., “Population-level analysis of gut microbiome variation”. *Science,* (2016) Apr 29;352(6285)560-4. doi: 10.1126/science.aad3503.

[80] Makki K., Deehan E.C., Walter J., Bäckhed F., “The impact of dietary fiber on gut microbiota in host health and disease”, *Cell Host Microbe.* (2018) Jun 13;23(6)705-715. doi: 10.1016/j.chom.2018.05.012.

[81] Sonnenburg E.D., Sonnenburg J.L., “Starving our microbial self: the deleterious consequences of a diet deficient in microbiota-accessible carbohydrates”, *Cell Metab.* (2014) Nov 4;20(5)779-786. doi: 10.1016/j.cmet.2014.07.003.

[82] Yatsunenko T., Rey F.E., Manary M.J., Trehan I., Dominguez-Bello M.G., Contreras M., Magris M., Hidalgo G., Baldassano R.N., Anokhin A.P., Heath A.C., Warner B., Reeder J., Kuczynski J., Caporaso J.G., Lozupone CA, Lauber C, Clemente J.C., Knights D., Knight R., Gordon J.I., et al. “Human gut microbiome viewed across age and geography.” *Nature* vol. 486,7402 222-7. (2012) doi:10.1038/nature11053

[83] Schnorr S., Candela M., Rampelli S., et al. "Gut microbiome of the Hadza hunter-gatherers". *Nat Commun* 5, 3654 (2014) doi:/10.1038/ncomms4654

[84] Sonnenburg J.L., Bäckhed F. “Diet-microbiota interactions as moderators of human metabolism”, *Nature,* (2016) Jul 7;535(7610)56-64. doi: 10.1038/nature18846.

[85] Wu G.D., Compher C., Chen E.Z., Smith S.A., Shah R.D., Bittinger K., Chehoud C., Albenberg L.G., Nessel L., Gilroy E., et al. “Comparative metabolomics in vegans and omnivores reveal constraints on diet-dependent gut microbiota metabolite production”, *Gut*. (2016); 65:63–72. doi: 10.1136/gutjnl-2014-308209.

[86] Neal E.G., et al. “The ketogenic diet for the treatment of childhood epilepsy: a randomised controlled trial.” *The Lancet. Neurology* vol. 7,6 (2008) 500-6. doi:10.1016/S1474-4422(08)70092-9

[87] Winesett S.P., et al. “The ketogenic diet in pharmacoresistant childhood epilepsy.” *Expert review of neurotherapeutics* vol. 15,6 (2015) 621-8. doi:10.1586/14737175.2015.1044982

[88] Xie G., Zhou Q,. Qiu C.Z., et al. “Ketogenic diet poses a significant effect on imbalanced gut microbiota in infants with refractory epilepsy”. *World J Gastroenterol*. (2017) 23(33):6164-6171. doi:10.3748/wjg.v23.i33.6164

[89] Lindefeldt M., Eng A., Darban H., Bjerkner A., Zetterström C.K., Allander T., et al., “The ketogenic diet influences taxonomic and functional composition of the gut microbiota in children with severe epilepsy”, *npj Biofilms Microbiomes.* 5**,** 5 (2019). doi:10.1038/s41522-018-0073-2

[90] Dahlin M, Prast-Nielsen S., “The gut microbiome and epilepsy”, *EBioMedicine* 44 741–746, (2019) doi:https://doi.org/10.1016/j.ebiom.2019.05.024

[91] Qin J., et al. “A human gut microbial gene catalogue established by metagenomic sequencing.” *Nature* vol. 464,7285 (2010) 59-65. doi:10.1038/nature08821

[92] Grenham S., Clarke G., Cryan J.F., Dinan T.G., “Brain-gut-microbe communication in health and disease”, *Front Physiol.* (2011) Dec 7;2:94. doi: 10.3389/fphys.2011.00094.

[93] Cryan J.F. Dinan T.G., “Mind-altering microorganisms: the impact of the gut microbiota on brain and behaviour”, *Nat Rev Neurosci.* 13, 701–712 (2012). doi: 10.1038/nrn3346

[94] Mulle J.G., Sharp W.G., Cubells J.F., “The gut microbiome: a new frontier in autism research”, *Curr Psychiatry Rep.* (2013) Feb;15(2)337. doi: 10.1007/s11920-012-0337-0.

[95] Jangi S., Gandhi R., Cox L.M., Li N., von Glehn F., Yan R., et al., “Alterations of the human gut microbiome in multiple sclerosis”, *Nat Commun.* (2016) Jun 28;7:12015. doi: 10.1038/ncomms12015.

[96] Barichella M., Severgnini M., Cilia R., Cassani E., Bolliri C., Caronni S., Ferri V., Cancello R., Ceccarani C., Faierman S., Pinelli G., et al. “Unraveling gut microbiota in Parkinson's disease and atypical parkinsonism”. *Mov Disord*. (2019) Mar;34(3):396-405. doi: 10.1002/mds.27581.

[97] Jiang C., Li G., Huang P., Liu Z., Zhao B., “The gut microbiota and alzheimer's disease”, *J Alzheimers Dis.* (2017) 58(1)1-15. doi: 10.3233/JAD-161141

[98] Rothschild D., Weissbrod O., Barkan E., Kurilshikov A., Korem T., Zeevi D., et al., “Environment dominates over host genetics in shaping human gut microbiota”, *Nature.* 555**,** 210–215 (2018). https://doi.org/10.1038/nature25973

[99] Montiel-Castro A.J., González-Cervantes R.M., Bravo-Ruiseco G., Pacheco-López G. “The microbiota-gut-brain axis: neurobehavioral correlates, health and sociality”. *Front Integr Neurosci*. (2013) Oct 7;7:70. doi: 10.3389/fnint.2013.00070.

[100] Singh R.K., Chang H.W., Yan D., et al., “Influence of diet on the gut microbiome and implications for human health”, *J Transl Med*. (2017) 15:73. doi:10.1186/s12967-017-1175-y

[101] Bailey M.T., Dowd S.E., Galley J.D., Hufnagle A.R., Allen R.G., Lyte M., “Exposure to a social stressor alters the structure of the intestinal microbiota: implications for stressor-induced immunomodulation”, *Brain Behav Immun*. (2011) 25:397–407. doi: 10.1016/j.bbi.2010.10.023

[102] Collins S.M., Surette M., Bercik P, “The interplay between the intestinal microbiota and the brain”, *Nat Rev Microbiol*. (2012) 10:735–742. doi:10.1038/nrmicro2876

[103] Sudo N., Chida Y., Aiba Y., et al., “Postnatal microbial colonization programs the hypothalamic-pituitary-adrenal system for stress response in mice”, *J Physiol*. (2004) 558(Pt 1)263–275. doi: 10.1113/jphysiol.2004.063388

[104] Sgritta M., Dooling S.W., Buffington S.A., et al., “Mechanisms underlying microbial-mediated changes in social behavior in mouse models of autism spectrum disorder”, *Neuron*. (2019) 101:246–259.e6. doi: 10.1016/j. neuron.2018.11.018

[105] Berer K., Mues M., Koutrolos M., Rasbi Z.A., Boziki M., Johner C., Wekerle H., Krishnamoorthy G. “Commensal microbiota and myelin autoantigen cooperate to trigger autoimmune demyelination”. *Nature* (2011) Oct 26;479(7374):538-41. doi: 10.1038/nature10554.

[106] Simren M, et al. “Intestinal microbiota in functional bowel disorders: a Rome foundation report”. *Gut* (2013) Jan; 62(1): 159–176. doi: 10.1136/gutjnl-2012-302167

[107] de Vos W.M., de Vos E.A., “Role of the intestinal microbiome in health and disease: from correlation to causation”, *Nutr Rev*. (2012) Aug;70 Suppl 1:S45-56. doi: 10.1111/j.1753-4887.2012.00505.x.

[108] Stilling R.M., Bordenstein S.R., Dinan T.G., Cryan J.F., “Friends with social benefits: host-microbe interactions as a driver of brain evolution and development?” *Front Cell Infect Microbiol*. (2014) Oct 29;4:147. doi: 10.3389/fcimb.2014.00147.

[109] Mayer E.A., Tillish K., Gupta A., “Gut/brain axis and the microbiota”, *J Clin Invest*. (2015) 125(3)926-938. doi: 10.1172/JCI76304.

[110] Lozupone C.A., Stombaugh J.I., Gordon J.I., Jansson J.K., Knight R., “Diversity, stability and resilience of the human gut microbiota”, *Nature*. (2012) 489:220–230. doi: 10.1038/nature11550

[111] Hornef M., “Pathogens, commensal symbionts, and pathobionts: discovery and functional effects on the host”, *ILAR J*. (2015) 56:159–162. doi:10.1093/ilar/ilv007

[112] Yoon M.Y., Lee K., Yoon S.S., “Protective role of gut commensal microbes against intestinal infections”, *J Microbiol. (*2014) 52:983–989. doi:10.1007/s12275-014-4655-2

[113] He Z., Cui B-T, Zhang T., Li P., Long C-Y, Ji. G-Z, et al., “Fecal microbiota transplantation cured epilepsy in a case with Crohn's disease: The first report” *World J Gastroenterol* (2017) May 21;23(19)3565–8.

[114] Paoli A., Mancin L., Bianco A., Thomas E., Mota J. F., & Piccini F., “Ketogenic Diet and Microbiota: Friends or Enemies?”, *Genes* (2019) 10(7), 534. doi.org/10.3390/genes10070534

[115] Behjati S., Patrick S.T. “What is next generation sequencing?.” *Archives of disease in childhood*. Education and practice edition vol. 98,6 (2013) 236-8. doi:10.1136/archdischild-2013-304340

[116] Sanger F., et al. “DNA sequencing with chain-terminating inhibitors.” *Proceedings of the National Academy of Sciences of the United States of America* vol. 74,12 (1977) 5463-7. doi:10.1073/pnas.74.12.5463

[117] Luo C., et al. “Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample.” *PloS one* vol. 7,2 (2012) e30087. doi:10.1371/journal.pone.0030087

[118] Lahens N.F., et al. “A comparison of Illumina and Ion Torrent sequencing platforms in the context of differential gene expression.” *BMC genomics* vol. 18,1 602. (2017) doi:10.1186/s12864-017-4011-0

[119] Eid J., et al. “Real-time DNA sequencing from single polymerase molecules.” Science vol. 323,5910 (2009) 133-8. doi:10.1126/science.1162986

[120] Shendure J., and Hanlee J. “Next-generation DNA sequencing.” *Nature biotechnology* vol. 26,10 (2008): 1135-45. doi:10.1038/nbt1486

[121] Ng P.C., and Ewen F.K. “Whole genome sequencing.” *Methods in molecular biology* (Clifton, N.J.) vol. 628 (2010) 215-26. doi:10.1007/978-1-60327-367-1_12

[122] www.illumina.com/systems/miseq/performance_specifications.html

[123] Holt R.A., Jones S.J. “The new paradigm of flow cell sequencing”. *Genome Res*. (2008) Jun;18(6):839-46. doi: 10.1101/gr.073262.107

[124] Caporaso J.G., Kuczynski J., Stombaugh J., Bittinger K., Bushman, F.D., Costello E.K., Fierer, N. Peña A.G., Goodrich J.K., Gordon J.I., et al., “QIIME allows analysis of high-throughput community sequencing data”, *Nat. Methods* (2010) May;7(5)335-6. doi:10.1038/nmeth.f.303.

[125] Weinstock G.M., “Genomic approaches to studying the human microbiota” *Nature (*2012) 489(7415), 250–6. http://doi.org/10.1038/nature11553

[126] Edgar R.C., Haas B.J., Clemente J.C., Quince C. & Knight R. “UCHIME improves sensitivity and speed of chimera detection”, *Bioinformatics* (2011) Aug 15;27(16)2194-200. doi: 10.1093/bioinformatics/btr381.

[127] Schloss P.D., Gevers D., and Westcott S.L., “Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies”. *PLoS ONE* (2011);6(12)e27310. doi: 10.1371/journal.pone.0027310.

[128] Ewing B., Hillier L; Wendl MC; Green P. "Base-calling of automated sequencer traces using phred. I. Accuracy assessment". *Genome Research*. 8 (3): 175–185. (1998) doi:10.1101/gr.8.3.175. PMID 9521921.

[129] Ewing B, Green P. "Base-calling of automated sequencer traces using phred. II. Error probabilities". *Genome Research.* 8 (3): 186–194. (1998) doi:10.1101/gr.8.3.186. PMID 9521922.

[130] Haas B.J., Gevers D., Earl A.M., Feldgarden M.,Ward D.V., Giannoukos G., Ciulia D., Tabbaa D., Highlander S.K., Sodergren E., “Chimeric 16S rRNA

sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons", *Genome Res*. (2011) Mar;21(3)494-504. doi:10.1101/gr.112730.110.

[131] https://www.drive5.com/usearch/manual/uparseotu_algo.jpg

[132] Di Bella J.M., Bao Y., Gloor G.B., Burton J.P., & Reid G., "High throughput sequencing methods and analysis for microbiome research", *Journal of Microbiological Methods*, (2013) 95(3), 401–14. doi:10.1016/j.mimet.2013.08.011

[133] Gilbert J.A., Field D., Swift P., Thomas S., Cummings D., Temperton B., Mühling M. "The taxonomic and functional diversity of microbes at a temperate coastal site: a "multi-omic" study of seasonal and diel temporal variation", *PloS One*, (2010) 5(11), e15545. doi:10.1371/journal.pone.0015545

[134] Martin J., et al., "Optimizing read mapping to reference genomes to determine composition and species prevalence in microbial communities" *PloS ONE* (2012) 7(6) e36427. doi:10.1371/journal.pone.0036427

[135] Desantis T.Z., Hugenholtz P., Larsen N., Rojas M., Brodie E.L., Keller K., Huber T., Dalevi D., Hu P., and Andersen G.L. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and environmental microbiology*, *72*(7), 5069–5072. (2006) https://doi.org/10.1128/AEM.03006-05

[136] Pruesse E., Quast C., Knittel K., et al., "SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB", *Nucleic Acids Res,* (2007) 35(21)7188-7196. doi:10.1093/nar/gkm864

[137] Maidak B.L., Olsen G.J., Larsen N., Overbeek R., McCaughey M.J., Woese C.R., "The RDP (Ribosomal Database Project)." *Nucleic Acids Res*. (1997) ;25(1)109-111. doi:10.1093/nar/25.1.109

[138] Altschul S.F., Gish W., Miller W., Myers E.W. & Lipman D.J. "Basic local alignment search tool." *J. Mol. Biol.* (1990) 215:403-410.

[139] Wang Q., Garrity G.M., Tiedje J.M., Cole J.R., "Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy", *Appl Environ Microbiol*. (2007) 73(16)5261-5267. doi:10.1128/AEM.00062-07

[140] Kuczynski J., Stombaugh J., Walters W.A., González A., Caporaso J.G., Knight R., "Using QIIME to Analyze 16S rRNA Gene Sequences from Microbial Communities", *Current Protocols in Microbiology*, (2012) 1E.5.1–1E.5.20. doi:10.1002/9780471729259.mc01e05s27

[141] Whittaker R.H. "Evolution and Measurement of Species Diversity". *Taxon,* Vol. 21, No. 2/3 (1972) pp. 213-251 doi:10.2307/1218190

[142] Kim B.R., Shin J., Guevarra R., et al., “Deciphering Diversity Indices for a Better Understanding of Microbial Communities”, *J Microbiol Biotechnol*. 2017;27(12)2089-2093. doi:10.4014/jmb.1709.09027

[143] Chao A. “Non-parametric estimation of the number of classes in a population.” *Scandinavian Journal of Statistics* Vol. 11, No. 4 (1984) pp. 265-270 https://www.jstor.org/stable/4615964

[144] Shannon P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T., “Cytoscape: a software environment for integrated models of biomolecular interaction networks” *Genome research 13*(11), (2003) 2498–2504. https://doi.org/10.1101/gr.1239303

[145] Lozupone C., Knight R. “UniFrac: a new phylogenetic method for comparing microbial communities”. *Appl Environ Microbiol*. (2005) Dec;71(12):8228-35. doi: 10.1128/AEM.71.12.8228-8235.2005.

[146] Lozupone C.A., Hamady M., Kelley S.T., Knight R. “Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities.” *Appl Environ Microbiol*. (2007) Mar;73(5):1576-85. doi: 10.1128/AEM.01996-06.

[147] Lozupone C., Lladser M.E., Knights D., Stombaugh J. & Knight R. “UniFrac: an effective distance metric for microbial community comparison”, *ISME* J. (2011) Feb;5(2)169-72. doi: 10.1038/ismej.2010.133.

[148] Adonis test.
https://www.rdocumentation.org/packages/vegan/versions/2.4-2/topics/adonis

[149] Claesson M., Jeffery I., Conde S. et al. “Gut microbiota composition correlates with diet and health in the elderly”. *Nature* 488, 178–184 (2012). https://doi.org/10.1038/nature11319

[150] Langley P., “The changing science of machine learning” *Machine Learning*. 82 (3) 275–279. (2011) doi:10.1007/s10994-011-5242-y.

[151] Knights D., Costello E.K., Knight R., “Supervised classification of human microbiota.” *FEMS microbiology reviews* vol. 35,2 (2011) 343-59. doi:10.1111/j.1574-6976.2010.00251.x

[152] Hastie T., Tibshirani R, Friedman J. 2009a “The Elements of Statistical Learning, Second Edition: Data Mining, Inference, and Prediction. 2nd end”, *Springer*, Berlin, 20pp.

[153] Borghi E., Borgo F., Severgnini M., Savini M.N., Casiraghi M.C., Vignoli A. “Rett Syndrome: A Focus on Gut Microbiota”, *Int J Mol Sci*. (2017) 18(2)344. doi:10.3390/ijms18020344

[154] Neul J.L., Kaufmann W.E., Glaze D.G., Christodoulou J., Clarke A.J., Bahi-Buisson N., Leonard H., Bailey M.E.S., Schanen N.C., Zappella M., et al., “Rett syndrome: Revised diagnostic criteria and nomenclature”, *Ann. Neurol.* (2010) Dec;68(6)944-50. doi: 10.1002/ana.22124.

[155] Masella A.P., Bartram A.K., Truszkowski J.M., Brown D.G., Neufeld J.D., “PANDAseq: Paired-end assembler for illumina sequences”, *BMC Bioinform.* (2012) Feb 14;13:31. doi: 10.1186/1471-2105-13-31.

[156] Edgar R.C., “Search and clustering orders of magnitude faster than BLAST”, *Bioinformatics* (2010) Oct 1;26(19)2460-1. doi: 10.1093/bioinformatics/btq461.

[157] Oksanen J., Blanchet F.G., Kindt R., Legendre P., Minchin P.R., O'Hara R.B. et al. “The vegan package. Community ecology package”. (2015) *R package version 2.3-2. Available at:* http://CRAN.R-project.org/package=vegan

[158] RStudio Team 2018. “RStudio: Integrated Development for R. RStudio” Inc., Boston, MA http://www.rstudio.com/

[159] R Core Team 2013. R: “A language and environment for statistical computing”, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0 http://www.R-project.org/

[160] https://cran.r-project.org/web/packages/rstatix/index.html

[161] Wickham H., et al., “Welcome to the tidyverse”. *Journal of Open Source Software,* (2019) 4(43), 1686, https://doi.org/10.21105/joss.01686

[162] Shannon P., et al. “Cytoscape: a software environment for integrated models of biomolecular interaction networks.” *Genome research* vol. 13,11 (2003) 2498-504. doi:10.1101/gr.1239303

[163] Henke M.T., Kenny D.J., Cassilly C.D., Vlamakis H., Xavier R.J., Clardy J. "Ruminococcus gnavus, a member of the human gut microbiome associated with Crohn's disease, produces an inflammatory polysaccharide". *Proceedings of the National Academy of Sciences.* 116 (26): 12672–12677. (2019) doi:10.1073/pnas.1904099116

[164] Vacca M., Celano G., Calabrese F.M., Portincasa P., Gobbetti M., and De Angelis M. “The Controversial Role of Human Gut Lachnospiraceae”. *Microorganisms*, (2020) 8(4), 573. https://doi.org/10.3390/microorganisms8040573

[165] Pedersen H.K., et al. “Human gut microbes impact host serum metabolome and insulin sensitivity.” *Nature* vol. 535,7612 (2016) 376-81. doi:10.1038/nature18646

[166] Leite A.Z., Rodrigues N.C., Gonzaga M.I., Paiolo J.C.C., de Souza C.A., Stefanutto N.A.V., et al. "Detection of increased plasma interleukin-6 levels and

prevalence of prevotella copri and Bacteroides vulgatus in the feces of Type 2 diabetes patients". *Front. Immunol.* (2017) 8:1107. doi: 10.3389/fimmu.2017.01107

[167] Waters J.L., and Ruth E.L. “The human gut bacteria Christensenellaceae are widespread, heritable, and associated with health.” *BMC biology* vol. 17,1 83. 28 Oct. (2019) doi:10.1186/s12915-019-0699-4

[168] Fernandes J., Su W., Rahat-Rozenbloom S. Wolever T.M., Comelli E.M. "Adiposity, gut microbiota and faecal short chain fatty acids are linked in adult humans". *Nutr. Diabetes* (2014), 4, e121. doi: 10.1038/nutd.2014.23

[169] Maier L., et al. “Extensive impact of non-antibiotic drugs on human gut bacteria.” *Nature* vol. 555,7698 (2018), 623-628. doi:10.1038/nature25979

[170] Shin J.-H., et al. “Serum level of sex steroid hormone is associated with diversity and profiles of human gut microbiome.” *Research in microbiology* vol. 170,4-5 (2019): 192-201. doi:10.1016/j.resmic.2019.03.003

APPENDIX

Supplementary tables

Supplementary Table 1. Taxonomy comparison between HC and all the Patients at genus level. Data are expressed as relative abundance (standard deviation). P-values have been adjusted for multiple comparisons with Bonferroni correction.

Genus	HC	Patients	P-value
Bacteroides	26.1 (11.5)	26.1 (15.5)	--
Faecalibacterium	7.9 (8.9)	5.8 (5.2)	--
Subdoligranulum	6.2 (6.2)	7.6 (7.2)	--
Alistipes	5.3 (8.5)	4.5 (5.3)	--
Dialister	3.8 (4.5)	3.2 (3.8)	--
Agathobacter	3.1 (3.2)	2.5 (2.7)	--
Christensenellaceae R-7 group	2.9 (3.9)	1.4 (2.1)	--
Lachnospiraceae NK4A136 group	2.5 (3.4)	1.8 (3.3)	--
Ruminococcaceae UCG-002	2.4 (2.3)	1.9 (2.3)	--
Ruminococcus 2	2.3 (3.2)	1.5 (2.8)	--
Bifidobacterium	2.1 (4.3)	2.4 (3.0)	--
Parabacteroides	2.0 (2.4)	1.8 (2.8)	--
[Eubacterium] coprostanoligenes group	1.8 (1.8)	1.7 (2.4)	--
Barnesiella	1.7 (3.3)	0.7 (0.9)	--
Roseburia	1.6 (1.8)	1.0 (1.1)	--
Akkermansia	1.6 (2.7)	4.1 (7.5)	--
Blautia	1.3 (3.1)	1.1 (1.3)	--
Ruminococcaceae UCG-014	1.2 (1.8)	1.5 (3.6)	--
Ruminococcus 1	1.2 (1.4)	0.8 (1.0)	--
Ruminococcaceae UCG-005	0.8 (0.9)	0.6 (0.8)	--

Supplementary Table 2. Taxonomy comparison at phylum level between healthy controls (HC), drug-naive patients (DN), epileptic patients in drug therapy (EPI = DT4, DT12 and SDT), patients affected by Rett syndrome. Data are expressed as relative abundance (standard deviation). P-values have been adjusted for multiple comparisons with Bonferroni correction.

Phylum	HC	DN	EPI	RTT	P-value
Firmicutes	54.7 (19.3)	56.8 (13.2)	53.2 (14.5)	35.6 (10.6)	0.044 (HC-RTT)
Bacteroidetes	38.2 (17.7)	29 (17.5)	35.9 (15.7)	52.4 (11.7)	0.047 (DN-RTT)
Proteobacteria	2.7 (3.3)	1.9 (2.0)	5.4 (9.2)	5.6 (5.1)	--
Actinobacteria	2.3 (4.3)	3.4 (4.2)	2.1 (2.3)	3.3 (4.0)	--
Verrucomicrobia	1.7 (2.7)	8.7 (10.8)	3.1 (6.7)	2.2 (3.4)	0.020 (HC-DN)
Euryarchaeota	0.1 (0.3)	0.1 (0.2)	0.0 (0.0)	0.0 (0.0)	--
Tenericutes	0.1 (0.4)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	--

Supplementary Table 3. Taxonomy comparison at family level between healthy controls (HC), drug-naive patients (DN), epileptic patients in drug therapy (EPI = DT4, DT12 and SDT), patients affected by Rett syndrome. Data are expressed as relative abundance (standard deviation). P-values have been adjusted for multiple comparisons with Bonferroni correction.

Family	HC	DN	EPI	RTT	P-value
Ruminococcaceae	28.2 (15.2)	30.9 (13.2)	30.0 (12.2)	8.8 (8.1)	0.006 (HC-RTT) 0.0134 (DN-RTT) 0.003 (EPI-RTT)
Bacteroidaceae	26.1 (11.5)	22.8 (16.7)	22.6 (14.5)	40.7 (8.1)	0.0134 (EPI-RTT)
Lachnospiraceae	16.0 (8.6)	14.9 (6.3)	14.2 (6.6)	13.9 (11.0)	--
Rikenellaceae	5.3 (8.5)	2.9 (2.3)	5.2 (5.3)	5.7 (9.2)	--
Veillonellaceae	4.3 (4.8)	6.0 (5.8)	3.9 (3.6)	5.4 (5.3)	--
Christensenellaceae	3.0 (4.0)	2.2 (2.7)	1.7 (2.1)	1.0 (0.1)	--
Bifidobacteriaceae	2.1 (4.3)	3.2 (4.0)	1.9 (2.3)	2.9 (3.9)	--
Tannerellaceae	2.1 (2.4)	1.3 (2.2)	1.1 (1.9)	4.4 (4.5)	0.026 (EPI-RTT)
Barnesiellaceae	1.9 (4.2)	0.9 (1.0)	0.8 (0.9)	0.6 (0.9)	--
Prevotellaceae	1.6 (4.3)	0.5 (0.9)	5.5 (13.2)	0.2 (0.3)	--
Akkermansiaceae	1.6 (2.7)	8.7 (10.8)	3.1 (6.7)	2.2 (3.4)	0.020 (HC-DN)
Acidaminococcaceae	1.3 (2.3)	1.1 (1.3)	1.9 (4.2)	4.3 (4.1)	--
Enterobacteriaceae	1.3 (2.7)	1.0 (1.3)	3.3 (9.2)	9.2 (1.0)	--
Marinifilaceae	0.9 (1.4)	0.2 (0.2)	0.3 (0.3)	0.5 (0.6)	--
Burkholderiaceae	0.6 (0.8)	0.5 (0.7)	0.9 (1.3)	1.8 (2.7)	--
Desulfovibrionaceae	0.5 (0.8)	0.1 (0.1)	0.4 (0.6)	0.8 (0.8)	--
Streptococcaceae	0.4 (0.3)	0.4 (0.5)	0.3 (0.3)	0.3 (0.4)	--
Erysipelotrichaceae	0.4 (0.4)	0.5 (0.7)	0.3 (0.5)	1.6 (1.5)	<0.001 (HC-RTT) 0.0128 (DN-RTT) <0.001 (EPI-RTT)
Peptostreptococcaceae	0.3 (0.5)	0.5 (0.4)	0.4 (0.7)	0.1 (0.1)	--
Flavobacteriaceae	0.2 (0.6)	0.0 (0.1)	0.0 (0.0)	0.0 (0.0)	--

Supplementary Table 4. Distribution and significances of the co-abundant groups (CAGs) within the experimental groups healthy controls (HC), drug-naive patients (DN), epileptic patients in drug therapy (EPI = DT4, DT12 and SDT), and patients affected by Rett syndrome.

Cluster	HC	DN	EPI	RTT	P-value
Bacteroides CAG	37.1 (14.9)	43.4 (14.7)	36.2 (18.6)	58.0 (13.4)	0.003 (HC-RTT) 0.008 (EPI-RTT)
Subdoligranulum CAG	19.8 (12.0)	24.7 (14.4)	20.8 (13.2)	3.4 (6.8)	0.002 (HC-RTT) 0.003 (DN-RTT) 0.002 (EPI-RTT)
Faecalibacterium CAG	13.9 (10.8)	10.3 (5.4)	12.5 (7.7)	2.2 (1.9)	0.002 (HC-RTT) 0.003 (DN-RTT) <0.001 (EPI-RTT)
Alistipes CAG	9.0 (12.0)	5.0 (3.4)	6.6 (5.3)	10.5 (8.2)	
Other	20.2 (7.6)	16.6 (7.0)	24.0 (15.6)	25.9 (9.0)	

Supplementary Table 5. Taxonomy comparison at phylum (A) and family (B) level between healthy controls (HC), drug-naive patients (DN), epileptic patients in drug therapy after 4 months (DT4) and after 12 months (DT12). Data are expressed as relative abundance (standard deviation). P-values have been adjusted for multiple comparisons with Bonferroni correction.

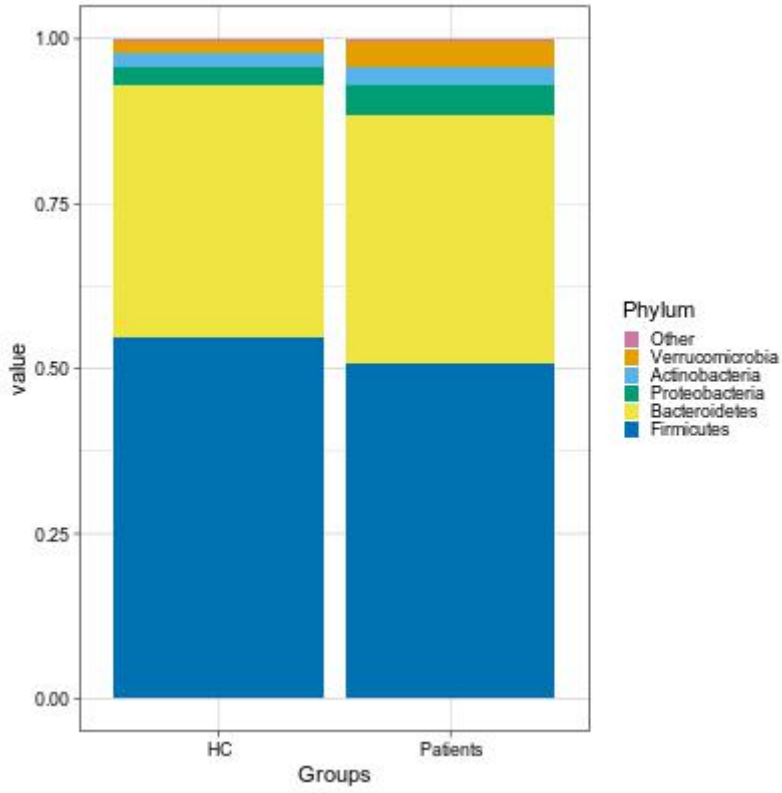
A

Phylum	HC	DN	DT4	DT12	P-value
Firmicutes	66.4 (17.7)	56.8 (13.2)	53.4 (11.9)	57.7 (18.7)	--
Bacteroidetes	27.5 (15.6)	29.0 (17.5)	42.0 (11.4)	34.8 (19.6)	--
Actinobacteria	3.1 (5.4)	3.4 (4.2)	1.2 (1.5)	2.9 (2.6)	--
Verrucomicrobia	2.1 (3.5)	8.7 (10.8)	0.5 (0.5)	3.2 (5.5)	--
Proteobacteria	0.5 (0.5)	1.9 (2.0)	2.9 (2.2)	1.3 (0.7)	0.014 (HC-DT4)
Euryarchaeota	0.2 (0.5)	0.1 (0.2)	0.0 (0.0)	0.0 (0.0)	--

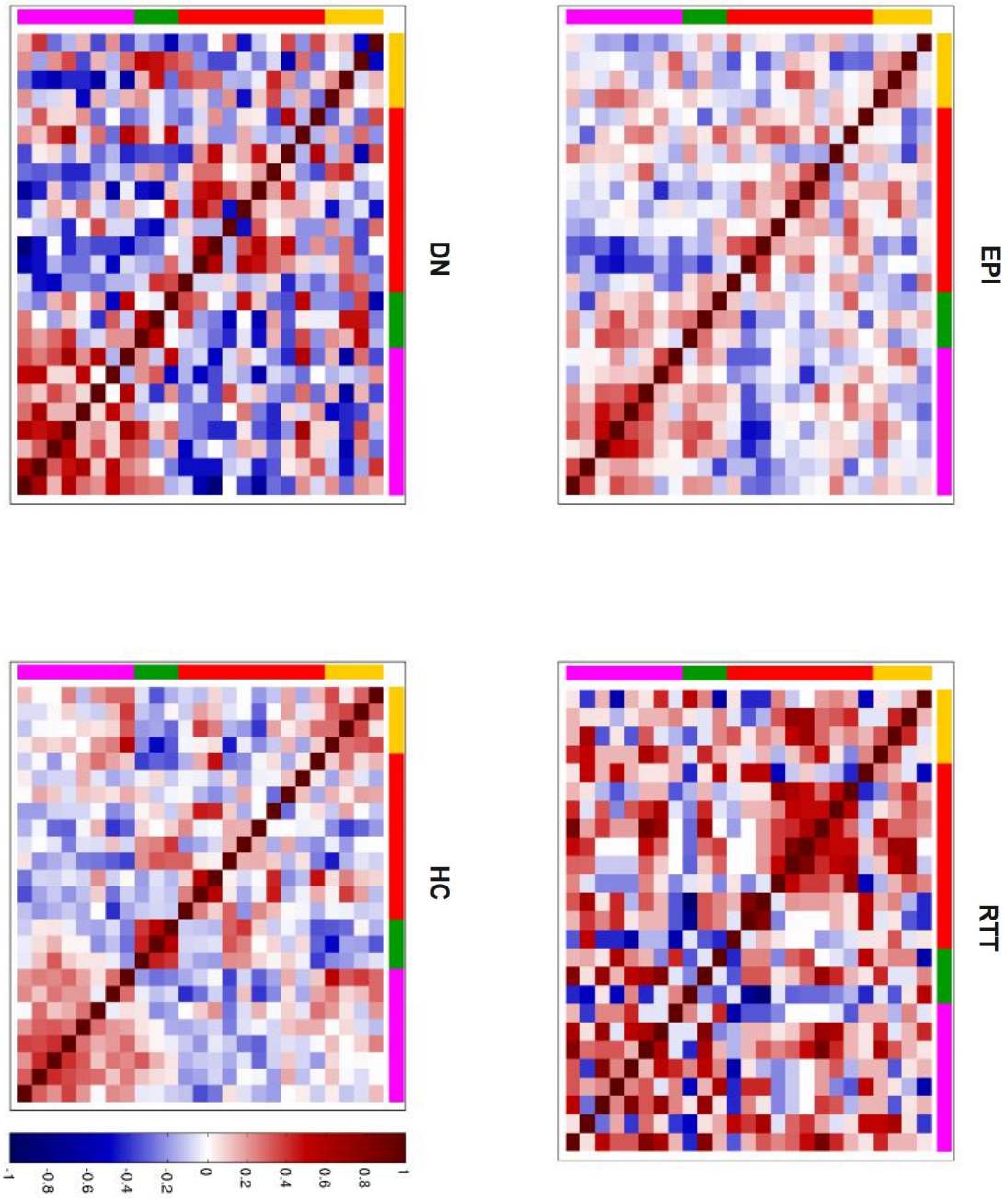
B

Family	HC	DN	DT4	DT12	P-value
Ruminococcaceae	38.7 (12.7)	30.9 (13.2)	29.9 (11.7)	36.6 (14.4)	--
Bacteroidaceae	21.5 (14.5)	22.8 (16.7)	28.1 (15.7)	23.7 (14.5)	--
Lachnospiraceae	19.1 (8.0)	14.9 (6.3)	15.9 (5.9)	11.2 (4.0)	--
Veillonellaceae	3.5 (3.2)	6.0 (5.8)	2.6 (2.3)	5.7 (4.8)	--
Bifidobacteriaceae	2.9 (5.3)	3.2 (4.0)	1.1 (1.4)	2.8 (2.5)	--
Rikenellaceae	2.9 (2.3)	2.9 (2.3)	5.2 (4.8)	3.7 (3.0)	--
Christensenellaceae	2.7 (2.8)	2.2 (2.7)	2.3 (2.5)	2.4 (2.4)	--
Akkermansiaceae	2.1 (3.5)	8.7 (10.8)	0.5 (0.5)	3.2 (5.5)	--
Marinifilaceae	0.9 (2.0)	0.2 (0.2)	0.3 (0.3)	0.2 (0.2)	--
Barnesiellaceae	0.8 (0.9)	0.9 (1.0)	0.9 (1.1)	0.5 (0.7)	--
Tannerellaceae	0.8 (0.7)	1.3 (2.2)	1.7 (2.7)	0.3 (0.5)	--
Prevotellaceae	0.7 (1.4)	0.5 (0.9)	5.4 (13.2)	6 (14.4)	--
Streptococcaceae	0.5 (0.4)	0.4 (0.5)	0.2 (0.1)	0.6 (0.5)	--
Erysipelotrichaceae	0.4 (0.3)	0.5 (0.7)	0.2 (0.2)	0.2 (0.2)	--
Acidaminococcaceae	0.4 (0.6)	1.1 (1.3)	1.7 (2.3)	0.2 (0.3)	--
Peptostreptococcaceae	0.3 (0.4)	0.2 (0.1)	0.2 (0.3)	0.4 (0.3)	--
Methanobacteriaceae	0.2 (0.5)	0.1 (0.2)	0.0 (0.0)	0.0 (0.0)	--
Enterococcaceae	0.2 (0.4)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	--
Clostridiaceae 1	0.2 (0.1)	0.1 (0.1)	0.2 (0.2)	0.3 (0.3)	--
Burkholderiaceae	0.2 (0.2)	0.5 (0.7)	1.4 (1.9)	0.4 (0.5)	--

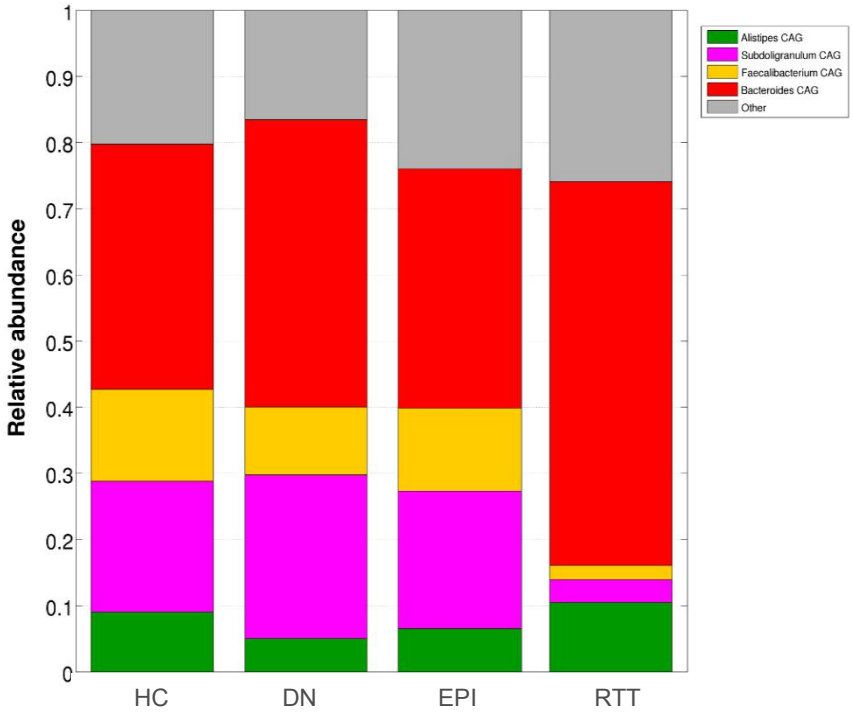
Supplementary figures



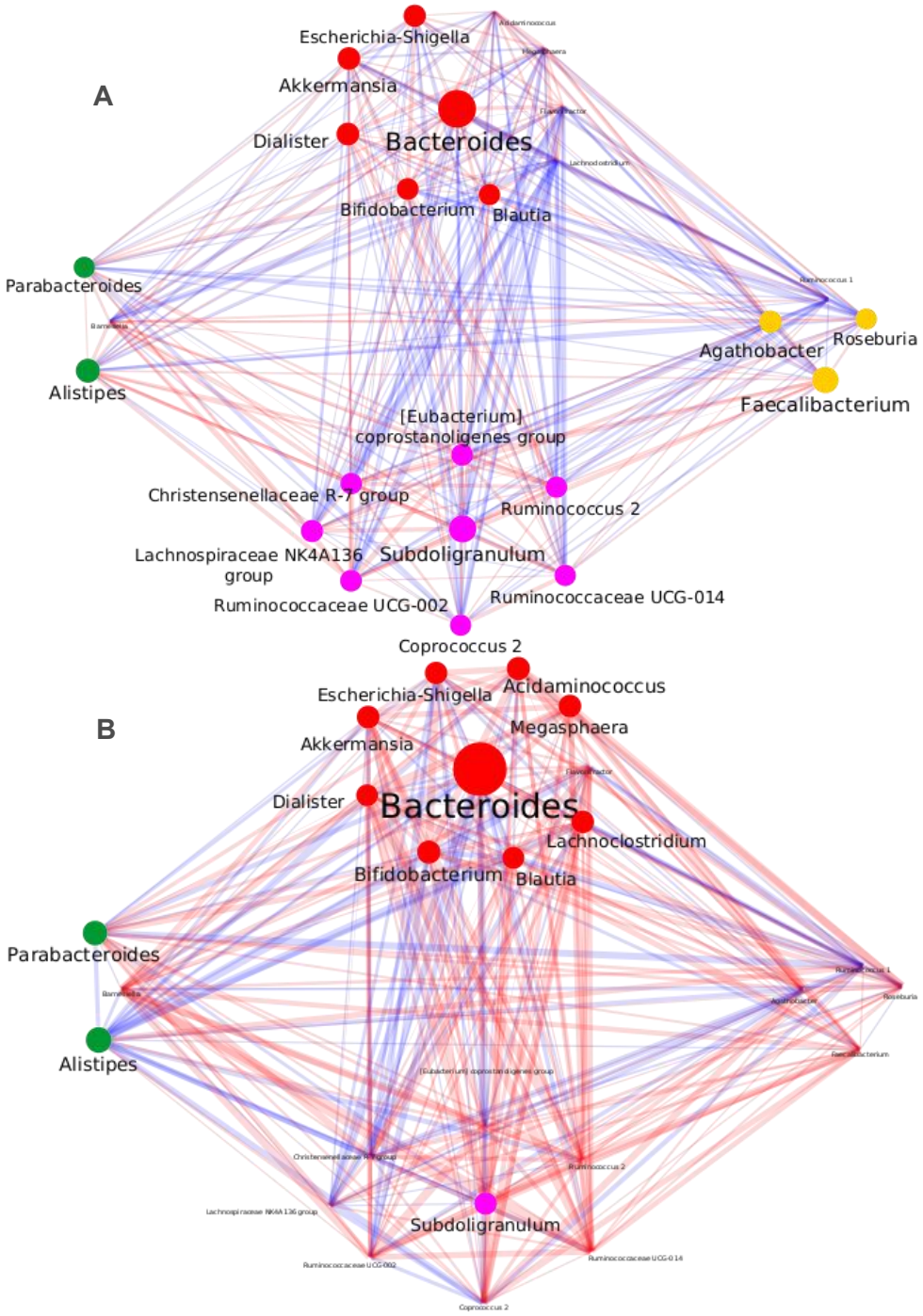
Supplementary Figure 1. Relative abundance at phylum level for the comparison between healthy controls (HC) and all the patients. The main 20 phyla, with a relative abundance higher than 1% among all the groups, are represented; those with less than the set threshold were clustered into the "Other" group. Bacterial phyla are ordered after HC abundances.



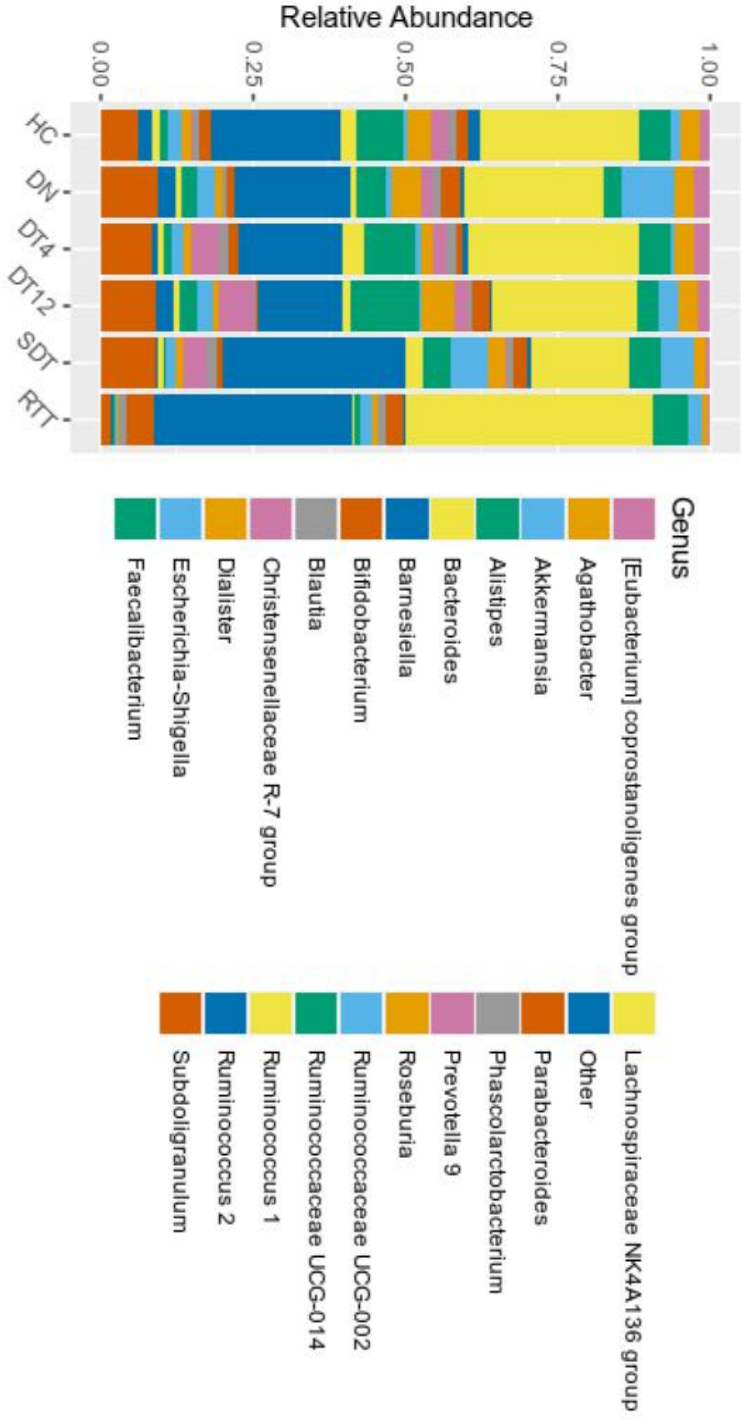
Supplementary Figure 2. Correlation matrix of co-abundance at genus level divided by experimental group. The clusters (CAG) correspond to the ones calculated from the matrix from the entire dataset. X- and Y-axis are the same, but mirrored as it is a symmetrical heatmap.



Supplementary Figure 3. Relative abundance of the Co-Abundant Groups (CAGs) obtained from Figure 11. The most abundant 20 genera have been analyzed in the co-abundance matrix and grouped in the Alistipes (green), Subdoligranulumn (pink), Faecalibacterium (yellow), and Bacteroides CAG (red); the genera not included in the matrix were clustered in the "Other" group.



Supplementary Figure 4. Correlation networks at genus level for (A) EPI and (B) RTT groups. Relative abundances of the singular genera represented by the size of the nodes and of the labels. The arches width indicate a higher or lower correlation value according to the Spearman test. The arches colors indicate a positive (red) or negative (blue) interaction. Bacterial genera are divided into CAGs deduced by the co-abundance heatmap in *Figure 10*.



Supplementary Figure 5. Relative abundance at genus level for all the experimental groups in the dataset. The main 20 genera, with a relative abundance higher than 1% among all the groups, are represented; those with less than the set threshold were clustered into the "Other" group. Bacterial genera are ordered after HC abundances.

SCIENTIFIC PRODUCTION

- Barichella M & Severgnini M, Cilia R, Cassani E, Bolliri C, Caronni S, Ferri V, Cancello R, **Ceccarani C**, et al. (2018) Unraveling Gut Microbiota in Parkinson's Disease and Atypical Parkinsonism. *Movement Disorders.* Dec 21. doi:10.1002/mds.27581

- Cremonesi P & **Ceccarani C**, Curone G, Severgnini M, Pollera C, Bronzo V, et al. (2018) Milk microbiome diversity and bacterial group prevalence in a comparison between healthy Holstein Friesian and Rendena cows. *PLoS ONE* 13 (10): e0205054. doi:10.1371/journal.pone.0205054

- Bassanini G & **Ceccarani C**, Borgo F, Severgnini M, Rovelli V, Morace G, Verduci E & Borghi E. (2019) Phenylketonuria Diet Promotes Shifts in Firmicutes Populations. *Front. Cell. Infect. Microbiol.* 9:101. doi:10.3389/fcimb.2019.00101

- **Ceccarani C**, Foschi C, Parolin C, D'Antuono A, Gaspari V, Consolandi C, Laghi L, Camboni T, Vitali B, Severgnini M & Marangoni A. (2019) Diversity of vaginal microbiome and metabolome during genital infections. *Scientific Reports* 9:14095 doi:10.1038/s41598-019-50410-x

- **Ceccarani C** & Marangoni A, Severgnini M, Camboni T, Laghi L, Gaspari V, D'Antuono A, Foschi C, Re MC and Consolandi C. (2019) Rectal Microbiota Associated With Chlamydia trachomatis and Neisseria gonorrhoeae Infections in Men Having Sex With Other Men. *Front. Cell. Infect. Microbiol*. 9:358. doi:10.3389/fcimb.2019.00358

- **Ceccarani C** & Bassanini G, Montanari C, Casiraghi MC, Ottaviano E, Morace G, Biasucci G, Paci S, Borghi B & Verduci E. (2020) Proteobacteria Overgrowth and Butyrate-Producing Taxa Depletion in the Gut Microbiota of Glycogen Storage Disease Type 1 Patients. *Metabolites* 10, 133; doi:10.3390/metabo10040133

- Marangoni A & **Ceccarani C**, Camboni T, Consolandi C, Foschi C, Salvo M, et al. (2020) Pharyngeal microbiome alterations during Neisseria gonorrhoeae infection. *PLoS ONE* 15(1):e0227985. doi:10.1371/journal.pone.0227985

- Cremonesi P, Morandi S, **Ceccarani C**, Battelli G, Castiglioni B, Cologna N, Goss A, Severgnini M, Mazzucchi M, Partel E, Tamburini A, Zanini L, Brasca M. (2020) Raw milk microbiota modifications as affected by different cleaning procedures: The Trentingrana PDO case. *Front. Microbiol.* doi:10.3389/fmicb.2020.564749

RINGRAZIAMENTI

La produzione di questa tesi non sarebbe stata possibile senza alcuni aiuti, ed intendo ringraziare coloro che me li hanno forniti.

Ringrazio la dottoressa Aglaia Vignoli e le sue specializzande Ilaria Viganò e Maria Gaia Redaelli per la disponibilità e l'impegno nell'arruolare i pazienti nelle prime fasi del progetto.

Ringrazio le mie colleghe di dottorato, le dottoresse Giulia Bassanini ed Emerenziana Ottaviano, per l'amicizia, i bei momenti, e la collaborazione durante tutto il triennio.

Ringrazio il dottor Joao B. Xavier per avermi accolta nel suo laboratorio al Memorial Sloan Kettering di New York ed avermi insegnato con pazienza diverse tecniche di analisi ed avermi aiutata a migliorare molto come bioinformatica, nonostante la pandemia da CoViD-19. Ringrazio i dottori Bradford P. Taylor, Chen Liao, Ana Djukovic, Michelle Wu, e tutti i colleghi dello Xavier Lab per avermi fatto sentire subito parte del gruppo ed avermi aiutata molto durante tutto il semestre a New York, tanto in ambito lavorativo quanto in quello personale.

Infine, il ringraziamento più grande va ai miei tutori: alla professoressa Elisa Borghi e al dottor Marco Severgnini, per l'incrollabile sostegno dimostrato da entrambi non solo per questa tesi, ma in ogni progetto, insegnandomi molto sul mondo della ricerca e non solo.