Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

SCIENZE BIOTECNOLOGICHE,
BIOCOMPUTAZIONALI, FARMACEUTICHE E
FARMACOLOGICHE

Ciclo XXXIII

Settore Concorsuale: 05/E1–BIOCHIMICA GENERALE E BIOCHIMICA CLINICA

Settore Scientifico Disciplinare: BIO/10 - BIOCHIMICA

COMPUTATIONAL RESOURCES FOR STRUCTURAL AND METAGENOMIC CHARACTERIZATION OF FUNCTIONS IN BACTERIA AND BACTERIAL COMMUNITIES.

ANTIMICROBIAL RESISTANCE, PATHWAYS FOR XENOBIOTIC DEGRADATION AND RELATIONSHIP BETWEEN GUT MICROBIOME AND AGEING

Presentata da: Teresa Tavella

Coordinatore Dottorato **Supervisore**

Prof.ssa Maria Laura BOLOGNESI **Prof. Pier Luigi MARTELLI**

Co-Supervisore

Prof.ssa Patrizia BRIGIDI

Esame finale anno 2021

ALMA MATER STUDIORUM – UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN
SCIENZE BIOTECNOLOGICHE, BIOCOMPUTAZIONALI, FARMACEUTICHE E FARMACOLOGICHE
CICLO XXXIII

Abstract

COMPUTATIONAL RESOURCES FOR STRUCTURAL AND METAGENOMIC CHARACTERIZATION OF FUNCTIONS IN BACTERIA AND BACTERIAL COMMUNITIES

ANTIMICROBIAL RESISTANCE, PATHWAYS FOR XENOBIOTIC DEGRADATION AND RELATIONSHIP BETWEEN GUT MICROBIOME AND AGEING

by Teresa TAVELLA

Prokaryotic organisms are one of the most successful forms of life, they are present in all known ecosystems. The deluge diversity of bacteria reflects their ability to colonise every environment. Also, human beings host trillions of microorganisms in their body districts, including skin, mucosae, and gut. This symbiosis is active for all other terrestrial and marine animals, as well as plants. With the term holobiont we refer, with a single word, to the systems including both the host and its symbiotic microbial species. The coevolution of bacteria within their ecological niches reflects the adaptation of both host and guest species, and it is shaped by complex interactions that are pivotal for determining the host state. Nowadays, thanks to the current sequencing technologies, Next Generation Sequencing (NGS), we have unprecedented tools for investigating the bacterial life by studying the prokaryotic genome sequences. This is feasible at large scale because of the increase of throughput and decrease of sequencing costs. NGS revolution has been sustained by the advancements in computational performance, in terms of speed, storage capacity, algorithm development and hardware costs decreasing following the Moore's Law. Bioinformaticians and computational biologists design and implement *ad hoc* tools able to analyse high-throughput data and extract valuable biological information. Metagenomics requires the integration of life and computational sciences and it is uncovering the deluge diversity of the bacterial world.

The present thesis work focuses mainly on the analysis of prokaryotic genomes under different aspects. Being supervised by two groups at the University of Bologna, the Biocomputing group and the group of Microbial Ecology of Health, I investigated three different topics: *i)* antimicrobial resistance, particularly with respect to missense point mutations involved in the resistant phenotype, *ii)* bacterial mechanisms involved in xenobiotic degradation via the computational analysis of metagenomic samples, and *iii)* the variation of the human gut microbiota through ageing, in elderly

and longevous individuals. During my work, I addressed these topics with specific computational approaches described in the following chapters.

Chapter 1 provides an introduction to the antimicrobial resistance and briefly reviews the molecular mechanisms involved and the state-of-the-art of experimental and computational approaches implemented to study this phenomenon.

Chapter 2 describes PVAR3D, an open-source database of protein variations involved in antibiotic resistance. PVAR3D provides a tool for analysing variations in the context of 3D-structure, integrating functional annotation from different sources. Major features characterising variations involved in antibiotic resistance are analysed.

Chapter 3 introduces the main ideas and techniques of metagenomics, a science that is determining a paradigm shift in the microbiological field.

Chapter 4 describes XenoPath, an open-source tool for analysing whole-genome shotgun metagenomic data (WGS), uncovering the xenobiotic degradation potential associated with the communities of bacteria identified in a given sample. XenoPath reports the functional profile of an environmental or host community.

Chapter 5 summarises the current knowledge on the human gut microbiota, with a focus on the composition of the different niches of the gastrointestinal tract, its role in the pathophysiology of the host, and its development in humans at different life stage.

Chapter 6 and 7 describe two studies on the analysis of the gut microbiota in elderly and centenarians. The first project aims to relate the health status of elderlies with visceral fat and their gut microbiota. The study involves an Italian cohort from the European project NU-AGE, analysed through 16S rRNA gene-based NGS. The second original work is devoted to the characterization of fecal samples from long-lived individuals, centenarians and semi-supercentenarians (*i.e.*, people aged 105 years or older), specifically unveiling their antibiotic resistance profile, in comparisons to younger individuals. In this case, we analysed WGS data of an Italian cohort.

Chapter 8 regards a published study on the latter cohort, investigating the xenobiotic degradation potential of metagenomic communities from the stool of centenarians. Lastly, Chapter 9 presents the overall conclusions of this thesis work.

Acknowledgements

These chapters are the work carried out in these three wonderful years spent at the University of Bologna. First of all, I would like to thank the former and the present PhD coordinators Prof. Santi Mario Spampinato and Prof. Maria Laura Bolognesi.

A huge thanks to my two supervisors for funding my scholarship, my Supervisor, Prof. Pier Luigi Martelli and my Co-Supervisor Prof. Patrizia Brigidi, for following and supporting my work. Without them nothing would have been possible!

I would like to thank Prof. Rita Casadio and all the members of the Biocomputing group for welcoming me in their lab. Many thanks to Dr. Castrense Savojardo, for his patience and clear explanations, thanks to Dr. Giulia Babbi for sharing her knowledge and wisdom and for the computational support.

A great thanks to the primordial members of the group, Dr. Giuseppe Profiti and Dr. Samuele Bovo for welcoming me from the very first day and for all their support and encouragement.

Thanks also to Dr. Giacomo Tartari and Dr. Davide Baldazzi for the support and the irreplaceable coffee breaks.

I would like to thank all the members of the Microbiology lab, which was also my home for the last years. Huge thanks to Dr. Simone Rampelli, and to Dr. Silvia Turroni who followed me during this whole period. Both patiently guiding me in the analyses and helping revising these works.

A huge thanks to Prof. Marco Candela and Dr. Elena Biagi for involving me in other studies and always welcoming me in their group meeting.

Important components of this path are my fellow colleagues, Dr. Monica Barone, my old desk fellow Phd Federica D'Amico, PhD Margherita Musella, PhD Giorgia Palladino, PhD Daniel Scicchitano. Thanks for the shared work and the joyful breaks!

Finally, I would like to thank my mother and my father, Bianca and Giuseppe, my sister Stella, and my boyfriend Vincenzo for the great support and patience. Last but not least, my whole family, my grandmothers and aunts!!

Contents

To my Family

Chapter 1

Antimicrobial resistance

1.1 Antimicrobial vs Antibiotic definition

Antimicrobial resistance (AMR) is the ability of bacteria to survive to the bactericidal effects of certain drugs. Under the broad definition of antimicrobial, here we both consider antibiotic and antimicrobial drugs. Specifically, antibiotics are molecules of natural origin synthetized by bacteria and fungi against other populations of co-resident prokaryotes (*i.e.*, penicillin), while antimicrobial drugs identify compounds of synthetic (*i.e.*, quinolones) and semi-synthetic (*i.e.*, methicillin) origin. Often and for sake of simplicity, the term "antibiotic" is used as a synonymous of antimicrobial and can include the compounds of synthetic and semi-synthetic origin (Blair et al., 2015).

1.2 Milestones in antimicrobials discoveries

Long before the discovery of the first antibacterial agent, the adoption of microbes/fungi able to produce antibiotics has been traced back to 2000 years ago in human populations, in remedies involving mouldy bread and soil (Aboelsoud, 2010). Antibiotic-producing bacteria were also hypothesized by Louis Pasteur, and in the 1890s Emmerich and Löw adopted an extract from Pseudomonas aeruginosa to threat infections. In 1907 in the laboratory of Paul Ehrlich the first screening of synthetized molecules led to the discovery of salvarsan, a compound with antimicrobial activity against the bacterium Treponema pallidum, responsible for the syphilis disease. Starting from the accidental discovery of penicillin, by Alexander Fleming in 1928, and up to the '60, we identify a period defined as the *golden age* of antibiotic discoveries, important for the finding of novel natural compounds with antimicrobial effect. In this context, an important role was played by the Oxford group, including Norman Heatley, Howard Florey, Ernst Chain and colleagues. In 1945 Dorothy Hodgkin solved the penicillin structure, identifying the characteristic beta-lactam ring. Following these findings, Selman Waksman started a systematic study of soil bacteria. In particular, he focused on Actinomycetes, leading to the discovery of streptomycin and neomycin, the former used against tuberculosis. Remarkably, 64% of known natural compounds come from Actinomycetes (Hutchings, Truman, and Wilkinson, 2019).

Antibacterial compounds have contributed to cure pneumonia and other infectious diseases, and are at the basis of modern medicine. They have extended the life expectancy, reducing mortality rate in childhood, in post-surgery and traumas, being also pivotal for chemotherapeutic treatments and transplants. Combined to the clinical practice, antibiotics are also administered in animal farming, for both therapeutic and non-therapeutic purposes, such as promoting animal growth, prophylaxis after surgery, and methaphilaxis for preventing a potential outbreak (Munita and Arias, 2016).

1.3 Increasing antimicrobial resistance

Bacteria have an extremely variable genetic material, they can successfully overtake the action of antimicrobials, evading their mechanism of action, by shifting their phenotype from susceptible to resistant. Antimicrobial resistance was first observed in clinical environments. Initially, it involved only penicillin and the discovery of methicillin in 1960 lowered the alarm on this phenomenon. Later on, the resistance towards methicillin promptly emerged, leaving us today to face methicillin-resistant *Staphylococcus aureus* (MRSA), among other threatening pathogens (Hutchings, Truman, and Wilkinson, 2019).

By the year 2050, a UK survey has estimated a high mortality rate due to antibiotic resistance, with the death of ten million people per year (O'Neill, 2016). It has been described a high correlation between antibiotic usage and increasing antibiotic resistance, and countries with higher administering and usage in farms and medicine, show higher levels of resistance (Kollef and Fraser, 2001; Christiaens, Digranes, and Baerheim, 2002; Chokshi et al., 2019). Furthermore, in our interconnected world, any effort made by countries promoting a moderate drug usage could be vain, due the ability of resistance to spread in the environment. To avoid the emergence of new types of resistance, we need to use antimicrobial in a better and systematic manner. The current medical practice should treat the infections identifying the microbes first and then administering the antibiotic best tailored to kill the infection causing bacteria. However, before the results of microbiological tests, infections are treated using the empirical medical knowledge. To keep the pace with the emergence of new resistant pathogens, there is a need to efficiently developing drugs. The production of synthetic antibiotic can be costly and requires a continuous effort: it is estimated that one out of five antibiotics tested finally passes successfully all the phases of the clinical trials.

1.3.1 Superbugs

Of great interest is the development of drugs against the ESKAPE pathogens. Bacteria under this definition are six Gram-negative pathogens responsible for nosocomial infections. Notably this are: *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* spp. These pathogens are a critical threat since they can present a multidrug resistance phenotype and an extended spectrum of resistance to β-lactamase (ESBL) and carbapenemase. Multidrug resistant pathogens, are more dangerous, with enhanced morbidity and mortality, requiring often higher antibiotic dosage, the usage of multiple compounds and an extended period of treatment in clinical care.

1.4 Recent discoveries of natural compounds

Besides the discovery of synthetic antibiotic with clinical effect, another research field regards the discovery of antimicrobial compounds from natural origin. Interestingly, unexplored environmental niches, from a bacteria taxonomic perspective, have been proved to be a resource of new natural products. To cite one unexpected source, *Staphylococcus lugdunensis* isolated from the nasal microbiota, produces lugdunin, which prevents *S. aureus* colonization of the nasal tract (Zipperer et al., 2016).

Many of the natural compounds produced by bacteria in their ecological niche are switched off in vitro. In this setting, bacteria are in a different artificial environment, where they do not receive the same stimuli, and most of the pathways are repressed. For instance, *Clostridium cellulolyticum* isolated from a decaying grass compost, when cultivated in laboratory does not produces the antibiotic closthioamide, except when an acqueous solution extracted from the same niche is added to the cultivated isolate (Lincke et al., 2010). This underlies the difficulties in identify new natural compounds, and another major problem is the limitation concerning the possibility of growing microbes in standard laboratory conditions (Hutchings, Truman, and Wilkinson, 2019).

1.5 Antimicrobial Susceptibility testing and AMR detection methods

Antimicrobial susceptibility testing and AMR detection methods are important for the identification of resistance and help clinicians in the appropriate choice of antimicrobial drugs. Rapid methods are of utmost importance since they can limit the empiric administration of antimicrobial therapy. However, many methods require overnight incubation and can need up to 72h to give a result.

Thanks to the high generation rate of bacterial populations, laboratory experiments as the antimicrobial susceptibility testing (AST) can assess in vitro the antibiotic dosage tolerated by the clones and identify the resistant phenotypes. The measure test is the Minimal Inhibitory Concentration (MIC) of an antibiotic preventing bacterial growth in vitro. The Clinical and Laboratory Standards Institute (CLSI) and the European Committee on AST (EUCAST) define standards procedures for performing these tests. Beside the quantitative methods, other qualitative assays have been developed, such as the disk diffusion and broth dilution assays. In these cases, it is possible to visually identify the growth of colonies in the medium (*i.e.*, solid agar, broth) divided in zone and characterized by different concentration of the antibiotic. Interestingly, different automated instruments can perform high-throughput screenings via qualitative assays.

An intrinsic bottleneck of the phenotypic assay, both quantitative and qualitative, is due to the growth time needed for bacteria - lag phase and log phase - and their response time to the antimicrobial drug. Importantly, cell quantification in the inoculum is another limiting factor affecting the detection of MIC (Smith and Kirby, 2018)

Detection of resistance can be also performed with rapid detection methods, based on molecular, genetic, and genomic assays. Immunochromatographic assay is a molecular diagnostic test that detects proteins putatively involved in AMR (see next section) by means of protein arrays with immobilized specific antibodies. On this basis, several lateral flow assays (LFAs) have been developed, in particular for the detection of beta-lactamases proteins. At a larger scale, matrix-assisted laser desorption ionization time-of-flight mass spectroscopy (MALDI- TOF-MS) allows a complete profiling of the proteome (Belkum et al., 2020).

Genetic methods aim at detecting the presence of specific genes and gene variants in the bacterial genome: DNA amplification (PCR) and sequencing are the basic techniques. Genotypic AMR detection methods have the advantage to be faster than phenotypic methods, since there is no necessity of growing bacteria (Bayot ML, 2020). In recent years, whole-genome shotgun sequencing (WGS) technology has become a powerful tool for the identification of resistance genes, since it enables to investigate the whole genome at the same time, instead of few genes. Genomics allows a rapid and deep characterization of isolates and Bioinformatics allows their characterization by mining specific AMR databases. Despite the technological advancement offered by AMR detection methods, these in silico characterizations of resistance need to be further characterized in vitro.

Notably, rapid test are fluorescence in situ hybridization (FISH) and microfluidics-based techniques (Yilmaz and Demiray, 2007; Choi et al., 2014).

Recently, metagenomics led to the discovery of resistance determinants from uncultured soil bacteria. In details, DNA-libraries are generated from non-cultured mixed bacterial samples and fragments are packed within lambda phage that mediates the transduction of the vectors into a susceptible strain. The deriving strains are then selected for AMR on selective media. By sequencing the surviving recombinant strains, it is possible to select sequences possibly carrying AMR determinants (Torres-Cortés et al., 2011).

1.6 Antimicrobials spectrum and cellular targets

Based on the bacterial targets, antimicrobials are classified as *broad spectrum* when they are effective against both Gram-positive and Gram-negative bacteria (*i.e.*, fluoroquinolones, tetracyclines and others), and *narrow spectrum*, when their bactericidal action is restricted to specific bacteria (*i.e.*, glycopeptides, bacitracin drugs).

Antimicrobials can also be classified on the basis of their mode of action, determined by their chemical structure and affinity with a target site. The main targets of antimicrobial resistance are cell wall, DNA, and ribosomes. In detail, these drugs can be classified as:

i. Inhibitors of cell wall synthesis. Drugs, such as penicillins, cephalosporins, bacitracin and vancomycin are inhibitors of cell wall synthesis. One of the major differences between the cells of animal eukaryotes and prokaryotes is the presence of an external cell wall. As an example, the group of penicillins, a type of beta-lactams antibiotic, characterized by a beta-lactam and thiazolidine ring, inactivates DD-transpeptidase (EC 3.4.16.4) by covalently binding a serine residue that participate to the active site of this enzyme end that is essential for crosslinking peptidoglycans of the cell wall.

ii. Inhibitors of Cell membrane function. Many peptide antibiotics interfere with the cell membrane permeability. For instance, colistin, effective against Gram-negative bacteria, is characterized by both hydrophilic and hydrophobic functional groups. Colistin works by disrupting the cell permeability and causing the cell lysis.

iii. Inhibitors of the synthesis of nucleic acids. Quinolones, metronidazole, and rifampin, are examples of inhibitors of enzymes involved in synthesis of nucleic acids, such as DNA gyrase, RNA polymerase. These compounds hamper the replication process of the cell.

iv. Protein synthesis inhibitors. Aminoglycosides, macrolides, lincosamides, streptogramins, chloramphenicol, tetracyclines inhibit protein synthesis, by targeting either the 30S or 50S subunits of the intracellular ribosomes. Sulfonamides and trimethoprim inhibit the folic acid pathway, the first by binding to the dihydropteroate synthase, while the second by inhibiting dihydrofolate reductase.

1.7 The genetic basis of antibiotic resistance

Antibacterial compounds attack core functions of the bacterial cell. Resistance is the microbial defence: different types of responses emerge within a population undergoing selective pressure and these mechanisms can lead to bacterial adaptation. There are two types of genomic changes at the basis of antimicrobial resistance: horizontal gene transfer (HTG) and genome variations, both types can result in an alteration of the gene expression.

1.7.1 Horizontal gene transfer

HGT is a mechanisms discovered in 1940s and corresponds to the acquisition of genetic material not inherited from parent to child cells. In prokaryotes this transfer of genetic material between cells can occur via conjugation, transformation, or transduction. The first mechanism requires a recipient and a donor cell, able to transfer the genetic material via a conjugation pilus. Transformations is the acquisition of exogenous genetic material from the environment, and the third method, transduction, concerns the delivery of DNA via a phage. The presence of new genetic material could be beneficial and could confers an advantageous phenotype to the beneficiary cell (Soucy, Huang, and Gogarten, 2015).

1.7.2 Variation

In absence of antibiotics, the variations (single nucleotide substitutions or small indels) occur at each cell division. Variations arising spontaneously, can be selected in a population of bacteria if they increase the fitness of the cell when exposed to stressing environment, including antibiotics (Woodford and Ellington, 2007). Conversely, if the population is not facing antibiotic pressure, the presence of variations within the population is regulated by natural selection and genetic drift. The rate of insurgence of AMR mutations is then dependent on the antibiotic dose, and even sublethal concentrations can boost the selection of resistant variants in a population. Some genes, as those encoding for the ribosomal subunit, are present in multiple copies in the bacterial genome. If by chance, the population carries a copy with a variation beneficial upon the exposure to linezolid, the population could be selected to carry more mutated copies of the gene. This is the case of some known variations in *S. pneumoniae* and *S. aureus* (Durão, Balbontín, and Gordo, 2018).

Non synonymous single nucleotide variations determining protein variants are characterized by diverse mechanisms of resistance. Residue variations on proteins can: *i)* alter the binding affinity of an antibiotic for its protein target, affecting the activity of the compound, without altering the protein function; *ii)* influence the drug uptake of the cell, when occurring in transport proteins; *iii)* influence the expression of other proteins, such as the efflux pump transporters able to extrude the antimicrobial molecule (Munita and Arias, 2016).

Of particular interest are the variations that decrease the affinity of the drug for the binding site while not affecting the catalytic activity of the enzyme (Floss and Yu, 2005). One of the best-known examples, is the resistance related to the rifampin antibiotic class. Ryfamicin, a rifampin type of antibiotic, blocks the wild-type form of the rpoB gene encoding for the β-subunit of the RNA polymerase, and therefore hampers bacterial transcription. Variations in the rpoB, does not affect the normal function of the enzyme, essential for the cell function, but drastically decreases the affinity of ryfamicin for the protein. Similar mechanisms are at the basis of the resistance to fluoroquinolones, molecule that target DNA gyrase (gyrA-gyrB) and the topoisomerase IV (parC-parE), both essential for DNA replication (Hooper, 2002). It is of upmost importance the analysis of AMR variations in the context of the 3D structure of a protein, in relation to its active and binding sites and to interactions with other molecules.

1.8 Resistance mechanisms: intrinsic vs. acquired

Bacteria can be intrinsically resistant because of the presence/absence of a specific target gene, or can become resistant after acquiring a genetic element. Organism like *Mycobacterium tuberculosis* and *Pseudomonas aeruginosa*, are resistant to several antimicrobials, making it difficult to clinically treat these bacterial infections. For instance, members of the genus *Pseudomonas*, Gram-negative bacteria, are intrinsically resistant to the triclosan targeting the Enoyl-ACP reductase (FabI), an enzyme involved in the fatty acid biosynthesis, by encoding a mutated FabI protein. Another case of intrinsic resistance relies on the composition of the outer membrane, defining the Gram-positive and negative bacteria. Few examples are the drug daptomycin, a lipopeptide, and the vancomycin, a glycopeptide antibiotic, which are not effective against Gram-negative bacteria, being unable to diffuse through their outer membrane.

Acquired resistance can be mainly categorized in *i)* mechanisms that tend to extrude the antibiotic, minimizing its concentration inside the cell, *ii)* the presence of genetic mutation in the targeted protein or new gene, *iii)* other translational modification of the antibiotic target, *iv)* inactivation of the antibiotic, by hydrolysis or other modifications.

Antimicrobial extrusion determines a reduced permeability. Gram-negative bacteria are less permeable to antibiotics because of the presence of an outer membrane, however hydrophilic compounds can diffuse through the presence of protein membranes. Study in the *Enterobacteriaceae* family, have shown how downregulating the expression of porin channels contributes to certain type of resistance, this is the case of carbapenems, and cephalosporins. Accumulation of mutations in the porin genes and in the genes that regulate their expression, has been associate with antibiotic exposure to carbapenems in *E. coli*, *Enterobacter* spp. and in *Klebsiella pneumoniae* (Blair et al., 2015). Another mechanism preventing the antibiotic to reach its target, involves the presence of efflux-pump proteins, transporting the antibiotic compound out of the cell. Efflux pumps specificity goes from being able to target a compound (tetracycline-specific pumps, Tet) to a broader type of substrate (multidrug resistance - MDR efflux pumps) and their overexpression plays a crucial role in the resistance. In the case of *Escherichia coli*, the AcrB efflux pump presents two binding pockets allowing the interaction with different type of chemical substrates.

Furthermore, resistance due to the transport of antimicrobials out of the cell, can be due to the presence of mutations emerging within the network regulating the expression of the transporters. While example of modification and protection of the target protein is the erythromycin ribosome methylase (erm) able to methylate the 16S rRNA, thus altering the binding site for the MLS antibiotics macrolides, lincosamides and streptogramins (Leclercq, 2002).

Lastly, direct modification of antibiotic can be due to hydrolysis. An example are beta-lactamases which can break the beta-lactam ring. Another type of antibiotic modification is the modification due to the addition of a chemical group to the antibiotic (*i.e.*, acyl groups, nucleotidyl, phosphate). These additions result in changing the drug affinity of the antibiotic for its binding target site. A know case are the aminoglycoside antibiotics, that can be targeted on the exposed hydroxyl and amide groups by the acetyltransferases, phosphotransferases and nucleotidyltransferase enzymes (Munita and Arias, 2016).

In chapter 7, it is described the resistome of the gut microbiota of an Italian cohort, also in relation to the mechanisms of resistance annotated from the metagenomic sequences analysed.

1.9 The Antibiotic Resistance Ontology (ARO)

Owing the necessity to comprehensively describe the complex phenomenon of antibiotic resistance and the related mechanisms, in the 2013 an ontology was generated for providing a controlled vocabulary of all the concepts (components and phenomena) involved in AMR and for describing the logical relations among these concepts by means of a graph (McArthur AG et al., 2013). The increasing volume of studies and data generated on this subject requires an rigorous and flexible approach for collecting and classifying the current biological knowledge, and for supplementing new information leveraging already defined concepts or creating new ones whenever a new molecular mechanism or antibiotic class is discovered. The ontology is an organized schema, a resource readily explorable, in general representing semantically the knowledge over a bioscientific topic, and in this specific case describing the molecular properties of the antibiotic resistance (Antezana E. et al, 2009). Ontologies can be represented as hierarchical graphs, directed (a *is_parent* node of b), and acyclic, with nodes representing the terms, and the edges representing the different types of relation connecting them. The Antibiotic Resistant Ontology, collects terms on genes, organism, resistant phenotypes, mechanisms of resistance, families of resistant genes. It formalizes these terms under a hierarchical schema controlled by rules. All the terms within the bio-ontology are associated to a unique identifier. This is crucial when querying a term in different databases, for ensuring that concepts maintain the same meaning in different resources: this ensures their cross-interoperability (Bard JB. et al., 2004). The ontology not only contains the words and the definitions, as a vocabulary, but it also describes the relationships between different terms. The ARO follows the standard format of Open Biological Ontologies (.obo) as archiving design. An example of annotation for the gyrA gene, with accession ARO_3000733, is formalized as: *is_a*, antibiotic sensitive DNA topoisomerase subunit gyrA - ARO_3003254, *targeted_by* fluoroquinolone antibiotic - ARO_0000001.

The mechanisms of antibiotic resistance introduced earlier in this chapter are formalized in seven terms within the ontology. Here, the terms are briefly reported with a description and a known case of encoded protein exhibiting the molecular mechanisms of resistance annotated:

i. antibiotic target alteration (ARO:0001001). This definition includes the presence of missense mutations, discussed earlier in this chapter, resulting in an altered gene product as well as modification of the protein target which results in antibiotic resistance. Cases are, modified antibiotic targets resulting in lower binding affinities for the antibiotic, the deactivation of repressors that result in increased expression of genes that pump out antibiotics. The mechanism of antibiotic target alteration is one of most widespread mechanisms of resistance extended to different antimicrobial compounds.

ii. antibiotic target replacement (ARO:0001002). Replacement of the antibiotic target refers to the expression of a protein performing the same function as the target, but with a lower affinity to the antibiotic; an example is an alternative dihydrofolate reductase (ARO:3003425), plasmid encoded, and less sensitive to trimethoprim (Brolund et al., 2010). Other relevant clinical cases are the methicillin resistant phenotype in *S. aureus* due to the expression of an alternative penicillin binding protein (PBP2a) encoded by mecA gene, exhibiting low affinity for some beta-lactams penicillin, some cephalosporins and carbapenems (Fuda et al., 2004).

ii. antibiotic target protection (ARO:0001003). Protection of the target proteins from antibiotic binding, resulting in antibiotic resistance. These proteins are mostly plasmid mediated, few examples are tetracycline resistance determinants ribosomal protection protein Tet(M), Tet(O) and others, quinolone resistant protein (Qnr) (Hegde et al., 2005) and fusidic acid resistant determinant (FusB and FusC/D/F) (Chen et al., 2011).

iii. antibiotic inactivation (ARO:0001004). Enzymatic modification on the antibiotic compound, by the addition of chemical moieties, via phosphorylation (aminoglycosides, chloramphenicol), acetylation (aminoglycosides, chloramphenicol, streptogramins), adenylation (aminoglycosides, lincosamides) which result in changing its chemical structure and thus inactivating the drug, or by hydrolysing bonds breaking the antimicrobial molecule. Beta-lactamases are enzymes (EC 3.5.2.6), discussed earlier, are part of this category, they act by breaking the amide bond of the four-atom beta-lactam ring (De Pascale and Wright, 2010).

iv. antibiotic efflux (ARO:0010000). Efflux pump proteins on the membrane are responsible for transporting the antibiotic out of the cell. An example is tetracycline resistance mediated by efflux pump of the major facilitator superfamily (MFS) class. This mechanism of resistance is also broad, since it can affect different types of antimicrobials, tetracyclines, fluoroquinolones, β-lactams, carbapenems and polymyxins and others (Li and Nikaido, 2009).

v. reduced permeability to antibiotic (ARO:3000244). Lower expression of porins decreases the uptake of the antibiotic which need to be located in the intracellular compartment to exert the antimicrobial effect. This mechanism is effective against hydrophilic molecules, some fluoroquinone, tetracycline and beta-lactam entering through the water channel (Delcour, 2009).

vi. resistance by absence (ARO:3003764). Deletion and silencing of genes encoding for porins, reducing the cell permeability to antibiotic. This can occur by insertion sequences (IS) within a target gene or promoter (Poirel et al., 2015).

These strategies have evolved within bacteria in order to encompass the killing purpose of antimicrobial molecules. Lack of susceptibility can be due to multiple types of mechanisms, that can also have an additive effect. The Gram-negative and Gram-positive bacteria activate different biochemical pathways to withstand to beta-lactam compounds: in Gram-negative bacteria this is achieved via the expression of β-lactamases, while the preferred resistance mechanism in Gram-positive is the target site modification acquired with mutations in the penicillin-binding proteins (PBPs).

An example of diverse resistance mechanisms evolved for the same compound regard fluoroquinolones. In this case, resistance can occur via mutations in the DNA gyrase and topoisomerase IV, it can be due to an increasing expression of the efflux pumps transporters, and also via protection of the target site thanks to the presence of the *qnr* coding gene. Another interesting case of resistance is the presence of proteins carrying out double enzymatic reactions, as is the case of the bifunctional enzyme AAC(6′)-APH(2″), with 6′-N-acetyltransferase and 2″-O-phosphotransferase activities, conferring resistance to various aminoglycoside substrates (Smith et al., 2014).

1.10 Resources to study antimicrobial resistance

Advancements in sequencing technologies and bioinformatics methods have created a new field of clinical metagenomics aiming at the identification of AMR genes. The sequenced short reads can be either annotated towards a reference database or first assembled into contiguous fragments (contigs) before annotation (see Chapter 3 for metagenomics). Reads or contigs are searched for Open Reading Frames (ORFs) for gene prediction. Among the plethora of tools available for this task the most accredited is the JGI pipeline using a combination of gene-prediction tools (GeneMark.hmm, MetaGeneAnnotator, Prodigal and FragGeneScan) aiming at reducing the false discoveries (Huntemann et al., 2016). The predicted ORFs can be aligned with BLAST (Altschul et al., 1990a) and DIAMOND (Buchfink, Xie, and Huson, 2015) algorithms against a database of resistance determinants. The choice of the reference database should be guided by the type of sample under analysis.

Many databases focus on human pathogens and thus their adoption ends up in an underestimation of the environmental resistance determinants.

There are web-tools with in-house database that can accept both short reads and contigs for the annotation. Cases are ResFinder, and Pointfinder which is also able to detect mutations for a set of pathogens (*Campylobacter*, *Escherichia coli*, *Mycobacterium tuberculosis*, *Neisseria gonorrhoeae*, *Plasmodium falciparum* and *Salmonella*. Another relevant database in this field, is CARD presenting the Resistance Gene Identifier (RGI) tool for reads annotation and the identification of mutation (Jia et al., 2017).

A strategy for the identification of remote homologous sequences is the use Hidden Markov Models. ResFams is a derived database of annotated resistant proteins with an associated profile HMM for each protein-family obtained (Gibson, Forsberg, and Dantas, 2015). ResFams can be adopted for the study of the human gut as well as the soil, and water resistome.

Machine learning based methods, aiming at identifying resistance have also being developed in the recent years. One of the earliest is Rapid Annotation using Subsystem Technology (RAST), an AdaBoost classifiers based on the k-mer substring derived from the contigs of the genomes deposited in PATRIC database (Davis et al., 2016). The classifiers are built on a binary matrix reporting the k-mers presence/absence in antibiotic susceptible/resistant strains, an information derived from antimicrobial susceptibility test and collected in the PATRIC database. Another tool is DeepArgs applying deep learning for the identification of resistance determinants in both short reads and assembled contigs (Arango-Argoty et al., 2018). These machine learning based approaches can be used for the annotation of resistance determinants, yet they are not meant to substitute the experimental AST as a diagnostic tool.

Importantly, the feasibility of sequencing at large-scale have enabled large comparative studies of human and environmental samples. These studies have been important in characterizing the sample resistomes (*i.e.*, the set of resistance determinant sequence) and in monitoring the transmission of resistance between different ecological niches, giving an unprecedent mean to inspect the spreading of AMR globally. Identifying the presence of specific resistance determinants and the rise of new mutations is pivotal for monitoring the spreading of resistance. Organizing accurately the data obtained on AMR is an extremely important task.

1.11 Summary and perspectives

Antibiotic resistance is a natural ecological event driven by evolution. Being antibiotic compounds naturally synthetized by bacteria against other cells, - independently from human activity and intensive usage of such compounds after industrialization - prokaryotes have started to early evolve mechanisms able to escape the drugs activity, as studies on pristine sites have confirmed (Dcosta et al., 2011). This is a on ongoing natural example of evolution following Darwin's principles. Antibiotic resistance was identified even in permafrost, in remote caves and in the gut of non-industrialized human population (Clemente et al., 2015a; Rampelli et al., 2015). Competing different species of microorganism, have divergently evolved the ability to produce different compounds with antimicrobial activity. On the other side, bacteria have evolved the ability to evade these drugs that impact the core machinery of the cell. Horizontal gene transfer causes a larger spreading of this ability to other cell in the same population, or even to different species. Horizontal transfer of resistant genes is possible even between different ecological niches, since it has been found a high matching profile between the resistome from environmental sources - soil, indoor spaces – and the human gut microbiota. Because of the intensive use and misuse of antibacterials in many sectors of our industrialized society - from medicine, to animal farm and horticulture - the evolutionary pressure is rising, and the resistance mechanisms are becoming widespread.

The rapid evolution of the strategies put in act by bacteria to circumvent the antimicrobial action is the one of the more threatening public issues of the 21st century, further exacerbated by the recent coronavirus outbreak, given the antibiotic use in the first stages, when the virus went undetected. The advancement in the genomic and structural biology field have contributed at pinpointing the resistance mechanism at a molecular level, this knowledge is pivotal for the medical practice. In this setting, it is important to research towards the development of new compounds. A pivotal issue is the need to archive the information, organize the resources collecting the available knowledge. Moreover, computational data analysis is required to dig out the mechanisms at the molecular level and computational tools are required to analyse new data in relation to the available information. Extensive biocuration and creation of standards is needed for facilitating investigation. This effort requires the development and update of databases and ontologies, focused on different aspects of this complex system. Interoperability is a major issue when different resources and tools have to be integrated. To this aim it is relevant to mention the effort of research groups within the ELIXIR EU infrastructure to define guidelines and tools to improve interoperability of computational resources in computational Biology.

Interestingly, many tools and database available are focusing only on the identification of resistance sequence, not reporting the rise of resistance mechanisms due to *de novo* mutations.

To date, no resource is available to analyse the AMR mechanisms in the context of the 3D structure of proteins. However, this is an important level to be investigated, in particular in the case of protein variations: as said, variations often change the 3D interaction with the drug. In order to fill this gap we developed PVAR3D a new database for the curation of sequences, structures and resistance phenotypes associated to protein variations and presented in the next chapter.

Chapter 2

PVAR3D: a bioinformatic resource on Proteins with missense Variations driving Antimicrobial Resistance and mapped on 3D structures

2.1 Abstract

Antimicrobial resistance (AMR) is an emerging issue in public health and increasing efforts are devoted to discover genomic mechanisms involved in the acquisition of AMR traits. A relevant role is played by polymorphisms promoting single residue variations (SRV) in bacterial proteins. CARD and UniProtKB databases collect different variations at the sequence level. We integrated information from these resources in PVAR3D (Proteins with missense Variations driving Antimicrobial Resistance with 3D structures) with the specific aim to contextualize variations on protein structures. Currently, PVAR3D includes 1178 variations on 241 proteins. Variations are involved in the phenotypic resistance to 113 different compounds (16 drug classes) in 97 organisms (36 genera). 3D structures are available in PDB for 56 proteins covering 552 variations, 125 comparative models have been built and cover another 409 variations. Proteins are annotated at the functional level with Gene Ontology (GO) terms and Enzyme Commission (EC) classification. PVAR3D is available at http://pvar3d.biocomp.unibo.it. It supports queries by gene name, UniProtKB or NCBI protein accession, species or genera, antibiotic name, GO term, and EC number. PVAR3D interface provides a user-friendly visualization of variations in the context of 3D structures, highlighting active sites and, when available in PDB, ligands and functional quaternary structure. In the present work we describe the resource, how to retrieve the information and the analysis we performed on the data.
Keywords: antimicrobial resistance, antibiotics, pathogens, resistome, resistant mutation, homology modelling, functional annotation, modified target-site, target-alteration.

2.2 Introduction

Antibiotic resistance due to single residue variations (SRV) occur in proteins interacting directly or indirectly with the antimicrobial compound, hampering the protein-ligand or protein-nucleic acids interactions. As introduced in chapter 1, proteins conferring resistance to some antibiotic upon SRV belong to different classes, including enzymes, transporters, efflux pumps and regulatory proteins (*i.e.*, transcription factors). The last class is peculiar since proteins does not directly interact with antibiotic compounds but they can influence the expression of protein directly involved in antibiotic influx, efflux or metabolism. This is the case of variation on MarA transcription factor that regulates the expression of an efflux pump (Woodford and Ellington, 2007). For a SRV to confer resistance and to be maintained in a population, it must be permissive and should not significantly hamper its function, resulting then non-lethal to the cell. This constraint determines different mutation probabilities associated to each position along the protein sequence. When analysed at the level of the coding DNA sequence, the nucleotide variation can be influenced by its context surrounding (*i.e.*, presence of repeated stretches of nucleotide enhances the variation probability). Furthermore, bacteria growth condition has a determinant role for the rise of novel variations. It has been suggested that missense changes can occur not only due to DNA replication error, but also as an adaptive response in non-dividing cells (Martinez and Baquero, 2000).

Thanks to NGS is now possible to analyse a community of bacteria in a high-throughput manner, without the need to isolate and culture each single microorganism in a specific medium. Organizing, analyzing and updating the data generated is essential to keep up with the knowledge of the biochemical activity of compounds also in relation to the mechanism, spreading and evolution of new resistances (Boolchandani, D'Souza, and Dantas, 2019). Multiple efforts in organizing data on antimicrobial resistance have been released as freely available resources on the web. Some of them are summarized in Table 2.1. Each one focuses on different aspects and levels of investigation. Besides general purpose databases (e.g. CARD, MEGARes (Jia et al., 2017; Lakin et al., 2017)) specific resources are devoted to a single protein families (e.g. CBMAR for beta-lactamase; Srivastava et al., 2014 or a single pathogens e.g. UCARE for Pseudomonas Genome DB, (Saha, Uttam, and Verma, 2015; Winsor et al., 2016)).

One of the first efforts in the collection and classification of antibiotic resistance sequencing data, was the ARDB database (Liu and Pop, 2009). Presently the database is not maintained, but the data are curated under the CARD database. CARD provides curated annotations on resistance gene, covering also the ones involved in the target alteration via mutations, and contains both DNA and protein sequences. Other databases are MEGARes reporting nucleic acids sequences involved in antibiotic resistance mechanisms and PATRIC a resource collecting the pathogens genome and their AMR phenotype (Wattam et al., 2017).

TABLE 2.1: **Summary of antimicrobial resistance databases.**

Database	No. entities	Antibiotic class	Availability
CARD	3057 sequences	29	Updated monthly
ResFams	177 profile HMM	18	Last updated 2015
MEGARes	8,000 sequences	27	Last update 2016
BLDB	6971 enzymes	1 (beta-lactams)	Last update 2020
PATRIC	382137 genomes	18	Last update 2020
CBMAR	2390 sequences	1	Last update 2014
UCARE	57 strains	15	Last update 2015

Information on SRV causing resistance is then sparse in different resources, also due to past experimental limitations, that allowed assaying only a small number of genes. Nowadays, analyses are performed at the genome scale and they are crucial to monitor the emergence of new mutations, responsible for resistance phenotype. A first goal of this work is therefore to collect all the available knowledge on SRVs in a single repository, called PVAR3D. It is worth noticing that none of the available resources, but Beta-Lactamase DataBase – BLDB (Naas et al., 2017) and CBMAR, integrate structural information of proteins. However, mechanisms of resistance induced by SRVs strongly relies on structural determinants, since in most cases the molecular effect reside in a different interaction between the target protein and the antibiotic or other molecules influencing the antibiotic action. For this reason, we provided whenever possible the structural context where the SRV occurs.

Here we describe PVAR3D, an open-access database reporting the 3D structures of protein with a resistant/wild type profile to antimicrobial drugs: It allows to localise SRV in protein structure, harmonizing the knowledge already available and integrating new data. The database conveys information not only on the altered sites, and active sites but also on functional and structural features of the protein, thus offering a biological view on different level of biological complexity from the variation on the protein sequences to their structure.

2.3 Materials and methods

2.3.1 Data sources

The list of proteins and relative SRVs involved in antibiotic resistance derives from merging the information comprised in UniprotKB (The UniProt Consortium, 2017), CARD (Alcock et al., 2020), and Protein Data Bank (PDB) (Berman et al., 2000).

2.3.1.1 Uniprot

We collected from UniProtKB (version 2020_04) sequences, by querying queried for bacterial protein, filtering for the mutagenesis and natural variant field in order to gather the highest number of sequences and related positions annotated in Uniprot.

2.3.1.2 CARD

We collected from CARD (release 2020_04) sequences with single and multiple resistance mutations annotated. We collected also the associated Antibiotic resistance Ontology (ARO) terms, relative to the branches 'confers_resistance_to', 'confers_resistance_to_drug'. In order to group all the SRVs of a protein and to avoid redundant information with records deriving from UniProt, the Refseq/GenBank identification code was used as sorting key. Furthermore, we manually curated a set of 226 variations from CARD showing inconsistencies on SRV position and/or wild type residue respect to the reported sequences.

2.3.2 Annotation of proteins included in PVAR3D

PVAR3D integrates different functional and structural annotations from different resources. In particular, we collected the following information:

i. Gene Ontology terms (GO) (Ashburner et al., 2000), from all the sub-ontologies molecular function, biological process and cellular component; each one links to the database QuickGo (Binns et al., 2009);

ii. Protein family domain (Pfam) (El-Gebali et al., 2019), for the presence of functional and structural domains in the protein;

iii. Metabolic pathways information as derived from link to: KEGG pathways (Kanehisa et al., 2017), Biocyc (Karp et al., 2018), Metacyc (Caspi et al., 2018) and Ecocyc (Keseler et al., 2017);

iv. Enzymatic classification through the Enzyme Commission number;

v. Protein-protein interaction, annotation through the database STRING (Szklarczyk et al., 2017), MINT (Zanzoni et al., 2002) and INTACT (Aranda et al., 2009);

vi. Information on the active sites and binding sites, obtained from UniprotKB;

vii. PubMed link for each relation antibiotic/resistance to a specific antibiotic;

viii. NCBI id (RefSeq/EMBL/GenBank/DDBJ);

ix. Link to the identification of Patric database (Wattam et al., 2017);

x. Antimicrobial resistance ontology (ARO) for the protein coming from CARD;

xi. Link to DrugBank, a chemoinformatic database (Wishart et al., 2018), Chembl (Mendez et al., 2019);

xii. 3D protein structure, as collected from PDB or modelled (see section 2.3.2) , All available structures for a protein are retained to offer an exhaustive information on possible conformational changes in different conditions.

2.3.3 Modelling 3D structure

Experimentally determined structures are present in the PDB only for 56 proteins, out of the 241 included in PVAR3D. To supplement the structural information, we adopted a comparative modelling procedure that allowed to compute the probable three-dimensional of another 125 proteins. The modelling workflow is depicted in Figure 2.1.

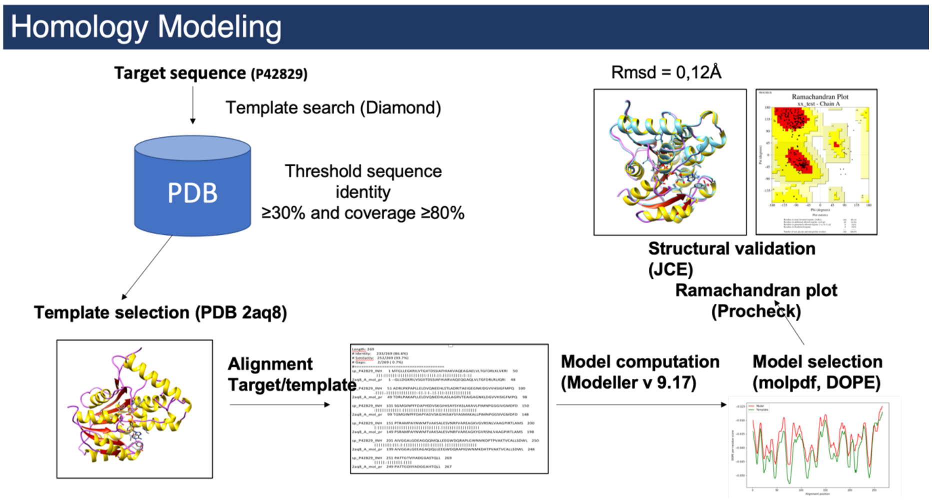


FIGURE 2.1: **Workflow of the comparative modelling procedure.**

In details, for each protein without PDB structure:

i. we searched for a suitable template in the PDB (version nov-2019), considering only structures of bacterial proteins. The Diamond software (Buchfink B et al., 2015) was adopted for database search and sequence alignment. Hits were further analysed and a template was chosen only if fulfilling the following criteria: a sequence identity $\geq 30\%$ with a coverage $\geq 80\%$. Whenever more than a template meet the conditions, the one with the lowest resolution is considered.

ii. All the models were obtained with Modeller (version 9.17) (Webb and Sali, 2016). Five models were generated for each query and the one reporting the lowest Modeller objective function score was retained.

The quality of the models were assessed in different ways:

i. structural superimposition between the model and the template 3D structure performed with JCE (Prlić et al., 2010): if the RMSD was lower than 3 Å the model was accepted (Figure 2.6).

ii. The stereochemical configurations was validated through the Ramachandran plot as computed by PROCHECK (Laskowski, MacArthur, and Thornton, 1992) (Figure 2.7).

These data can be downloaded together with the model itself. All proteins important sites retrieved from the corresponding Uniprot association where mapped to the model, exceptions are the presence of gap in the template structure.

2.3.4 Clustering of structures

Groups of similar proteins were obtained by structurally aligning pairwise the PDB-chains associated to each proteins (JCE version) and identifying cluster of proteins within 3.0 Å and z-score > 3. The cluster were manually checked for inconsistency with respect to the EC number in the same cluster and domain.

2.3.5 Protein family domain characterization of proteins in PVAR3D

The hidden Markov models of the known protein families domains (Pfam32.0, `https://pfam.xfam.org`) were used to scan the protein sequences within PVAR3D database. Hmmer version 3.2 was adopted with E-value cut-off 0.001, non-overlapping and significant Pfam domains were detected for each query protein. From this result, we mapped the mutated position falling within each identified domain.

2.3.6 Multiple sequence alignments

Each protein in the database has been aligned with the bacterial sequences retrieved from Uniprot, retaining the sequences with at least the 60% of sequence identity, and 80% of coverage, generating a graph where a node represents a single protein in the database and the edge the sequence identity. For each protein it was possible to identify a connected component, For each connected component, a multiple sequence alignment (MSA) is generated, mapping the mutation from the sequence seed on the MSA.

2.3.7 Analysis of secondary structure and solvent accessibility

The DSSP program (Kabsch and Sander, 1983) was adopted to compute secondary structure and the solvent accessibility area of protein residues. Secondary structures were grouped as follow: helix (including alpha-helices, 3-10 helices and pi-helices), Beta (β-sheet, β-bridge) and coil (β-turn, bend, loops and irregular structures). Relative solvent accessibility values were obtained by dividing the Solvent accessible area by the residue-specific maximal area as reported in (Rost and Sander, 1994) as in 2.1.

$$RSA\_ri = ACC\_ri/MaxAcc\_r \tag{2.1}$$

Where ACC is the solvent accessibility measure of the residue r in position i, calculated with DSSP, and MaxAcc (Å^2) is the accessibility value in the Sander scale. A threshold of 20% relative solvent accessibility was used to classify residues as buried or exposed.

2.3.8 Database and web server implementation

PVAR3D data are organized in a relational database implemented in PostgreSQL. The web server is built with Django Python Web framework (`https://www.djangoproject.com`) and adopts JavaScript library jQuery and Bootstrap (version 4) for generating a user friendly interface. The tables in the web pages are visualized with DataTable library enabling sorting and searching.

2.4 Results and Discussion

2.4.1 Database content

PVAR3D integrates the biological knowledge for sequences with missense residue variations and offers a framework for the visualization of the variations within the protein. It reports the protein functional annotation and protein-protein interaction via cross-references to STRING. Each protein is associated to the antibiotic(s) it confer resistance to. A summary statistic of the PVAR3D content is reported in Table 2.2.

TABLE 2.2: **Summary statistic of PVAR3D.**

	#proteins	#SRVs
Proteins with PDB structure	50	552
Proteins with modelled structure	125	409
Proteins without structure	60	217
Total	241	1178

Overall, PVAR3D includes 1178 variations on 241 proteins out of 97 organisms (36 genera). Different proteins are encoded by orthologous genes in different organisms: the total number of genes is 101. Variations are involved in the phenotypic resistance to 113 different compounds (16 drug classes). 3D structures are available in PDB for 56 proteins covering 552 variations, 125 comparative models have been built and cover another 409 variations.

2.4.2 Complexity of the relationship between organisms, drugs and genes involved in resistance upon mutation

The relationship between genes and compounds is complex. Different drugs are available for the same compound and drugs with different targets are known. The genes conferring resistance to the highest number of drugs are *parE, parC, gyrA and gyrB, nalD, nalC, folP, mexR and soxS*. Fig 2.2 depicts this complexity. The more connected drugs are fluoroquinolone, tetracycline, monobactam, phenicol, peptide, carbapenem, cephalosporin and cephamycin. The relationship between organisms and drug is also complex as depicted in Supplementary Figure S1.

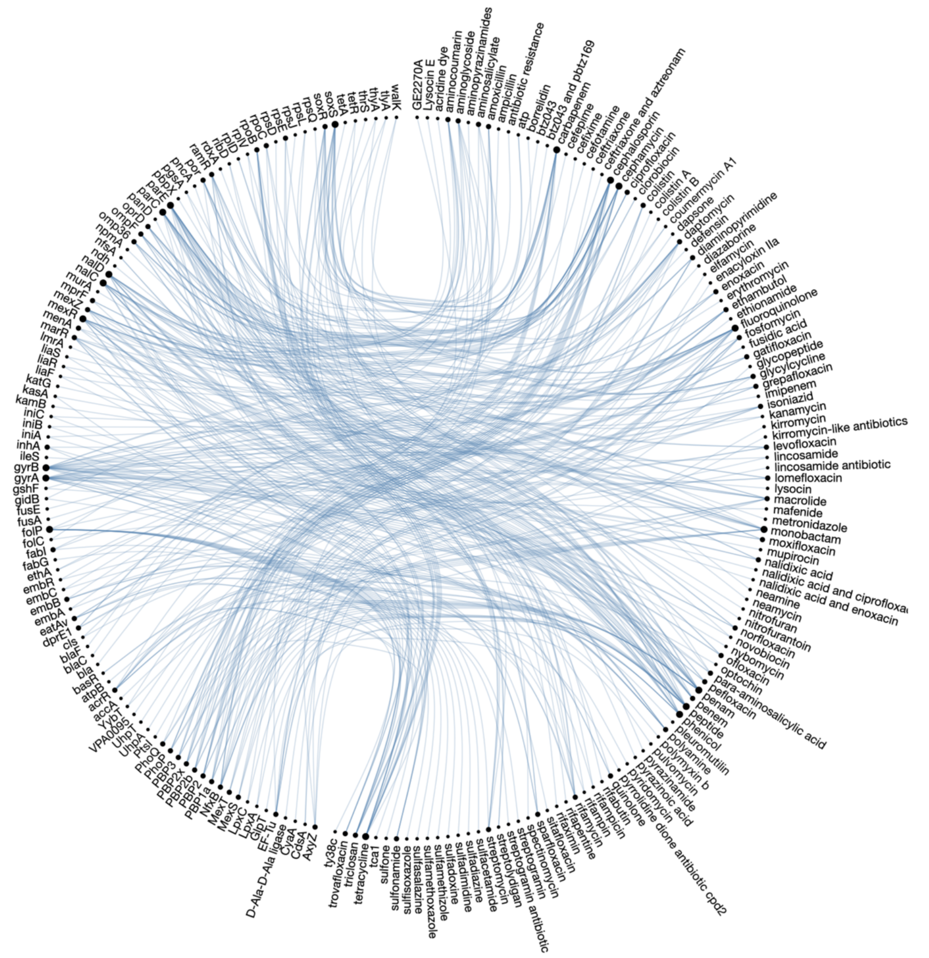


FIGURE 2.2: **Relationship between the gene name and the drug in PVAR3D.** Entities, in the gene (left) and drug (right) groups, are depicted as nodes and linked if in the database, proteins with a given gene name are annotated as resistant to the drug. The resistance is due to the presence of SRVs. The size of each node is proportional to the node degree. By means of this representation, the resistance is highlighted by drug-gene specific associations. *GyrA*, *gyrB*, *parE* and *parC* are the most connected nodes (interactive graph on the web site: http://pvar3d.biocomp.unibo.it/statistics/).

2.4.3 Distribution of the SRVs per protein and per gene

In general the large majority of proteins carries only few resistance related SRVs. When grouping different proteins with respect to the orthologous gene, the distribution of the number of SRVs per gene is showed in Figure 2.3. The number of variations per gene is highly variable ranging from 1 to 150. When analysing the distribution of variations, 84% of the proteins have less than 10 resistance related SRVs. All but two gene are associated to less than 30 variations. katG and pncA are outliers, totalizing more than 100 resistance related SRVs each. katG is a gene encoding for catalase-preoxidase (E.C. 1.11.1.21), associated to one entry in PVAR3D. Many of these mutations are associated to loss of function, thus conferring resistance by their inability to activate the prodrug isoniazid in one case and pyrazinamide in the case of the second protein (Ando et al., 2010; Lemaitre et al., 1999).

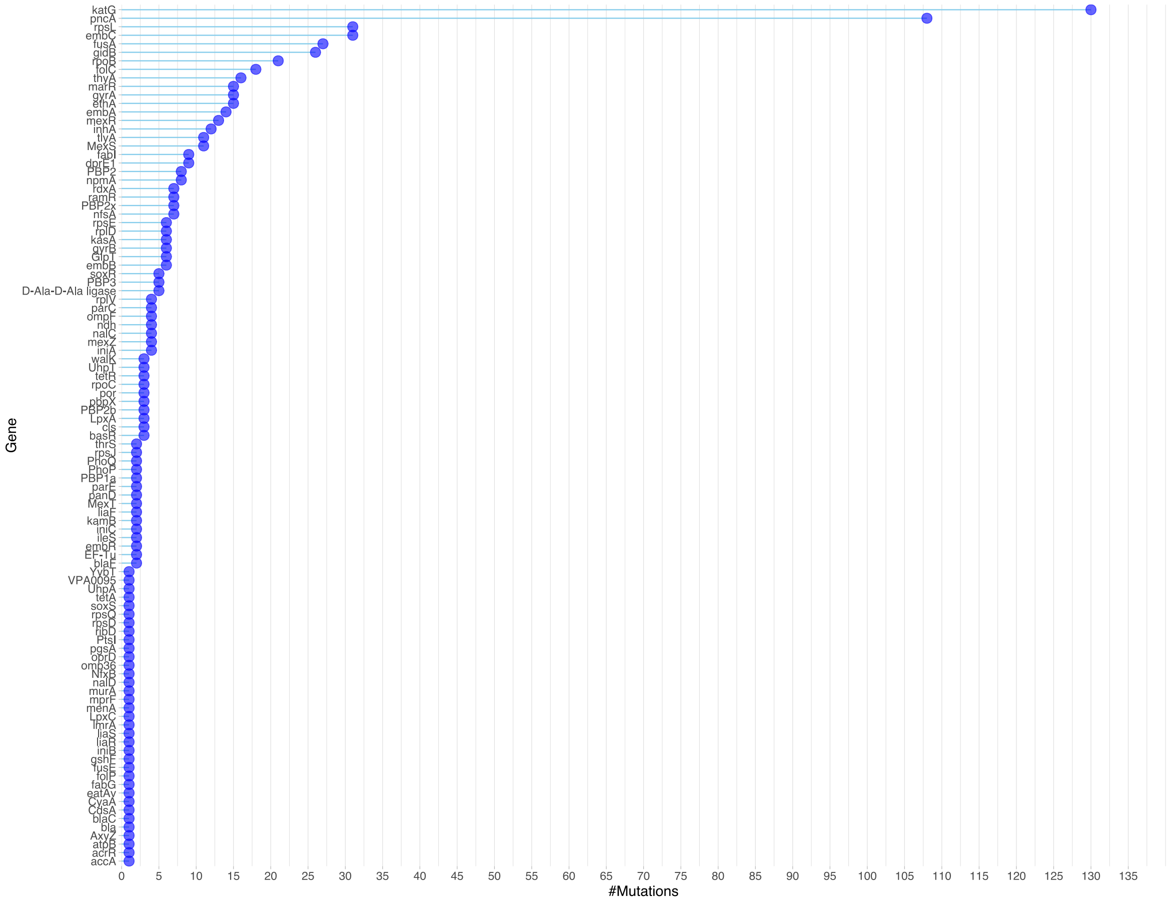


FIGURE 2.3: **Number of mutation per gene.** The plot depicts the number of mutations per protein in the PVAR3D database.

2.4.4 PDB structures

706 PDB identifiers are associated to 56 proteins (and 45 genes) in our database. 260 unique compounds co-crystalized in the PDBs and with studied antimicrobial activity are contained in the retrieved files. 65% (168) can be identified in known drugs already adopted as therapeutic, as in the case of quinolones (CPF - ciprofloxacin, GFN - gatifloxacin), beta-lactams (*i.e.*, AXL - amoxicillin, AIX - ampicillin, CB9 - carbenicillin), and antimicrobial agent (*i.e.*, 6KA, GSK625). 35% (92) of the compounds in the crystalized structure are novel lead compounds (*i.e.*, EEH, 841, 9NU, 5T0, 6G9) investigated for their antimicrobial activity.

This analysis gives a bound to intra-protein variability. We compared the structures in our dataset, selecting one representative PDB chain for each one of the 56 proteins endowed with experimental structure: in particular, the highest coverage and best resolution structure was retained. We pairwise superimposed all the representative structures and computed the corresponding RMSD. As shown in Figure 2.4 different groups of proteins can be superimposed with RMSD lower than 3 Å. 36 clusters were identified. In particular, different proteins corresponding to the same gene in different species cluster together. The structural analysis is based on the structures with lover resolution, identifying unique PDBs chains associated to the proteins in our database. Cluster of pairwise similar crystal structures are organized in 36 groups (Figure2.4 and Supplementary Table 1).

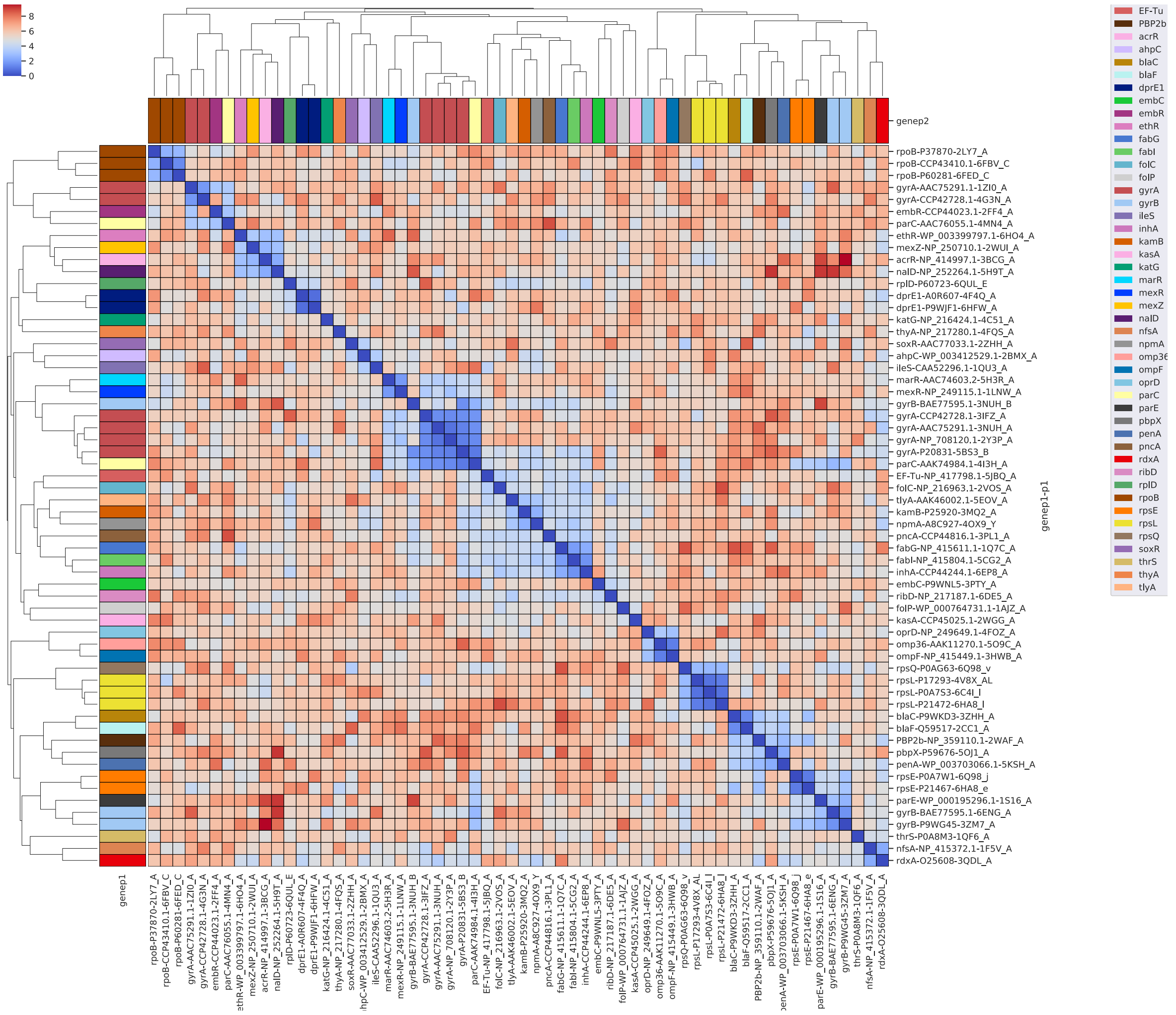


FIGURE 2.4: **Pairwise structural comparison.** Heatmap showing the similarity between structures experimentally determined and included in PVAR3D. One structure for each protein was selected. Clusters of similar structures were obtained by means of hierarchical clustering computed with average linkage method. Specular dendograms are shown in horizontal and vertical axes. Each structure is colored by protein name, reporting the legend on the right of the plot. The most populated group is the number 15 (no. proteins 5, gyrA-parC). Different domains in the same protein, crystalized in different PDB chains are organized in different clusters. Pairwise structural alignment was computed between each pair of structure, chosen from a non-redundant set of proteins and with the lower resolution.

2.4.5 Modelled structures

In total we modelled the 3D structure of 125 protein sequences. in order to assess the functional characterization we remapped the active sites, DNA binding sites and metal binding sites annotated from the template structure. The 75% of the PDBs used as template was not included in the database of proteins with experimental structures.

We built models for 125 proteins whose 3D structure is not available. They are based on 90 PDB chains. Figure 2.5 shows the sequence identity and the coverage of the alignment used for building the models. We retained only models showing less than 3 Å RMSD with respect to the template. Average RMSD is 0.89 Å(Figure 2.6). Moreover quality check performed with PROCHECK shows that modelling procedure was successful, being the rate of residues in the favoured regions higher than 80% in all but few cases (Figure 2.7)

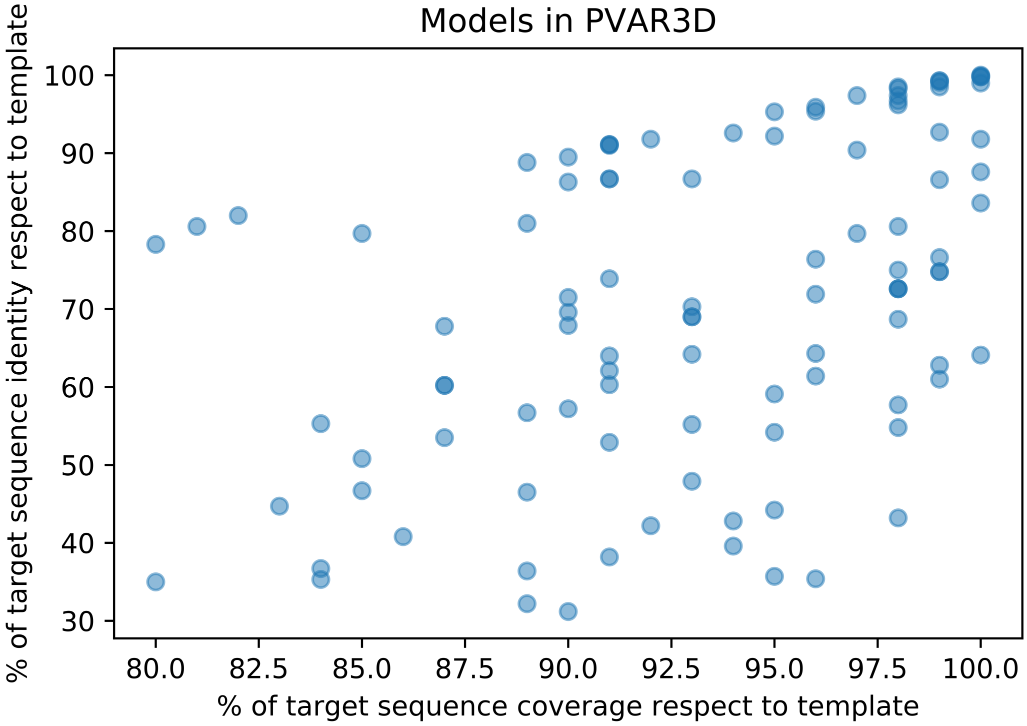


FIGURE 2.5: **Models coverage and percentage of sequence identity.** Scatterplot depicting the target percentage sequence identity and coverage respect to the template sequence for each model in the PVAR3D database.

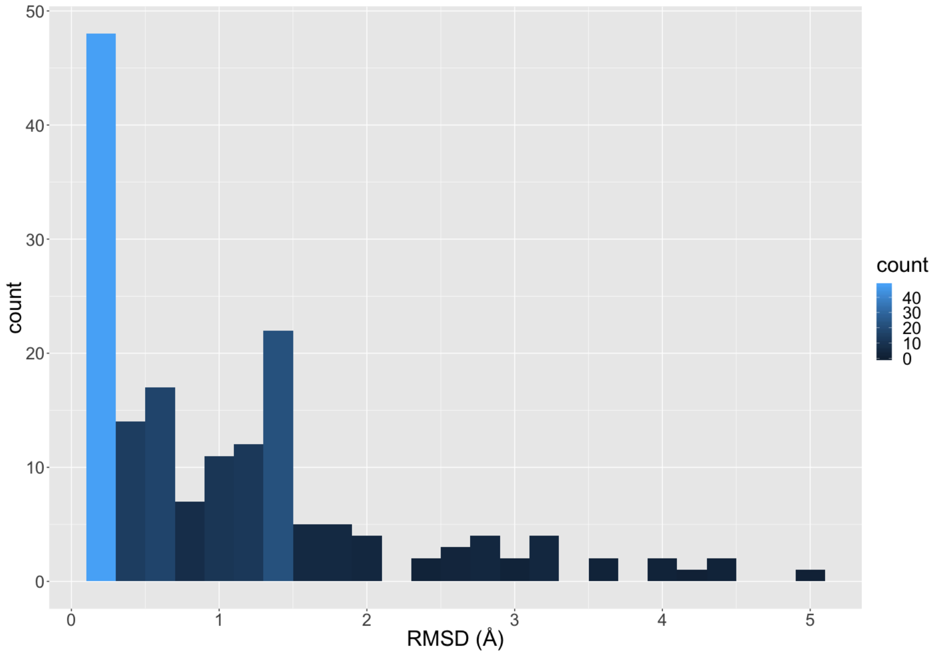


FIGURE 2.6: **Barplot showing the calculated root mean Square deviation between the model and the template backbone, calculate via structural superimposition (JCE).** The 70% of the retained models have a RMSD below 2 Å. Only RMSD below 3 Å are finally retained in the database

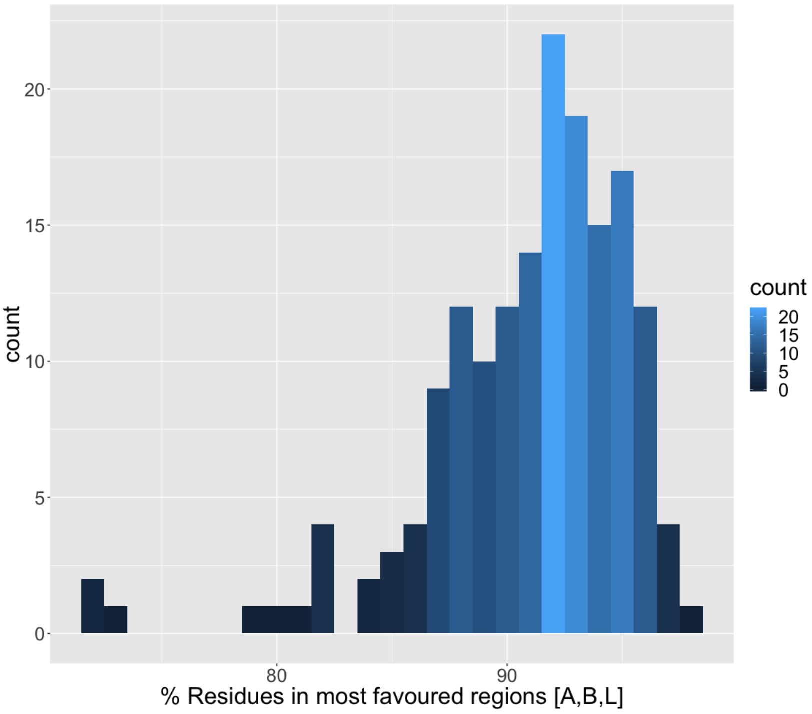


FIGURE 2.7: **Percentage of residues in the favourable regions.** Barplot showing the number of models with a given percentage of residues falling in the favourable protein secondary regions, as calculated by PROCHECK (https://www.ebi.ac.uk/thornton-srv/software/PROCHECK/) and reported in the Ramachandran plot.

2.4.6 PVAR3D native vs. mutated structural comparison

In order to investigate the structural effects of variations responsible of AR, we collected a set of six proteins for which crystallographic structure of both the wild-type and the variant forms are available in PDB (and reported in PVAR3D, Table 2.3).

TABLE 2.3: **Summary statistic of PVAR3D.**

protein	position	wild type	SRV
BAE77595.1	136	R	H
CCP42728.1	90	A	S
CCP44244.1	148	D	G
CCP44244.1	21	I	V
CCP44244.1	47	I	T
CCP44244.1	94	S	A
NP_415449.1	119	G	D
NP_415449.1	132	R	A
NP_415449.1	132	R	P
NP_415804.1	93	G	A
NP_415804.1	93	G	S
NP_415804.1	93	G	V
WP_003703066.1	501	A	T
WP_003703066.1	551	P	S

The pairwise comparisons between a native and a mutated PDB for the same protein lead to an average RMSD of 0.53 Å. We investigated the differences under several aspects:

i. Change of solvent accessibility (Table 2.4, Table 2.5): in total, 11 positions can be retrieved for this analysis (with 14 variations, *i.e.*, D115S, D115A) and mapped on 165 PDBs (of which 27 mutated and 138 non-mutated), considering in total 434 chains. None of the investigated positions is annotated as active site or binding site. The variations have been studied considering the relative solvent accessibility (Rost and Sander, 1994).

ii. Change of secondary structure: on the same dataset it is possible to assess that also secondary structure is only slightly affected by the variations. Together with the previous findings, it seems that also local variations promoted by the variation are limited.

Previous findings confirm the high complexity of the mechanisms at the basis of antibiotic resistance, that must investigated at the level of local structure variability and difference in interactions with other proteins or ligands.

TABLE 2.4: **Differences in the mean RSA of the native versus the RSA from mutated structure.**

Protein id and position	RSA native - RSA mutated	WT_POS_SRV
NP_415449.1_141	-0.51	G141D
CCP44244.1_94	-0.04	S94A
CCP44244.1_148	-0.08	D148G
CCP42728.1_90	-0.01	A90S
NP_415449.1_154	0.04	R154P, R154A
BAE77595.1_136	0.16	R136H
CCP44244.1_47	0.01	I47T
CCP44244.1_21	0.03	I21V
NP_415804.1_93	0.05	G93A, G93S, G93V
WP_003703066.1_501	0.24	A501T
WP_003703066.1_551	0.27	P551S

TABLE 2.5: **Table of changes in the RSA, considering 0.2 as the threshold for exposed residues.**

	#positions	Protein_position
WT exposed to SRV buried	2 (18.18%)	BAE77595.1_136, WP_003703066.1_501
WT buried to SRV exposed	1 (9.09%)	NP_415449.1_141
WT exposed to SRV exposed	3 (27.27%)	CCP42728.1_90, NP_415804.1_93, WP_003703066.1_551
WT buried to SRV buried	5 (45.45%)	NP_415449.1_154, CCP44244.1_94,_148,_47,_21
tot	11	

2.4.7 Functional important sites.

We evaluated if the mutations known were mapped on functionally important site for the protein and found that only in the case of catalase-peroxidase (katG – R104L) and 3-ketoacyl-acyl carrier protein reductase (fabG – Y151V) the variations fall on active sites. Notably, for the following proteins: soxR, acrR, nalD, AxyZ, mexZ, the variations can occur in DNA binding site.

2.4.8 Variation frequency

The most frequent variation types, involved in antibiotic resistance, are characterized by substitution of apolar with polar residues and vice-versa, also apolar with apolar residues, with a change in steric hindrance (Figure 2.8).

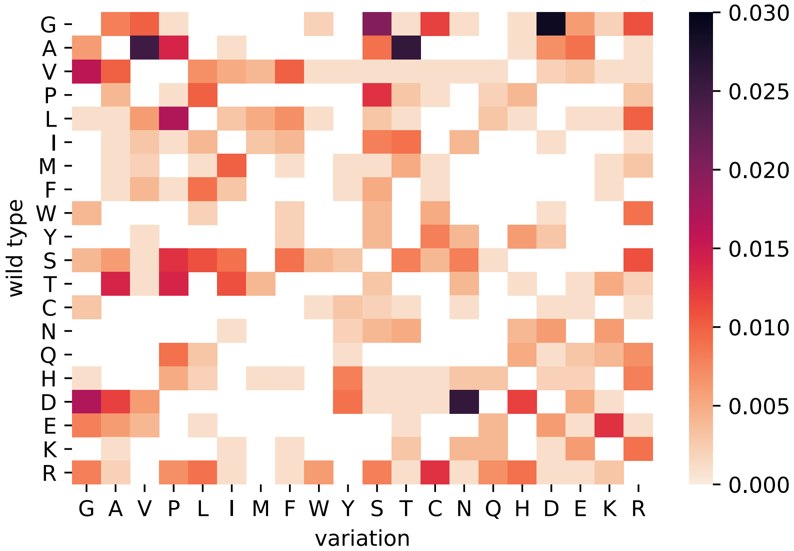


FIGURE 2.8: **Antibiotic resistant variations from PVAR3D.** The heatmap shows the frequency of each variation type in the collection on antibiotic resistant amino acid changes from the database. The residues are listed in order from the groups: apolar (G,A,V,P,L,I,M), aromatic (F,W,Y), polar (S,T,C,N,Q) and charged (D,E,K,R). Other changes are found too, apolar to charged, charged to apolar/polar.

2.4.9 Protein families

By annotating the Pfam domains within the sequences in our database we obtain a total number of 675 Pfams, and 96 Pfam with at least a mapped mutated position, which can be assigned to 1772 pdb-chain unique list. The protein family domain with an high number of mapped mutation (freq ≥ 0.01) are here reported (Supplementary Figure S2):

i. PF00141.23 PEROXIDASE (apolar to apolar, apolar to charged, apolar to polar, polar to apolar);

ii. PF00521.20 DNA topoisoIV (apolar to apolar, apolar to polar, charged to apolar, polar to apolar);

iii. PF00857.20 Isochorismatase (apolar to apolar);

iv. PF00905.22 Transpeptidase (apolar to apolar);

v. PF04565.16 RNA pol Rpb2_3 (apolar to apolar, charged to polar, polar to apolar, polar to polar);

vi. PF04602.12 Arabinose trans (apolar to apolar, apolar to charged, apolar to apolar).

2.4.10 Web interface and data visualization

PVAR3D database has a user-friendly interface and can be queried through different biological identifiers code. The web site main pages are Home, Browse, Statistics

Tutorial and Download, from each page is accessible the search bar. The database allows queries by gene symbol, Genus (obtaining all the pathogenic member at lower taxonomic levels), drug, Uniprot, NCBI, Antimicrobial Resistance Ontology identifier of the sequence, Gene Ontology terms, KEGG, pathway, Enzyme Commission number (EC) and String identifier. The query outcome is reported in a responsive table with all the entry bearing the queried term and the respective link to the single record page.

For each pathogen it is possible to retrieve the relative resistome, referring to the set of proteins involved in the mechanism of resistance to antibiotic, due to point mutation. The Home view offer a short path to the proteins for the ESKAPE pathogens via a link. The database visualization is composed of:

i. the protein entry panel with the protein identification code from CARD, through the Antibiotic Resistance Ontology (ARO), NCBI, Uniprot ID;

ii. the cross-references panel with link to other databases;

iii. the structural visualization;

iv. the mutation panel;

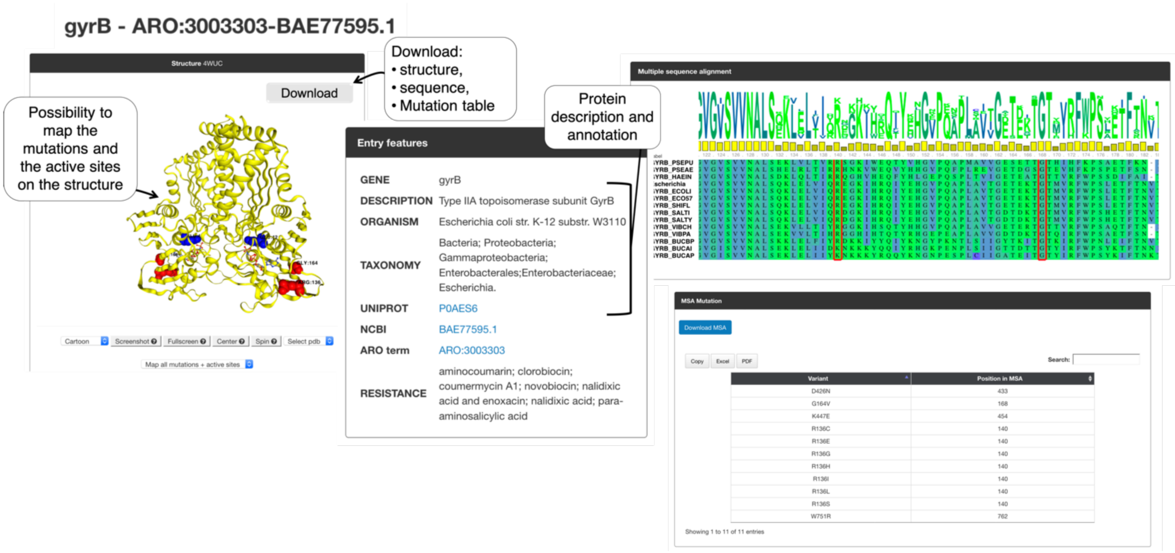
v. the sequence feature panel with mutations and their phenotype;

vi. the multiple sequence alignment panel, with the cluster of orthologous sequences.

vii. table of mutations remapped on the MSA.

The view of 3D structure and the ligand is integrated via the application web NGL Viewer (Rose et al., 2018). Mapping the variations directly on the structure (functional biological unit), is possible via the 'Mutation panel' of the entry page. The annotated active sites can be mapped on the structure, alone or together with the mutations annotated. The visualization of the primary sequence and of the mutations with relative phenotype is possible via the application Feature Viewer (Paladin et al., 2020). On the page bottom, it is visualized the multiple sequence alignment of the protein sequence in PVAR3D database and the pool of orthologous sequences, via the application MSA viewer (Yachdav et al., 2016). (Figure 2.3)

For each entry it is possible to download the related information about the resistance and the variations associated, as found in the 'Mutation panel' of the single entries, as well as the fasta file of the sequence and MSA if available. In case where the three dimensional structure was obtained via homology modelling, it is possible to download the model, the .pir file and the alignment with the respective template, the superimposition data and the PROCHECK statistics on the model. Furthermore, a fasta file of the whole database is also accessible in the Download page. Each sequence reports the name of the dataset of origin from which it was retrieved.


gyrB - ARO:3003303-BAE77595.1
Structure 4WUC
Download
Possibility to map the mutations and the active sites on the structure
Download:
• structure,
• sequence,
• Mutation table
Protein description and annotation
GLY:164
ARG:136
Cartoon
Screenshot
Fullscreen
Center
Spin
Select pdb
Map all mutations + active sites
Multiple sequence alignment
Download MSA
Copy
Excel
PDF
Search:

Entry features

GENE	gyrB
DESCRIPTION	Type IIA topoisomerase subunit GyrB
ORGANISM	Escherichia coli str. K-12 substr. W3110
TAXONOMY	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacterales;Enterobacteriaceae; Escherichia.
UNIPROT	P0AES6
NCBI	BAE77595.1
ARO term	ARO:3003303
RESISTANCE	aminocoumarin; clorobiocin; coumermycin A1; novobiocin; nalidixic acid and enoxacin; nalidixic acid; para-aminosalicylic acid

MSA Mutation

Variant	Position in MSA
D426N	433
G164V	168
K447E	454
R136C	140
R136E	140
R136G	140
R136H	140
R136I	140
R136L	140
R136S	140
W751R	762

Showing 1 to 11 of 11 entries

FIGURE 2.9: **PVAR3D page visualization.** Example of a *gyrB* belonging to *Escherichia coli*. On the left, the protein structure visualization, the protein description and the cross-references annotated. On the right the multiple sequence alignment (MSA), using as seed sequence the *gyrB*, and mapping the mutated position on the MSA (in red).

2.4.11 Application of PVAR3D

The database is freely consultable, and the data are accessible for downloading. The repository offers a comprehensive resource of variation resulting in antibiotic resistance. It can be used to design new experimental tests for resistance, and together with other bioinformatic resources, these data can be the starting point for the analysis of new antimicrobial molecules via docking and through the analysis of engineered mutations. This curated database can also be a dataset for sequencing data, mapping the resistome from environmental source or host.

2.4.12 Conclusions and future directions

Antimicrobial resistance due to target alterations is one of the major pathogenicity drivers. PVAR3D is a curated resource targeting protein coding genes and presenting the literature-based, variational landscape of antibiotic resistance. The database, visualized by an interactive and dynamic interface, provides a compact overview of the resistance phenotype, protein variants and related structural data. The information in the database can be recovered and used to characterize, *in silico*, microbial resistant phenotypes, thus narrowing further experimental testing needed to assess the drug resistance. The database will be updated regularly, also integrating the possibility for the user to research a protein sequence. PVAR3D aims to fill the gap between experimentally identified variation and the resistant phenotype, in the context of studying the physical-chemical changes induced by the variations at the protein 3D level.

Chapter 3

Metagenomics: sequencing technology and computational approaches

3.1 Bacteria shape life

Bacteria are one of the most ancient and diverse living forms, inhabiting the extreme ecological niches of this planet. In addition to be the most adaptable cell type, outperforming multicellular animals, except probably for tardigrades, bacteria are important for many aspects concerning life. Microbes are involved in the chemical cycles of the key elements of the biosphere - carbon, nitrogen, oxygen, and sulfur (Pedros-Ali, Carlos, 2006). Notably, microbes can perform bioremediation of contaminated environments from anthropic activities (Jansson and Hofmockel, 2018). Furthermore, animal metabolism relies on microbial metabolism for the uptake of nutrients, and also in plants microbes are important for the host health (Sekirov et al., 2010; Berendsen, Pieterse, and Bakker, 2012). These activities are conducted by complex communities of bacteria that by living in symbiosis, interact with the different ecosystems, creating distinct networks of plant-microbes, insect-microbes, soil-microbes, and animal/human-microbes interactions. However, in the past we have struggled to define the interactions between bacteria and their role with respect to their ecological niche. It has been estimated that we could study only 1% of microbial species with respect to the deluge diversity of bacteria, by means of classical microbiology. In this regard, a mathematical modelling approach of synthetic microbial communities can be used to predict these interactions (Medlock et al., 2018). Other experimentally based techniques regard microfluidic assays that by mimicking a micro-niche are also useful to reduce the complexity of the web of associations (Massalha et al., 2017). Despite these advancing approaches, profiling a microbial community at both taxonomical and functional level is extremely challenging, and in order to deconvolute these layers of complexities, metagenomics and other omics techniques in association with computational approaches can help in the interpretation of these interactions.

3.1.1 Studying bacteria has revolutionized molecular biology

The study of bacteria is not only important to disclose their role within an ecosystem, but more often discoveries from bacterial genomics have led to important applications in molecular biology. One of the most important advancements in this field is due to the extremophile *Thermus aquaticus*, isolated in 1969 from hot springs at Yellowstone Park. Extremophiles are bacteria able to live in unimaginable conditions, and known cases are methanogenic bacteria found in the permafrost

(Rivkina et al., 2004). Due to the evolutionary circumstances, the enzymes of this bacterium are extremely thermostable and this has revealed its DNA polymerase as a perfect candidate for dealing with recurrent cycles of DNA denaturation. In 1983 Kary B. Mullis invented the Polymerase Chain Reaction (PCR) for the amplification of DNA fragments. The protocol relied on the DNA polymerase of *E. coli* and due to the necessity of increasing the temperature for inducing DNA denaturation, the enzyme needed to be introduced at each cycle. The purification of the thermostable DNA polymerase from *T. acquaticus* in 1986, also known as Taq polymerase, greatly simplified the PCR protocol.

3.2 Microbiology milestones

One of the first attempts to understand microorganisms' physiology was made by Robert Koch who first tried to isolate and cultivate them on potato slices and on solid phase of gelatine (Blevins and Bronze, 2010). In 1876, his discovery of the anthrax bacillus posed the basis of the medical bacteriology field. Other pathogen discoveries (*Mycobacterium tuberculosis*, *Staphylococcus*, *Vibrio cholerae*) are due to his studies as well as the pathogenicity concept, and the implication of microbes in infectious disease affecting both human and animal health. With the study of cell morphology, he advanced the hypothesis of bacterial species.

Isolation of bacteria made possible the visualization of the cells, by means of a microscope. A great contribution to the microbial visualization and characterization was achieved with the implementation of staining techniques (i.e., Gram, Ziehl–Neelsen, and Schaeffer and Fulton) (Beveridge, 2001). Having noticed a difference in the number of bacteria counted from a sample and the number of clones achieved after plating and cultivation, the idea that bacteria need specific media for an optimal growth condition soon started. This observation is known as the 'Great Plate Count Anomaly' (Staley and Konopka, 1985). In this context, the scientist Sergei Winogradsky initiated his pioneering work on selective media. An important contribution was made by Robert Hungate and colleagues cultivating anaerobic microbes, from the cattle rumen ecosystem, they posed the base for the development of the microbial ecology field.

3.2.1 From morphology to genotyping

The early study of microorganisms was limited to their morphological profile, their growth rate, and their metabolism. In 1977 Carl Woese moved the concept of microorganism classification by morphology to their genotypic classification. He proposed the gene coding for the ribosomal RNA (i.e., 16S, 18S rRNA), a ubiquitous gene in the domains of life, as a marker gene for the computation of phylogenetic trees. The 16S rRNA genes are considered the molecular clock of life, due to the presence of highly conserved nucleic acid regions interspersed within 9 hypervariable regions (V1-V9). Another important reason for it being a marker gene, is the conserved function in the different domains of life, from prokaryotes to eukaryotes and archaea, being the assembly of proteins at the base of life (Woese and Fox, 1977). In particular, the 16S rRNA subunit became soon a staple gene for the classification of bacteria, representing a new classification method with respect to the morphological assay, by means of it now expanded to uncultured bacteria. Furthermore, it has become subsequently cheap for sequencing and suitable for the analysis of a bacterial community.

3.2.2 Technological breakthroughs

An important breakthrough was the DNA sequencing technique developed by Frederick Sanger in 1977, that in combination with Woese's idea of adopting rRNA as a marker gene, revolutionized microorganism classification. In particular, Sanger's protocol for determining DNA sequences was based on the chain termination method. The principal elements of the sequencing mixture were the DNA polymerase, the fragments to be sequenced, and chemically modified nucleic acids, dideoxyribonucleotides (ddNTPs). Dideoxyribonucleotides, by missing the 3′-OH group, block the formation of a phosphodiester bond between the free hydroxyl of the last nucleotide and the 5′-phosphate group of the next, generating sequencing fragments of variable lengths. The four mixtures, containing one of the four different ddNTPs (A, C, T, G) where the polymerase reaction had occurred, are then loaded in four channels of a gel matrix and chain-terminated oligonucleotides could be separated by size via gel electrophoresis. The sequenced fragments could then be read considering the gel bands for the four channels (Sanger, Nicklen, and Coulson, 1977). Later, an automated version of the Sanger machine enabled to perform the reaction in one mixture containing all the fluorescently-labelled ddNTPs. The fragments are then separated by capillary gel electrophoresis and detected by laser excitation, reading the sequencing output from a chromatogram.

Subsequent important discoveries included the aforementioned PCR boosting the efficiency of Sanger sequencing. Improved versions of Sanger sequencing can yield 10^2 sequences with length in the range of 600-900 bp, which can be useful for metagenomics studies (Logares et al., 2012). Other key technological breakthroughs were the rRNA gene cloning and sequencing techniques, fluorescent in situ hybridization (FISH), denaturing gradient gel electrophoresis (DGGE and its variant, temperature gradient gel electrophoresis - TGGE), and restriction fragment length polymorphism (RFLP) (Escobar-Zepeda, Leon, and Sanchez-Flores, 2015). However, with the introduction of high-throughput sequencing approaches and *in silico* functional characterization by sequence similarity, all the aforementioned techniques are outdated.

3.2.3 Sequencing generation

The first genome platform after Sanger was 454 Roche, representing the beginning of the second-generation technologies. The 454 machine has been slowly dismissed since the end of 2016. This technology is cheaper than Sanger and yields a higher number of sequences per run, but shorter in length. In a nutshell, this technique relies on the light emission after the incorporation of labelled pyrophosphate during the DNA polymerase reaction. When a nucleotide is integrated into the newly synthesized chain, the pyrophosphate is released and by interacting with luciferin, it generates visible light. These emissions are recorded in a pyrogram (Margulies et al., 2005). Another technology in the second-generation series is Ion torrent. It relies on a similar concept, without labelling the nucleotide, but recording the pH change each time a hydrogen ion is released by nucleotide incorporation during the chain synthesis (Rusk, 2011).

One of the most popular second-generation technologies is the Illumina platform, based on reversible terminated chemistry and adopting fluorescent nucleotides. DNA fragments are ligated to the flow cell and the fluorescence is detected by a charged coupled device (CCD) when then nucleotide is incorporate on the newly synthesized chain (Bentley et al., 2008). The synthesis can occur from one side of the complementary DNA, determining single-end reads, or both ends of the fragment, generating a couple of read fragments, paired-end reads. All the aforementioned techniques are widely used for metagenomics studies.

The third-generation sequencing technology, in particular Oxford nanopore and PacBio technologies, by reading single DNA molecules can lead to longer sequences. Importantly, they avoid the biases induced by the amplification step. In the nanopore sequencing, a DNA polymerase is allocated within a protein nanopore embedded in a synthetic membrane. Without labelling the nucleotides, the change in the electronic potential recorded due to the nucleotide incorporation is detected by a sensor chip. The second technology relies on the Single-Molecule Real-Time (SMRT) sequencing method. In brief, a zero-mode waveguide (ZMW) is able to detect the fluorescence of the tag cleaved at each nucleotide incorporation. The synthesis reaction occurs at the bottom of the ZMW, where a DNA polymerase is attached and works by reading a single molecule of DNA as a template (Niedringhaus et al., 2011). The main advantage of these new technologies is the length of the sequences obtained. Longer sequences yield to higher coverage of the genomic portion of interest, in this way it is possible to reconstruct more confidently a genome. One pitfall of these technologies is the higher error rate with respect to second-generation sequencing (Gupta, 2008).

3.2.4 Sequencing quality control

Assessing the quality of the sequences is a crucial step in the analysis. Different biases of the technologies are imputable to the way each platform detects the incorporation of the nucleotide in the newly synthesized chain. In general, it is good practice to check the output of the reads sequenced in terms of length, GC percentage, number of sequences per sample and presence of overrepresented reads, trimming low-quality bases at the ends of the reads, and filtering low-quality reads.

3.2.5 The study of microbial communities

Starting from the definition by Begon et al. (1986), microbial communities are organisms that share the same space and coexist at a given time. In 1988 Whipps and colleagues working on the plant rhizosphere, defined the composite term microbiome (Whipps, Karen, and Cooke, 1988). In such a term, *micro* stands for the community of microscopic organisms, and *biome*, living together in the same environment with defined physico-chemical properties, biotic and abiotic factors, also accounting for their activities defining the entire ecological niche. These organisms can be bacteria, but it is widely accepted by researchers that fungi, archaea, algae and protists are part of this definition, while phages, viruses, plasmids, prions and viroids are not considered living organisms. There are other views on the definition of the microbiome but the first reported is the one that continues to be vastly accepted.

In the current definition, we further distinguish between the microbiota, as the members of a community, and the metagenome as their genetic material. However, sometimes microbiome and metagenome are interchangeably used. As an example, the gut microbiome is the ensemble of the genomes of the microbiota, the living members inhabiting the intestinal tract.

The study of microbial communities can be addressed at different molecular levels via omics techniques (*i.e.*, meta-genomics, -transcriptomics, -proteomics, -lipidomics) and other types of metadata (Berg et al., 2020). Exploiting the microbiome has gained the spotlight since the advancement of sequencing technologies, in relation to the promising applications in medicine. In addition, the study of environmental microbiomes and the possibility to engineer them is a new frontier of bioremediation, food processing and safety regarding agriculture and aquaculture.

3.3 Metagenomics

Metagenomics analysis has been developed in order to identify the type of microorganisms in a community, *which type of bacteria can be identified in a sample?* Later on, with the improvement of the sequencing technology and of computational approaches, we were also able to identify which type of genes they are coding for and which type of enzymatic activity these bacteria possess. By means of sequencing technologies relying on different molecular techniques, we can address both topics, regarding taxonomy and function, by *i)* amplicon sequencing and *ii)* whole-genome shotgun (WGS) metagenomics.

3.3.1 Amplicon sequencing profiling

Often we find this technique defined as metagenomics, even though it is targeting one or few marker genes, a more precise definition would be metaprofiling. According to this approach, specific marker genes are amplified, usually 16S rRNA for prokaryotes, 18S rRNA for eukaryotes, and internal transcribed spacer (ITS) region for the specific characterization of fungi. Furthermore, instead of sequencing the full gene, for the taxonomic task it is necessary to target only informative regions. Metaprofiling is feasible with all the aforementioned sequencing techniques and platforms, and it is a means for researchers to study the taxonomy and phylogenetic profile of a sample, also feasible and cheap for longitudinal studies. The first software aimed at analysing 16S rRNA sequences were Mothur and QIIME (Schouls, Schot, and Jacobs, 2003; Caporaso et al., 2010). Since the adoption of this marker gene in phylogenetic studies by Carl Woese, the scientific community has worked in collecting and organizing these sequences, creating databases such as Greengenes, SILVA and the Ribosomal Database Project (DeSantis et al., 2006; Pruesse et al., 2007; Cole et al., 2005). These databases collect both prokaryotic and eukaryotic data.

Once assessed the reads quality, the sequences are grouped by similarity in order to identify the set of 'species' in the analysed sample. For a long time the taxonomic classification in ecology relied on the concept of Operational Taxonomic Unit (OTUs). OTUs are obtained by grouping similar sequences with the same feature up to 97% of sequence identity, which conceptually identifies the same taxon (at least the same genus), and choosing one representative sequence for each cluster. Nowadays, Amplicon Sequence Variants (ASVs) are the preferable standard unit for defining the concept of species through a marker gene. ASVs contrary to OTUs resolve taxonomy discriminating up to single nucleotide differences.

Once sequences redundancy has been reduced, the information is summarized by counting the number of each sequence artifact (OTU, ASV) found in each of the samples sequenced, producing a table of counts. Next, each of these features is going to be assigned to a taxon. For this classification task, different computational approaches have been adopted, such as BLAST (Altschul et al., 1990b), the RDP classifier, a naïve Bayesian classifier trained on genus-level oligonucleotide frequencies (Wang et al., 2007), and UCLUST algorithm (Edgar, 2010).

From the table of counts it is possible to characterize the diversity of a microbial community. The characterization of diversity is a problem that can be posed in different views, alpha diversity is defined within the community, while beta diversity between communities. The former can be simply described by the number of species found in a sample, also referred to as the species richness (number of observed species, Chao1 measure of diversity). Another way to estimate alpha diversity is by taking into account the structure of the composition, the evenness, considering how even is the estimated presence of the counted species within a sample (e.g. Shannon's index). Furthermore, alpha diversity measures can take into account phylogeny (Faith's phylogenetic diversity) (Knight et al., 2018). On the other side, beta diversity measures pairwise the similarity between communities, generating a matrix of distances. Also in this case, it can be defined by different approaches. It can be computed by taking into account the phylogeny of the community and thus the presence or absence of a taxon (*i.e.*, unweighted Unifrac) or by considering the taxon abundances in the calculation (*i.e.*, weighted Unifrac, Bray-Curtis). Ordination techniques, such as Principal Component Analysis (PCA) and Principal Coordinate Analysis (PCoA) can be adopted for visualizing beta diversity. By condensing the distance matrix into a two/three-dimensional plot, it is possible to depict individual samples and consider their proximity to identify a similar trend in the composition of the OTUs/ASVs found.

3.3.2 Whole-genome shotgun (WGS) metagenomic

In this approach, there is no need for the use of specific primers targeting specific genes. Once extracted, DNA is randomly fragmented and sequenced, which allows identifying both coding and non-coding regions of the genome and their metabolic potential. This is a powerful way to analyse microbial diversity in different types of habitat, and can uncover new species, identifying clusters of genes with specific functions, also by studying sequence evolution in a whole microbial community. In addition to providing information on all genomes within a sample, WGS data can also be specifically looked up for ribosomal genes, they can be mapped to a reference database for phylogenetic reconstruction, alone or by choosing multiple marker genes.

3.3.2.1 Assembly-based versus read-based approach

By assembly it is meant the *in silico* reconstruction of a genome, which starts from the generation of contigs from the sequenced reads. It can be obtained in the presence of a reference sequence, or can occur *de novo*, without a guiding sequence. The most used strategies rely on the De Bruijn graph method (Miller, Koren, and Sutton, 2010), which infers reads overlapping by using k-mer substrings to speed the computation. The choice of the substring length is pivotal in the construction of the assembly, and can be chosen based on different parameters (*i.e.*, the reads error rate, genome size and coverage, repetitive sequences). The assembly generation is a highly challenging task, especially in metagenomics, requiring the construction of many genomes found in the same sample. The field is blooming with tools for metagenomics data analysis, well-known software developed for this task are MEGAHIT (Li and Nikaido, 2009), metaSPADES (Nurk et al., 2017), and IDBA-UD (Peng et al., 2012). Some advantages of analysing metagenomic data with assemblies are the increased possibility of finding full genes and operons, these are even more advantageous with long reads. By grouping together the contigs, we obtain bins that can be classified both taxonomically and functionally. The binning method can be achieved either by using a reference genome, therefore identifying similar contigs in a supervised way, or by grouping together contigs based on the sequences features (*i.e.*, GC content and oligonucleotide frequencies). In alternative, there are algorithms that use a combination of both methods. Refined bins, filtered from host contaminant sequences are called Metagenome-Assembled Genomes (MAGs).

Functional annotation can be achieved by sequence homology with respect to a database collecting a wide set of functions, cases can be Pfam (El-Gebali et al., 2019) and InterPro (Mitchell et al., 2019). From the functional annotation assignment, it is possible to retrieve the pathways where the proteins are known to operate and to reconstruct metabolic pathways from metagenomic data. Known tools are MinPath (Ye and Doak, 2011) and MetaPath (Liu and Pop, 2011)(Liu B et al., 2011), using metabolic information from databases as KEGG (Kanehisa et al., 2017) and Metacyc (Caspi et al., 2018).

The assembly free-method has the advantage of being faster and less computationally expensive than the assembly approach. High-quality metagenomic reads can be looked up for taxonomic assignments by software as Centrifuge (Kim et al., 2016) and Kaiju (Menzel, Ng, and Krogh, 2016), which use a different implementation of the Burrows–Wheeler transform FM index method (Ferragina and Manzini, 2005) for matching query reads to a reference database. Also for the read-based approach, there is a plethora of tools available for the taxonomic and functional annotation of reads (Pérez-Cobas, Gomez-Valero, and Buchrieser, 2020).

3.3.3 The dark metagenome

One of the main limitations of the metagenomic annotation based on homology search is the narrow possibility to identify sequences based on the reference databases available. A large part of these reads will be unutilized because of missed hit matching. Unassigned sequences are referred to as orphans, many of which can be due to short sequences unable to find a homologous in the reference database (Sberro et al., 2019). In order to confirm a novel gene, it is possible to first take into account the secondary and tertiary structure of the translated protein. However, in order to confirm its function, more experimental procedures are requested. The unknown domain functions are estimated to outnumber the acknowledged functional domain over time (Baric et al., 2016).

3.4 Technological limitations

The advancement in new sequencing techniques and the specialization in other fields (chemical, informatic) has defined the transition between classical and modern microbiology, led by genomics techniques. NGS technologies have made a concrete contribution to microbial genomic research by defining a new paradigm shift. Notable advancements have been the simplification of the DNA library preparation, the discarded necessity of a DNA cloning vector and a bacterial host, decreasing the problem of contamination from the host organism DNA, and the ability to parallelize at high-throughput many samples.

Nevertheless, there are several limiting factors to metagenomics approaches, from the extraction of microbial DNA to the informatic pre-processing and analysis. From the data preparation point of view, there are a series of technical biases, from designing adapters and barcodes to samples handling. Despite the effort in the implementation of pipeline/software freely available, there are several limitations due to reproducibility.

One of the pitfalls of the 16S rRNA technique is the difficulty of obtaining reproducible results when sequencing different hypervariable regions within the 16S rRNA gene. Furthermore, amplicon sequencing does not provide resolution at the species level, it does not take into account the possibility of horizontal gene transfer of the 16S rRNA gene, nor the difference in the gene copy number between organisms. Another limitation occurs when sequencing samples with low-abundance genomes, considering that ribosomal genes represent only a small fraction of the total bacterial genome (Schouls, Schot, and Jacobs, 2003).

The choice of the different reference databases, filtering cut-offs and algorithms strongly influences the data analysis results. Furthermore, a classification bias can reside on relic DNA, which is extracellular DNA not part of a living microbe that can be sampled and sequenced, and is estimated to comprise up to 40% in soil. Moreover, under-sampled taxa that might play an important role in a bacterial community, are often filtered out and neglected in the analysis. In the process of generating data, computational resources and bioinformatics are the bottlenecks of the investigation, requiring dedicated infrastructure and personnel.

3.4.1 Conclusions and perspectives

Omics techniques have impacted many biological fields. Importantly, the integration of several omics techniques (*i.e.*, transcriptomics, metabolomics, proteomics) can mine different layers of biological complexity. The amount of data generated requires informatics expertise in order to answer questions related to microbiome research. Metagenomics has outdated the classical culturing techniques, marking the entrance of microbiology into the generation of big data, where astronomical science is still at the top. In brief, metagenomics approaches can picture a microbial community at a given time, by profiling the taxonomy and genomic function from the microbial DNA of the collected samples. Metagenomics science is becoming a useful tool for monitoring environments and wastewater for potential outbreaks. Importantly, mining the genomic and functional information of microbial communities associated with hosts, as well as the complex interactions between their members, can uncover the settings that define the hosts' health (Berg et al., 2020).

Chapter 4

XenoPath: tool for profiling the bacterial species with xenobiotic metabolizing enzymes from sequenced data

4.1 Abstract

In recent years, the adoption of culture independent approaches has led to a size escalation of the biological data collected, expanding the knowledge over the microbial diversity. These approaches offer new means to identify microbial bioremediation processes, known as the ability of bacteria to participate in the degradation of xenobiotic compounds, by encoding in their genome, and potentially express, xenobiotic metabolizing enzymes. In this respect, computational methods can support the expansion of the microbial genomics field, with discoveries that range from the identification of species and genes present within a microbial community. With XenoPath we propose to adopt an advantageous computational framework to screen metagenomic samples, targeting taxa and xenobiotic degradation functions, by defining the meta-phenotype of the community analysed. XenoPath has been benchmarked, by means of *in silico* metagenomics communities, and with respect to the state-of-the-art of available tools with similar aim. We reported its usability in analysing a soil metagenomic community, in a study case. Furthermore, we provided a module for the visualization of the meta-phenotype at species level. The tool can be used for an accurate identification of new microbial partners, and their functions from metagenomics data and for these reasons it represents a key opportunity for accelerating the discovery of novel bioremediation solutions. Finally, the proposed framework, facilitates the analysis of metagenomics data, by allowing the identification of untapped bacteria and their potential role in the degradation of xenobiotics.
Keywords: xenobiotics, bioremediation, metagenomics, environment, meta-phenotype

4.2 Introduction

The microbial genomics is an expanding research field, aiming to characterize microbial communities within a niche, organism or environment, for the identification of key components and functions in relation to the changes investigated. The identification of bacterial species within contaminated environments, where the bacterial activity is responsible of a variety of enzymatic functions with biodegradation potential, is of particular interest in the microbial ecology area (Haiser and

Turnbaugh, 2013; Koppel, Rekdal, and Balskus, 2017). In this context, xenobiotics are compounds not normally produced or found in an organism or in an environment. These compounds are outsider to the enzymatic system of a determined niche and potentially harmful when accumulating.

Remediation mediated by biological entities, can transform hazardous to less harmful compound, using natural biological activity, and thus reverting the accumulation of these compound in the environment. Nevertheless, the activity of degradative microbes, is subjected to a series of factors (aerobic or anaerobic conditions, pH, substrate as electron donor or acceptors, temperatures and the presence of other inhibitory compounds), that together can determine favourable or unfavourable physicochemical conditions. For instance, the degradation of the benzenoid aromatic ring compounds are described by different pathways based on the presence/absence of oxygen. Also the aerobic and anaerobic metabolism differs for the BTEX compounds (benzene, toluene, ethylbenzene and xylenes). In other cases, some reaction have been identified in aerobic bacteria, but could hypothetically be carried out also otherwise (Ellis and Wackett, 2012).

For many reactions we still lack the metabolic information. The key to the biodecontamination process is the presence of microbes harbouring specific enzymatic activities, although in some cases, new intermediate compounds are generated by spontaneous chemical reaction. The ability to degrade certain compounds in bacteria could be already encoded in their genome, or as in case of molecules of anthropogenic nature, as atrazine, it is most probable new enzymatic function have evolved over relatively recent time under the selective pressure of the pesticide (Russell et al., 2011). The process of adaptation makes bacteria initially not able to degrade a compound to later acquire or to alter enzymes with novel metabolic activities. This can occur mostly by acquisition of mutations or new genetic material. Interestingly, these enzymes have been extensively found in plasmids (Van Der Meer et al., 1992).

There is a knowledge gap, between all the chemical compounds that can be synthesized and the known degradative ability of the microbial diversity. Despite the increasing need of finding novel bioremediation solution and contextually to identify the chemistry of biodegradation, these is a shortcoming of computational approaches. We propose XenoPath as a tool for mining the metagenomes and leading to the characterization of sample's molecular signatures, with the possibility of enzymes and species identification.

4.3 Material and methods

XenoPath is a free and opensource pipeline, written in bash and python 3, tailored for shotgun metagenomics data type. The software accepts pre-processed reads both for single, paired-end type or a combination of both. The .fastq/.fna can be in compressed format and the software can processes multiple samples. XenoPath is composed of three modules, in the first, the software identifies the taxonomy for each read and in the second retrieves the enzymatic function associated, enriching the xenobiotic degradation pathways (Table 4.1) present in the XenoPath database. Finally, the visualization module allows to generate the maps of pathways enriched in the dataset, also summarizing the information regarding the EC number associated to a given species.

TABLE 4.1: **List of xenobiotics and pathways name included in XenoPath**

Xenobiotics and pathways name
Food conservatives and antiseptic (*i.e.*, benzoate, aminobenzoate and fluorobenzoate producing carcinogenic compounds)
Pharmaceutical additives and solvents (*i.e.*, chloroalkane and chloroalkene, toluene also found in petroleum)
Insecticides (*i.e.*, chlorocyclohexane and chlorobenzene)
Herbicides (*i.e.*, atrazine)
Pesticides (*i.e.*, nitrotoluene also in plastic, pharmaceuticals and explosives, dioxin)
Plastics, paints additives (*i.e.*, stryrene, ethylbenzene, caprolactam, bisphenol, dioxin)
Refrigerants, insulation (dioxin)
Insecticides (naphthalene)
Combustion products (*i.e.*, polycyclic aromatic hydrocarbons, anthracene)
Product of fermentations (*i.e.*, furfural)

The bacterial classification task is obtained by means of Kaiju algorithm (Menzel, Ng, and Krogh, 2016), with the advantage of choosing one of the database provided by software. The software parameters can be tailored based on a configuration file. As result the tool gives table by Phylum/Family/Genus and Species level with the total number of reads counted per taxon. Diamond is adopted for the local alignment of the Open Reading Frames translated from the reads, with respect to the database of degradation enzyme constructed from (Swissprot release 2020-06), filtering for reads profiled with at least the 40% of sequence identity and the 80% query coverage. An enzymatic function is not exclusively associated to a single degradation pathway, for this reason the tool adopts the minimum pathways algorithm from MinPath tool (Ye and Doak, 2011) assigning the EC numbers to an enriched pathway of the KEGG database (Kanehisa et al., 2017). The final tables can be filtered setting a minimum threshold.

XenoPath visualization module, relies on biojs-kegg, and is tailored to annotate all the degradation pathways as shown in KEGG. For each of the pathway identified in the analysed community, it is generated an HTML file visually summarizing the EC number stratified by the species found in the dataset.

The metabolic network, based on KEGG, gives a species-specific annotation per mapped enzyme. The enzymes are highlighted by species in the pathway, but the information is also summarized at the level of a table and other graphs. Furthermore, a list of species co-occurring in the same pathways is given in output.

4.4 Performance evaluation

For the bacterial community (Table 4.3) we randomly generated 10 million *in silico* synthesized reads randomly for the mock community tested, setting reads length of 100 bases (BBmap - sourceforge.net/projects/bbmap). The annotation of each reads produced in silico was obtained from the organism respective gff file. We evaluated the functional annotation of the mock community, thus the correct number of entries in the community annotated with their correct name (TP,TN,FP,FN). We evaluated the accuracy of the method by calculating the accuracy (F1) with the formula:

1. Recall or Sensitivity=TP/(TP+FN)
2. Precision=TP/(TP+FP)
3. F1 = 2*(precision*recall)/(precision + recall)

4.5 Results and Discussion

An in silico metagenomic community was created, retrieving the genome of 22 organisms from RefSeq and randomly selecting sequences fragments of 100 bases, generating 10 million reads. The ability of XenoPath to correctly identify a read, both at the taxonomic and functional level, was retrospectively considered, taking into account the correct number of entries correctly annotated (True positive - TP, True Negative - TN, False Positive - FP, False Negative - FN). We evaluated the accuracy of the method by calculating the accuracy (F1) calculated as described in the methods section. The tool was further benchmarked against HUMAnN 2.0. The benchmarking of the taxonomy classification task leads to an higher degree of accuracy in the case of XenoPath, in particular XenoPath correctly identified 12/22 species with F1 ≥ 0.5, on the other hand HUMAnN only 8/22 species with F1 ≥ 0.5 (Figure 4.1). At the genus level, XenoPath identified 9/17 genera with F1 ≥ 0.5 and 5 of the remaining with F1 ≥ 0.25. On the other side, HUMAnN classified 6/17 genera with F1 ≥ 0.5 and 1 of the remaining with F1 ≥ 0.25 (Figure 4.1 a). Regarding the functional annotation of the reads, HUMAnN 2.0 identified only 10479 reads (55% of the total reads in the mock community known to be in a xenobiotic pathway), while XenoPath identified 16459 reads (87% of the total reads). Notably, XenoPath correctly detected 39 EC numbers against the 29 identified by HUMAnN 2.0 (Figure 4.1 b).

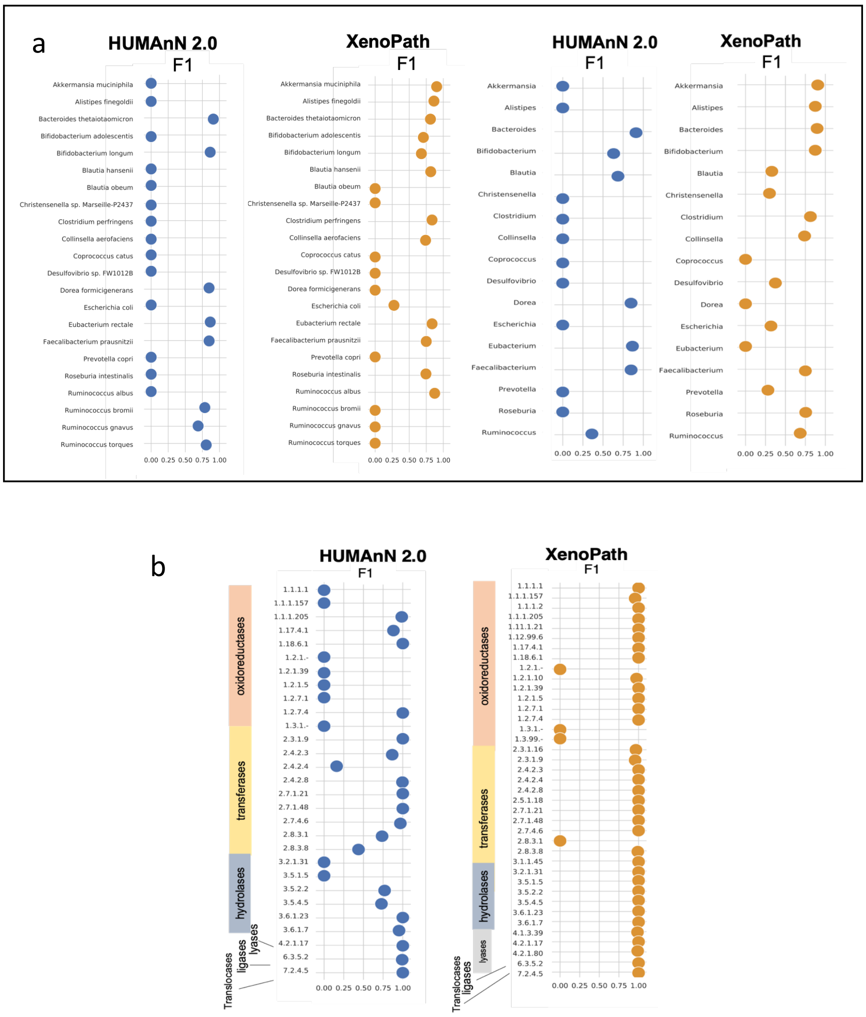


FIGURE 4.1: **Benchmarking of XenoPath at the taxonomic and functional level.**
a. Accuracy of classification at species and genus level reported as F1, for each member of the prototypical microbial community. b. Performance evaluation of XenoPath and HUMAnN 2.0 at the functional level, reporting the F1 value calculated for each enzyme commission number associated to the reads of the in silico metagenomic community.

4.6 Study case

In order to test and to describe the potential of XenoPath, we identified a recent metagenomic dataset from Weigold et al. (2016), specifically studying soil (de)halogenation potential of a German forest site. Halogenated compounds are naturally produced compounds, by some plants, macroalgae and wood root fungi. These data were evaluated looking at a broader spectrum of xenobiotic degradation potential in the three soil layers (Of 0-1 cm, Ah 1-15 cm and II.P 15-40 cm of depth). The sequences were retrieved from MGRast (ID number 11442), already filtered for adapters and eventual contaminant DNA. Paired and unpaired reads were stored in different files and prepared to be analysed by XenoPath pipeline. The first task was to check the taxonomic composition of the dataset, looking for similar microbial compositional patterns to what obtained in the aforementioned study. Protobacteria, as already revealed by the original work, are the predominant Phylum in the soil niche, followed by Actinobacteria and Acidobacteria. Firmicutes, are less abundant respect to a human gut sample.

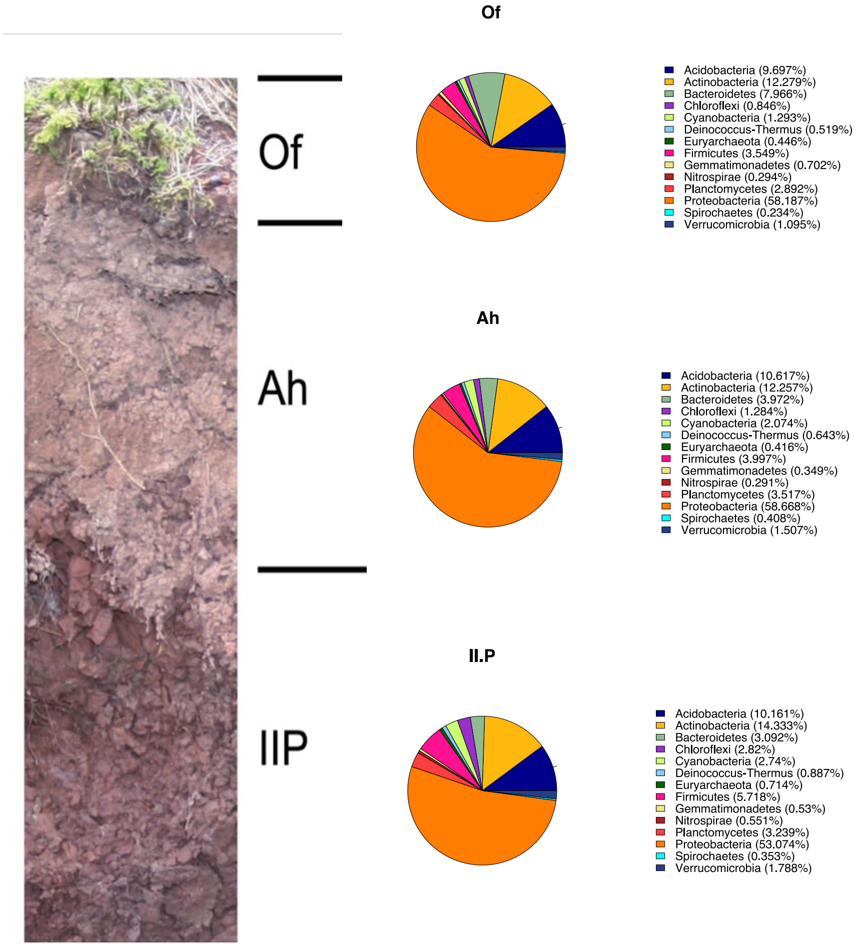


FIGURE 4.2: **Taxonomic annotation by soil layers (Of 0-1 cm, Ah 1-15 cm and II.P 15-40 cm of depth).**
Soil stratification image taken from the work of Weigold et al. (2016), while the pie chart data were obtained from Kaiju results

As in the work of Weigold et al. (2016), haloalkane dehalogenase enzyme (EC: 3.8.1.5 - dhaA) was identified to be highly abundant in the two more profound soil horizons. The second most abundant function identified by XenoPath corresponds to the enzyme 1.2.7.1 participating in the nitrotoluene degradation pathways, while the third most relevant function in the deepest soil layer is acetophenone carboxylase (EC: 6.4.1.8) annotated for ethylbenzene degradation.

4.7 Conclusions

XenoPath was implemented with the aim of enhancing our knowledge over bacteria metabolism related to a set of compounds (*i.e.*, pesticides, insecticides), also seeking promising bioremediation applications. In light of the promising results achieved benchmarking the tool with the method available in literature, for which we aim to further test the tool with more randomly generated mock communities, this framework offers new opportunities in identifying bioremediation microbial partners. XenoPath allows to uncover both the taxonomy and the metabolism of the microbial community under investigation. The functional profile limitations regard the actual expression of the enzymes in the microbial community and the initial annotation bounded by the current database adopted as reference.

TABLE 4.3: **List species in the mock community**

Species
Akkermansia muciniphila ATCC BAA-835
Alistipes finegoldii DSM 17242
Bacteroides thetaiotaomicron VPI-5482
Bifidobacterium adolescentis ATCC 15703
Bifidobacterium longum subsp. longum GT15
Blautia hansenii DSM 20583
Christensenella sp. Marseille-P2437 strain Marseille-P2437T
Clostridium perfringens F262
Collinsella aerofaciens ATCC 25986
Coprococcus catus GD/7
Desulfovibrio sp. FW1012B
Dorea formicigenerans ATCC 27755
Escherichia coli str. K-12 substr. MG1655
Eubacterium rectale ATCC 33656
Faecalibacterium prausnitzii SL3/3
Prevotella copri DSM 18205
Roseburia intestinalis XB6B4
Ruminococcus albus 7
Ruminococcus bromii strain L2-36
Ruminococcus gnavus AGR2154
Ruminococcus obeum A2-162
Ruminococcus torques L2-14

Chapter 5

The human gut microbiota

5.1 Gut microbiota 101

Since the technological advancement of microbiome research over the past 20 years, we are accumulating valuable insights into the structure and function of microbiotas in different ecosystems (Garrido-Cardenas and Manzano-Agugliaro, 2017). Among the multiple fields of application, a great effort has regarded the characterization of the human being. As for other animals, humans host trillions of bacteria, networking with each other and with the host. Among the internal and external surfaces covered by these microbes, the gastrointestinal tract is the most colonized compartment, with an estimated presence of up to 1014 prokaryotic cells (Rajilić-Stojanović, Smidt, and De Vos, 2007). Following this number, more than one million is the estimated number of genes present in the cumulative genome of the gut microbiota, about 150 times higher than those coded in the human genome (Lepage et al., 2013). This tremendous difference indicates a set of metabolic characteristics of exclusive bacterial prerogative, which complement and confer to the host important new functions, possibly determining a higher adaptability and resilience potential. Notably, the gut dwelling bacteria share the gastrointestinal tract with members of other domains of life - Eukarya and Archaea - that are contemplated within the microbiota definition as given previously in Chapter 3.

5.2 Hallmarks of gut microbiota ecology

Distinct physiological conditions along the gastrointestinal tract determine the presence of compartmentalized microbial habitats (Donaldson, Lee, and Mazmanian, 2016). Starting from the oral cavity, down to the esophagus and the stomach, bacteria can already encounter different environmental settings. The stomach, once believed to be a sterile niche, is highly challenging for bacterial life, due to the low pH of the gastric juice. The stomach in fact exhibits the lowest diversity and abundance in terms of bacterial species. *Helicobacter pylori* can colonize the gastric mucosa, but it is possible to encounter other commensal bacteria, members of the following phyla, Proteobacteria, Firmicutes, Actinobacteria, Bacteroidetes and Fusobacteria (Ferreira et al., 2018).

Interestingly, several other factors influence the presence of bacteria colonizing distinct niches of the gastrointestinal tract - including the small intestine, the cecum and the large intestine. In this respect, the availability of glycans and the presence of antimicrobials peptides (AMPs) play an import role (El Kaoutari et al., 2013). Shifts in chemicals, presence of digestive enzymes and nutrient gradients, as well as the host immunity response and the presence of mucus, vastly contribute to shape this variation along the small intestine down to the colon. Importantly, the time of transit of the chyme from the stomach to the first tract of the small intestine, the duodenum, is significantly shorter compared to its permanence in the colon. This aspect is believed to be pivotal for bacterial colonization. Furthermore, the small intestinal tract is characterized by a higher level of oxygen and bile acids. Notably, primary bile acids, produced by the liver and reversed in the duodenum through the bile duct, are able to dissolve the outer membrane of certain bacterial species. Given the limiting physiological configurations of the small intestine, the dominant families in this tract are *Lactobacillaceae* and *Enterobacteriaceae* (Sengupta et al., 2013).

On the other hand, the cecum and colon are densely colonized, representing the most populated tracts of the human body, also characterized by a high species diversity. Human metabolism also influences the gut microbiota biogeography. Being simple sugars and fatty acids absorbed in the small intestine, polysaccharide fermentative bacteria colonize the large intestine. Furthermore, a set of features of the colon, namely the slower transit time of digested food, the decreased presence of AMP and primary bile acids, contribute to the growth of anaerobic bacteria specialized in the degradation of complex polysaccharides. Bacteria with this fermentative potential mainly belong to the families *Bacteroidaceae* and *Clostridiaceae* (Saffarian A et al., 2019).

5.2.1 Gut microhabitats and colonization

The mucus acts as a physical barrier. It is produced by specialized cells, the goblet cells, and forms a layer covering the epithelium of the small and large intestine. In this respect, the difference regarding these two macro-areas lies in the thickness of the mucus layer. A mono-stratum protects the epithelium of the small intestine, whereas two layers are present in the colon, one packed on the other, shielding the epithelium. In the colon, the inner mucus layer, in contact with the lamina propria, is almost devoid of bacteria compared to the outer layer, as seen with FISH technique (Johansson et al., 2008). Together with the mucus, also the presence of AMPs, secretory immunoglobulins A (sIgAs) and a high oxygen concentration concur to the selection of bacteria able to grow in these conditions. Typically, small intestine colonizers are adherent species, namely *Helicobacter* spp. and *Lactobacillaceae*. In the colon, we can find mucin-degrading species as *Bacteroides acidifaciens*, *Bacteroides fragilis*, *Akkermansia muciniphila*, *Ruminococcus gnavus* and members of *Bifidobacteriaceae* (Donaldson, Lee, and Mazmanian, 2016).

The crypts of the colon, as well as the inner mucus and the appendix, serve as a protective barrier for microbial cells, primarily against fecal stream, and also against different types of perturbations, which can be due to changes in diet, intestinal motility and antibiotic intake. Such perturbations can deplete the lumen from bacteria, whereas these micro-niches represent a reservoir of species capable of recolonizing the lumen, providing a means for the recovery of the gut microbiota.

Regarding gut colonization, bacteria have evolved several mechanisms to evade the host immune response. *Haemophilus influenzae*, known to cause respiratory infections, is also responsible for gastrointestinal diseases. Like the chameleon, *H. influenzae* is capable of shielding its cellular surface with host sialic acids, reducing the possibility of being spotted by the host immune system (Severi et al., 2005). Gram positive and Gram negative bacteria use different strategies to evade the bactericidal effect of antimicrobial peptides. In both cases, they act modifying the external cellular surface, overall eliminating the potential electrostatic interaction with antimicrobial peptides (Lysenko et al., 2000; Saar-Dover et al., 2012).

5.3 Host-bacteria crosstalk

The gut microbiota is considered as the 'forgotten organ', able to modulate the host health. The host-microbiota symbiotic relationship is important for regulating host metabolism, for interaction with the immune system and for communicating with human cells, at both the enteric and systemic level (Rooks and Garrett, 2016). This interaction is highly dynamic and shaped by different factors that determine on the one hand the fate of bacterial colonization of the gastrointestinal tract and on the other hand, the maintenance or vice versa the disruption of human homeostasis.

Notably, the gut microbiota can be seen as an endocrine system, due to its ability to produce and regulate metabolites that can reach other distal organs and systems through the bloodstream. A relevant case are short-chain fatty acids (e.g., butyrate and propionate), products of the carbohydrate metabolism by the gut microbiota, which perform key signalling functions, discussed later in this chapter. Furthermore, the gut microbiota can modulate tryptophan plasma concentrations, important for the synthesis of the neurotransmitter serotonin. Many of the chemicals produced by the gut microbiota are neuroactive, as the main inhibitory transmitter of the brain, γ-aminobutyric acid (GABA), and in addition, dopamine, noradrenaline and serotonin, collectively defining a gut-brain communication axis. Overall, the biochemical possibility and complexity of the gut microbiota far exceed those of our endocrine system and are still not fully elucidated (Clarke et al., 2014).

5.3.1 Immunomodulation

The gut microbiota plays a central role in the development of the immune system by training it to identify commensal and pathogenic bacteria. In the first case, by suppressing the inflammatory reaction, and in the second case by promoting it, thus refining the adaptative responses of our immune system during the course of our life. When this fine tuning is impaired, as in a dysbiotic state (occurring for a variety of reasons, e.g., antibiotic intake, diet change, reduced colonization of the gastrointestinal tract, pathologies, etc.), our immune responses are compromised, and this can lead to an increased risk of disease. In particular, alterations of this fine balance can induce the breakdown of the mucosal barrier with consequent translocation of microorganisms in the lamina propria underneath the epithelium. Secretory immunoglobulin A (sIgAs) are one of the sentinels of our immune system in the gut. sIgAs are highly present in mucus monitoring the microbiota and mediating host homeostasis. Certain pathogenic bacteria are entirely coated by sIgAs (Palm et al., 2014). This assemblage could contribute to the formation of a biofilm, acting as a barrier against the adherence of bacteria to the epithelium (Mathias and Corthésy, 2011). Non-pathogenic strains *B. fragilis* and *Bifidobacterium breve* are well-studied cases of specific immunomodulation. These bacteria have evolved mechanisms to stimulate the production of the anti-inflammatory cytokine IL-10 by immune regulatory T cells. As a result of avoiding a pro-inflammatory reaction, they make their way through the mucus layer of the colon (Round et al., 2011). Another sophisticated mechanism evolved by commensal bacteria for ensuring gut colonization by modulation of the host immune response, is the stimulation of a set of T helper cells, Th17 cells, playing an important protective role against pathogens (Ivanov et al., 2009). On the other hand, also pathogenic bacteria have evolved mechanisms to reduce the host pro-inflammatory response by limiting the immune response (Monack, Mueller, and Falkow, 2004).

5.3.2 Gut microbiota metabolism

In addition to stimulating and alerting our immune system, commensal bacteria provide an invaluable extension to our catabolic activities, with special regard to complex carbohydrates that our gut would otherwise be unable to digest, thus complementing our saccharolytic activity. This function is performed by specific degradation enzymes encoded in the bacterial genome, known as Carbohydrate-Active enZYmes (CAZymes) (Huang et al., 2017). Bacteria belonging to the genus *Bacteroides* are the most involved in these metabolic pathways (Koropatkin, Cameron, and Martens, 2012). In particular, *B. thetaiotaomicron* has the ability to metabolize both diet and host-derived glycans from intestinal mucus. Furthermore, bacteria are involved in the anabolism and absorption of secondary bile acids, lipids, amino acids and vitamins (Flint et al., 2015).

5.3.2.1 Short-chain fatty acids

From the fermentation of polysaccharides, occurring in the colon, bacteria produce short-chain fatty acids (SCFAs). These end-products of the anaerobic fermentation of sugars, particularly butyrate, propionate and acetate, are important signaling molecules that play a key multifactorial role in human physiology. For example, they are known to stimulate the expansion and differentiations of regulatory T cells (Tregs) (Smith et al., 2013; Atarashi et al., 2013; Arpaia et al., 2013). This interaction can be an indirect way for the immune system to assess, through the sensing of fermentation end-products, the presence of commensal bacteria, while otherwise acting against pathogens. Well-studied effects of SCFA chemical communication are inhibition of histone deacetylases (HDACs) and activation of G protein-coupled receptors (GPCRs). Notably, GPCRs are involved in the regulation of metabolism, inflammation and neurological homeostasis - with a series of cascading consequences, some of which are not yet well disclosed (Sivaprakasam, Prasad, and Singh, 2016; Sun et al., 2017).

5.3.2.2 Bile acids

Bile acids (BAs) are hydroxylated steroids synthesized from cholesterol in the liver. BAs are a component of bile and they facilitate the solubilization of fat molecules, lipids and fat-soluble vitamins. Only 5% of BAs is lost daily, and the rest is reabsorbed in the ileum and transported back to the liver. Bacteria in the colon have evolved bile acid metabolism. From the hydrolysis of primary bile acids, through bile salt hydrolase (BSH), hydroxyl group dehydrogenation and 7-dehydroxylation, bacteria can generate secondary bile acids, of which the two most important are deoxycholic and lithocholic acid (Ridlon, Kang, and Hylemon, 2006). Metagenomics analysis has revealed an enrichment of enzymes catalysing these reactions in the gut microbiota in comparison to other ecosystems. Importantly, in a study in mice, increased levels of deoxycholic acid have been associated with obesity and cancer (Yoshimoto et al., 2013). BAs are important signaling molecules, by binding to the cytoplasmic G protein-coupled receptor TGR5 (TGR5/M-BAR) and nuclear farnesoid X receptor (FXR), they participate in glucose regulation and lipid metabolism (Ramírez-Pérez et al., 2017). FXR is expressed by many tissues and its ability to ligate to several BAs, including secondary ones, reveals the role of the gut microbiota in the modulation of FXR signaling. The study of BAs has also highlighted the role of this receptor in many diseases, and the importance of diet for the host physiopathological state (Wahlström et al., 2016).

5.4 Onset of colonization of the human gastrointestinal tract

The gut is at the center of human health, whose connection with other organs determines the gut-brain, gut-lung and gut-liver axes. This complex and dynamic ecosystem is shaped since human birth (Milani et al., 2017). However, the very starting point of bacterial colonization in humans is still somehow debated. Few studies, supporting the theory of prenatal colonization, have identified traces of bacterial DNA in the amniotic fluid, umbilical cord blood and placenta (Bearfield, 2002; Jiménez et al., 2005; Aagaard et al., 2014). In this debate, a difference is undoubtedly detected in the gut microbiota composition of naturally born babies compared to those delivered by C-section (Dominguez-Bello et al., 2010). Overall, the infant microbiome is largely shaped by the maternal microbiota. Disruption of the maternal-offspring microbiota exchange, either via C-section delivery or via the usage of antibiotics during pregnancy, is associated with a series of health-related issues. Particularly, these include asthma, increased risk of obesity, celiac disease and type 1 diabetes (Kero et al., 2002; Huh et al., 2012; Decker, Hornef, and Stockinger, 2011; Algert et al., 2009). After birth, the gut microbiota is shaped by a milk-based diet, enriching the intestinal community with *Bifidobacterium* species able to degrade milk polysaccharides (Marques et al., 2010).

Most of the development of the gut microbiota occurs when the infant changes the diet to solid food, marking an increase in bacterial richness, with increased abundance of the genera *Bacteroides*, *Clostridium* and *Ruminococcus* (Marques et al., 2010; Bäckhed et al., 2015; Arrieta et al., 2014). These changes are also mirrored by the metabolism, updated from simple to complex carbohydrate degradation, and the integrated ability to synthetize SCFAs as fermentation end-products. A major shaping of the gut microbiota occurs in the early years of infant life up to the age of three, with a prominent impact from the environment and lifestyle, but its development continues even during adolescence (Agans et al., 2011; Derrien, Alvarez, and Vos, 2019).

5.5 Characterization of adult-like gut microbiota

The shaping of an adult-like microbiota begins with the transition to solid food. The gut ecosystems is different from any other habitat, with several bacterial species exclusively living in our gut. The gastrointestinal tract is estimated to be colonized by one thousand different bacterial species (Bäckhed, 2012). The vast majority has been determined only by *in silico* characterization of the 16S rRNA gene, thus lacking a classical microbiological approach of isolation and cultivation (Rajilić-Stojanović, Smidt, and De Vos, 2007). Despite the high diversity revealed at the species level, these taxa are classified within ten major phyla of the hundred known (Peterson et al., 2008). Importantly, the main components belong to Firmicutes and Bacteroidetes, constituting more than 80% of the human gut microbiota.

5.5.1 Gut microbiota stability and variability

In 1998, a study on sixteen adults, based on 16S rRNA gene amplification and TGGE, demonstrated the uniqueness of the human gut microbiota. Furthermore, by studying the microbiota of two of those individuals at two time points, the researchers found that the microbiota constituted a stable fingerprint for at least six months (Zoetendal, Akkermans, and De Vos, 1998). Subsequent researches, extended to the characterization and monitoring of the microbiota of other body sites, have since then improved our understanding of microbiota diversity in adulthood. In 2005, another milestone study regarded the analysis of 13,335 16S rRNA sequences amplified from multiple samplings of the colonic mucosa and also from the feces of three individuals. This was the first step in the adoption of stools as a proxy for the characterization of the gut microbiota (Eckburg et al., 2005). The results from this work confirmed the inter-individual variation, highlighting the dominant presence of members of the phyla Firmicutes and Bacteroidetes, followed by other minor components, such as Actinobacteria, Proteobacteria and Verrucomicrobia.

In the following years, another study focused on 27 different body sites, in seven individuals sampled at four time points, adding another important piece to our understanding of the composition of the human microbiota. In this work, the interpersonal variability across body niches was highlighted, thus stressing the high variability of the bacterial composition between and within the human host (Costello et al., 2009). With regards to intra-individual variability, distinct phases of human life are characterized by specific microbiota configurations. Overall, various endogenous and exogenous factors contribute in shaping the individual gut microbiota, these can be lifestyle - also considering the variability due to ethnicity and geography – the use of antibiotics and medications, diet and the inevitable ageing process, and only minimally the host genotype. In particular, diet is one of the strongest influences in modeling bacterial composition and function (Zmora, Suez, and Elinav, 2019).

5.5.2 Metagenomics milestone studies

Since 2005, researchers have progressed from the study of three, to seven up to 124 fecal samples and 576.6 Gb of sequences, as in 2010 with the international project MetaHIT (Metagenomes of the Human Intestinal Tract). The MetaHIT project aimed to characterize the diversity of the human gut microbiota, at an unprecedented higher sampling scale, and also to identify a core human microbiota. This project published a gene catalogue of the gut microbiota, identifying 18 species shared in the whole cohort. However, further studies have highlighted the presence of common key functionality in the gut microbiome of humans rather than key species (Qin et al., 2010).

In this context, it is important to mention the Human Microbiome Project (HMP) Consortium (Consortium et al., 2010) that, by doubling the number of individuals and sampling eighteen body sites, further confirmed the previous hypotheses. Microbiota stability was also evaluated for the first time by reconstructing the longest trajectory, from two individuals across multiple body sites for one year, confirming individual variability across months (Caporaso et al., 2010). A second phase of this study, the HMP 2.0 (Integrative et al., 2019), has focused on following the dynamics of the gut microbiota in health and diseased states (*i.e.*, pregnancy and preterm birth, inflammatory bowel disease (IBD), dietary perturbations and infectious diseases that affect individuals with prediabetes). Importantly, in this second phase, the project is based on a multiomics approach, with the inclusion of metagenomics, metatranscriptomics, metaproteomics, metabolomics as well as host genomics, epigenomics and antibody profile.

Another important longitudinal study evaluated the long-term stability of the gut microbiota across five years (Faith et al., 2013). In this time span, the authors identified 60% of the gut microbiota to remain stable over time. This data was confirmed by 16S rRNA amplicon analysis and by strain isolation and genome sequencing from fecal samples, showing the individual stability of the strains cultured over 5 years of sampling. Furthermore, this study highlighted the higher stability of members of the phylum Bacteroidetes and Actinobacteria, compared to Firmicutes and Proteobacteria.

These milestone studies have strengthened the idea of a human microbiota fingerprint. Microbiota configuration in adults tends to be stable (Faith et al., 2013; M. and W.M., 2014; Franzosa et al., 2015; Palleja et al., 2018), but dynamic enough to adapt to major changes, e.g. changes in diet or other environmental factors (David et al., 2014), antibiotics and immune system response.

More recently, several studies, aimed at investigating the changes and the signature features of the gut microbiota in health and disease (*i.e.*, type 2 diabetes, metabolic syndrome, obesity and fatty liver), have pinpointed a higher predominance of envinomental factors (such as geography and lifestyle habits) in shaping the gut microbiota composition, rather than the genetic component (Rothschild et al., 2018; He et al., 2018; Deschasaux et al., 2018).

Further large and longitudinal studies will be needed to enhance our knowledge over the physiological changes shaping the gut microbiota through ageing, including elderhood and longevity, aiming to discriminate bacterial functions and species that characterize a healthy microbiota.

5.6 Gut microbiota alterations

Dysbiosis is defined as an altered composition of the gut microbiota with a detrimental effect on the host pathophysiology. Overall, a dysbiotic state is usually characterized by a reduced diversity of the microbiota, with a decreased presence of health-associated commensals and an increase in pathogens or pathobionts. The administration of antibiotics is another major factor shaping the gut microbiota, creating opportunities for pathogen colonization (Antonopoulos et al., 2009).

A common signature of any metabolic disorder is the unbalanced abundance of the Firmicutes and Bacteroidetes phyla. Known cases of an altered gut microbiota are inflammatory bowel disease (IBD) (Franzosa et al., 2019), type 1 and 2 diabetes (Han et al., 2018; Sharma and Tripathi, 2019), and obesity (Gomes, Hoffmann, and Mota, 2018). The early definition regarding pathogenicity in all 'black' or all 'white' is changing as new studies have highlighted the importance of host-bacteria interactions in the emergence of pathogenicity (J.S. et al., 2017). The host genetics, as well as the environment, play a crucial role. In this context, a known case is represented by Chron's disease (Cho, 2008). Dysfunctional Paneth cells, responsible for host production of AMPs, are one of the proposed causes contributing to the Chron's dysbiotic state (Bevins and Salzman, 2011).

Previous findings suggested that the rise of a pathogenic state is induced by the presence of bacteria in direct contact with the epithelial cells of the intestine. In particular, pathogens can cause the disruption of the mucosal barrier by secreting toxins capable of breaking cell junctions. This causes reduced growth of epithelial cells with a consequent decrease in mucus production. The breakdown of the mucosal barrier allows pathogens to translocate into the lamina propria. However, recent works have established the ability of certain bacteria to penetrate the mucosal barrier and get in contact with the epithelial cells even in healthy individuals (Swidsinski et al., 2005). Both commensal and pathogenic bacteria can reach the epithelium by degrading mucus with mucinases and proteases, and adhere to the tissue through pili, lectins and other outer-membrane proteins. Certain bacteria are also able to use the glycans in the mucus to facilitate their attachment, as in the case of some *E. coli* strains expressing the fimbrial protein FimH, able to interact with mannose residues in glycans (Krogfelt, Bergmans, and Klemm, 1990).

5.7 Conclusions and perspectives on ageing and gut microbiota

To summarize, the gut microbiota is stratified both horizontally and vertically in micro-niches and its composition is highly influenced by nutrients, chemical gradient and host immune response. The mucus itself represents an important reservoir of bacteria, also providing the necessary nutrients to bacterial species, through its glycan components, scavenged in case of limited food resources. Given the importance of the gut microbiota in shaping human health and disease, metagenomics studies have gained an important place in deciphering host physiology. Being non-invasive, fecal microbiota profiling though omics techniques provides a valuable tool for determining composition, function and dynamics of the microbiota community associated with a specific state. Several metagenomics studies on the gut microbiota have contributed to highlight the main microbial actors in health and disease. This chapter has emphasized the gut microbiota composition of human beings and, importantly, how it is shaped by the ageing process. Many age-related lifestyle changes have been found to contribute to the re-configuration of the gut microbiota, with potential health consequences (Kim and Jazwinski, 2018), first of all, changes in nutrition and usage of medications. Contributing factors are also a decrease in locomotion, leading to a frailty state, and changes in the host immune response. Importantly, the residential status and long stay in facilities have a detrimental impact on the microbiota configurations, as well as hospitalization (O'Toole and Jeffery, 2015; Araos et al., 2019). The microbiota configuration is even more challenged at its core in extremely longevous individuals after 100 years of symbiotic relationship. Although this topic is of relevant importance to our society, limited studies have been conducted for the characterization of the gut microbiota and its importance in healthy ageing. The following chapters regard three original studies The first describing the gut microbiota of an elderly cohort in relation to the health state, particularly visceral adiposity, the second characterizing the gut resistome over the course of life, including centenarians and semi-supercentenarians, and the last investigating the gut metagenome of the aforementioned cohort in relation to the metabolic pathways involved in xenobiotic degradation.

Chapter 6

Gut microbiota in elderly. Elevated gut microbiome abundance of *Christensenellaceae*, *Porphyromonadaceae* and *Rikenellaceae* is associated with reduced visceral adipose tissue and healthier metabolic profile in Italian elderly

6.1 Abstract

Ageing is accompanied by physiological changes affecting body composition and functionality, including accumulation of fat mass at the expense of muscle mass, with effects upon morbidity and quality of life. The gut microbiome has recently emerged as a key environmental modifier of human health that can modulate healthy ageing and possibly longevity. However, its associations with adiposity in old age are still poorly understood. Here we profiled the gut microbiota in a well-characterized cohort of 201 Italian elderly subjects from the NU-AGE study, by 16S rRNA amplicon sequencing. We then tested for association with body composition from dual-energy X-ray absorptiometry (DXA), with a focus on visceral and subcutaneous adipose tissue. Dietary patterns, serum metabolome and other health-related parameters were also assessed. This study identified distinct compositional structures of the elderly gut microbiota associated with DXA parameters, diet, metabolic profiles and cardio-metabolic risk factors.

Keywords: microbiota, ageing, visceral adipose tissue, inflammaging, diet, serum metabolome

6.2 Introduction

The gut microbiome is a crucial component of the individual health, by means of its impact on food degradation, energy intake, and regulation of immune system functionality (Cani et al., 2019; Spanogiannopoulos et al., 2016; Kim, Zeng, and

Núñez, 2017). Recently, the human gut microbiome has also been proposed as a determinant of healthy ageing, by counteracting inflammaging (i.e., the low-grade chronic inflammation characterizing the advancement of age), immunosenescence, intestinal permeability, and the decline in cognitive and bone health, thus helping to preserve homeostasis (Claesson et al., 2012; Biagi et al., 2010; Ventura et al., 2017; Nicoletti, 2015; Villa, Ward, and Comelli, 2017; Leung and Thuret, 2015).

Overall, the aged-type gut microbiome is reported to be characterized by altered diversity, with increased representation of opportunistic bacteria and potential pathobionts, and reduced relative abundance of microbes capable of producing short-chain fatty acids (SCFAs), key signalling molecules for host metabolic and immunological homeostasis (Biagi et al., 2010; Rampelli et al., 2013). Interestingly, while the aforementioned microbiome modifications have been found to persist in longevity, in those individuals who reach the extreme limit of human lifespan (*i.e.*, semi-supercentenarians, aged >105 years), some peculiarities have emerged, that is an enrichment and/or greater prevalence of health-associated taxa, *Bifidobacterium*, *Akkermansia* and *Christensenellaceae* (Biagi et al., 2016). Though it is not yet clear how and especially when these age-related microbiome structures are established, it is worth noting that the increased representation of health-promoting microbes in extremely old people appears to be robust to geography, as a sort of universal microbiome signature of healthy ageing and longevity (Santoro et al., 2018b).

Ageing involves a series of changes in body composition, which generally result in higher levels of fat mass at the expense of muscle mass, with critical implications in terms of morbidity and quality of life (Santoro et al., 2018a; St-Onge and Gallagher, 2010; Reinders, Visser, and Schaap, 2017; Bazzocchi et al., 2013; Ponti et al., 2020). Previous works, exploring how ageing affects body mass distribution, have shown that muscle tissue and high-metabolic rate organs such as brain, kidneys, liver and spleen, decrease in mass with increasing age, while the abdominal area is more prone to fat deposits (St-Onge and Gallagher, 2010; Conte et al., 2019).

The elderly indeed tend to accumulate fat in the muscles, liver and viscera as lipid droplets, while losing subcutaneous fat mass (Reinders, Visser, and Schaap, 2017). The age-related accumulation of fat deposition has been associated with an increase in a pro-inflammatory state that may contribute to the onset of cardiovascular disease, insulin resistance and type 2 diabetes (Fried, Bunkin, and Greenberg, 1998; Kanda et al., 2006). In particular, evidence has shown that excess visceral adipose tissue (VAT) rather than accumulation of subcutaneous adipose tissue (SAT), represents the cause of atherosclerotic cardiovascular events and is the key contributor to metabolic syndrome (Sato et al., 2018; Gómez et al., 2019).

The accumulation of fat mass is also well known to be linked to the gut microbiota. Landmark studies in animal models revealed that microbial transplantation from obese to lean mice was able to induce weight gain (Turnbaugh et al., 2006; Turnbaugh et al., 2008). More recently, findings in human subjects showed that lean and obese individuals have a particular gut microbial signature in terms of composition and diversity with also differences between men and women (Ridaura VK et al., 2013; Rampelli et al., 2018; Cancello et al., 2019; Min et al., 2019). Furthermore, in one of the largest gut microbiota-obesity studies to date, conducted in a cohort of twins, the authors suggested the existence of heritable microbes that could play a major role in components of adiposity relevant for cardiovascular risk (Beaumont et al., 2016). However, these studies have mostly dealt with individuals with a wide age range, so the associations between microbiome and fat distribution in the elderly are still poorly understood.

Measures specifically assessing visceral fat could help better explore the contribution of the gut microbiome to abdominal adiposity. In this regard, different imaging methods such as ultrasound, computed tomography (CT) and magnetic resonance, are able to assess VAT and SAT, being CT the "gold standard" technique for assessing this. However, CT is limited by radiation exposure and availability and MRI is limited in terms of availability. Dual energy X-ray absorptiometry (DXA) is considered the "gold standard" technique for body composition assessment at molecular level – translated into a 3-compartment model of fat mass, non-bone lean mass and bone mineral content, and certain DXA devices have embedded algorithms to specifically estimate the amount of VAT and SAT in the android region (Ponti et al., 2020; Guglielmi and Bazzocchi, 2020). Unlike indirect measures, such as the body mass index (BMI), DXA allows rapid, sensitive and accurate, yet non-invasive, characterization of body composition, including levels of fat and lean mass, and bone density (Messina et al., 2020). In an attempt to better reveal the associations between abdominal adiposity and gut microbiome in old age, here we analysed the multivariate relationship between DXA-derived measures of VAT and SAT and the gut microbiota structure, as profiled by 16S rRNA gene-based sequencing, in a cohort of 201 Italian seniors, enrolled within the EU FP7 NU-AGE project. Dietary data, collected by 7-day food records, and serum metabolomics data, generated by ultra-performance liquid chromatography coupled to quadrupole-time-of-flight mass spectrometer, were also analysed to explore associations of macro/micronutrients and metabolites with abdominal adiposity-related microbiota profiles. We find that distinct gut bacterial taxa are associated with reduced VAT, as well as with peculiar profiles of circulating metabolites and food intake. Monitoring and possibly modulating the gut microbiota, in addition to promoting healthy eating habits, could therefore represent an additional tool to support healthy ageing and possibly longevity.

6.3 Materials and Methods

6.3.1 Study cohort

NU-AGE (https://cordis.europa.eu/project/id/266486) is a multicentre EU FP7 project, ended in 2016, which involved 30 partners from 16 European countries, working in the field of nutrition, gerontology, immunology and molecular biology. NU-AGE objective was to study the effects of a 12-month customized Mediterranean diet (registered with clinicaltrials.gov, NCT01754012) on the ageing process, including cognitive decline, bone density, muscle mass, digestive health, immune and cardiovascular systems. Enrolment of participants has been described in detail previously (Santoro et al., 2014; Marseglia et al., 2019). Briefly, after screening for inclusion/exclusion criteria, 1,279 free-living healthy elderly aged 65 to 79 years were enrolled across five EU countries (Poland, Netherlands, UK, France, Italy) and thoroughly characterized for anthropometry, nutritional status, body composition, health and medical status, cognitive and physical functions, and a series of biochemical and inflammatory measures (Fried et al., 2001). Participants were classified according to their frailty status based on the five criteria proposed by Fried and colleagues, including weight loss, weakness (*i.e.*, poor handgrip strength), self-reported exhaustion, slowness (*i.e.*, slow gait speed), and low physical activity. Only non-frail (absence of all the above 5 criteria) and pre-frail (presence of 1 or 2 of the criteria) subjects were included in the study (Marseglia et al., 2018).

Written informed consent was collected from all participants prior to their inclusion in the study, in accordance with the Declaration of Helsinki. NU-AGE was approved by the ethics committee of the coordinator centre — the Independent Ethics Committee of the S. Orsola-Malpighi Hospital Bologna (Italy) — and by the local/national ethics committees of all the other four recruiting centres.

As the analysis of VAT and SAT by DXA was only available for the Italian elderly, here we focused only on this cohort, of which we profiled the faecal microbiome by means of next-generation sequencing, and sought correlations with DXA variables, especially VAT and SAT, as well as with dietary habits and circulating metabolites (please see the paragraphs below). As for stool collection, each participant was asked to collect a faecal sample. The samples were immediately stored at -20°C and delivered to the Department of Experimental, Diagnostic and Specialty Medicine (University of Bologna, Bologna, Italy) where they were stored at -80°C until processing.

6.3.2 Anthropometric, physical, cardiovascular, clinical and cognitive function assessment

Height was measured by a stadiometer to the nearest 0.1 cm. Weight was measured to the nearest 0.1 kg with a calibrated scale while wearing light clothes. Body Mass Index (BMI) was calculated as weight/height2 (kg/m2). Waist circumference was measured either at the narrowest circumference of the torso or at the midpoint between the lower ribs and the iliac crest. Hip circumference was measured horizontally at the level of the largest lateral extension of the hips or over the buttocks. Hand grip strength was measured three times in the dominant hand using the Scandidact Smedley's Hand Dynamometer® (Odder, Denmark) to the nearest 0.1 kg. Physical performance was evaluated by the sum score of 6-minute walking distance, Activities of Daily Living (ADL) scale, Instrumental Activities of Daily Living (IADL) scale and PASE questionnaire. Cognitive function was assessed by the administration of the Cambridge Mental Disorders of the Elderly Examination (CAMDEX) subjective memory score, the Geriatric Depression Scale (GDS) score and Mini Mental State Examination (MMSE) score as previously reported (Jennings et al., 2019). Blood pressure and heart rate were measured using the automated and calibrated electronic monitor Omron, M2 compact (Milan, Italy) as previously reported (Santoro et al., 2019). The use of prescribed medicines and supplements and the clinical history were collected by a questionnaire and verified by interviewers. All measures and questionnaires were taken by trained research assistants.

6.3.3 Body composition assessment

Direct measurements of total and regional body composition were obtained by performing whole body DXA scan (Lunar iDXA, GE Healthcare, Madison, WI – enCORETM 2011 software version 13.6 and upgrade to estimate VAT and SAT). The scanners are compliant with standard quality control procedures and were re-calibrated daily following the manufacturers' instructions. DXA scans were performed by trained personnel, removing all metal items prior to densitometry and placing the subjects in supine position with arms resting on the side of the participant's body, leaving some space with respect to the trunk and centred on the scanning field. DXA scanned the following regions: total body, trunk, upper limbs, lower limbs, android region (from the two superior iliac crests and extended cranially and covering 20% of the distance to the chin) and gynoid region (from the greater trochanter of the femur directed caudally and covering two times the distance in the android region). For each scanned region, the weight (in g) of the total mass, fat mass, non-bone lean mass and bone mineral content were obtained. Measurements of VAT and SAT were obtained at android level with CoreScan software.

This work includes variables related to total fat and lean mass distribution and bone mineral content, with a focus on fat measures of the abdominal area including subcutaneous and visceral adiposity: whole body fat mass (FM, g), whole body fat mass index (FMI, g/m^2), whole body fat mass to lean mass ratio (FM/LM), whole body non-bone lean mass (LM, g), whole body non-bone lean mass index (LMI, g/m^2), skeletal mass index (SMI, i.e., the appendicular lean mass to total body mass ratio), whole body bone mineral content (BMC, g), whole body bone mineral density (BMD, g/cm^2), whole body T-score (T-score), android fat mass to android lean mass ratio (AF/AL), android fat mass to gynoid fat mass ratio (AF/GF), VAT (g) and SAT (g) and their ratio.

6.3.4 Blood sampling and biochemical parameters

Blood samples were obtained after participants had fasted (at least 8 h) and had avoided heavy exercise and alcohol in the prior 24 h. Samples were centrifuged after sitting for 30 min at room temperature and separated into plasma and serum according to a standardized operating procedure, then aliquoted and stored at -80°C until analysis.

Methods for inflammatory parameters assessment are reported in Santoro et al. (Santoro et al., 2019). Briefly, C-reactive protein (CRP), leptin and adiponectin were measured by ProcartaPlex^{TM} Immunoassay (Thermo Fisher Scientific, Waltham, MA, USA), performed using Luminex 200 instrumentation (Luminex Corporation, Austin, TX, USA), according to the manufacturer's instructions. Ghrelin and Pentraxin-3 were measured in multiplex with Bio-Plex Pro human diabetes and Pro human inflammation assay (Bio-Rad, Hercules, CA, USA), respectively. Plates were read and analysed by Bio-Plex Manager Software (Bio-Rad). Plasma homocysteine was measured by an enzymatic assay using an Olympus AU400 clinical chemistry platform (Beckman Coulter, High Wycombe, UK). Serum glucose and serum insulin were determined by biochemical assay and chemiluminescent immunoassay, respectively. Insulin resistance status was calculated according to the homeostasis model assessment of insulin resistance (HOMA-IR) using the following formula: insulin (mIU/mL) × glucose (mmol/L)/22.5 (Matthews et al., 1985). Plasma albumin was analysed using the VITROS ALB slides (Ortho-Clinical Diagnostics, UK) on a VITROS 5.1/FS analyzer. Plasma total, High Density Lipoprotein (HDL) and Low Density Lipoprotein (LDL) cholesterol and triglycerides were measured on a konelab system and reagents were from Thermo Scientific (Asnières sur Seine, France). All the other biochemical analyses, including creatinine, uric acid, alkaline phosphatase (ALP), gamma-glutamyl transpeptidase (GGT), aspartate aminotransferase (AST) and alanine transaminase (ALT) were measured on frozen blood and frozen urine (urea) in a centralized centre with standard methodologies.

6.3.5 Dietary intake data

Dietary intake was assessed by 7-day food records as reported elsewhere (Ostan et al., 2018). Briefly, participants were trained one to one by the interviewer receiving exhaustive instructions to correctly fill in the food diary. Food records were provided in a structured format, with tables for each day and eating occasion, time/hour, location, foods and drinks consumed, quantity and recipes in order to record all details of the meals. Participants were recommended to record data at the time the foods were eaten/consumed and not to change eating habits during the week of registration. At the end of the recorded period, the 7-day food records were accurately checked to obtain more detailed information about types of foods, dressings, preparation methods and recipes, to estimate portion sizes by using real-life models and pictures and to probe the possible consumption of forgotten foods. Consumed foods were coded according to standardized procedures and translated into nutrients by the use of WinFood® software exploiting local food composition tables: INRAN (National Institute for Research on Food and Nutrition, Italy) and IEO (European Institute of Oncology, Italy). Energy (kcal), total carbohydrate (g), total protein (g), animal and plant protein (g), total, saturated and unsaturated fat (g), fibre (g) cholesterol (g), water (g), vitamin (mg: biotin, B1, B2, B3, B6, C, E; µg: folic acid, b-carotene, A, B12, D), and calcium (mg) intake normalized on body weight (g/kg BW), were used in the analysis together with the intake of food groups (white and whole grains, fruits and vegetables, legumes, dairy products, cheese, red and processed meat, white meat, nuts and seeds, potatoes, eggs and egg products, butter and animal fat, olive oil, other vegetable oils, sugar and sweetened beverages, sugar, honey and artificial sweeteners, sweet, chocolate and snacks) (g/day) normalized on body weight. The dietary intakes from the 7-day food records were added/summed to the intakes of related dietary supplements as assessed by a specific vitamin/mineral supplements questionnaire.

6.3.6 Serum metabolomics analysis

Untargeted metabolomics was performed following the procedure described in Pujos-Guillot et al. (Pujos-Guillot et al., 2019). Briefly, serum samples (100 µL) were deproteinized using cold methanol. After evaporation under nitrogen, the dry residues were redissolved in 50/50 (v/v) acetonitrile/water containing 0.1% formic acid. Pooled quality-control samples were prepared by mixing 20 µL from each of the serum samples and prepared similarly. Metabolic profiles were then determined using an ultra-performance liquid chromatography coupled to quadrupole-time-of-flight mass spectrometer (Bruker Impact HD2), equipped with an electrospray source. Separations were carried out using an Acquity HSS T3 column (Waters). Data were acquired in positive and negative ion modes with a scan range from 50 to 1,000 mass-to-charge ratio (m/z). Data were processed under the Galaxy web-based platform Worflow4metabolomics using first XCMS, followed by quality checks and signal drift correction (Giacomoni et al., 2015). The remaining unknown compounds were identified on the basis of their exact masses which were compared to those registered in the Human Metabolome Database (HMDB) or in Kyoto Encyclopedia of Genes and Genomes (KEGG) database. Database results were confirmed using appropriate standards when available, isotopic patterns, and mass fragmentation analyses, performed on a Thermo Scientific LTQ Orbitrap Velos hybrid mass spectrometer (Thermo Fisher Scientific, San José, CA, USA) using high resolution, at 100,000 resolving power.

6.3.7 Microbial DNA extraction

Microbial DNA was extracted from 250 mg of faecal material using the repeated bead-beating plus column method (Yu and Morrison, 2004). Briefly, samples were suspended in 1 mL of lysis buffer (500 mM NaCl, 50 mM Tris-HCl pH 8, 50 mM EDTA, and 4% SDS) and bead-beaten three times in the presence of four 3-mm glass beads and 0.5 g of 0.1-mm zirconia beads (BioSpec Products, Bartlesville, OK, USA), in a FastPrep instrument (MP Biomedicals, Irvine, CA, USA) at 5.5 movements/s for 1 min. Afterwards, the samples were incubated at 95°C for 15 min and subsequently centrifuged for 5 min at 13,000 rpm. The supernatant was added with 260 µL of 10 M ammonium acetate and incubated in ice for 5 min. After a further centrifugation step, one volume of isopropanol was added to the supernatant and incubated in ice for 30 min. Precipitated DNA was washed with 70% ethanol and resuspended in 100 µL of TE buffer. The samples were depleted of RNA and proteins with 2 µL of 10 mg/mL DNase-free RNase at 37°C for 15 min and 15 µL of proteinase K (QIAGEN, Hilden, Germany) at 70°C for 10 min, respectively. Final DNA purification was performed using the QIAamp DNA Stool Mini Kit (QIAGEN). The purified nucleic acids were quantified with the NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA).

6.3.8 16S rRNA gene amplification and sequencing

The V3–V4 hypervariable region of the 16S rRNA gene was PCR amplified with 341F and 805R primers with Illumina overhang adapter sequences as previously reported (Barone et al., 2019). The PCR thermal cycle was as follows: denaturation at 95 °C for 3 min, followed by 25 cycles of denaturation at 95°C for 30 s, annealing at 55°C for 30 s, then extension at 72°C for 30 s, and the last extension step at 72°C for 5 min. The Agencourt AMPure XP magnetic beads (Beckman Coulter, Brea, CA, USA) were used to clean PCR amplicons. Indexed libraries were obtained by limited-cycle PCR using Nextera technology. After a second clean-up as described above, libraries were pooled at equimolar concentration, denatured with 0.2 N NaOH and diluted to 6 pM. For sequencing, an Illumina MiSeq (Illumina, San Diego, CA, USA) platform was used with a 2 x 250 bp paired-end protocol, following the manufacturer's instructions. Sequencing data are available at NCBI SRA under the BioProject ID: PRJNA661289.

6.3.9 Bioinformatics and biostatistics

Sequencing read quality was assessed with Fastqc tool. High-quality read couples were joined together in a single read with PANDAseq tool and the resultant reads with length lower than 350 bp and greater than 500 bp were filtered out. Single-end reads were further pre-processed with DADA2, in order to reduce the noise of the dataset, eliminating chimera sequences and duplicates, and cluster them into amplicon sequence variants (ASVs) (Masella et al., 2012; Callahan et al., 2016). The algorithm VSEARCH was applied to scan the representative feature sequences against the precomputed clusters from SILVA database (128 version) at 99% of sequence identity, and to assign the taxonomy with a confidence score > 0.5 (Quast et al., 2013). The ASVs table was normalized by the minimum number of feature sequences in a sample. Read pre-processing and taxonomic classification were performed in QIIME 2 (release 2018) framework (Bolyen et al., 2019). The R packages Phyloseq and Vegan were used for statistical analysis. Beta diversity was calculated with unweighted, weighted and generalized UniFrac metrics (GUniFrac package), and the function adonis was used to test the significance of beta diversity-based sample separation in Principal Coordinates Analysis (PCoA) (McMurdie and Holmes, 2013; Philip, 2003). The separation of the three microbiome groups (G1 to G3) as found in the unweighted UniFrac-based PCoA was verified by means of hierarchical clustering with Ward as the linkage method. The stability of the clusters was assessed by using average Jaccard similarities from the clusterboot function in the R package fpc. Alpha diversity was estimated using the number of observed ASVs and Chao1 index. Power calculation was computed with micropower R package (Kelly et al., 2015); we found that the size of G1 to G3 microbiome groups allowed 90% power to detect an ω2 of 0.014.

To find associations between the gut microbiota profiles and host characteristics, we adopted the sparse partial least square (sPLS) regression analysis as implemented in the mixOmics package in R, modelling the genus-level relative abundances to the DXA measures or metabolite classes via multiple regressions (Lê Cao et al., 2009). The number of components was tuned to 2, retaining all DXA/metabolomic variables and all taxa in the model. Bacterial abundances were transformed as Centered Log Ratio (CLR). The associations between genera and DXA/metabolomic matrices were visualized projecting the variables inside a correlation circle plot (plotVar), with associated variables projected in the same direction (González et al., 2012). Hierarchical clustering (cim function) on the sPLS regression model was plotted with Pearson correlation as distance and complete linkage method. As for diet, the food groups most contributing to the PCoA ordination space were identified using the function "envfit" of vegan. Significant differences among the microbiota groups identified by PCoA in taxon relative abundance as well as in measures of DXA-related variables, dietary and metabolomics data and other health-related parameters, were assessed using the Kruskal-Wallis test. Wilcoxon test was adopted as a post-hoc test to check for differences between each pair of groups, adjusting p values for multiple testing via Benjamini–Hochberg method. A p value ≤ 0.05 was considered statistically significant.

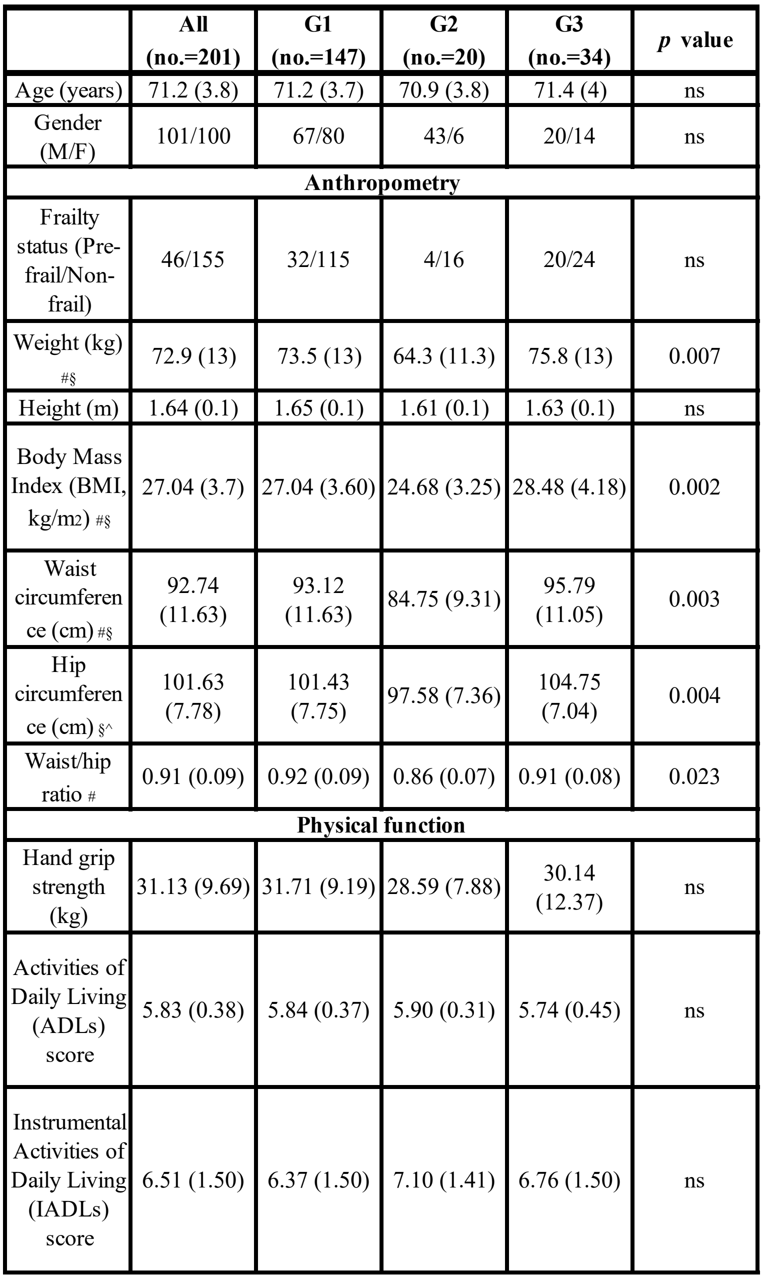
6.4 Results

To explore microbiome links to abdominal adiposity in the elderly, we profiled the faecal microbiome and searched for its correlations with DXA-derived parameters describing fat, in particular visceral and subcutaneous adiposity, and lean mass composition, in a cohort of 201 elderly subjects (101 females, 100 males; age range, 65-79 years, mean age, 71.2 years) from the Emilia-Romagna region (Italy), in the context of the NU-AGE FP7 EU project (see Table 6.1 for cohort description). Our microbiota dataset was composed of 15,167,630 high-quality reads with an average of 75,460 (± 64,658, SD) 300-bp paired-end reads per sample.

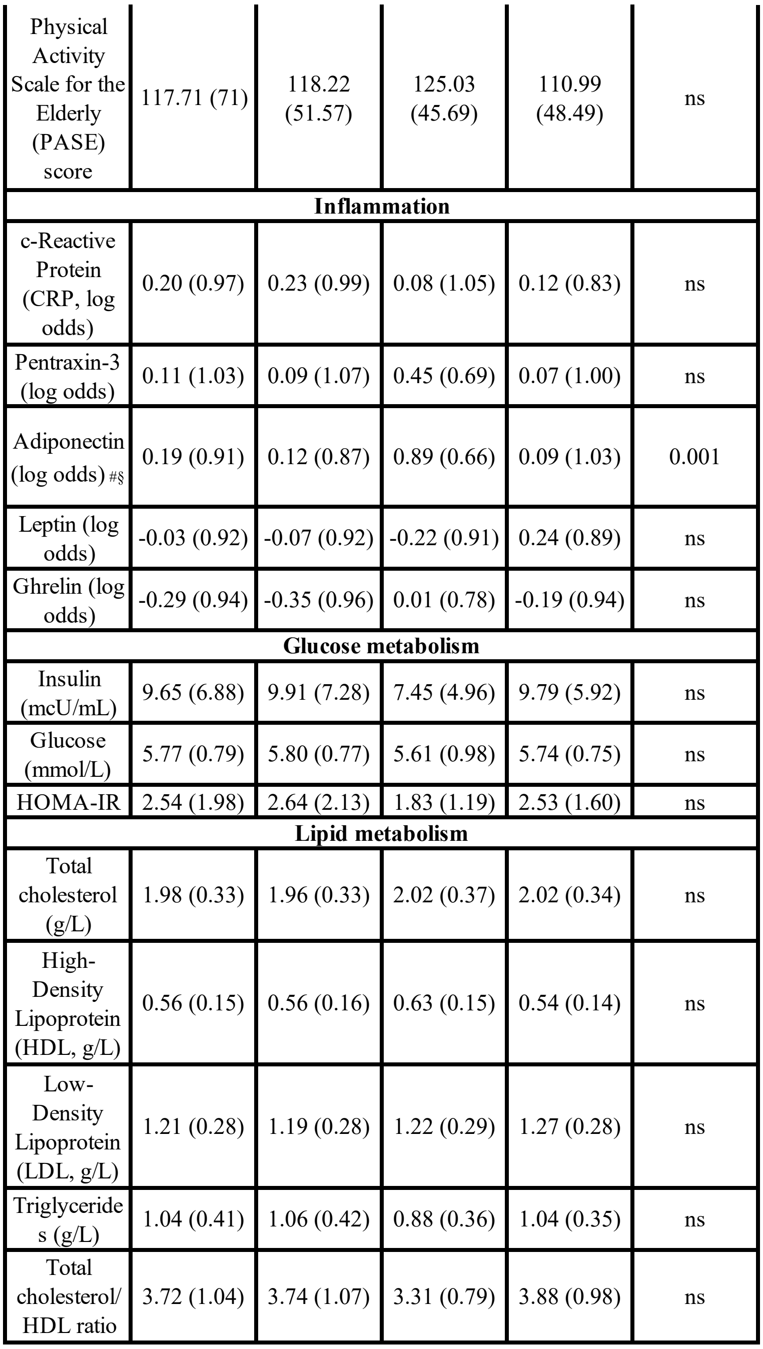
6.4.1 Taxonomic profile

Overall, the gut microbial profiles showed a high representation of the phylum *Firmicutes* (mean relative abundance, 80%), along with *Bacteroidetes* (8.9%) and *Actinobacteria* (7.4%). The *Ruminococcaceae* (37.5%) and *Lachnospiraceae* (27.6%) families, both belonging to *Firmicutes*, were the most represented in the dataset. At the genus level, *Subdoligranulum* (12.5%), *Faecalibacterium* (7.8%) and *Bifidobacterium* (4.6%) were identified as the most abundant taxa.

TABLE 6.1: **Demographic, anthropometric, biochemical and other health-related parameters in a cohort of 201 Italian elders.** Data are shown for the entire cohort as well as for the three microbiome groups (G1 to G3), as identified by PCoA of unweighted UniFrac distances (see Figure 6.1 a). Values are expressed as mean (SD), unless otherwise indicated. P values were determined by Kruskal-Wallis test, followed by post-hoc Wilcoxon test. HOMA-IR, homeostasis model assessment of insulin resistance. ns, not significant.

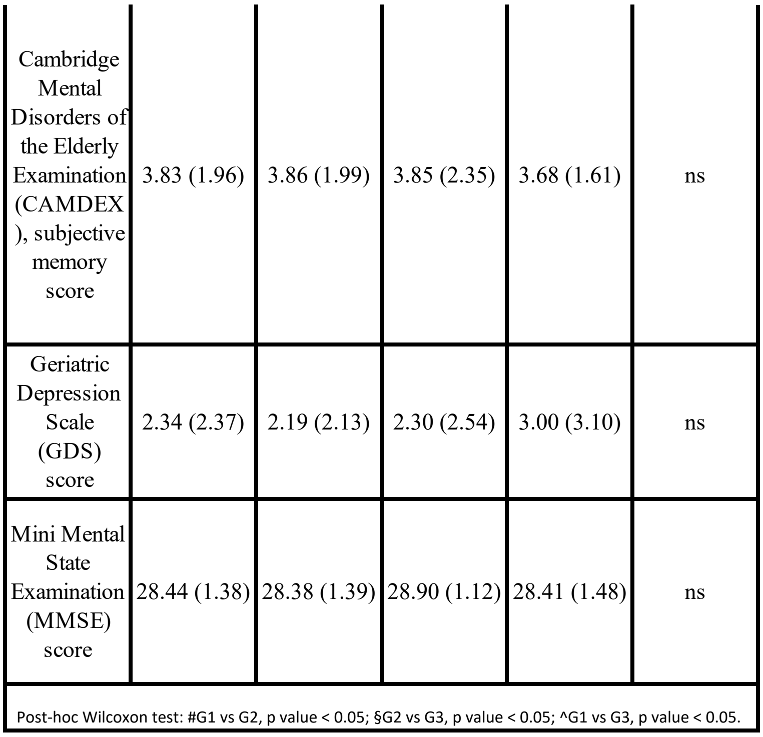


	All (no.=201)	**G1 (no.=147)**	**G2 (no.=20)**	**G3 (no.=34)**	***p* value**
Age (years)	71.2 (3.8)	71.2 (3.7)	70.9 (3.8)	71.4 (4)	ns
Gender (M/F)	101/100	67/80	43/6	20/14	ns
Anthropometry					
Frailty status (Pre-frail/Non-frail)	46/155	32/115	4/16	20/24	ns
Weight (kg) #§	72.9 (13)	73.5 (13)	64.3 (11.3)	75.8 (13)	0.007
Height (m)	1.64 (0.1)	1.65 (0.1)	1.61 (0.1)	1.63 (0.1)	ns
Body Mass Index (BMI, kg/m$_2$) #§	27.04 (3.7)	27.04 (3.60)	24.68 (3.25)	28.48 (4.18)	0.002
Waist circumference (cm) #§	92.74 (11.63)	93.12 (11.63)	84.75 (9.31)	95.79 (11.05)	0.003
Hip circumference (cm) §^	101.63 (7.78)	101.43 (7.75)	97.58 (7.36)	104.75 (7.04)	0.004
Waist/hip ratio #	0.91 (0.09)	0.92 (0.09)	0.86 (0.07)	0.91 (0.08)	0.023
Physical function					
Hand grip strength (kg)	31.13 (9.69)	31.71 (9.19)	28.59 (7.88)	30.14 (12.37)	ns
Activities of Daily Living (ADLs) score	5.83 (0.38)	5.84 (0.37)	5.90 (0.31)	5.74 (0.45)	ns
Instrumental Activities of Daily Living (IADLs) score	6.51 (1.50)	6.37 (1.50)	7.10 (1.41)	6.76 (1.50)	ns



Physical Activity Scale for the Elderly (PASE) score	117.71 (71)	118.22 (51.57)	125.03 (45.69)	110.99 (48.49)	ns
Inflammation					
c-Reactive Protein (CRP, log odds)	0.20 (0.97)	0.23 (0.99)	0.08 (1.05)	0.12 (0.83)	ns
Pentraxin-3 (log odds)	0.11 (1.03)	0.09 (1.07)	0.45 (0.69)	0.07 (1.00)	ns
Adiponectin (log odds) #§	0.19 (0.91)	0.12 (0.87)	0.89 (0.66)	0.09 (1.03)	0.001
Leptin (log odds)	-0.03 (0.92)	-0.07 (0.92)	-0.22 (0.91)	0.24 (0.89)	ns
Ghrelin (log odds)	-0.29 (0.94)	-0.35 (0.96)	0.01 (0.78)	-0.19 (0.94)	ns
Glucose metabolism					
Insulin (mcU/mL)	9.65 (6.88)	9.91 (7.28)	7.45 (4.96)	9.79 (5.92)	ns
Glucose (mmol/L)	5.77 (0.79)	5.80 (0.77)	5.61 (0.98)	5.74 (0.75)	ns
HOMA-IR	2.54 (1.98)	2.64 (2.13)	1.83 (1.19)	2.53 (1.60)	ns
Lipid metabolism					
Total cholesterol (g/L)	1.98 (0.33)	1.96 (0.33)	2.02 (0.37)	2.02 (0.34)	ns
High-Density Lipoprotein (HDL, g/L)	0.56 (0.15)	0.56 (0.16)	0.63 (0.15)	0.54 (0.14)	ns
Low-Density Lipoprotein (LDL, g/L)	1.21 (0.28)	1.19 (0.28)	1.22 (0.29)	1.27 (0.28)	ns
Triglycerides (g/L)	1.04 (0.41)	1.06 (0.42)	0.88 (0.36)	1.04 (0.35)	ns
Total cholesterol/ HDL ratio	3.72 (1.04)	3.74 (1.07)	3.31 (0.79)	3.88 (0.98)	ns

Cardiovascular function					
Systolic pressure (mmHg)	133.25 (16.36)	133.65 (16.43)	129.68 (15.98)	133.64 (16.51)	ns
Diastolic pressure (mmHg) #	73.43 (9.23)	74.43 (9.43)	69.32 (7.71)	71.51 (8.35)	0.014
Heart rate (bpm)	68.77 (9.59)	68.70 (9.86)	69.78 (8.24)	68.48 (10.14)	ns
Homocysteine (μmol/L)	15.06 (8.51)	14.37 (3.57)	14.36 (3.55)	18.44 (19.01)	ns
Renal function					
Albumin (g/L)	43.84 (2.61)	44.05 (2.49)	42.73 (3.05)	43.60 (2.71)	ns
Creatinine (μmol/L) #	77.35 (18.21)	79.09 (17.89)	69.45 (16.98)	74.56 (19.21)	0.022
Uric acid (mg/24 h) #§	312.07 (69.99)	320 .09 (66.57)	254.83 (56.77)	311.51 (76.98)	0.000
Liver function					
Alkaline phosphatase (ALP, μkat/L)	1.24 (0.29)	1.23 (0.30)	1.21 (0.25)	1.30 (0.25)	ns
Gamma-glutamyl transpeptidase (GGT, μkat/L)	0.43 (0.24)	0.42 (0.23)	0.44 (0.35)	0.43 (0.21)	ns
Aspartate aminotransferase (AST, μkat/L)	0.43 (0.10)	0.44 (0.11)	0.44 (0.09)	0.42 (0.08)	ns
Alanine transaminase (ALT, μkat/L)	0.57 (0.15)	0.58 (0.16)	0.57 (0.13)	0.55 (0.11)	ns
Cognitive function					



Cambridge Mental Disorders of the Elderly Examination (CAMDEX), subjective memory score	3.83 (1.96)	3.86 (1.99)	3.85 (2.35)	3.68 (1.61)	ns
Geriatric Depression Scale (GDS) score	2.34 (2.37)	2.19 (2.13)	2.30 (2.54)	3.00 (3.10)	ns
Mini Mental State Examination (MMSE) score	28.44 (1.38)	28.38 (1.39)	28.90 (1.12)	28.41 (1.48)	ns
Post-hoc Wilcoxon test: #G1 vs G2, p value < 0.05; §G2 vs G3, p value < 0.05; ^G1 vs G3, p value < 0.05.					

6.4.2 Beta-diversity of NU-AGE cohort

When examining the beta diversity of microbial communities by PCoA of unweighted UniFrac distances, we identified three distinct groups of individuals, *i.e.*, G1 to G3 (Figure 6.1 a). Within a range of microbiota profiles, these groups represent clusters of subjects who have a significantly different microbiota structure from each other, as demonstrated by the permutation multivariate analysis of variance (permutational test with pseudo-F ratio, R2 = 0.25, p value = 0.0001). The separation of the three groups was further verified by hierarchical clustering with Ward as the linkage method. The stability of the clusters was supported by average Jaccard similarities from 1000 bootstrapping of 0.96 (G1), 0.95 (G2) and 0.92 (G3). The groupings were also evaluated by weighted and generalized UniFrac metrics, by verifying rejection of the null hypothesis (*i.e.*, no difference between the three pre-defined clusters) (p value = 0.0001).

6.4.3 Alpha diversity

The three groups were also found to differ in alpha diversity, with G2 and G3 samples showing the highest values, according to the number of observed Amplicon Sequence Variants (ASVs) and the Chao1 index (Kruskal-Wallis test, p value = 0.05; post-hoc Wilcoxon test, p value = 0.045 for both G1 vs G2, and G1 vs G3). Such values were associated with the microbiome space, meaning that the sample coordinates on the PCoA plot of Figure 6.1 a also mirrored the differences in diversity among the three groups (linear regression analysis, p value = 0.02 for the number of observed ASVs, and p value = 0.01 for Chao1 index). In particular, we observed a gradual increase in diversity along PCo1, from the lowest values in G1 microbiomes to the highest values in the samples belonging to the G2-G3 clusters.

6.4.4 Taxonomic profile at family level

As mentioned above for the entire cohort, the faecal microbiota of all groups was largely dominated by only two families, *Ruminococcaceae* and *Lachnospiraceae*, even if with different proportions (Figure 6.1 b). The relative abundance of *Ruminococcaceae* was in fact significantly greater in G3 compared to G1, while that of *Lachnospiraceae* was lower in G2 compared to G1 (Wilcoxon test, p value ≤ 0.01). The three groups were also found to differ in subdominant families. In particular, G2 showed an enrichment in *Christensenellaceae* and *Rikenellaceae* compared to G1, as well as in *Porphyromonadaceae* compared to G3 (p value ≤ 0.05).

6.4.5 Cohort description

It is important to note that the stratification in the three groups was independent of gender and frailty (pairwise Fisher's exact test, p value > 0.05) as well as age (Kruskal-Wallis test, p value > 0.05) (see Table 6.1 and Figure 6.2). Moreover, as shown in Table 6.1, the three groups were similar for the majority of the measured parameters related to physical function, glucose and lipid metabolism, liver and cognitive function (p value > 0.05). However, the elderly subjects in G2 group compared to G1 and G3 showed significantly lower values for anthropometric measures (*i.e.*, BMI, waist and hip circumference and waist to hip ratio), cardiovascular risk factors (diastolic

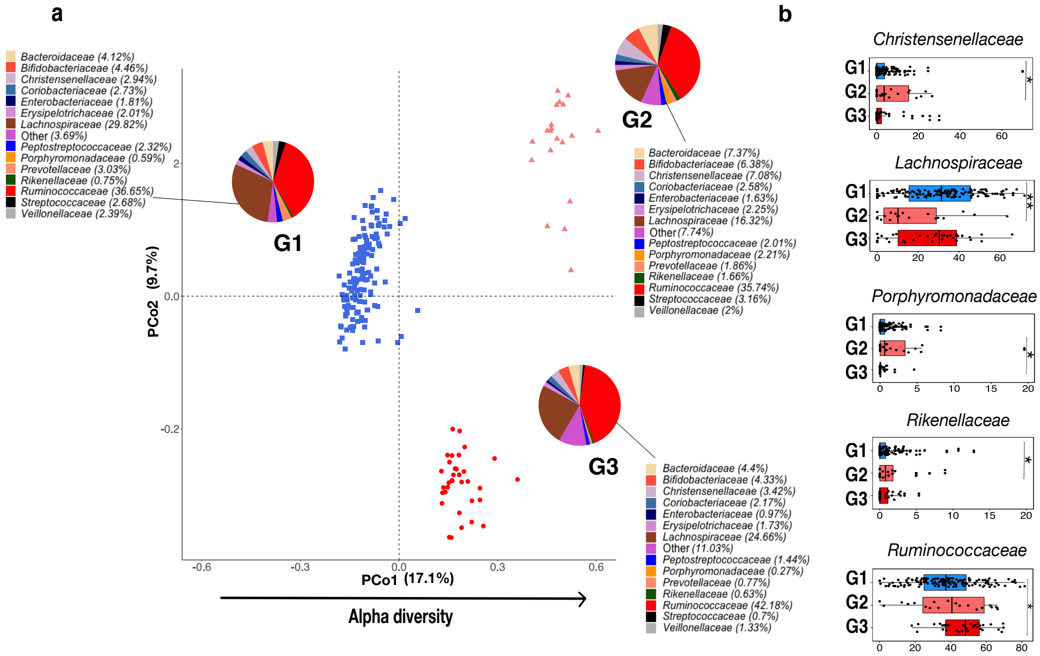


FIGURE 6.1: **Gut microbiome profiles in Italian elders**.
a. Principal Coordinates Analysis (PCoA) plot based on the unweighted UniFrac distances of the faecal microbiota profiles of 201 elderly Italians. Three groups with a significantly different microbial community structure were identified (permutational test with pseudo-F ratio, p value = 0.0001). We refer to them as G1, G2, G3 based on their self-organization in the PCoA space (*i.e.*, left, right upper and lower right quadrant, respectively). The pie charts in the plot summarize the family-level relative abundances in the three groups, considering only the taxa present in at least 20% of the samples (*i.e.*, 40 individuals) with $\geq 0.1\%$ relative abundance. The arrow at the bottom of the PCoA plot represents the alpha diversity gradient, estimated as number of observed ASVs. b. Box-and-whisker plots of relative abundances of families differentially represented among the three microbiome groups. **,* respectively p value < 0.01 and < 0.05, Wilcoxon test.

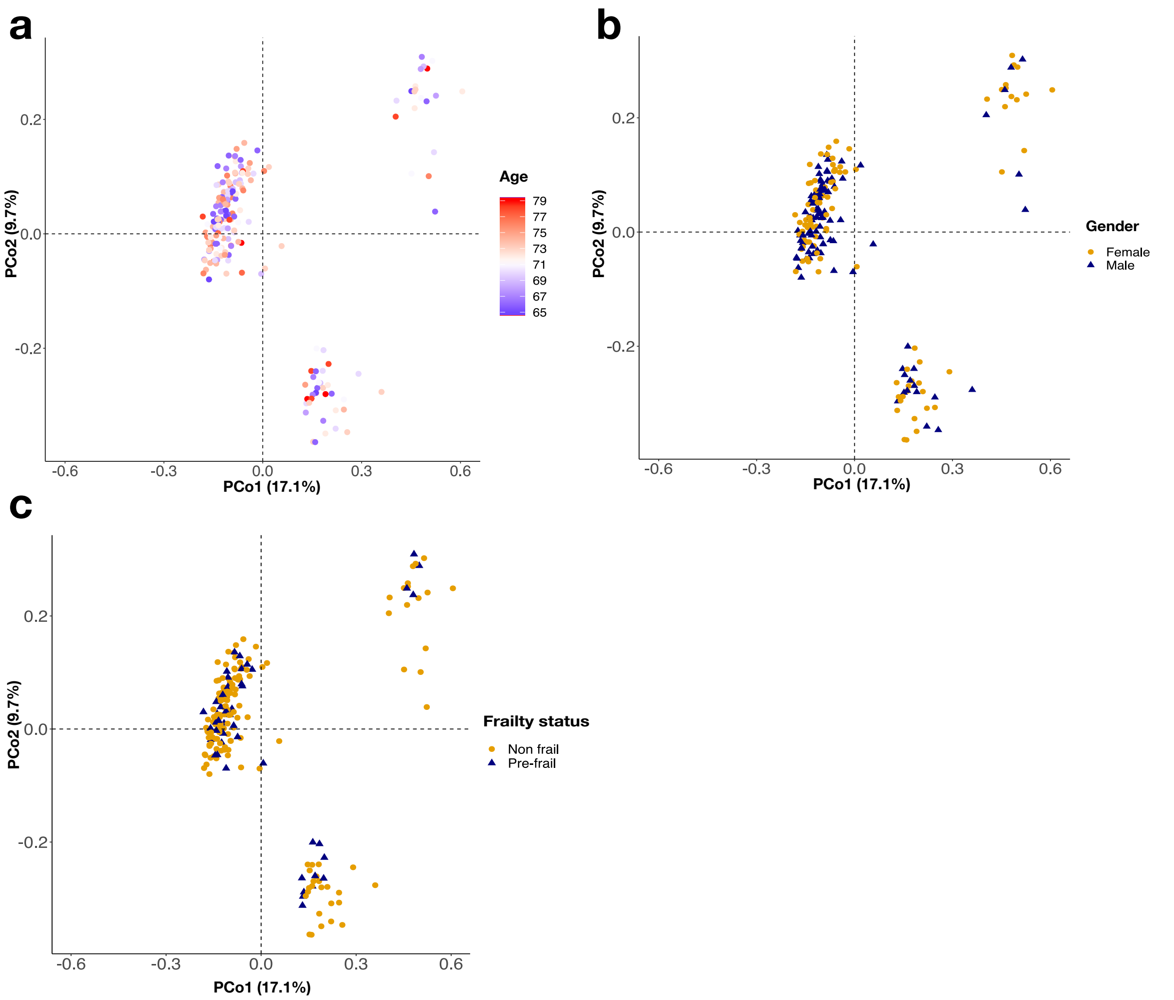


FIGURE 6.2: **Age, gender and frailty status distribution in the three microbiome groups (G1 to G3).**
Unweighted UniFrac-based Principal Coordinates Analysis showing the age (a), gender (b), and frailty.

pressure), renal function markers (creatinine and uric acid) and higher values for adiponectin (an adiposity-related cytokine with anti-inflammatory effects) (post-hoc Wilcoxon test, p value $<$ 0.05) (Table 6.1). The number of subjects taking medicines was similar for antihypertensive drugs (G1: 77.5%, G2: 80.0% and G3: 73.5%) in the three groups, while a higher number of elderly in the G3 group had taken statins (G1: 17.0%, G2: 15.0% and G3: 26.5%), and no elderly in the G2 group had taken anti-diabetic drugs (G1: 8.8%, G2: 0% and G3: 5.9%).

6.4.6 Association between Genera and DXA variables

We then explored the associations between the three microbiome profiles and the DXA-derived body composition variables, with a specific focus on abdominal fat. The elderly in G2 group showed significantly lower levels of VAT than G1 and G3 subjects (p value = 0.003) while no differences were observed with respect to SAT (p value = 0.2) (Figure 6.2 a). Accordingly, the VAT/SAT ratio was significantly lower in G2 compared to G1 and G3 subjects (p value $<$ 0.05) (Figure 6.2 a). Consistent results were also observed for the other DXA variables related to adiposity. Correlations between DXA metadata and relative abundances of genus-level taxa were then specifically sought by means of sPLS regression (Figure 6.2 b and 6.2 c). Genera belonging to *Christensenellaceae* (*Christensenellaceae R7 group*), *Porphyromonadaceae* (*Parabacteroides*) and *Rikenellaceae* (*Alistipes*), *i.e.*, the families found to be enriched in the faecal microbiota of G2 subjects, were inversely associated with a number of DXA variables, including VAT. On the other hand, members of the *Lachnospiraceae* family (*Eubacterium rectale group*, *Fusicatenibacter* and *Blautia*), whose proportions were far lower in the G2 group, were positively correlated with the vast majority of the considered DXA measures, including especially those related to fat mass distribution (*i.e.*, whole body fat mass (FM), whole body fat mass index (FMI), android fat mass to android lean mass ratio (AF/AL), android fat mass to gynoid fat mass ratio (AF/GF) and VAT). Discordant data were instead observed for *Ruminococcaceae*, an overrepresented family in the G3 group, with three genera (*Ruminococcaceae UCG 014, 002, 005*) negatively and three others (*Faecalibacterium*, *Subdoligranulum* positively correlated with most of the adiposity-related DXA variables. Finally, it should be noted that the lean mass parameter SMI (skeletal mass index, i.e., the appendicular lean mass to total body mass ratio) differed from all the others, both for the position in the correlation circle plot and for the direction of correlations. In particular, a direct correlation was observed for *Christensenellaceae R7 group*, as well as for three *Ruminococcaceae* genera (*Ruminococcaceae UCG 014, 002, 005*). On the other hand, consistent with the observations above, *Ruminococcus 2*, *Subdoligranulum* and the *Lachnospiraceae* members *Fusicatenibacter* and *Blautia* negatively correlated with SMI.

6.4.7 Diet

Seven-day food records were used to assess dietary intake. The results were normalized to the body weight of the individuals, facilitating a comparison among the three microbiome groups. The dietary intakes of nutrients for G1, G2 and G3 are shown in Table 6.2. No significant differences were found for the total intake of energy, saturated and unsaturated fatty acids, protein and fibre, and also for the majority of vitamins and calcium, as normalized by body weight. Interestingly, the elderly belonging to the G2 group showed a significantly higher carbohydrate intake (Kruskal-Wallis test, p value = 0.025; post-hoc Wilcoxon test, p value < 0.05 for G1 vs G2 and G2 vs G3) and a trend to higher levels of water (Kruskal-Wallis test, p value = 0.055), b-carotene (p value = 0.052) and vitamin C (p value = 0.054). Moreover, by comparing the average daily intake (normalized to body weight) of different food groups among the three microbiome clusters, the elderly in the G2 group showed a significantly lower intake of potatoes than G1 (0.10 g/day vs 0.27 g/day for G1, post-hoc Wilcoxon test, p value < 0.05) and a trend to higher intake of fruit plus vegetables when compared with G3 (9.41 g/day vs 6.20 g/day for G3; Wilcoxon test, p value = 0.058). A superimposition analysis of the average daily intakes of food groups on the PCoA plot of Figure 6.1 a confirmed an association between the microbiota profile of the G2 group and a lower consumption of potatoes along with cheese (permutational correlation test, p value ≤ 0.025).

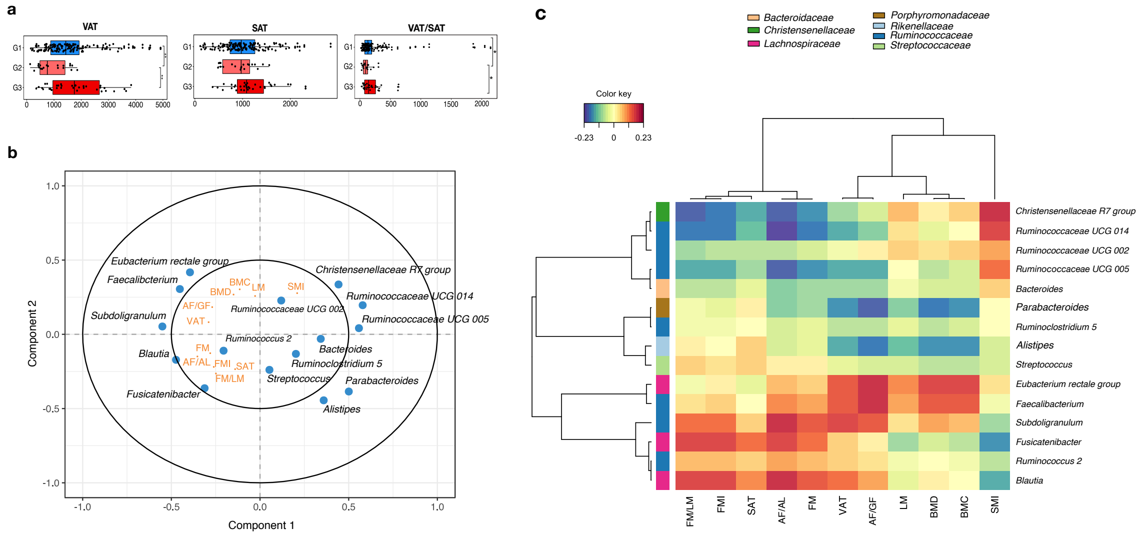
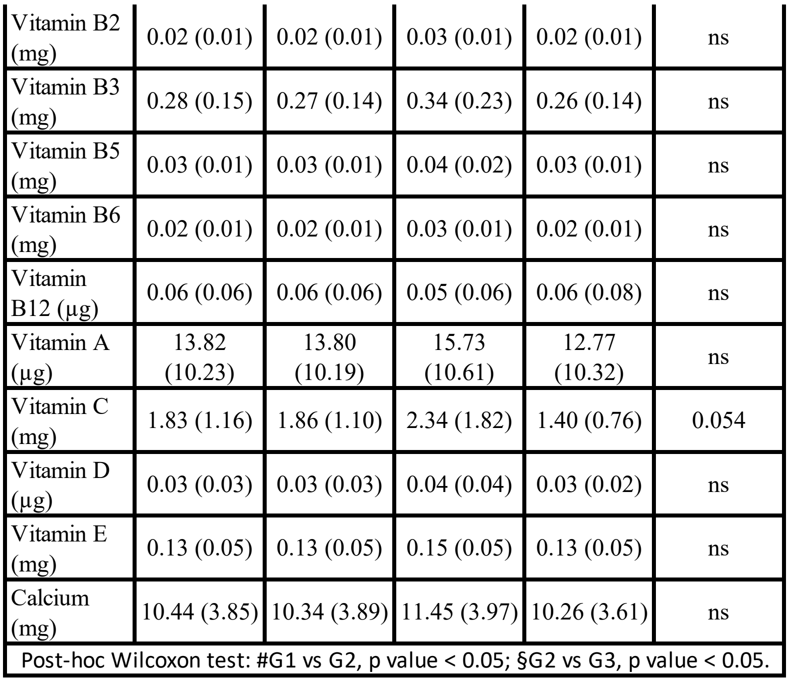


FIGURE 6.3: **Associations between the elderly gut microbiome and body composition.**

a. Box-and-whisker plots of visceral adipose tissue (VAT, g) and subcutaneous adipose tissue (SAT, g) measures, and their ratio, for the three microbiome groups identified by unweighted UniFrac-based PCoA of Figure 6.1a (*i.e.*, G1 to G3). **,* respectively p value < 0.01 and 0.05, Wilcoxon test. b. Sparse partial least square (sPLS) regression of microbial abundances at the genus level and DXA variables. Correlation circle plot for the first two sPLS components with correlations depicted for $<$ -0.2 and $>$ 0.2. The two circumferences show correlation coefficient radii at 0.5 and 1.0. The farther from the centre a bacterial genus or DXA measure is, the greater the association with the component. Variables projected in the same direction of the plot are positively correlated, while variables in diametrically opposite position are negatively correlated. Variables located perpendicular to each other are not correlated. The variance explained by the genera is 10% on the first component and 5% on the second component, while the variance explained by the DXA variables is 37% on the first component and 42% on the second component. c. Hierarchical clustering obtained with complete linkage method and Pearson correlation as distance, was performed on the sPLS regression model retaining the variables shown in the correlation circle plot. For each genus, family level assignment is also shown (see colour legend). The abbreviated names of the DXA variables correspond to whole body fat mass (FM), whole body fat mass index (FMI), whole body fat mass to lean mass ratio (FM/LM), whole body lean mass (LM), appendicular lean mass to total body mass ratio (SMI), whole body bone mineral content (BMC), whole body bone mineral density (BMD), android to gynoid fat mass ratio (AF/GF), android fat mass to lean mass ratio (AF/AL), visceral adipose tissue (VAT), and subcutaneous adipose tissue (SAT).

TABLE 6.2: **Average daily intake of energy and nutrients.** All values were normalized to body weight (kg). Data are shown for the entire cohort as well as for the three microbiome groups (G1 to G3), as identified by PCoA of unweighted UniFrac distances (see Figure 1a). Values are expressed as mean (SD). P values were determined by Kruskal-Wallis test, followed by post-hoc Wilcoxon test. MUFA, monounsaturated fatty acids; PUFA, polyunsaturated fatty acids. ns, not significant.

	All (no.=201)	**G1 (no.=147)**	**G2 (no.=20)**	**G3 (no.=34)**	***p* value**
Total energy (kcal)	24.44 (6.14)	24.25 (5.94)	27.41 (7.54)	23.52 (5.75)	ns
Total carbohydrates (g)#§	3.15 (0.95)	3.11 (0.87)	3.76 (1.23)	2.99 (0.95)	0.025
Total fats (g)	0.87 (0.24)	0.87 (0.24)	0.90 (0.26)	0.87 (0.25)	ns
Total saturated fatty acids (g)	0.27 (0.08)	0.27 (0.08)	0.28 (0.08)	0.27 (0.08)	ns
Total MUFA (g)	0.39 (0.12)	0.38 (0.12)	0.43 (0.15)	0.39 (0.10)	ns
Total PUFA (g)	0.12 (0.05)	0.12 (0.05)	0.12 (0.05)	0.12 (0.06)	ns
omega 3 PUFA (g)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	ns
omega 6 PUFA (g)	0.07 (0.04)	0.07 (0.04)	0.08 (0.03)	0.08 (0.05)	ns
Total proteins (g)	0.95 (0.22)	0.95 (0.22)	1.00 (0.24)	0.90 (0.21)	ns
Animal proteins (g)	0.46 (0.14)	0.46 (0.15)	0.45 (0.13)	0.44 (0.13)	ns
Vegetal proteins (g)	0.35 (0.14)	0.35 (0.15)	0.39 (0.16)	0.32 (0.11)	ns
Total dietary fiber (g)	0.31 (0.15)	0.31 (0.15)	0.34 (0.16)	0.28 (0.14)	ns
Starch (g)	1.43 (0.55)	1.42 (0.56)	1.64 (0.61)	1.31 (0.45)	ns
Cholesterol (g)	2.89 (0.97)	2.91 (0.99)	2.96 (1.10)	2.77 (0.84)	ns
Water (g)	26.37 (9.41)	0.23 (0.99)	32.45 (12.19)	25.06 (7.93)	0.055
Biotin (mg)	0.25 (0.16)	0.26 (0.16)	0.23 (0.08)	0.24 (0.17)	ns
Folic acid (µg)	3.97 (1.76)	3.99 (1.68)	4.84 (2.53)	3.40 (1.36)	ns
b-carotene (µg)	25.10 (19.27)	24.72 (19.38)	34.41 (24.41)	21.26 (13.30)	0.052
Vitamin B1 (mg)	0.01 (0.01)	0.01 (0.01)	0.02 (0.01)	0.01 (0.01)	ns



Vitamin B2 (mg)	0.02 (0.01)	0.02 (0.01)	0.03 (0.01)	0.02 (0.01)	ns
Vitamin B3 (mg)	0.28 (0.15)	0.27 (0.14)	0.34 (0.23)	0.26 (0.14)	ns
Vitamin B5 (mg)	0.03 (0.01)	0.03 (0.01)	0.04 (0.02)	0.03 (0.01)	ns
Vitamin B6 (mg)	0.02 (0.01)	0.02 (0.01)	0.03 (0.01)	0.02 (0.01)	ns
Vitamin B12 (μg)	0.06 (0.06)	0.06 (0.06)	0.05 (0.06)	0.06 (0.08)	ns
Vitamin A (μg)	13.82 (10.23)	13.80 (10.19)	15.73 (10.61)	12.77 (10.32)	ns
Vitamin C (mg)	1.83 (1.16)	1.86 (1.10)	2.34 (1.82)	1.40 (0.76)	0.054
Vitamin D (μg)	0.03 (0.03)	0.03 (0.03)	0.04 (0.04)	0.03 (0.02)	ns
Vitamin E (mg)	0.13 (0.05)	0.13 (0.05)	0.15 (0.05)	0.13 (0.05)	ns
Calcium (mg)	10.44 (3.85)	10.34 (3.89)	11.45 (3.97)	10.26 (3.61)	ns

Post-hoc Wilcoxon test: #G1 vs G2, p value < 0.05; §G2 vs G3, p value < 0.05.

6.4.8 Metabolomics data

Finally, serum metabolomics data were analysed in order to find out metabolites that discriminated gut microbiome structures. The G2 group was characterized by significantly lower circulating levels of some serum metabolites, including mineral (sulphur), BCAAs (isoleucine, leucine and valine), fatty acids (myristic acid - C14:0) and methyl ester fatty acids (methyl-hexadecenoic acid, and tetramethyl-dihydroxy-octadecahexaenoic acid) (post-hoc Wilcoxon test for G2 vs G1 and/or G3, p value *leq* 0.05), and a tendency to lower levels of primary bile acid, chenodeoxycholic acid (p value > 0.05). On the other hand, methyl-heptadecadienoic acid was found to be significantly higher in G2 compared to G3, but also in G1 compared to G3 (p value *leq* 0.05). Based on a sPLS regression analysis of relative abundances at the genus level and metabolites (Fig. 6.4), genera belonging to the families identified as signatures of the G2 group, *i.e.*, *Christensenellaceae R7 group*, *Alistipes* and *Parabacteroides*, inversely correlated with BCAAs, while some *Lachnospiraceae* and *Ruminococcaceae* members, distinctive of the G1 and G3 profiles, showed opposite correlations. Similarly, for fatty acids, we observed negative correlations between *Christensenellaceae R7 group* or *Alistipes* and the majority of systemic fatty acids retained in the sPLS regression model. In contrast, mostly positive correlations were found with *Lachnospiraceae* taxa, along with *Bifidobacterium*, *Streptococcus*, *Ruminococcus 1* and *Erysipelotrichaceae UCG 003*.

6.4.9 Discussion

In the present study, we identified three significantly different groups (G1 to G3) of elderly Italian individuals harbouring distinct gut microbiome structures, which correlate with body composition and other health-related parameters. In particular, the G1 group was characterized by higher abundance of *Lachnospiraceae*, the G2 group was enriched in *Christensenellaceae*, *Porphyromonadaceae* and *Rikenellaceae*, and G3 in *Ruminococcaceae*. The three profiles were also characterized by different biodiversity, with G2 and G3 showing the highest level followed by G1. When we explored the connections between the gut microbiome and body composition, we found that the G2 microbiome cluster had the lowest median value of VAT, a specific measure of abdominal obesity.

Unlike a recent study in a Chinese adult population, which reported a sex-specific association between the gut microbiome layout and fat distribution, using DXA data for android and gynoid fat, the microbial communities defining the three elderly groups in our work are neither sex-related nor age-driven (Min et al., 2019). However, it should be noted that our dataset is constrained to old age and characterized by a lower range of android/gynoid to whole fat mass ratio (the lowest and highest value in our cohort, 4.1 and 21.9, respectively; in the Chinese cohort, 6.6 and 26.6, respectively).

In line with the available literature, the microbial footprints of the G2 group (*i.e.*, the greater proportion of *Christensenellaceae*, *Porphyromonadaceae* and *Rikenellaceae*) could contribute to a reduced amount of visceral fat mass (Beaumont et al., 2016; Tamura et al., 2017). Indeed, the family *Christensenellaceae* has been consistently reported as negatively related to visceral fat mass and indicated as a marker of lean phenotype, (Beaumont et al., 2016; Tamura et al., 2017; Goodrich et al., 2014) as also shown by our sPLS regression. On the other hand, *Porphyromonadaceae* and *Rikenellaceae* members, both belonging to the Bacteroidetes phylum, could play a role

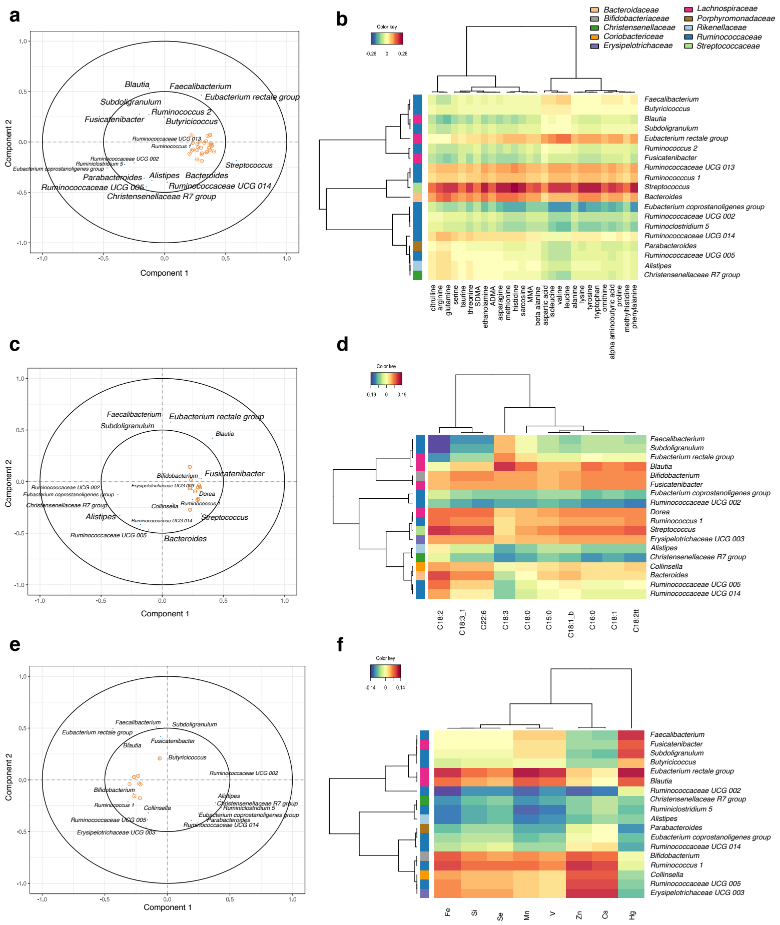


FIGURE 6.4: **Associations between the elderly gut microbiome and the serum metabolome.**

Sparse partial least square (sPLS) regression of microbial abundances at the genus level and metabolites (a-b, amino acids; c-d, fatty acids; e-f, minerals). Left, Correlation circle plot for the first two sPLS components. The two circumferences show correlation coefficient radii at 0.5 and 1.0. The farther from the centre a bacterial genus or metabolite is, the greater the association with the component. Variables projected in the same direction of the plot are positively correlated, while variables in diametrically opposite position are negatively correlated. Variables located perpendicular to each other are not correlated. The variance explained by the genera is 10% on the first component and 5% on the second component. The variance explained by the metabolites is as follows: amino acids, 49% on component 1 and 30% on component 2; fatty acids, 34% on component 1 and 9% on component 2; minerals, 11% on component 1 and 7% on component 2. Right, Hierarchical clustering obtained with complete linkage method and Pearson correlation as distance, was performed on the sPLS regression model retaining the variables shown in the correlation circle plot. For each genus, family level assignment is also shown (see colour legend).

as adiposity modulators through the production of the SCFAs acetate and propionate (Lu et al., 2016). Specifically, it has been shown that acetate contributes to adiposity reduction in mice, by upregulating the genes involved in fatty acid oxidation in the liver (Kondo et al., 2009). Furthermore, the abundances of *Christensenellaceae* and *Rikenellaceae* have recently been found to be highly correlated with each other and significantly higher in lean than obese subjects (Oki et al., 2016).

Conversely, the family *Lachnospiraceae*, found to be distinctive of the low-diverse, higher VAT-related microbiome profile G1, has been connected to dietary lipid metabolism, and genera belonging to this family, e.g. *Blautia*, have been associated with higher amounts of VAT, (Beaumont et al., 2016; Just et al., 2018; Ozato et al., 2019), as in our correlation analysis. On the other hand, for members of *Ruminococcaceae* (marker of the high-diverse but VAT-related G3 group) we found conflicting trends of association with visceral fat, which are however generally consistent with what is available in the literature. For example, *Faecalibacterium*, which in our cohort positively correlated with VAT, has sometimes been found at increased levels in obese subjects, despite the known anti-inflammatory and immunomodulating properties, probably due to its ability to increase energy harvesting from otherwise unabsorbable carbohydrates (Balamurugan et al., 2010; Del Chierico et al., 2018).

In contrast, the genera *Ruminococcaceae UCG 014* and *Ruminococcaceae UCG 005* have both been negatively associated with adiposity (Wutthi-in et al., 2020; Zhao et al., 2017), in line with our sPLS regression. It is also worth noting that *Subdoligranulum*, which in our dataset showed positive associations with all DXA-related variables considered except SMI (a lean mass parameter), has recently been identified as one of the few key species associated with both faecal and blood metabolic profiles, therefore likely to play a major role in the gut-systemic metabolic interplay (Visconti et al., 2019).

Interestingly, while the entire cohort is composed of apparently healthy elderly subjects with almost all risk parameters in their normal range, the elderly in the G2 group, compared to G1 and G3, have a significantly lower level of several anthropometric, metabolic, cardiovascular and renal risk factors, such as BMI, waist and hip circumference and waist to hip ratio, diastolic pressure, creatinine and uric acid, and higher levels of adiponectin, an adipose-related cytokine with anti-inflammatory effects. These findings are of particular interest, also because *Christensenellaceae*, specifically enriched in G2 subjects, are considered an important component of the gut microbiome of centenarians and semi-supercentenarians, and of a healthier profile in general, thus potentially representing a marker of healthy ageing and longevity since the early old age (60-70 years) (Biagi et al., 2016; Waters and Ley, 2019; Le Roy et al., 2019).

Furthermore, compared to G1 and G3 groups, the elderly in G2 showed lower serum levels of BCAAs, which are known to be associated with insulin-deficient and -resistant disorders and have already been shown to correlate positively with VAT (Holeček, 2018; Rietman et al., 2016; Lackey et al., 2013).

The G2 group was also characterized by lower circulating levels of methyl ester fatty acids and myristic acid, for which an inverse association with HDL cholesterol levels was demonstrated in an Italian population following a Mediterranean diet (Noto et al., 2016). Although cholesterol levels are not discrete in the three groups, the mean values are lower for G2. As expected, both fatty acid and BCAA levels were found to inversely correlate with genera belonging to the families identified as signatures of the G2 group, *i.e.*, *Christensenellaceae R7 group*, *Alistipes* and *Parabacteroides*, while positively with some *Lachnospiraceae* and *Ruminococcaceae* members, distinctive of the G1 and G3 profiles. It is also worth noting that the elderly in the G2 group showed a tendency to lower levels of chenodeoxycholic acid, whose impact on cholesterol metabolism is not yet conclusive but could be unfavorable (Porez et al., 2012). This could suggest a greater capacity of production of secondary bile acids by the G2-related gut microbiota. Although this ability must be verified by appropriate methods, including metagenomics, metatranscriptomics and, not least, stool metabolomics, previous reports have shown that G2 discriminating taxa, especially Bacteroidetes members, are capable of deconjugation and metabolism of primary bile acids into secondary ones (Hirano and Masuda, 1982; Ishii et al., 2014; Gu et al., 2017; Yao et al., 2018). Further strengthening this hypothesis, a strong positive correlation has recently been found between secondary bile acid metabolism and *Christensenellaceae*, another distinctive taxon of the G2 profile (Alemán et al., 2018). Based on a search in the PFAM and NCBI database, *Christensenellaceae* species actually exhibit both bile salt hydrolase (EC 3.5.1.24) and bile-acid 7-alpha-dehydratase (EC 4.2.1.106) activity, participating in the 7-dehydroxylation process associated with bile acid degradation (El-Gebali et al., 2019; Sayers et al., 2010).

Consistent with the above assumptions of better metabolic health for the elderly in the G2 group, their dietary pattern was also healthier, with lower consumption of potatoes and a trend to higher average daily intake of fruit and vegetables than the other groups. Interestingly, it has been demonstrated that increasing the consumption of fruit and vegetables and reducing the intake of potatoes can reduce the risk of ischemic stroke (Hansen et al., 2020). Furthermore, the G2-related microbiota profile was found to be associated with a lower intake of cheese, another well-known product to increase cardiovascular risk through adiposity and lipid pathways (Trichia et al., 2020).

In summary, here we advance the hypothesis that distinctive high-diverse structures of the gut microbiome of the elderly may contribute to a reduced amount of VAT. In particular, our results suggest the relevance of high amounts of *Christensenellaceae*, *Porphyromonadaceae* and *Rikenellaceae* as protective of cardiovascular and metabolic diseases related to visceral fat and, thus, potential markers of healthy ageing and, possibly, longevity. This hypothesis is supported by a healthier dietary intake and metabolic profile, and overall better health for the elderly harbouring this microbial layout. We can therefore argue that favourable compositions of the gut microbiota of older people could contribute to reduce metaflammation, a specific metabolism-induced inflammation, mostly overlapping with inflammaging, triggering obesity-induced insulin resistance and type 2 diabetes (Franceschi et al., 2018).

Further studies in larger cohorts, possibly from different geographical locations, via shotgun metagenomics combined with metabolomics, will be needed to confirm our findings and provide insights on the mechanisms underlying the relationship between gut microbes and VAT, and their role in modulating adiposity and promoting a healthy life. Such mechanisms should possibly be validated in an animal model. Similarly, additional work, possibly also through culturomics approaches, is required to better understand the dynamics and ecological rules within the gut microbiota that lead to the establishment of different networks. It is reasonable to expect that in the near future the targeted manipulation of the elderly intestinal microbiota, the feasibility of which has been recently demonstrated in the context of NU-AGE

(Ghosh et al., 2020), will become an integral component of current strategies aimed at contrasting age-related deterioration in body composition and multiple bodily functions, thus supporting healthy ageing.

6.4.10 Author contributions

Aurelia Santoro and Simone Rampelli, conceptualization; Aurelia Santoro, Simone Rampelli and Silvia Turroni, project administration; Patriza Brigidi and Claudio Franceschi, resources; Monica Barone and Silvia Turroni, library preparation and sequencing; Teresa Tavella and Simone Rampelli, bioinformatic and biostatistic analysis; Alberto Bazzocchi Giuseppe Battista and Chiara Gasperini, DXA analysis; Claudio Nicoletti, Fawzi Kadi and Paul W. O'Toole collected and contributed to data; Aurelia Santoro, Stefano Salvioli and Giulia Guidarelli, analysis of demographic, biochemical, nutritional and other health-related data; Estelle Pujos-Guillot and Blandine Comte, serum metabolomics; Teresa Tavella, Simone Rampelli, Silvia Turroni and Aurelia Santoro, writing - original draft; Estelle Pujos-Guillot and Blandine Comte, writing - review and editing. All authors discussed the results and commented on the manuscript.

6.4.11 Acknowledgments

The NU-AGE project was supported by the European Union's Seventh Framework Programme under grant agreement no. 266486 ('NU-AGE: New dietary strategies addressing the specific needs of the elderly population for healthy ageing in Europe').

Chapter 7

Longevity and antibiotic resistance. The human gut resistome up to extreme longevity

7.1 Abstract

Antibiotic resistance (AR) is indisputably a major health threat, which has drawn much attention in recent years. In particular, the gut microbiome has been shown to act as a pool of AR genes, potentially available to be transferred to opportunistic pathogens. Herein, we investigated for the first time, changes in the human gut resistome across ageing, up to extreme longevity, including an exceptional cohort of individuals aged till 109 years. According to our findings, some AR genes are similarly represented in all subjects regardless of age, potentially forming part of the core resistome. Interestingly, ageing was found to be associated with an overall higher burden of AR genes, including especially proteobacterial genes encoding multidrug efflux pumps. Our results warn of possible health implications and pave the way for further investigations aimed at containing AR accumulation, with the ultimate goal of promoting healthy ageing.

Keywords: ageing; extreme longevity; metagenome; microbiome; antibiotic resistance, resistome

7.2 Introduction

Infection rates with antibiotic-resistant microorganisms continue to rise worldwide (Hashiguchi et al., 2019). In Italy, 2015 data showed that over 30% of infections were caused by bacteria resistant to antimicrobial treatment for eight priority antibiotic-bacterium combinations, with more than 10,000 of 33,000 attributable deaths in Europe per year (Cassini et al., 2019). Antibiotic resistance (AR) is therefore considered a critical public health threat (OECD, 2018). Antibiotics use and abuse are a major cause of spread and increase in AR (WHO, 2014), together with their routine usage in the food industry, for both the veterinary sector and agriculture, as they are essential in controlling the state of infestation (Hendriksen et al., 2019; Munk et al., 2018).

An important evidence of the impact of antibiotic misuse/abuse and environmental exposure on the development of AR was provided by the comparison of the gut resistome (*i.e.*, the set of genes/proteins conferring AR in the gut microbiome) of Western populations with that of traditional communities (Rampelli et al., 2015; Clemente et al., 2015b). Interestingly, while some AR genes are shared among all sampled populations regardless of lifestyle, natural environments being the first unquestionable reservoir of AR (Allen et al., 2010), the Western-type resistome pattern supports a "farm to fork" aetiology of resistance transmission. In other words, the habitual use of antibiotics in food production and medicine in the Western world strongly affects the AR profiles of the gut microbiome, favouring the emergence of new resistances (not limited to the antibiotics to which we are exposed) and boosting their expansion through horizontal gene transfer (Rampelli et al., 2015; Forslund et al., 2014; Huddleston, 2014). This has profound implications for health, because the acquisition of AR genes by the gut microbiome may also involve pathobionts, *i.e.*, minor microbiome components with pathogenic potential, which can cause or promote disease under certain circumstances (Escudeiro et al., 2019; Francino, 2016; Jochum and Stecher, 2020).

The relevance of antibiotic-resistant gut bacteria as an immediate and long-lasting threat to human health is well recognized, especially in compromised individuals such as preterm infants receiving early-life antibiotics (Gasparrini et al., 2019), elderly people with a debilitated immune system (Araos et al., 2019) and patients with cancer or autoimmune disorders (D'Amico et al., 2019). However, little information is available to date on the variation of the human intestinal resistome along healthy ageing.

In an attempt to bridge this gap, here we profiled the human gut resistome of an exceptional cohort of semi-supercentenarians, *i.e.*, extremely long-lived individuals over the age of 105, compared to young adults, elderly and centenarians. Specifically, we characterized type and target of resistances, and related bacterial taxa. In addition to providing a fine characterization of antibiotic resistance within the gut microbiome at distinct times of life, our study emphasizes a progressive age-related burden in AR genes assigned to potential pathobionts, possibly as a result of life-long exposure to antibiotics.

7.3 Materials and Methods

7.3.1 Subjects and study groups

Herein, we analysed shotgun metagenomics reads of 62 faecal samples from Italian subjects, generated in (Rampelli et al., 2020).All subjects were enrolled from the same geographic area (Emilia Romagna region, Italy). The subjects' age ranged from 22 to 109 years, with an average age of 85 years. In line with previous studies (Biagi et al., 2016; Rampelli et al., 2020), subjects were stratified into four age groups: semi-supercentenarians over 105 (group S: 23 subjects), centenarians aged 99 to 105 (group C: 15 subjects), elderly individuals aged 65 to 98 (group E: 13 subjects), and younger adults aged 22 to 48 (group Y: 11 subjects).

7.3.2 Quality assessment: reads pre-processing, quality filtering and contaminant removal

For DNA extraction and library preparation, please refer to Rampelli et al. (2020) (Rampelli et al., 2020). Sequencing data is available at the SRA repository (`https://www.ncbi.nlm.nih.gov/bioproject/PRJNA553191`). Reads were in silico depleted of host DNA, using the NCBI Homo sapiens assembly19 as a reference genome, identified with bmtagger software (`ftp://ftp.ncbi.nlm.nih.gov/pub/agarwala/bmtagger`) and removed with BBMap tool (http://sourceforge.net/projects/bbmap/). Raw reads were processed with Trimmomatic (Bolger, Lohse, and Usadel, 2014) for adapter removal (Nextera adapters) and quality trimming. Reads were scanned by evaluating the sequences over a 4-base sliding window and setting the average quality score below Q20 for trimming. We set Trimmomatic parameters enabling reads dropping if their length was less than 35 bp. Moreover, PCR duplicates were estimated and removed with the Picard tool EstimatedLibraryComplexity (version used by the International Human Microbiome Standards project and described in their standard operating procedures). The quality of the reads was inspected before and after the pre-processing steps (FastQC) (Andrews, 2010).

7.3.3 Bioinformatics and statistical analysis

The taxonomic classification of high-quality reads was performed with Kaiju (Menzel, Ng, and Krogh, 2016) (version 1.6.3, greedy algorithm) using as a reference NCBI RefSeq (March 2019 release). On the other hand, we identified resistance proteins in our dataset with Diamond (Buchfink, Xie, and Huson, 2015), using the taxonomically assigned reads and the Deeparg dataset (Arango-Argoty et al., 2018) (14,957 entries) as a reference, which contains Swissprot, Trembl, (Bateman et al., 2017), CARD (Jia et al., 2017) and ARDB (Liu and Pop, 2009). Antibiotic-resistant protein families were curated from Antibiotic Resistance Ontology (ARO) obtained by CARD database. The member of each read couple showing at least 35% sequence identity and 80% read coverage against hit sequence and e-value $<$ 0.001, was annotated as the respective mapped protein in the Deeparg dataset. A table of counts was generated, summarizing the read counts per million (CPM) for each identified protein, normalized by sequencing depth. Similarly, the taxa-level CPM were computed. The taxonomic classification of reads assigned to resistant determinants was summarized at phylum, family, genus and species level, retaining only taxa with a relative abundance of at least 0.1% in 20% of the dataset. The Kruskal-Wallis test followed by Wilcoxon test was adopted to test for differences in relative abundance between the four age groups. Bonferroni correction for multiple testing was applied. A corrected p value ≤ 0.05 was considered statistically significant.

The PCoAs were obtained in R (R Core Team, 2013) with the function cmdscale (vegan package) (Oksanen et al., 2018) and the ordination was computed with the Bray-Curtis dissimilarities. Age group-specific antibiotic resistance determinants (ARDs) were identified by Wald test as implemented in the DEseq2 package (Love, Huber, and Anders, 2014) in R, on the AR counts mapped per sample. Hits with less than 10 reads across the dataset were removed from the final table. Differences were assessed for each pair of groups, considering as differentially abundant proteins with log2 fold change $\leq$ -2 and $\geq$ 2 and a p value threshold of 0.05 corrected with Bonferroni. The hierarchical clustering, showing the ARD relative abundances in the samples (rlog function in DEseq2), was computed with the Ward linkage method over the Euclidean distances (pheatmap package) (Kolde, 2019).

Correlation analysis was performed between ARDs and the taxonomic profile of the gut resistome at the species level, retaining only species with a minimum relative abundance of 0.1% in 20% of the dataset. The table of correlations, obtained with Spearman method (hmisc package) (Harrell Jr, Charles Dupont, and others., 2019), retaining only associations with rho greater than 0.8 and adjusted p values of 0.001

(Bonferroni method), was visualized as a network (R package igraph, gephi v. 0.9.2) (Bastian, Heymann, and Jacomy, 2009).

7.4 Results

We previously identified considerable taxonomic and functional variability in the gut microbiome structure of 62 individuals, including 11 young adults (Y, mean age: 32 years), 13 younger elderly (E, mean age: 73 years), 15 centenarians (C, mean age: 100 years), and 23 semi-supercentenarians (S, mean age: 106 years) (Rampelli et al., 2020). Here, we profiled their gut resistome by analysing 1 billion metagenomic reads from shotgun sequencing of faecal samples, with an average of 16 million (± 4 million SD) reads per subject. The reads were mapped to a collection of AR proteins, which summarize and organize the resistance databases of Uniprot, CARD and ARDB (see Materials and Methods). A total of 1,746 proteins were returned as best hits (from 12,567,041 mapped reads with an average of 202,694 ± 60,145 SD reads per sample), then reduced to 377 after eliminating those with less than 10 counts and filtering for subject prevalence of at least 40%.

7.4.1 The resistome taxonomic composition

The four age groups show separation in the Principal Coordinates Analysis (PCoA) based on the Bray-Curtis distance between the taxonomic profiles of the gut resistome at both family and genus level (p value = 0.0015 and 0.001, respectively, permutation test with pseudo-F ratios) (Figure 7.1a and d). The family-level core structure of the faecal resistome is dominated by a few taxa that normally abound in the human gut microbiota, *i.e.*, *Lachnospiraceae* (mean Counts per Million (CPM) in percentage per group ± SD, 25.8% ± 9.8% in Y, 28.2% ± 13.7% in E, 17.9% ± 8.9% in C and 17.6% ± 13.6% in S) and *Ruminococcaceae* (15.2% ± 7.2% in Y, 13.1% ± 5.6% in E, 12.6% ± 7.0% in C, 8.5% ± 6.7% in S), as well as *Bifidobacteriaceae* (16% ± 12.4% in Y, 8.9% ± 12.8% in E, 13.0% ± 11.9% in C, 20.5% ± 19.5% in S) (Figure 7.1b). Compared to younger individuals, extremely long-lived people show decreased contribution of AR reads assigned to *Bacteroidaceae*, *Eubacteriaceae*, *Prevotellaceae* and *Veillonellaceae*, along with a progressive increase in AR reads assigned to *Enterobacteriaceae* (p value *leq* 0.01, Kruskal-Wallis test) (Figure 7.1c). The age-related decrease in AR reads assigned to Bacteroidetes members was already evident at the phylum level (p value = 0.018) (Supplementary Figure S4).

At the genus level, the core resistome structure is essentially dominated by *Bifidobacterium* (18.4% ± 13.4% in Y, 11.0% ± 15.8% in E, 15.1% ± 13.0% in C, 22.5% ± 21.9% in S), *Faecalibacterium* (13.9% ± 7.8% in Y, 10.0% ± 6.6% in E, 7.7% ± 4.9% in C, 6.1% ± 7.8% in S) and *Collinsella* (7.4% ± 5.2% in Y, 3.2% ± 2.2% in E, 4.4% ± 4.6% in C, 4.7% ± 5.6% in S), but with a decreased proportion of Faecalibacterium-assigned AR reads along with ageing (S vs Y, p value = 0.02, Wilcoxon test) (Figure 7.1e and f). Compared to younger subjects, the semi-supercentenarian group also shows reduced contribution of AR reads assigned to other typically health-associated, short-chain fatty acid-producing taxa, *i.e.*, *Roseburia* and *Ruminococcus*, as well as to Eubacterium, and Megasphaera (p value ≤ 0.03). On the other hand, the faecal resistome of extremely long-lived people is enriched in AR reads assigned to *Eggerthella* (p value = 0.01) (Figure 7.1f).

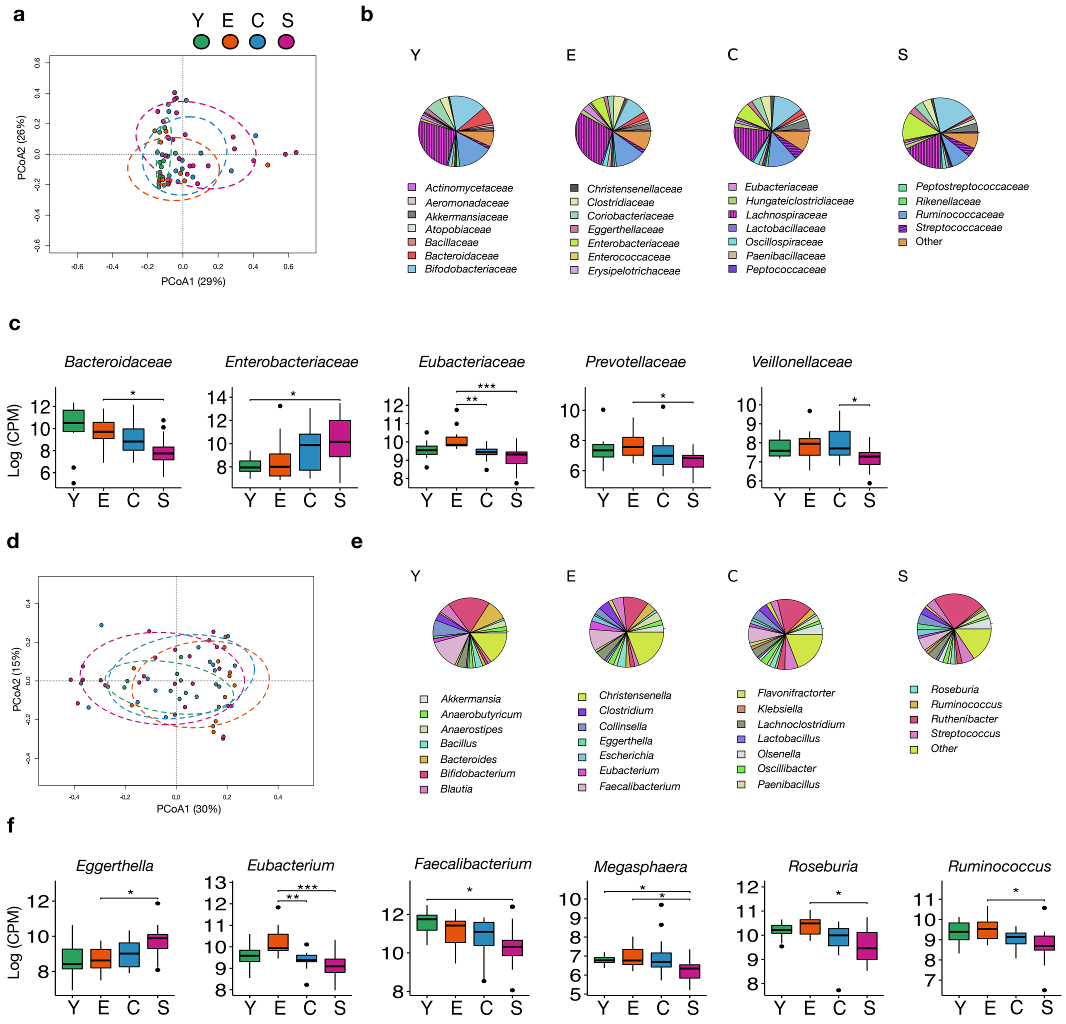


FIGURE 7.1: **Age-related variation in the taxonomic composition of the human gut resistome.** Principal Coordinates Analysis of the Bray-Curtis dissimilarities between the family (**a**) and genus-level (**d**) profiles of the gut resistome of young adults (Y), younger elderly (E), centenarians (C) and semi-supercentenarians (S). Significant separation was found at both phylogenetic levels (p value = 0.0015 and p value = 0.001 respectively family and genus, permutation test with pseudo-F ratios). Pie charts of the 25 most abundant families (**b**) and genera (**e**) for each group (Y, E, C, S). Boxplots showing the relative abundance distribution of significantly differentially represented families (**c**) and genera (**f**) between age groups (Wilcoxon test, Bonferroni corrected p value). Statistical significance is reported as "***", "**", "*", corresponding to the following p value thresholds 0.0001, 0.001, 0.05.

7.4.2 Mechanism of resistance annotated in the gut resistome

Consistent with taxonomic data, the Bray-Curtis PCoA on protein-level resistome profiles provides evidence of an age-related trajectory (p value= 0.012, permutation test with pseudo-F ratios) (Figure 7.2a), suggesting the presence of age group-specific AR determinants (ARDs). As for the resistome profiling, a summary of the results in terms of mechanisms of resistance (using the Antibiotic Resistance Ontology – ARO) and antibiotics is shown in Figure 7.2b. In particular, antibiotic efflux (ARO:0010000) constitutes the prevailing resistance mechanism accounting for 44.1% of the mechanisms identified. Alteration of the antibiotic target (ARO:0001001) is the second most represented mechanism (29.1%), including both mutations and enzymatic modification of the target site. Finally, antibiotic inactivation by bacterial enzymes (ARO:0001004) is the third most represented mechanism, accounting for 12.3% of the total. A core set of 8 ARDs, with mean relative abundance above 2% across all groups, was then identified (Figure 7.2c). In particular, 6 core ARDs encode for ATP-binding cassette (ABC) antibiotic efflux pumps (macB, bcrA, efrA, efrB, sav1866, msbA), 1 for a quinolone resistance protein (mfd) and 1 for an isoleucyl-tRNA synthetase (ileS) conferring resistance to mupirocin.

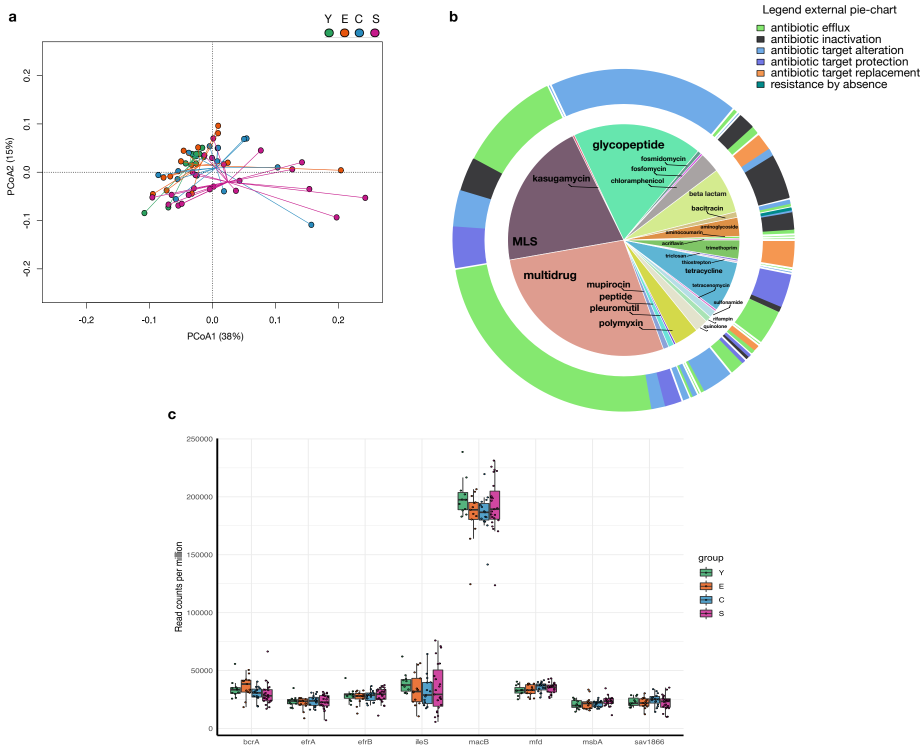


FIGURE 7.2: **The human gut resistome across ageing**. (**a**) Principal Co-ordinates Analysis of the Bray-Curtis dissimilarities between the gut resistome profiles of young adults (Y), younger elderly (E), centenarians (C) and semi-supercentenarians (S). A significant separation was found (p value = 0.012, permutation test with pseudo-F ratios). (**b**) Hierarchical pie plot, the external pie chart depicts the resistance mechanisms of the whole gut-resistome dataset, while the internal donut recalls the type of antibiotics for each mechanism, annotation obtained from the Antibiotic Resistance Ontology. (**c**) The core human gut resistome consists of 8 antibiotic resistance determinants with mean relative abundance above 2% across all age groups.

7.4.3 The resistome profile of the cohort

A total of 39 age group-specific ARDs were next identified, intended as AR determinants that were differentially represented in at least one comparison between two age groups (with log2 fold change $\leq$ -2 and $\geq$ 2) (Figure 7.3). Twenty-nine of these discriminating ARDs were annotated as coding for efflux pumps, 7 as antibiotic-inactivating, 2 as antibiotic target-modifying, and 1 as target-protecting. The full list of age group-specific ARDs, including a description of their mechanisms, is reported in Supplementary Table S2. Interestingly, compared to the younger group, the faecal resistome of elderly, centenarians and semi-supercentenarians shows an overall higher amount of ARDs for sulfonamide (leuO) and multidrug, particularly bcr, emrD, emrY, mdfA, mdtG, mdtL, robA and tolC (p value ≤ 0.03, Wald test) (Figure 3). Seven multidrug ARDs are specifically discriminatory for semi-supercentenarians (cpxA, mdtA, mdtB, mdtC, mdtD, mdtK and mdtN) (p value ≤ 0.008). Furthermore, the semi-supercentenarian group shows higher levels of ARDs conferring resistance to rifampin (rphB) and tetracycline (tcr3 and tetD) (p value ≤ 0.02). On the other hand, compared to semi-supercentenarians, the resistome of young adults is enriched in ARDs for beta-lactam antibiotics (Bl2e_cepa, cblA-1, OXA-34), whereas that of elderly subjects in ARDs for macrolide-lincosamide-streptogramin (ermB, ermF) (p value ≤ 0.05).

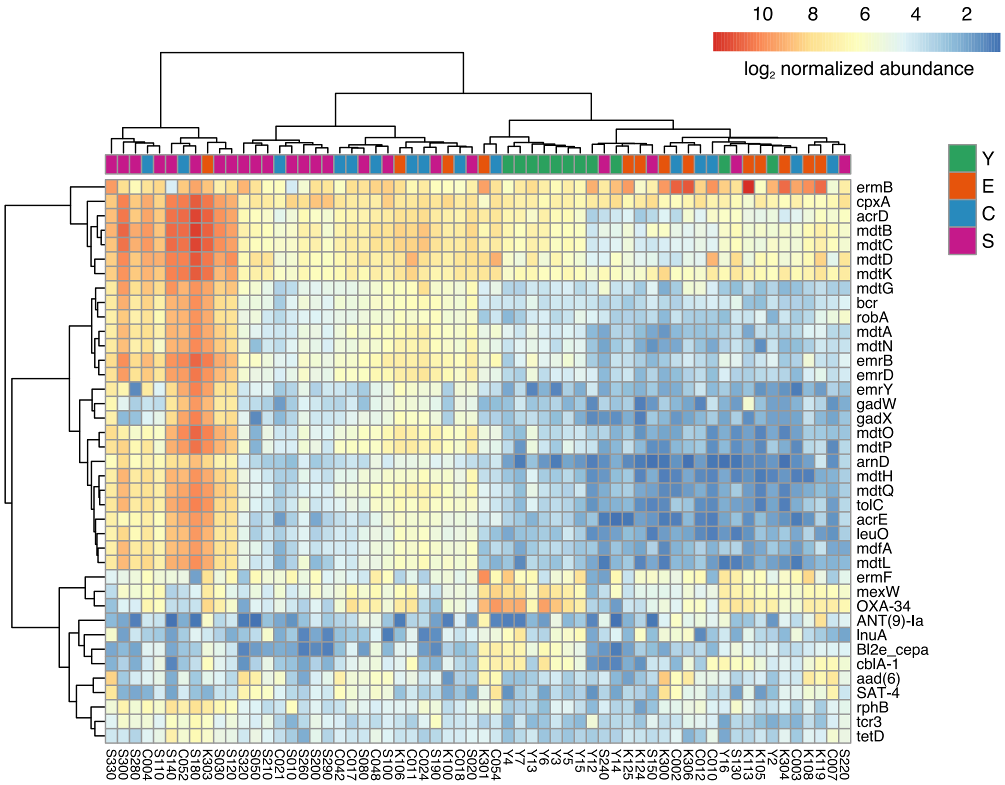


FIGURE 7.3: **Age group-specific antibiotic resistance determinants.** The heatmap is computed on the abundances of counts normalized by library size and log2 scaled. The colour code ranges from blue (low abundance) to red (high abundance). The samples are in column while on the rows there are the gene names of the ARDs significantly differentially represented between age groups as identified by the Wald test (Bonferroni corrected p value ≤ 0.05). Clustering is computed over Euclidean distances with Ward linkage method. For more information on age group-specific antibiotic resistance determinants, see Supplementary Table S2

7.4.4 Co-occurrence patterns among ARDs and species

When looking at the taxonomic classification of these ARDs, we found that they were differently assigned even at the phylum level, depending on the age group (Figure 7.4), hinting at the establishment of age group-specific topological patterns in the gut resistome. In particular, we observed a higher contribution of ARDs from *Proteobacteria* (S vs Y, p value = 0.002; S vs E, p value = 0.02, Wilcoxon test) and a lower contribution of ARDs from *Bacteroidetes* (S vs Y, p value = 0.01; S vs E, p value = 0.02) in semi-supercentenarians than in younger adults. It is worth noting that 46.3% of proteobacterial ARDs encode for multidrug efflux pumps of the species *Escherichia coli*. On the other hand, the most represented ARDs in the faecal resistome of young adults are beta-lactamases from Bacteroidetes species (Y vs S, p value = 0.03). These findings were further highlighted by a correlation network analysis, showing associations between microbial species and ARDs. Specifically, the network in Figure 7.5 shows the co-occurrence patterns between the age group-specific ADRs as identified above, and the resistome relative abundances summarized at species level. The nodes represent the ARD entities or the species. Six connected components (*i.e.*, nodes interconnected by edges and spatially separated by other groups of nodes and edges) with more than two nodes were identified. The most populated connected component contains 87% of nodes resistant to multidrug, of which 96% are annotated with antibiotic efflux mechanism and 4% (arnD) with antibiotic target protection mechanism (ARO:0001003). The AR protein families linked to E. coli are the major facilitator superfamily (MFS), accounting for 91.3% of E. coli links, and the resistance-nodulation-cell division (RND) antibiotic efflux pump family (4.3% for acrD). *E. coli* was found to have the highest diversity in terms of ARDs, with links also to determinants conferring resistance to sulfonamide and glicopeptide drugs. On the other hand, Bacteroidetes species, including *Alistipes* (from *Rikenellaceae* family), are linked to beta-lactamases, with Bl2e_cepa, cblA-1 and OXA-34 accounting for 80% of the nodes, and to the RND antibiotic efflux pump family (mexW), accounting for the remaining 20%.

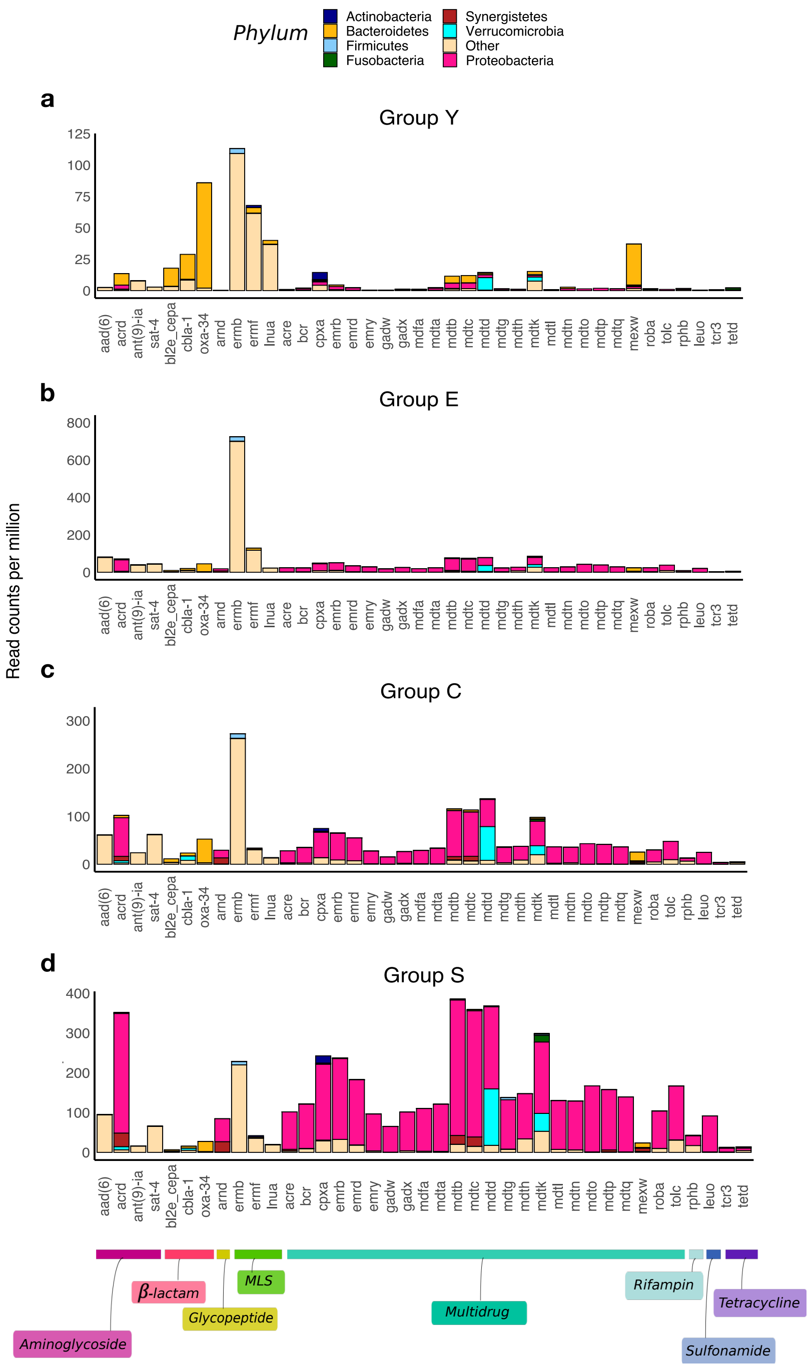


FIGURE 7.4: **Ageing-related distribution and phylum-level assignment of age group-specific antibiotic resistance determinants.** Bar plots showing the normalized abundance (read counts per million normalized by sequencing depth) of age group-specific resistance determinants (ARDs) in the gut resistome of young adults (Y, panel **a**), younger elderly (E, **b**), centenarians (C, **c**) and semi-supercentenarians (S, **d**). Phylum-level assignment of ARD reads is also shown. ARDs are organized by drug to which resistance is conferred.

7.5 Discussion

To the best of our knowledge, this is the longest metagenomic trajectory of the human gut resistome along with ageing, up to extreme longevity, which includes data

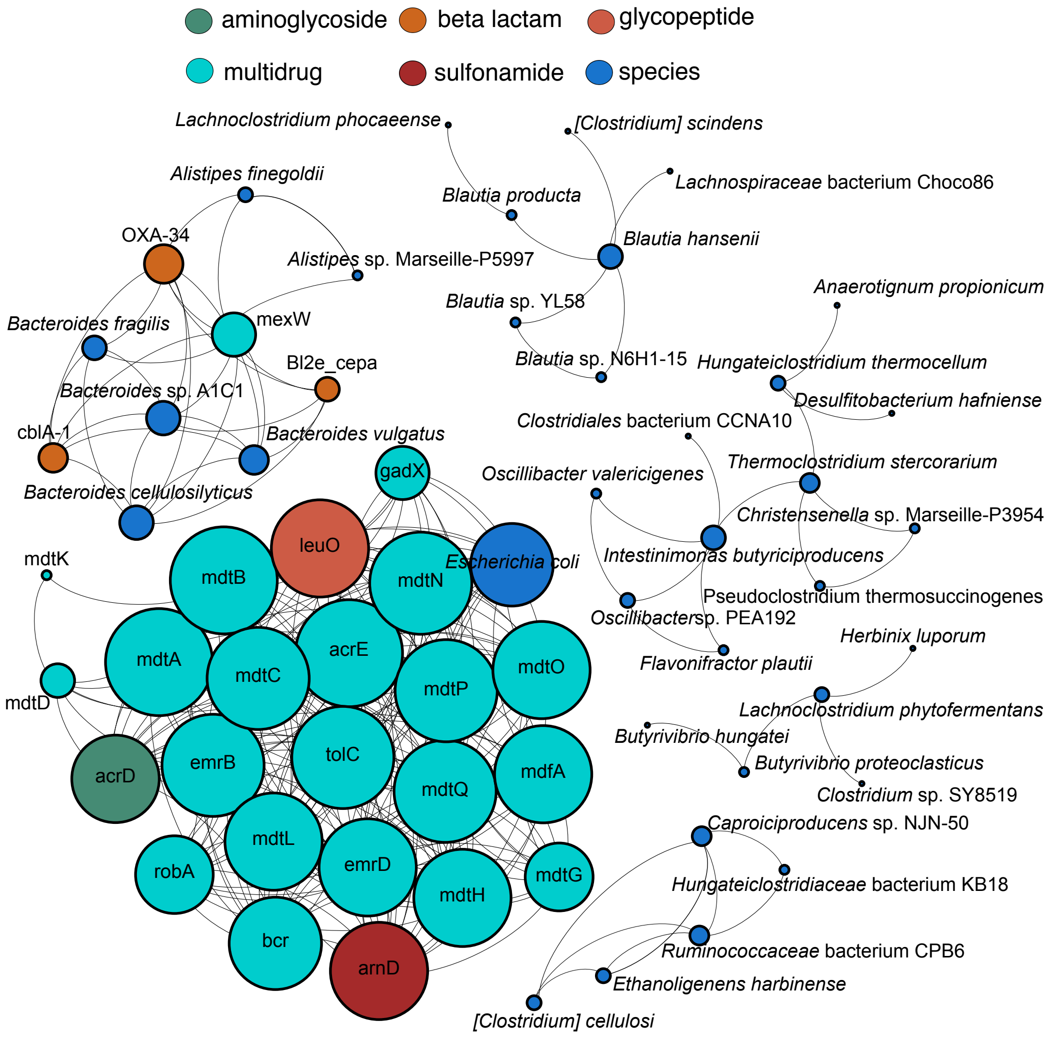

aminoglycoside
beta lactam
glycopeptide
multidrug
sulfonamide
species
Lachnoclostridium phocaeense
[Clostridium] scindens
Lachnospiraceae bacterium Choco86
Alistipes finegoldii
Blautia producta
Blautia hansenii
OXA-34
Alistipes sp. Marseille-P5997
Anaerotignum propionicum
Blautia sp. YL58
mexW
Bacteroides fragilis
Bl2e_cepa
Blautia sp. N6H1-15
Hungateiclostridium thermocellum
Bacteroides sp. A1C1
Desulfitobacterium hafniense
cblA-1
Bacteroides vulgatus
Clostridiales bacterium CCNA10
gadX
Thermoclostridium stercorarium
Oscillibacter valericigenes
Bacteroides cellulosilyticus
Christensenella sp. Marseille-P3954
leuO
mdtK
Escherichia coli
Intestinimonas butyriciproducens
mdtB
mdtN
Pseudoclostridium thermosuccinogenes
Oscillibactersp. PEA192
acrE
Herbinix luporum
mdtA
mdtC
mdtO
Flavonifractor plautii
mdtP
mdtD
Lachnoclostridium phytofermentans
Butyrivibrio hungatei
tolC
emrB
acrD
mdfA
Butyrivibrio proteoclasticus
mdtQ
Clostridium sp. SY8519
Caproiciproducens sp. NJN-50
mdtL
robA
emrD
mdtG
Hungateiclostridiaceae bacterium KB18
mdtH
bcr
Ruminococcaceae bacterium CPB6
arnD
Ethanoligenens harbinense
[Clostridium] cellulosi

FIGURE 7.5: **Co-occurrence network between age group-specific antibiotic resistance determinants and related bacterial species.** Nodes identify ARDs (colour-coded by the antibiotic to which they confer resistance) or the bacterial species assigned to them. Only connected components with more than 2 nodes are depicted. The node size is proportional to the node degree (*i.e.*, the number of edges connected to the node) and the links are Spearman correlations (see Materials and Methods).

from the gut microbiome of semi-supercentenarians (>105 years of age). Our cohort is representative of the population of the Emilia Romagna region (Northern Italy), whose gut microbiome has previously been characterized in terms of taxonomic and functional structure (Biagi et al., 2016; Rampelli et al., 2020).

We found that the taxonomic structure of the resistome largely overlaps with that of the microbiome, as antibiotic resistance (AR)-coding genes are mainly harboured by the dominant families of the gut microbial ecosystem, such as *Lachnospiraceae* and *Ruminococcaceae*, along with *Bifidobacteriaceae*. As recently argued (Schaik, 2015), this could be the result of extensive microorganism-microorganism crosstalk within the gut microbiome, with the spread of AR genes via horizontal gene transfer, potentially fuelled by antibiotic exposure. On the other hand, in line with gut microbiota data (Biagi et al., 2016), the resistome of extremely long-lived people was found to be depleted in AR reads assigned to beneficial, short-chain fatty acid-producing taxa while enriched in those assigned to potential pathobionts, such as *Enterobacteriaceae* members and *Eggerthella*. In light of the increased vulnerability of older people to infectious diseases (Liang and Mackowiak, 2007), the emergence of resistant taxa with pathogenic potential could pose a serious threat to health, as well as stress the need for resistome mapping in clinical practice, for improved efficacy of antimicrobial treatments, as has recently been discussed (D'Amico et al., 2019).

As for resistance mechanisms, the gut resistome is mainly composed of genes conferring resistance through antibiotic efflux, along with alteration of the antibiotic target and antibiotic inactivation by bacterial enzymes. In particular, 6 AR determinants (ARDs) involved in antibiotic efflux are similarly represented in all subjects regardless of age, likely being part of the core human gut resistome. Fascinatingly, we found that ageing is associated with an overall increasing abundance of AR genes, including in particular ARDs for multidrug and sulfonamide. This is especially true for semi-supercentenarians, who showed the highest load of multidrug ARDs as well as ARDs conferring resistance to rifampin and tetracycline. We speculate that this may represent an adaptive response of the human holobiont to lifelong exposure to antibiotics, including those used through the food chain and for health reasons. Although they are a model of healthy ageing, long-lived people are indeed very likely to have been more exposed to antimicrobials, also due to ageing-related physiological processes, such as immunosenescence, which contributes to increased susceptibility to infections, potentially implying a greater need of medicines, including antibiotics (Franceschi et al., 2018). On the other hand, AR is an ancient and inherent bacterial trait that predates the human use of antibiotics (Dcosta et al., 2011), and AR genes are well known to be widely distributed in any environment inhabited by bacteria, including soil, air and even household dust (Schaik, 2015; Li et al., 2018; Maamar et al., 2020). In particular, built environments have recently been appointed as an overlooked reservoir for AR, with exposure to cleaning chemicals leading to accumulation of AR genes, especially those involved in antibiotic efflux, along with loss of microbial diversity and an overall higher level of virulence (Hartmann et al., 2016; Mahnert et al., 2019). It is thus tempting to speculate that the greater abundance of AR genes in the gut microbiome of centenarians and semi-supercentenarians, i.e., people with reduced mobility who spend more time at home (Rampelli et al., 2020), is the result of a top-down selection process connected not only to health status and past medical history, but also to lifestyle habits, including stable and constant living settings within homes, with longer and more extensive exposure to various chemicals.

Consistently, we previously found that their gut microbiome is enriched in several pathways of degradation of pervasive xenobiotics in Western societies, including those contained in common consumer and other indoor products (Rampelli et al., 2020).On the other hand, it is worth noting that a higher abundance of AR genes for beta-lactam antibiotics, mainly harboured by *Bacteroidetes* members, characterizes the gut resistome of young adults. As previously discussed, genes conferring beta-lactam resistance are frequently present in Bacteroides spp. and among the most abundant AR genes in the human gut microbiome (Forslund et al., 2013; Hu et al., 2013). Interestingly, these genes do not appear to transfer to opportunistic pathogens, such as *Enterobacteriaceae* (Sommer, Dantas, and Church, 2009), possibly explaining their poor representation in the gut resistome of older people.

In conclusion, our work for the first time sheds some light on the trajectory of the human intestinal resistome along ageing and draws attention to the progressive age-related accumulation of AR genes with potentially severe repercussions on human health. In addition to stressing the relevance of resistome surveys for more effective therapies, our results pave the way for further studies aimed at reconsidering our behaviours with the ultimate goal of containing the spread of AR, thus supporting healthy ageing.

7.5.1 Author contributions

Conceptualization (Teresa Tavella, Silvia Turroni, Simone Rampelli), bioinformatic and biostatistic analysis by Teresa Tavella, validation Silvia Turroni, writing of the original draft by Teresa Tavella, Silvia Turroni, Simone Rampelli, writing, review, and editing of the manuscript by Patrizia Brigidi, Marco Candela. All authors discussed the results and commented on the manuscript.

Chapter 8

Longevity and xenobiotics

AMERICAN SOCIETY FOR MICROBIOLOGY | mSystems®

RESEARCH ARTICLE
Host-Microbe Biology

Shotgun Metagenomics of Gut Microbiota in Humans with up to Extreme Longevity and the Increasing Role of Xenobiotic Degradation

Simone Rampelli,[a] Matteo Soverini,[a] Federica D'Amico,[a] Monica Barone,[a] Teresa Tavella,[a] Daniela Monti,[b] Miriam Capri,[c,d,e] Annalisa Astolfi,[f] Patrizia Brigidi,[a,g] Elena Biagi,[a] Claudio Franceschi,[h,i] Silvia Turroni,[a] Marco Candela[a,g]

[a]Unit of Microbial Ecology of Health, Department of Pharmacy and Biotechnology, University of Bologna, Bologna, Italy
[b]Department of Experimental and Clinical Biomedical Sciences Mario Serio, University of Florence, Florence, Italy
[c]Department of Experimental, Diagnostic and Specialty Medicine, University of Bologna, Bologna, Italy
[d]CIG–Interdepartmental Center Galvani, University of Bologna, Bologna, Italy
[e]CSR–Centro di Studio per la Ricerca dell'Invecchiamento, University of Bologna, Bologna, Italy
[f]Giorgio Prodi Cancer Research Center, University of Bologna, Bologna, Italy
[g]Interdepartmental Centre of Industrial Agrifood Research (CIRI-Agrifood), University of Bologna, Cesena, Italy
[h]IRCCS, Institute of Neurologic Sciences of Bologna, Bellaria Hospital, Bologna, Italy
[i]Department of Applied Mathematics, Institute of Information Technology, Lobachevsky University of Nizhny Novgorod, Nizhny Novgorod, Russia

ABSTRACT The gut microbiome of long-lived people display an increasing abundance of subdominant species, as well as a rearrangement in health-associated bacteria, but less is known about microbiome functions. In order to disentangle the contribution of the gut microbiome to the complex trait of human longevity, we here describe the metagenomic change of the human gut microbiome along with aging in subjects with up to extreme longevity, including centenarians (aged 99 to 104 years) and semisupercentenarians (aged 105 to 109 years), i.e., demographically very uncommon subjects who reach the extreme limit of the human life span. According to our findings, the gut microbiome of centenarians and semisupercentenarians is more suited for xenobiotic degradation and shows a rearrangement in metabolic pathways related to carbohydrate, amino acid, and lipid metabolism. Collectively, our data go beyond the relationship between intestinal bacteria and physiological changes that occur with aging by detailing the shifts in the potential metagenomic functions of the gut microbiome of centenarians and semisupercentenarians as a response to progressive dietary and lifestyle modifications.

IMPORTANCE The study of longevity may help us understand how human beings can delay or survive the most frequent age-related diseases and morbidities. In this scenario, the gut microbiome has been proposed as one of the variables to monitor and possibly support healthy aging. Indeed, the disruption of host-gut microbiome homeostasis has been associated with inflammation and intestinal permeability as well as a general decline in bone and cognitive health. Here, we performed a metagenomic assessment of fecal samples from semisupercentenarians, i.e., 105 to 109 years old, in comparison to young adults, the elderly, and centenarians, shedding light on the longest compositional and functional trajectory of the human gut microbiome with aging. In addition to providing a fine taxonomic resolution down to the species level, our study emphasizes the progressive age-related increase in degradation pathways of pervasive xenobiotics in Western societies, possibly as a result of a supportive process within the molecular continuum characterizing aging.

KEYWORDS microbiome, metagenome, extreme longevity, xenobiotics, aging

Citation Rampelli S, Soverini M, D'Amico F, Barone M, Tavella T, Monti D, Capri M, Astolfi A, Brigidi P, Biagi E, Franceschi C, Turroni S, Candela M. 2020. Shotgun metagenomics of gut microbiota in humans with up to extreme longevity and the increasing role of xenobiotic degradation. mSystems 5:e00124-20. https://doi.org/10.1128/mSystems.00124-20.

Editor Lawrence A. David, Duke University

Address correspondence to Simone Rampelli, simone.rampelli@unibo.it.

This metagenomic assessment of fecal samples from semi-supercentenarians, i.e., 105-109 years old, in comparison to young adults, elderly and centenarians, sheds light on the longest trajectory of the human gut microbiome along aging!

Received 11 February 2020
Accepted 6 March 2020
Published 24 March 2020

Longevity has been described as the result of a complex combination of variables, deriving from genetics, lifestyle, and environment (1, 2). In this context, the intestinal microbiome has been proposed as a possible mediator of healthy aging that preserves host-environment homeostasis by counteracting inflammaging (3, 4), intestinal permeability (5), and deterioration of cognitive and bone health (5, 6). Correlations have been previously found between age-related gut microbiota dysbioses and levels of proinflammatory cytokines, hospitalization, poor diet, and frailty in the elderly (7). More recently, the longest human gut microbiota trajectory with aging has been built by comparing the fecal bacterial taxa from healthy adults and older individuals, including semisupercentenarians, i.e., people aged 105 to 109 years (8, 9). However, the functional changes that occur in the gut microbiome along with aging are still largely unexplored. In an attempt to provide some glimpses in this direction and to advance our knowledge on whether and how the gut microbiome may support the maintenance of health in extreme aging, we here characterized the fecal microbiome of 62 individuals, with ages ranging from 22 to 109 years, by shotgun metagenomics. According to our findings, aging is characterized by an increased number of genes involved in xenobiotic degradation, as well as by rearrangements in metabolic pathways related to carbohydrate, amino acid, and lipid metabolism. These microbiome features are boosted even more in semisupercentenarians, probably representing the result of a lifelong remodeling response to progressive changes in diet and lifestyle.

RESULTS

We previously found considerable age-related variability in fecal microbiota composition of 69 people, including centenarians and semisupercentenarians, from the Emilia Romagna region of Italy and the surrounding area (8). In an attempt to go further, unraveling the functional and species-level taxonomic links between the gut microbiome and extreme aging, we applied shotgun metagenomics to a subset of 62 DNA samples derived from the same data set previously analyzed (8). Specifically, we characterized the gut microbiome from 11 young adults (group Y, 6 females and 5 males, aged 22 to 48 years [mean age, 32.2 years]), 13 younger elderly (group K, 6 females and 7 males, aged 65 to 75 years [mean age, 72.5 years]), 15 centenarians (group C, 14 females and 1 male, aged 99 to 104 years [mean age, 100.4 years]), and 23 semisupercentenarians (group S, 17 females and 6 males, aged 105 to 109 years [mean age, 106.3 years]). A total of 1.3 billion sequences were generated, with an average of 20 million reads (±5 million reads standard deviation [SD]) per subject.

We first confirmed that the fecal microbiota in all age groups is dominated by a few bacterial families (i.e., *Bifidobacteriaceae*, *Bacteroidaceae*, *Lachnospiraceae*, and *Ruminococcaceae*) whose relative abundance decreases with age (mean relative abundance ± SD: group Y, 73% ± 3%; group K, 65% ± 4%; group C, 62% ± 4%; group S, 58% ± 6%). When focusing our attention at the species level, we found that these contributions were mainly accounted for by 13 bacterial species: *Bifidobacterium adolescentis*, *Bifidobacterium longum*, *Bacteroides uniformis*, *Faecalibacterium prausnitzii*, *Ruminococcus bromii*, *Subdoligranulum* sp., *Anaerostipes hadrus*, *Blautia obeum*, *Ruminococcus torques*, *Coprococcus catus*, *Coprococcus comes*, *Dorea longicatena,* and *Roseburia* sp. Bray-Curtis principal-coordinate analysis (PCoA) of species-level relative abundance profiles provided evidence of an age-related trajectory ($P < 0.05$, permutation test with pseudo-F ratios), involving the establishment of age group-specific topological patterns in the taxonomic and functional microbiome structure, as shown by network plots (Fig. 1) and bar plots (see Fig. S1 in the supplemental material). However, the species-level compositional structure of the gut microbiota from the younger elderly group overall matches that from young adults ($P = 0.2$), suggesting that the physiology of the aging process may not involve gross changes in gut microbiome species and their relative abundance. On the other hand, gut microbiota from centenarians and semisupercentenarians feature a distinctive rearrangement in their taxonomic configurations (Fig. 2A). In particular, compared with younger individuals, long-lived people show a decreased contribution of *B. uniformis*, *Eubacterium rectale*, *C. comes*, and *F. prausnitzii*,

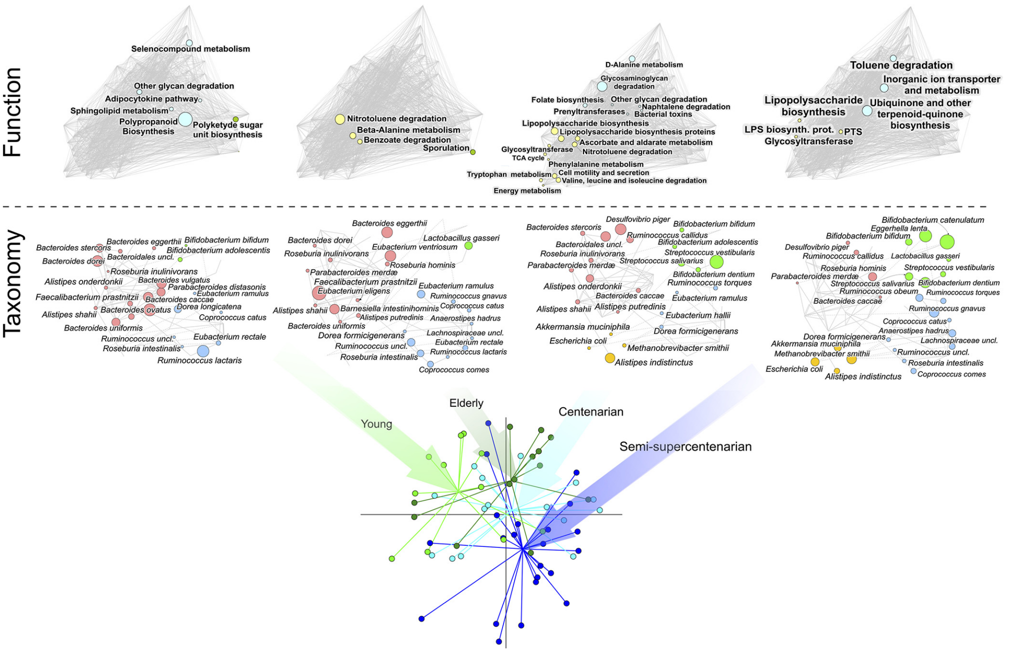


FIG 1 Gut microbiome variation with aging. (Top) Network plots showing the taxonomic and functional configurations of the gut microbiome of four age groups: 11 young adults (aged 22 to 48 years; group young), 13 younger elderly (aged 65 to 75 years; group elderly), 15 centenarians (aged 99 to 104 years; group centenarian), and 23 semisupercentenarians (aged 105 to 109 years; group semisupercentenarian). Disc sizes indicate species or functional pathway overabundances relative to the average abundance of the whole cohort. Lines indicate significant positive correlations between the values of the discs. (Bottom) PCoA plot of Bray-Curtis dissimilarity between the species-level relative abundance data sets of the four age groups.

along with a progressive increase of *Escherichia coli*, *Methanobrevibacter smithii*, *Akkermansia muciniphila,* and *Eggerthella lenta* ($P < 0.05$, Kruskal-Wallis test). These trends have already been reported in previous 16S rRNA gene-based microbiome works in the same subjects (3, 8), as well as in Chinese centenarians (10), further strengthening that the observed gut microbiome variations may be part of the extreme aging process, regardless of environmental variables, such as geographical origin and cultural habits (i.e., diet and lifestyle) (11).

Interestingly, when we focused our analysis at a functional scale, we found a progressive age-related increase in the number of reads for genes devoted to xenobiotic biodegradation and metabolism, and a simultaneous decrease in genes involved in carbohydrate metabolism (Fig. 2B and C; Fig. S2). This functional rearrangement is even more pronounced in the gut microbiome of centenarians and semisupercentenarians, where we observed a reduced contribution of pathways for starch and sucrose (KEGG pathway no. ko00500), pentose phosphate (ko00030), and amino sugar and nucleotide sugar (ko00520) metabolism and a concomitant increase in toluene (ko00623), ethylbenzene (ko00642), caprolactam (ko00930), and chlorocyclohexane and chlorobenzene (ko00361) degradation pathways. While the changes related to carbohydrate metabolism have already been reported in previous studies and suggested to be associated with age-related changes in dietary habits (7, 9), the increase in genes for xenobiotic metabolism is reported here for the first time and appears particularly intriguing.

Ethylbenzene, chlorobenzene, chlorocyclohexane, and toluene are pervasive chemicals mainly deriving from industrial manufacturing and municipal discharges and are

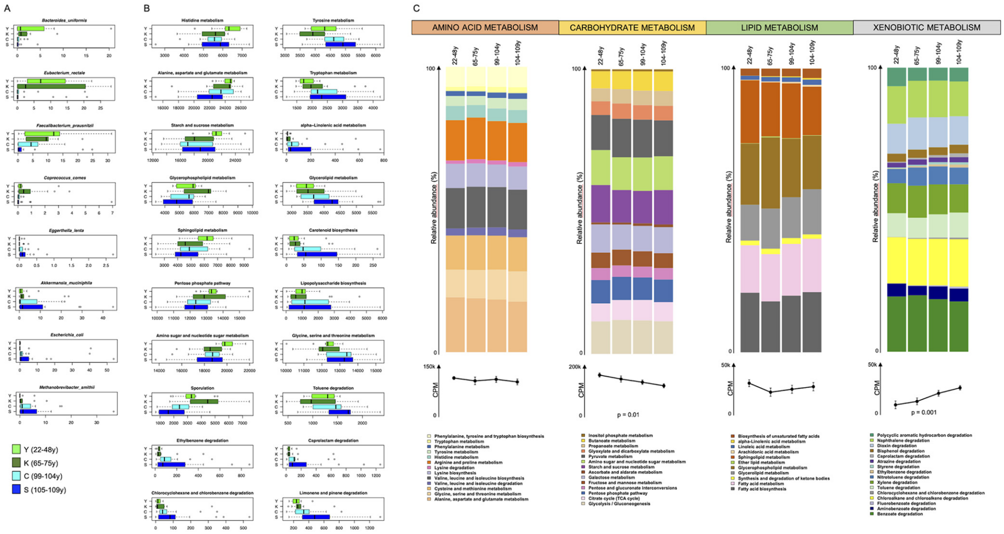


FIG 2 Aging-related trajectories of gut microbiome species and functional pathways. (A) Box plots of the normalized relative abundances of bacterial species differentially represented among the four age groups (Y, young adults; K, younger elderly; C, centenarians; S, semisupercentenarians) ($P < 0.05$, Kruskal-Wallis test). (B) Box plots of the normalized abundance (assigned reads per million sequences, i.e., counts per million [CPM]) of KEGG pathways differentially represented among age groups ($P < 0.05$, Kruskal-Wallis test). (C) Bar plots at the top show the KEGG pathway-classified metabolic configurations for amino acid, carbohydrate, lipid, and xenobiotic metabolism as the mean relative contribution of each pathway to the total normalized number of reads assigned to each specific metabolism. At the bottom of the panel, the average number of normalized reads (CPM ± standard error of the mean [SEM error bar]) assigned to each specific metabolism is shown. Significant differences among age groups are shown on the graphs.

under monitoring all over the world as part of the main environmental contaminants of the atmosphere, due to their toxic effects (12–14). The primary man-made sources of these molecules are indeed the emissions from motor and exhaust vehicles, as well as cigarette smoke. Furthermore, they are known to be generated during the processing of refined petroleum products, such as plastics, and to be contained in common consumer products, such as paints and lacquers, thinners, and rubber products (14). As regards caprolactam, it is the raw material of nylon, used for the production of many indoor products, such as synthetic fibers, resins, synthetic leather, and plasticizers. Previous studies have demonstrated the higher indoor burden of these molecules than in the outdoor environment and emphasized the exceptional importance of indoor exposure on human health (15, 16). It is a matter of fact that living in environments under strong anthropic pressures, such as the Emilia Romagna region in Italy (17, 18), results in the continuous and constant exposure to these pervasive xenobiotic substances, favoring their maintenance and progressive accumulation in body tissues, including the gut (19–22). We believe that this could create the appropriate conditions for the human host to select for gut microbiome components capable of detoxifying such chemical compounds, with a mutual benefit in terms of microbiome and host fitness in anthropic environments. Indeed, recent works have shown that the human-associated microbial communities of urban Western populations are functionally suited to the degradation of xenobiotic molecules, including caprolactam (23–25). Further supporting the importance of human microbiomes in providing a response to xenobiotic exposure, in another recent work the upper airway microbiome of nonasthmatic individuals has been found to possess greater ability to metabolize caprolactam than that of asthmatic people (25). According to the authors, the selection of caprolactam-degrading microbes in the airway microbiome would decrease host exposure to indoor air pollutants, providing an ultimate impact on human health.

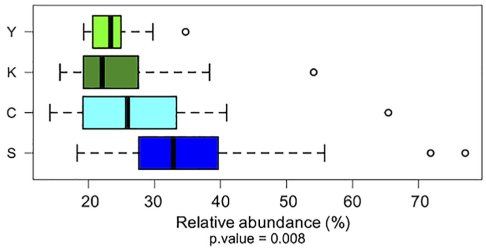


FIG 3 The contribution of commensal bacteria to xenobiotic degradation is significantly higher in long-lived individuals. Box plots show the percentages of bacteria of the human core gut microbiome that harbor genes for xenobiotic degradation. Members of the core microbiome were defined based on previous works (26–30).

Centenarians and semisupercentenarians are long-lived individuals who, as such, may boast an important history of exposure to xenobiotic stressors. Furthermore, as they have reduced mobility, these subjects tend to spend more time in their own houses than younger people (Fig. S3), with increased exposure to indoor pollutants. It is thus tempting to speculate that their microbiome is better equipped for the degradation of these xenobiotics as a result of a process driven by the more lasting and assiduous exposure to these chemicals. It is also worth noting that these metabolic functionalities are possessed by commensal bacteria belonging to the human core microbiome, i.e., microbial taxa that have been found to be shared by the microbiome of all human populations sampled to date (26–30) (Fig. 3). This raises important open questions on the biological mechanisms that lead to the consolidation and enrichment of xenobiotic-degrading abilities in centenarian and semisupercentenarian gut microbiomes. Here, we speculate that the highest contribution to xenobiotic degradation by commensals in long-lived people might be the result mainly of a top-down selection process related to the lifestyle habits of these exceptionally old individuals, i.e., stable and constant living settings within their own homes, together with a longer exposure and consequent accumulation of these chemicals in the host tissues due to their longer life.

Besides xenobiotic-degrading genes and those involved in carbohydrate metabolism, we also found age-related differences in other metabolic pathways, including those associated with lipid metabolism. In particular, centenarians and semisupercentenarians show more reads for alpha-linoleic acid (KEGG pathway no. ko00592) and glycerolipid (ko00561) metabolism; on the other hand, younger people show a greater contribution of genes involved in sphingolipid (ko00600) and glycerophospholipid (ko00564) metabolism. Given that glycerophospholipids and sphingolipids are known to be more abundant in animal-derived foods (31, 32), while alpha-linoleic acid is derived mainly from plant foods (33), these profiles may be related to eating habits and, in particular, to the higher intake of plant-derived fats than animal fats by long-lived individuals than by younger people (Fig. S4). Moreover, when looking at functional pathways involved in amino acid metabolism, we found a progressive increase with age in genes for the metabolism of tryptophan (ko00380), tyrosine (ko00350), glycine, serine, and threonine (ko00260). On the other hand, genes for alanine, aspartate, and glutamate metabolism (ko00250) were found to be more abundant in younger individuals. These evidences are in agreement with our previous study (9), in particular with regard to the metabolism of tryptophan and tyrosine as an indicator of enhanced proteolytic metabolism. Furthermore, these findings fit with metabolite measures in the centenarians of our cohort, i.e., the decrease of the bioavailability of tryptophan in serum (34), as well as the increased urinary levels of phenolic metabolites, deriving from the metabolism of tyrosine (35). Finally, we found a progressive increase with aging of genes for lipopolysaccharide biosynthesis (ko00540), which can be associated with the presence of pathobionts (i.e., members of the *Enterobacteriaceae* family) and the low levels of chronic inflammation (i.e., inflammaging), as previously demonstrated in long-lived people (3, 8, 9).

DISCUSSION

Here we described—as far as we know, for the first time—the metagenomic changes of the human gut microbiota that occur with aging, up to extreme longevity, by characterizing the microbiome of semisupercentenarians, i.e., demographically very uncommon subjects who reach the extreme limit of the human life span (>105 years of age). In addition to confirming the known taxonomic features of an aging microbiota, we extended the definition of the human core gut microbiota down to the species level and provided an accurate depiction of the functional changes occurring along with aging. In a sort of continuum line with our previous study, where we demonstrated that the intestinal microbiome of Italian adults is equipped for the degradation of xenobiotics, probably as a functional response to exposure to these compounds (24), we here advance the fascinating hypothesis that aging in Western urban environments progressively selects for commensal microbiome strains with metabolic abilities toward specific xenobiotics. We speculate that this could represent an adaptive response of the human holobiont to the increased exposure to, and accumulation of, xenobiotic substances along the aging process. As recently discussed (36), future studies should be aimed at better understanding the complex interplay between xenobiotic exposure and the human gut microbiome. The individual gut microbiome structure will have to be matched with the personal exposure level, with the latter being dissected by monitoring xenobiotics in feces and body fluids. Long-term longitudinal studies must be conceived, with the aim of highlighting the mechanisms underlying this potential microbiome adaptive variation, as a result of a top-down selection process of microbiome functions for xenobiotic detoxification and the ultimate impact in terms of host health protection. Given that the xenobiotics that emerged in the present study are now ubiquitous in modern urban areas, it would also be interesting to assess the xenobiotic degradation capacity of ancient microbial communities by analyzing samples from the preindustrial era, in order to fully understand the effects of these molecules on the evolutionary history of the human holobiont. Studies of this type would help to shed light on whether the peculiar functional profiles of the gut microbiome of extremely long-lived hosts, as found in our work, are the result of an adaptive and remodeling process inherent to the physiology of human aging in modern urban societies and thus capable of supporting a new homeostasis.

MATERIALS AND METHODS

Subjects and study groups. The study used genomic DNA from 62 fecal samples collected for a study by Biagi et al. (8). Subjects were enrolled in the Emilia Romagna region (Italy) and categorized as follows: 11 young adults (group Y, 6 females and 5 males, aged 22 to 48 years [mean age, 32.2 years]), 13 younger elderly (group K, 6 females and 7 males, aged 65 to 75 years [mean age, 72.5 years]), 15 centenarians (group C, 14 females and 1 male, aged 99 to 104 years [mean age, 100.4 years]), and 23 semisupercentenarians (group S, 17 females and 6 males, aged 105 to 109 years [mean age, 106.3 years]). See Table S1 in the supplemental material for further information about the cohort. The study protocol was approved by the Ethics Committee of Sant'Orsola-Malpighi University Hospital (Bologna, Italy) under EM/26/2014/U (with reference to 22/2007/U/Tess).

Evaluation of the time spent indoors and outdoors by the elderly. Elderly participants signed the informed consent before undergoing the questionnaires with an interviewer as previously described (37). The participants were asked how often they left their homes (daily, weekly, monthly, etc.) and based on seven different answers were assigned a score: those who never went out, the lowest frequency, were given a score of 1, while those who left their homes "daily," the highest frequency, were given a score of 7. The answers, treated as a continuous scale (arbitrary scores of 1 to 7), were used to determine the frequency of movement outside home (FMOH) score.

Library preparation and shotgun sequencing. DNA libraries were prepared using the QIAseq FX DNA library kit (Qiagen, Hilden, Germany) in accordance with the manufacturer's instructions. Briefly, total microbial DNA was quantified by a Qubit fluorometer (Invitrogen, Waltham, MA, USA), and 100 ng of each sample was fragmented to a 450-bp size, end-repaired, and A-tailed using FX enzyme mix with the following thermal cycle: 4°C for 1 min, 32°C for 8 min, and 65°C for 30 min. Samples were then incubated at 20°C for 15 min in the presence of DNA ligase and Illumina adapter barcodes for adapter ligation. After two purification steps with Agencourt AMPure XP magnetic beads (Beckman Coulter, Brea, CA, USA), a 10-cycle PCR amplification and a further step of purification as described above, the final library was obtained by pooling the samples at equimolar concentrations of 4 nM. Sequencing was performed on an Illumina NextSeq platform using a 2 × 150-bp paired-end protocol, in accordance with the manufacturer's instructions (Illumina, San Diego, CA, USA). High-quality paired-end sequences were uploaded to the SRA repository.

Bioinformatics and biostatistics. The functional annotation of the sequences deriving from the 62 genomic DNA samples (8) was conducted as previously described (9). In brief, shotgun reads were first filtered by quality and human sequences. This last step was achieved using the human sequence removal pipeline and the WGS read processing procedure of the Human Microbiome Project (HMP) (38). The obtained reads were taxonomically characterized at the species level by MetaPhlAn2 (39) and assigned for functionality at different levels of the KEGG database (40), using Metagenome Composition Vector (MetaCV) with default parameters (41). The resulting table consisted of multiple matrices, with sample identification numbers (IDs) in the columns and annotations at the species level or at different levels of the KEGG database in the rows.

PCoA analysis was carried out using vegan (https://cran.r-project.org/web/packages/vegan/index.html) in R. Significance testing and permutation analysis were performed using the R package stats and vegan. Data separation in the PCoA was tested using a permutation test with pseudo-F ratios (function adonis in the vegan package). When appropriate, *P* values were adjusted for multiple comparisons using the Benjamini-Hochberg correction. A false discovery rate (FDR) of <0.05 was considered statistically significant.

Network plots were determined as previously described (24). In brief, associations between KEGG pathway abundances were evaluated by the Kendall correlation test, displayed with hierarchical Ward linkage clustering based on the Spearman correlation coefficients, and then used to define pathway groups (circles with the same color). Significant associations were verified for multiple testing using the *q* value method (http://www.bioconductor.org/packages/release/bioc/html/qvalue.html) ($P < 0.05$). Permutational multivariate analysis of variance was used to determine whether the pathway groups were significantly different from each other. The network plots were created using Cytoscape software (42). Circle size represents the normalized overabundance of the pathway relative to the background. Connections between nodes represent significant positive Kendall correlations between KEGG pathways (FDR < 0.05).

Assignment of functions for xenobiotic degradation to commensal bacteria. Reads with assignment to xenobiotic degradation functions were further inspected for taxonomy. Where present, the species-level classification of MetaCV (41) was retrieved, and the taxon ID in the NCBI taxonomy database was obtained using the web interface of the NCBI Taxonomy Browser tool (https://www.ncbi.nlm.nih.gov/Taxonomy/TaxIdentifier/tax_identifier.cgi). In order to retrieve the entire phylogeny of the assignment, we transformed the NCBI taxonomy IDs into the full lineage by using the ETE3 toolkit (43). Hits for xenobiotic degradation were then split based on their taxonomy and collected in a new table containing the values for each sample. We finally identified the proportion of functions assigned to commensal bacteria of the human core gut microbiome, i.e., microbial taxa that have been found to be shared by all human populations sampled to date (26–30), by specifically looking for their abundance across samples and visualizing them by box plots using the R software.

Analysis of nutritional data. Dietary information for the elderly subjects of groups K, C, and S were provided and discussed in our previous publications (1, 8). As regards group Y, the subjects were asked to compile 24-h dietary recalls to retrieve information on the composition of their diet, as previously reported by Barone and colleagues (44). Dietary data for semisupercentenarians (8) were converted to a numeric frequency, in order to infer the daily consumption of each food category. Total daily calorie intake as well as macro- and micronutrient contributions for individuals in groups Y and S were estimated through the MètaDieta software version 3.7 (Meteda, Rome, Italy).

Data availability. High-quality paired-end sequences were uploaded to the SRA repository under BioProject number PRJNA553191.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, PDF file, 0.7 MB.

FIG S2, PDF file, 0.2 MB.

FIG S3, PDF file, 0.02 MB.

FIG S4, PDF file, 0.01 MB.

TABLE S1, XLSX file, 0.01 MB.

ACKNOWLEDGMENTS

We thank Valentina Indio for her technical support in sequencing.

Author contributions included library preparation and sequencing by F.D., M.B., and A.A., bioinformatic and biostatistic analysis by S.R., M.S., and T.T., analysis of nutritional data by M.B., writing of the original draft by S.R., F.D., S.T., and M. Candela, and writing, review, and editing of the manuscript by M.S., D.M., M. Capri, A.A., E.B., P.B., and C.F. All authors discussed the results and commented on the manuscript.

REFERENCES

1. Franceschi C, Ostan R, Santoro A. 2018. Nutrition and inflammation: are centenarians similar to individuals on calorie-restricted diets? Annu Rev Nutr 38:329–356. https://doi.org/10.1146/annurev-nutr-082117-051637.
2. Giuliani C, Garagnani P, Franceschi C. 2018. Genetics of human longevity within an eco-evolutionary nature-nurture framework. Circ Res 123:745–772. https://doi.org/10.1161/CIRCRESAHA.118.312562.
3. Biagi E, Nylund L, Candela M, Ostan R, Bucci L, Pini E, Nikkïla J, Monti D, Satokari R, Franceschi C, Brigidi P, De Vos W. 2010. Through ageing, and beyond: gut microbiota and inflammatory status in seniors and centenarians. PLoS One 5:e10667. https://doi.org/10.1371/journal.pone.0010667.
4. Franceschi C, Garagnani P, Parini P, Giuliani C, Santoro A. 2018. Inflammaging: a new immune-metabolic viewpoint for age-related diseases. Nat Rev Endocrinol 14:576–590. https://doi.org/10.1038/s41574-018-0059-4.
5. Nicoletti C. 2015. Age-associated changes of the intestinal epithelial barrier: local and systemic implications. Expert Rev Gastroenterol Hepatol 9:1467–1469. https://doi.org/10.1586/17474124.2015.1092872.
6. Villa CR, Ward WE, Comelli EM. 2017. Gut microbiota-bone axis. Crit Rev Food Sci Nutr 57:1664–1672. https://doi.org/10.1080/10408398.2015.1010034.
7. Claesson MJ, Jeffery IB, Conde S, Power SE, O'Connor EM, Cusack S, Harris HMB, Coakley M, Lakshminarayanan B, O'Sullivan O, Fitzgerald GF, Deane J, O'Connor M, Harnedy N, O'Connor K, O'Mahony D, van Sinderen D, Wallace M, Brennan L, Stanton C, Marchesi JR, Fitzgerald AP, Shanahan F, Hill C, Ross RP, O'Toole PW. 2012. Gut microbiota composition correlates with diet and health in the elderly. Nature 488:178–184. https://doi.org/10.1038/nature11319.
8. Biagi E, Franceschi C, Rampelli S, Severgnini M, Ostan R, Turroni S, Consolandi C, Quercia S, Scurti M, Monti D, Capri M, Brigidi P, Candela M. 2016. Gut microbiota and extreme longevity. Curr Biol 26:1480–1485. https://doi.org/10.1016/j.cub.2016.04.016.
9. Rampelli S, Candela M, Turroni S, Biagi E, Collino S, Franceschi C, O'Toole PW, Brigidi P. 2013. Functional metagenomic profiling of intestinal microbiome in extreme ageing. Aging (Albany NY) 5:902–912. https://doi.org/10.18632/aging.100623.
10. Wang F, Yu T, Huang G, Cai D, Liang X, Su H, Zhu Z, Li D, Yang Y, Shen P, Mao R, Yu L, Zhao M, Li Q. 2015. Gut microbiota community and its assembly associated with age and diet in Chinese centenarians. J Microbiol Biotechnol 25:1195–1204. https://doi.org/10.4014/jmb.1410.10014.
11. Santoro A, Ostan R, Candela M, Biagi E, Brigidi P, Capri M, Franceschi C. 2018. Gut microbiota changes in the extreme decades of human life: a focus on centenarians. Cell Mol Life Sci 75:129–148. https://doi.org/10.1007/s00018-017-2674-y.
12. Bruno P, Caselli M, de Gennaro G, Scolletta L, Trizio L, Tutino M. 2008. Assessment of the impact produced by the traffic source on VOC level in the urban area of Canosa di Puglia (Italy). Water Air Soil Pollut 193:37–50. https://doi.org/10.1007/s11270-008-9666-3.
13. Buczynska AJ, Krata A, Stranger M, Locateli Godoi AF, Kontozova-Deutsch V, Bencs L, Naveau I, Roekens E, Van Grieken R. 2009. Atmospheric BTEX-concentrations in an area with intensive street traffic. Atmos Environ 43:311–318. https://doi.org/10.1016/j.atmosenv.2008.09.071.
14. Leusch F, Bartkow M. 2010. A short primer on benzene, toluene, ethylbenzene and xylenes (BTEX) in the environment and hydraulic fracturing fluids. Smart Water Research Centre, Griffith University, Queensland, Australia. https://environment.des.qld.gov.au/management/coal-seam-gas/pdf/btex-report.pdf.
15. Massolo L, Rehwagen M, Porta A, Ronco A, Herbarth O, Mueller A. 2010. Indoor-outdoor distribution and risk assessment of volatile organic compounds in the atmosphere of industrial and urban areas. Environ Toxicol 25:339–349. https://doi.org/10.1002/tox.20504.
16. Esplugues A, Ballester F, Estarlich M, Llop S, Fuentes-Leonarte V, Mantilla E, Iñiguez C. 2010. Indoor and outdoor air concentrations of BTEX and determinants in a cohort of one-year old children in Valencia, Spain. Sci Total Environ 409:63–69. https://doi.org/10.1016/j.scitotenv.2010.09.039.
17. Larsen BR, Gilardoni S, Stenström K, Niedzialek J, Jimenez J, Belis CA. 2012. Sources for PM air pollution in the Po Plain, Italy: II. Probabilistic uncertainty characterization and sensitivity analysis of secondary and primary sources. Atmos Environ 50:203–213. https://doi.org/10.1016/j.atmosenv.2011.12.038.
18. Zauli Sajani S, Marchesi S, Trentini A, Bacco D, Zigola C, Rovelli S, Ricciardelli I, Maccone C, Lauriola P, Cavallo DM, Poluzzi V, Cattaneo A, Harrison RM. 2018. Vertical variation of PM2.5 mass and chemical composition, particle size distribution, NO2, and BTEX at a high rise building. Environ Pollut 235:339–349. https://doi.org/10.1016/j.envpol.2017.12.090.
19. Heinrich-Ramm R, Jakubowski M, Heinzow B, Molin Christensen J, Olsen E, Hertel O. 2000. Biological monitoring for exposure to volatile organic compounds (VOCs) (IUPAC Recommendations 2000). Pure Appl Chem 72:385–436. https://doi.org/10.1351/pac200072030385.
20. Galloway TS. 2015. Micro- and nano-plastics and human health, p 343–366. *In* Bergmann M, Gutow L, Klages M (ed), Marine anthropogenic litter. Springer International Publishing, New York, NY.
21. Sutic I, Bulog A, Sutic I, Pavisic V, Mrakovcic-Sutic I. 2016. Changes in the concentration of BTEX (benzene, toluene, ethylbenzene, m/p-xylene and o-xylene) following environmental and occupational exposure to vapors. JMESS 2:1014–1018.
22. Wright SL, Kelly FJ. 2017. Plastic and human health: a micro issue? Environ Sci Technol 51:6634–6647. https://doi.org/10.1021/acs.est.7b00423.
23. Wu J, Peters BA, Dominianni C, Zhang Y, Pei Z, Yang L, Ma Y, Purdue MP, Jacobs EJ, Gapstur SM, Li H, Alekseyenko AV, Hayes RB, Ahn J. 2016. Cigarette smoking and the oral microbiome in a large study of American adults. ISME J 10:2435–2446. https://doi.org/10.1038/ismej.2016.37.
24. Rampelli S, Schnorr SL, Consolandi C, Turroni S, Severgnini M, Peano C, Brigidi P, Crittenden AN, Henry AG, Candela M. 2015. Metagenome sequencing of the Hadza hunter-gatherer gut microbiota. Curr Biol 25:1682–1693. https://doi.org/10.1016/j.cub.2015.04.055.
25. Lee JJ, Kim SH, Lee MJ, Kim BK, Song WJ, Park HW, Cho SH, Hong SJ, Chang YS, Kim BS. 2019. Different upper airway microbiome and their functional genes associated with asthma in young adults and elderly individuals. Allergy 74:709–719. https://doi.org/10.1111/all.13608.
26. Moeller AH, Li Y, Mpoudi Ngole E, Ahuka-Mundeke S, Lonsdorf EV, Pusey AE, Peeters M, Hahn BH, Ochman H. 2014. Rapid changes in the gut microbiome during human evolution. Proc Natl Acad Sci U S A 111:16431–16435. https://doi.org/10.1073/pnas.1419136111.
27. Zhang J, Guo Z, Xue Z, Sun Z, Zhang M, Wang L, Wang G, Wang F, Xu J, Cao H, Xu H, Lv Q, Zhong Z, Chen Y, Qimuge S, Menghe B, Zheng Y, Zhao L, Chen W, Zhang H. 2015. A phylo-functional core of gut microbiota in healthy young Chinese cohorts across lifestyles, geography and ethnicities. ISME J 9:1979–1990. https://doi.org/10.1038/ismej.2015.11.
28. Lloyd-Price J, Abu-Ali G, Huttenhower C. 2016. The healthy human microbiome. Genome Med 8:51. https://doi.org/10.1186/s13073-016-0307-y.
29. Groussin M, Mazel F, Sanders JG, Smillie CS, Lavergne S, Thuiller W, Alm EJ. 2017. Unraveling the processes shaping mammalian gut microbiomes over evolutionary time. Nat Commun 8:14319. https://doi.org/10.1038/ncomms14319.
30. Mancabelli L, Milani C, Lugli GA, Turroni F, Ferrario C, van Sinderen D, Ventura M. 2017. Meta-analysis of the human gut microbiome from urbanized and pre-agricultural populations. Environ Microbiol 19:1379–1390. https://doi.org/10.1111/1462-2920.13692.
31. Vesper H, Schmelz EM, Nikolova-Karakashian MN, Dillehay DL, Lynch DV, Merrill A. 1999. Sphingolipids in food and the emerging importance of sphingolipids to nutrition. J Nutr 129:1239–1250. https://doi.org/10.1093/jn/129.7.1239.
32. Castro-Gómez P, Garcia-Serrano A, Visioli F, Fontecha J. 2015. Relevance of dietary glycerophospholipids and sphingolipids to human health. Prostaglandins Leukot Essent Fatty Acids 101:41–51. https://doi.org/10.1016/j.plefa.2015.07.004.
33. Stark AH, Crawford MA, Reifen R. 2008. Update on alpha-linolenic acid. Nutr Rev 66:326–332. https://doi.org/10.1111/j.1753-4887.2008.00040.x.
34. Collino S, Montoliu I, Martin FP, Scherer M, Mari D, Salvioli S, Bucci L, Ostan R, Monti D, Biagi E, Brigidi P, Franceschi C, Rezzi S. 2013. Metabolic signatures of extreme longevity in northern Italian centenarians reveal a complex remodeling of lipids, amino acids, and gut microbiota metabolism. PLoS One 8:e56564. https://doi.org/10.1371/journal.pone.0056564.
35. Moco S, Candela M, Chuang E, Draper C, Cominetti O, Montoliu I, Barron

D, Kussmann M, Brigidi P, Gionchetti P, Martin FP. 2014. Systems biology approaches for inflammatory bowel disease: emphasis on gut microbial metabolism. Inflamm Bowel Dis 20:2104–2114. https://doi.org/10.1097/MIB.0000000000000116.

36. Collins SL, Patterson AD. 2020. The gut microbiome: an orchestrator of xenobiotic metabolism. Acta Pharm Sin B 10:19–32. https://doi.org/10.1016/j.apsb.2019.12.001.

37. Bucci L, Ostan R, Giampieri E, Cevenini E, Pini E, Scurti M, Vescovini R, Sansoni P, Caruso C, Mari D, Ronchetti F, Borghi MO, Ogliari G, Grossi C, Capri M, Salvioli S, Castellani G, Franceschi C, Monti D. 2014. Immune parameters identify Italian centenarians with a longer five-year survival independent of their health and functional status. Exp Gerontol 54: 14–20. https://doi.org/10.1016/j.exger.2014.01.023.

38. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. 2007. The human microbiome project. Nature 449:804–810. https://doi.org/10.1038/nature06244.

39. Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, Tett A, Huttenhower C, Segata N. 2015. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nat Methods 12:902–903. https://doi.org/10.1038/nmeth.3589.

40. Wixon J, Kell D. 2000. The Kyoto encyclopedia of genes and genomes—KEGG. Yeast 17:48–55. https://doi.org/10.1002/(SICI)1097-0061(200004)17:1<48::AID-YEA2>3.0.CO;2-H.

41. Liu J, Wang H, Yang H, Zhang Y, Wang J, Zhao F, Qi J. 2013. Composition-based classification of short metagenomic sequences elucidates the landscapes of taxonomic and functional enrichment of microorganisms. Nucleic Acids Res 41:e3. https://doi.org/10.1093/nar/gks828.

42. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504. https://doi.org/10.1101/gr.1239303.

43. Huerta-Cepas J, Serra F, Bork P. 2016. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. Mol Biol Evol 33:1635–1638. https://doi.org/10.1093/molbev/msw046.

44. Barone M, Turroni S, Rampelli S, Soverini M, D'Amico F, Biagi E, Brigidi P, Troiani E, Candela M. 2019. Gut microbiome response to a modern Paleolithic diet in a Western lifestyle context. PLoS One 14:e0220619. https://doi.org/10.1371/journal.pone.0220619.

Chapter 9

Conclusions

The study of bacteria and their interactions with the host/environment is central in the matter of circular health. In order to decipher the different aspects of the nature of these symbiosis, we can benefit from omics techniques for unveiling the bacteria genotypes, transcripts, proteins, metabolites. Furthermore, we can use computational resources for extensively annotating both sequences and protein structures. Central to this context is the implementation of databases and tools for both the functional annotation of biological entities and the biocuration of this assembled knowledge.

The first two chapters of this dissertation present the topic of antimicrobial resistance. In particular, the second chapter focuses on point variations involved in antibiotic-resistant bacteria. Given the gap between sequences and the structural annotations of these proteins, I developed the database PVAR3D to study in 3D the variations inducing antimicrobial resistance. The database provides information on sequences, variations, structures, pathogens, antibiotics, and the literature. Furthermore, PVAR3D is annotated with functional features derived from other databases among which Pfam, GO, String, Kegg, EC. In the fight against pathogenic bacteria, structural information is pivotal. From this we can understand the proteins mechanisms of action and interactions with other molecules, aiding in the design of new means/drugs able to counteract and interfere with the evolved mechanisms of resistance. The study of bacteria within different ecological niches has revealed the huge taxonomic diversity and functional profiles characteristic of each ecosystem. In chapter 4, we presented XenoPath a pipeline for the analysis of metagenomic data aiming at uncovering the xenobiotic degradation potential of the bacterial community. Xenobiotics are compounds non normally found in an ecosystem: drugs, food conservatives, additives, and also hydrocarbons, plastic and dyes, oil, and in general hazardous waste. Certain bacteria can degrade these compounds thanks to the expression of specific enzymes. Presently, we lack a complete overview of the degradation potential of microorganisms, and in this XenoPath can help to describe the meta-phenotype of a community. In particular, it identifies the microbial taxa, the functions, and pathways involved in the bioremediation processes.

In chapters 5-8, the thesis focus goes on the study of the gut microbiota in ageing, with an emphasis with respect to human health. Particularly, chapter 6 studies the Italian cohort of the European project NU-AGE. The associations between the gut microbiome and fat distribution in the elderly are still poorly characterized, despite the recognized importance of these factors in determining healthy aging. This study shows that specific gut microbial consortia with distinct compositional traits are associated with healthier metabolic profiles, low visceral adiposity, and diet habits. The presented findings represent a step forward in understanding the role of the gut microbiota in supporting healthy living.

The study of antimicrobial resistance, in chapter 7, is viewed with respect to the gut microbiota. In this study, it has been analyzed the resistome of an Italian cohort of an extremely aged population (> 105 years). Antibiotic resistance is widespread among different ecosystems, and in human it plays a key role in reshaping the composition of the gut microbiota, enhancing the ecological fitness of certain bacterial populations when exposed to antibiotics. A considerable component of the definition of healthy aging and longevity is associated to the structure of the gut microbiota. In this respect, the presence of antibiotic-resistant bacteria is critical to many pathologies befalling with ageing. In this regard, the characterization of the resistome has not been sufficiently elucidated. The results of this work revealed specific host-gut resistome compositionality through longevity.

In chapter 8, it is presented a published paper investigating the xenobiotic degradation potential of centenarians' gut microbiota. Particularly, we confirmed a decreased taxonomic diversity along with ageing. Furthermore, we found a marked difference between the functional profile between the young and the centenarians microbial community. The latter being characterized by a higher presence of sequences annotated in the ethylbenzene, chlorobenzene, chlorocyclohexane, and toluene degradation pathways.

In conclusion, this thesis work has focused on developing and adopting computational resources aiming at investigating protein structures in pathogens and profiling taxonomy and functions of bacterial communities, with a focus over human and environmental health. Importantly, the database, available through the web server, can be adopted for structural and functional characterization on antimicrobial-resistant variations. Taken together, these data provide an unprecedented hub of annotation over structures and missense variations, that are freely downloadable and can be adopted by independent researchers. This work also describes how the metagenomic approaches have determined a paradigm shift in the characterization of microbial communities. In this respect, the focus was on the gut microbiota in ageing and longevity. Furthermore, here is presented a tool for the analysis of metagenomic data and specifically for describing the set of enzymes involved in the xenobiotic degradation pathways in metagenomic sequences.

Given the blast of big data in biology, Bioinformatics and Computational biology can successfully face the current challenge by providing computational means: algorithms, databases, and tools. To this aim, it is highly relevant the role of biocuration, data mining and statistics, able to extract, associate and reveal different types of biological information. Taken together these approaches can span different layers of biological complexity: going from genotype to phenotype, from structure to function, helping researchers in organizing the biological information, annotating biological entities, and formulating new hypothesis.

Appendix A

Appendix - Supplementary material

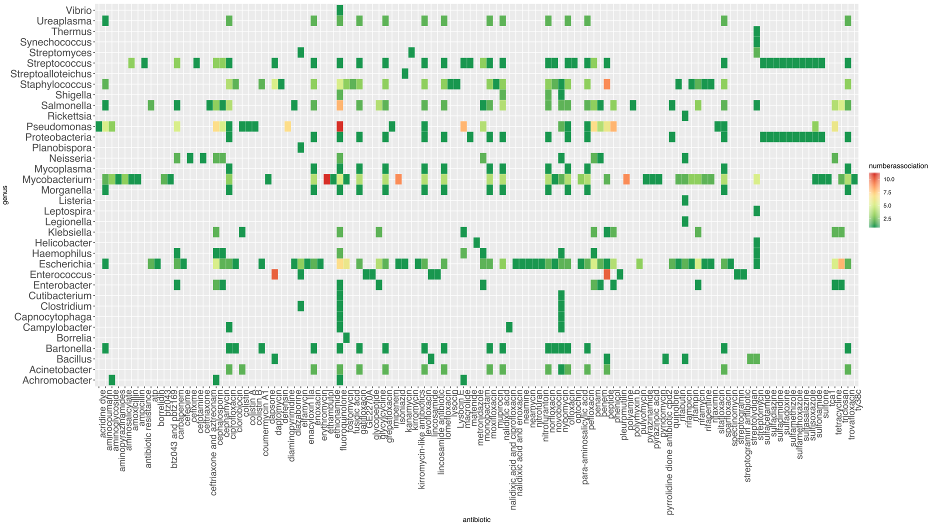


FIGURE S1: **Drug-Genus specific associations for proteins with target alterations.** The color code shows the number of associations for mutated proteins of a given Genus, with respect to the drug they have been annotated as resistant. The heatmap gives an overview on the type of resistance per Genus based on the data in PVAR3D.

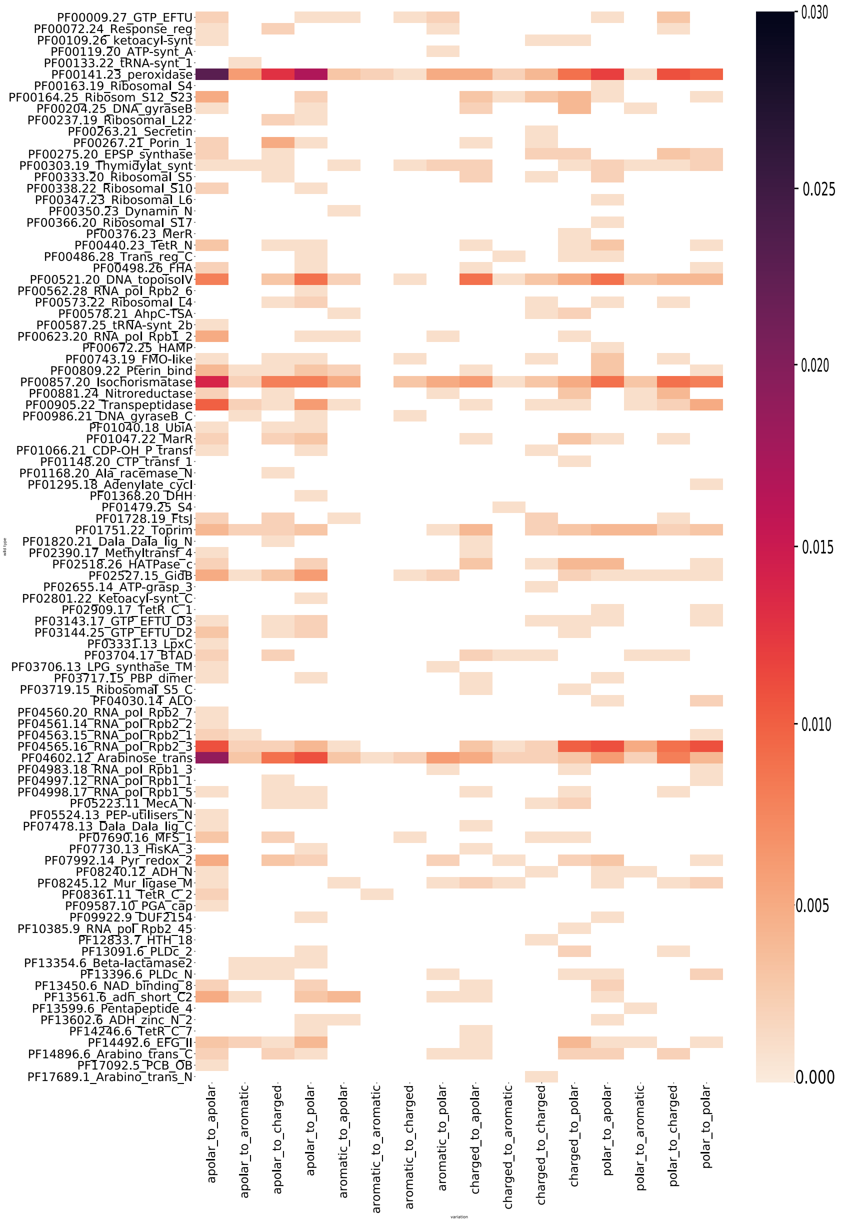


FIGURE S2: **Heatmap showing the frequency of the type of variations mapped within each protein family domain identified in PVAR3D.** Domains are on the rows and the type of amino acid changes on the columns, high frequency is depicted by increasing tones of red.

TABLE S1: **List of protein in the clusters defined by pairwise structural alignment with PDB in PVAR3D.**

clusters	protein	Accession-PDB
1	rpoB	CCP43410.1-6FBV_C
1	rpoB	P37870-2LY7_A
1	rpoB	P60281-6FED_C

Table S1 continued from previous page

clusters	protein	Accession-PDB
2	gyrA	AAC75291.1-1ZI0_A
2	gyrA	CCP42728.1-4G3N_A
3	embR	CCP44023.1-2FF4_A
4	parC	AAC76055.1-4MN4_A
5	acrR	NP_414997.1-3BCG_A
5	mexZ	NP_250710.1-2WUI_A
5	nalD	NP_252264.1-5H9T_A
6	ethR	WP_003399797.1-6HO4_A
7	dprE1	A0R607-4F4Q_A
7	dprE2	P9WJF1-6HFW_A
8	rplD	P60723-6QUL_E
9	katG	NP_216424.1-4C51_A
10	thyA	NP_217280.1-4FQS_A
11	ahpC	WP_003412529.1-2BMX_A
12	ileS	CAA52296.1-1QU3_A
13	soxR	AAC77033.1-2ZHH_A
14	marR	AAC74603.2-5H3R_A
14	mexR	NP_249115.1-1LNW_A
15	gyrA	AAC75291.1-3NUH_A
15	gyrA	CCP42728.1-3IFZ_A
15	gyrA	NP_708120.1-2Y3P_A
15	gyrA	P20831-5BS3_B
15	parC	AAK74984.1-4I3H_A
16	gyrB	BAE77595.1-3NUH_B
17	EF-Tu	NP_417798.1-5JBQ_A
18	folC	NP_216963.1-2VOS_A
19	npmA	A8C927-4OX9_Y
19	kamB	P25920-3MQ2_A
20	tlyA	AAK46002.1-5EOV_A
21	fabG	NP_415611.1-1Q7C_A
21	fabI	NP_415804.1-5CG2_A
21	inhA	CCP44244.1-6EP8_A
22	pncA	CCP44816.1-3PL1_A
23	embC	P9WNL5-3PTY_A
24	ribD	NP_217187.1-6DE5_A
25	folP	WP_000764731.1-1AJZ_A
26	kasA	CCP45025.1-2WGG_A
27	rpsL	P0A7S3-6C4I_l
27	rpsL	P17293-4V8X_AL
27	rpsL	P21472-6HA8_l
27	rpsQ	P0AG63-6Q98_v
28	rpsA	CCP44394.1-4NNG_A
29	omp36	AAK11270.1-5O9C_A
29	ompF	NP_415449.1-3HWB_A
30	oprD	NP_249649.1-4FOZ_A
31	blaC	P9WKD3-3ZHH_A
31	blaF	Q59517-2CC1_A

Table S1 continued from previous page

clusters	protein	Accession-PDB
32	PBP2b	NP_359110.1-2WAF_A
32	pbpX	P59676-5OJ1_A
32	penA	WP_003703066.1-5KSH_A
33	rpsE	P0A7W1-6Q98_j
33	rpsE	P21467-6HA8_e
34	parE	WP_000195296.1-1S16_A
34	gyrB	P9WG45-3ZM7_A
34	gyrB	BAE77595.1-6ENG_A
35	nfsA	NP_415372.1-1F5V_A
35	rdxA	O25608-3QDL_A
36	thrS	P0A8M3-1QF6_A

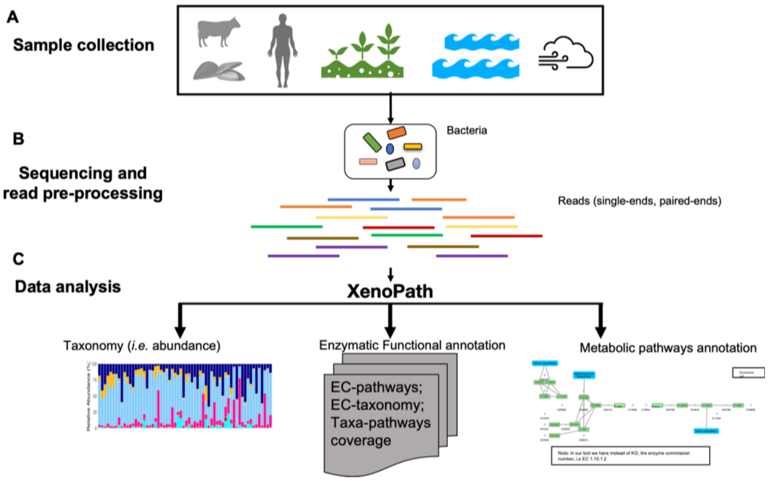


FIGURE S3: **Analysing xenobiotic pathways in a microbial community.** A) DNA extracted from samples is B) sequenced and the reads are pre-processed, the tool XenoPath C) can give in output the reads taxonomy, the functional annotation and the metabolic pathways.

TABLE S2: **Age group-specific antibiotic resistant determinants (ARDs).**
For each ARD, identified by the Wald test (Bonferroni corrected p value ≤ 0.05), the name of the gene, the family, the resistance mechanism and the antibiotic to which the resistance is conferred are reported. The annotation was manually curated from the Antibiotic Resistant Ontology (ARO) from the CARD database.

Genes	**AR family name**	**AR mechanism**	**antibiotic**
aad(6)	Aminoglycoside nucleotidyltransferase	antibiotic inactivation	aminoglycoside
acrD	resistance-nodulation-cell division (RND) antibiotic efflux pump	antibiotic efflux	aminoglycoside
acrE	resistance-nodulation-cell division (RND) antibiotic efflux pump	antibiotic efflux	multidrug
ANT(9)-Ia	Aminoglycoside nucleotidyltransferase ANT(9)	antibiotic inactivation	aminoglycoside
arnD	hydrolase (lipopolysaccharide biosynthesis)	antibiotic target protection	glycopeptide
bcr	transmembrane transporter activity	antibiotic efflux	multidrug
Bl2e_cepa	Beta-lactamase	antibiotic inactivation	beta_lactam
cblA-1	Beta-lactamase	antibiotic target alteration	beta_lactam
cpxA	membrane-localized sensor kinase promoting efflux complex expression	antibiotic efflux	multidrug
emrB	translocase of antibiotic efflux pump: in the emrB -TolC efflux protein	antibiotic efflux	multidrug
emrD	antibiotic efflux pump: major facilitator superfamily (MFS)	antibiotic efflux	multidrug
emrY	antibiotic efflux pump: major facilitator superfamily (MFS)	antibiotic efflux	multidrug
ermB	antibiotic efflux pump: major facilitator superfamily (MFS)	antibiotic efflux	macrolide-lincosamide-streptogramin
ermF	erm 23S ribosomal RNA methyltransferase	antibiotic target alteration	macrolide-lincosamide-streptogramin
gadW	AraC-family regulator of mdtEF/resistance-nodulation-cell division (RND) antibiotic efflux pump	antibiotic efflux	multidrug
gadX	AraC-family regulator of mdtEF/resistance-nodulation-cell division (RND) antibiotic efflux pump	antibiotic efflux	multidrug
leuO	transcriptional activator: LySR family transcriptot factor, activator of the MdtNOP efflux pump	antibiotic efflux	sulfonamide
lnuA	lincosamide nucleotidyltransferase (LNU)	antibiotic inactivation	macrolide-lincosamide-streptogramin
mdfA	transmembrane transport	antibiotic efflux	multidrug
mdtA	resistance-nodulation-cell division (RND) antibiotic efflux pump	antibiotic efflux	multidrug
mdtB	resistance-nodulation-cell division (RND) antibiotic efflux pump	antibiotic efflux	multidrug
mdtC	resistance-nodulation-cell division (RND) antibiotic efflux pump	antibiotic efflux	multidrug
mdtD	transmembrane transporter activity	antibiotic efflux	multidrug
mdtG	antibiotic efflux pump: major facilitator superfamily (MFS)	antibiotic efflux	multidrug
mdtH	antibiotic efflux pump: major facilitator superfamily (MFS)	antibiotic efflux	multidrug
mdtK	transmembrane transporter activity	antibiotic efflux	multidrug
mdtL	transmembrane transporter activity	antibiotic efflux	multidrug
mdtN	antibiotic efflux pump: major facilitator superfamily (MFS)	antibiotic efflux	multidrug
mdtO	antibiotic efflux pump: major facilitator superfamily (MFS)	antibiotic efflux	multidrug
mdtP	antibiotic efflux pump: major facilitator superfamily (MFS)	antibiotic efflux	multidrug
mdtQ	transmembrane transporter activity	antibiotic efflux	multidrug
mexW	resistance-nodulation-cell division (RND) antibiotic efflux pump	antibiotic efflux	multidrug
OXA-34	Beta-lactamase	antibiotic inactivation	beta_lactam
robA	positive regulator of acrAB efflux	antibiotic efflux	multidrug
rphB	phosphotransferase	antibiotic inactivation	rifampin
SAT-4	Streptothricine-acetyl-transferase	antibiotic inactivation	aminoglycoside
tcr3	tetracycline efflux pump	antibiotic efflux	tetracycline
tetD	antibiotic efflux pump: major facilitator superfamily (MFS)	antibiotic efflux	tetracycline
tolC	transmembrane transporter activity	antibiotic efflux	multidrug

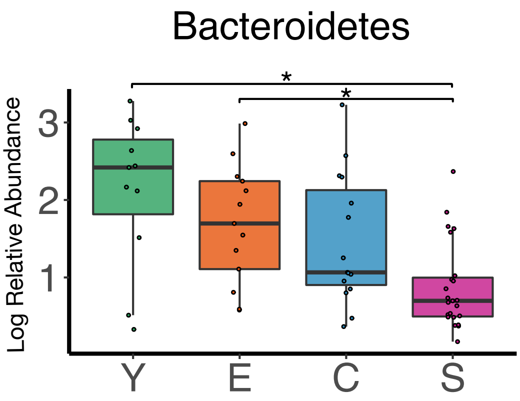


FIGURE S4:
Relative abundance of AR reads assigned to the Bacteroidetes phylum across ageing. Boxplots showing the relative abundance distribution of AR reads assigned to Bacteroidetes in the gut resistome of young adults (Y), younger elderly (E), centenarians (C) and semi-supercentenarians (S). An age-related decrease was found (Wilcoxon test, Bonferroni corrected p value = 0.05, "*").

Bibliography

Blair, Jessica M.A. et al. (2015). "Molecular mechanisms of antibiotic resistance". In: *Nature Reviews Microbiology* 13.1, pp. 42–51. ISSN: 17401534. DOI: 10.1038/nrmicro3380.

Aboelsoud, Neveen H (2010). "Herbal medicine in ancient Egypt". In: *Journal of Medicinal Plants Research* 4.2, pp. 082–086.

Hutchings, Matthew I., Andrew W. Truman, and Barrie Wilkinson (2019). "Editorial overview: Antimicrobials: Tackling AMR in the 21st century". In: *Current Opinion in Microbiology* 51, pp. iii–v. ISSN: 18790364. DOI: 10.1016/j.mib.2019.11.004.

Munita, Jose M. and Cesar A. Arias (2016). "Mechanisms of Antibiotic Resistance". In: *Virulence Mechanisms of Bacterial Pathogens*, pp. 481–511. DOI: 10.1128/9781555819286.ch17.

O'Neill, Jim (2016). "Report on Antimicrobial Resistance". In: URL: https://amr-review.org..

Kollef, Marin H and Victoria J Fraser (2001). "Antibiotic resistance in the intensive care unit". In: *Annals of internal medicine* 134.4, pp. 298–314. DOI: 10.7326/0003-4819-134-4-200102200-00014.

Christiaens, Thierry CM, Asbjørn Digranes, and Anders Baerheim (2002). "The relation between sale of antimicrobial drugs and antibiotic resistance in uropathogens in general practice". In: *Scandinavian journal of primary health care* 20.1, pp. 45–49. DOI: 10.1080/028134302317282743.

Chokshi, Aastha et al. (2019). "Global contributors to antibiotic resistance". In: *Journal of global infectious diseases* 11.1, p. 36.

Zipperer, Alexander et al. (2016). "Human commensals producing a novel antibiotic impair pathogen colonization". In: *Nature* 535.7613, pp. 511–516. ISSN: 14764687. DOI: 10.1038/nature18634.

Lincke, Thorger et al. (2010). "Closthioamide: An Unprecedented Polythioamide Antibiotic from the Strictly Anaerobic Bacterium Clostridium cellulolyticum". In: *Angewandte Chemie* 122.11, pp. 2055–2057. ISSN: 1521-3757. DOI: 10.1002/ange.200906114.

Smith, Kenneth P. and James E. Kirby (2018). "The inoculum effect in the era of multidrug resistance: Minor differences in inoculum have dramatic effect on MIC Determination". In: *Antimicrobial Agents and Chemotherapy* 62.8. ISSN: 10986596. DOI: 10.1128/AAC.00433-18.

Belkum, Alex van et al. (2020). "Innovative and rapid antimicrobial susceptibility testing systems". In: *Nature Reviews Microbiology* 18.5, pp. 299–311. ISSN: 17401534. DOI: 10.1038/s41579-020-0327-x.

Bayot ML, Bragg BN (2020). *Antimicrobial susceptibility testing*. DOI: https://www.ncbi.nlm.nih.gov/books/NBK539714/.

Yilmaz, Özlem and Ebru Demiray (2007). "Clinical role and importance of fluorescence in situ hybridization method in diagnosis of H pylori infection and determination of clarithromycin resistance in H pylori eradication therapy". In:

World Journal of Gastroenterology 13.5, pp. 671–675. ISSN: 10079327. DOI: 10.3748/wjg.v13.i5.671.

Choi, Jungil et al. (2014). "A rapid antimicrobial susceptibility test based on single-cell morphological analysis". In: *Science translational medicine* 6.267, 267ra174–267ra174.

Torres-Cortés, Gloria et al. (2011). "Characterization of novel antibiotic resistance genes identified by functional metagenomics on soil samples". In: *Environmental Microbiology* 13.4, pp. 1101–1114. ISSN: 14622912. DOI: 10.1111/j.1462-2920.2010.02422.x.

Soucy, Shannon M., Jinling Huang, and Johann Peter Gogarten (2015). "Horizontal gene transfer: Building the web of life". In: *Nature Reviews Genetics* 16.8, pp. 472–482. ISSN: 14710064. DOI: 10.1038/nrg3962.

Woodford, Neil and M. J. Ellington (2007). "The emergence of antibiotic resistance by mutation". In: *Clinical Microbiology and Infection* 13.1, pp. 5–18. ISSN: 14690691. DOI: 10.1111/j.1469-0691.2006.01492.x.

Durão, Paulo, Roberto Balbontín, and Isabel Gordo (2018). "Evolutionary mechanisms shaping the maintenance of antibiotic resistance". In: *Trends in microbiology* 26.8, pp. 677–691.

Floss, Heinz G and Tin-Wein Yu (2005). "Rifamycin mode of action, resistance, and biosynthesis". In: *Chemical reviews* 105.2, pp. 621–632.

Hooper, David C (2002). "Fluoroquinolone resistance among Gram-positive cocci". In: *The Lancet infectious diseases* 2.9, pp. 530–538. DOI: 10.1016/S1473-3099(02)00369-9.

Leclercq, Roland (2002). "Mechanisms of resistance to macrolides and lincosamides: Nature of the resistance elements and their clinical implications". In: *Clinical Infectious Diseases* 34.4, pp. 482–492. ISSN: 10584838. DOI: 10.1086/324626.

Brolund, Alma et al. (2010). "Molecular characterisation of trimethoprim resistance in Escherichia coli and Klebsiella pneumoniae during a two year intervention on trimethoprim use". In: *PLoS ONE* 5.2. ISSN: 19326203. DOI: 10.1371/journal.pone.0009233.

Fuda, Cosimo et al. (2004). "The basis for resistance to β-lactam antibiotics by penicillin-binding protein 2a of methicillin-resistant Staphylococcus aureus". In: *Journal of Biological Chemistry* 279.39, pp. 40802–40806.

Hegde, Subray S et al. (2005). "A fluoroquinolone resistance protein from Mycobacterium tuberculosis that mimics DNA". In: *Science* 308.5727, pp. 1480–1483. DOI: 10.1126/science.1110699.

Chen, Hsiao-Jan et al. (2011). "Identification of fusB-mediated fusidic acid resistance islands in Staphylococcus epidermidis isolates". In: *Antimicrobial agents and chemotherapy* 55.12, pp. 5842–5849. DOI: 10.1128/AAC.00592-11.

De Pascale, Gianfranco and Gerard D Wright (2010). "Antibiotic resistance by enzyme inactivation: from mechanisms to solutions". In: *Chembiochem* 11.10, pp. 1325–1334. DOI: 10.1002/cbic.201000067.

Li, Xian Zhi and Hiroshi Nikaido (2009). "Efflux-mediated drug resistance in bacteria: An update". In: *Drugs* 69.12, pp. 1555–1623. ISSN: 00126667. DOI: 10.2165/11317030-000000000-00000.

Delcour, Anne H (2009). "Outer membrane permeability and antibiotic resistance". In: *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* 1794.5, pp. 808–816.

Poirel, Laurent et al. (2015). "The mgrB gene as a key target for acquired resistance to colistin in Klebsiella pneumoniae". In: *Journal of Antimicrobial Chemotherapy* 70.1, pp. 75–80. ISSN: 14602091. DOI: 10.1093/jac/dku323.

Smith, Clyde A. et al. (2014). "Structure of the bifunctional aminoglycoside-resistance enzyme AAC(6')-Ie-APH(2")-Ia revealed by crystallographic and small-angle X-ray scattering analysis". In: *Acta Crystallographica Section D: Biological Crystallography* 70.10, pp. 2754–2764. ISSN: 13990047. DOI: 10.1107/S1399004714017635.

Huntemann, Marcel et al. (2016). "The standard operating procedure of the DOE-JGI Metagenome Annotation Pipeline (MAP v.4)". In: *Standards in Genomic Sciences* 11.1. ISSN: 19443277. DOI: 10.1186/s40793-016-0138-x.

Altschul, S. et al. (1990a). "Basic local alignment search tool". In: *Journal of Molecular Biology* 215, pp. 403–410. DOI: 10.1016/S0022-2836(05)80360-2.

Buchfink, Benjamin, Chao Xie, and Daniel H Huson (2015). "Fast and sensitive protein alignment using DIAMOND". In: *Nature Methods* 12.1, pp. 59–60. ISSN: 1548-7091. DOI: 10.1038/nmeth.3176. URL: http://www.nature.com/articles/nmeth.3176.

Jia, Baofeng et al. (2017). "CARD 2017: Expansion and model-centric curation of the comprehensive antibiotic resistance database". In: *Nucleic Acids Research* 45.D1, pp. D566–D573. ISSN: 13624962. DOI: 10.1093/nar/gkw1004.

Gibson, Molly K, Kevin J Forsberg, and Gautam Dantas (2015). "Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology". In: *The ISME journal* 9.1, pp. 207–216. DOI: 10.1038/ismej.2014.106.

Davis, James J et al. (2016). "Antimicrobial resistance prediction in PATRIC and RAST". In: *Scientific reports* 6, p. 27930. DOI: https://doi.org/10.1038/srep27930.

Arango-Argoty, Gustavo et al. (2018). "DeepARG: A deep learning approach for predicting antibiotic resistance genes from metagenomic data". In: *Microbiome* 6.1, pp. 1–15. DOI: 10.1186/s40168-018-0401-z.

Dcosta, Vanessa M. et al. (2011). "Antibiotic resistance is ancient". In: *Nature* 477.7365, pp. 457–461. ISSN: 00280836. DOI: 10.1038/nature10388.

Clemente, Jose C. et al. (2015a). "The microbiome of uncontacted Amerindians". In: *Science Advances* 1.3. ISSN: 23752548. DOI: 10.1126/sciadv.1500183.

Rampelli, Simone et al. (2015). "Metagenome Sequencing of the Hadza Hunter-Gatherer Gut Microbiota". In: *Current Biology* 25.13, pp. 1682–1693. ISSN: 0960-9822. DOI: 10.1016/J.CUB.2015.04.055.

Martinez, J. L. and F. Baquero (2000). "Mutation frequencies and antibiotic resistance". In: *Antimicrobial Agents and Chemotherapy* 44.7, pp. 1771–1777. ISSN: 00664804. DOI: 10.1128/AAC.44.7.1771-1777.2000.

Boolchandani, Manish, Alaric W. D'Souza, and Gautam Dantas (2019). "Sequencing-based methods and resources to study antimicrobial resistance". In: *Nature Reviews Genetics* 20.6, pp. 356–370. ISSN: 14710064. DOI: 10.1038/s41576-019-0108-4.

Lakin, Steven M. et al. (2017). "MEGARes: An antimicrobial resistance database for high throughput sequencing". In: *Nucleic Acids Research* 45.D1, pp. D574–D580. ISSN: 13624962. DOI: 10.1093/nar/gkw1009.

Srivastava, Abhishikha et al. (2014). "CBMAR: A comprehensive b-lactamase molecular annotation resource". In: *Database* 2014. ISSN: 17580463. DOI: 10.1093/database/bau111.

Saha, Saurav B., Vishwas Uttam, and Vivek Verma (2015). "U-CARE: User-friendly comprehensive antibiotic resistance repository of Escherichia coli". In: *Journal* *of Clinical Pathology* 68.8, pp. 648–651. ISSN: 14724146. DOI: 10.1136/jclinpath-2015-202927.

Winsor, Geoffrey L. et al. (2016). "Enhanced annotations and features for comparing thousands of Pseudomonasgenomes in the Pseudomonas genome database". In: *Nucleic Acids Research* 44.D1, pp. D646–D653. ISSN: 13624962. DOI: 10.1093/nar/gkv1227.

Liu, B. and M. Pop (2009). "ARDB–Antibiotic Resistance Genes Database". In: *Nucleic Acids Research* 37.Database-issue, pp. D443–D447. ISSN: 0305-1048. DOI: 10.1093/nar/gkn656.

Wattam, Alice R. et al. (2017). "Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center". In: *Nucleic Acids Research* 45.D1, pp. D535–D542. ISSN: 13624962. DOI: 10.1093/nar/gkw1017.

Naas, Thierry et al. (2017). "Beta-lactamase database (BLDB) – structure and function". In: *Journal of Enzyme Inhibition and Medicinal Chemistry* 32.1, pp. 917–919. DOI: 10.1080/14756366.2017.1344235.

Alcock, Brian P. et al. (2020). "CARD 2020: Antibiotic resistome surveillance with the comprehensive antibiotic resistance database". In: *Nucleic Acids Research* 48.D1, pp. D517–D525. ISSN: 13624962. DOI: 10.1093/nar/gkz935.

Berman, Helen M. et al. (Jan. 2000). "The Protein Data Bank". In: *Nucleic Acids Research* 28.1, pp. 235–242. ISSN: 0305-1048. DOI: 10.1093/nar/28.1.235. eprint: https://academic.oup.com/nar/article-pdf/28/1/235/9895144/280235.pdf. URL: https://doi.org/10.1093/nar/28.1.235.

Ashburner, Michael et al. (2000). "Gene ontology: tool for the unification of biology". In: *Nature genetics* 25.1, pp. 25–29. DOI: doi.org/10.1038/75556.

Binns, David et al. (2009). "QuickGO: A web-based tool for Gene Ontology searching". In: *Bioinformatics* 25.22, pp. 3045–3046. ISSN: 13674803. DOI: 10.1093/bioinformatics/btp536.

El-Gebali, Sara et al. (2019). "The Pfam protein families database in 2019". In: *Nucleic Acids Research* 47.D1, pp. D427–D432. ISSN: 13624962. DOI: 10.1093/nar/gky995.

Kanehisa, Minoru et al. (2017). "KEGG: New perspectives on genomes, pathways, diseases and drugs". In: *Nucleic Acids Research* 45.D1, pp. D353–D361. ISSN: 13624962. DOI: 10.1093/nar/gkw1092.

Karp, Peter D. et al. (2018). "The BioCyc collection of microbial genomes and metabolic pathways". In: *Briefings in Bioinformatics* 20.4, pp. 1085–1093. ISSN: 14774054. DOI: 10.1093/bib/bbx085.

Caspi, Ron et al. (2018). "The MetaCyc database of metabolic pathways and enzymes". In: *Nucleic acids research* 46.D1, pp. D633–D639.

Keseler, Ingrid M. et al. (2017). "The EcoCyc database: Reflecting new knowledge about Escherichia coli K-12". In: *Nucleic Acids Research* 45.D1, pp. D543–D550. ISSN: 13624962. DOI: 10.1093/nar/gkw1003.

Szklarczyk, Damian et al. (2017). "The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible". In: *Nucleic Acids Research* 45.D1, pp. D362–D368. ISSN: 13624962. DOI: 10.1093/nar/gkw937.

Zanzoni, Andreas et al. (2002). "MINT: A Molecular INTeraction database". In: *FEBS Letters* 513.1, pp. 135–140. ISSN: 00145793. DOI: 10.1016/S0014-5793(01)03293-8.

Aranda, B. et al. (2009). "The IntAct molecular interaction database in 2010". In: *Nucleic Acids Research* 38.SUPPL.1. ISSN: 03051048. DOI: 10.1093/nar/gkp878.

Wishart, David S. et al. (2018). "DrugBank 5.0: A major update to the DrugBank database for 2018". In: *Nucleic Acids Research* 46.D1, pp. D1074–D1082. ISSN: 13624962. DOI: 10.1093/nar/gkx1037.

Mendez, David et al. (2019). "ChEMBL: Towards direct deposition of bioassay data". In: *Nucleic Acids Research* 47.D1, pp. D930–D940. ISSN: 13624962. DOI: 10.1093/nar/gky1075.

Webb, Benjamin and Andrej Sali (2016). "Comparative protein structure modeling using MODELLER". In: *Current Protocols in Protein Science* 2016, pp. 2.9.1–2.9.37. ISSN: 19343663. DOI: 10.1002/cpps.20.

Prlić, Andreas et al. (2010). "Pre-calculated protein structure alignments at the RCSB PDB website". In: *Bioinformatics* 26.23, pp. 2983–2985. ISSN: 13674803. DOI: 10.1093/bioinformatics/btq572.

Laskowski, R A, M W MacArthur, and J M Thornton (1992). "PROCHECK: a program to check the stereochemical quality of protein structures". In: *Journal of applied crystallography* 26, pp. 283–291. URL: papers://3a44b00c-3356-4f71-8e28-59fa753ece5f/Paper/p226.

Kabsch, Wolfgang and Christian Sander (1983). "Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features". In: *Biopolymers* 22.12, pp. 2577–2637. ISSN: 10970282. DOI: 10.1002/bip.360221211.

Rost, Burkhard and Chris Sander (1994). "Conservation and prediction of solvent accessibility in protein families". In: *Proteins: Structure, Function, and Bioinformatics* 20.3, pp. 216–226. ISSN: 10970134. DOI: 10.1002/prot.340200303.

Ando, Hiroki et al. (2010). "Identification of katG mutations associated with high-level isoniazid resistance in Mycobacterium tuberculosis". In: *Antimicrobial agents and chemotherapy* 54.5, pp. 1793–1799. DOI: ando2010identification.

Lemaitre, Nadine et al. (1999). "Characterization of new mutations in pyrazinamide-resistant strains of mycobacterium tuberculosisand identification of conserved regions important for the catalytic activity of the pyrazinamidase PncA". In: *Antimicrobial agents and chemotherapy* 43.7, pp. 1761–1763. DOI: 10.1128/AAC.43.7.1761.

Rose, Alexander S et al. (2018). "NGL viewer: web-based molecular graphics for large complexes". In: *Bioinformatics* 34.21, pp. 3755–3758. DOI: 10.1093/bioinformatics/bty419.

Paladin, Lisanna et al. (2020). "The Feature-Viewer: a visualization tool for positional annotations on a sequence". In: *Bioinformatics* 36.10, pp. 3244–3245. DOI: 10.1093/bioinformatics/btaa055.

Yachdav, Guy et al. (2016). "MSAViewer: Interactive JavaScript visualization of multiple sequence alignments". In: *Bioinformatics* 32.22, pp. 3501–3503. ISSN: 14602059. DOI: 10.1093/bioinformatics/btw474.

Pedros-Ali, Carlos (2006). "Genomics and marine microbial ecology". In: *International Microbiology* 9, pp. 191–197.

Jansson, Janet K and Kirsten S Hofmockel (2018). "The soil microbiome—from metagenomics to metaphenomics". In: *Current opinion in microbiology* 43, pp. 162–168. DOI: 10.1016/j.mib.2018.01.013.

Sekirov, Inna et al. (2010). "Gut microbiota in health and disease". In: *Physiological reviews* 90.3, pp. 859–904. DOI: 10.1152/physrev.00045.2009.

Berendsen, Roeland L., Corné M.J. Pieterse, and Peter A.H.M. Bakker (2012). "The rhizosphere microbiome and plant health". In: *Trends in Plant Science* 17.8, pp. 478–486. ISSN: 13601385. DOI: 10.1016/j.tplants.2012.04.001.

Medlock, Gregory L. et al. (2018). "Inferring Metabolic Mechanisms of Interaction within a Defined Gut Microbiota". In: *Cell Systems* 7.3, 245–257.e7. ISSN: 24054720. DOI: 10.1016/j.cels.2018.08.003.

Massalha, Hassan et al. (2017). "Live imaging of root-bacteria interactions in a microfluidics setup". In: *Proceedings of the National Academy of Sciences of the United States of America* 114.17, pp. 4549–4554. ISSN: 10916490. DOI: 10.1073/pnas.1618584114.

Rivkina, E et al. (2004). "Microbial life in permafrost". In: *Advances in Space Research* 33.8, pp. 1215–1221. DOI: 10.1016/j.asr.2003.06.024.

Blevins, Steve M. and Michael S. Bronze (2010). "Robert Koch and the 'golden age' of bacteriology". In: *International Journal of Infectious Diseases* 14.9. ISSN: 12019712. DOI: 10.1016/j.ijid.2009.12.003.

Beveridge, Terry J (2001). "Use of the Gram stain in microbiology". In: *Biotechnic & Histochemistry* 76.3, pp. 111–118. DOI: 10.1080/bih.76.3.111.118.

Staley, J. T. and A. Konopka (1985). "Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats." In: *Annual review of microbiology* 39, pp. 321–346. ISSN: 00664227. DOI: 10.1146/annurev.mi.39.100185.001541.

Woese, Carl R and George E Fox (1977). "Phylogenetic structure of the prokaryotic domain: the primary kingdoms". In: *Proceedings of the National Academy of Sciences* 74.11, pp. 5088–5090. DOI: 10.1073/pnas.74.11.5088.

Sanger, F., S. Nicklen, and A. R. Coulson (1977). "DNA sequencing with chain-terminating inhibitors." In: *Proceedings of the National Academy of Sciences of the United States of America* 74.12, pp. 5463–5467. ISSN: 00278424. DOI: 10.1073/pnas.74.12.5463.

Logares, Ramiro et al. (2012). "Environmental microbiology through the lens of high-throughput DNA sequencing: Synopsis of current platforms and bioinformatics approaches". In: *Journal of Microbiological Methods* 91.1, pp. 106–113. ISSN: 01677012. DOI: 10.1016/j.mimet.2012.07.017.

Escobar-Zepeda, Alejandra, Arturo Vera-Ponce de Leon, and Alejandro Sanchez-Flores (2015). "The road to metagenomics: from microbiology to DNA sequencing technologies and bioinformatics". In: *Frontiers in genetics* 6, p. 348.

Margulies, Marcel et al. (2005). "Genome sequencing in microfabricated high-density picolitre reactors". In: *Nature* 437.7057, pp. 376–380. DOI: 10.1038/nature03959.

Rusk, Nicole (2011). "Torrents of sequence". In: *Nature Methods* 8.1, p. 44. ISSN: 15487091. DOI: 10.1038/nmeth.f.330.

Bentley, David R et al. (2008). "Accurate whole human genome sequencing using reversible terminator chemistry". In: *Nature* 456, pp. 53–59.

Niedringhaus, Thomas P. et al. (2011). "Landscape of next-generation sequencing technologies". In: *Analytical Chemistry* 83.12, pp. 4327–4341. ISSN: 00032700. DOI: 10.1021/ac2010857.

Gupta, Pushpendra K (2008). "Single-molecule DNA sequencing technologies for future genomics research". In: *Trends in biotechnology* 26.11, pp. 602–611.

Whipps, J. M., L. Karen, and R. C. Cooke (1988). "Mycoparasitism and plant disease control". In: *Fungi in biological control systems*, pp. 161–187. ISSN: 1880-6805. URL: http://www.jphysiolanthropol.com/content/34/1/23.

Berg, Gabriele et al. (2020). "Microbiome definition re-visited: old concepts and new challenges". In: *Microbiome* 8.1. ISSN: 20492618. DOI: 10.1186/s40168-020-00875-0.

Schouls, Leo M., Corrie S. Schot, and Jan A. Jacobs (2003). "Horizontal Transfer of Segments of the 16S rRNA Genes between Species of the Streptococcus anginosus Group". In: *Journal of Bacteriology* 185.24, pp. 7241–7246. ISSN: 00219193. DOI: 10.1128/JB.185.24.7241-7246.2003.

Caporaso, J Gregory et al. (2010). "QIIME allows analysis of high-throughput community sequencing data". In: *Nature methods* 7.5, pp. 335–336.

DeSantis, Todd Z et al. (2006). "Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB". In: *Applied and environmental microbiology* 72.7, pp. 5069–5072.

Pruesse, Elmar et al. (2007). "SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB". In: *Nucleic acids research* 35.21, pp. 7188–7196. DOI: 10.1093/nar/gkm864.

Cole, James R et al. (2005). "The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis". In: *Nucleic acids research* 33.suppl_1, pp. D294–D296. DOI: 10.1093/nar/gki038.

Altschul, Stephen F et al. (1990b). "Basic local alignment search tool". In: *Journal of molecular biology* 215.3, pp. 403–410. DOI: 10.1016/S0022-2836(05)80360-2.

Wang, Qiong et al. (2007). "Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy". In: *Applied and environmental microbiology* 73.16, pp. 5261–5267. DOI: 10.1128/AEM.00062-07.

Edgar, Robert C (2010). "Search and clustering orders of magnitude faster than BLAST". In: *Bioinformatics* 26.19, pp. 2460–2461.

Knight, R. et al. (2018). "Best practices for analysing microbiomes". In: *Nature Reviews Microbiology*.

Miller, Jason R., Sergey Koren, and Granger Sutton (2010). "Assembly algorithms for next-generation sequencing data". In: *Genomics* 95.6, pp. 315–327. ISSN: 08887543. DOI: 10.1016/j.ygeno.2010.03.001.

Nurk, Sergey et al. (2017). "MetaSPAdes: A new versatile metagenomic assembler". In: *Genome Research* 27.5, pp. 824–834. ISSN: 15495469. DOI: 10.1101/gr.213959.116.

Peng, Yu et al. (2012). "IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth". In: *Bioinformatics* 28.11, pp. 1420–1428. ISSN: 13674803. DOI: 10.1093/bioinformatics/bts174.

Mitchell, Alex L. et al. (2019). "InterPro in 2019: Improving coverage, classification and access to protein sequence annotations". In: *Nucleic Acids Research* 47.D1, pp. D351–D360. ISSN: 13624962. DOI: 10.1093/nar/gky1100.

Ye, Yuzhen and Thomas G Doak (2011). "A parsimony approach to biological pathway reconstruction/inference for metagenomes". In: *Handbook of Molecular Microbial Ecology I: Metagenomics and Complementary Approaches*, pp. 453–460. DOI: 10.1002/9781118010518.ch52.

Liu, Bo and Mihai Pop (2011). "MetaPath: identifying differentially abundant metabolic pathways in metagenomic datasets". In: 5.2, pp. 1–12. DOI: 10.1186/1753-6561-5-S2-S9.

Kim, Daehwan et al. (2016). "Centrifuge: rapid and sensitive classification of metagenomic sequences". In: *Centrifuge: Rapid and sensitive classification of metagenomic sequences*, p. 054965. ISSN: 1549-5469. DOI: 10.1101/054965.

Menzel, Peter, Kim Lee Ng, and Anders Krogh (2016). "Fast and sensitive taxonomic classification for metagenomics with Kaiju". In: *Nature Communications* 7. ISSN: 20411723. DOI: 10.1038/ncomms11257.

Ferragina, Paolo and Giovanni Manzini (2005). “Indexing compressed text”. In: *Journal of the ACM (JACM)* 52.4, pp. 552–581.

Pérez-Cobas, Ana Elena, Laura Gomez-Valero, and Carmen Buchrieser (2020). “Metagenomic approaches in microbial ecology: An update on whole-genome and marker gene sequencing analyses”. In: *Microbial Genomics* 6.8, pp. 1–22. ISSN: 20575858. DOI: 10.1099/mgen.0.000409.

Sberro, Hila et al. (2019). “Large-Scale Analyses of Human Microbiomes Reveal Thousands of Small, Novel Genes”. In: *Cell* 178.5, 1245–1259.e14. ISSN: 10974172. DOI: 10.1016/j.cell.2019.07.016.

Baric, Ralph S. et al. (2016). “Next-generation high-throughput functional annotation of microbial genomes”. In: *mBio* 7.5. ISSN: 21507511. DOI: 10.1128/mBio.01245-16.

Haiser, Henry J and Peter J Turnbaugh (2013). “Developing a metagenomic view of xenobiotic metabolism”. In: *Pharmacological research* 69.1, pp. 21–31.

Koppel, Nitzan, Vayu Maini Rekdal, and Emily P. Balskus (2017). “Chemical transformation of xenobiotics by the human gut microbiota”. In: *Science* 356.6344, pp. 1246–1257. ISSN: 10959203. DOI: 10.1126/science.aag2770.

Ellis, Lynda BM and Lawrence P Wackett (2012). “Use of the University of Minnesota Biocatalysis/Biodegradation Database for study of microbial degradation”. In: *Microbial informatics and experimentation* 2.1, p. 1.

Russell, Robyn J. et al. (2011). “The evolution of new enzyme function: Lessons from xenobiotic metabolizing bacteria versus insecticide-resistant insects”. In: *Evolutionary Applications* 4.2, pp. 225–248. ISSN: 17524563. DOI: 10.1111/j.1752-4571.2010.00175.x.

Van Der Meer, Jan Roelof et al. (1992). “Molecular mechanisms of genetic adaptation to xenobiotic compounds.” In: *Microbiology and Molecular Biology Reviews* 56.4, pp. 677–694.

Garrido-Cardenas, Jose Antonio and Francisco Manzano-Agugliaro (2017). “The metagenomics worldwide research”. In: *Current genetics* 63.5, pp. 819–829.

Rajilić-Stojanović, Mirjana, Hauke Smidt, and Willem M. De Vos (2007). “Diversity of the human gastrointestinal tract microbiota revisited”. In: *Environmental Microbiology* 9.9, pp. 2125–2136. ISSN: 14622912. DOI: 10.1111/j.1462-2920.2007.01369.x.

Lepage, Patricia et al. (2013). “A metagenomic insight into our gut’s microbiome”. In: *Gut* 62.1, pp. 146–158. ISSN: 00175749. DOI: 10.1136/gutjnl-2011-301805.

Donaldson, Gregory P, S Melanie Lee, and Sarkis K Mazmanian (2016). “Gut biogeography of the bacterial microbiota”. In: *Nature Reviews Microbiology* 14.1, pp. 20–32.

Ferreira, Rui M et al. (2018). “Gastric microbial community profiling reveals a dysbiotic cancer-associated microbiota”. In: *Gut* 67.2, pp. 226–236.

El Kaoutari, Abdessamad et al. (2013). “The abundance and variety of carbohydrate-active enzymes in the human gut microbiota”. In: *Nature Reviews Microbiology* 11.7, pp. 497–504.

Sengupta, Ranjita et al. (2013). “The role of cell surface architecture of lactobacilli in host-microbe interactions in the gastrointestinal tract”. In: *Mediators of Inflammation* 2013. ISSN: 09629351. DOI: 10.1155/2013/237921.

Johansson, Malin E.V. et al. (2008). “The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria”. In: *Proceedings of the National Academy of Sciences of the United States of America* 105.39, pp. 15064–15069. ISSN: 00278424. DOI: 10.1073/pnas.0803124105.

Severi, Emmanuele et al. (2005). "Sialic acid transport in Haemophilus influenzae is essential for lipopolysaccharide sialylation and serum resistance and is dependent on a novel tripartite ATP-independent periplasmic transporter". In: *Molecular Microbiology* 58.4, pp. 1173–1185. ISSN: 0950382X. DOI: 10.1111/j.1365-2958.2005.04901.x.

Lysenko, Elena S. et al. (2000). "Bacterial phosphorylcholine decreases susceptibility to the antimicrobial peptide LL-37/hCAP18 expressed in the upper respiratory tract". In: *Infection and Immunity* 68.3, pp. 1664–1671. ISSN: 00199567. DOI: 10.1128/IAI.68.3.1664-1671.2000.

Saar-Dover, Ron et al. (2012). "D-Alanylation of Lipoteichoic Acids Confers Resistance to Cationic Peptides in Group B Streptococcus by Increasing the Cell Wall Density". In: *PLoS Pathogens* 8.9. ISSN: 15537366. DOI: 10.1371/journal.ppat.1002891.

Rooks, Michelle G. and Wendy S. Garrett (2016). "Gut microbiota, metabolites and host immunity". In: *Nature Reviews Immunology* 16.6, pp. 341–352. ISSN: 14741741. DOI: 10.1038/nri.2016.42.

Clarke, Gerard et al. (2014). "Minireview: gut microbiota: the neglected endocrine organ". In: *Molecular endocrinology* 28.8, pp. 1221–1238. DOI: 10.1210/me.2014-1108.

Palm, Noah W. et al. (2014). "Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease". In: *Cell* 158.5, pp. 1000–1010. ISSN: 10974172. DOI: 10.1016/j.cell.2014.08.006.

Mathias, Amandine and Blaise Corthésy (2011). "N-glycans on secretory component: Mediators of the interaction between secretory IgA and gram-positive commensals sustaining intestinal homeostasis". In: *Gut Microbes* 2.5, pp. 287–293. ISSN: 19490984. DOI: 10.4161/gmic.2.5.18269.

Round, June L. et al. (2011). "The toll-like receptor 2 pathway establishes colonization by a commensal of the human microbiota". In: *Science* 332.6032, pp. 974–977. ISSN: 00368075. DOI: 10.1126/science.1206095.

Ivanov, Ivaylo I. et al. (2009). "Induction of Intestinal Th17 Cells by Segmented Filamentous Bacteria". In: *Cell* 139.3, pp. 485–498. ISSN: 00928674. DOI: 10.1016/j.cell.2009.09.033.

Monack, D, A Mueller, and S Falkow (2004). "Persistent bacterial infections: the interface of the pathogen and the host immune system". In: *Nature Reviews. Microbiology* 2.9, pp. 747–765. ISSN: 1740-1526. DOI: 10.5167/uzh-31197.

Huang, Le et al. (Oct. 2017). "dbCAN-seq: a database of carbohydrate-active enzyme (CAZyme) sequence and annotation". In: *Nucleic Acids Research* 46.D1, pp. D516–D521. ISSN: 0305-1048. DOI: 10.1093/nar/gkx894. eprint: https://academic.oup.com/nar/article-pdf/46/D1/D516/23162649/gkx894.pdf. URL: https://doi.org/10.1093/nar/gkx894.

Koropatkin, Nicole M., Elizabeth A. Cameron, and Eric C. Martens (2012). "How glycan metabolism shapes the human gut microbiota". In: *Nature Reviews Microbiology* 10.5, pp. 323–335. ISSN: 17401526. DOI: 10.1038/nrmicro2746.

Flint, Harry J et al. (2015). "Links between diet, gut microbiota composition and gut metabolism". In: *Proceedings of the Nutrition Society* 74.1, pp. 13–22.

Smith, Patrick M et al. (2013). "The microbial metabolites, short-chain fatty acids, regulate colonic Treg cell homeostasis". In: *Science* 341.6145, pp. 569–573. DOI: 10.1126/science.1241165.

Atarashi, Koji et al. (2013). "Treg induction by a rationally selected mixture of Clostridia strains from the human microbiota". In: *Nature* 500.7461, pp. 232–236. ISSN: 00280836. DOI: 10.1038/nature12331.

Arpaia, Nicholas et al. (2013). "Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation". In: *Nature* 504.7480, pp. 451–455. ISSN: 00280836. DOI: 10.1038/nature12726.

Sivaprakasam, Sathish, Puttur D Prasad, and Nagendra Singh (2016). "Benefits of short-chain fatty acids and their receptors in inflammation and carcinogenesis". In: *Pharmacology and Therapeutics* 164, pp. 144–151. ISSN: 1879016X. DOI: 10.1016/j.pharmthera.2016.04.007.

Sun, Mingming et al. (2017). "Microbiota metabolite short chain fatty acids, GPCR, and inflammatory bowel diseases". In: *Journal of Gastroenterology* 52.1. ISSN: 14355922. DOI: 10.1007/s00535-016-1242-9.

Ridlon, Jason M., Dae Joong Kang, and Phillip B. Hylemon (2006). "Bile salt biotransformations by human intestinal bacteria". In: *Journal of Lipid Research* 47.2, pp. 241–259. ISSN: 00222275. DOI: 10.1194/jlr.R500013-JLR200.

Yoshimoto, Shin et al. (2013). "Obesity-induced gut microbial metabolite promotes liver cancer through senescence secretome". In: *Nature* 499.7456, pp. 97–101. DOI: 10.1038/nature12347.

Ramírez-Pérez, Oscar et al. (2017). "The Role of the Gut Microbiota in Bile Acid Metabolism". In: *Annals of hepatology* 16, S21–S26. ISSN: 16652681. DOI: 10.5604/01.3001.0010.5672.

Wahlström, Annika et al. (2016). "Intestinal Crosstalk between Bile Acids and Microbiota and Its Impact on Host Metabolism". In: *Cell Metabolism* 24.1, pp. 41–50. ISSN: 19327420. DOI: 10.1016/j.cmet.2016.05.005.

Milani, Christian et al. (2017). "The First Microbial Colonizers of the Human Gut: Composition, Activities, and Health Implications of the Infant Gut Microbiota". In: *Microbiology and Molecular Biology Reviews* 81.4. ISSN: 1092-2172. DOI: 10.1128/mmbr.00036-17.

Bearfield, C (2002). "Possible association between amniotic fluid micro-organism infection and microflora in the mouth". In: *BJOG: An International Journal of Obstetrics and Gynaecology* 109.5, pp. 527–533. ISSN: 14700328. DOI: 10.1016/s1470-0328(02)01349-6.

Jiménez, Esther et al. (2005). "Isolation of commensal bacteria from umbilical cord blood of healthy neonates born by cesarean section". In: *Current Microbiology* 51.4, pp. 270–274. ISSN: 03438651. DOI: 10.1007/s00284-005-0020-3.

Aagaard, Kjersti et al. (2014). "The placenta harbors a unique microbiome". In: *Science translational medicine* 6.237, 237ra65–237ra65. DOI: 10.1126/scitranslmed.3008599.

Dominguez-Bello, Maria G et al. (2010). "Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns". In: *Proceedings of the National Academy of Sciences* 107.26, pp. 11971–11975. DOI: 10.1073/pnas.1002601107.

Kero, Jukka et al. (2002). "Mode of Delivery and Asthma – Is There a Connection?" In: *Pediatric Research* 52.1, pp. 6–11. ISSN: 0031-3998. DOI: 10.1203/00006450-200207000-00004.

Huh, Susanna Y. et al. (2012). "Delivery by caesarean section and risk of obesity in preschool age children: A prospective cohort study". In: *Archives of Disease in*

Childhood 97.7, pp. 610–616. ISSN: 00039888. DOI: 10.1136/archdischild-2011-301141.

Decker, Evalotte, Mathias Hornef, and Silvia Stockinger (2011). "Cesarean delivery is associated with celiac disease but not inflammatory bowel disease in children". In: *Gut microbes* 2.2, pp. 91–98.

Algert, CS et al. (2009). "Perinatal risk factors for early onset of Type 1 diabetes in a 2000–2005 birth cohort". In: *Diabetic medicine* 26.12, pp. 1193–1197. DOI: 10.1111/j.1464-5491.2009.02878.x.

Marques, Tatiana Milena et al. (2010). "Programming infant gut microbiota: Influence of dietary and environmental factors". In: *Current Opinion in Biotechnology* 21.2, pp. 149–156. ISSN: 09581669. DOI: 10.1016/j.copbio.2010.03.020.

Bäckhed, Fredrik et al. (2015). "Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life". In: *Cell Host & Microbe* 17.6, p. 852. ISSN: 19313128. DOI: 10.1016/j.chom.2015.05.012.

Arrieta, Marie Claire et al. (2014). "The intestinal microbiome in early life: Health and disease". In: *Frontiers in Immunology* 5.AUG. ISSN: 16643224. DOI: 10.3389/fimmu.2014.00427.

Agans, Richard et al. (2011). "Distal gut microbiota of adolescent children is different from that of adults". In: *FEMS Microbiology Ecology* 77.2, pp. 404–412. ISSN: 01686496. DOI: 10.1111/j.1574-6941.2011.01120.x.

Derrien, Muriel, Anne-Sophie Alvarez, and Willem M de Vos (2019). "The gut microbiota in the first decade of life". In: *Trends in microbiology* 27.12, pp. 997–1010. DOI: 10.1016/j.tim.2019.08.001.

Bäckhed, Fredrik (2012). "Host responses to the human microbiome". In: *Nutrition Reviews* 70.SUPPL. 1. ISSN: 00296643. DOI: 10.1111/j.1753-4887.2012.00496.x.

Peterson, Daniel A. et al. (2008). "Metagenomic Approaches for Defining the Pathogenesis of Inflammatory Bowel Diseases". In: *Cell Host and Microbe* 3.6, pp. 417–427. ISSN: 19313128. DOI: 10.1016/j.chom.2008.05.001.

Zoetendal, Erwin G., Antoon D.L. Akkermans, and Willem M. De Vos (1998). "Temperature gradient gel electrophoresis analysis of 16S rRNA from human fecal samples reveals stable and host-specific communities of active bacteria". In: *Applied and Environmental Microbiology* 64.10, pp. 3854–3859. ISSN: 00992240. DOI: 10.1128/aem.64.10.3854-3859.1998.

Eckburg, Paul B et al. (2005). "Diversity of the human intestinal microbial flora". In: *science* 308.5728, pp. 1635–1638.

Costello, Elizabeth K et al. (2009). "Bacterial community variation in human body habitats across space and time". In: *Science* 326.5960, pp. 1694–1697.

Zmora, Niv, Jotham Suez, and Eran Elinav (2019). "You are what you eat: diet, health and the gut microbiota". In: *Nature Reviews Gastroenterology and Hepatology* 16.1, pp. 35–56. ISSN: 17595053. DOI: 10.1038/s41575-018-0061-2.

Qin, Junjie et al. (2010). "A human gut microbial gene catalogue established by metagenomic sequencing". In: *Nature* 464.7285, pp. 59–65. ISSN: 14764687. DOI: 10.1038/nature08821.

Consortium, Human Microbiome Jumpstart Reference Strains et al. (2010). "A catalog of reference genomes from the human microbiome". In: *Science* 328.5981, pp. 994–999. DOI: 10.1126/science.1183605.

Integrative, HMP et al. (2019). "The integrative human microbiome project". In: *Natur* 569.7758, pp. 641–648. DOI: 10.1038/s41586-019-1238-8.

Faith, Jeremiah J et al. (2013). "The long-term stability of the human gut microbiota". In: *Science* 341.6141. DOI: 10.1126/science.1237439.

M., Rajilić-Stojanović and de Vos W.M. (2014). "The first 1000 cultured species of the human gastrointestinal microbiota". In: *FEMS Microbiology Reviews* 38.5, pp. 996–1047. ISSN: 1574-6976. DOI: 10.1111/1574-6976.12075.

Franzosa, Eric A et al. (2015). "Identifying personal microbiomes using metagenomic codes". In: *Proceedings of the National Academy of Sciences* 112.22, E2930–E2938. DOI: 10.1073/pnas.1423854112.

Palleja, Albert et al. (2018). "Recovery of gut microbiota of healthy adults following antibiotic exposure". In: *Nature microbiology* 3.11, pp. 1255–1265. DOI: 10.1038/s41564-018-0257-9.

David, Lawrence A et al. (2014). "Diet rapidly and reproducibly alters the human gut microbiome". In: *Nature* 505.7484, pp. 559–563. DOI: 10.1038/nature12820.

Rothschild, Daphna et al. (2018). "Environment dominates over host genetics in shaping human gut microbiota". In: *Nature* 555.7695, pp. 210–215. DOI: 10.1038/nature25973.

He, Yan et al. (2018). "Regional variation limits applications of healthy gut microbiome reference ranges and disease models". In: *Nature medicine* 24.10, pp. 1532–1535. DOI: 10.1038/s41591-018-0164-x.

Deschasaux, Mélanie et al. (2018). "Depicting the composition of gut microbiota in a population with varied ethnic origins but shared geography". In: *Nature medicine* 24.10, pp. 1526–1531. DOI: 10.1038/s41591-018-0160-1.

Antonopoulos, Dionysios A. et al. (2009). "Reproducible community dynamics of the gastrointestinal microbiota following antibiotic perturbation". In: *Infection and Immunity* 77.6, pp. 2367–2375. ISSN: 00199567. DOI: 10.1128/IAI.01520-08.

Franzosa, Eric A et al. (2019). "Gut microbiome structure and metabolic activity in inflammatory bowel disease". In: *Nature microbiology* 4.2, pp. 293–305.

Han, Hui et al. (2018). "Gut microbiota and type 1 diabetes". In: *International journal of molecular sciences* 19.4, p. 995. DOI: 10.3390/ijms19040995.

Sharma, Sapna and Prabhanshu Tripathi (2019). "Gut microbiome and type 2 diabetes: where we are and where to go?" In: *The Journal of nutritional biochemistry* 63, pp. 101–108. DOI: 10.1016/j.jnutbio.2018.10.003.

Gomes, Aline Corado, Christian Hoffmann, and João Felipe Mota (2018). "The human gut microbiota: Metabolism and perspective in obesity". In: *Gut microbes* 9.4, pp. 308–325.

J.S., Messer et al. (2017). "Evolutionary and ecological forces that shape the bacterial communities of the human gut". In: *Mucosal Immunology* 10.3, pp. 567–579. ISSN: 1935-3456. DOI: 10.1038/mi.2016.138.

Cho, Judy H (2008). "The genetics and immunopathogenesis of inflammatory bowel disease". In: *Nature Reviews Immunology* 8.6, pp. 458–466.

Bevins, Charles L. and Nita H. Salzman (2011). "Paneth cells, antimicrobial peptides and maintenance of intestinal homeostasis". In: *Nature Reviews Microbiology* 9.5, pp. 356–368. ISSN: 17401526. DOI: 10.1038/nrmicro2546.

Swidsinski, Alexander et al. (2005). "Spatial organization and composition of the mucosal flora in patients with inflammatory bowel disease". In: *Journal of Clinical Microbiology* 43.7, pp. 3380–3389. ISSN: 00951137. DOI: 10.1128/JCM.43.7.3380-3389.2005.

Krogfelt, K. A., H. Bergmans, and P. Klemm (1990). "Direct evidence that the FimH protein is the mannose-specific adhesion of Escherichia coli type 1 fimbriae". In:

Infection and Immunity 58.6, pp. 1995–1998. ISSN: 00199567. DOI: 10.1128/iai.58.6.1995-1998.1990.

Kim, Sangkyu and S. Michal Jazwinski (2018). "The Gut Microbiota and Healthy Aging: A Mini-Review". In: *Gerontology* 64.6, pp. 513–520. ISSN: 14230003. DOI: 10.1159/000490615.

O'Toole, Paul W and Ian B Jeffery (2015). "Gut microbiota and aging". In: *Science* 350.6265, pp. 1214–1215. DOI: 10.1126/science.aac8469.

Araos, Rafael et al. (2019). "Fecal Microbiome Characteristics and the Resistome Associated With Acquisition of Multidrug-Resistant Organisms Among Elderly Subjects". In: *Frontiers in Microbiology* 10, p. 2260. DOI: 10.3389/fmicb.2019.02260.

Cani, Patrice D. et al. (2019). "Microbial regulation of organismal energy homeostasis". In: *Nature Metabolism* 1.1, pp. 34–46. ISSN: 25225812. DOI: 10.1038/s42255-018-0017-4.

Spanogiannopoulos, Peter et al. (2016). "The microbial pharmacists within us: A metagenomic view of xenobiotic metabolism". In: *Nature Reviews Microbiology* 14.5, pp. 273–287. ISSN: 17401534. DOI: 10.1038/nrmicro.2016.17.

Kim, Donghyun, Melody Y Zeng, and Gabriel Núñez (2017). "The interplay between host immune cells and gut microbiota in chronic inflammatory diseases". In: *Experimental & molecular medicine* 49.5, e339–e339. DOI: 10.1038/emm.2017.24.

Claesson, Marcus J et al. (2012). "Gut microbiota composition correlates with diet and health in the elderly". In: *Nature* 488.7410, pp. 178–184.

Biagi, Elena et al. (2010). "Through ageing, and beyond: Gut microbiota and inflammatory status in seniors and centenarians". In: *PLoS ONE* 5.5. ISSN: 19326203. DOI: 10.1371/journal.pone.0010667.

Ventura, Maria Teresa et al. (2017). "Immunosenescence in aging: Between immune cells depletion and cytokines up-regulation". In: *Clinical and Molecular Allergy* 15. ISSN: 14767961. DOI: 10.1186/s12948-017-0077-0.

Nicoletti, Claudio (2015). "Age-associated changes of the intestinal epithelial barrier: Local and systemic implications". In: *Expert Review of Gastroenterology and Hepatology* 9.12, pp. 1467–1469. ISSN: 1747-4132. DOI: 10.1586/17474124.2015.1092872.

Villa, Christopher R., Wendy E. Ward, and Elena M. Comelli (2017). "Gut microbiota-bone axis". In: *Critical Reviews in Food Science and Nutrition* 57.8, pp. 1664–1672. ISSN: 15497852. DOI: 10.1080/10408398.2015.1010034.

Leung, Katherine and Sandrine Thuret (2015). "Gut Microbiota: A Modulator of Brain Plasticity and Cognitive Function in Ageing". In: *Healthcare* 3.4, pp. 898–916. ISSN: 2227-9032. DOI: 10.3390/healthcare3040898.

Rampelli, Simone et al. (2013). "Functional metagenomic profiling of intestinal microbiome in extreme ageing". In: *Aging* 5.12, pp. 902–912. ISSN: 19454589. DOI: 10.18632/aging.100623.

Biagi, Elena et al. (2016). "Gut Microbiota and Extreme Longevity". In: *Current Biology* 26.11, pp. 1480–1485. DOI: 10.1016/j.cub.2016.04.016.

Santoro, Aurelia et al. (2018b). "Gut microbiota changes in the extreme decades of human life: a focus on centenarians". In: *Cellular and Molecular Life Sciences* 75.1, pp. 129–148. ISSN: 14209071. DOI: 10.1007/s00018-017-2674-y.

Santoro, Aurelia et al. (2018a). "A cross-sectional analysis of body composition among healthy elderly from the European NU-AGE study: Sex and country specific features". In: *Frontiers in Physiology* 9. ISSN: 1664042X. DOI: 10.3389/fphys.2018.01693.

St-Onge, Marie Pierre and Dympna Gallagher (2010). “Body composition changes with aging: The cause or the result of alterations in metabolic rate and macronutrient oxidation?” In: *Nutrition* 26.2, pp. 152–155. ISSN: 08999007. DOI: 10.1016/j.nut.2009.07.004.

Reinders, Ilse, Marjolein Visser, and Laura Schaap (2017). “Body weight and body composition in old age and their relationship with frailty”. In: *Current Opinion in Clinical Nutrition and Metabolic Care* 20.1, pp. 11–15. ISSN: 14736519. DOI: 10.1097/MCO.0000000000000332.

Bazzocchi, Alberto et al. (2013). “Health and ageing: A cross-sectional study of body composition”. In: *Clinical Nutrition* 32.4, pp. 569–578. ISSN: 02615614. DOI: 10.1016/j.clnu.2012.10.004.

Ponti, Federico et al. (2020). “Aging and Imaging Assessment of Body Composition: From Fat to Facts”. In: *Frontiers in Endocrinology* 10, p. 861. ISSN: 1664-2392. DOI: 10.3389/fendo.2019.00861.

Conte, Maria et al. (2019). “The dual role of the pervasive "Fattish" Tissue remodeling with age”. In: *Frontiers in Endocrinology* 10.FEB. ISSN: 16642392. DOI: 10.3389/fendo.2019.00114.

Fried, Susan K., Dove A. Bunkin, and Andrew S. Greenberg (1998). “Omental and subcutaneous adipose tissues of obese subjects release interleukin-6: Depot difference and regulation by glucocorticoid”. In: *Journal of Clinical Endocrinology and Metabolism* 83.3, pp. 847–850. ISSN: 0021972X. DOI: 10.1210/jc.83.3.847.

Kanda, Hajime et al. (2006). “MCP-1 contributes to macrophage infiltration into adipose tissue, insulin resistance, and hepatic steatosis in obesity”. In: *Journal of Clinical Investigation* 116.6, pp. 1494–1505.

Sato, Fumi et al. (2018). “Association of epicardial, visceral, and subcutaneous fat with cardiometabolic diseases”. In: *Circulation Journal* 82.2, pp. 502–508. ISSN: 13474820. DOI: 10.1253/circj.CJ-17-0820.

Gómez, Maria Pilar Aparisi et al. (2019). “Correlation between DXA and laboratory parameters in normal weight, overweight, and obese patients”. In: *Nutrition* 61, pp. 143–150. ISSN: 18731244. DOI: 10.1016/j.nut.2018.10.023.

Turnbaugh, Peter J et al. (2006). “An obesity-associated gut microbiome with increased capacity for energy harvest”. In: *nature* 444.7122, p. 1027.

Turnbaugh, Peter J. et al. (2008). “Diet-Induced Obesity Is Linked to Marked but Reversible Alterations in the Mouse Distal Gut Microbiome”. In: *Cell Host and Microbe* 3.4, pp. 213–223. ISSN: 19313128. DOI: 10.1016/j.chom.2008.02.015.

Ridaura VK et al. (2013). “Gut microbiota from twins discordant for obesity modulate metabolism in mice.” In: *Science.* 341, p. 1241214.

Rampelli, Simone et al. (2018). “Pre-obese children’s dysbiotic gut microbiome and unhealthy diets may predict the development of obesity”. In: *Communications Biology* 1.1. ISSN: 23993642. DOI: 10.1038/s42003-018-0221-5.

Cancello, Raffaella et al. (2019). “Effect of short-term dietary intervention and probiotic mix supplementation on the gut microbiota of elderly obese women”. In: *Nutrients* 11.12. ISSN: 20726643. DOI: 10.3390/nu11123011.

Min, Yan et al. (2019). “Sex-specific association between gut microbiome and fat distribution”. In: *Nature Communications* 10.1. ISSN: 20411723. DOI: 10.1038/s41467-019-10440-5.

Beaumont, Michelle et al. (2016). “Heritable components of the human fecal microbiome are associated with visceral fat”. In: *Genome biology* 17.1, pp. 1–19.

Guglielmi, Giuseppe and Alberto Bazzocchi (2020). “Body composition imaging”. In: *Quantitative Imaging in Medicine and Surgery* 10.8, pp. 1576–1579. ISSN: 22234306. DOI: 10.21037/QIMS-2019-BC-13.

Messina, Carmelo et al. (2020). “Body composition with dual energy X-ray absorptiometry: From basics to new tools”. In: *Quantitative Imaging in Medicine and Surgery* 10.8, pp. 1687–1698. ISSN: 22234306. DOI: 10.21037/QIMS.2020.03.02.

Santoro, Aurelia et al. (2014). “Combating inflammaging through a Mediterranean whole diet approach: The NU-AGE project’s conceptual framework and design”. In: *Mechanisms of Ageing and Development* 136-137, pp. 3–13. ISSN: 18726216. DOI: 10.1016/j.mad.2013.12.001.

Marseglia, Anna et al. (2019). “Participating in mental, social, and physical leisure activities and having a rich social network reduce the incidence of diabetes-related dementia in a cohort of Swedish older adults”. In: *Diabetes Care* 42.2, pp. 232–239. ISSN: 19355548. DOI: 10.2337/dc18-1428.

Fried, L. P. et al. (2001). “Frailty in older adults: Evidence for a phenotype”. In: *Journals of Gerontology - Series A Biological Sciences and Medical Sciences* 56.3. ISSN: 10795006. DOI: 10.1093/gerona/56.3.m146.

Marseglia, Anna et al. (2018). “Effect of the NU-AGE diet on cognitive functioning in older adults: A randomized controlled trial”. In: *Frontiers in Physiology* 9.APR. ISSN: 1664042X. DOI: 10.3389/fphys.2018.00349.

Jennings, Amy et al. (2019). “Mediterranean-style diet improves systolic blood pressure and arterial stiffness in older adults: Results of a 1-year european multicenter trial”. In: *Hypertension* 73.3, pp. 578–586. ISSN: 15244563. DOI: 10.1161/HYPERTENSIONAHA.118.12259.

Santoro, Aurelia et al. (2019). “Gender-specific association of body composition with inflammatory and adipose-related markers in healthy elderly Europeans from the NU-AGE study”. In: *European Radiology* 29.9, pp. 4968–4979. ISSN: 14321084. DOI: 10.1007/s00330-018-5973-2.

Matthews, DR et al. (1985). “Homeostasis model assessment: insulin resistance and β-cell function from fasting plasma glucose and insulin concentrations in man”. In: *Diabetologia* 28.7, pp. 412–419.

Ostan, Rita et al. (2018). “Cross-sectional analysis of the correlation between daily nutrient intake assessed by 7 -day food records and biomarkers of dietary intake among participants of the NU-AGE study”. In: *Frontiers in Physiology* 9.OCT. ISSN: 1664042X. DOI: 10.3389/fphys.2018.01359.

Pujos-Guillot, Estelle et al. (2019). “Identification of Pre-frailty Sub-Phenotypes in Elderly Using Metabolomics”. In: *Frontiers in Physiology* 10.JAN. ISSN: 1664042X. DOI: 10.3389/fphys.2018.01903.

Giacomoni, Franck et al. (2015). “Workflow4Metabolomics: A collaborative research infrastructure for computational metabolomics”. In: *Bioinformatics* 31.9, pp. 1493–1495. ISSN: 14602059. DOI: 10.1093/bioinformatics/btu813.

Yu, Zhongtang and Mark Morrison (2004). “Improved extraction of PCR-quality community DNA from digesta and fecal samples”. In: *Biotechniques* 36.5, pp. 808–812. DOI: 10.2144/04365ST04.

Barone, Monica et al. (2019). “Gut microbiome response to a modern Paleolithic diet in a Western lifestyle context”. In: *PLoS ONE* 14.8. ISSN: 19326203. DOI: 10.1371/journal.pone.0220619.

Masella, Andre P. et al. (2012). "PANDAseq: Paired-end assembler for illumina sequences". In: *BMC Bioinformatics* 13.1. ISSN: 14712105. DOI: 10.1186/1471-2105-13-31.

Callahan, Benjamin J. et al. (2016). "DADA2: High-resolution sample inference from Illumina amplicon data". In: *Nature Methods* 13.7, pp. 581–583. ISSN: 15487105. DOI: 10.1038/nmeth.3869. URL: https://www.google.com/search?q=High-resolution+sample+inference+from+Illumina+amplicon+data{\&}rlz=1C1GCEA{_}enBR853BR853{\&}oq=High-resolution+sample+inference+from+Illumina+amplicon+data{\&}aqs=chrome..69i57j0l2.511j0j7{\&}sourceid=chrome{\&}ie=UTF-8.

Quast, Christian et al. (2013). "The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools". In: *Nucleic Acids Research* 41.D1. ISSN: 03051048. DOI: 10.1093/nar/gks1219.

Bolyen, Evan et al. (2019). "Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2". In: *Nature biotechnology* 37.8, pp. 852–857.

McMurdie, Paul J and Susan Holmes (2013). "phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data". In: *PloS one* 8.4, e61217.

Philip, Dixon (2003). "VEGAN, a package of R functions for community ecology". In: *Journal of Vegetation Science* 14, pp. 927–930.

Kelly, Brendan J. et al. (2015). "Power and sample-size estimation for microbiome studies using pairwise distances and PERMANOVA". In: *Bioinformatics* 31.15, pp. 2461–2468. ISSN: 14602059. DOI: 10.1093/bioinformatics/btv183.

Lê Cao, Kim Anh et al. (2009). "Sparse canonical methods for biological data integration: Application to a cross-platform study". In: *BMC Bioinformatics* 10. ISSN: 14712105. DOI: 10.1186/1471-2105-10-34.

González, Ignacio et al. (2012). "Visualising associations between paired 'omics' data sets". In: *BioData Mining* 5.1. ISSN: 17560381. DOI: 10.1186/1756-0381-5-19.

Tamura, Motoi et al. (2017). "Quercetin metabolism by fecal microbiota from healthy elderly human subjects". In: *PLoS ONE* 12.11. ISSN: 19326203. DOI: 10.1371/journal.pone.0188271.

Goodrich, Julia K. et al. (2014). "Human genetics shape the gut microbiome". In: *Cell* 159.4, pp. 789–799. ISSN: 10974172. DOI: 10.1016/j.cell.2014.09.053.

Lu, Yuanyuan et al. (2016). "Short chain fatty acids prevent high-fat-diet-induced obesity in mice by regulating g protein-coupled receptors and gut Microbiota". In: *Scientific Reports* 6. ISSN: 20452322. DOI: 10.1038/srep37589.

Kondo, Tomoo et al. (2009). "Acetic acid upregulates the expression of genes for fatty acid oxidation enzymes in liver to suppress body fat accumulation". In: *Journal of Agricultural and Food Chemistry* 57.13, pp. 5982–5986. ISSN: 00218561. DOI: 10.1021/jf900470c.

Oki, Kaihei et al. (2016). "Comprehensive analysis of the fecal microbiota of healthy Japanese adults reveals a new bacterial lineage associated with a phenotype characterized by a high frequency of bowel movements and a lean body type". In: *BMC Microbiology* 16.1, pp. 1–13. ISSN: 14712180. DOI: 10.1186/s12866-016-0898-x.

Just, Sarah et al. (2018). "The gut microbiota drives the impact of bile acids and fat source in diet on mouse metabolism". In: *Microbiome* 6.1. ISSN: 20492618. DOI: 10.1186/s40168-018-0510-8.

Ozato, Naoki et al. (2019). "Blautia genus associated with visceral fat accumulation in adults 20–76 years of age". In: *npj Biofilms and Microbiomes* 5.1. ISSN: 20555008. DOI: 10.1038/s41522-019-0101-x.

Balamurugan, Ramadass et al. (2010). "Quantitative differences in intestinal Faecalibacterium prausnitzii in obese Indian children". In: *British Journal of Nutrition* 103.3, pp. 335–338. ISSN: 00071145. DOI: 10.1017/S0007114509992182.

Del Chierico, Federica et al. (2018). "Gut microbiota markers in obese adolescent and adult patients: Age-dependent differential patterns". In: *Frontiers in Microbiology* 9.JUN. ISSN: 1664302X. DOI: 10.3389/fmicb.2018.01210.

Wutthi-in, Montree et al. (2020). "Gut Microbiota Profiles of Treated Metabolic Syndrome Patients and their Relationship with Metabolic Health". In: *Scientific Reports* 10.1. ISSN: 20452322. DOI: 10.1038/s41598-020-67078-3.

Zhao, Le et al. (2017). "A combination of quercetin and resveratrol reduces obesity in high-fat diet-fed rats by modulation of gut microbiota". In: *Food and Function* 8.12, pp. 4644–4656. ISSN: 2042650X. DOI: 10.1039/c7fo01383c.

Visconti, Alessia et al. (2019). "Interplay between the human gut microbiome and host metabolism". In: *Nature Communications* 10.1. ISSN: 20411723. DOI: 10.1038/s41467-019-12476-z.

Waters, Jillian L. and Ruth E. Ley (2019). "The human gut bacteria Christensenellaceae are widespread, heritable, and associated with health". In: *BMC Biology* 17.1. ISSN: 17417007. DOI: 10.1186/s12915-019-0699-4.

Le Roy, Caroline I. et al. (2019). "Dissecting the role of the gut microbiota and diet on visceral fat mass accumulation". In: *Scientific Reports* 9.1. ISSN: 20452322. DOI: 10.1038/s41598-019-46193-w.

Holeček, Milan (2018). "Branched-chain amino acids in health and disease: Metabolism, alterations in blood plasma, and as supplements". In: *Nutrition and Metabolism* 15.1. ISSN: 1743-7075. DOI: 10.1186/s12986-018-0271-1.

Rietman, Annemarie et al. (2016). "Associations between plasma branched-chain amino acids, β-aminoisobutyric acid and body composition". In: *Journal of Nutritional Science* 5. ISSN: 20486790. DOI: 10.1017/jns.2015.37.

Lackey, Denise E. et al. (2013). "Regulation of adipose branched-chain amino acid catabolism enzyme expression and cross-adipose amino acid flux in human obesity". In: *American Journal of Physiology - Endocrinology and Metabolism* 304.11. ISSN: 01931849. DOI: 10.1152/ajpendo.00630.2012.

Noto, Davide et al. (2016). "Myristic acid is associated to low plasma HDL cholesterol levels in a Mediterranean population and increases HDL catabolism by enhancing HDL particles trapping to cell surface proteoglycans in a liver hepatoma cell model". In: *Atherosclerosis* 246, pp. 50–56. ISSN: 18791484. DOI: 10.1016/j.atherosclerosis.2015.12.036.

Porez, Geoffrey et al. (2012). "Bile acid receptors as targets for the treatment of dyslipidemia and cardiovascular disease". In: *Journal of Lipid Research* 53.9, pp. 1723–1737. ISSN: 00222275. DOI: 10.1194/jlr.R024794.

Hirano, S. and N. Masuda (1982). "Enhancement of the 7α-dehydroxylase activity of a gram-positive intestinal anaerobe by Bacteroides and its significance in the 7-dehydroxylation of ursodeoxycholic acid". In: *Journal of Lipid Research* 23.8, pp. 1152–1158. ISSN: 00222275. DOI: 10.1016/S0022-2275(20)38052-4.

Ishii, Makoto et al. (2014). "Gastrectomy increases the expression of hepatic cytochrome P450 3A by increasing lithocholic acid-producing enteric bacteria in mice". In: *Biological and Pharmaceutical Bulletin* 37.2, pp. 298–305. ISSN: 09186158. DOI: 10.1248/bpb.b13-00824.

Gu, Yanyun et al. (2017). "Analyses of gut microbiota and plasma bile acids enable stratification of patients for antidiabetic treatment". In: *Nature Communications* 8.1. ISSN: 20411723. DOI: 10.1038/s41467-017-01682-2.

Yao, Lina et al. (2018). "A selective gut bacterial bile salt hydrolase alters host metabolism". In: *eLife* 7. ISSN: 2050084X. DOI: 10.7554/eLife.37001.

Alemán, José O. et al. (2018). "Fecal microbiota and bile acid interactions with systemic and adipose tissue metabolism in diet-induced weight loss of obese postmenopausal women". In: *Journal of Translational Medicine* 16.1. ISSN: 14795876. DOI: 10.1186/s12967-018-1619-z.

Sayers, Eric W. et al. (2010). "Database resources of the National Center for Biotechnology Information." In: *Nucleic acids research* 38.Database issue. ISSN: 13624962. DOI: 10.1093/nar/gkp967.

Hansen, Mette Damborg et al. (2020). "Substitutions between potatoes and other vegetables and risk of ischemic stroke". In: *European Journal of Nutrition*, pp. 1–9.

Trichia, Eirini et al. (2020). "The associations of longitudinal changes in consumption of total and types of dairy products and markers of metabolic risk and adiposity: Findings from the European Investigation into Cancer and Nutrition (EPIC)-Norfolk study, United Kingdom". In: *American Journal of Clinical Nutrition* 111.5, pp. 1018–1026. ISSN: 19383207. DOI: 10.1093/ajcn/nqz335.

Franceschi, Claudio et al. (2018). "Inflammaging: a new immune–metabolic viewpoint for age-related diseases". In: *Nature Reviews Endocrinology* 14.10, pp. 576–590. ISSN: 17595037. DOI: 10.1038/s41574-018-0059-4.

Ghosh, Tarini Shankar et al. (2020). "Mediterranean diet intervention alters the gut microbiome in older people reducing frailty and improving health status: The NU-AGE 1-year dietary intervention across five European countries". In: *Gut* 69.7, pp. 1218–1228. ISSN: 14683288. DOI: 10.1136/gutjnl-2019-319654.

Hashiguchi, Tiago Cravo Oliveira et al. (2019). "Resistance proportions for eight priority antibiotic-bacterium combinations in OECD, EU/EEA and G20 countries 2000 to 2030: A modelling study". In: *Eurosurveillance* 24.20. ISSN: 15607917. DOI: 10.2807/1560-7917.ES.2019.24.20.1800445.

Cassini, Alessandro et al. (2019). "Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis". In: *The Lancet Infectious Diseases* 19.1, pp. 56–66. ISSN: 14744457. DOI: 10.1016/S1473-3099(18)30605-4.

OECD (2018). "Stemming the Superbug Tide: Just A Few Dollars More, OECD Health Policy Studies". In: *OECD Publishing, Paris,*

WHO (2014). "Antimicrobial resistance: global report on surveillance 2014". In: pp. 1–257. ISSN: 0019-6061. DOI: 9789241564748.

Hendriksen, Rene S. et al. (2019). "Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage". In: *Nature Communications* 10.1, p. 1124. DOI: 10.1038/s41467-019-08853-3.

Munk, P et al. (2018). "Abundance and diversity of the faecal resistome in slaughter pigs and broilers in nine European countries". In: *Nature Microbiology* 3.8, pp. 898–908. DOI: 10.1038/s41564-018-0192-9.

Clemente, Jose C et al. (2015b). "The microbiome of uncontacted Amerindians". In: *Science advances* 1.3, e1500183.

Allen, Heather K. et al. (2010). "Call of the wild: Antibiotic resistance genes in natural environments". In: *Nature Reviews Microbiology* 8.4, pp. 251–259. ISSN: 17401526. DOI: 10.1038/nrmicro2312.

Forslund, Kristoffer et al. (2014). "Metagenomic insights into the human gut resistome and the forces that shape it". In: *BioEssays* 36.3, pp. 316–329. ISSN: 02659247. DOI: 10.1002/bies.201300143.

Huddleston, Jennifer R. (2014). "Horizontal gene transfer in the human gastrointestinal tract: Potential spread of antibiotic resistance genes". In: *Infection and Drug Resistance* 7, pp. 167–176. ISSN: 11786973. DOI: 10.2147/IDR.S48820.

Escudeiro, Pedro et al. (2019). "Antibiotic resistance gene diversity and virulence gene diversity are correlated in human gut and environmental microbiomes". In: *mSphere* 4.3. DOI: 10.1101/298190.

Francino, M. P. (2016). "Antibiotics and the human gut microbiome: Dysbioses and accumulation of resistances". In: *Frontiers in Microbiology* 6. DOI: 10.3389/fmicb.2015.01543.

Jochum, Lara and Bärbel Stecher (2020). "Label or Concept – What Is a Pathobiont?" In: *Trends in Microbiology* 28.10, pp. 789–792. ISSN: 18784380. DOI: 10.1016/j.tim.2020.04.011.

Gasparrini, Andrew J. et al. (2019). "Persistent metagenomic signatures of early-life hospitalization and antibiotic treatment in the infant gut microbiota and resistome". In: *Nature Microbiology* 4.12, pp. 2285–2297. DOI: 10.1038/s41564-019-0550-2.

D'Amico, Federica et al. (2019). "Gut resistome plasticity in pediatric patients undergoing hematopoietic stem cell transplantation". In: *Scientific Reports* 9.1. ISSN: 20452322. DOI: 10.1038/s41598-019-42222-w.

Rampelli, Simone et al. (2020). "Shotgun Metagenomics of Gut Microbiota in Humans with up to Extreme Longevity and the Increasing Role of Xenobiotic Degradation". In: *mSystems* 5.2. ISSN: 2379-5077. DOI: 10.1128/msystems.00124-20.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel (2014). "Trimmomatic: A flexible trimmer for Illumina sequence data". In: *Bioinformatics*. ISSN: 14602059. DOI: 10.1093/bioinformatics/btu170.

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data.* URL: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Bateman, Alex et al. (2017). "UniProt: The universal protein knowledgebase". In: *Nucleic Acids Research* 45.D1, pp. D158–D169. ISSN: 13624962. DOI: 10.1093/nar/gkw1099.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: http://www.R-project.org/.

Oksanen, Jari et al. (2018). *vegan: Community Ecology Package*. R package version 2.5-3. URL: https://CRAN.R-project.org/package=vegan.

Love, Michael I, Wolfgang Huber, and Simon Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." In: *Genome biology* 15.12, p. 550. ISSN: 1474-760X. DOI: 10.1186/s13059-014-0550-8.

Kolde, Raivo (2019). *pheatmap: Pretty Heatmaps*. R package version 1.0.12. URL: https://CRAN.R-project.org/package=pheatmap.

Harrell Jr, Frank E, with contributions from Charles Dupont, and many others. (2019). *Hmisc: Harrell Miscellaneous*. R package version 4.2-0. URL: https://CRAN.R-project.org/package=Hmisc.

Bastian, Mathieu, Sebastien Heymann, and Mathieu Jacomy (2009). "Gephi: An Open Source Software for Exploring and Manipulating Networks". In: URL: `http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154`.

Schaik, Willem van (2015). "The human gut resistome." In: *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 370.1670, p. 20140087. ISSN: 1471-2970. DOI: `10.1098/rstb.2014.0087`.

Liang, Stephen Y. and Philip A. Mackowiak (2007). "Infections in the Elderly". In: *Clinics in Geriatric Medicine* 23.2, pp. 441–456. ISSN: 07490690. DOI: `10.1016/j.cger.2007.01.010`.

Li, Jing et al. (2018). "Global Survey of Antibiotic Resistance Genes in Air". In: *Environmental Science and Technology* 52.19, pp. 10975–10984. ISSN: 15205851. DOI: `10.1021/acs.est.8b02204`.

Maamar, Sarah Ben et al. (2020). "Mobilizable antibiotic resistance genes are present in dust microbial communities". In: *PLoS Pathogens* 16.1. ISSN: 15537374. DOI: `10.1371/journal.ppat.1008211`.

Hartmann, Erica M. et al. (2016). "Antimicrobial Chemicals Are Associated with Elevated Antibiotic Resistance Genes in the Indoor Dust Microbiome". In: *Environmental Science and Technology* 50.18, pp. 9807–9815. ISSN: 15205851. DOI: `10.1021/acs.est.6b00262`.

Mahnert, Alexander et al. (2019). "Man-made microbial resistances in built environments". In: *Nature Communications* 10.1. ISSN: 20411723. DOI: `10.1038/s41467-019-08864-0`.

Forslund, Kristoffer et al. (2013). "Country-specific antibiotic use practices impact the human gut resistome". In: *Genome Research* 23.7, pp. 1163–1169. ISSN: 10889051. DOI: `10.1101/gr.155465.113`.

Hu, Yongfei et al. (2013). "Metagenome-wide analysis of antibiotic resistance genes in a large cohort of human gut microbiota". In: *Nature Communications* 4. ISSN: 20411723. DOI: `10.1038/ncomms3151`.

Sommer, Morten O.A., Gautam Dantas, and George M. Church (2009). "Functional characterization of the antibiotic resistance reservoir in the human microflora". In: *Science* 325.5944, pp. 1128–1131. ISSN: 00368075. DOI: `10.1126/science.1176950`.